UNIVERSIDADE ESTADUAL DE CAMPINAS

INSTITUTO DE BIOLOGIA

João Paulo Gomes Viana

Genômica da conservação de *Casearia sylvestris* Sw. no Cerrado e Mata Atlântica do Estado de São Paulo

Conservation genomics of *Casearia sylvestris* Sw. in the Brazilian Savannah and Atlantic Forest of the State of São Paulo

CAMPINAS

2017

João Paulo Gomes Viana

Genômica da conservação de *Casearia sylvestris* Sw. no Cerrado e Mata Atlântica do Estado de São Paulo

Conservation genomics of *Casearia sylvestris* Sw. in the Brazilian Savannah and Atlantic Forest of the State of São Paulo

Tese apresentada ao Instituto de Biologia da Universidade Estadual de Campinas como parte dos requisitos exigidos para obtenção do Título de Doutor em Genética e Biologia Molecular na área de Genética Vegetal e Melhoramento.

Thesis presented to the Institute of Biology of the University of Campinas in partial fulfillment of the requirements for the degree of Doctor in Genetics and Molecular Biology in the area Plant Genetics and Genetic Breeding.

ESTE ARQUIVO DIGITAL CORRESPONDE À VERSÃO FINAL DA TESE DEFENDIDA PELO CANDIDATO JOÃO PAULO GOMES VIANA E ORIENTADO PELA DRA. MARIA IMACULADA ZUCCHI.

Orientadora: Drª. Maria Imaculada Zucchi

CAMPINAS

2017

Agência(s) de fomento e nº(s) de processo(s): FAPESP, 2013/05762-6; CNPq, 141039/2013-2; FAPESP, 2011/50296-8; FAPESP, 2016/05765-3; CAPES
ORCID: http://orcid.org/0000-0002-9217-9604

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Biologia
Mara Janaina de Oliveira - CRB 8/6972

V654g

Viana, João Paulo Gomes,1989-
Genômica da conservação de *Casearia sylvestris* Sw. no Cerrado e Mata Atlântica do Estado de São Paulo/ João Paulo Gomes Viana . – Campinas, SP : [s.n.], 2017.

Orientador: Maria Imaculada Zucchi.
Tese (doutorado) – Universidade Estadual de Campinas, Instituto de Biologia.

1. Genética da conservação. I. Zucchi, Maria Imaculada. II. Universidade Estadual de Campinas. Instituto de Biologia. III. Título.

Informações para Biblioteca Digital

Título em outro idioma: Conservation genomics of *Casearia sylvestris* Sw. in the Brazilian Savannah and Atlantic Forest of the State of São Paulo
Palavras-chave em Inglês:
Conservation genetics
Área de concentração: Genética Vegetal e Melhoramento
Titulação: Doutor em Genética e Biologia Molecular
Banca examinadora:
Maria Imaculada Zucchi [Orientador]
Gustavo Maruyama Mori
Karina Martins
Prianda Rios Laborda
Mariana Freitas Nery
Data da defesa: 28-08-2017
Programa de Pós Graduação: Genética e Biologia Molecular

Campinas, 28 de agosto de 2017

Comissão Examinadora

Prof. Dr. Gustavo Maruyama Mori

Profª. Drª. Karina Martins

Drª. Prianda Rios Laborda

Profª. Drª. Mariana Freitas Nery

Drª. Maria Imaculada Zucchi (Orientadora)

Os membros da Comissão Examinadora acima assinaram a Ata de Defesa, que se encontra no processo de vida acadêmica do aluno.

Dedico à minha amada família e amigos.

AGRADECIMENTOS

Agradeço à Deus pelo dom da vida e oportunidade de apreciar o milagre de sua criação;

Agradeço à Universidade Estadual de Campinas (Unicamp) pela excelente formação que recebi no Programa de Pós-Graduação em Genética e Biologia Molecular do Instituto de Biologia desta maravilhosa instituição;

Agradeço muito honrosamente à minha orientadora Drª. Maria Imaculada Zucchi pela oportunidade de crescimento e evolução científica no grupo de pesquisa sob sua coordenação;

Agradeço ao Dr. Marcelo Mattos Cavallari pelas valiosas contribuições científicas desde o delineamento a finalização deste trabalho;

Agradeço à Profª. Drª. Anete Pereira de Souza, Danilo Augusto Sforça e Aline da Costa Lima Moraes pelas importantíssimas contribuições nas etapas experimentais desta pesquisa em especial durante a construção das bibliotecas genômicas e sequenciamento;

Agradeço ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela bolsa de estudos (Proc. Nº 141039/2013-2) concedida nas primeiras etapas desta pesquisa e a a bolsa de produtividade científica que apoiou este estudo (Proc. Nº 310446/2015-5);

Agradeço à Fudação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) pela bolsa de estudos (Proc. Nº 2013/05762-6) e apoio financeiro em todas as etapas de realização e divulgação deste trabalho e pela bolsa de estágio e pesquisa no exterior (BEPE Proc. Nº 2016/05765-3);

Agradeço à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo apoio financeiro (Proc. Nº 23038.010032/2013-14) nas etapas finais desta pesquisa;

Agradeço a banca examinadora pelas maravilhosas contribuições que resultaram no enriquecimento dos estudos presentes neste documento;

Agradeço aos meus pais João Evangelista Viana e Maria Cícera Gomes dos Santos e ao meu irmão Albert Isaac Gomes Viana pelo alicerce dado durante minha caminhada para que eu atingisse meus objetivos;

Agradeço à Nelma Neylanne Pinho Muniz Oliveira pelo companheirismo incondicional e suporte constante na vida pessoal e acadêmica;

Agradeço à Profª. Drª. Regina Lucia Ferreira Gomes e ao Prof. Dr. Sérgio Emílio dos Santos Valente pelas valiosas orientações ante ao ingresso nesta etapa que se conclui na minha vida acadêmica;

Agradeço a meu amigo Prof. Dr. Hendrie Ferreira Nunes (formalização proposital) pela incomensurável parceria e suporte em vários momentos dentro e fora da vida acadêmica;

Agradeço às minhas amigas Ellida de Aguiar Silvestre e Mariana Novello pelas valiosas contribuições no desenvolvimento desta pesquisa bem como no compartilhamento de momentos de alegria fora da vida acadêmica;

Agradeço ao Miklos Maximiliano Bajay, Alessandro Alves-Pereira, Vitor Antônio Correa Pavinato e Patricia Sanae Sujii pelos modelos e inspirações acadêmicas no decorrer deste curso;

Agradeço à Camila Menezes Trindade Macrini, Carolina Grando, Fabiano Lucas Araujo, Fabricio J. B. Francischini, Gustavo M. Mori, Jaqueline Bueno de Campos, Kaiser Dias Schwarcz, Marcos V. B. M. Siqueira e todos os outros membros do grupo de Genética e Genômica da Conservação pela amizade, convívio e contribuições científicas;

Agradeço ao Prof. Dr. José Baldin Pinheiro pelos valiosos conselhos e colaboração;

Agradeço à Ana Paula Mendes-Silva, Camila Campêlo de Sousa, Diane Simon Rozzetto, Eleonora Zambrano Blanco, Fabiana Freitas Moreira, Fabiani da Rocha, Felipe Bermudez Pereira, Júlia Morosini, Kênia Carvalho de Oliveira, Maísa Curtolo, Mônica Christina Ferreira, Vanessa Rizzi e todos os outros membros do laboratório de Diversidade Genética e Melhoramento pelos momentos de amizade e aprendizado mútuo;

Agradeço ao José Ribamar de Assunção Filho e a Josilane Souza da Penha pelo suporte e amizade;

Agradeço à coordenação de todas as unidades de conservação que nos apoiaram na realização desta maravilhosa pesquisa;

Agradeço a todos que contribuíram diretamente e indiretamente na realização deste trabalho.

RESUMO

VIANA, J. P. G. Genômica da conservação de *Casearia sylvestris* Sw. no Cerrado e Mata Atlântica do Estado de São Paulo. **Tese de doutorado**. Universidade Estadual de Campinas, Campinas – SP, 2017.

As restaurações florestais e o estabelecimento de unidades de conservação têm sido os principais meios para a conservação dos biomas Cerrado e Mata Atlântica no Brasil mas pouca atenção tem sido dada à diversidade genética nas populações de espécies vegetais destas áreas. Por isso, realizou-se dois estudos com os objetivos de: 1) Avaliar a diversidade e estruturação genética de *Casearia sylvestris* Sw. em restaurações florestais e em remanescentes naturais de Mata Atlântica; 2) Estudar a diversidade e estruturação genética de populações naturais de *C. sylvestris* do Cerrado e Mata Atlântica do Brasil e verificar a ocorrência de seleção positiva nestas populações. Construiram-se e foram sequenciadas bibliotecas genômicas da espécie *C. sylvestris* Sw. utilizando as técnicas de genotyping-by-sequencing (GBS) e double-digested restriction associated DNA (ddRADseq). Calcularam-se estimativas de diversidade e diferenciação genética para todas as populações (número total de alelos, heterozigosidades, número de alelos privados, coeficientes de endogamia e índices de fixação) e o estudo de estruturação genética foi realizado através da análise discriminante de componentes principais (DAPC). Não houveram diferenças significativas entre as estimativas de diversidade e diferenciação genética calculadas no primeiro estudo indicando que as restaurações florestais têm oferecido condições de conservação da diversidade genética. No segundo estudo foi observado que as populações de *C. sylvestris* do Cerrado e Mata Atlântica estão geneticamente estruturadas. Além disso, relatou-se ineditamente a ocorrência de seleção positiva nas populações destes biomas. Conclui-se com estes dois estudos que o modelo o qual as restaurações florestais foram implementadas no Brasil tem possibilitado a conservação da diversidade genética das populações e que as diferenças nas características ambientais entre o Cerrado e Mata Atlântica podem levar a ocorrência de eventos microevolutivos que levam a adaptação local das populações a estes biomas.

Palavras-chave: Genética da Conservação, ddRADseq, GBS, Adaptação Local, Genética da Restauração

ABSTRACT

VIANA, J. P. G. Conservation genomics of *Casearia sylvestris* Sw. in Cerrado and Atlantic Forest from State of São Paulo. **PhD thesis**. Universidade Estadual de Campinas, Campinas - SP, 2017.

Forest restorations and the establishment of conservation units have been the main ways for the conservation of Brazilian Savannah (Cerrado) and Atlantic Forest biomes in Brazil, but little attention has been given to genetic diversity in the populations of plant species of these areas. Therefore, two studies were carried out with the objectives of: 1) To evaluate the genetic diversity and structure of *Casearia sylvestris* Sw. In forest restorations and in natural remnants of Atlantic Forest; 2) To study the genetic diversity and structure of natural populations of *C. sylvestris* from the Cerrado and Atlantic Forest of Brazil and verify the occurrence of positive selection in these populations. Genomic libraries of *C. sylvestris* were built and sequenced using genotyping-by-sequencing (GBS) and double-digested restriction associated DNA (ddRADseq) techniques. Estimates of genetic diversity and genetic differentiation were calculated for all populations (total number of alleles, heterozygosities, number of private alleles, coefficients of inbreeding and fixation indices) and the genetic structuring study was performed through the discriminant analysis of components (DAPC). There were no significant differences between the estimates of diversity and genetic differentiation calculated in the first study, indicating that forest restorations have offered conservation conditions of genetic diversity. In the second study it was observed that the populations of *C. sylvestris* from Cerrado and Atlantic Forest are genetically structured. In addition, the occurrence of positive selection in the populations of these biomes was reported. It is concluded with these two studies that the model that the forest restorations were implemented in Brazil has made possible the conservation of the genetic diversity of the populations and that the differences in the environmental characteristics between the Cerrado and Atlantic Forest can lead to the occurrence of microevolutionary events that lead local adaptation of populations to these biomes.

Keywords: Conservation Genetics, ddRADseq, GBS, Local Adaptation, Genetics of Restoration

SUMÁRIO

INTRODUÇÃO GERAL

O processo de degradação ambiental, que consiste na perda da capacidade do meio ambiente de produzir recursos naturais essenciais para sustentação da vida, na maioria dos casos é precedido pela perda ou constantes agressões à cobertura vegetal natural que por sua vez tem como principais causas a expansão das fronteiras agrícolas e áreas urbanas (Pagiola et al., 2007; Tabarelli et al., 2010), aumento do processo de industrialização (Butler e Laurance, 2008) e desmatamento. Além da perda da biodiversidade e contribuição para o aumento da emissão de gases relacionados ao efeito estufa, a degradação ambiental nos priva de vários recursos não madeireiros como mananciais de qualidade, possibilidade de manejo agroflorestal, ecoturismo, dentre outros (Donohoe, 2003; Lamb, Erskine e Parrotta, 2005; Wright, 2010).

Como em muitos países, a degradação ambiental, principalmente por meio do desmatamento, tem sido um dos principais problemas ambientais no Brasil. Conhecido por deter um alto nível de diversidade biológica, apresentando a maior biodiversidade terrestre do planeta (MMA, 2017a), uma decorrência dos diferentes biomas e ecossistemas que se encontram no país, o Brasil possui duas das 25 áreas críticas para conservação mundial pelo alto grau de endemismo florístico: o Cerrado e a Mata Atlântica (Myers et al., 2000).

Antes do processo de colonização do Brasil que se intensificou a partir do século XVI, a Mata Atlântica ocupava uma área no país de cerca de 1.315.460 km^2 e se estendia ao longo do território que atualmente corresponde a 17 estados da federação (SOS Mata Atlântica e INPE 2015). Naquele período, as ações antrópicas sobre a vegetação da Mata Atlântica eram realizadas em baixa intensidade pelos povos indígenas que viviam nas proximidades do litoral brasileiro. Estas ações antrópicas se intensificaram com a chegada dos portugueses que inicialmente viram no extrativismo do Pau-Brasil (Cardoso et al. 1998; Rocha, Presotto e Cavalheiro 2007), espécie arbórea nativa da Mata Atlântica, a oportunidade de se realizar na então colônia portuguesa a primeira de uma série de atividades humanas que culminaram na degradação da cobertura vegetal original da Mata Atlântica (SOS Mata Atlântica, 2017). Após séculos de exploração através do extrativismo, da expansão das fronteiras agrícolas, da caça, da industrialização e da urbanização

aproximadamente 8% de sua área original permanece preservada (SOS Mata Atlântica e INPE, 2015). Ao incluirmos florestas secundárias e pequenos fragmentos (<100 ha), esta percentagem aumenta para aproximadamente 11% (Ribeiro et al., 2009).

A Mata Atlântica (Figura 1A) é reconhecida por sua grande biodiversidade, estima-se que contenha cerca de 1 a 8% das espécies do mundo (Myers et al., 2000), incluindo mais de 8000 espécies endêmicas (Tabarelli et al., 2003). Esta riqueza é reflexo dos inúmeros gradientes ambientais em que este bioma é encontrado já que considerando apenas o Brasil, o complexo de ecossistemas que constitui a Mata Atlântica pode ser encontrado ao longo de 30 graus de latitude e desde o nível do mar até altitudes superiores a 2000 metros (Joly et al., 1999; SOS Mata Atlântica, 2017). Devido esta diversidade de ecossistemas, do ponto de vista ecológico não se costuma a se referir à Mata Atlântica como um bioma único, por outro lado, é mais adequado referencia-la como um domínio de natureza formado por diferentes fitofisionomias (Ab'Saber, 2007). Entre estas fitofisionomias destacam-se a Floresta Ombrófila (Densa, Aberta e Mista), Floresta Estacional (Semidecidual e Decidual), Vegetação de Mangue, Restingas e Campos de Altitude (Joly et al., 1999).

O Cerrado (Figura 1B) constitui o segundo maior bioma do Brasil pois com uma área de 2.036.448 km^2 ocupa cerca de 22% de seu território (Medeiros e Young, 2011; MMA, 2017b). Diferente da Mata Atlântica, o processo de degradação ambiental das áreas de Cerrado se iniciou de forma mais tardia com a prática de atividades mineradoras, seguido mais recentemente pela expansão das fronteiras agrícolas e utilização de recursos madeireiros. Estas atividades fizeram do Cerrado o segundo domínio brasileiro que mais sofreu alterações devido a interferência humana (Ribeiro et al., 2005). Estima-se que 57% da vegetação original foi completamente destruída e a metade das áreas remanescentes esteja bastante alterada (Machado et al., 2004). Além disso, ao serem considerados apenas a lista de *hotspots* de biodiversidade, o Cerrado é o domínio que apresenta o menor percentual de áreas sob proteção integral, contando com apenas 8,21% de sua abrangência em áreas protegidas, sendo menos de 3% em áreas de proteção integral (IPEA, 2011).

O conhecimento sobre a origem biogeográfica do Cerrado evidencia a importância da conservação destas áreas. O alto grau de endemismo florístico no Cerrado é associado à alta incidência de incêndios. Durante sua formação, espécies capazes de resistir a incêndios recorrentes foram selecionadas (Simon et al., 2009).

Além disso, assim como na Mata Atlântica, devido ao vasto gradiente ambiental, o Cerrado apresenta diferentes fitofisionomias com formações florestais, savânicas e campestres que juntas possuem rica biodiversidade. Sua flora é composta por mais de 11.600 espécies (Mendonça et al., 2008; MMA, 2017b) das quais estima-se que 44% sejam endêmicas (Klink e Machado, 2005).

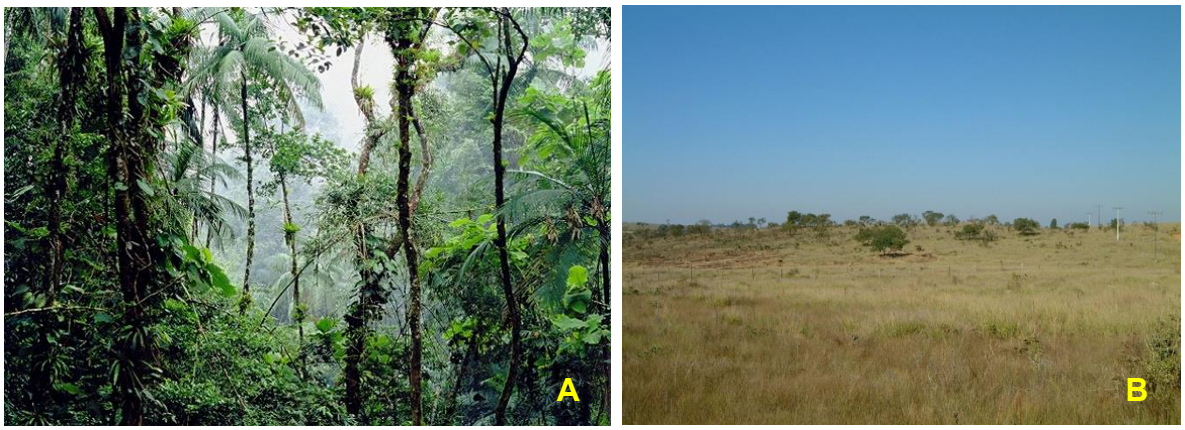

Figura 1 – Aspectos morfológicos de fitofisionomias dos domínios de Mata Atlântica (A) e Cerrado (B). Estes domínios diferem em várias características bióticas e abióticas, criando condições para existência de diferentes pressões seletivas. Fotos: Instituto Florestal do Estado de São Paulo, 2017.

A proteção dos remanescentes florestais no Brasil, incluindo os ameaçados biomas do Cerrado e Mata Atlântica, é regida por um conjunto de leis que estipulam normas para preservação e manejo destas áreas em propriedades públicas e privadas. Dentre estas áreas, as unidades de conservação (UCs) correspondem a ambientes delimitados, contendo as riquezas do meio natural, incluindo propriedades ambientais importantes, as quais possuem a finalidade de garantir a constituição de porções expressivas e ecologicamente duradouras dos distintos demes, conservando a riqueza biológica presente (MMA, 2017c). No entanto, as UCs que representam os biomas Cerrado e Mata Atlântica no Brasil são insuficientes para atender a estas metas de proteção pois não representam uma parcela satisfatória da área ocupada pelo bioma original (IPEA, 2011). Além disso, deve-se ressaltar que embora estas estratégias sejam fundamentais para conservação da biodiversidade, o modelo de implementação de UCs, dissociadas de uma perspectiva mais abrangente da paisagem tal como se fossem ilhas isoladas meio a áreas urbanas e de

desmatamento, pode prejudicar a resiliência em longo prazo destas áreas causando perda das funções ecológicas e consequentemente de biodiversidade (Delelis, Rehder e Cardoso, 2010; Ganem, 2010).

Para sanar as consequências da degradação ambiental em muitos países tem sido adotada a restauração florestal (Chazdon, 2008; Rodrigues et al., 2009) que além de possibilitar reverter os efeitos do desmatamento pela recuperação quase completa da forma e função da floresta original, oferece alternativas para o manejo destas áreas de forma sustentável. Além disso, as áreas de restauração florestal podem favorecer a conectividade entre os fragmentos florestais possibilitando a manutenção do fluxo gênico entre as diferentes populações de uma espécie e reduzindo possíveis efeitos deletérios intensificados pela endogamia ou deriva genética.

No entanto, muitos projetos de restauração florestal no Brasil se iniciaram muito antes do reconhecimento da importância da diversidade genética para sobrevivência em longo prazo das populações (Rodrigues et al., 2009). A principal meta ecológica tem sido a recuperação da estrutura florestal e da diversidade de espécies arbóreas, com reduzido foco na diversidade genética. Recentemente, muitos estudos têm apontado a importância da diversidade genética para o sucesso de uma restauração florestal (Ritchie e Krauss, 2012; Kettenring et al., 2014; Thomas et al., 2014; Mijangos et al., 2015). No entanto, poucos destes estudos têm mensurado a diversidade genética em áreas de restauração florestal o que tem limitado o conhecimento sobre a genética da restauração.

Do ponto de vista genético, muitos estudos têm chamado atenção para as consequências da fragmentação e isolamento destas populações. Entre os principais temas abordados nestes estudos, destacam-se os efeitos da ausência de fluxo gênico entre populações de diferentes fragmentos, endogamia e deriva genética em decorrência do reduzido tamanho populacional (Ezard e Travis, 2006; Jump e Penuelas, 2006; Yuan et al., 2012; Mona et al., 2014; Christie e Knowles, 2015; Peñaloza-Ramírez et al., 2016). Tais fenômenos podem levar à redução da diversidade genética e consequente redução do *fitness* destas populações (Allard, Jain e Workman, 1968; Hughes et al., 2008; Star e Spencer, 2013; Balick et al., 2015). Além disso, a diversidade genética é a matéria-prima para que ocorra a seleção natural, na ausência desta diversidade, diante do atual cenário de mudanças climáticas globais, as populações perdem o potencial de adaptação a estas mudanças

levando a um comprometimento da sobrevivência destas populações em longo prazo (Fischer e Matthies, 1998; Buza, Young, e Thrall, 2000; Wisely et al., 2002; Reed e Frankham, 2003; Willi et al., 2007; Werf et al., 2009; Markert et al., 2010). Entretanto, outro importante fato é em relação ao aspecto genético da fragmentação dos *habitats*, mas que até o momento tem recebido pouca atenção é o efeito desta fragmentação na direção e intensidade da seleção natural (Willi, Buskirk e Hoffmann, 2006). Devido à limitada capacidade de migração para novos ambientes, deve-se dar maior atenção à ocorrência destes fenômenos em populações de espécies arbóreas.

Atualmente, a grande maioria dos estudos genéticos em populações naturais têm sido desenvolvidos com uso de poucos marcadores moleculares, geralmente acessando apenas a variação genética neutra. O estudo da variação neutra elucida como fenômenos tais como a deriva genética e fluxo gênico afetam as frequências alélicas no genoma como um todo. No entanto, para entender em que intensidade e direção a seleção pode agir na diferenciação das populações é necessário acessar a variação genética adaptativa que é específica de locos; esta pode ser acessada por marcadores com ampla cobertura do genoma como AFLP e SNPs (Allendorf, Hohenlohe e Luikart, 2010; Kirk e Freeland, 2011; Funk et al. 2012; Andrews et al., 2016).

Com o desenvolvimento das técnicas de DNA associadas a sítios de restrição (do inglês, RAD) e do sequenciamento em larga escala, que permitem a descoberta de grandes quantidades de polimorfismos de nucleotídeo único (do inglês, SNPs), tem sido possível a condução de estudos em escala genômica em populações de espécies não modelo (Ekblom e Galindo, 2010). Os SNPs têm o potencial para um escaneamento em larga escala do genoma para descoberta tanto da variação neutra quanto da variação adaptativa. Porém, mesmo que as técnicas baseadas em RAD possam ser aplicadas a qualquer espécie não modelo, casos de estudo ainda são necessários (Xu et al., 2014), especialmente na ausência de genomas de referência, um fato que ocorre com a maioria das espécies nativas do Cerrado e Mata Atlântica.

As questões apresentadas até o momento foram abordadas detalhadamente neste trabalho em dois capítulos. O estudo abordado no primeiro capítulo teve o objetivo avaliar a diversidade e estrutura genética em populações de *Casearia sylvestris*, uma espécie arbórea nativa de ocorrência natural em várias regiões do continente americano, em restaurações florestais e remanescentes naturais de Mata Atlântica com intuito de avaliar a qualidade das restaurações

florestais em nível de diversidade genética. Já o estudo abordado no segundo capítulo teve o objetivo de encontrar sinais de adaptação em populações de *C. sylvestris* com ocorrência no Cerrado e Mata Atlântica a estes respectivos biomas com a finalidade de relatar a ocorrência de seleção positiva nestes ambientes.

Capítulo 1 - Genetic diversity is similar between Atlantic Forest restorations and natural remnants for the native tree *Casearia sylvestris* Sw.

ABSTRACT

The primary focus of tropical forest restoration has been the recovery of forest structure and tree taxonomic diversity, with limited attention given to genetic conservation. Populations reintroduced through restoration plantings may have low genetic diversity and be genetically structured due to founder effects and genetic drift, which limit the potential of restoration to recover ecologically resilient plant communities. Here, we studied the genetic diversity and genetic differentiation using single nucleotide polymorphisms (SNPs) between restored populations (CO and IR) and natural populations (SG and TT) of the native tree *Casearia sylvestris* in the Atlantic Forest of Brazil. We sampled leaves from approximately 24 adult individuals in each of the study sites: two restoration plantations (27 and 62 years old) and two natural forest remnants. We prepared and sequenced a genotyping-by-sequencing library, SNP markers were identified *de novo* using *Stacks* pipeline, and genetic parameters and structure analyses were then estimated for populations. Genetic diversity was similar between restored (CO: A_R = 1.72, H_O = 0.13, H_E = 0.17, F_{IS} = 0.19; IR: A_R = 1.73, H_O = 0.14, H_E = 0.17, F_{IS} = 0.15) and natural populations (SG: A_R = 1.73, H_O = 0.14, H_E = 0.17, F_{IS} = 0.14; TT: A_R = 1.72, H_O = 0.13, H_E = 0.16, F_{IS} = 0.17), which were not structured (F_{ST} = 0) among population but they were structured within populations. About 99% of the alleles were shared with at least two populations. Despite of the favorable genetic diversity for conservation, the studied populations had high F_{IS} values which might be related to the high level of forest fragmentation. In general, contrary to our expectations that expected low genetic diversity in restoration plantations, they appear then be effective for conserving tree genomic diversity in human-modified tropical landscapes, but more efforts might be necessary to reduce the fragmentation level of those forests. Furthermore, we demonstrate that genotyping-by-sequencing can be an useful tool in restoration genetics.

Keywords: Conservation genetics, genotyping-by-sequencing, forest fragments, gene flow, genetic differentiation, native trees, restoration genetics

1.1 INTRODUCTION

Increasing environmental degradation has led to the promotion of ecological restoration worldwide, particularly for forest ecosystems (Chazdon, 2008; Rodrigues et al., 2009; Holl, 2017). Forest restoration methods are selected based upon the degree of ecosystem degradation (Holl and Aide, 2011; Lamb, 2011), the availability of native forest fragments to act as seed sources in the landscape (Letcher and Chazdon, 2009; De Rezende et al., 2015) the goals of restoration interventions (Lamb, 2011), and funding constraints (Brancalion et al., 2012; Kimball et al., 2015). For all restoration approaches, forest succession must proceed in time, which is the primary ecological process controlling the recovery of forest ecosystems (Arroyo-Rodriguez et al., 2017). However, several factors drive tropical forest succession. Although some of these factors are unpredictable and difficult to manipulate at the site level, such as dispersal by fauna, others can be controlled at some level during restoration planning, implementation, or adaptive management; these are the factors that receive special attention from practitioners. Genetic diversity of reintroduced populations is a key factor for the long-term persistence of native species in forest restoration sites that can be manipulated by practitioners, based on the selection of natural populations and seed-trees for producing planting stocks (Brancalion et al., 2012; Thomas et al., 2014; Mijangos et al., 2015).

Genetic diversity is defined as the variety of alleles and genotypes in a population, which drives morphological, physiological and behavioral differences in individuals and populations (Frankham, Ballou and Briscoe, 2010). It is important for the long-term viability of populations because of the influence on adaptive potential, inbreeding, and genetic drift (Reed and Frankham, 2003; Bell and Collins, 2008; Kahilainen, Puurtinen and Kotiaho, 2014). Moreover, the risk of extinction of natural populations increases at reduced genetic diversity (Kahilainen, Puurtinen and Kotiaho, 2014). One of the causes of increased vulnerability is the limitation of evolutionary potential to respond to environmental changes (Toro and Caballero, 2005), a critical issue in the Anthropocene (Ganzhorn et al., 2015). Tree species are highly vulnerable to reduced genetic diversity in founder populations, because of the lack of mobility and susceptibility to significant environmental changes within a lifetime (Schaberg et al.,

2008). Despite the numerous conceptual arguments for considering genetic diversity in biodiversity conservation, a large gap occurs between theory and practice.

Forest restoration projects, for example, began long before the recognition of the importance of genetic diversity for their long-term sustainability (Rodrigues et al., 2009). The primary ecological goals have been the recovery of forest structure and woody species diversity, with less focus on genetic diversity. More recently, many studies have addressed the importance of genetic diversity for restoration success (Ritchie et al., 2012; Kettenring et al., 2014; Thomas et al., 2014; Mijangos et al., 2015). However, few studies have measured genetic diversity in restored areas, which limits our understanding of the genetics of restoration.

Forest restoration plantings are often implemented using seedlings from few mother trees, and the level of genetic diversity and kinship among planted seedlings has received little attention (Burgarella et al., 2007; Kettle et al., 2008; Liu et al., 2012). Populations reintroduced through restoration plantings may have low genetic diversity and be genetically structured due to founder effects and genetic drift, which limit the potential of restoration to recover ecologically resilient plant communities. These limitations are a concern because inbreeding can lead to reduced fertility and individual survival in some species. However, natural remnants located near forest restoration areas can conserve genetic diversity (Otálora et al., 2011) and supply a diverse set of alleles for the population undergoing restoration, acting as stepping stones that promote genetic connectivity among forest fragments across human-modified landscapes (Leidner and Haddad, 2011; Lloyd and Marsden, 2011). Consequently, the expectation is that populations established through restoration plantations will have low genetic diversity and be structured due to founder effects, as observed for regenerating palm populations in second-growth tropical forests (Sezen, 2005).

The study of genetic diversity in natural populations can be performed using molecular markers, such as single nucleotide polymorphisms (SNPs). The advent of high throughput sequencing of short-reads, which can assess large-scale sets of SNPs, has eased the study of population genetics in a genomic scale using populations of non-model species (Etter and Johnson, 2012), such as tropical trees. Additionally, SNPs offer the potential for large-scale scanning of the genome by providing a large number of markers that improve the accuracy of population genetic estimates (Luikart et al., 2003). Among the methodologies used for SNP discovery, Elshire et al. (2011) proposed a method that uses restriction enzymes to reduce the complexity of the

genome. This genotyping-by-sequencing (GBS) approach is a low cost, highly specific, and reproducible method to detect SNPs and can be applied in non-model species.

In this context, to understand the current situation of forest restoration plantations and natural remnants, the objective of this study was to assess the genetic diversity and genetic structure of populations of the native tree *C. sylvestris* in these types of fragments of the Brazilian Atlantic Forest to verify if there are differences in genetic diversity level between these types of forest fragments and the potential of forest restoration plantations in recovering genetic diversity.

1.2 MATERIALS AND METHODS

1.2.1 Study species

Casearia sylvestris Sw. (Saliaceae) is a pioneer tree species highly abundant in secondary forests (Tabarelli, Villani and Mantovani, 1993; Silva et al., 2003; Marmontel et al., 2013; Dias et al., 1998), distributed from Mexico to Argentina (Sleumer, 1980). The mating system has not been studied using molecular markers approach yet, but the flowers are hermaphroditic, and the species is self-compatible (Ramalho, 2004; Barbosa, 1997). Similar to other species of this genus, *C. sylvestris* is usually pollinated by various species of flies, whereas seed dispersal occurs by small birds over short distances (Wanderley, 2002).

1.2.2 Study areas and sampling

We performed the study in the Brazilian Atlantic Forest, a global hot spot for biodiversity conservation that is estimated to contain from 1% to 8% of the world's terrestrial species (Myers et al., 2000) and more than 8000 endemic species (Tabarelli et al., 2003). However, less than 8% of the original area remains within continuous forests (SOS Mata Atlantica and INPE, 2015) and only 11-16% when secondary forests and small fragments (<100 ha) are considered (Ribeiro et al., 2009). We selected two natural remnants (located in the Campinas and Tietê municipalities) and two forest restoration plantations (located in the Cosmópolis and Iracemápolis municipalities) in São Paulo State, southeastern Brazil, for this study (Figure 1.1). The natural remnant in Campinas (Mata de Santa Genebra, hereafter SG, 22º49.36’ S, 47º6.60’ W) is 252-ha and is embedded within one of the largest urban centers in Brazil, in a landscape dominated by pastures, sugarcane plantations, and small-scale natural remnants. Tietê (hereafter TT, 23º0.11’ S, 47º43.66’ W) is a 70-ha forest fragment surrounded by extensive pastures, similar to most natural remnants of the Brazilian Atlantic Forest. Cosmópolis (hereafter CO, 22º40.30’ S, 47º12.25’ W) is a 25-ha forest restoration plantation implemented between 1955 and 1960 in the riparian buffer of the Jaguari River in a landscape matrix dominated by sugarcane plantations, with few natural fragments remaining. The Iracemápolis (hereafter IR, 22º34.60 S, 47º30.18’ W) study

area is a 50-ha forest restoration plantation established between 1988 and 1990 at the border of a water reservoir that supplies drinking water to the city of Iracemápolis (Brancalion et al., 2012). All four forests had a high abundance of *C. sylvestris* individuals, particularly near the borders and in the understory. The CO population was established with nursery-grown seedlings produced with seeds harvested in an old landscaping planting in the "Luiz de Queiroz" College of Agriculture of the University of São Paulo in Piracicaba – SP, whereas the IR population was established with nursery-grown seedlings bought from a forest nursery in Campinas – SP and another in São Paulo - SP. Leaves from approximately 24 adult individuals, separated by approximately 12 m, were collected from each of the four populations. Using GPS equipment, a minimum distance of 12 m was established between individuals, with an error of a maximum of 4 m. In the forest restorations, we tried to select individuals of the same generation but as *C. sylvestris* is a rapid growing plant species, both planted (i.e., adult individuals growing in the original planting lines) and regenerating individuals (i.e., adult individuals growing in between the original planting lines) may have been sampled.

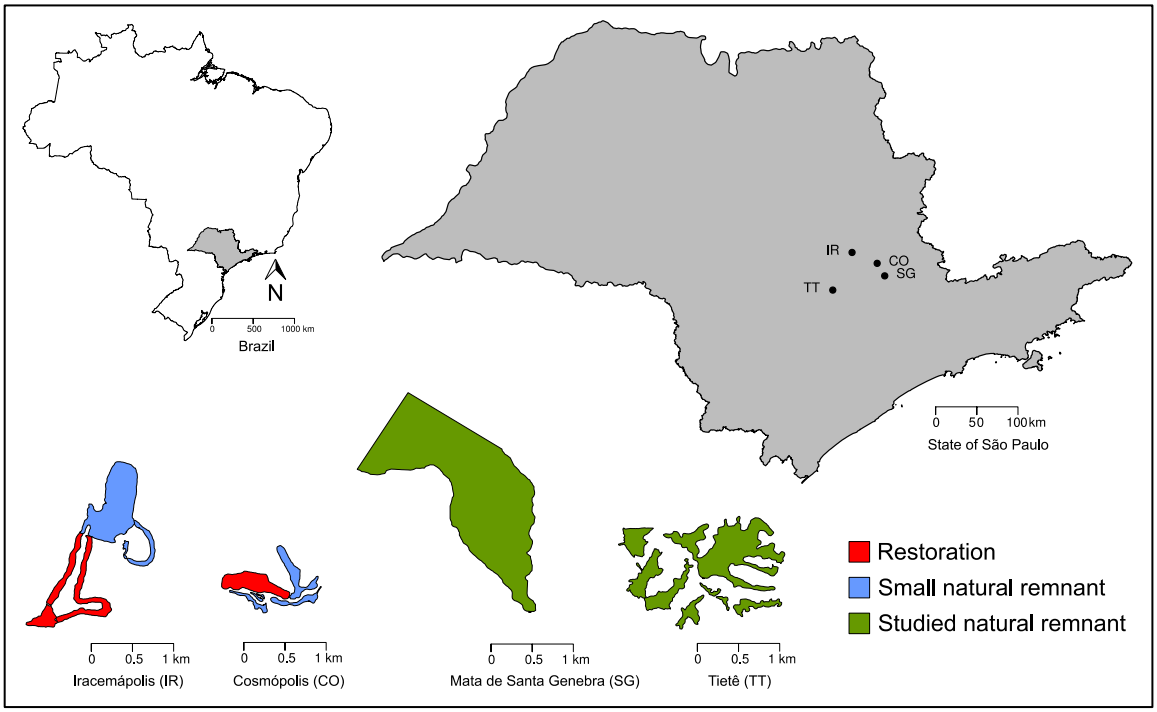

Figure 1.1– Distribution and relative size of Atlantic Forest fragments used in the study of genomic diversity and genomic structure of *Casearia sylvestris* populations of natural remnants and forest restorations.

1.2.3 SNP discovery and data processing

1.2.3.1 Library preparation and sequencing

DNA was extracted from leaf tissue samples of individuals following the method described by Doyle and Doyle (1987). To increase the chances of obtaining the same proportion of reads per individual, the samples with high quality DNA were diluted to the same concentration based on electrophoresis in agarose 1% gel. Preliminary tests with different restriction enzymes were performed. PstI digestion showed the highest concentration of fragments with the greatest distribution in size (data not shown). The library was then prepared according to the method described by Elshire et al. (2011), using barcodes with at least three nucleotides of difference to identify each individual. Finally, single-end sequencing of the 95 samples was performed in a single lane of an Illumina HiSeq 2000 (Illumina, San Diego, CA) producing sets of 100bp reads.

1.2.3.2 Data processing

Sequences of each sample were demultiplexed and cleaned using the *process_radtags* program from the *Stacks* package v. 1.44 (Catchen et al., 2013). A quality score filter (-q, *process_radtags*) was applied across each sample to retain only high-quality sequencing reads. Reads with any non-called base were discarded (-c, *process_radtags*). Barcodes were recovered when they had a maximum of one mismatch (-r, *process_radtags*). Because a reference genome is not available for *C. sylvestris*, the loci were built *de novo*. Sequences from each sample were aligned using a minimum depth of 3 and allowing up to 2 mismatches with the *ustacks* program (-m 3, -M 2, *ustacks*). Subsequently, a catalog of loci was built from subsets of samples, allowing 2 mismatches (-n 2, *cstacks*) between loci from distinct samples. All samples were matched against the catalog, using *sstacks*, to determine the set of loci in each. We then implemented a correction step with *rxstacks*, discarding loci when the mean log likelihood of the locus in the metapopulation was less than -10; confounded loci were also removed. Using the populations program, a filter was applied across loci and individuals to retain loci with a maximum of 20% missing data.

To separate the outlier loci in the set of SNPs, we used an F_{ST}-based approach to assess population pairwise comparisons using the infinite alleles model and a false discovery rate (FDR) of 0.05, as implemented in the *Lositan* software (Antao et al., 2008). Subsequently, neutral candidate loci were separated from outlier loci. We identified candidate loci for positive selection ($P > 0.995$) and candidate loci for balancing selection ($P < 0.005$) separating them of neutral candidate loci. In subsequent steps, only the set of neutral candidate loci (hereafter neutral loci) was used for the population analyses.

1.2.4 Data analysis

1.2.4.1 Genetic diversity

To compare the genetic diversity in forest restoration populations and natural remnants, we used the following population parameters: number of different alleles (A), number of private alleles (P), observed heterozygosity (H_O), expected heterozygosity (H_E), and the coefficient of inbreeding (F_{IS}). We tested for significant differences in heterozygosities and F_{IS} using confidence intervals calculated based on 1000 bootstrap resamples. Estimates of population genetic parameters were calculated using the *PopGenKit* (Paquette, 2012) and *hierfstat* (Goudet and Jombart, 2015) packages. Venn diagrams were generated using the *VennDiagram* package (Chen, 2016). All packages were developed for use in the *R* software (R Core Team, 2017).

1.2.4.1 Genomic differentiation and structure

To understand the partitioning of genetic diversity, we calculated overall and pairwise F-statistics for neutral loci and candidate loci for positive selection using the diveRsity package (Keenan et al., 2013). The significance of these estimates was determined using 95% confidence intervals based on 1000 bootstrap resamples across individuals. A Discriminant Analysis of Principal Components *(DAPC)* described by Jombart et al., (2008) was conducted over the set of neutral loci using Bayesian Information Criterion (BIC) to explore the genetic structure in these populations using

adegenet (Jombart, 2008; Jombart and Ahmed, 2011). We selected 26 principal components in DAPC retaining 56% of total variation.

The spatial genetic structure (SGS) analysis was performed over neutral loci based on the kinship coefficient proposed by Loiselle et al. (1995). This kinship coefficient is robust and unbiased by the frequency of rare alleles. The number and distance intervals of classes were defined to suit each population. To summarize the spatial genetic structure, we used the *Sp-statistic* proposed by Vekemans and Hardy (2004), which is calculated as $S_p = -\beta_{log} / 1 - F_1$; where β_{log} is the regression slope of the kinship coefficient on the log of geographic distance, and F_1 is the mean of the kinship coefficient between individuals estimated for the first distance class. For each distance class, the 95% confidence interval was obtained based on 1000 permutations. For all analyses of SGS within the populations, we used the *SPAGeDi* program developed by Hardy and Vekemans (2002), whereas the spatial autocorrelation graphs were plotted with the *R* software (R Core Team, 2017).

1.3 RESULTS

1.3.1 SNP discovery and data processing

As result of the sequencing we obtained 317,802,888 reads. After cleaning, demultiplexing and filtering of missing data, 861 high-quality biallelic SNPs were identified, which were used to characterize the genetic diversity of 80 individuals distributed across the four populations. For these results, the mean sequencing depth per individual ranged from 1.25 to 43.40, whereas average depth per locus ranged from 4.70 to 83.22 (SD = 8.58). After all filtering steps it was possible to retain 80 individuals for the subsequent analyses. Using the F_{ST} approach in pairwise population comparisons to detect outlier loci, we identified 662 neutral candidate loci, 64 candidate loci for positive selection, and 135 candidate loci for balancing selection.

1.3.2 Genetic diversity

We found no significant differences in observed and expected heterozygosities and coefficient of inbreeding between the natural remnants and forest restoration populations (Table 1.1).

Table 1.1 – Genetic diversity parameters for *Casearia sylvestris* populations in natural remnants and restoration plantations in the Atlantic Forest of southeastern Brazil.

Pop	N	A	P	H_O	H_E	F_{IS}
CO	22	1217	2	0.13 (0.13 – 0.14)	0.17 (0.16 – 0.17)	0.19 (0.15 – 0.23)
IR	18	1213	1	0.14 (0.13 – 0.16)	0.17 (0.15 – 0.17)	0.15 (0.10 – 0.19)
SG	21	1223	4	0.14 (0.13 - 0.15)	0.17 (0.15 – 0.17)	0.14 (0.10 – 0.18)
TT	19	1216	2	0.13 (0.12 – 0.15)	0.16 (0.15 – 0.17)	0.17 (0.12 – 0.22)

CO – Forest restoration in Cosmópolis, SP; IR – Forest restoration in Iracemápolis, SP; SG – Natural remnant in Campinas, SP; TT - Natural remnant in Tietê, SP. N – Sample size; A – Number of different alleles; P – Number of private alleles; H_O – Observed heterozygosity; H_E – Expected heterozygosity; F_{IS} – Within population fixation index. Confidence intervals are represented between parentheses.

Of the 1324 different alleles from neutral loci detected in the four populations, all populations shared approximately 79%, 11% corresponded to alleles shared by at least three populations, 9% were in at least two populations, and less than 1% were exclusive to a population (Figure 1.2). These results are consistent with those presented in Table 1.1, indicating no genomic differences among the forest restoration and natural remnant populations.

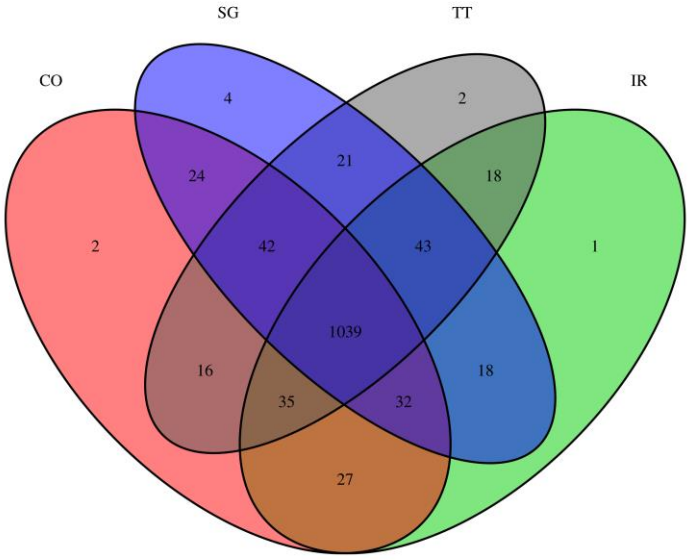

Figure 1.2 – Venn diagram showing the number of shared alleles among groups of *Casearia sylvestis* populations from restoration plantations and natural remnants in the Atlantic Forest of southeastern Brazil. CO – Samples from forest restoration in Cosmópolis, SP; IR – Samples from forest restoration in Iracemápolis, SP; SG – Samples from natural forest in Campinas, SP; TT – Samples from natural forest in Tietê, SP. Different colors represent each of the four studied populations.

1.3.3 Genetic differentiation and structure

The populations were not genetically differentiated (F_{ST} = 0). Moreover, all pairwise F_{ST} estimates (Table 1.2, lower triangular) were not significantly different from zero (P > 0.05). Based on these estimates, which consider the set of neutral loci, no population genetic structure was detected. Additionally, DAPC did not show any structure by pre-defined population subdivision (Figure 1.3).

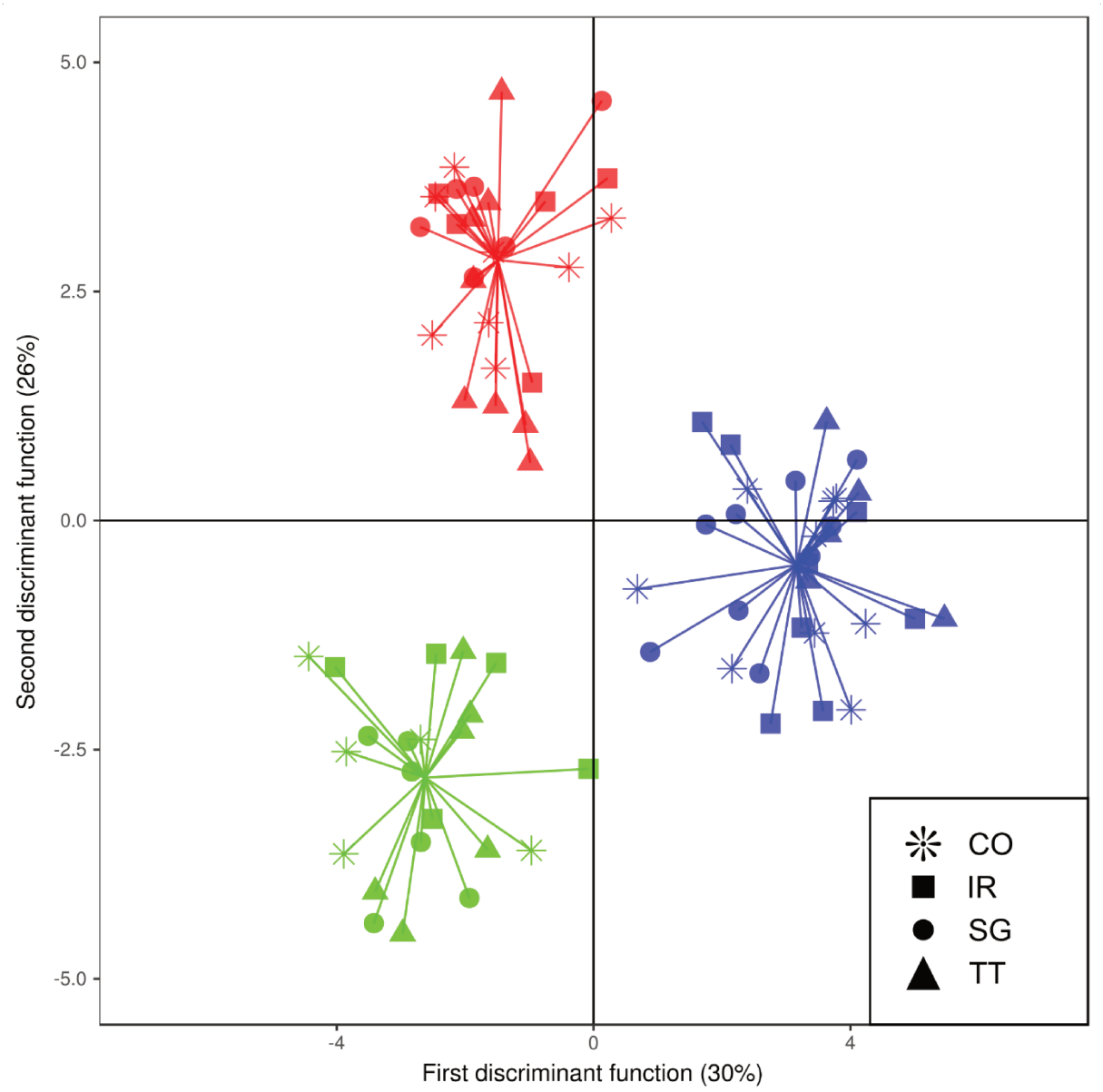

Figure 1.3 – Scatterplot from DAPC based on neutral loci using BIC showing the two main discriminant functions for *Casearia sylvestris* populations from restoration plantations and natural remnants in the Atlantic Forest of southeastern Brazil. Samples from all pre-defined populations are represented in all clusters indicating that genetic structure was not based on population subdivision. Different colors represent tree different clusters found using BIC of DAPC analysis. Different symbols represent each pre-defined population.

Table 1.2 – Pairwise Nei's F_{ST} based on neutral loci (lower triangular). Values based on neutral loci were not significantly different from zero based on confidence intervals calculated from 1000 bootstrap resamplings.

	CO	IR	SG	TT
CO	-	-	-	-
IR	-0.0006	-	-	-
SG	0.0011	0.0020	-	-
TT	0.0004	-0.0004	0.0008	-

CO – Forest restoration in Cosmópolis, SP; IR – Forest restoration in Iracemápolis, SP; SG – Natural remnant in Campinas, SP; TT - Natural remnant in Tietê, SP.

Each of the groups generated based on Bayesian information criterion (BIC) was formed by samples from all populations (Figure 1.4), reinforcing the lack of genomic structure resulting from population pre-defined subdivision.

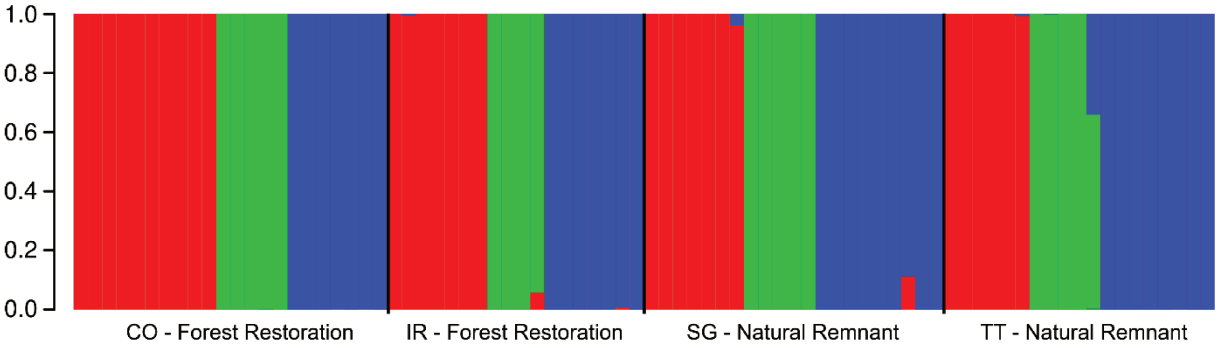

Figure 1.4 – Bar plot of membership probability from Discriminant Analysis of Principal Components for *Casearia sylvestis* populations from restoration plantations and natural remnants in the Atlantic Forest of southeastern Brazil. Groups were composed of samples from all populations; thus, the genetic structure was not based on population subdivision. CO – Samples from forest restoration in Cosmópolis, SP; IR –

Samples from forest restoration in Iracemápolis, SP; SG – Samples from natural forest in Campinas, SP; TT – Samples from natural forest in Tietê, SP. Different colors represent tree different clusters found using BIC of DAPC analysis.

We found no significant positive spatial autocorrelation at any distance interval (Figure 1.5). Moreover, no evidence of spatial genetic structure was detected in the forest fragments (Table 1.3). The regression slope of the kinship coefficient on the log of distance was not significantly different from zero (P > 0.05). Therefore, individuals located close to one another were not necessarily more genetically related than individuals separated by greater distances.

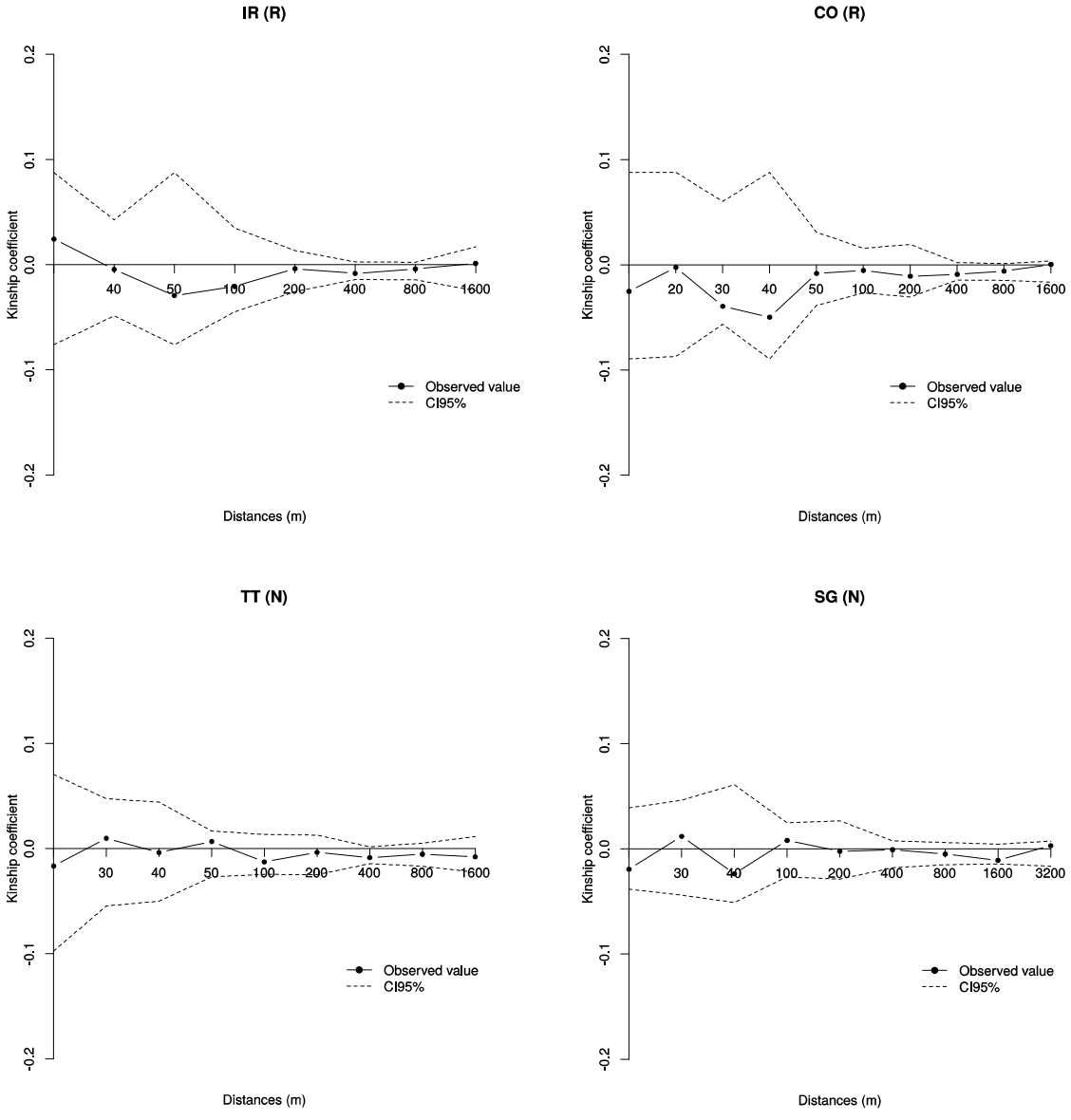

Figure 1.5 – Kinship correlograms (Loiselle et al., 1995) for *Casearia sylvestris* populations from restoration plantations and natural remnants in the Atlantic Forest of southeastern Brazil.

Table 1.3 – Estimates of fine-scale spatial genetic structure for all *C. sylvestris* populations.

Pop.	β_{log}	P-value (β_{log})	F_1	*Sp*
CO	0.0038	0.1968	0.0646	-0.0041
IR	0.0007	0.8641	0.0144	-0.0007
SG	-0.0016	0.5165	-0.0065	0.0015
TT	-0.0025	0.4026	-0.0035	0.0025

β_{log} – Regression slope of kinship coefficient on log of spatial distance; F_1 – Kinship coefficient for the first distance class; *Sp* - Vekemans and Hardy's (2004) estimator of fine-scale SGS intensity. CO – Forest restoration in Cosmópolis, SP; IR – Forest restoration in Iracemápolis, SP; SG – Natural remnant in Campinas, SP; TT - Natural remnant in Tietê, SP

1.4 DISCUSSION

The value of research involving population genomics using non-model species without previous knowledge of the genome size and complexity has been previously demonstrated in conservation genomics. With the advent of inexpensive sequencing and the advancement of techniques that reduce genome complexity, such as GBS, the study of genomic diversity and structure in natural populations can be based on hundreds of SNPs (Narum et al., 2013; Andrews et al., 2016). Here, we demonstrate the usefulness of this approach for restoration genetics.

We found no differences in the estimates of genetic diversity for *C. sylvestris* populations in natural remnants and forest restorations. The similar values of the genetic diversity estimates between forest restorations and natural remnants are promising results for Brazilian Atlantic Forest conservation, because they indicate the effectiveness of forest restoration in maintaining the genetic diversity of the species, and consequently, the potential to self-perpetuate in the restored communities. Heterozygosities are good indicators of individual fitness in a population (Reed and Frankham, 2003; Greenbaum et al., 2014) and represent an initial estimate of the adaptive potential of a population to respond to environmental changes (Caballero and Garcia-Dorado, 2013). In previous studies developed by Cavallari et al. (2010) on population relationships among *C. sylvestris* varieties in different ecosystems of Brazil, a similar level of genetic diversity was found in almost all populations. Thus, as measured by genetic diversity, the forest restoration plantations studied here were as conserved as forest remnants.

Casearia sylvestris is a pioneer species that is very abundant in disturbed fragments of the region and is frequently dispersed by birds, which may promote gene flow between restoration plantations and remnants at the landscape level. Otálora et al. (2011), studying the diversity and genetic structure of Mediterranean forests, suggested that as long as habitat quality and minimum connectivity remain among fragments, forest fragments can preserve the genetic diversity of original forests. In this context, forest restoration plantations can improve landscape connectivity and facilitate gene flow through seeds and pollen, which can mitigate the deleterious effects of inbreeding. In spite of the favorable results for genetic conservation, the studied populations had high F_{IS} values. Despite the mating system of *C. sylvestris* that allow

self-fertilization, the differences found for these estimates might be related to the overwhelming level of forest fragmentation, as found in other studies conducted in fragmented landscapes, such as those of Dick et al. (2003) and Vranckx et al. (2012).

We expected that the reduction of population size in the studied forests and the isolation of the trees in the fragments contributed to the increase in F_{IS}. Andrianoelina et al. (2006), studying the effects of fragmentation on the genetic diversity of an arboreal species in Madagascar, suggest that an increase of F_{IS} in fragmented populations can be explained by high rates of self-fertilization caused by the isolation of adult individuals after population fragmentation.

The results for the general F_{ST} and the pairwise F_{ST} estimates among populations also suggested that the genetic diversity in natural remnants and forest restoration plantations was not structured through population subdivision on the other hand the observed structure is present within populations. The DAPC results are consistent with the F_{ST} estimates and reinforce the potential gene flow among the studied forests and with other fragments of the region. Martins et al. (2016) conducted an important study on the role of small natural remnants of Atlantic Forest for conserving genetic diversity of the tree *Copaifera langsdorffii*, and these authors suggest that small forest fragments preserve the genetic diversity that existed before fragmentation. Therefore, forest fragments are crucial for *in situ* species conservation and the maintenance of long-term genetic diversity in human-modified tropical landscapes. Moreover, by establishing restoration plantations with high genomic diversity, the risk of extinctions of tropical trees threatened by reproductive isolation in fragmented landscapes can be reduced.

To explain the absence of spatial genetic structure, we suggest that seed and pollen dispersal in these populations might play a role (Hamrick and Nason, 1996). This result is consistent with previous discussions of genetic differentiation and structure for *C. sylvestris* populations in Brazilian Atlantic Forest fragments (Araujo et al., 2017). Populations of *C. sylvestris*, which is self-compatible, may be experiencing gene flow within and among populations in such a way that genetic diversity is uniformly distributed in space.

Sato et al. (2006), studying the effects of gene flow on SGS for a riparian canopy tree species in Japan, suggest extensive gene flow occurs within and among populations. Furthermore, they argue that high levels of gene flow can increase the local effective size of a population and maintain genetic diversity. Bittencourt and

Sebbenn (2007) also suggest that pollen dispersal into forest fragments promotes genetic variation, reduces kinship and inbreeding, and increases the effective population size.

Therefore, we suggest that efforts are required to not only conserve large natural remnants but also the small forest fragments that are most of the remaining remnants in the Brazilian Atlantic Forest (Rodrigues et al., 2009). These remnants preserve not only the genetic diversity but also the adaptive characteristics that will be important in the establishment and maintenance of new populations under future climate change scenarios. Additionally, we emphasize the importance of forest restoration as a way to connect forest fragments across the landscape, which will contribute to the long-term resilience of tropical forests embedded within human-modified landscapes in the current scenario of global environmental change.

Capítulo 2 – SNP discovery and local adaptation study in a native tree from Brazilian Atlantic Forest and Brazilian Savannah

ABSTRACT

Brazil is the country with the greatest terrestrial biological diversity in the world due to the different biomes and ecosystems present in its territory. Many plant species occur in more than one biome and are subject to different selective pressures. *Casearia sylvestris* is an example of a species with wide occurrence in different Brazilian biomes. Two varieties of this species are recognized, one with preferential occurrence to the Cerrado (BSV) biome and the other with preferential occurrence to the Brazilian Atlantic Forest (BAF) biome. In the ecotones between these ecosystems it is possible to find large proportions of individuals of the two varieties. The genomic bases that define the adaptation of these varieties to the environments in which they are found have not yet been reported. The objective of this study was to evaluate the genetic diversity and structure of these varieties to understand the genetic basis of the adaptation to the BSV and BAF biomes. For this, genomic libraries were built using ddRADseq and sequenced in an Illumina HiSeq 4000. Estimates of heterozygosities, private alleles, coefficient of inbreeding and genetic differentiation based on sets of neutral loci and loci under selection that may be related to adaptation were then calculated. The results indicate that differences in the levels of neutral genetic diversity among populations are due to differences in the conservation status of the area. In addition, the results indicate strong neutral and adaptive genetic structuring by variety. In the ecotones where the occurrence of the two varieties in sympatry is observed, it was also possible to verify that these varieties are genetically distinct. Therefore, it can be concluded in this study that microevolutionary events can lead to the local adaptation of populations of species occurring in the BSV and BAF biomes, since the structure of *C. sylvestris* varieties is accentuated when considering loci under selection that can present alleles related to the adaptation of these varieties to these different biomes.

Keywords: Conservation genetics, ddRADseq, local adaptation, gene flow, genetic differentiation, native trees

2.1 INTRODUCTION

Brazil has a high level of biological diversity, a result of the different biomes and ecosystems that characterize this country. On its territory, there is the largest vegetal biodiversity in the world, presenting more than 55 thousand species already cataloged that correspond to approximately 22% of the total plant world terrestrial biota (MMA, 2017a). Among these biomes, the importance of the Brazilian Savannah also known as Cerrado (hereafter BSV) and the Brazilian Atlantic Forest (hereafter BAF), due to the current rate of devastation and the high degree of floristic endemism, stand out as two of the 25 critical areas for world conservation (Myers et al., 2000).

The BSV is the second largest territorial biome in South America, occupying around 21% of the Brazilian territory (Medeiros and Young, 2011; MMA, 2017b). It has rich biodiversity, being considered the most biodiverse tropical savannah in the world. Its flora is composed of more than 11,600 species (Mendonça et al., 2008; MMA, 2017b), which it is estimated that 44% are endemics (Klink and Machado, 2005). Despite this, 57% of the original vegetation was completely destroyed and half of the remaining areas are heavily altered (Machado et al., 2004) due to an annual deforestation rate estimated at 1.5% (about 3 million hectares per year).

The BAF is another biome recognized for its great biodiversity, estimated to contain about 1 to 8% of the world's species (Myers et al., 2000), including more than 8000 endemic species (Tabarelli et al., 2003). After centuries of exploration through extractivism, the expansion of agricultural frontiers, industrialization and urbanization, less than 8% of its original area remains preserved (SOS Mata Atlantica and INPE, 2015). When we include secondary forests and small fragments (<100 ha), this percentage increases to just over 11% (Ribeiro et al., 2009).

The protection of forest remnants in Brazil, including the threatened BSV and BAF biomes, is ruled by a set of laws that stipulate normalizations for the preservation and management of these areas in public and private areas. Among these areas, conservation units (CUs) are territorial areas, including their environmental resources, with relevant natural characteristics, which have the function of ensuring the representativeness of significant and ecologically viable samples of the different populations, preserving the existing biological heritage (MMA, 2017c). However, the CUs that represent the BSV and BAF biomes are insufficient to reach these protection

goals because they do not represent a satisfactory portion of the area occupied by the original biome (IPEA, 2011). In addition, it should be emphasized that although these strategies are fundamental to biodiversity conservation, the model to which CUs have been implemented, disassociated from a more comprehensive perspective of the landscape as if they were isolated islands surrounded by urban areas and deforestation, may not reach the long-term sustainability of these areas causing loss of ecological functions and consequently biodiversity (Delelis, Rehder and Cardoso, 2010; Ganem, 2010).

Many studies have focused on the consequences of forest fragmentation and isolation to these populations. Among the main topics addressed in these studies are the effects of absence of gene flow among populations of different fragments, inbreeding and genetic drift due to the small population size (Ezard and Travis, 2006; Mona et al., 2014; Christie and Knowles, 2015; Peñaloza-Ramírez et al., 2016). Such situation may lead to a reduction in genetic diversity and a consequent reduction in the overall fitness of these populations (Allard, Jain and Workman, 1968; Hughes et al., 2008; Star and Spencer, 2013; Balick et al., 2015). In addition, genetic diversity is the source for natural selection, in the absence of this diversity, in the face of the environmental change, populations may lose the potential to adapt to these changes again leading to a compromise of the survival (Werf et al., 2009), as well as the long-term (Fischer and Matthies, 1998; Buza, Young and Thrall, 2000; Wisely et al., 2002; Reed and Frankham, 2003; Willi et al., 2007; Werf et al., 2009). In addition, another important genetic aspect of habitat fragmentation that has so far received little attention is the effect of this fragmentation on the direction and intensity of natural selection (Willi, Buskirk and Hoffmann, 2006). Due to the limited migration capacity to new environments, the occurrence of positive selection should be given more attention in populations of native tree species.

The intensity and direction of selection in populations of natural remnants of BSV and BAF has not been clearly reported. However, in populations of the tree species *Casearia sylvestris* Sw. (Salicaceae), which has a large occurrence in the BSV and BAF, two morphologically distinct varieties have been observed, one occurring preferentially in the BAF (var. *sylvestris*) and another preferably found in the BSV (var. *lingua*) (Sleumer, 1980). The distinction between the varieties is very tenuous (Torres and Yamamoto, 1986; Silva et al., 2006), and many individuals with intermediate morphology can be found, especially in ecotones between the BSV and the BAF

(Cavallari et al., 2010). The most recent taxonomic revision of the genus in the State of São Paulo, Brazil (Wanderley et al., 2002) does not consider the existence of intraspecific levels in *C. sylvestris* but results obtained through microsatellite markers indicate two genetically distinct varieties (Cavallari et al., 2008; Cavallari et al., 2010).

The genomic bases that define the two *C. sylvestris* varieties (var. *sylvestris* and var. *lingua*) have not been fully elucidated. Knowledge about the mechanisms that may have led to the occurrence of this genomic divergence may be of great utility for understanding local adaptation in populations of BSV and BAF. Thus, the hypothesis tested in our study is that prior to the fragmentation of these domains, a forest *continuum* favored gene flow among the populations of these domains by delaying the effect of positive selection on alleles related to local adaptation to these environments. After fragmentation and consequently reduction of ecotones areas, the gene flow restriction probably contributed to selection and fixation of alleles adapted to each of these domains.

Until recently, most genetic studies in natural populations have been developed with the use of few molecular markers, usually accessing only neutral genetic variation. The study of neutral variation elucidated as phenomena such as genetic drift and gene flow affect the allelic frequencies of whole genome. However, to understand in what intensity and direction the selection may have acted in the differentiation of the populations of native species of the BSV and BAF, it is necessary to access the adaptive genetic variation that is locus-specific and can be accessed by markers with wide coverage of the genome such as AFLP and SNPs (Allendorf, Hohenlohe and Luikart, 2010; Kirk and Freeland, 2011; Funk et al., 2012; Andrews et al., 2016). The advantage of using SNPs to study adaptive variation is that it is a codominant marker with the possibility of knowing the sequence associated with the marker. This information is useful because it may allow the discovery of the biological processes, structure or molecular mechanisms involved in the local adaptation of these populations. The utility and techniques for obtaining such markers have been known but the high costs for obtaining them limited the application of SNPs markers only to humans or to model species.

However, since the development of restriction-associated DNA (RAD) techniques and large-scale sequencing, which allow the discovery of large numbers of SNPs, it has been possible to conduct genomic-scale studies in populations of non-model species (Ekblom and Galindo, 2010). SNPs have the potential for large-scale

genome scanning for the discovery of both neutral and adaptive variation. However, even RAD-based techniques can be applied to any non-model species, case studies are still needed (Xu et al., 2014), especially in the absence of reference genomes, a fact that occurs with most of the native species of the BSV and BAF.

The objective of the present study was to understand the genomic relations between the *C. sylvestris* varieties (var. *sylvestris* and var. *lingua*). in the BSV and BAF biomes to gain insights on the microevolutionary processes that may be involved in the differentiation *between C. sylvestris* varieties and adaptation to these different biomes.

2.2 MATERIALS AND METHODS

2.2.1 Studied species and sampling

C. sylvestris, a native species commonly found in the BSV and BAF, was used for this study. This species belongs to the family Salicaceae, the same family in which the genus *Salix* and *Populus* are classified, widely used in studies of population genomics. It has a wide distribution in the Americas, occurring from Mexico to Argentina (Sleumer, 1980) and can be found in shrub or tree form, depending on the location of the biome of occurrence. It is a pioneer species (Tabarelli, Villani and Mantovani, 1993; Dias et al., 1998; Silva et al., 2003) and its reproductive system has not yet been studied using molecular markers, but its flowers are hermaphrodite, and self-fertilization is compatible (Barbosa, 1997; Ramalho, 2004). Like other species of this genus, *C. sylvestris* is usually pollinated by different species of flies while seed dispersal is carried out by small birds that distribute seeds at small distances (Wanderley et al., 2002).

It has been observed in *C. sylvestris* individuals that there are two morphologically distinct varieties, *C. sylvestris* var. *sylvestris* (Figure 2.1 A) with preferential distribution in the BAF and the variety *C. sylvestris* var. *lingua* (Figure 2.1 B) with preferential distribution in the BSV (Sleumer, 1980). The distinction between the varieties is very tenuous and many individuals with intermediate morphology can be found, especially in transition formations between the BSV and the BAF (Torres and Yamamoto 1986; Wanderley et al., 2002).

Figure 2.1 – *Casearia sylvestris* varieties found in the biomes BAF and BSV. *Casearia sylvestr*is var. *sylvestris* (A) and *Casearia sylvestris* var. *lingua* (B). Pictures: Cavallari (2010).

Twelve individuals were collected from five different populations *of C. sylvestri*s located in areas of occurrence of BSV, BAF and in the transition zone between these two domains known as ecotones (ECO). In order to assure the correct taxonomic determination of the collected individuals, each sample was evaluated by Roseli Buzanelli Torres, the specialist in the *Casearia* genus of the State of São Paulo, and the vouchers of each sample were deposited in the herbarium of the Instituto Agronômico de Campinas, Campinas, Brazil.

The first population was collected in the Botucatu State Forest (Botucatu-SP, Brazil, 22º 56' S, 48º 27' W), which is a 33.8-ha area dominated by biome BSV *strictu sensu*. It has a dry winter mesothermic climate (CWB), annual average temperature of 19.4º Celsius and average annual precipitation of 1,314 millimeters. The second population was collected at the Ecological Station of Assis (Assis-SP, Brazil, 22º 34' S, 50º 24' W), which has 1,760.64-ha of vegetation characteristic of biome BSV *lato sensu* and *strictu sensu*. It has a transition climate between cwa and cfa in the classification of Köppen, mean annual temperature of 21º Celsius and average annual rainfall of 1,413 millimeters.

The third population was collected at the Caetetus Ecological Station (Galia-SP, Brazil, 22º 23' S, 49º 42' W), which is an area of 2,178.84-ha with vegetation characteristic of the Semideciduous Seasonal Forest (BAF). It has climate classified as CWA and average annual precipitation of 1,480 millimeters. The fourth population was collected in the Anchieta Island State Park (Ubatuba - SP, Brazil, 23º 32' S, 45º 03' W), which has an area of 828-ha with a vegetative mosaic composed mostly of Restinga, Ombrophylous Rala Forest and Dense Ombrophilous Forest (BAF).

By observing individuals with intermediate characteristics to the *varieties C. sylvestris* var. *sylvestris* and *C. sylvestris* var. *lingua* in transitional regions between the BSV and BAF biomes, the fifth population was collected at the Mogi Guaçu Ecological Station (Mogi Guaçu - SP, Brazil, 22º 15' S, 47º 11' W). Is an area of 980.71-ha with transitional vegetation between the BSV and BAF. Its climate is classified as CWA with annual average temperature of 20.5º Celsius and average annual precipitation of 1335 millimeters.

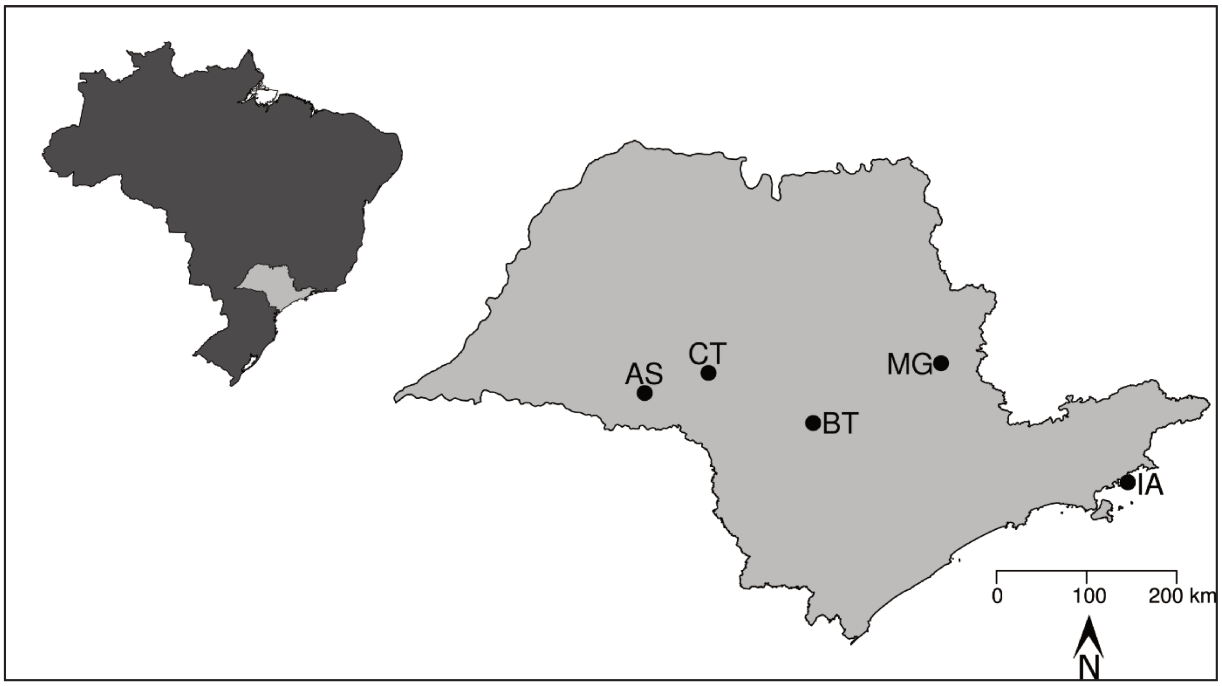

Figure 2.2– Localization of each Casearia sylvestris population considered in this study. AS – Assis Ecological Station, Assis, SP; BT – Botucatu State Forest, Botucatu, SP; CT – Caetetus Ecological Station, Galia, SP; IA – Ilha Anchieta State Park, Botucatu, SP; MG – Mogi Guaçu Ecological Station, Mogi Guaçu.

2.2.2 Genomic library development and sequencing

Leaf tissue DNA from adult samples was extracted using the protocol described by Doyle and Doyle (1987). The samples were used in the genomic libraries development using the double digest RAD-seq (ddRADseq) method proposed by Peterson et al. (2012), which consists of the construction of genomic libraries using two restriction enzymes. This method offers the potential to obtain large-scale SNPs in non-model species with relatively low cost when compared to other methods based on RAD-seq.

DNA from each sample was digested using the restriction enzymes EcoRI and MseI. To the fragments generated from the samples were bonded adapters containing specific barcodes for each sample. Fragments ranging from 300-400 bp

were selected on agarosis gel and purified by electroelution in which a electric charge was applied to the agarosis gel containing the selected fragments and these fragments migrated to a special membrane where they were collected and eluted in TAE 1x. The libraries were then validated using Agilent Bioanalyzer 2100 equipment and then quantified by qPCR to estimate the dilution required for sequencing. The libraries were sequenced in two lanes of Illumina HiSeq 4000 producing a large ammount of 80 bp reads length.

2.2.3 SNP calling and raw data processing

There is not enough information on the genome size and ploidy of the *C. sylvestris* to date. The only evidence we have for ploidy results from the application of microsatellites that indicated diploidy for this species. Like any RAD approach, ddRADseq allows the discovery and genotyping of thousands of SNPs across the genome, in regions close to the restriction site, without the need for an already sequenced reference genome (Seeb et al., 2011). In the SNP calling stage we used the *Stacks* v. 1.45 (Catchen et al., 2011) which includes a set of programs that through a model-based approach allows the construction of loci SNP even in the absence of reference genomes or when the ploidy is unknown.

Samples from each lane were processed (*process_radtags*) for removal of low-quality reads (-q) and reads presenting non-called base (-c). The files corresponding to a same sample were then concatenated to constitute a set of unique samples. Barcodes have been recovered as long as they have a maximum of one mismatch (-r, process_radtags). Since the reference genome for *C. sylvestris* is not available, the loci were constructed *de novo*. The sequences of each sample were aligned using a minimum depth of 3 and allowing up to 2 mismatches with the ustacks program (-m 3, -M 2, ustacks). Then, the loci catalog was constructed from all samples, allowing 2 mismatches (-n 2, cstacks) between loci of different samples. All samples were compared to the catalog, using sstacks, to determine the set of loci in each. Next, we implemented a correction step with rxstacks, disconnecting loci when the mean log probability of the locus in the metapopulation was less than -10; Confused loci were also removed. Using the population program, a filter was applied at loci and individuals to retain loci with a maximum of 20% missing data.

2.2.4 Data analysis

2.2.4.1 Descriptive analysis

Descriptive analyses regarding the quality and quantity of SNPs detected with ddRADseq were extracted during the SNP calling process that was performed using *Stacks* package. With this information, plots were generated to summarize the results obtained with the application of ddRADseq for the detection of SNPs in *C. sylvestris*. Microsatellite screening was performed against the SNP catalog using MISA tool for Unix system.

2.2.4.2 Genetic diversity, differentiation and structure

First, outlier loci were identified in the set of SNPs using an F_{ST}-based approach using the infinite alleles model and a false discovery rate (FDR) of 0.05, as implemented in the *Lositan* software (Antao et al., 2008). Three SNPs categories were identified using this method: candidate loci to positive selection, candidate loci to balancing selection and candidate loci to neutrality. Our focus is on the genetic diversity, genetic differentiation and genetic structure explained by neutral loci and genetic differentiation and structure ruled by loci under positive selection. Therefore, we analyzed these two sets of SNPs separately: candidate loci to neutrality (hereafter neutral loci) and candidate loci to positive selection (hereafter outlier loci). Filters were applied to both set of SNPs to remove loci and individuals with more than 20% of missing data. To perform the studies of genetic diversity in BSV and BAF, we used the following population parameters: number of different alleles (A), number of private alleles (P), observed heterozygosity (H_O), expected heterozygosity (H_E), and coefficient of inbreeding (F_{IS}). We tested significant values for F_{IS} using confidence intervals based on 1000 bootstrap resamples over individuals. Estimates of genetic diversity parameters were calculated using the *PopGenKit* (Paquette, 2012) and *hierfstat* (Goudet and Jombart, 2015) packages. Plots were generated using *ggplot2* (Wickham, 2009). All packages were developed for use in the *R* (R Core Team, 2017).

We calculated F-statistics for neutral loci and outlier loci using the *diveRsity* package (Keenan et al., 2013). The significance of these estimates was verified using 95% confidence intervals based on 1000 bootstrap resamples over individuals. The

Discriminant Analysis of Principal Components (DAPC) described by Jombart et al. (2008) were conducted over neutral loci and outlier loci using Bayesian Information Criterion (BIC) to explore the genetic structure in these populations using *adegenet* (Jombart, 2008; Jombart and Ahmed, 2011).

2.2.4.3 Annotation of outlier loci

DNA sequences corresponding to candidate loci for positive selection were compared against the National Center for Biotechnology Information (NCBI) non-redundant protein database using *blastx* algorithm (Gish and States, 1993) with e-value threshold of 10^{-3}. We then assigned gene ontology (GO) names using *Blast2GO* software (Conesa et al., 2005) and plotted the results using *R* (R Core Team, 2017).

2.3 RESULTS

2.3.1 Descriptive analyzes

The sequencing of ddRADseq libraries for *C. sylvestris* were useful in obtaining large amounts of reads. The descriptive analysis regarding the amount and distribution of reads per samples after the processing of the sequencing data are shown in Table 2.1.

Table 2.1 - Read summary after read processing (*process_radtags*) of ddRADseq libraries sequencing data of *Casearia sylvestris*.

Parameter	Value
Total de Reads	280,858,051
Retained RAD-tags	231,953,337
Average RAD-tags	5,154,519 ± 659,339.9

RAD – Restriction-site associated DNA

It was observed that approximately 49 million reads were discarded due to failures in the restriction site sequence (No RAD-tag) or low sequence quality and by observation of error value it was found that probably due to stochastic differences the remaining reads were not distributed equally among the samples. Only samples with at least 300,000 reads were considered and the maximum number of reads per sample obtained was 13,802,363. The first implication of these differences lies in the depth of coverage of the stacks that are groups of aligned reads that serve as the source for the discovery of SNPs in a raw sequence. In this regard, the stacks frequency per depth of coverage found in our study is shown in Figure 2.3. It is possible to observe that the proportion of stacks constituted by a few reads (< 10 reads) is more frequent than stacks constituted by many reads (> 10 reads). The mean depth of coverage was 20.82 and the threshold used to recover the stacks in the sequencing data was enough to proceed with a good quality analysis.

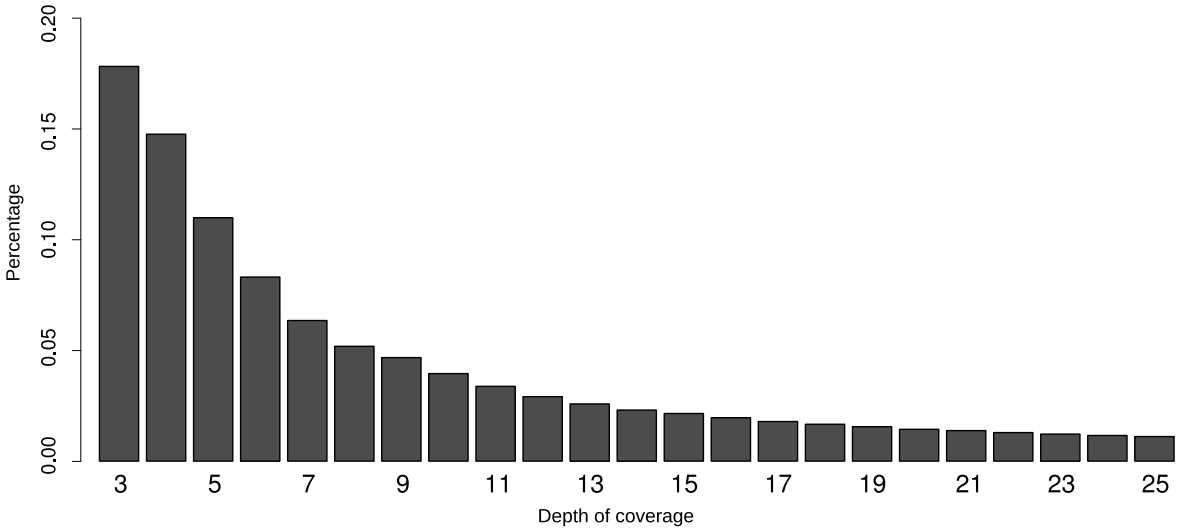

Figure 2.3 – Depth of coverage distribution found in processed RAD-tags of *Casearia sylvestris* libraries.

The first catalog SNPs generated before the correction step (*rxstacks*) contained 1,060,364 RAD-tag loci. After the corrections were applied, the new generated catalog presented 986,938 RAD-tag loci. The distribution of the log likelihood values found in the first catalog and which was used as the basis for choosing the thresholds applied in the correction step is showed in Figure 2.4.

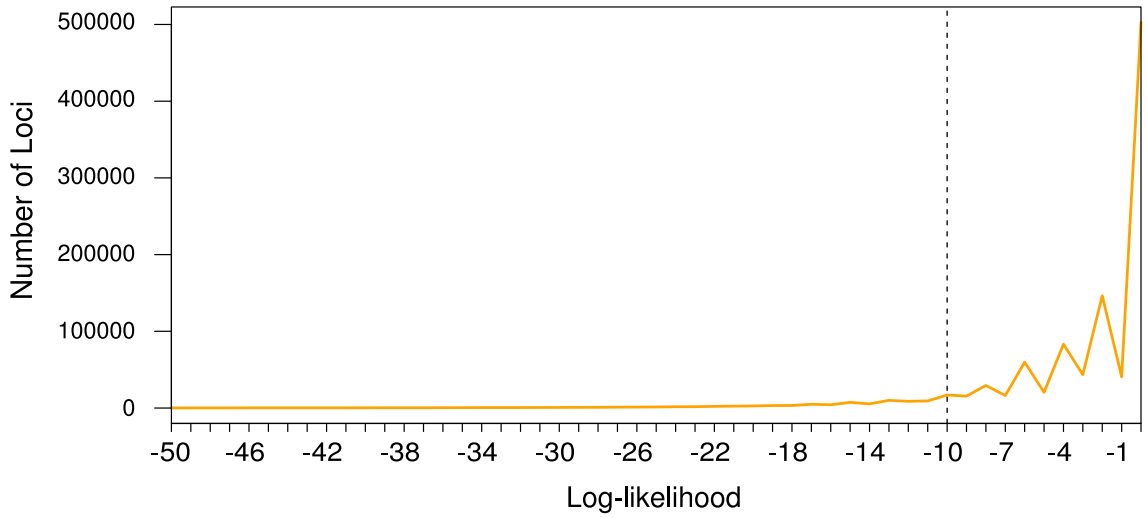

Figure 2.4 – Log-likelihood distribution found in the first catalog generated based on *C. sylvestris* samples from BSV, BAF its ecotones. Dashed line represents the threshold value. Continuous line represents the distribution of number of loci by Log-likelihood values.

The SNPs discovered in our study are distributed throughout the genome of *C. sylvestris*, including in microsatellite regions. In this catalog, it was possible to find 10,782 SNPs flanking microsatellite regions and the proportions of motifs found are shown in Table 2.2. Among the microsatellite regions found, it was possible to recover two previously described *C. sylvestris* in previous studies by Cavallari et al. (2008).

Table 2.2 – Absolute and relative frequency of microsatellites motif sizes found in *Casearia sylvestris* SNP catalog.

SSR Type	Absolute Frequency	Relative Frequency (%)
Di-nucleotides	6605	0.6125
Tri-nucleotides	2809	0.2605
Tetra-nucleotides	589	0.0546
Penta-nucleotides	163	0.0151
Hexa-nucleotides	85	0.0078
Composite Interrupted	489	0.0453
Composite Perfect	42	0.0038

After searching for polymorphisms against the last generated catalog, it was possible to identify 4,695 high quality SNP markers in the samples from five populations of *C. sylvestris*. After using an F_{ST} based approach on the set of SNPs obtained, it was possible to identify subsets of 201 candidate loci to balancing selection, 279 loci candidates to positive selection (outlier loci) and 4,215 candidate loci to neutrality (neutral loci). In Figure 2.5 the candidate loci to balancing selection are plotted in the yellow area of the graph, this area corresponds to the lower limits of the confidence intervals at 99%. The loci candidates to positive selection (outlier loci) are plotted in the red area that is delimited by the upper limits of the confidence intervals at 99%. Candidate loci to neutrality (neutral loci) are plotted between these two boundaries and constitute most polymorphic loci found.

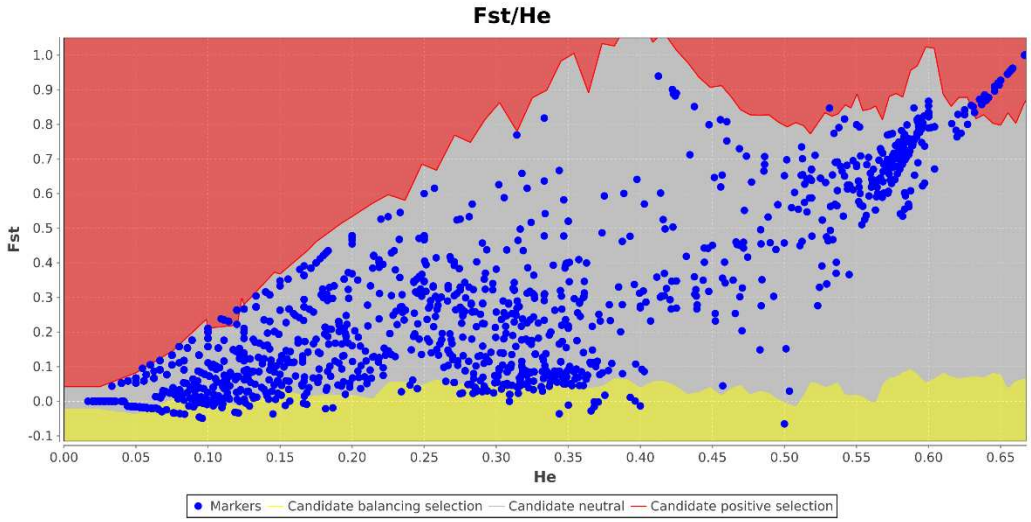

Figure 2.5 - Loci distribution of *Casearia sylvestris* ddRADseq data using F_{ST} approach implemented in Lositan (Antao, 2008). Loci plotted in red area are candidate loci to positive selection while loci plotted in yellow area are candidate loci to balancing selection. Most loci are candidate to neutrality.

2.3.2 Genetic diversity

The calculated values of the estimates of the population parameters based on the set of neutral loci are showed in Table 2.3. These results are further evidence of the potential of the ddRADseq to explore population differences.

Table 2.3 - Genetic diversity parameters estimates based on neutral loci *of Casearia sylvestris* populations from BSV, BAF and its ecotones.

Biome	Pop.	N	A	P	H_O	H_E	F_{IS}
BSV	AS	10	5,279	640	0.111	0.165	0.174
BSV	BT	12	4,170	56	0.059	0.085	0.386
BAF	CT	8	5,011	452	0.084	0.122	0.309
BAF	IA	7	4,091	326	0.054	0.060	0.069
Ecotone	MG	8	4,511	308	0.148	0.168	0.058

AS – Assis Ecological Station, Assis, SP; BT – Botucatu State Forest, Botucatu, SP; CT – Caetetus Ecological Station, Galia, SP; IA – Ilha Anchieta State Park, Botucatu, SP; MG – Mogi Guaçu Ecological Station, Mogi Guaçu, SP; BSV – Brazilian Savannah

(Cerrado); BAF – Brazilian Atlantic Forest; Pop – Population; N – Sample size; A – Number of different alleles; P – Number of private alleles; H_O – Observed heterozygosity; H_E – Expected heterozygosity; F_{IS} – Within population fixation index. Confidence intervals are represented between parentheses.

There are differences in the amount of neutral genetic diversity between the populations of BSV and BAF as well as different levels of coefficient of inbreeding in these populations. Of the 6,958 different alleles found, 37% are represented in all populations and 26% are unique to each one of five populations studied. The number of shared alleles among each group of populations are showed in Figure 2.6.

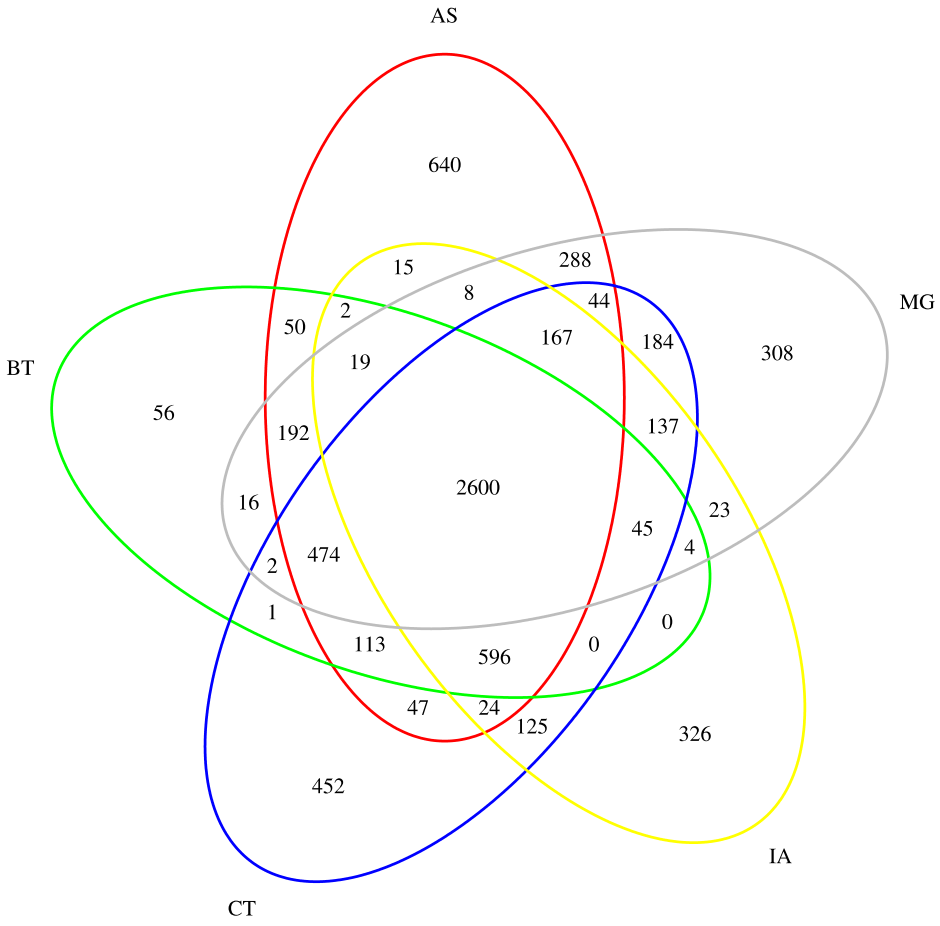

Figure 2.6 – Venn diagram showing the number of shared alleles among groups of *Casearia sylvestis* populations from BSV, BAF and its ecotones. AS – Assis Ecological Station, Assis, SP; BT – Botucatu State Forest, Botucatu, SP; CT – Caetetus

Ecological Station, Galia, SP; IA – Ilha Anchieta State Park, Botucatu, SP; MG – Mogi Guaçu Ecological Station, Mogi Guaçu, SP. Different colors represent each of the five studied populations.

2.3.3 Genetic differentiation and genetic structure

In general, when considering the set of neutral loci, it is possible to observe that the studied populations are genetically different ($F_{ST\ neutral}$ = 0.249) and these differences are even more pronounced when considering the set of outlier loci ($F_{ST\ outlier}$ = 0.805). Pairwise F_{ST} estimates for neutral loci (upper triangular) and outlier loci (lower triangular) are showed in Table 2.4.

Table 2.4– Pairwise Nei's F_{ST} based on neutral loci (upper triangular) and outlier loci (lower triangular). Values based on neutral loci were significantly different from zero based on confidence intervals calculated from 1000 bootstrap resamplings.

Biome	Pop.	AS	BT	CT	IA	MG
BSV	AS	-	0.111	0.277	0.375	0.070
BSV	BT	0.448	-	0.404	0.508	0.200
BAF	CT	0.854	0.972	-	0.120	0.097
BAF	IA	0.848	0.959	0.264	-	0.195
ECO	MG	0.423	0.645	0.342	0.355	-

AS – Assis Ecological Station, Assis, SP; BT – Botucatu State Forest, Botucatu, SP; CT – Caetetus Ecological Station, Galia, SP; IA – Ilha Anchieta State Park, Botucatu, SP; MG – Mogi Guaçu Ecological Station, Mogi Guaçu, SP. BSV – Brazilian Savannah (Cerrado); BAF – Brazilian Atlantic Forest; Pop – Population;

These populations of *C. sylvestris* from BSV, BAF and its ecotones are genetically structured by population and variety when considering neutral loci and when outlier loci are considered as well. In Figure 2.7 it is possible to observe the scatterplot showing the pattern of genetic structure discovered with neutral loci, while in Figure 2.8 it is possible to observe the pattern of genetic structure discovered using the set of outlier loci.

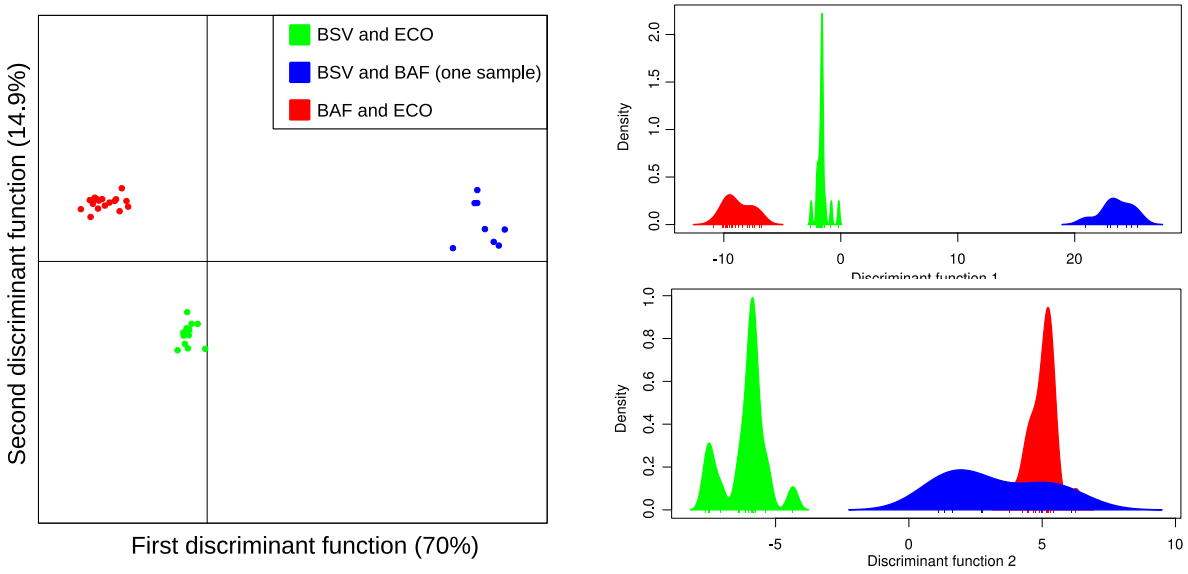

Figure 2.7 – Scatterplot from DAPC based on neutral loci using BIC showing the two main discriminant functions for *Casearia sylvestis* populations from Brazilian Savannah, Brazilian Atlantic Forest and its ecotones. Different colors represent tree different clusters found using BIC of DAPC.

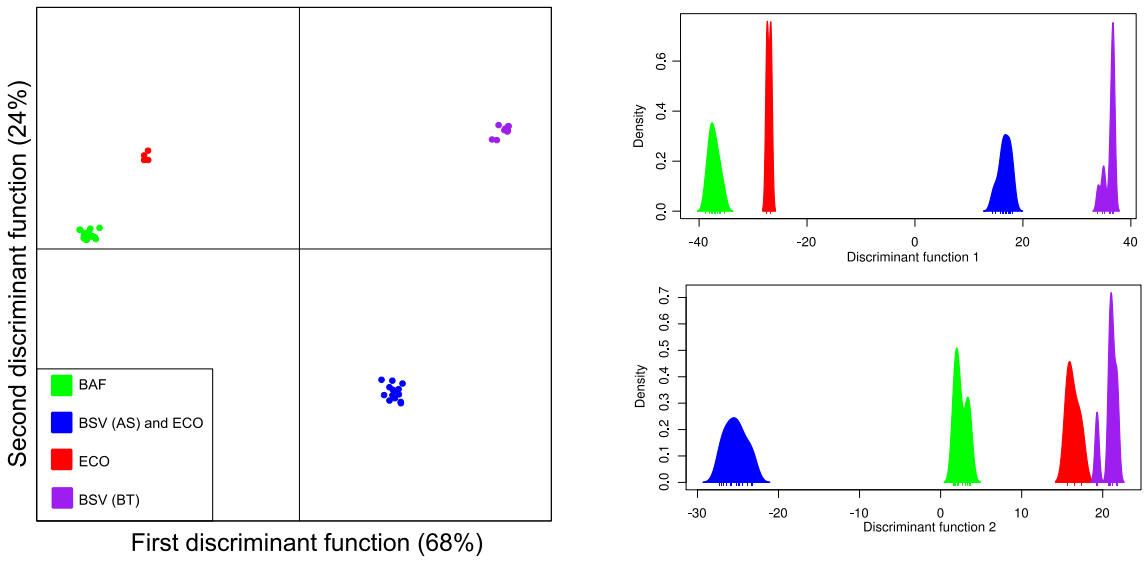

Figure 2.8 – Scatterplot from DAPC based on outlier loci using BIC showing the two main discriminant functions for *Casearia sylvestis* populations from Brazilian Savannah, Brazilian Atlantic Forest and its ecotones. Different colors represent four different clusters found using BIC of DAPC.

The membership probabilities of assigning the individuals to each of the groups shown in the scatterplots based on neutral loci and outlier loci are shown respectively in Figure 2.9 and Figure 2.10.

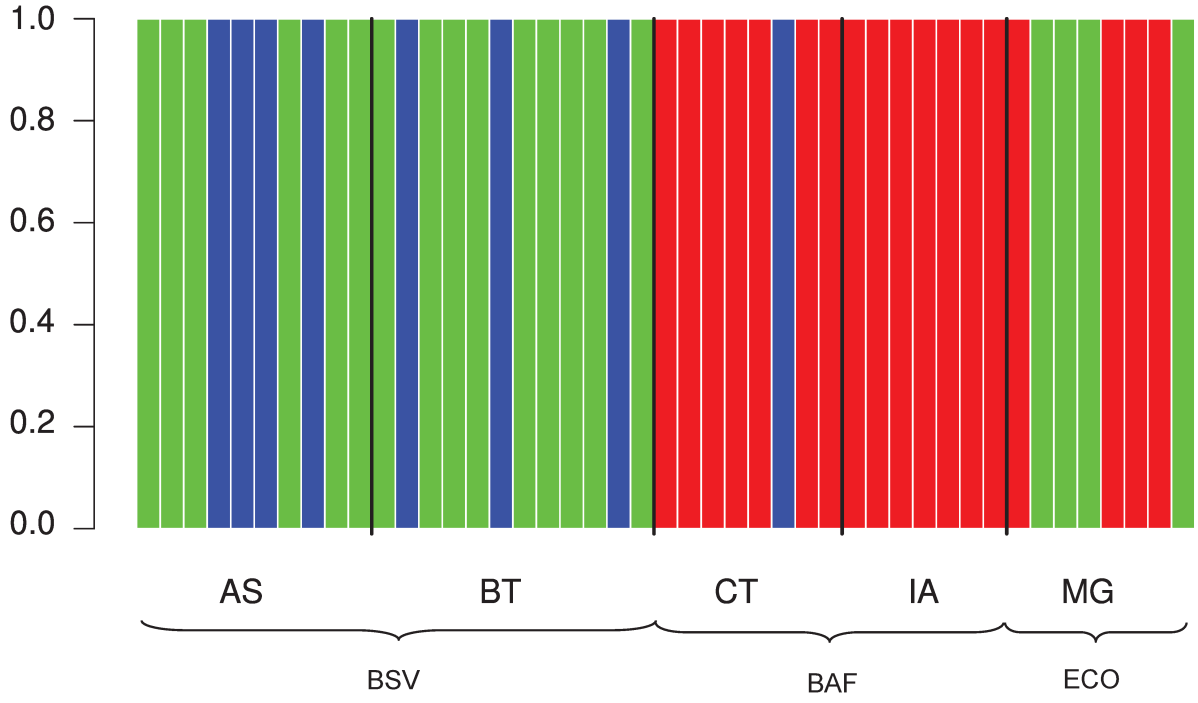

Figure 2.9 – Bar plot of membership probability from Discriminant Analysis of Principal Components based on neutral loci of *Casearia sylvestis* populations from Brazilian Savannah, Brazilian Atlantic Forest and its ecotones. AS – Assis Ecological Station, Assis, SP; BT – Botucatu State Forest, Botucatu, SP; CT – Caetetus Ecological Station, Galia, SP; IA – Ilha Anchieta State Park, Botucatu, SP; MG – Mogi Guaçu Ecological Station, Mogi Guaçu, SP; BSV – Brazilian Savannah (Cerrado); BAF – Brazilian Atlantic Forest; ECO – Ecotone. Different colors represent tree different clusters found using BIC of DAPC analysis.

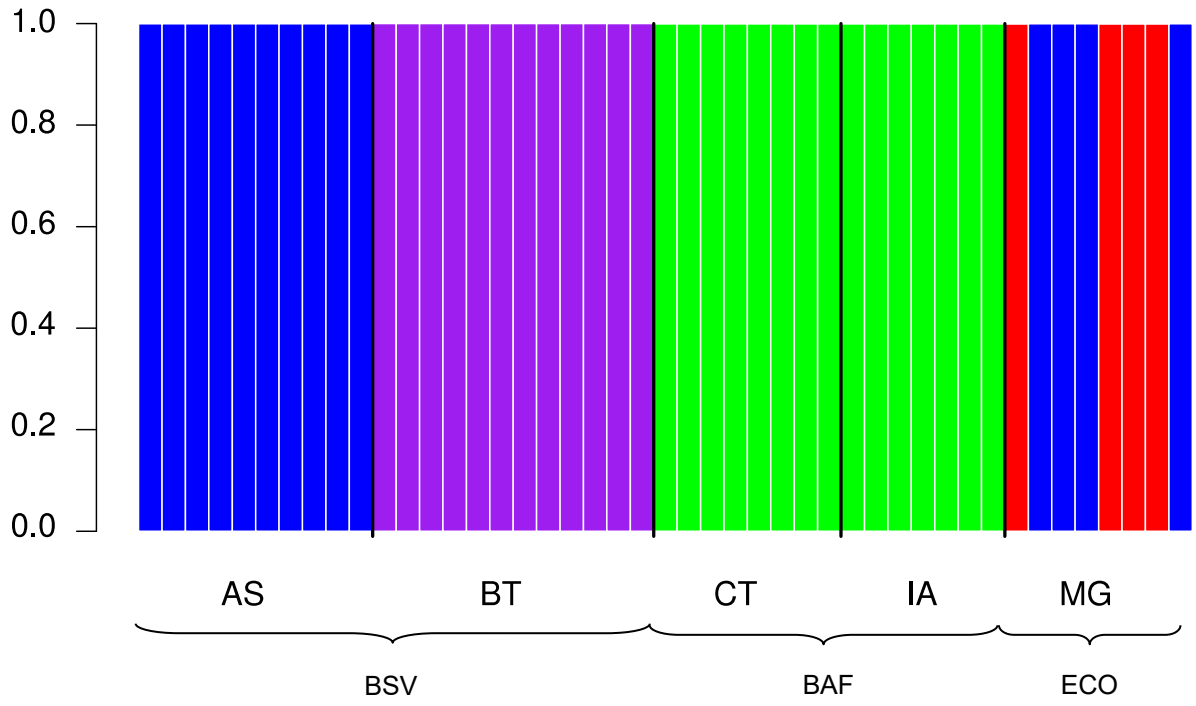

Figure 2.10 – Bar plot of membership probability from Discriminant Analysis of Principal Components based on outlier loci of *Casearia sylvestis* populations from Brazilian Savannah, Brazilian Atlantic Forest and its ecotones. AS – Assis Ecological Station, Assis, SP; BT – Botucatu State Forest, Botucatu, SP; CT – Caetetus Ecological Station, Galia, SP; IA – Ilha Anchieta State Park, Botucatu, SP; MG – Mogi Guaçu Ecological Station, Mogi Guaçu, SP; BSV – Brazilian Savannah (Cerrado); BAF – Brazilian Atlantic Forest; ECO – Ecotone. Different colors represent four different clusters found using BIC of DAPC analysis.

2.3.4 Annotation of outlier loci

Of the 279 outlier loci candidates for positive selection, we found 16 loci that presented an e-value < 10^{-5}, which are similar to protein-binding sequences in the NCBI database. This represents 5% of the outlier loci candidates for positive selection found. The distribution of the species found in the blastx hits are shown in Figure 2.11.

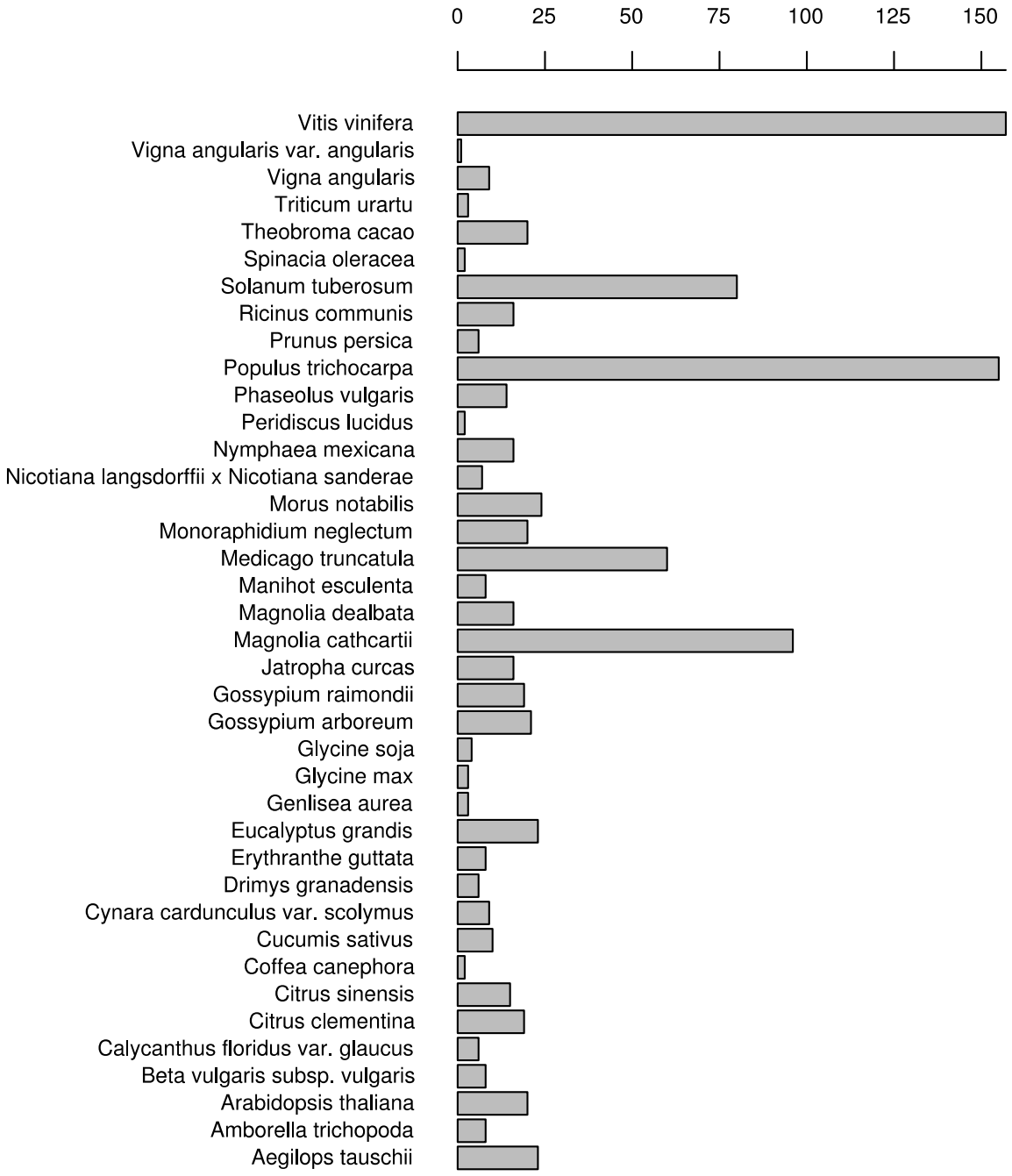

Figure 2.11 - Species distribution of *blastx* hits. Many hits matched against *Populus trichocarpa* species.

Among the species that received the greatest numbers of hits, the *Populus trichocarpa* species belongs to the family Salicaceae, the same to which *Casearia sylvestris* is currently classified. In general, most of the hits found were in GO Terms corresponding to the subdomain "Molecular Function", followed by "Cellular

Component" and "Biological Process" (Figure 2.12). However, it is important to highlight that among the GO Terms associated with the highest number of hits, the regions that are related to the chloroplasts genome corresponds to about 15% of the total hits obtained stand out.

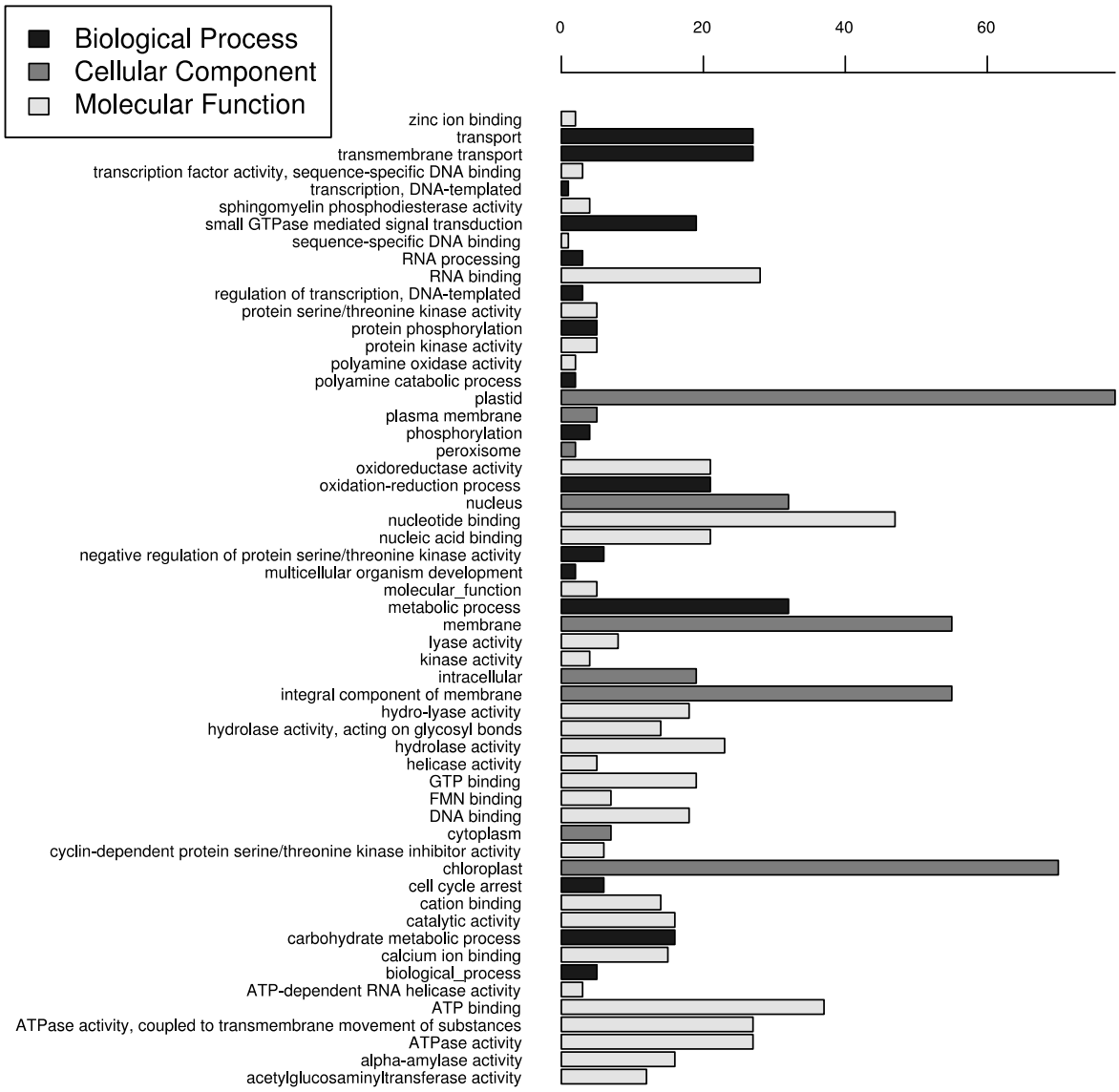

Figure 2.12 - Frequency of GO Terms associated with biological domains found with *blastx* hits. The bar plot indicates the absolute frequency of GO terms of subdomains corresponding to Biological Process, Molecular Function or Cellular Component.

2.4 DISCUSSION

The success of *de novo* sequencing of any genome region of a species depends on several factors such as genome size, species ploidy, frequency of repetitive regions. The quality of sequencing and the distribution of reads per sample, for example, may have greater implications for the interpretation of *a posteriori* results because, when there is no uniformity in the distribution of reads per sample, the detection of SNPs in regions with low depth may be impaired since the high frequency of sequencing errors may result in false polymorphisms or addition of erroneous consensus sequences to the SNPs catalog (Minoche, Dohm and Himmelbauer, 2011). These aspects of sequencing quality are important factors to be considered in population genomics studies, so these differences may compromise the number of SNPs detected and the quality of these SNPs in terms of missing data rates.

Regarding the SNPs discovery process in the raw sequencing data used in this study, the first implication of the differences in the distribution of reads per sample is the depth of coverage of the stacks. The stacks correspond to the set of exactly-matching or putative alleles short-reads aligned to each other (Catchen et al., 2011). Since the depth of coverage consists of the number of good quality reads which is represented in the sequencing data (Sims et al., 2014), this parameter is inversely correlated with the size and complexity of the genome, so it is possible to infer based on our results on the distribution of depth of coverage that *C. sylvestris* probably has a large genome.

Solutions to reduce possible biases due to sequencing errors or detection of SNPs in regions of low depth of coverage are the corrections applied in the catalog after an examination throughout the population allowed, and the discarding of thousands of low quality loci that could compromise the accuracy of population studies. In our study, one of the main corrections applied in this step was regarding the discard of loci whose log likelihood values were highly negative. These loci usually may have low coverage or may have many sequencing errors and it is recommended that they be discarded in the subsequent steps of catalog assembly (Catchen et al., 2011).

SNPs detected by the ddRADseq may be distributed throughout the genome of the species close to the restriction sites of the enzymes used to build the genomic library, some regions flanking microsatellites may be included. Microsatellites

are important tools for the study of neutral genetic variation in natural populations and, as genetic markers, they are important for estimating population size and kinship because they are able to detect a large intralocus variation (Kirk and Freeland, 2011; Freeland, Kirk and Petersen, 2011). The ddRADseq demonstrated efficiency in the screening of microsatellite regions in native species, since even without any sequence enrichment step it was possible to recover microsatellites in many reads.

The main evidence of the potential of ddRADseq for genomic studies in natural populations was the large number of high-quality SNPs found in the *C. sylvestris* populations used in our investigation, since even using a mean sample size of 9 individuals per population it was possible to identify polymorphism in enough number of SNPs markers to develop studies at a genome wide scale. Unlike other molecular markers generally classified as first or second generation (Schlötterer 2004), such as microsatellites, where 20 to 30 individuals are generally required to obtain accurate estimates of genetic diversity and differentiation (Pruett and Winker, 2008), the SNPs markers have been very useful tools to calculate these estimates with precision even when few individuals per population are considered (Nazareno et al., 2017).

One of the main advantages of using RADseq-based methods is the possibility of extensive genome scan (Luikart et al., 2003; Allendorf, Hohenlohe and Luikart 2010). Since populations of different biomes were used in these studies, it is justified the detection of SNPs located in both neutral and expressed regions because different environmental conditions allow the establishment of alleles related to local adaptation by positive selection. This made it possible to extend genetic diversity studies in populations of native species of BSV and BAF on a genomic scale not only to detect phenomena of allelic frequency change on a genomic scale, such as genetic drift and gene flow, but also the detection of specific effects of locus such as natural selection (Holderegger, Kamm and Gugerli, 2006; Kirk and Freeland, 2011).

The differences observed in these populations based on the set of neutral candidate loci, which are not subject directly to selection, may be a consequence of habitat fragmentation that led to gene flow restriction. Many authors have been reported the role of habitat fragmentation as a gene flow barrier (Slatkin, 1985, 1987; Milligan, Leebens-Mack and Strand, 1994; Hamrick, 2004) and intensification of inbreeding and genetic drift (Luikart et al., 2003). Since there is a correlation between reduced genetic diversity evidenced by neutral markers and a negative impact on

overall population fitness (Reed and Frankham, 2003), it is suggested that more efforts are needed to preserve the genetic diversity in these populations. In the case of the IA population, despite relatively low genetic diversity, the population is apparently well established and few crosses between related individuals are happening when we consider the relatively low coefficients of inbreeding. However, there are probably other factors that are responsible for the isolation of the population of BT because in addition to low genetic diversity this population has a high coefficient of inbreeding. Among these factors, it is suggested that the absence of forest fragments close to the BT population has led to restriction of gene flow and loss of neutral genetic diversity. The high genetic differentiation of this population even when compared to the AS population of the same biome is indicative that the occurrences of these phenomena that led to the reduction of neutral genetic diversity in BT may not be exclusively related to environmental differences or differences among varieties, but may also be related to the general conservation status of this population. The other populations considered in our study presented higher levels of genetic diversity, even if they correspond to environments with different characteristics. Therefore, these results indicate that the differences among these populations in terms of neutral genetic diversity amount are due to aspects inherent in population size and conservation status.

Corroborating with results found in our study, Leonardi et al. (2012) studied the effects of habitat fragmentation on genetic diversity and genetic structure in natural populations of *Fagus sylvatica* L. in Italy using alloenzymatic markers and highlighted the influence of the distance among populations and forest fragments sizes on the genetic differentiation of the populations besides evidencing the intensification of the effect of genetic drift in populations with higher levels of fragmentation. The low genetic diversity levels found in the populations of *C. sylvestris* from BT and IA highlight the importance in the conservation of the natural remnants of BSV and BAF because as found in our study there may be many fragments of both biomes that present low genetic diversity. Reichmann et al. (2017) studied genetic diversity in populations of *Maytenus dasyclada* in natural remnants of the BAF and verified the importance of these fragments for the maintenance of gene flow among these populations and consequently for the preservation of genetic diversity.

All populations studied are different in terms of neutral genetic diversity and mainly in terms of outlier genetic diversity calculated on the basis of loci candidates for positive selection. When we consider the pairwise estimates of genetic differentiation

among populations it is possible to observe that there is a greater genetic differentiation among the populations of different biomes than among the populations of the same domain, considering both sets of loci. Hey and Pinho (2012) suggested the F_{ST} value = 0.35 as threshold to determine whether population differences are due to intraspecific or interspecific differences. The vast genetic differences were between the populations BT and IA. In this context, we highlight the population of Mogi-Guaçu that corresponds to a transition zone between BSV and BAF where the population of this area intermediate genetic differentiation in relation to the populations from different nature domains. These results support the hypothesis that before the process of environmental fragmentation, gene flow between the lingua and sylvestris ecotype of *C. sylvestris* was more frequent in transition zones between these biomes. Environmental fragmentation isolated the natural remnants of BSV and BAF, reducing the contact zones between these two domains and consequently hindering the gene flow between these varieties.

Moreover, the high levels of genetic differentiation calculated based on outlier loci reinforce that the genetic differentiation between the populations are more related to the local adaptation of the *C. sylvestris* varieties to the BSV and BAF biomes. Cavallari et al., (2010) achieved similar results when the genetic relationship among *C. sylvestris* populations from these same areas using microsatellite markers and suggested that the genetic differentiation between *C. sylvestris* populations of the BSV and BAF is related not only to the distance between the areas but also to the existence of two distinct varieties of this species. Brousseau et al. (2015) studied the neutral and adaptive genetic variation in populations of the *Eperua falcata* species within an environmental *continuum* in French Guiana using AFLP markers and suggested that genetic differentiation can occur both at large geographic scales and between the closest populations since there are environmental differences that limit gene flow.

As expected, this genetic differentiation among different populations of *C. sylvestris* varieties and biomes resulted in a genetic structure pattern by population and by variety. The genetic structure observed when considering the neutral loci highlights the relation of the *C. sylvestris* varieties without the influence of positive selection. Our results using neutral loci corroborate with Cavallari et al. (2010) regarding the genetic nature of the differences between *C. sylvestris* varieties. When considering only the outlier loci, the genetic structure by variety was even more intense than that observed in the set of neutral loci. In evaluating these aspects, it was possible

to detect the role of ecotones in the preservation as adaptive genetic variation as well neutral genetic variation in the contrasting environments found in the BSV and BAF. These results support the hypothesis that an environmental *continuum* favored the gene flow between individuals of both varieties since it is possible to find in the ecotones the genetic variation that occurs in the populations of the two biomes but in each biome, occurs almost exclusively individuals of a single group.

The existence of an environmental *continuum* that favors gene flow between varieties implies that even though a variety occurs more frequently in the environment which is adapted due to positive selection, it is not exclusive to the environment that is inserted. Field observations and our results support this discussion. Furthermore, with fragmentation of the environments populations are more subject to genetic drift and gene flow constraints (Reed and Frankham, 2003; Leonardi et al., 2012). Although positive selection favors adapted alleles, genetic drift then intensified by habitat fragmentation may act against alleles that promote local adaptation (Lenormandi, 2002; Andrews, 2010), reducing the general fitness of the population.

Studies on local adaptation are of particular importance mainly because they provide information that can contribute to the planning of conservation strategies (Funk et al., 2012). Since populations of different biomes with contrasting environmental conditions were used, it was expected that outlier loci should correspond to regions that control biological processes or other mechanisms that lead to the adaptation of these varieties to their domain of origin. With the availability of a reference genome it would be possible to map the sequences corresponding to the outlier loci and to identify to which biological components they are attached or directly linked. At the present moment, as for most native species of the Atlantic Forest and Cerrado the reference genome for *Casearia sylvestris* is unavailable, but it is possible to perform the study of outlier loci assuming homology between the sequence associated to outlier loci and genomes for other plant species available in the NCBI database through the GO annotation (Conesa et al., 2005).

Among the species that received the high frequency of *blastx* hits during the GO annotation, *Populus trichocarpa* stands out that corresponds to a species of the same family of *Casearia sylvestris*. Except the terms associated to chloroplast genome, those that received the clear majority of hits were those responsible for regulating cellular components and biological process, such as cellular membrane components and transmembrane transport. Our results point to a new and important

genetic evidence that the populations of Cerrado and Atlantic Forest are probably in the process of adaptation to the conditions of these biomes. These environments differ in a range of abiotic conditions such as temperature, soil composition, humidity and light intensity, and relationships have been found between cellular membrane components and transmembrane transport may be associated with some adaptation to soil nutrient uptake (Sánchez-Calderón et al., 2013; Rao et al., 2016). The *blastx* hits associated with differences in chloroplast genome between *C. sylvestris* varieties from Cerrado and Atlantic Forest also should be considered. Since light intensity is an important difference between Cerrado and Atlantic Forest, the *blastx* hits associated with chloroplast may represent adaptation to these conditions seeing that these cellular components may respond to abiotic stress (Spetea and Rintama 2014; Bobik and Burch-Smith, 2015). Therefore, it is suggested that these abiotic conditions operated selective pressures that acted on the populations of *C. sylvestris* of the different biomes giving rise to two distinct varieties of this species.

Conservation units of integral protection in Brazil that guarantee the preservation of the biomes and the genetic diversity of their populations correspond to less than 3% of their territory and only about 4% of this area is destined to preserve the ecotones between the among different biomes that occurs in the country (Arruda, 2003). The results presented in this study can be extrapolated to other species that have a large occurrence in different biomes. In our study, the occurrence of the microevolution process is reported, a phenomenon that usually occurs gradually in nature (Reznick and Ricklefs, 2009), but that with the fragmentation of habitats due to anthropic actions may be intensified or impaired. Efforts should therefore be made not only to preserve areas that correspond to well defined biomes, but also to consider transitional regions between biomes that maintain both neutral genetic variation and adaptive genetic variation.

Therefore, based on this study, it can be concluded that the application of ddRADseq is effective in the discovery of high-quality SNPs for genomic studies in native species. In addition, it is suggested that the populations of *C. sylvestris* var. *lingua* is adapted to the environmental conditions present in the Cerrado, while the variety *C. sylvestris sylvestris* is adapted to the environmental conditions present in the Atlantic Forest. Furthermore, on the basis of these results, it is suggested that, like *C. sylvestris*, other species may be subject to positive selection, therefore, an increase in the representativity of the areas defined for the conservation of these biomes is

necessary in order to preserve these varieties without loss of genetic diversity.

DISCUSSÃO GERAL

As pesquisas descritas nestes dois capítulos correspondem a abordagens pioneiras na utilização de espécie não modelo, nativa dos biomas Cerrado e Mata Atlântica do Brasil, para estudos genéticos populacionais aplicados à biologia da conservação. Com o advento das técnicas de redução de complexidade do genoma pelo uso de enzimas de restrição, tal como GBS (Elshire et al., 2011) e ddRADseq (Peterson et al., 2012), associado as técnicas de sequenciamento em larga escala, estudos de diversidade e estruturação genética populacional puderam ser realizados a partir da avaliação de centenas a milhares de marcadores SNPs com baixo custo (Narum et al., 2013; Andrews et al., 2016). Nestes trabalhos foram demonstrados a aplicação de marcadores SNPs descobertos por técnicas de preparo e sequenciamento de short-reads na genética da restauração e na genômica da adaptação.

A espécie *Casearia sylvestris* utilizada nestes estudos é uma espécie pioneira, tolerante a autofecundação (Barbosa, 1997; Ramalho, 2004) e de rápido estabelecimento em áreas de clareiras, florestas esparsas e bordas dos fragmentos florestais (Tabarelli, Villani e Mantovani, 1993; Dias et al., 1998; Silva et al., 2003; Marmontel et al., 2013). Os frutos contendo suas sementes são dispersados por pássaros (Wanderley, 2002) fato que pode ter contribuído para sua ampla ocorrência e distribuição que se estende do México à Argentina (Sleumer, 1980). Para esta espécie são reconhecidas duas variedades morfologicamente (Sleumer, 1980; Torres e Yamamoto, 1986) e geneticamente distintas (Cavallari et al., 2010) que possuem ocorrência preferencial a biomas específicos (Torres e Yamamoto, 1986). Apesar de até o momento não haver sido desenvolvido genomas de referência para esta espécie, suas características biológicas a fizeram uma excelente escolha para realização dos estudos aqui apresentados.

No primeiro capítulo, tratou-se do estudo de diversidade e estruturação genética em populações de *C. sylvestris* de restaurações florestais e remanescentes naturais de Mata Atlântica pelo uso da técnica de GBS (Elshire et al., 2011) para descoberta de SNPs. Nesse estudo, observou-se que os níveis de diversidade genética em remanescentes naturais e restaurações florestais é semelhante e que

não existe estruturação genética por população pré-definida (remanescentes naturais e restaurações florestais).

Em estudos anteriores desenvolvidos por Cavallari et al. (2010) em populações de *C. sylvestris*, níveis similares de diversidade genética foram encontrados entre quase todas as populações consideradas por estes autores. Sugere-se que o enriquecimento genético e a redução da estruturação genética entre as populações de remanescentes naturais e restaurações florestais tem ocorrido através do fluxo gênico entre as áreas de restauração florestal e pequenos fragmentos de remanescentes naturais próximos a estas restaurações pois desde que o fragmento florestal tenha tamanho adequado e um mínimo de conectividade com outros fragmentos, estes podem preservar a diversidade genética que outrora existia nas florestas originais antes do processo de degradação (Otálora et al., 2011).

Esses resultados são promissores para conservação da Mata Atlântica pois eles indicam a eficiência das restaurações florestais em manter a diversidade genética das espécies e, consequentemente, o potencial para resiliência das populações em áreas restauradas. As estimativas de diversidade genética consideradas neste estudo são bons indicadores do *fitness* geral de uma população e representam um parâmetro inicial do potencial de uma população para responder às mudanças climáticas globais (Reed e Frankham, 2003; Greenbaum et al., 2014).

Já no segundo capítulo, tratou-se do estudo sobre adaptação local em populações de variedades de *C. sylvestris* do Cerrado e Mata Atlântica pelo uso da técnica ddRADseq (Peterson et al., 2012) para descoberta de SNPs. Neste estudo foi possível observar que as populações de *C. sylvestris* var. *sylvestris* podem estar localmente adaptadas às condições ambientais da Mata Atlântica enquanto as populações *de C. sylvestris* var. *lingua* podem estar localmente adaptadas às condições ambientais do Cerrado.

Os marcadores SNPs descobertos pela técnica ddRADseq aplicada no segundo capítulo foram submetidos a um teste de neutralidade (Beaumont e Nichols, 1996) implementado no software Lositan (Antao et al., 2008) que levou à construção de dois subconjuntos de marcadores SNPs. O primeiro subconjunto corresponde aos marcadores que podem não estar sujeitos a seleção natural, aqui chamados de loci neutros, e que, portanto, podem ser melhor utilizados para investigar a ocorrência de fenômenos como deriva genética e fluxo gênico (Holderegger, Kamm e Gugerli, 2006; Kirk e Freeland, 2011). O segundo subconjunto corresponde aos marcadores que

podem estar sujeitos a seleção positiva, aqui chamados de loci outliers, e que podem ser utilizados para estudos de adaptação local (Holderegger, Kamm e Gugerli, 2006; Kirk e Freeland, 2011).

As diferenças observadas nas estimativas de diversidade, diferenciação e estruturação genética das populações das duas variedades de *Casearia sylvestris* do Cerrado e Mata Atlântica, quando considerados os loci neutros, podem ser consequência da fragmentação dos *habitats* que pode ter levado à restrição do fluxo gênico entre estas áreas. Muitos autores associaram a fragmentação do *habitat* como uma das principais barreiras do fluxo gênico (Slatkin, 1985, 1987, Milligan, Leebens-Mack e Strand, 1994, Hamrick, 2004) e intensificação da endogamia e deriva genética (Luikart et al., 2003). Já os níveis de diferenciação e estruturação genética considerando os loci outliers indicam que a diferenciação genética entre as populações também está bastante relacionada à adaptação local das variedades de *C. sylvestris* aos biomas Cerrado e Mata Atlântica, não apenas à fragmentação dos *habitats*. Tal diferenciação pode ter se intensificado devido a eventos microevolutivos capazes de levar à adaptação local das diferentes variedades aos biomas Cerrado e Mata Atlântica. Cavallari et al., (2010) obtiveram resultados semelhantes quanto à relação genética entre estas variedades utilizando marcadores microssatélites e sugeriu que a diferenciação genética entre as populações de *C. sylvestris* do Cerrado e Mata Atlântica está relacionada não apenas com a distância entre as áreas, mas também a existência de destas distintas variedades nas populações desta espécie. Com a fragmentação destes biomas, as populações estão mais sujeitas a restrições de fluxo genético e consequentemente maior diferenciação genética (Reed e Frankham, 2003; Leonardi et al., 2012).

De forma geral os dois estudos forneceram importantes informações para conservação das populações de espécies que apresentam características biológicas semelhantes à espécie *Casearia sylvestris*. A diferença no número de marcadores SNPs para *Casearia sylvestris* descobertos pela técnica de GBS e pela técnica de ddRADseq se deve principalmente à presença de uma etapa de seleção de fragmentos com tamanho específico na técnica de ddRADseq. Esta seleção permite que estes fragmentos sejam sequenciados repetidas vezes aumentando a profundidade de cobertura dos loci polimórficos encontrados. No entanto, esta diferença não implicou na precisão das estimativas calculadas e atendeu aos objetivos propostos para cada pesquisa.

CONCLUSÃO GERAL

Considerando o primeiro estudo, conclui-se que as restaurações florestais têm contribuído para manutenção de níveis similares de diversidade genética em populações de *Casearia sylvestris* quando comparados aos remanescentes naturais estudados. Além disto, devido a influência do método de reflorestamento adotado na implementação dessas restaurações florestais na diversidade genética das populações, sugere-se que os níveis similares de diversidade genética e ausência de estruturação genética por população sejam devidos ao fluxo gênico entre as restaurações florestais e outros fragmentos florestais de remanescentes naturais próximos a essas áreas de restauração.

No segundo estudo foi possível concluir que a forte estruturação genética encontrada entre as populações de *Casearia sylvestris* do Cerrado e Mata Atlântica, principalmente quando considerados os loci candidatos a seleção positiva, pode ser sinal da ocorrência de eventos microevolutivos capazes de levar a adaptação local destas populações. Além disso, foi possível encontrar relação entre as sequências desses loci candidatos a seleção positiva com atividades biológicas que podem estar sujeitas a pressões seletivas diferentes nos biomas considerados neste estudo.

REFERÊNCIAS

Ab'Saber, A., 2007. **Os domínios de natureza no Brasil**: Potencialidades paisagísticas. Edited by Plínio Martins Filho. 4th ed. São Paulo: Ateliê Editorial;

Allard, R. W.; Jain, S. K.; Workman, P. L., 1968. "The genetics of inbreeding populations." **Advances in Genetics**, v. 14, p. 55–131. DOI:10.1016/S0065-2660(08)60425-3;

Allendorf, F. W.; Hohenlohe, P. A.; Luikart, G., 2010. Genomics and the future of conservation genetics. **Nature Reviews Genetics**, v. 11, n. 10, p. 697-709. DOI: 10.1038/nrg2844;

Andrews, C. A., 2010. Natural selection, genetic drift, and gene flow do not act in isolation in natural populations. **Nature Education Knowledge,** v. 3, n. 10, p. 5;

Andrews, K. R.; Good, J. M.; Miller, M. R.; Luikart, G.; Hohenlohe, P. A., 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. **Nature Reviews Genetics**, v. 17, n. 2, p. 81-92. DOI: 10.1038/nrg.2015.28;

Andrianoelina, O.; Rakotondraoelina, H.; Ramamonjisoa, L.; Maley, J.; Danthu, P.; Bouvet, J. M., 2006. Genetic diversity of *Dalbergia monticola* (Fabaceae) an endangered tree species in the fragmented oriental forest of Madagascar. **Biodiversity & Conservation**, v. 15, n. 4, p. 1109-1128. DOI: 10.1007/978-1-4020-5208-8_4;

Antao, T.; Lopes, A.; Lopes, R. J.; Beja-Pereira, A.; Luikart, G., 2008. LOSITAN: A workbench to detect molecular adaptation based on a F_{ST}-outlier method. **BMC Bioinformatics**, n. 9:323. DOI: 10.1186/1471-2105-9-323;

Araujo, F. L.; Siqueira, M. V. B. M.; Grando, C.; Viana, J. P. G.; Pinheiro, J. B.; Alves-Pereira, A.; Campos, J. B.; Brancalion, P. H.; Zucchi, M. I., 2017. Genetic diversity of *Casearia sylvestris* populations in remnants of the Atlantic Forest. **Genetics and Molecular Research**, v. 16, n. 1, p. 4393-4396. DOI: 10.4238/gmr16019105;

Arroyo-Rodríguez, V.; Melo, F. P. L.; Martínez-Ramos, M.; Bongers, F.; Chazdon, R. L.; Meave, J. A.; Norden, N.; Santos, B. A.; Leal, I. R.; Tabarelli, M., 2017. Multiple successional pathways in human-modified tropical landscapes: new

insights from forest succession, forest fragmentation and landscape ecology research. **Biological Reviews**, v. 92, n. 1, p. 326-340. DOI: 10.1111/brv.12231;

Arruda, M. B., 2003. Representatividade ecológica com base na biogeografia de biomas e ecorregiões continentais do Brasil: O caso do bioma Cerrado. **Tese de Doutorado**. Universidade de Brasília, Brasília – DF;

Balick, D. J.; Ron Do, C. A.; Cassa, D. R.; Shamil, R. S., 2015. Dominance of deleterious alleles controls the response to a population bottleneck. **PLoS Genetics**, v. 11, n. 8, p. 1–23. DOI:10.1371/journal.pgen.1005436;

Barbosa, A. A. A., 1997. Biologia reprodutiva de uma comunidade de Campo Sujo, Uberlandia/MG. **Tese de Doutorado**. Universidade Estadual de Campinas. Available from: http://repositorio.unicamp.br/jspui/handle/REPOSIP/316111;

Beaumont, M. A.; Nichols, R. A., 1996. Evaluating loci for use in the genetic analysis of population structure. **Proceedings of the Royal Society of London B,** v. 263, n. 1377, p. 1619-1626. DOI: 10.1098/rspb.1996.0237;

Bell, G.; Collins, S., 2008. Adaptation, extinction and global change. **Evolutionary Applications**, v. 1, p. 3-16. DOI: 10.1111/j.1752-4571.2007.00011.x;

Bittencourt, J. V. M.; Sebbenn, A. M., 2007. Patterns of pollen and seed dispersal in a small, fragmented population of the wind-pollinated tree *Araucaria angustifolia* in southern Brazil. **Heredity**, v. 99, n. 6, p. 580-591. DOI: 10.1038/sj.hdy.6801019;

Bobik, K.; Burch-Smith, T. M., 2015. Chloroplast signaling within, between and beyond cells. **Frontiers in Plant Science**, v. 6, n. 781, p. 1–26. DOI:10.3389/fpls.2015.00781;

Brancalion, P. H. S.; Viani, R. A. G.; Aronson, J.; Rodrigues, R. R.; Nave, A. G., 2012. Improving planting stocks for the Brazilian Atlantic Forest restoration through community-based seed harvesting strategies. **Restoration Ecology**, v. 20, n. 6, p. 704-711. DOI: 10.1111/j.1526-100X.2011.00839.x;

Brousseau, L.; Foll, M.; Scotti-Saintagne, C.; Scotti, I., 2015. Neutral and adaptive drivers of microgeographic genetic divergence within continuous populations: The case of the neotropical tree *Eperua falcata* (Aubl.). **PLoS One**, v. 10, n. 3, p. e0121394. DOI: 10.1371/journal.pone.0121394;

Burgarella, C.; Navascués, M.; Soto, Á.; Lora, Á.; Fici, S., 2007. Narrow genetic base in forest restoration with holm oak (*Quercus ilex* L.) in Sicily. **Annals of Forest Science**, v. 64, n. 7, p. 757-763. DOI: 10.1051/forest:2007055;

Butler, R. A.; Laurance, W. L., 2008. New strategies for conserving tropical forests. **Trends in Ecology and Evolution**, v. 23, n. 9, p. 469–72. DOI:10.1016/j.tree.2008.05.006;

Buza, L.; Young, A.; Thrall, P., 2000. Genetic erosion, inbreeding and reduced fitness in fragmented populations of the endangered tetraploid pea *Swainsona recta*. **Biological Conservation**, v. 93, n. 2, p. 177–86. DOI:10.1016/S0006-3207(99)00150-0;

Caballero, A.; García-Dorado, A., 2013. Allelic diversity and its implications for the rate of adaptation. **Genetics**. 195: 1373-1384. DOI: 10.1534/genetics.113.158410;

Cardoso, M. A.; Provan, J.; Powell, W.; Ferreira, P. C. G.; De Oliveira, D. E., 1998. High genetic differentiation among remnant populations of the endangered *Caesalpinia echinata* Lam. (Leguminosae-Caesalpinioideae). **Molecular Ecology**, v. 7, n. 5, p. 601–608. DOI:10.1046/j.1365-294x.1998.00363.x;

Catchen, J. M.; Amores, A.; Hohenlohe, P.; Cresko, W.; Postlethwait, J. H.; De Koning, D. J., 2011. Stacks: building and genotyping loci de novo from short-read sequences. **G3: Genes | Genomes | Genetics**, v. 1, n. 3, p. 171–182. DOI:10.1534/g3.111.000240;

Catchen, J.; Hohenlohe, P. A.; Bassham, S.; Amores, A.; Cresko, W. A., 2013. Stacks: An analysis tool set for population genomics. **Molecular Ecology**, v. 22, n. 11, p. 3124-3140. DOI: 10.1111/mec.12354;

Cavallari, M. M.; Billot, C.; Bouvet, J. M.; Favreau, B.; Zucchi, M. I.; Palmieri, D. A.; Gimenes, M. A., 2008. Isolation and characterization of microsatellite markers for *Casearia sylvestris* Sw. (Salicaceae), a neotropical medicinal tree. **Molecular Ecology Resources**, v. 8, n. 4, p. 802–804. DOI:10.1111/j.1755-0998.2007.02069.x;

Cavallari, M. M.; Gimenes, M. A.; Billot, C.; Torres, R. B.; Zucchi, M. I.; Cavalheiro, A. J.; Bouvet, J. M., 2010. Population genetic relationships between *Casearia sylvestris* (Salicaceae) varieties occurring sympatrically and allopatrically in different ecosystems in south-east Brazil. **Annals of Botany**, v. 106, n. 4, p. 627-636. DOI: 10.1093/aob/mcq151;

Chazdon, R. L., 2008. Beyond deforesttation: Restoring forest and ecosystem services on degraded lands. **Science**, v. 320, p. 1458-1460. DOI: 10.1126/science.1155365;

Chen, H., 2016. **VennDiagram: Generate high-resolution Venn and Euler plots**. Available from: https://cran.r-project.org/package=VennDiagram;

Christie, M. R.; Knowles, L. L., 2015. Habitat corridors facilitate genetic resilience irrespective of species dispersal abilities or population sizes. **Evolutionary Applications**, v. 8, n. 5, p. 454–463. DOI:10.1111/eva.12255;

Conesa, A.; Götz, S.; García-Gómez, J. M.; Terol, J.; Talón, M.; Robles, M., 2005. Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. **Bioinformatics**, v. 21, p. 3674-3676. DOI: 10.1093/bioinformatics/bti610;

De Rezende, C. L.; Uezu, A.; Scarano, F. R.; Araujo, D. S. D., 2015. Atlantic Forest spontaneous regeneration at landscape scale. **Biodiversity Conservation**, v. 24, p. 2255-2272. DOI: 10.1007/s10531-015-0980-y;

Delelis, C. J.; Rehder, T.; Cardoso, T. M., 2010. **Mosaico de Áreas Protegidas: Reflexões E Propostas Da Cooperação Franco-Brasileira**. 1ed ed. Brasília.

Dias, M. C.; Vieira, A. O. S.; Nakajima, J. N.; Pimenta, J. A.; Lobo, P. C., 1998. Composição florística e fitossociologia do componente arbóreo das florestas ciliares do rio Lapó, na bacia do rio Tibagi, Tibagi, PR. **Brazilian Journal of Botany**, v. 21, n. 2, p. 183-195. DOI: 10.1590/S0100-84041998000200011;

Dick, C. W.; Etchelecu, G.; Austerlitz, F., 2003. Pollen dispersal of tropical trees (*Dinizia excelsa*: Fabaceae) by native insects and African honeybees in pristine and fragmented Amazonian rainforest. **Molecular Ecology**, v. 12, n. 3, p. 753-764. DOI: 10.1046/j.1365-294X.2003.01760.x;

Donohoe, M., 2003. Causes and health consequences of environmental degradation and social injustice. **Social Science and Medicine**, v. 56, n. 3, p. 573–587. DOI:10.1016/S0277-9536(02)00055-2;

Doyle, J. J.; Doyle, J. L., 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. **Phytochemical Bulletin**, v. 19, p. 11–15;

Ekblom, R.; Galindom J., 2010. Applications of next generation sequencing in molecular ecology of non-model organisms. **Heredity**, v. 107, n. 1, p. 1–15. DOI:10.1038/hdy.2010.152;

Elshire, R. J.; Glaubitz, J. C.; Sun, Q.; Poland, J. A.; Kawamoto, K.; Buckler, E. S. et al., 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species**. PLoS One**, v. 6, p. 1-10. DOI: 10.1371/journal.pone.0019379;

Etter, P. D.; Johnson, E., 2012. RAD paired-end sequencing for local de novo assembly and SNP discovery in non-model organisms. In: **Methods in Molecular Biology**, Clifton – NJ, p. 135-151. DOI: 10.1007/978-1-61779-870-2_9;

Ezard, T. H. G.; Travis, J. M. J., 2006. The impact of habitat loss and fragmentation on genetic drift and fixation time. **Oikos**, v. 114, n. 2, p. 367–375;

Fischer, M.; Matthies, D., 1998. RAPD variation in relation to population size and plant fitness in the rare *Gentianella germanica* (Gentianaceae). **American Journal of Botany**, v. 85, n. 6, p. 811. http://www.ncbi.nlm.nih.gov/pubmed/21684965.

Frankham, R.; Ballou, J. D.; Briscoe, D. A., 2010. **Introduction to Conservation Genetics**. 2nd ed. Cambridge: Cambridge University Press; 644p;

Freeland, J. R.; Kirk, H.; Petersen, S., 2011. **Molecular Ecology**. Chichester, UK: John Wiley & Sons, Ltd. DOI:10.1002/9780470979365;

Funk, W. C.; McKay, J. K.; Hohenlohe, P. A.; Allendorf, F. W., 2012. Harnessing genomics for delineating conservation units. **Trends in Ecology & Evolution**, v. 27, p. 489-496. DOI: 10.1016/j.tree.2012.05.012;

Ganem, R. S., 2010. **Conservação da biodiversidade legislação e políticas públicas**, 437p;

Ganzhorn, S. M.; Perez-Sweeney, B.; Thomas, W. W.; Gaiotto, F. A.; Lewis, J. D., 2015. Effects of fragmentation on density and population genetics of a threatened tree species in a biodiversity hotspot. **Endangered Species Research**, v. 26, p. 189-199. DOI: 10.3354/esr00645;

Gish, W.; States, D. J., 1993. Identification of protein coding regions by database similarity search. **Nature Genetics**, v. 3, p. 266-272. DOI: 10.1038/ng0393-266;

Goudet, J.; Jombart, T., 2015. **hierfstat: Estimation and tests of hierarchical F-statistics**. Available from: https://cran.r-project.org/package=hierfstat;

Greenbaum, G.; Templeton, A. R.; Zarmi, Y.; Bar-David, S., 2014. Allelic richness following population founding events - A stochastic modeling framework incorporating gene flow and genetic drift. **PLoS One**, v. 9, p. 1-23. DOI: 10.1371/journal.pone.0115203;

Hamrick, J. L., 2004. Response of Forest Trees to Global Environmental Changes. **Forest Ecology and Management**, v. 197, p. 323–335. DOI:10.1016/j.foreco.2004.05.023;

Hamrick, J. L.; Nason, J. D., 1996. Consequence of dispersal in plants. In: **Rhodes, O. E.; Ronald, K. C.; Smith, M. H. Population dynamics in ecological space and time**. Chicago; p. 203-235;

Hardy, O. J.; Vekemans, X., 2002. SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. **Molecular Ecology Resources**, v. 2, n.4, p. 618-620. DOI: 10.1046/j.1471-8286.2002.00305.x;

Hey, J.; Pinho, C., 2012. Population genetics and objectivity in species diagnosis. **Evolution**, v. 66, n. 5, p. 1313-1429;

Holderegger, R.; Kamm, U.; Gugerli, F., 2006. Adaptive vs. neutral genetic diversity: Implications for landscape genetics. **Landscape Ecology**, v. 21, n. 6, p. 797–807. DOI:10.1007/s10980-005-5245-9;

Holl, K. D., 2017. Restoring tropical forests from the bottom up. **Science**, v. 355, p. 455-456. DOI: 10.1126/science.aam5432;

Holl, K. D.; Aide, T. M., 2011. When and where to actively restore ecosystems? **Forest Ecology and Management**, v. 261, n. 10, p. 1558-1563. DOI: 10.1016/j.foreco.2010.07.004;

Hughes, A. R.; Inouye, B. D.; Johnson, M. T. J.; Underwood, N.; Vellend, M., 2008. Ecological consequences of genetic diversity. **Ecology Letters**, v. 11, n. 6, 609–623. DOI:10.1111/j.1461-0248.2008.01179.x;

IPEA, 2011. Código Florestal: Implicações do PI 1876/99 nas áreas de reserva legal. **Comunicados do IPEA**;

Joly, C. A.; Aidar, M.; Klink, C.; Mc Grath, D. G.; Moreira, A. G.; Moutinho, P.; Nepstad, D. C. et al., 1999. Evolution of the Brazilian phytogeography classification

systems: Implications for biodiversity conservation. **Ciência e Cultura**, v. 51, p. 331–348. http://ecologia.ib.usp.br/ecovegetal/leituras/CienCultJolyet.pdf;

Jombart, T., 2008. Adegenet: A R package for the multivariate analysis of genetic markers. **Bioinformatics**, v. 24, n. 11, p. 1403-1405. DOI: 10.1093/bioinformatics/btn129;

Jombart, T.; Ahmed, I., 2011. adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. **Bioinformatics**, v. 27, n. 21, p. 3070-3071. DOI: 10.1093/bioinformatics/btr521;

Jump, A. S.; Penuelas, J., 2006. Genetic effects of chronic habitat fragmentation in a wind-pollinated tree. **Proceedings of the National Academy of Sciences**, v. 103, n. 21, p. 8096–8100. DOI:10.1073/pnas.0510127103;

Kahilainen, A.; Puurtinen, M.; Kotiaho, J. S., 2014. Conservation implications of species-genetic diversity correlations. **Global Ecology and Conservation**, v. 2, p. 315-323. DOI: 10.1016/j.gecco.2014.10.013;

Keenan, K.; Mcginnity, P.; Cross, T. F.; Crozier, W. W.; Prodöhl, P. A., 2013. diveRsity: An R package for the estimation and exploration of population genetics parameters and their associated errors. **Methods in Ecology and Evolution**, v. 4, n. 8, p. 782-788. DOI: 10.1111/2041-210X.12067;

Kettenring, K. M.; Mercer, K. L.; Adams, C. R.; Hines, J., 2014. Application of genetic diversity-ecosystem function research to ecological restoration**. Journal of Applied Ecology**, v. 51, n. 2, p. 339-348. DOI: 10.1111/1365-2664.12202;

Kettle, C. J.; Ennos, R. A.; Jaffré, T.; Gardner, M.; Hollingsworth, P. M., 2008. Cryptic genetic bottlenecks during restoration of an endangered tropical conifer. **Biological Conservation**, v. 141, n. 8, p. 1953-1961. DOI: 10.1016/j.biocon.2008.05.008;

Kimball, S.; Lulow, M.; Sorenson, Q.; Balazs, K.; Fang, Y. C.; Davis, S. J. et al., 2015. Cost-effective ecological restoration. **Restoration Ecology**, v. 23, p. 800-810. DOI: 10.1111/rec.12261;

Kirk, H.; Freeland J. R., 2011. Applications and Implications of Neutral versus Non-Neutral Markers in Molecular Ecology. **International Journal of Molecular Sciences**, v. 12, n. 6, p. 3966–3988. DOI:10.3390/ijms12063966;

Klink, C. A.; Machado, R. B., 2005. Conservation of the Brazilian Cerrado. **Conservation Biology**, v. 19, n. 3, p. 707–713. DOI:10.1111/j.1523-1739.2005.00702.x;

Lamb, D., 2011. **Regreening the bare hills**. 1st ed. Dordrecht: Springer Netherlands; 550p. DOI: 10.1007/978-90-481-9870-2;

Lamb, D.; Erskine, P.; Parrotta, J., 2005. Restoration of Degraded Tropical Forest Landscapes. **Science**, v. 310, n. 5754, p. 1628–1632. DOI:10.1126/science.1111773;

Leidner, A. K.; Haddad, N. M., 2011. Combining measures of dispersal to identify conservation strategies in fragmented landscapes. **Conservation Biology**, v. 25, n. 5, p. 1022-1031. DOI: 10.1111/j.1523-1739.2011.01720.x;

Lenormandi, T., 2002. Gene flow and the limit to natural selection. **Trends in Ecology & Evolution**, v. 17, n. 4, p. 183-189. DOI: 10.1016/S0169-5347(02)02497-7;

Leonardi, S.; Piovani, P.; Scalfi, M.; Piotti, A.; Giannini, R.; Menozzi, P., 2012. Effect of habitat fragmentation on the genetic diversity and structure of peripheral populations of beech in Central Italy. **Journal of Heredity**, v. 103, n. 3, p. 408-417. DOI: 10.1093/jhered/ess004;

Letcher, S. G.; Chazdon, R. L., 2009. Rapid recovey of biomass, species richness and species composition in a forest chronosequence in Northeastern Costa Rica. **Biotropica**, v. 41, p. 608-617. DOI: 10.1111/j.1744-7429.2009.00517.x;

Liu, Q.; Guo, Y.; Li, J.; Long, J.; Zhang, B.; Shyr, Y., 2012. Steps to ensure accuracy in genotype and SNP calling from Illumina sequencing data. **BMC Genomics**, v. 13. DOI: 10.1186/1471-2164-13-S8-S8;

Lloyd, H.; Marsden, S. J., 2011. Between-Patch bird movements within a High-Andean polylepis woodland/matrix landscape: Implications for habitat restoration. **Restoration Ecology**, v. 19, p. 74-82. DOI: 10.1111/j.1526-100X.2009.00542.x;

Loiselle, B. A.; Sork, V. L.; Nason, J.; Graham, C., 1995. Spatial genetic structure of a tropical understory shrub, *Psychotria officinalis* (Rubiaceae). **American Journal of Botany**, v. 82, n. 11, p. 1420-1425. Available from: http://www.jstor.org/stable/2445869;

Luikart, G.; England, P. R.; Tallmon, D.; Jordan, S.; Taberlet, P., 2003. The power and promise of population genomics: from genotyping to genome typing. **Nature Reviews Genetics**, v. 4, n. 12, p. 981-994. DOI: 10.1038/nrg1226;

Machado, R. B.; Neto, M. G. P.; Caldas, E. F.; Gonçalves, D. A.; Santos, N. A.; Tabor, K.; Steininger, M., 2004. Estimativas de perda da área do Cerrado brasileiro. **Conservation International do Brasil**, p. 1–23. DOI:10.1590/S0104-42302007000600013;

Markert, J. A.; Champlin, D. M.; Gutjahr-Gobell, R.; Grear, J. S.; Kuhn, A.; McGreevy, T. J.; Roth, A.; Bagley, M. J.; Nacci, D. E., 2010. Population genetic diversity and fitness in multiple environments. **BMC Evolutionary Biology**, v. 10, n. 1, p. 205. DOI:10.1186/1471-2148-10-205;

Marmontel, C. V. F.; Rodrigues, V. A.; Martins, E. T.; Starzynski. R; de Carvalho, J. L., 2013. Caracterização da vegetação secundária do bioma mata atlântica com base em sua posição na paisagem. **Bioscience Journal**, v. 29, n. 6, p. 2042-2052;

Martins, K.; Kimura, R. K.; Francisconi, A. F.; Gezan, S.; Kainer, K.; Christianini, A. V., 2016. The role of very small fragments in conserving genetic diversity of a common tree in a hyper fragmented Brazilian Atlantic forest landscape. **Conservation Genetics**, v. 17, p. 509-520. DOI: 10.1007/s10592-015-0800-7;

Medeiros, R.; Young, C. E. F., 2011. **Contribuição Das Unidades de Conservação Brasileiras Para a Economia Nacional: Relatório Final**;

Mendonça, R C.; Felfili, J. M.; Walter, B. M. T.; Silva Jr., M. C.; Rezende, A. V.; Filgueiras, T. S.; Nogueira, P. E.; Fagg, C. W., 2008. Flora vascular do bioma Cerrado: Checklist com 12.356 espécies. In **Cerrado: Ecologia e Flora,** p. 421–1279. Planaltina - DF: Embrapa Cerrados;

Mijangos, J. L.; Pacioni, C.; Spencer, P. B. S.; Craig, M. D., 2015. Contribution of genetics to ecological restoration. **Molecular Ecology**, v. 24, n. 1, p. 22-37. DOI: 10.1111/mec.12995;

Milligan, B. G.; Leebens-Mack, J.; Strand, A. E., 1994. Conservation Genetics: Beyond the maintenance of marker diversity. **Molecular Ecology**, v. 3, n. 4, p. 423–435. DOI:10.1111/j.1365-294X.1994.tb00082.x;

Minoche, A. E.; Dohm, J. C.; Himmelbauer, H., 2011. Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and Genome Analyzer Systems. **Genome Biology**, v. 12, n. 11. BioMed Central Ltd: R112. DOI:10.1186/gb-2011-12-11-r112;

MMA, 2017a. "**Biodiversidade**." http://www.mma.gov.br/biodiversidade;

MMA, 2017b. "**Cerrado**." http://www.mma.gov.br/biomas/cerrado;

MMA, 2017c. "**Unidades de Conservação."** http://www.mma.gov.br/areas-protegidas/unidades-de-conservacao;

Mona, S.; Ray, N.; Arenas, M.; Excoffier, L., 2014. Genetic consequences of habitat fragmentation during a range expansion. **Heredity**, v. 112, n. 3, p. 291–299. DOI:10.1038/hdy.2013.105;

Myers, N.; Mittermeier, R. A.; Mittermeier, C. G.; da Fonseca, G. A. B.; Kent, J. Biodiversity hotspots for conservation priorities. **Nature**. 2000; 403 (6772): 853-858. DOI: 10.1038/35002501;

Narum, S. R.; Buerkle, C. A.; Davey, J. W.; Miller, M. R.; Hohenlohe, P. A., 2013. Genotyping-by-sequencing in ecological and conservation genomics. **Molecular Ecology**. 2013; 22: 2841-2847. DOI: 10.1111/mec.12350;

Nazareno, A. G.; Bemmels, J. B.; Dick, C. W.; Lohmann, L. G., 2017. Minimum Sample Sizes for Population Genomics: An Empirical Study from an Amazonian Plant Species. **Molecular Ecology** Resources 38 (1): 42–49. doi:10.1111/1755-0998.12654;

Otálora, M. G.; Martínez, I.; Belinchón, R.; Widmer, I.; Aragón, G.; Escudero, A. et al., 2011. Remnants fragments preserve genetic diversity of the old forest lichen Lobaria pulmonaria in a fragmented Mediterranean mountain forest. **Biodiversity Conservation**. 20: 1239-1254. DOI: 10.1007/s10531-011-0025-0;

Pagiola, S.; Ramírez, E.; Gobbi, J.; Haan, C.; Ibrahim, M.; Murgueitio, E.; Ruízm J. P., 2007. "Paying for the Environmental Services of Silvopastoral Practices in Nicaragua." **Ecological Economics** 64 (2): 374–85. doi:10.1016/j.ecolecon.2007.04.014;

Paquette, S. R., 2012. **"PopGenKit: Useful Functions for (Batch) File Conversion and Data Resampling in Microsatellite Datasets."** https://cran.r-project.org/package=PopGenKit;

Peñaloza-Ramírez, J. M.; Aguilar-Amezquita, B.; Núñez-Farfán, J.; Pérez-Nasser, N.; Albarrán-Lara, A. L.; Oyama, K., 2016. Consequences of Habitat Fragmentation on Genetic Structure of Chamaedorea Alternans (Arecaceae) Palm Populations in the Tropical Rain Forests of Los Tuxtlas, Veracruz, Mexico. **Revista Mexicana de Biodiversidad** 87 (3): 990–1001. doi:10.1016/j.rmb.2016.07.004;

Peterson, B. K.; Weber, J. N.; Kay, E. H.; Fisher, H. S.; Hoekstra, H. E., 2012. Double Digest RADseq: An Inexpensive Method for de Novo SNP Discovery and Genotyping in Model and Non-Model Species**. PLoS ONE** 7 (5). doi:10.1371/journal.pone.0037135;

Pruett, C. L.; Winker, K., 2008. The Effects of Sample Size on Population Genetic Diversity Estimates in Song Sparrows Melospiza Melodia. **Journal of Avian Biology** 39 (June 2007): 252–56. doi:10.1111/j.2008.0908-8857.04094.x;

R Core Team, 2017. **R: A Language and Environment for Statistical Computing**. Vienna, Austria: R Foundation for Statistical Computing. Available from: https://www.r-project.org;

Ramalho, M. 2004. Stingless Bees and Mass Flowering Trees in the Canopy of Atlantic Forest: A Tight Relationship. **Acta Botanica Brasilica** 18 (1): 37–47. doi:10.1590/S0102-33062004000100005;

Rao, I. M.; Miles, J. W.; Beebe, S. E.; Horst, W. J., 2016. Root adaptations to soils with low fertility and aluminium toxicity. **Annals of Botany**, 118(4), 593–605. http://doi.org/10.1093/aob/mcw073;

Reed, D. H.; Frankham, R., 2003. Correlation between Fitness and Genetic Diversity. **Conservation Biology**. 17 (1): 230-237. DOI: 10.1046/j.1523-1739.2003.01236.x;

Reichmann, M. C.; Zanella, C. A.; Valério Júnior, C.; Borges, A. C. P.; Sausen, T. L.; Paroul, N.; Mielniczki-Pereira, A. A.; Teixeira, A. J.; Budke, J. C.; Mossi, A. J.; Cansian, R. L., 2017. Genetic diversity in populations of Maytenus dasyclada (Celastraceae) in forest reserves and unprotected Araucaria forest remnants. **Acta Botanica Brasilica**, 31(1), 93-101. https://dx.doi.org/10.1590/0102-33062016abb0428;

Reznick, D.; Ricklefs, R. E. 2009. Darwin's bridge between microevolution and macroevolution. **Nature** 457:837– 842;

Ribeiro, J. F.; Bridgewater, S.; Ratter, J. A.; Sousa-Silva, J. C., 2005. Ocupação do bioma Cerrado e conservação da sua diversidade vegetal. In: Scariot, A.; Sousa-Silva, J. C.; Felfili, J. M. (Org.). **Cerrado: ecologia, biodiversidade e conservação**. Brasília: Ministério do Meio Ambiente. p.383-399;

Ribeiro, M. C.; Metzger, J. P.; Martensen, A. C.; Ponzoni, F. J.; Hirota, M. M., 2009. The Brazilian Atlantic Forest: How much is left, and how is the remaining forest distributed? Implications for conservation. **Biological Conservation**. 142 (6): 1141-1153. DOI: 10.1016/j.biocon.2009.02.021;

Ritchie, A. L.; Krauss, S. L., 2012. A Genetic assessment of ecological restoration success in banksia attenuata. **Restoration Ecology**. 20 (4): 441-449. DOI: 10.1111/j.1526-100X.2011.00791.x;

Rocha, Y. T.; Presotto, A.; Cavalheiro, F., 2007. The Representation of Caesalpinia Echinata (Brazilwood) in Sixteenth-and-Seventeenth-Century Maps. **Anais da Academia Brasileira de Ciências** 79 (4): 751–65. doi:10.1590/S0001-37652007000400014;

Rodrigues, R. R.; Lima, R. A. F.; Gandolfi, S.; Nave, A. G., 2009. On the restoration of high diversity forests: 30 years of experience in the Brazilian Atlantic Forest. **Biological Conservation**. 142 (6): 1242-1251. DOI: 10.1016/j.biocon.2008.12.008;

Sánchez-Calderón, L.; Ibarra-Cortés, M. E.; Zepeda-Jazo, I., 2013. Root Development and Abiotic Stress Adaptation, **Abiotic Stress - Plant Responses and Applications in Agriculture**, Dr. Kourosh Vahdati (Ed.), InTech, DOI: 10.5772/55043;

Sato, T.; Isagi, Y.; Sakio, H.; Osumi, K.; Goto, S., 2006. Effect of gene flow on spatial genetic structure in the riparian canopy tree Cercidiphyllum japonicum revealed by microsatellite analysis. **Heredity**. 96: 79-84. DOI: 10.1038/sj.hdy.6800748;

Schaberg, P. G.; DeHayes, D. H.; Hawley, G. J.; Nijensohn, S. E., 2008. Anthropogenic alterations of genetic diversity within tree populations: Implications for forest ecosystem resilience. **Forest Ecology Management**. 256: 855-862. DOI: 10.1016/j.foreco.2008.06.038;

Schlötterer, C., 2004. Opinion: The Evolution of Molecular Markers — Just a Matter of Fashion? **Nat Rev Gen** 5 (1): 63–69. doi:10.1038/nrg1249;

Seeb, J. E.; Carvalho, G.; Hauser, L.; Naish, K.; Roberts, S.; Seeb, L. W., 2011. Single-Nucleotide Polymorphism (SNP) Discovery and Applications of SNP Genotyping in Nonmodel Organisms. **Molecular Ecology Resources** 11 (SUPPL. 1): 1–8. doi:10.1111/j.1755-0998.2010.02979.x;

Sezen, U. U., 2005. Genetic Consequences of Tropical Second-Growth Forest Regeneration. **Science**. 307: 891-891. DOI: 10.1126/science.1105034;

Silva, A. F.; Oliveira, R. V.; Santos, N. R. L.; De Paula, A., 2003. Composição florística e grupos ecológicos das espécies de um trecho de floresta semidecídua submontana da Fazenda São Geraldo, Viçosa-MG. **Revista Árvore**. 27 (3): 311-319. DOI: 10.1590/S0100-67622003000300006;

Silva, M. A. S.; Ming, L. C.; Pereira, A. M. S.; Bertoni, B. W.; Batistini, A. P.; Pereira, P. S., 2006. Phytochemical and Genetic Variability of *Casearia sylvestris* Sw. from Sao Paulo State Atlantic Forest and Cerrado Populations. **Revista Brasileira de Plantas Medicinais** 8: 159–66. http://www.sbpmed.org.br/download/issn_06_4/8esp_159_166.pdf;

Simon, M. F., R. Grether, L. P. de Queiroz, C. Skema, R. T. Pennington, and C. E. Hughes. 2009. “Recent Assembly of the Cerrado, a Neotropical Plant Diversity Hotspot, by in Situ Evolution of Adaptations to Fire.” **Proceedings of the National Academy of Sciences** 106 (48): 20359–64. doi:10.1073/pnas.0903410106;

Sims, D.; Sudbery, I.; Ilott, N. E.; Heger, A.; Ponting, C. P., 2014. Sequencing Depth and Coverage: Key Considerations in Genomic Analyses. **Nature Reviews Genetics**, 15 (2). Nature Publishing Group: 121–32. doi:10.1038/nrg3642;

Slatkin, M., 1985. Gene Flow in Natural Populations. **Annual Review of Ecology and Systematics** 16 (1): 393–430. doi:10.1146/annurev.es.16.110185.002141;

Slatkin, M., 1987. Gene Flow and the Geographic Structure of Natural Populations. **Science** 236 (4803): 787–92;

Sleumer, H. O., 1980. Flacourtiaceae. **In: Flora Neotropica**. New York: The New York Botanical Garden;

SOS Mata Atlantica, 2017. **Florestas - A Mata Atlântica**. Available from: https://www.sosma.org.br/nossa-causa/a-mata-atlantica/;

SOS Mata Atlantica, INPE, 2015. Atlas dos remanescentes florestais da Mata Atlântica período 2013-2014. **Instituto Nacional de Pesquisas Espaciais**. Available

from: http://mapas.sosma.org.br/site_media/download/atlas_2013-2014_relatorio_tecnico_2015.pdf;

Spetea, C.; Rintama, E., 2014. Changing the Light Environment: Chloroplast Signalling and Response Mechanisms. **Philosophical Transactions B**, 1: 1–3;

Star, B.; Spencer, H. G., 2013. Effects of Genetic Drift and Gene Flow on the Selective Maintenance of Genetic Variation. **Genetics** 194 (1): 235–44. doi:10.1534/genetics.113.149781;

Tabarelli, M.; Aguiar, A. V.; Ribeiro, M. C.; Metzger, J. P.; Peres, C. A., 2010. Prospects for Biodiversity Conservation in the Atlantic Forest: Lessons from Aging Human-Modified Landscapes. **Biological Conservation** 143 (10). Elsevier Ltd: 2328–40. doi:10.1016/j.biocon.2010.02.005;

Tabarelli, M.; Pinto, L. P.; Silva, J. M. C.; Costa, C. M. R., 2003. The Atlantic Forest of Brazil: Endangered Species and Conservation Planning. **In: The Atlantic Forest of South America: Biodiversity Status, Trends, and Outlook,** edited by C Galindo-Leal and I G Camara, 86–94. Washington, D.C.: Center for Applied Biodiversity Science, Island Press;

Tabarelli, M.; Villani, J. P.; Mantovani, W., 1993. Aspectos da sucessão secundária em trecho da Floresta Atlântica no Parque Estadual da Serra do Mar, SP. **Revista do Instituto Florestal**, São Paulo. 5 (1): 95-112. Available from: http://www.iflorestal.sp.gov.br/RIF/RevistaIF/RIF5-1/RIF5-1_99-112.pdf;

Thomas, E.; Jalonen, R.; Loo, J.; Boshier, D.; Gallo, L.; Cavers, S. et al., 2014. Genetic considerations in ecosystem restoration using native tree species. **Forest Ecology Management**. 333: 66-75. DOI: 10.1016/j.foreco.2014.07.015;

Toro, M. A.; Caballero, A., 2005. Characterization and conservation of genetic diversity in subdivided populations. **Philosophical Transactions of the Royal Society B: Biological Sciences.** 2005; 360: 1367-1378. DOI: 10.1098/rstb.2005.1680;

Torres, R. B.; Yamamoto, K., 1986. Taxonomy of the species of *Casearia* Jacq. (Flacourtiaceae) from the State of Sao Paulo. **Revista Brasileira de Botânica**, v. 9, n. 2, p. 239–258;

Vekemans, X.; Hardy, O. J., 2004. New insights from fine-scale spatial genetic structure analyses in plant populations. **Molecular Ecology**, v. 13, p. 921-935. DOI: 10.1046/j.1365-294X.2004.02076.x;

Vranckx, G.; Jacquemyn, H.; Muys, B.; Honnay, O., 2012. Meta-analysis of susceptibility of woody plants to loss of genetic diversity through habitat fragmentation. **Conservation Biology**, v. 26, p. 228-237. DOI: 10.1111/j.1523-1739.2011.01778.x;

Wanderley, M. G. L.; Shepherd, G. J.; Giulietti, A. M.; Melhem, T. S.; Kameyama, C.; Bittrich, V., 2002. **Flora Fanerogâmica do Estado de São Paulo**. Vol. 2. 386p. Available from: http://botanica.sp.gov.br/ffesp_online/;

Werf, J. V. D.; Graser, H.; Frankham, R.; Gondro, C., 2009. **Adaptation and fitness in animal populations**. Springer Ebooks. DOI:10.1007/978-1-4020-9005-9.

Wickham, H., 2009. **ggplot2: Elegant graphics for data analysis**. Springer-Verlag New York. Available from: http://ggplot2.org;

Willi, Y.; Buskirk, J. V.; Hoffmann, A. A., 2006. Limits to the adaptive potential of small populations. **Annual Review of Ecology, Evolution, and Systematics**, v. 37, n. 1, p. 433–58. DOI:10.1146/annurev.ecolsys.37.091305.110145;

Willi, Y.; Buskirk, J. V.; Schmid, B.; Fischer, M., 2007. Genetic isolation of fragmented populations is exacerbated by drift and selection. **Journal of Evolutionary Biology**, v. 20, n. 2, p. 534–542. DOI:10.1111/j.1420-9101.2006.01263.x;

Wisely, S. M.; Buskirk, S. W.; Fleming, M. A.; McDonald, D. B.; Ostrander, E. A., 2002. Genetic diversity and fitness in black-footed ferrets before and during a bottleneck. **The Journal of Heredity**, v. 93, n. 4, p. 231–237. DOI:10.1093/jhered/93.4.231;

Wright, S. J., 2010. The Future of Tropical Forests. **Annals of the New York Academy of Sciences**, v. 1195, p. 1–27. DOI:10.1111/j.1749-6632.2010.05455.x;

Xu, P.; Xu, S.; Wu, X.; Tao, Y.; Wang, B.; Wang, S.; Qin, D.; Lu, Z.; Li, G., 2014. Population genomic analyses from low-coverage RAD-Seq data: A case study on the non-model cucurbit bottle gourd. **Plant Journal**, v. 77, n. 3, p. 430–442. DOI:10.1111/tpj.12370;

Yuan, N.; Comes, H. P.; Mao, Y. R.; Qi, X. S.; Qiu, Y. X., 2012. Genetic effects of recent habitat fragmentation in the Thousand-Island Lake Region of Southeast China on the distylous herb *Hedyotis chrysotricha* (Rubiaceae). **American Journal of Botany**, v. 99, n. 10, p. 1715–1725. DOI:10.3732/ajb.1200054.

ANEXOS

Declaração

As cópias de artigos de minha autoria ou de minha co-autoria, já publicados ou submetidos para publicação em revistas científicas ou anais de congressos sujeitos a arbitragem, que constam da minha Dissertação/Tese de Mestrado/Doutorado, intitulada **Genômica da conservação de Casearia sylvestris Sw. no Cerrado e Mata Atlântica do Estado de São Paulo**, não infringem os dispositivos da Lei n.° 9.610/98, nem o direito autoral de qualquer editora.

Campinas, 20 de outubro de 2017

Assinatura : _____
Nome do(a) autor(a): **João Paulo Gomes Viana**
RG n.° 2.620.062

Assinatura : _____
Nome do(a) orientador(a): **Maria Imaculada Zucchi**
RG n.° 19.810.213-6

COORDENADORIA DE PÓS-GRADUAÇÃO
INSTITUTO DE BIOLOGIA
Universidade Estadual de Campinas
Caixa Postal 6109. 13083-970, Campinas, SP, Brasil
Fone (19) 3521-6378. email: cpgib@unicamp.br

DECLARAÇÃO

Em observância ao **§5º do Artigo 1º da Informação CCPG-UNICAMP/001/15**, referente a Bioética e Biossegurança, declaro que o conteúdo de minha Tese de Doutorado, intitulada "***Genômica da conservação de Casearia sylvestris Sw. no Cerrado e Mata Atlântica do Estado de São Paulo***", desenvolvida no Programa de Pós-Graduação em Genética e Biologia Molecular do Instituto de Biologia da Unicamp, não versa sobre pesquisa envolvendo seres humanos, animais ou temas afetos a Biossegurança.

Assinatura: _____
Nome do(a) aluno(a): João Paulo Gomes Viana

Assinatura: _____
Nome do(a) orientador(a): Maria Imaculada Zucchi

Data: 20 de outubro de 2017