

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE BIOCÊNCIAS
DEPARTAMENTO DE GENÉTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E BIOLOGIA MOLECULAR

**Análise da variação genética adaptativa em populações naturais de
Eugenia uniflora L. (Myrtaceae)**

NICOLE MOREIRA VETO

Orientadora: Prof.^a Dr.^a Andreia Carina Turchetto Zolet

Co-orientador: Prof.^o Dr. Rogerio Margis

Porto Alegre, novembro de 2015.

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE BIOCÊNCIAS
DEPARTAMENTO DE GENÉTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E BIOLOGIA MOLECULAR

**Análise da variação genética adaptativa em populações naturais de
Eugenia uniflora L. (Myrtaceae)**

NICOLE MOREIRA VETO

Dissertação submetida ao Programa de Pós-graduação em Genética e Biologia Molecular da Universidade Federal do Rio Grande do Sul como requisito parcial para a obtenção do grau de Mestre em Genética e Biologia Molecular.

Orientadora: Prof.^a Dr.^a Andreia Carina Turchetto Zolet

Co-orientador: Prof.^o Dr. Rogerio Margis

Porto Alegre, novembro de 2015.

“Desistir... Eu já pensei seriamente nisso, mas nunca me levei realmente a sério. É que tem mais chão nos meus olhos do que o cansaço nas minhas pernas, mais esperança nos meus passos do que tristeza nos meus ombros, mais estrada no meu coração do que medo na minha cabeça.”

(Geraldo Eustáquio de Souza)

INSTITUIÇÕES E FONTES FINANCIADORAS

O presente trabalho foi executado no Laboratório de Genética Vegetal do Departamento de Genética da Universidade Federal do Rio Grande do Sul (UFRGS) e teve como fontes financiadoras a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e o Programa de Pós-graduação em Genética e Biologia Molecular (PPGBM) da UFRGS.

AGRADECIMENTOS

Normalmente, os agradecimentos acabam sendo nossas palavras findas, ao término de cada trabalho. Acredito que deva ser para lembrar-nos que mesmo até o último minuto não trabalhamos sós.

Àqueles que estiveram ao meu lado diariamente no **Laboratório de Genética Vegetal**, por todas as minhas incansáveis perguntas técnicas e filosóficas. Pelas alegrias, pipocas, cafés, neuroses e aprendizado que foram das bactérias até os cromossomos. Em especial à Camila M. **Zanella**, pelos debates, risos, pela parceria, compreensão e pipetas. À Joana **Fochezatto** por compartilhar histórias de um mundo particular, Paula **Burchardt** por ter sido minha referi de última hora, sou muito feliz por ter conhecido vocês a tempo. Gabi **Fernandes** e Fernanda **Moreira** por todos os nossos carnavais e energia positiva que me manteve até agora. Aos professores que participaram ativamente da minha formação, meu orgulho por estar próxima de vocês. Aos ICs Débora e Krause pela cía e auxílio na bancada, foi muito bom tê-los comigo. Ao triângulo mais energético do PPGBM Tiago **Ferraz** e Giovanna **Giudicelli**, por tudo que fizemos e pelo que somos quando estamos juntos. À doce Caroline **Turchetto**, pelo exemplo em sua competência profissional, pelas sugestões acadêmicas, pelos seus cadernos fantásticos, além da “queridice” com a qual me recebe e me incentiva diariamente. À querida Fran **Kulcheski**, pela ajuda nas intermináveis análises de amplificação. Ao Gran Elmo **Cardoso** por ser esse cara sensacional, faça chuva ou faça sol. Ao PPGBM, UFRGS e Capes pelo auxílio financeiro e estrutura física que tornaram este trabalho possível.

Aos colegas e amigos geográficos do CPC que recarregaram minhas baterias durante estes dois anos e aos professores Ulisses **Bremer**, Francisco **Aquino** e Jefferson **Simões** por deixarem as portas abertas e por me motivarem a estar por perto, mesmo estando longe. Ricardo Burgo **Braga** pelo pontapé inicial e minha vida acadêmica e Denilson **Ribeiro**, pela ajuda com os tão lindos dados climáticos. Vocês (todos) não fazem ideia de o quanto sou feliz por conhecer e participar deste seletto grupo.

Aos amigos que foram os responsáveis pela mobilização em prol do “concerto” do meu braço direito, devo este trabalho a vocês também. Sem que ele estivesse tão bem operado, não conseguiria. GabiGabi, tu és uma dessas estrelas que fazem meu céu brilhar e uma das responsáveis por essa ser uma passagem que divide a minha vida entre “antes” e “depois”.

À todos os outros com quem tenho a sorte de conviver, Juliana de Paula, Faten, Tomazzinha, Hiago, Alexandre, Alisson, Luiza, Eduardo, GP6, Leonardo, Janine , Brites e Franklin: mesmo com toda minha chatice, faço TUDO por vocês. Um agradecimento especial à Ana Paula, minha amiga desde as séries iniciais, com quem tenho o prazer de compartilhar essa caminhada com tantos sonhos e objetivos, que tenho certeza que vamos realizar. Obrigada por serem o que são!

Ao meu co-orientador **Rogério Margis**, por trabalharmos novamente juntos e pelo que representa profissional e cientificamente. Ao **Frank Guzman**, pelo árduo trabalho de Bioinformática, que nos possibilitou encontrar os SNPs, tua participação foi fundamental! À **Andreia Turchetto**, como falar, minha amada? Ilustre? Sensacional? Adorada? Incrível? Excepcional? Orientadora e colega de profissão. Por termos retomado nossa relação, que vem desde os tempos da iniciação científica. Por todas as oportunidades de aprendizado, crescimento e amadurecimento, pela confiança e motivação, pelas doses diárias de bom humor, respeito, compreensão e dedicação. Pela oportunidade que me deste, de colocar em prática minha tão sonhada vida acadêmica, por entender e fazer parte dos meus sonhos de vida, objetivos pessoais e profissionais. Se fosse falar do tamanho da felicidade que vivi neste período, não caberia aqui.

Aos meus primos **Anne Marie** e **Luiz Gustavo**, que embora morem tão longe, sempre estiveram TÃO perto. Anne, eu sei, você sabe....”Desenha e vai”, nós fomos!!! Tavo, és um orgulho e um grande motivador que atua junto com aquele grito que vem de dentro, a força maior, o que nos move!

À minha família, guiada pela grandeza e luz de minha Mãe **Maria**, que me ensinou a sonhar, buscar, tentar e conseguir, que continua ao meu lado e a quem ainda espero despertar muitos sorrisos e orgulho. Ao meu pai Jean, pelo incentivo em ler e estudar para formarmos as bases sólidas que nós três formamos. Às minhas irmãs **Bruna** e **Rafaela** por tudo! Porque vocês são meu tudo. “Só enquanto eu respirar”. Às inclassificáveis sobrinhas Manu, Manúscula e BebêBetina, o amor incondicional e a gratidão de ser tia e surpresa do aprendizado com vocês, da reflexão sobre o mundo que vem pela frente e a auto crítica que isso tudo me causa. Juntos podemos tudo e o nosso amor é um exemplo disso.

Não posso, nem por um minuto, sentir-me sozinha, com tanta gente especial por perto. Vamos em frente!

SUMÁRIO

AGRADECIMENTOS.....	5
RESUMO	8
ABSTRACT	10
1. INTRODUÇÃO	11
1.1 Floresta Atlântica: Local de estudo	11
1.2 A família Myrtaceae: grupo de estudo.....	13
1.4 Genética adaptativa, seleção natural e genômica	18
2. OBJETIVOS	21
2.1 Objetivo geral	21
2.2 Objetivos específicos.....	21
3. ESTRUTURA E ORGANIZAÇÃO DA DISSERTAÇÃO	22
4. CAPÍTULO 1: Identifying signature of positive selection in <i>Eugenia uniflora</i> natural populations genes through SNP analysis	23
Nicole M. Veto, Débora Bublitz, Felipe Augusto Krause, Fabiano Salgueiro, Frank Lino Guzman Escudero, Denílson Ribeiro, Francisco Eliseu Aquino, Jefferson Cardia Simões, Giovanni G. Vendramin, Rogério Margis, Andréia Carina Turchetto Zolet.....	23
References.....	41
REFERÊNCIAS INTRODUÇÃO E DISCUSSÃO	53

RESUMO

O conhecimento sobre a história evolutiva de uma espécie é essencial para obter *insights* sobre processos que contribuíram para os padrões atuais de distribuição e diversidade dessa espécie. Entender a evolução de uma espécie envolve a compreensão de sua história demográfica, estruturação populacional e também os mecanismos envolvidos na adaptação local. Neste contexto, os avanços em genômica e tecnologia de sequenciamento de DNA estão revolucionando o entendimento da história evolutiva das espécies, pois possibilitam o desenvolvimento de plataformas de genotipagem de alto rendimento, facilitando o uso de análises multilocus para identificar as regiões genômicas responsáveis pela adaptação local e trazendo uma visão detalhada sobre a história demográfica e a estrutura de populações. Além disso, estes dados permitem a identificação de loci responsáveis ou associados com adaptações locais e compreender as estratégias de adaptação que possibilitam algumas plantas sobreviverem em diferentes condições ambientais. *Eugenia uniflora* L. é uma espécie pertencente à família Myrtaceae e que apresenta uma ampla distribuição ao longo dos domínios da Floresta Atlântica (DFA). Popularmente conhecida como pitanga ou cereja-brasileira, *E. uniflora* é bastante versátil e cresce em vários habitats diferentes, ocorrendo como um arbusto ou árvore pequena em ambientes arenosos de restinga na planície costeira, próximo ao oceano e também no sul do Brasil apresenta-se como árvore, em ambientes de mata ciliar. Essas características, aliadas ao fato de que esta espécie seja uma das espécies-chave presentes em diferentes regiões fitogeográficas que integram a DFA, fazem com que ela seja um excelente modelo para estudos de variação adaptativa. Dessa forma, a presente dissertação teve como objetivo o estudo da variação genética adaptativa em populações naturais de *E. uniflora*. Para isso, foram selecionados 30 genes candidatos envolvidos em diferentes vias de sinalização com as respostas ao estresse abiótico, através da análise do transcriptoma da folha de *E. uniflora*. As sequências desses genes foram amplificadas por PCR em 96 indivíduos provenientes de populações de ambientes contrastantes. Os fragmentos amplificados foram submetidos ao sequenciamento usando plataforma de Sequenciamento de Nova Geração para a identificação e genotipagem de marcadores SNPs (*Single Nucleotide Polimorphism*). Após a análise do sequenciamento, um total de 381 SNPs, distribuídos em 16 genes foram identificados e genotipados em 84 indivíduos. A análise de

F_{ST} outlier revelou dois loci sob seleção positiva, um no gene que codifica uma desidrina e outro no gene que codifica uma Lea14-A, genes estes previamente reportados estarem envolvidos com estresse de resposta à seca em plantas.

ABSTRACT

The knowledge about the evolutionary history of a species is essential to gain insight into processes that contributed to the current patterns of distribution and diversity of this species. It involves the understanding of demographic history, population structure and also the mechanisms involved in local adaptation. In this context, advances in genomics and DNA sequencing technology are revolutionizing the understanding of the evolutionary history of the species, as they allow the development of high-throughput genotyping platforms, facilitating the use of multilocus analysis to identify genomic regions responsible for local adaptation. It also gives detailed insight into the demographic history and population structure. Furthermore, these data allow the identification of loci responsible for or associated with local adaptations and comprises adaptation strategies that enable some plants to survive in different environmental conditions. *Eugenia uniflora* L. is a species belonging to Myrtaceae family, wide distributed over the Atlantic Forest Domain (AFD). Popularly known as Pitanga or Brazilian-cherry, it is quite versatile and grows in distinct habitats, occurring as a shrub or small tree in sandy environments (Restinga) on the coastal plain near the ocean and as a tree in southern Brazil (riparian forest). These characteristics, combined with the fact that this species is one of the key species present in different phytogeographic regions of the AFD, make this species an excellent model for studies of adaptation. Thus, the present work aims to study the adaptive genetic variation in natural populations of *E. uniflora*. For this, we selected 30 candidate genes involved in different signaling pathways with the responses to abiotic stress by transcriptome analysis of *E. uniflora*. The sequences of these genes were amplified by PCR in 96 individuals from populations of different environments. The amplified fragments were subjected to sequencing using Next Generation Sequencing platform for the identification and genotyping of SNP markers (Single Nucleotide Polymorphism). After sequencing analysis, a total of 381 SNPs spread on 16 genes were identified and genotyped in 84 individuals. The F_{ST} outlier analysis revealed two loci under positive selection, one present in a gene encoding a dehydrin and another in a gene encoding a Lea14-A. These two genes were previously reported to be involved in drought stress response in plants.

1. INTRODUÇÃO

1.1 Floresta Atlântica: Local de estudo

O continente Sul-americano possui ampla extensão territorial no sentido norte-sul e está localizado em uma importante área geográfica, a zona intertropical ocidental, pois recebe influência de importantes zonas climáticas. A América do Sul também apresenta uma variedade de climas: tropical, subtropical e temperado, representados nas porções equatorial, sul e extremo sudoeste, respectivamente (Fittkau 1969; Mayle et al. 2009; Aragón et al. 2011). Além disso, apresenta complexos perfis geomorfológicos, atualmente constituídos por planícies, planaltos e uma cadeia de montanhas, denominada Cordilheira dos Andes. Todas estas características refletem particularidades biogeográficas da América do Sul, diversamente composta por diferentes biomas e ecoregiões, que a torna o continente com maior biodiversidade da Terra (Morrone 2004; Morrone 2006; Aragón et al. 2011).

Compondo este interessante cenário, está a Floresta Atlântica, a qual é considerada a segunda maior floresta tropical na América do Sul, cobrindo uma área de mais de um milhão de quilômetros quadrados ao longo da costa brasileira e se estendendo até o leste do Paraguai e nordeste da Argentina (Figura 1) (Rotman 1995; Joly et al. 1999; Oliveira-Filho & Fontes 2000; Ribeiro et al. 2009). A Floresta Atlântica abriga uma das maiores riquezas e endemismo de espécies do mundo, incluindo uma grande diversidade de fauna e flora, distribuída em diferentes tipos de habitat. Conforme o Instituto Brasileiro de Geografia e Estatística (IBGE 1988) a Floresta Atlântica é formada por um conjunto de diferentes formações vegetacionais: Florestas Ombrófila Densa, Ombrófila Aberta e Mista, Floresta Estacional Decidual e Semidecidual, campos de altitude, Manguezais, Restingas e Dunas. Tais formações compõem o que vem sendo chamado de Domínios da Floresta Atlântica (DFA) (Joly et al. 1999). Pesquisas de ocorrências de plantas terrestres na Floresta Atlântica demonstraram que muitas espécies podem ocupar um ou mais nichos fitogeográficos dentro dos domínios da Floresta Atlântica, com incidência de altas taxas de endemismo local, mas também espécies que são amplamente distribuídas em mais de uma região fitoecológica (Stehmann 2009; Barros & Morim 2014). Mesmo considerando a importância da DFA, com parte da sua biodiversidade ainda desconhecida, existe um problema bastante sério em termos de conservação dessa ecoregião, pois ela tem sido

relatada como um dos ecossistemas mais ameaçados da Terra devido à perturbação intensa, tendo sido reduzida a apenas 7% de sua área original, que se mantém preservada (Colombo & Joly 2010). Por ser considerada um dos *hotspots* de biodiversidade do mundo e por esta perda considerável devido às pressões antrópicas, a Floresta Atlântica é considerada uma prioridade de conservação (da Fonseca 1985; Myers et al. 2000).

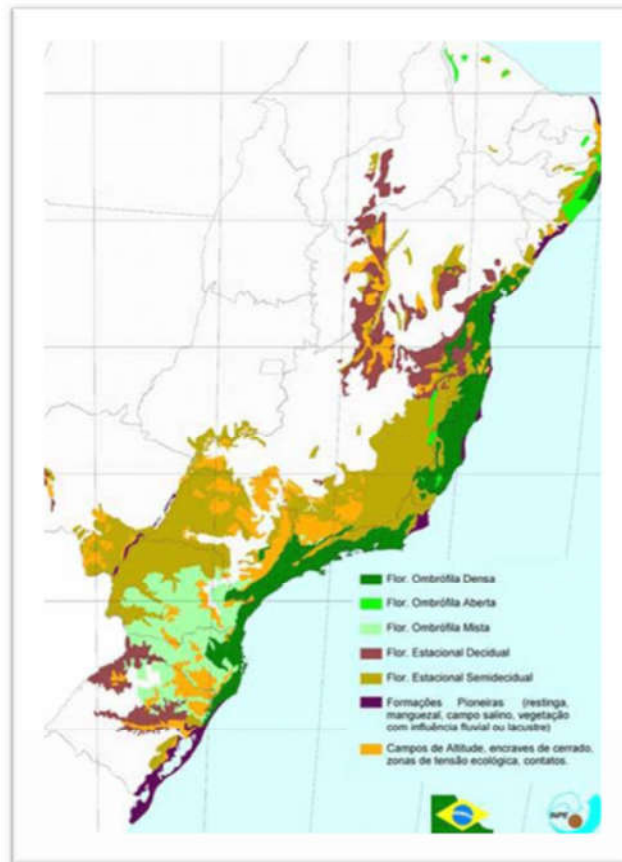


Figura 1. Cobertura original da Floresta Atlântica e seus diferentes domínios.

A Floresta Atlântica tem sido alvo de muitos estudos que visam entender a história evolutiva das espécies que habitam esse bioma, a origem da biodiversidade e endemismo, bem como o entendimento sobre a história biogeográfica nesta região. Estudos Paleopalínológicos (Behling 1997; Behling 1998; Lichte & Behling 1999; Behling 2002) e Paleoclimáticos (Carnaval & Moritz 2008; de Mello Martins 2011) têm mostrado interessantes padrões de fragmentação, retração e identificação de refúgios na Floresta Atlântica, influenciados pelos eventos de oscilações climáticas passadas e eventos orogênicos. Por outro lado, estudos filogeográficos e de genética de populações têm

contribuído para o entendimento sobre a história demográfica das populações, identificação de refúgios e zonas de contato, migração, identificação de centros de diversidade genética, barreiras para o fluxo gênico e respostas das espécies aos efeitos das oscilações climáticas passadas e eventos orogênicos (de Mello Martins 2011; Turchetto-Zolet et al. 2013). Apesar do crescente aumento nos estudos de espécies na Floresta Atlântica nos últimos anos, ainda existem muitas perguntas a serem respondidas. Uma questão interessante e que ainda tem sido pouco explorada é o entendimento sobre mecanismos adaptativos de espécies presentes nas diferentes regiões fitogeográficas que integram a DFA.

1.2 A família Myrtaceae: grupo de estudo

A família Myrtaceae inclui aproximadamente 142 gêneros e cerca de 5500 espécies (Wilson 2011). As espécies dessa família são predominantemente encontradas no hemisfério sul, incluindo Austrália, América do Sul, Sudeste da Ásia, África, Índia, Nova Caledônia, e muitas ilhas do Pacífico, que se estende para o norte até o Havaí. Até pouco tempo eram aceitas duas subfamílias: Leptospermoideae, caracterizada por frutos secos, folhas alternas e distribuição na Oceania; e Myrtoideae composta de frutos carnosos, folhas opostas e distribuição Pantropical (Legrand, Klein & Myrciaria 1978; Sobral 2003). Porém, estudos filogenéticos (Wilson et al. 2005; Wilson 2011) sugerem a reorganização da família, considerando como subfamílias Psiloxylloideae e Myrtoideae. A família Myrtaceae é bastante representativa em Biomas neotropicais incluindo a Floresta Atlântica (Figura 2), sendo uma das mais abundantes em espécies lenhosas no Brasil, contendo 927 espécies distribuídas em 24 gêneros. Stehmann (2009) e Sobral et al (2015) mostraram que Myrtaceae é a família de plantas lenhosas com mais endemismo na Floresta Atlântica. A ampla distribuição da família Myrtaceae e a fácil adaptabilidade nos diferentes domínios da Floresta Atlântica são características que facilitam o estudo e o entendimento das dinâmicas ecológicas e climáticas que envolvem a Floresta Atlântica. Essas características tornam as espécies dessa família modelos para diversos estudos, incluindo estudos de genética adaptativa em plantas.

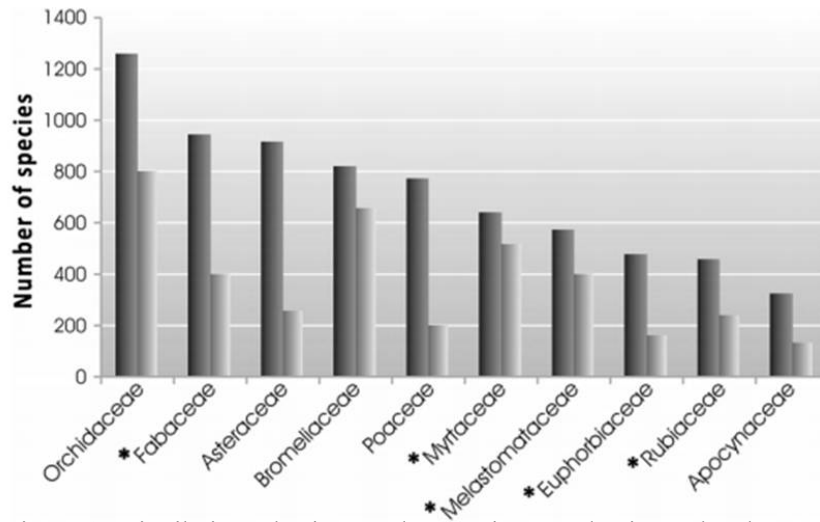


Figura 2. Distribuição da riqueza de espécies e endemismo das dez mais importantes famílias de angiospermas na Floresta Atlântica. As barras escuras representam total de espécies, barras claras representam espécies endêmicas, famílias lenhosas estão marcadas com um asterisco (Stehmann et al 2009; Lucas & Bünger 2015).

1.3 *Eugenia uniflora* L., uma espécie da família Myrtaceae modelo para estudos evolutivos na Floresta Atlântica

Eugenia L. é um dos gêneros da família Myrtaceae mais ricos em espécies, apresentando mais de 1050 espécies (Mazine et al. 2014). Este gênero inclui espécies de árvores ou arbustos com uma distribuição pantropical, apresenta dificuldades taxonômicas e uma grande importância ecológica. No Brasil, o gênero *Eugenia* possui aproximadamente 356 espécies, sendo cerca de 274 endêmicas (Sobral et al. 2015). É um gênero bastante representativo também na Floresta Atlântica com aproximadamente 241 espécies (Mazine 2006; Stehmann 2009; Mazine et al. 2014). *Eugenia uniflora* L. é uma espécie do gênero *Eugenia* proximamente relacionada com *Eugenia brasiliensis*, *E. neomyrtifolia* Sobral, *E. sulcata* Spring ex Mart. e *E. pitanga* Kiaersk. (Mazine et al. 2014). É uma espécie amplamente distribuída em ambientes tropicais e subtropicais da América do Sul. Nativa da Floresta Atlântica, ocorre naturalmente no Paraguai, Brasil Argentina e Uruguai. No Brasil, sua distribuição surge desde nordeste e sudeste, em toda a faixa litorânea em ambientes de restinga, até o estado do Rio Grande do Sul, onde está presente em matas ciliares e estepes (Rotman 1995; Rattmann et al. 2012). *Eugenia*

uniflora possui frutos comestíveis que são localmente conhecidos como pitanga ou cereja-brasileira (Figura 3).



Figura 3. Frutos maduros de *Eugenia uniflora*.

Ocorre naturalmente como arbusto ou subarbusto na vegetação da planície costeira arenosa (Restinga) no nordeste e sudeste do Brasil, ou como árvores presentes em vegetação ribeirinha na região sul do Brasil, Paraguai e Argentina. Esta espécie apresenta um grande espectro ecológico, sendo encontrada em variadas formações vegetacionais, incluindo florestas, restingas, áreas áridas e semi-áridas do nordeste brasileiro, ambientes que compõem diferentes domínios da Floresta Atlântica. Esses ambientes são fisicamente e climaticamente heterogêneos, tanto em suas formações geomorfológicas, quanto composição de solo, aporte de nutrientes e disponibilidade de água (Figura 4 A e B, Figura 5 A e B).



Figura 4. Subarbustos de *E. uniflora* em ambientes de restinga na praia de Grumari, RJ.



Figura 5. Árvores de *E. uniflora* em ambientes de mata ciliar na localidade de Frederico Westphalen, RS.

As restingas são ecossistemas costeiros, fisicamente determinados pelas condições edáficas (solo arenoso) e pela influência marinha, possuindo origem sedimentar recente (início no período Quaternário, ~2,6 milhões de anos). Os fatores físicos dominantes neste ecossistema são: uma elevada salinidade, extremos de temperatura, forte presença de ventos, escassez de água, solo instável e insolação forte e direta. Já as vegetações ribeirinhas, também chamadas rupícolas ou ripárias, são constituídas por diversas comunidades vegetais, ocupando os diferentes micro-habitats definidos por uma linha de

água (Ribeiro 2000). O ambiente ribeirinho está presente em vários ecossistemas e se apresenta de forma complexa e variada, resultado das interações locais entre vários elementos: clima, geologia, condições edáficas, topografia, hidrologia e vegetação limítrofe, mas caracteriza-se por uma maior disponibilidade hídrica e uma menor incidência luminosa. Dessa forma, as características da vegetação ciliar refletem as características do mosaico ambiental de onde *E. uniflora* pode ser encontrada.

A plasticidade adaptativa de *E. uniflora* e sua capacidade de se desenvolver em diferentes habitats, são características relevantes que conferem a esta espécie um importante papel ecológico. *Eugenia uniflora* é largamente utilizada em recuperação de áreas degradadas e também possui função essencial na alimentação de diversos animais, os quais agem como dispersores de suas sementes. Além disso, a pitangueira é explorada comercialmente, não só por conter óleos essenciais, muito utilizados pela indústria de cosméticos, mas também por possuir interesse farmacológico devido às suas propriedades antifúngicas, anti-inflamatórias e anti-hipertensivas (Schapoval et al. 1994; Consolini & Sarubbio 2002) extraídas dos extratos de suas folhas.

Alguns trabalhos sobre diversidade genética em populações dessa espécie utilizando marcadores neutros já foram realizados. Em uma escala micro-geográfica, estudos em populações próximas de *E. uniflora* utilizando marcadores AFLP (*amplified fragment length polymorphism*) mostraram que a maior percentagem de diversidade genética foi observada dentro das populações (Margis et al. 2002; Salgueiro et al. 2004; Ferreira-Ramos et al. 2008). Um estudo filogeográfico de *E. uniflora* ao longo de toda a sua distribuição (Turchetto-Zolet, dados não publicados), corroboram com estes resultados e revelaram uma alta variabilidade nas populações dessa espécie no extremo sul de sua distribuição, enquanto uma menor variabilidade foi detectada em populações nas regiões sudeste e nordeste. Além disso, esses resultados também revelaram dois padrões distintos de evolução e história demográfica, um no sul e outro no norte da distribuição, dentro dos domínios da Floresta Atlântica.

Diante da importância de compreender mecanismos evolutivos envolvidos na adaptação de espécies em resposta a variações ambientais, *E. uniflora* torna-se um excelente modelo deste tipo de estudo por possuir características importantes que podem auxiliar na elucidação destes processos, devido à ampla plasticidade fenotípica que

apresenta, potencial de distribuição em ambientes com características ambientais e climáticas distintas, além da fácil adaptabilidade a estes locais.

1.4 Genética adaptativa, seleção natural e genômica

A seleção natural pode ser definida como a contribuição diferencial de variantes genéticas para as futuras gerações, ou a força motriz da evolução Darwiniana (Darwin 1859). A variação genética adaptativa é um processo resultante da seleção natural que pode ser observado entre os genomas de indivíduos que compõem diferentes populações (Holderegger et al. 2006; Lowry 2010). Esta variabilidade evolui como resultado de fatores ambientais impondo pressão seletiva sobre os indivíduos que habitam uma paisagem heterogênea (Schoville et al. 2012). Um fator ambiental pode ser qualquer fator, biótico ou abiótico, que influencia os organismos vivos e é também conhecido como um fator ecológico ou ecofactor (Schoville et al. 2012). A compreensão de como o genoma de um organismo evolui e lida com esses desafios ambientais aos quais é exposto, é de extrema relevância em estudos de genômica ecológica e evolutiva (Feder 2003; Elmer & Meyer 2011; Rice et al. 2011).

O grau de divergência adaptativa entre populações naturais por muito tempo foi geneticamente determinado por medidas de caráter distintivo - por exemplo, a propriedade de monofilia recíproca em sequências de DNA mitocondrial ou plastidial. No entanto, esta distinção com base no DNA dessas organelas, ou alguns marcadores nucleares, pode em grande parte refletir somente a história demográfica das populações estudadas. Dessa forma, é necessário quantificar a especificidade das populações em termos da sua adaptação local, focando em genes com papéis funcionais (Black IV et al. 2001; Luikart et al. 2003). Encontrar regiões genômicas sob seleção é um dos primeiros passos necessários para preencher a lacuna entre o genótipo e fenótipo de características adaptativas e é, portanto, crucial para a compreensão do processo de adaptação.

Muitas são as estratégias que as plantas utilizam para driblar os efeitos causados pelos mais diversos tipos de estresses bióticos e abióticos, que prejudicam seu pleno desenvolvimento, reprodução e produtividade. Essas habilidades podem envolver até mesmo mecanismos evolutivos para a adaptação das plantas em diferentes tipos de ambientes. A regulação de genes responsivos a estresses exerce papel importante na

cascata de sinalização desses estresses, seja através de processos biológicos, componentes celulares ou funções moleculares. Mecanismos epigenéticos, produção de mRNAs e fatores de transcrição são exemplos de controle da expressão gênica, essencial para responder a estes estímulos (Madlung 2004; Chinnusamy & Zhu 2009; Angers et al. 2010; Gollack et al. 2011; Sahu et al. 2013).

Os avanços em genômica e tecnologia de sequenciamento de DNA estão revolucionando o entendimento da variação genética natural. Marcadores ou assinaturas associadas a processos de seleção natural podem ser detectados em padrões de polimorfismo de nucleotídeos dentro e entre as populações (Anderson et al. 2011). Mais precisamente, o desenvolvimento de tecnologias de Sequenciamento de Nova Geração (*Next Generation Sequencing - NGS*) tem possibilitado a disponibilidade de bases de dados de sequências de diversas espécies (De Wit et al. 2015), facilitando o uso de análises multilocus para identificar loci responsáveis ou associados à adaptação local e entender as estratégias de adaptação que permitem algumas plantas sobreviverem a uma determinada condição ambiental.

Uma abordagem bastante utilizada em estudos sobre variação genética adaptativa é a análise de polimorfismo de nucleotídeo único (*Single Nucleotide Polymorphisms - SNP*). SNP é uma variação nas sequências do DNA que ocorre em uma posição individual e difere no genoma entre membros da mesma espécie ou entre pares de cromossomos de um indivíduo, com frequência populacional de pelo menos 1% (Brookes 1999). Um dos meios de utilizar a abordagem de SNPs para a análise de variação adaptativa em populações naturais envolve a (i) seleção de genes candidatos para a característica de interesse, (ii) avaliação desses genes para segregação da variação na forma de polimorfismos de nucleotídeo único (SNPs) e (iii) uso de métodos estatísticos para testar a associação entre as variações genéticas (SNPs) com as variações ambientais. Outros métodos utilizam um sequenciamento de regiões randômicas do genoma na busca desta informação de polimorfismo.

Na última década, as abordagens de genes candidatos tornaram-se um método interessante para estimar o nível e a estrutura da variação genética adaptativa. Abordagens usando genes candidatos permitem a caracterização da variação genética em genes potencialmente envolvidos em características adaptativas, incluindo aqueles envolvidos em

resposta ao clima (Lalagüe et al. 2014). Projetos que envolvem genomas completos de espécies não modelo ainda são escassos, seja em função da complexidade envolvida na montagem e anotação de genomas completos ou em relação aos custos envolvidos nestas técnicas. Por isso, a análise de sequências expressas, através da utilização de dados provenientes de transcritomas tem facilitado a identificação de genes candidatos (Lalagüe et al. 2014; De Wit et al. 2015).

2. OBJETIVOS

2.1 Objetivo geral

O objetivo geral deste trabalho foi analisar o padrão de variação genômica em populações naturais de *Eugenia uniflora* L. e seu potencial envolvimento na adaptação dessa espécie aos diferentes ambientes em que ocorre ao longo de sua distribuição.

2.2 Objetivos específicos

(1) Identificar e selecionar genes candidatos para a análise da variação genética em populações de *E. uniflora*;

(2) Identificar SNPs nos genes selecionados em indivíduos de populações de *E. uniflora* através do sequenciamento dos mesmos em plataforma de Sequenciamento de Nova Geração;

(3) Realizar análises de seleção nos SNPs identificados;

(4) Realizar análises de quantificação da diversidade genética e estrutura populacional nas populações estudadas;

(5) Realizar análises de medidas de associação entre a variação genética encontrada e as diferenças ambientais.

3. ESTRUTURA E ORGANIZAÇÃO DA DISSERTAÇÃO

Esta dissertação foi elaborada contendo um capítulo, assim intitulado:

Capítulo 1:

Identifying signature of positive selection in *Eugenia uniflora* natural populations genes through SNP analysis

Nicole M. Veto, Débora Bublitz, Felipe Augusto Krause, Fabiano Salgueiro, Frank Lino Guzman Escudero, Denilson Ribeiro, Francisco Eliseu Aquino, Jefferson Cardia Simões, Giovanni G. Vendramin, Rogério Margis, Andreia Carina Turchetto Zolet*.

O Capítulo 1 está apresentado em forma de um artigo científico em preparação e que será posteriormente submetido a uma revista da área, contemplando os objetivos da dissertação. Por isso a formatação do capítulo ainda não segue o modelo de nenhuma revista, as figuras e tabelas principais estão distribuídas ao longo do texto para facilitar a leitura e interpretação. As figuras e tabelas suplementares encontram-se no final do capítulo após as referências bibliográficas.

4. CAPÍTULO 1: Identifying signature of positive selection in *Eugenia uniflora* natural populations genes through SNP analysis

Nicole M. Veto, Débora Bublitz, Felipe Augusto Krause, Fabiano Salgueiro, Frank Lino Guzman Escudero, Denílson Ribeiro, Francisco Eliseu Aquino, Jefferson Cardia Simões, Giovanni G. Vendramin, Rogério Margis, Andréia Carina Turchetto Zolet.

Artigo em preparação

Identifying signature of positive selection in *Eugenia uniflora* natural populations genes through SNP analysis

Nicole M. Veto¹, Débora Bublitz¹, Felipe Augusto Krause¹, Fabiano Salgueiro³, Frank Lino Guzman Escudero¹, Denílson Ribeiro⁴, Francisco Eliseu Aquino⁴, Jefferson Cardia Simões⁴, Giovanni G. Vendramin⁵, Rogério Margis², Andréia Carina Turchetto Zolet^{1*}

¹Laboratory of Plant Genetics , Department of Genetics, Universidade Federal do Rio Grande do Sul, P.O. Box 15053, Porto Alegre, Brazil

²Laboratory of Genomes Laboratory and Plant Populations, Biotechnology Center, Universidade Federal do Rio Grande do Sul, P.O. Box 15053, Porto Alegre, Brazil

³Biological Science Center, Department of Botany, Universidade Federal do Estado do Rio de Janeiro, Av. Pasteur, nº 458, 22290 240 Rio de Janeiro, RJ, Brazil

⁴Institute of Geosciences, Department of Geography, Universidade Federal do Rio Grande do Sul, P.O. Box 15001, Porto Alegre, Brazil

⁵Institute of Biosciences and BioResources, Division of Florence, National Research Council, Florence, Italy

*Corresponding author Andreia Carinna Turchetto Zolet, Phone 55 51 3308-9812, e-mail: aturchetto@gmail.com

ABSTRACT

The next generation sequencing (NGS) technologies has facilitated the studies of population structure and evolutionary adaptations in plants and allowed us to better understand adaptive and neutral genetic variation at the population level in a spatially explicit context. *Eugenia uniflora*, popularly known as Pitanga or Brazilian cherry, is a very versatile Myrtaceae and grows in different habitats, occurring as a shrub or small tree nearby vegetation coastal-plain sand to the sea in the southeast and northeast Brazil (adjacent Restinga), and as a tree in southern part (riparian forests) of the area of Atlantic forest domain (AFD). The main goal of this study was to identify adaptive loci through screening single nucleotide polymorphism (SNP) markers at different populations of *E. uniflora*, taking into account population structures and to detect genetic signatures of natural selection. We genotyped 84 individuals of *E. uniflora* from four populations representing the main environment of species occurrence. A total of 381 SNPs representing 16 genes related to abiotic stress were genotyped for these individuals. Bayesian clustering revealed two major clusters, corresponding to the main areas of *E. uniflora* occurrence, i.e. the restinga and riparian forest. To assess the role of selection in structuring genetic variation, we applied a Bayesian and coalescent-based F_{ST} -outlier method and identified two outlier loci.

Keywords: adaptive genetic variation, NGS, SNP, *E. uniflora*.

Introduction

Spatial environmental variation is ubiquitous, and populations of a wide range of species are adapted to local abiotic or biotic conditions (Savolainen et al. 2013). The persistence of a species in heterogenic environment depends on its vulnerability to environmental changes, which is determined by its genetic constitution and physiological tolerance (Potvin & Tousignant 1996). There are many strategies that species use to circumvent the effects caused by several types of biotic and abiotic stresses that impair its full development, reproduction and productivity. In this way, identifying genes and allelic variations involved in local adaptation can help us understand how species have adapted to their environment and characterize the underlying genetic and evolutionary basis of the adaptation. As a result of natural selection, adaptive genetic variation can be observed between the genomes of individuals that comprise populations from distinct landscapes (Holderegger et al. 2006; Lowry 2010).

Nowadays, the understanding of natural genetic variation has been improved because of the use of large molecular data that are identified and genotyped based on Next Generation Sequencing (NGS) technologies (Wit et al. 2015). This approach has enabled the use of multilocus analysis to identify responsible or associated loci with local adaptation in natural population. To this end, a good option, especially in studies of native and non-model species is to give special attention to expressed sequences, which in addition to reducing the complexity of assembly of these sequences, ensure greater accuracy in functional annotation (Wang et al. 2009; Gayral et al. 2013; Wit et al. 2015). Therefore, analysis of expressed sequences using transcriptome data has facilitated the identification of candidate genes (Lalagüe et al. 2014; Wit et al. 2015).

This is an excellent approach for dealing with species occurring in different environments also adopting traits that remind their local occurrence. The knowledge of genetic and evolutionary mechanisms involved in adaptation can be of major relevance at the present time, as it may highlight how species will respond to future climate changes.

In regions such as the Atlantic Forest Domain (AFD) studying the adaptation of species is of particular relevance, since complex factors, such as strong seasonality, sharp environmental gradients and orographic driven rainfall as a result of the easterly winds from the tropical Atlantic, make this ecoregion a diverse landscape (Joly et al. 1999; De Mello Martins, 2011). Also, AFD is considered one of the most diverse and threatened biomes on the planet with only approximately 7.3% of its primary forest currently intact. The ADF consists of a set of different vegetation formations, where species can occupy one or more of this phytogeographic environment.

This is the case of *Eugenia uniflora*, popularly known as Pitanga or Brazilian-cherry. This Myrtaceae species is quite versatile and grows in many different habitats, occurring as a shrub or small tree in the vegetation nearby coastal-plain sand to the ocean in the southeast and northeast of Brazil (adjacent Restinga), and like a tree in the southern part AFD (Atlantic rain forest, steppe grasslands, training Pioneer and riparian forests). In the south, it extends from the coast to as much as 400 km inland (in riparian forests). *Eugenia uniflora* is an important ecological study model because it grows in areas of medium and high levels of precipitation and can be found in different types of vegetation and ecosystems (Almeida-Faria & Silva, 2012). The Restinga are coastal ecosystems, physically determined by soil conditions (sandy soil) and by the marine influence, possessing recent sedimentary origin (beginning in the

Quaternary period). The dominant physical factors in this ecosystem are: high salinity, extremes temperature, strong presence of winds, water shortages, unstable soil, strong, direct sunlight, among others. The riparian vegetation, also called rupicolous or riparian forests, consists of several plant communities, occupying different microhabitats defined by a water line (Ribeiro 2000). The coastal environment is present in several ecosystems and presents a complex and varied way as a result of local interactions among distinct elements such as climate, geology, soil conditions, topography, hydrology and bordering vegetation. This environment is characterized by greater water availability and less light incidence. Thus, the riparian vegetation characteristics reflect the environmental characteristics of mosaic where it can be found.

In previous study of *Eugenia uniflora* using plastidial markers (Turchetto et al, submitted manuscript), we identified two well-structured genetic units in each: northern and southern regions of the species distribution. Despite these results, nothing is known about *E. uniflora* adaptive genetic variation and there are no reports on genes underlying local adaptation mediated by abiotic or biotic stress responses. In this way, the main question is: is there a genomic signature responsible for the adaptation of this species in such distinct environments or is this strong geographical structure previously reported merely a result of demographic history? To answer this question, the main goal of this study was to identify adaptive loci through screening SNP markers at different populations of *E. uniflora*, taking into account population structures and to detect genetic signatures of natural selection. SNP markers, representing 16 genes, were used to examine the role of genetic structure and

environmental variation in shaping the distribution of this species genetic variation and its adaptation.

Methods

Plant material and DNA extraction

A total of 96 adult individuals of *Eugenia uniflora* were sampled from four populations along distinct environments of natural distribution of this species (Fig. 1) (see description of species distribution on the introduction section). Leaf samples were collected from adult individuals of each population, which were selected considering both spatial distribution and environmental heterogeneity along of species distribution. The populations RSRF (riparian forest near the rivers), RSPA (gallery forest in the PAMPA savanna) and RSCO (Coastal plain near the ocean) are located in the southern border of Brazilian Atlantic Forest, representing distinct environment; while the RJRE (restinga) is located in the southeast of Atlantic forest, represent the restinga environment (Table 1 and Fig. 1). Our sampling covers also different soils, temperature range and precipitation patterns (see description in Climatic data). The geographic coordinates are reported in Table 1. The total genomic DNA was extracted from leaves dried in silica gel according to Doyle and Doyle (1990).

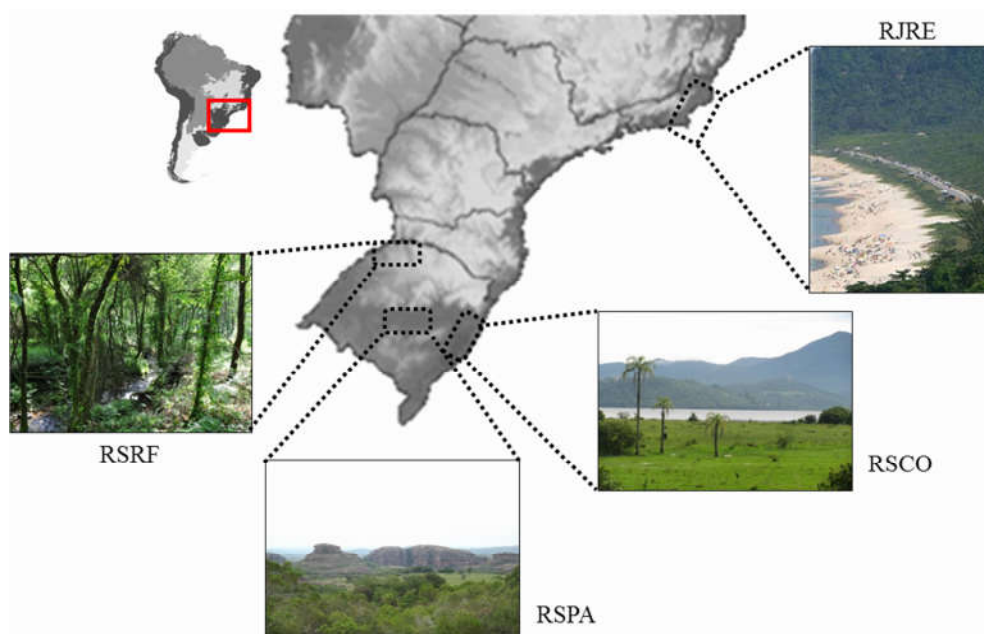


Figure 1. Populations of *Eugenia uniflora* showing the distinct environment where the species occurs. Dotted boxes represent the locality of each population.

Table 1: Information about population used in this study: number of collected individuals, geographic coordinates and soil type description.

Population	N	Lat	Lon	Elevation (m)	Soil
RJRE	32	-23.06	-43.53	11	Nutrient poor, sandy, salt and acid
RSRF	32	-27.32	-53.33	536	Nutrient rich, organic material and water accumulation
RSPA	14	-30.52	-53.49	378	Neosoil, shallow, very low permeability
RSCO	18	-30.19	-50.49	7	Sandy, water stress, low fertility, lagoon and sea

Environmental and climatic data

Four climatic variables were considered: 30 years (1979 – 2009) of average monthly mean temperature (MMT), monthly mean of maximum temperature (MMATM), monthly mean of minimum temperature (MMITM) and monthly mean precipitation (MP) (Table 2). These variables were obtained from INMET (Instituto Nacional de Meteorologia) for two stations referring to two major populations: Iraí

and Rio de Janeiro.

Table 2: Environment variables for two principal populations

Population	MMT (°C)	MMATM (°C)	MMITM (°C)	MP (mm)
RJRE	22,7	29,0	21,6	117,08
RSRF	21,0	27,5	15,2	161,46

MMT - Average mean temperature, MMATM - Average maximum temperature, MMITM - Average minimum temperature, MP - Average precipitation.

Selection of candidate genes, primers design and PCR amplification

The transcriptome of *Eugenia uniflora* (Guzman et al. 2014) was used to identify candidate genes involved in abiotic stress responses (hereafter, stress genes response to drought, salinity, temperature and light). We firstly compiled a nonexhaustive list of proteins/genes that are involved in plant response to abiotic stress from the literature (2014-2015 ISI Web of Knowledge using the following keywords: “candidate gene”, “drought”, “salinity”, “cold”, “temperature” “protein” and “gene expression”). Secondly, we used the unigenes of *E. uniflora* transcriptome to perform the identification of Gene Ontology and annotation of candidate genes with BLAST2go. Finally, based on literature and in the annotation we selected 45 candidate genes for primer designing. The primers were designed in the 5' UTR and 3' UTR region, using Primer3.

The regions of 45 candidate genes were amplified using Real-time polymerase chain reaction (PCR) in the 96 DNA samples. The PCR reaction were conducted using 150 ng of genomic DNA, 50 mM magnesium chloride (MgCl₂), 5 mM deoxyribonucleotide triphosphates (dNTP) mix, 1 x PCR buffer, 0.3 U of Platinum Taq DNA polymerase (Invitrogen) and 5 μM of each primer, in a total of 20 μL. After initial tests the following PCR conditions were able to amplify all fragments: an initial hot-start step at 95 °C for 5 min, followed by 35 cycles with

denaturation at 94 °C for 30 s, an annealing temperature of 60 °C for 30 s, and 3 m of elongation at 72 °C. All primer sequences used and conditions are described in Table 3. To check the amplification of the products we used the 7500 fast v2.0.3 software.

Sequencing, SNP discovery and genotyping

SNPs discover was based on Next Generation sequence of candidate genes previously amplified by PCR using Illumina Miseq 2000 platform. The Nextera XT DNA Sample Preparation Kit was selected for library preparation, it requires only 1 ng template DNA, can be performed in a relatively-short time frame, and multiple samples can be prepared simultaneously. After the verification of amplification, quality, amplified sequences of 30 candidate gene were pooled for each of the 96 individuals of *E. uniflora*. Sample libraries of 96 individuals were prepared according to manufacturer's instructions and sequenced using MiSeq v3 chemistry.

For the treatment of sequence data we used a pipeline as described in Fig. 2.

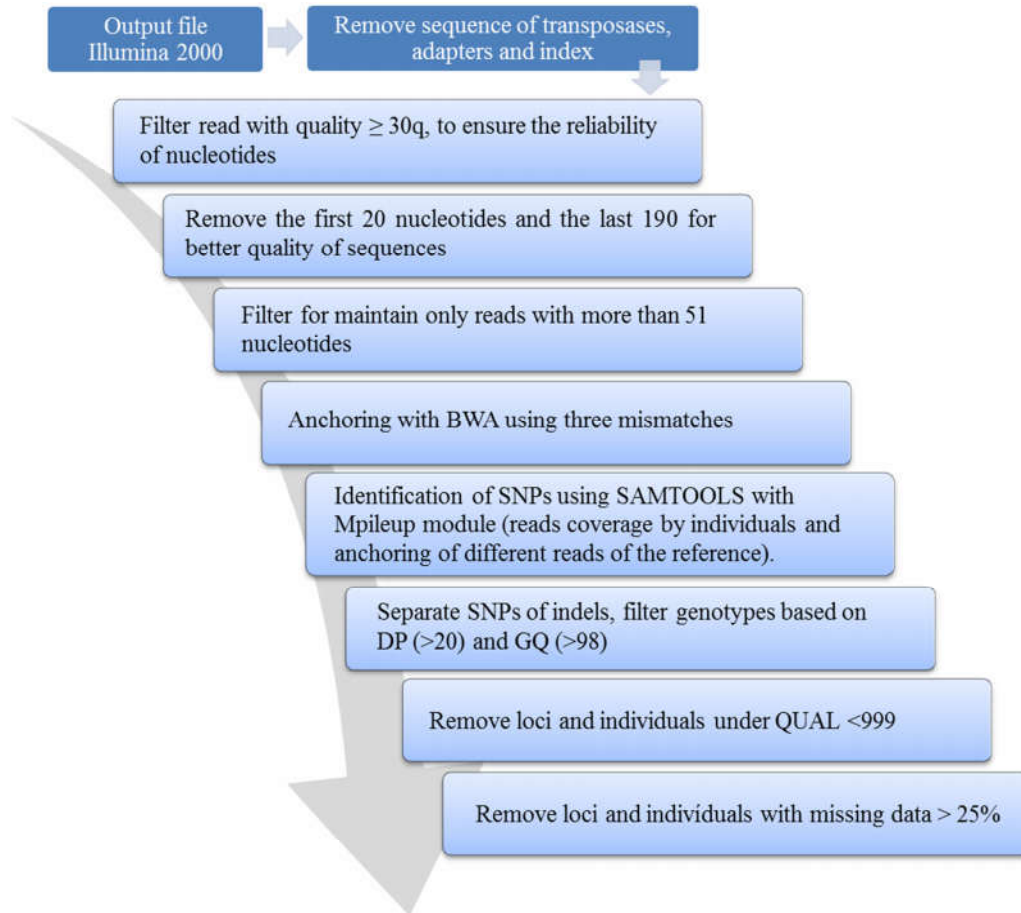


Figure 2. Pipeline used for processing output data of NGS in Illumina platform.

F_{ST} -outlier detection

To identify F_{ST} outliers, Bayesian approach was applied considering individual locus effect and specific population, focusing on a genome scan for positive and balancing selection, as implemented in BayeScan 2.1. The method tests two alternative models and assigns a Bayes factor each locus. We used a prior odd equal 10 and a false discovery rate FDR 50.001.

Population structure

Patterns of population structure were analyzed by Bayesian cluster analysis that was performed on the SNP data matrix using the program STRUCTURE ver.2.2.

STRUCTURE runs were performed with a Markov Chain Monte Carlo (MCMC) burn in of 2500,000 steps, followed by an MCMC of 1,000,000 steps. An admixture model was used in the simulations. Each analysis was replicated 10 times for each K, with K ranging from 1 to 10. The best K was assigned using the log likelihood value, and populations were assigned to each genetic cluster considering the assignment of the majority of individuals within each population.

Results and Discussion

Gene selection, sequencing and SNP genotyping

The BLAST2go analysis allowed us to identify the gene ontology annotation of 64 candidate genes related with abiotic stress (Table S1). Among these, 45 were selected for amplification and sequencing of 96 individuals of *E. uniflora*. Thirty of the 45 candidate genes were successfully amplified in total of 96 individuals and then were submitted to sequencing. Out of the 30 amplified genes, 16 were successfully sequenced from a set of 84 individuals. The BLAST2go showed that these genes are involved in different processes, such as molecular, biological or cellular processes (Fig. 3). In particular, these genes are involved in different signaling cascades that interact at the level of cellular components, mechanisms that respond to various abiotic stresses such as drought, oxidative stress, osmotic stress, senescence and seed development. After amplification and sequencing, a total of 381 SNPs across 16 genes were successfully genotyped. The Table 3 and Figure 4 represent the characteristics of sequencing and SNPs analysis. These genes have already been reported in studies of abiotic stress response (Amara et al. 2012; Giarola et al. 2015; Y. Ke et al. 2009; Melmaiee et al. 2015; Perdiguero et al. 2015; Tao et al. 2006). The use of candidate

genes has been successfully used in studies of selection and adaptation (Csilléry et al. 2014; Scalfi et al.2014; Lagüe et al. 2014; Grivet et al. 2011; Jaramillo-Correa et al. 2015), since candidate gene approaches allow the characterization of the genetic variation directly at genes that potentially code for adaptive traits, including those involved in response to climate.

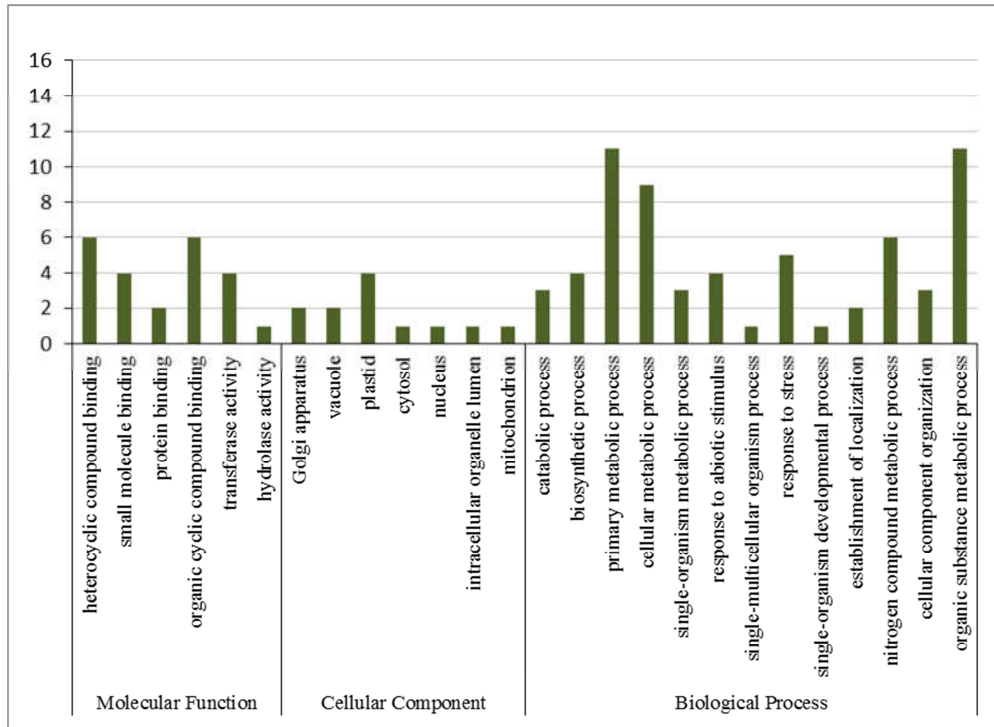


Figure 3. Blast2Go annotation of the 16 genes sequencing.

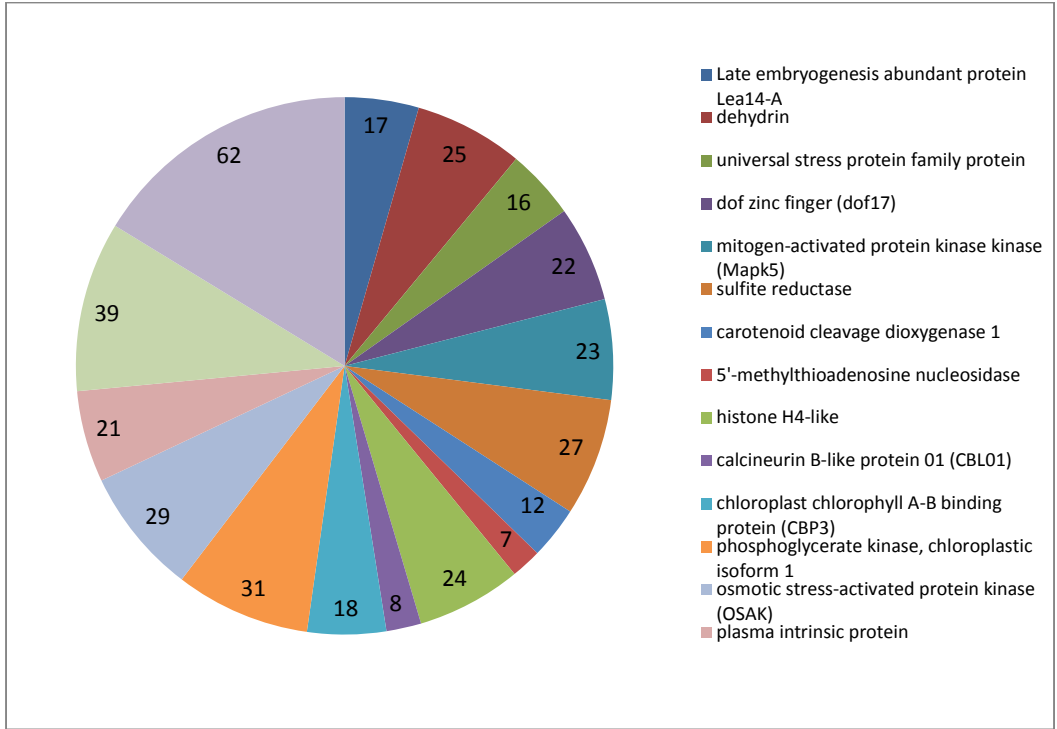


Figure 4. SNPs distribution in the 16 analyzed genes.

Table 3. Information about the 16 sequenced genes.

Gene	Locus identification	Predicted size	Sequenced coverage	Reads
dehydrin	<i>E. Uni 63</i>	744	1188	147.480
universal stress protein family protein	<i>E. Uni 7951</i>	480	1984	105.037
Late embryogenesis abundant protein Lea14-A	<i>E. Uni 15860</i>	453	701	253.077
dof zinc finger (dof17)	<i>Euni16779_1574</i>	828	1574	315.877
sulfite reductase	<i>Euni481_2782</i>	1098	2782	131.078
mitogen-activated protein kinase (Mapk5)	<i>Euni1727_1212</i>	585	1212	114.325
carotenoid cleavage dioxygenase 1	<i>Euni5318_779</i>	633	779	79.303
5'-methylthioadenosine nucleosidase	<i>Euni8928_1564</i>	642	1564	38.630
calcineurin B-like protein 01 (CBL01)	<i>Preuni17960_663</i>	477	663	28.981
chloroplast chlorophyll A-B binding protein (CBP3)	<i>Preuni31_2107</i>	426	2107	80.509
histone H4-like	<i>Preuni17184_671</i>	309	671	214.997
3-ketoacyl-CoA synthase 11	<i>Preuni962_2650</i>	1407	2650	266.509
osmotic stress-activated protein kinase (OSAK)	<i>Preuni5552_1340</i>	699	1340	153.326
phosphoglycerate kinase, chloroplastic isoform 1	<i>Preuni328_2200</i>	873	2200	155.070
plasma intrinsic protein	<i>Preuni7330_1053</i>	483	1053	106.777
pentatricopeptide repeat-containing protein At1g15510	<i>Preuni9300_1515</i>	1164	1515	431.094

Climatic and environmental characteristics

Environmental variables collected for the two main populations (RJRE and RSRF) shows the monthly average of 30 years for rainfall and temperature in two locations (Fig. 2, Table S2). The site of population RJRE show low temperature amplitude, maintaining mild throughout the year, a great difference in rainfall, volume extent 227mm average maximum and average minimum January-July 34mm. Otherwise occurs with the average for RSRF population. There are approximately 10°C temperature amplitude between the months of January and July, while the precipitation volumes appear more uniform. These data help us to understand how and what mechanisms are used by this species to be able to adapt in environments with such distinct characteristics, for both dry periods for RJRE as resistance to low temperatures as in RSRF.

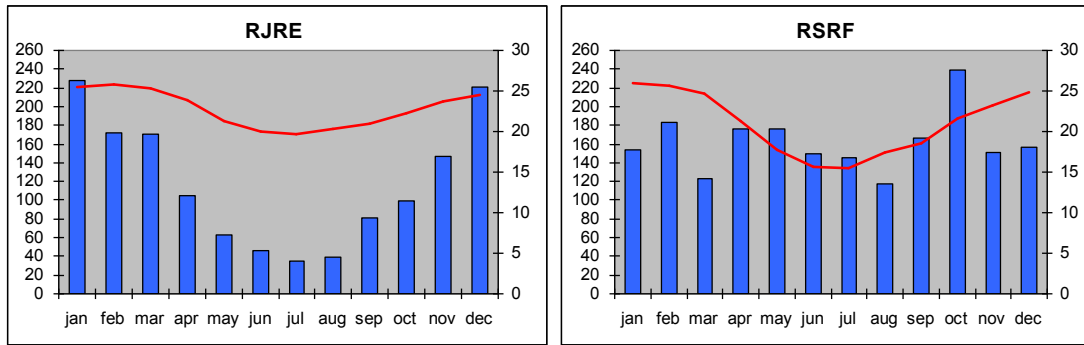


Figure 5. Graphic of the monthly average rainfall and temperature for 30 years (1979-2009) for the two main populations RJRE and RSRF. The blue bars indicate the averages of rainfall and the red lines average temperature.

F_{ST} -outliers detection

F_{ST} values were calculated for each locus among population and varied from 0.027 to 0.173. The BayeScan simulation was run using populations assigned according to their geographic position and two outlier loci were detected (Table 4). The outliers with the highest F_{ST} -value, 0.172, SNP locus E.Uni15860_635, that encodes a Late embryogenesis abundant

protein Lea14-A and 0.149 for SNP locus E.Uni63_556, that encodes a dehydrin. Interestingly, the population from restinga (RJRE) has a different allelic combination than other three populations located in the southern part of *E. uniflora* distribution (Fig. S1).

Table 4. Outlier detection using BayeScan results.

SNP	Putative function	Fst	All Populations			
			prob	log10(PO)	qval	alpha
E.Uni15860_635	Late embryogenesis abundant protein Lea14-A	0.17301	0.99500	22.988	0.0029006	19.389
E.Uni63_556	dehydrin	0.14978	0.99920	30.965	0.00080016	17.610

The SNP in this position represent a substitution of a Thymine (T) to a Cytosine (C), which causes aminoacid change in the protein sequence (Fig. 6). Amongst the numerous gene and transcription factor families involved in the molecular response to water stress in plants, late embryogenesis abundant (LEA) proteins, and more specifically dehydrins, the group II of LEA proteins, have been related with drought tolerance in numerous species (Hu et al. 2010; Lopez et al. 2003; Suprunova et al. 2004). A recent study that investigated the effects of adaptation to environment on population level genetic diversity of *Eucalyptus camaldulensis* using SNPs found outlier loci under diversifying selection in genes to response to abiotic stress, including dehydrin (Dillon et al. 2014). These results suggest that genes directly impacting water movement, dehydration or high salinity response may be involved in adaptation mechanisms. It corroborate with studies identifying dehydrins as candidates for adaptive evolution in response to water availability (Xia et al, 2010; Wachowiak et al. 2009).

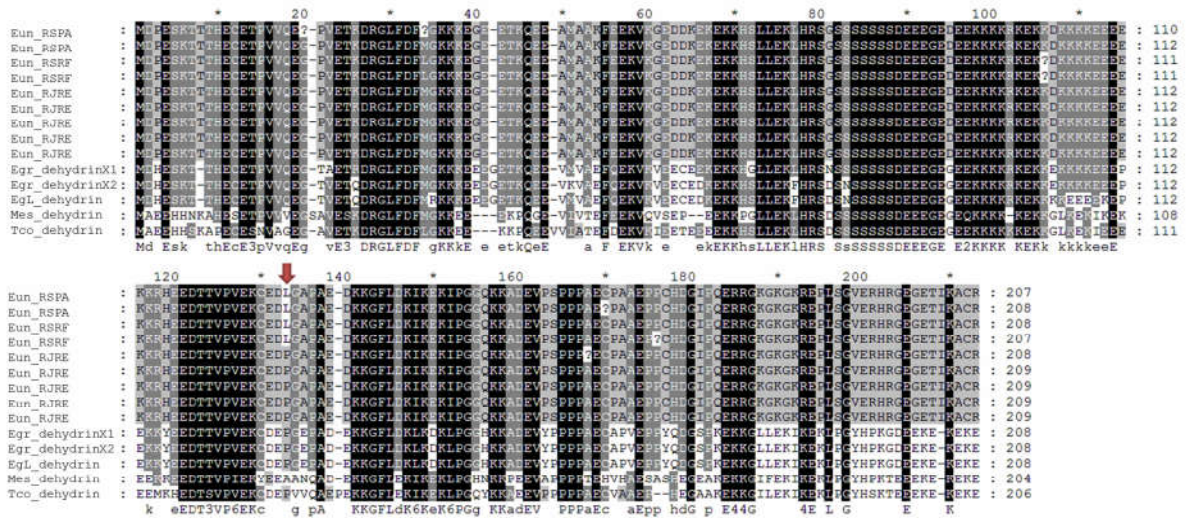


Figure 6. Dehydrin sequence alignment of *Eugenia uniflora* (Eun), *Eucalyptus grandis* (Egr), *Eucalyptus globulus* (EgL), *Manihot esculenta* (Mes) and *Theobroma cacao* (Tco). The red arrow indicate the outlier SNP position.

Population structure

To examine patterns of population structure, Bayesian clustering was applied using STRUCTURE. These analysis detected two main clusters (Fig. 7) associated with the two distinct environments where *E. uniflora* occurs (restinga and riparian forest). The first cluster included individuals from RJRE population from restinga, while the second cluster included the other three populations (RSRF, RSRF and RSCO). A previous study (Turchetto-Zolet et al. unpublished) using chloroplast markers showed two divergent lineages with an estimated divergence time at approximately 5 Mya with distinct evolutionary histories from southern to northern populations of *E. uniflora* in the AFD.

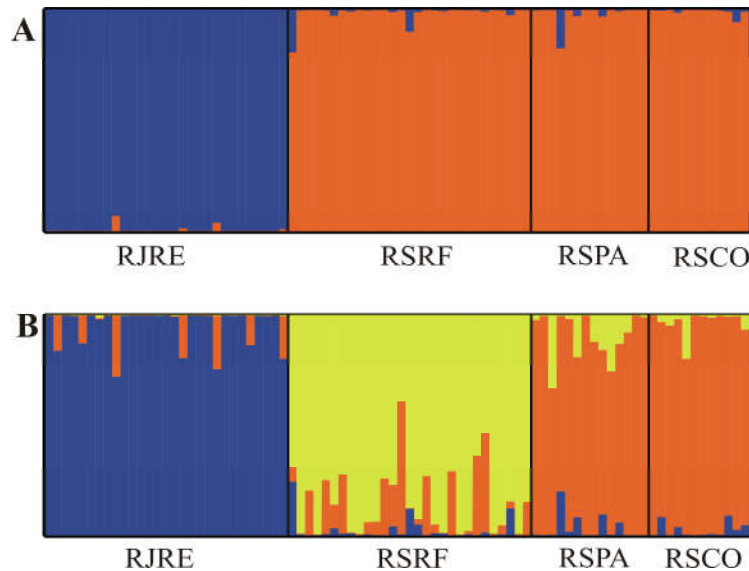


Figure 7. The Bayesian cluster analysis with STRUCTURE. A) Shows two clusters supported with the best $K=2$; B) $K=3$.

Conclusion

This research confirms the findings from previous study describing the genetic structure and phylogeography of *E. uniflora* at the Atlantic Forest Domain and highlights the importance of use the outlier detection studies to better understand the evolutionary history of this species. The results of this study contribute for the understanding of the genetic structure of populations in different environments along the AFD. The experimental design we used and the employment of different approaches in data analysis provided new insights into the underlying genes that may be responsible for local adaptation.

References

- Almeida, D. J., Faria, M. V., & Silva, P. R. (2012). Biologia experimental em Pitangueira: uma revisão de cinco décadas de publicações científicas. *Ambiência Guarapuava (PR)*, 8, 159-175.
- Amara, I., Odena, A., Oliveira, E., Moreno, A., Masmoudi, K., Pagés, M., & Goday, A. (2012). Insights into maize LEA proteins: from proteomics to functional approaches. *Plant and Cell Physiology*, 53(2), 312-329.
- Csilléry, K., Lalagüe, H., Vendramin, G. G., González-Martínez, S. C., Fady, B., & Oddou-Muratorio, S. (2014). Detecting short spatial scale local adaptation and epistatic selection

- in climate-related candidate genes in European beech (*Fagus sylvatica*) populations. *Molecular ecology*, 23(19), 4696-4708.
- De Mello Martins, Felipe (2011). Historical biogeography of the Brazilian Atlantic forest and the Carnaval–Moritz model of Pleistocene refugia: what do phylogeographical studies tell us?. *Biological Journal of the Linnean Society*, 104(3), 499-509.
- De Wit P, Pespeni MH & Palumbi SR (2015) SNP genotyping and population genomics from expressed sequences -current advances and future possibilities. *Mol Ecol* 24:n/a–n/a. doi: 10.1111/mec.13165
- Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. *Focus* 12:3
- Gayral P, Melo-Ferreira J, Glémin S et al. (2013) Reference-free population genomics from next-generation transcriptome data and the vertebrate-invertebrate gap. *PLoS Genetics*, 9, e1003457.
- Giarola, V., Challabathula, D., & Bartels, D. (2015). Quantification of expression of dehydrin isoforms in the desiccation tolerant plant *Craterostigma plantagineum* using specifically designed reference genes. *Plant Science*, 236, 103-115.
- Grivet, D., Sebastiani, F., Alía, R., Bataillon, T., Torre, S., Zabal-Aguirre, M., ... & González-Martínez, S. C. (2011). Molecular footprints of local adaptation in two Mediterranean conifers. *Molecular biology and evolution*, 28(1), 101-116.
- Guzman, F., Kulcheski, F. R., Turchetto-Zolet, A. C., & Margis, R. (2014). De novo assembly of *Eugenia uniflora* L. transcriptome and identification of genes from the terpenoid biosynthesis pathway. *Plant Science*, 229, 238-246.
- Holderegger R, Kamm U & Gugerli F (2006) Adaptive vs. neutral genetic diversity: implications for landscape genetics. *Landsc Ecol* 21:797–807. doi: 10.1007/s10980-005-5245-9.
- Hu L, Wang Z, Du H, Huang B (2010) Differential accumulation of dehydrins in response to water stress for hybrid and common bermudagrass genotypes differing in drought tolerance. *J Plant Physiol* 167:103–109. doi:10.1016/j.jplph.2009.07.008
- Jaramillo-Correa, J. P., Rodríguez-Quilón, I., Grivet, D., Lepoittevin, C., Sebastiani, F., Heuertz, M., ... & González-Martínez, S. C. (2015). Molecular Proxies for Climate Maladaptation in a Long-Lived Tree (*Pinus pinaster* Aiton, Pinaceae). *Genetics*, 199(3), 793-807.
- Joly, C. A., Aidar, M. P., Klink, C. A., McGrath, D. G., Moreira, A. G., Moutinho, P., ... & Sampaio, E. V. S. B. (1999). Evolution of the Brazilian phytogeography classification systems: implications for biodiversity conservation. *Ciencia e Cultura*(Sao Paulo), 51(5), 331-348.
- Ke, Y., Han, G., He, H., & Li, J. (2009). Differential regulation of proteins and phosphoproteins in rice under drought stress. *Biochemical and biophysical research communications*, 379(1), 133-138.
- Lalagüe H, Csilléry K, Oddou-Muratorio S, Safrana J, de Quattro C, Fady B, González-Martínez SC, Vendramin GG (2014) Nucleotide diversity and linkage disequilibrium at 58 stress response and phenology candidate genes in a European beech (*Fagus sylvatica* L.) population from southeastern France. *Tree Genet Genomes* 10:15–26. doi: 10.1007/s11295-013-0658-0.
- Lalagüe, H., Csilléry, K., Oddou-Muratorio, S., Safrana, J., de Quattro, C., Fady, B., ... & Vendramin, G. G. (2014). Nucleotide diversity and linkage disequilibrium at 58 stress

- response and phenology candidate genes in a European beech (*Fagus sylvatica* L.) population from southeastern France. *Tree genetics & genomes*, 10(1), 15-26.
- Lopez CG, Banowetz GM, Peterson CJ, Kronstad WE (2003) Dehydrin expression and drought tolerance in seven wheat cultivars. *Crop Sci* 43:577–582. doi:10.2135/cropsci2003.5770
- Lowry D (2010) Landscape evolutionary genomics. *Biol. Lett.* 6:502–4.
- Melmaiee, K., Anderson, M., Elavarthi, S., Guenzi, A., & Canaan, P. (2015). Transcriptional Analysis of Resistance to Low Temperatures in Bermudagrass Crown Tissues. *PLoS one*, 10(9), e0136433.
- Perdiguero, P., Soto, Á., & Collada, C. (2015). Comparative analysis of *Pinus pinea* and *Pinus pinaster* dehydrins under drought stress. *Tree Genetics & Genomes*, 11(4), 1-11.
- Potvin, C., & Tousignant, D. (1996). Evolutionary consequences of simulated global change: genetic adaptation or adaptive phenotypic plasticity. *Oecologia*, 108(4), 683-693
- Scalfi, M., Mosca, E., Di Pierro, E. A., Troglio, M., Vendramin, G. G., Sperisen, C., ... & Neale, D. B. (2014). Micro-and macro-geographic scale effect on the molecular imprint of selection and adaptation in Norway spruce. *PLoS one*, 9(12), e115499.
- Suprunova T, Krugman T, Fahima T, Chen G, Shams I, Korol A, Nevo E (2004) Differential expression of dehydrin genes in wild barley, *Hordeum spontaneum*, associated with resistance to water deficit. *Plant Cell Environ* 27:1297–1308. doi:10.1111/j.1365-3040.2004.01237.
- Tao, N. G., Ye, J. L., Juan, X. U., & Deng, X. X. (2006). Cloning and Characterization of a Novel cDNA Encoding Late Embryogenesis-Abundant Protein 5 Like (LEA-5) Gene from Cara Cara Navel Orange Fruit (*Citrus sinensis* Osbeck). *Agricultural Sciences in China*, 5(4), 272-276.
- Wachowiak, W., Balk, P. A., & Savolainen, O. (2009). Search for nucleotide diversity patterns of local adaptation in dehydrins and other cold-related candidate genes in Scots pine (*Pinus sylvestris* L.). *Tree Genetics & Genomes*, 5(1), 117-132.
- Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews. Genetics*, 10, 57–63.
- Xia H, Camus-Kulandaivelu L, Stephan W, Tellier A, Zhang ZW (2010) Nucleotide diversity patterns of local adaptation at drought-related candidate genes in wild tomatoes. *Mol Ecol* 19: 4144–4154.



Figure S1.: Snp in position 556 in dehydrin loci (*E. Uni 63*).

Table S1. Annotation BLAST2go of total of the candidate genes selected for test, related with abiotic stress.

Contig	Annotation blastx blast2go	Abundance
31_Locus_4321_Transcript_3_4_Confidence_0.643_Length_1992	ap2 erf domain-containing transcription factor	729
31_Locus_63_Transcript_66_66_Confidence_0.082_Length_1188	dehydrin	5999
51_Locus_668_Transcript_4_4_Confidence_0.571_Length_2716	ma3 domain-containing protein	1583
21_Locus_8101_Transcript_1_2_Confidence_0.667_Length_3385	protein far-red impaired response 1-like	497
51_Locus_7951_Transcript_2_4_Confidence_0.818_Length_1984	universal stress protein family protein	1608
61_Locus_9472_Transcript_1_3_Confidence_0.714_Length_3881	protein argonaute	1480
31_Locus_15860_Transcript_1_2_Confidence_0.667_Length_701	desiccation protectant protein lea14-like protein	98
21_Locus_1156_Transcript_4_6_Confidence_0.562_Length_2276	peptidylprolyl isomerase	2731
21_Locus_16779_Transcript_1_2_Confidence_0.833_Length_1574	dof zinc finger	89
51_Locus_845_Transcript_2_2_Confidence_0.667_Length_1105	h+-transporting atp synthase-like protein	2223
31_Locus_5330_Transcript_1_2_Confidence_0.800_Length_4243	phytochrome b	1214
41_Locus_481_Transcript_1_2_Confidence_0.833_Length_2782	sulfite reductase	1761
21_Locus_1727_Transcript_1_2_Confidence_0.500_Length_1212	mitogen-activated protein kinase kinase	192
31_Locus_5308_Transcript_1_1_Confidence_1.000_Length_1896	3-isopropylmalate dehydrogenase	284
51_Locus_5318_Transcript_2_9_Confidence_0.423_Length_779	carotenoid cleavage dioxygenase 1	1029
41_Locus_7324_Transcript_1_1_Confidence_1.000_Length_2739	potassium channel skor	1013
21_Locus_4637_Transcript_8_11_Confidence_0.345_Length_3702	cbl-interacting protein kinase 23	2095
31_Locus_8928_Transcript_1_9_Confidence_0.600_Length_1564	protein	354
31_Locus_17960_Transcript_1_2_Confidence_0.667_Length_663	calcineurin b-like protein 1	50
31_Locus_3_Transcript_1448_1450_Confidence_0.004_Length_2107	chlorophyll a b-binding protein cp29	4340
31_Locus_156_Transcript_147_150_Confidence_0.069_Length_4226	glyceraldehyde-3-phosphate dehydrogenase	14201
31_Locus_1117_Transcript_2_6_Confidence_0.615_Length_1118	chlorophyll a b-binding protein	1776

21_Locus_5928_Transcript_1_3_Confidence_0.130_Length_2404	c2-h2 zinc finger protein	533
31_Locus_1020_Transcript_3_3_Confidence_0.667_Length_1307	proteasome subunit alpha type-7	1330
41_Locus_17184_Transcript_1_1_Confidence_1.000_Length_671	histone h4	45
21_Locus_962_Transcript_24_27_Confidence_0.089_Length_2650	beta-ketoacyl-coa synthase family protein	482
21_Locus_6258_Transcript_2_2_Confidence_0.667_Length_5648	mediator of rna polymerase ii transcription subunit	794
41_Locus_5552_Transcript_5_7_Confidence_0.333_Length_1340	serine threonine protein kinase	631
21_Locus_328_Transcript_5_9_Confidence_0.581_Length_2200	phosphoglycerate kinase	4387
41_Locus_7330_Transcript_1_3_Confidence_0.400_Length_1053	plasma intrinsic protein	477
21_Locus_1012_Transcript_27_36_Confidence_0.081_Length_2701	probable lactoylglutathione chloroplast-like	2450
21_Locus_701_Transcript_3_10_Confidence_0.464_Length_2337	cbs domain protein	1982
51_Locus_3469_Transcript_5_6_Confidence_0.722_Length_3547	beta-glucosidase-like chloroplastic-like	3653
31_Locus_1536_Transcript_3_12_Confidence_0.281_Length_1455	endo- -beta-d-	4342
21_Locus_8597_Transcript_6_6_Confidence_0.815_Length_2115	serine threonine-protein kinase	668
21_Locus_4950_Transcript_1_3_Confidence_0.667_Length_1349	cold regulated 314 thylakoid membrane 2	521
21_Locus_4342_Transcript_1_1_Confidence_1.000_Length_2358	delta-1-pyrroline-5-carboxylate dehydrogenase 1 protein	540
21_Locus_7011_Transcript_1_1_Confidence_1.000_Length_3232	probable -trehalose-phosphate synthase	546
21_Locus_3427_Transcript_6_7_Confidence_0.500_Length_2484	sodium-dependent phosphate transport protein 1	2386
41_Locus_9300_Transcript_1_1_Confidence_1.000_Length_1515	pentatricopeptide repeat-containing protein chloroplastic-like	131
21_Locus_980_Transcript_4_4_Confidence_0.733_Length_1209	glutathione peroxidase	1402
31_Locus_1915_Transcript_2_4_Confidence_0.727_Length_2668	udp-glucose pyrophosphorylase	706
31_Locus_2262_Transcript_8_10_Confidence_0.500_Length_3123	bel1-like homeodomain 1	5474
41_Locus_8760_Transcript_1_1_Confidence_1.000_Length_4160	auxin response factor 7	566
41_Locus_690_Transcript_2_3_Confidence_0.625_Length_1236	pyrroline-5-carboxylate synthetase	323
31_Locus_4321_Transcript_4_4_Confidence_0.500_Length_1823	ap2 erf domain-containing transcription factor	1232
41_Locus_259_Transcript_8_11_Confidence_0.595_Length_4069	phototropin 2	5577
31_Locus_75_Transcript_18_18_Confidence_0.269_Length_3381	tcp-1 cpn60 chaperonin family protein	19887
21_Locus_440_Transcript_44_45_Confidence_0.148_Length_3260	aspartic proteinase	9240

51_Locus_10393_Transcript_1_2_Confidence_0.800_Length_1373	glutamate dehydrogenase 1	557
21_Locus_11883_Transcript_1_1_Confidence_1.000_Length_2659	pyrophosphate-energized vacuolar membrane proton pump-like	372
51_Locus_4632_Transcript_1_1_Confidence_1.000_Length_3417	cytochrome p450	1200
31_Locus_15558_Transcript_1_1_Confidence_1.000_Length_1644	calcineurin b-like protein 10	163
21_Locus_785_Transcript_6_6_Confidence_0.781_Length_1733	cbl-interacting serine threonine-protein	503
31_Locus_3642_Transcript_4_4_Confidence_0.700_Length_4703	alpha-glucan water chloroplastic-like	7285
51_Locus_3_Transcript_35_39_Confidence_0.049_Length_2144	monodehydroascorbate reductase	1180
21_Locus_1118_Transcript_3_6_Confidence_0.240_Length_3614	26s proteasome non-atpase regulatory subunit 8	1812
21_Locus_1395_Transcript_4_4_Confidence_0.500_Length_2576	26s proteasome subunit 4	5838
31_Locus_5919_Transcript_3_5_Confidence_0.688_Length_1540	cold regulated protein 27	232
41_Locus_220_Transcript_25_26_Confidence_0.365_Length_6774	heat shock protein 70	19917
61_Locus_1326_Transcript_3_4_Confidence_0.727_Length_2116	glycine hydroxymethyltransferase	13391
21_Locus_532_Transcript_3_3_Confidence_0.600_Length_1559	embryonic flower 2	508
21_Locus_246_Transcript_20_25_Confidence_0.260_Length_4301	cysteine proteinase	6270
21_Locus_1608_Transcript_3_4_Confidence_0.400_Length_2628	vernalization5 vin3-like protein	550

Table S2: Features and primers referring to 30 sequenced genes.

Gene	Euni_Contig Locus	Predicted size	Annealing temperature	Primers
dehydrin	<i>Preuni63_1188</i>	744	60°	<i>Forwad</i> 5'- AGC ACG TCA GTC TCT CTT TAT TCC CGT TCG -3' <i>Reverse</i> 5'- CAT AGA AGC ACC TTA TAA ACA ATA CTC GCG CAT G -3'
programmed cell death protein 4	<i>Preuni668_2716</i>	2118	60°	<i>Forwad</i> 5'- CAG AAG AAT CGG ATA CGC TTC GGT CTC -3' <i>Reverse</i> 5'- GAT ACA TAA CGT GAC AGG CTA CAA CAT AAG ACA G -3'
protein far-red impaired response	<i>Preuni8101_3385</i>	2472	60°	<i>Forwad</i> 5'- CTG TCG GAA ATC TTC TTT TGA TTG GGA TGA ATC -3' <i>Reverse</i> 5'- AAT GCC GGA AAT TGC TGG ACC AC -3'
universal stress protein family protein	<i>Preuni7951_1984</i>	480	60°	<i>Forwad</i> 5'- CTC GGA CCA AGA TCA ATC GGC G -3' <i>Reverse</i> 5'- AAC GTC ATT CTT TAT CCA AAA ACT TCA TTC TAG CAA -3'
Late embryogenesis abundant protein Lea14-A	<i>Preuni5860_701</i>	453	60°	<i>Forwad</i> 5'- GAA CGG GAA GGT CCG ACT TTT ATT CTT CTC -3' <i>Reverse</i> 5'- CGA GAC ATG TCG GAG TTG CTG GAG -3'
dof zinc finger (dof17)	<i>Preuni16779_1574</i>	828	60°	<i>Forwad</i> 5'- GTC CTT GTT GTG AAC TAG GAC ACA AGG G -3' <i>Reverse</i> 5'- GTT AGC TCT GAA CAG TCT CAG ACA TGG AGT AA -3'
ATP synthase delta chain, chloroplastic	<i>Preuni845_1105</i>	738	60°	<i>Forwad</i> 5'- CAC GAA ATC GCC TCC TCT CTC CTT -3' <i>Reverse</i> 5'- AAA AAT GAA AAA GAA AAT GCT ATC GCC GGC ATC -3'
sulfite reductase	<i>Preuni481_2782</i>	1098	60°	<i>Forwad</i> 5'- GTT TGT TGA TCG ACC ATG ACG ACG G -3' <i>Reverse</i> 5'- TAG TCT CCA ACC TAT GTG TTC TTC CCA TAC C -3'
mitogen-activated protein kinase kinase (Mapk5)	<i>Preuni1727_1212</i>	585	60°	<i>Forwad</i> 5'- GGC ATC TTC GAG AAG CCG AAC G -3' <i>Reverse</i> 5'- TCG GCT TCC GGG GTT TTC G -3'
3-isopropylmalate dehydrogenase	<i>Preuni5308_1896</i>	1218	54°	<i>Forwad</i> 5'- GAT GGC TTC CCT GCA ACT GAA TTT CAA ATC -3' <i>Reverse</i> 5'- ACC TTC GCG ATC CTT GTA ATT AGG AAA CAC -3'
carotenoid cleavage dioxygenase 1	<i>Preuni5318_779</i>	633	60°	<i>Forwad</i> 5'- TCG GTT TGT TCT ATT CAG GGA AAT GCT GTA AC -3'

					<i>Reverse</i>	5'- GGC TCA GCA TGC AGA TCA AAC TTC AC -3'
potassium channel skor	<i>Preuni7324_2739</i>	2490	54°	<i>Forwad</i>	5'- TCG GTT TTC CGG AAT CCG CG -3'	
				<i>Reverse</i>	5'- ACA ACA ACG GCT ATT GTT TCA CAA CAT AAT AGC T -3'	
5'-methylthioadenosine nucleosidase	<i>Preuni8928_1564</i>	642	60°	<i>Forwad</i>	5'- CAA CAA TAG AAT TCA GTA CCT ATC AGA TGC CCT CT -3'	
				<i>Reverse</i>	5'- TAA GAG AAG ATC AAG TGC GCA ACT AAA GCG -3'	
calcineurin B-like protein 01 (CBL01)	<i>Preuni17960_663</i>	477	60°	<i>Forwad</i>	5'- ACA CAA GGG ATT CAT CGA CTT CGG T -3'	
				<i>Reverse</i>	5'- CCA GGA ACT TAA TTG GCA TGG TTC TCA CA -3'	
chloroplast chlorophyll A-B binding protein (CBP3)	<i>Preuni31_2107</i>	426	60°	<i>Forwad</i>	5'- AGA GGT TCC GCG AGT GCG -3'	
				<i>Reverse</i>	5'- CAT CCT CAA TCG CGA CAG ATT CTG GA -3'	
proteasome subunit alpha type-7	<i>Preuni1020_1307</i>	747	60°	<i>Forwad</i>	5'- ACC CTC CGA AGC TAA AAA GCT CGA AT -3'	
				<i>Reverse</i>	5'- TGG AAA GCA TCC TCA GTT CGC ATC TAG -3'	
cold regulated 314 inner membrane 1	<i>Preuni4950_1349</i>	525	62°	<i>Forwad</i>	5'- TCA CCA CCT GCG AGC TTT CG -3'	
				<i>Reverse</i>	5'- GAA GCC CGA CTT AAA GCT CTG CC -3'	
histone H4-like	<i>Preuni17184_671</i>	309	60°	<i>Forwad</i>	5'- GAA AGA GGA AGC TCA AGC AAT TTT CCC G -3'	
				<i>Reverse</i>	5'- CCT CCG ACC CTG TAA ATT CAC TAG TCC -3'	
Cbs domain protein	<i>Preuni701_2337</i>	606	60°	<i>Forwad</i>	5'- TCC GTT CAA GAG TGA AAA GGC TGG A -3'	
				<i>Reverse</i>	5'- CCA TAA TTC AAG CAG GGC CAT CAT TGG -3'	
osmotic stress-activated protein kinase (OSAK)	<i>Preuni5552_1340</i>	699	60°	<i>Forwad</i>	5'- GCT TGA GTA TTA TGT AGC CAT GCC TAA CCA TTC -3'	
				<i>Reverse</i>	5'- AAT TGC CAA ACC GAC ATC TCT CTT CCT -3'	
phosphoglycerate kinase, chloroplastic isoform 1	<i>Preuni328_2200</i>	873	60°	<i>Forwad</i>	5'- CAT GTG GGT TCC AAG CTC CGA TC -3'	
				<i>Reverse</i>	5'- CGA AGC AGT TGC ATT TTC ACT TGC ACA TAT -3'	
plasma intrinsic protein	<i>Preuni7330_1053</i>	483	60°	<i>Forwad</i>	5'- GGG CTG TTC TTG GCG AGG A -3'	
				<i>Reverse</i>	5'- CTC GGA CAC ACC TCA AGC CAG -3'	
serine/threonine-protein kinase SAPK10 isoform 2	<i>Preuni8597_2115</i>	615	60°	<i>Forwad</i>	5'- GCT CGA TCA TGT GGA AGC TCC G -3'	
				<i>Reverse</i>	5'- TCA TAT ACG TGG ATC TCA GAC CTA TGA ATG GG -3'	
P5CS	<i>Preuni P5</i>	2199	56°	<i>Forwad</i>	5' - CGA CGG GAA CAT GGA CTC - 3'	
				<i>Reverse</i>	5' - CCC TTT ATC ACC ATC CAC CA - 3'	

trehalose_6-phosphate_synthase	<i>Preuni7011_3232</i>	2583	60°	<i>Forwad</i>	5' - CTC GGT TTT GTT CTG TGC TTC - 3'
				<i>Reverse</i>	5' - AAT TCC CAG ATG CCA ATC CT - 3'
sodium-dependent phosphate transport protein 1, chloroplastic	<i>Preuni3427_2484</i>	1170	60°	<i>Forwad</i>	5' - TGT TCA GGT TTG GTC ACA TTC T - 3'
				<i>Reverse</i>	5' - CCC AAC GCT CTT GTT CCT TA - 3'
pentatricopeptide repeat-containing protein At1g15510	<i>Preuni9300_1515</i>	1164	60°	<i>Forwad</i>	5' - TTC GGC AAA TGA AAC TCA GA - 3'
				<i>Reverse</i>	5' - GAG CGA GGA AGC ACA ACT CT - 3'
BEL1-like homeodomain protein 1	<i>Preuni2262_3123</i>	2112	56°	<i>Forwad</i>	5' - GGC GAC TTA CTT CCA CCA GA - 3'
				<i>Reverse</i>	5' - CAA CTC CGA TGA CGA AGA GAG - 3'
3-ketoacyl-CoA synthase 11	<i>Preuni962_2650</i>	1407	60°	<i>Forwad</i>	5' - TTC CTC TTC CTC TCC CCA TT - 3'
				<i>Reverse</i>	5' - TCG CTA CAA GTC CAC GAG AA - 3'
glutathione peroxidase (gpx)	<i>Preuni980_1209</i>	759	60°	<i>Forwad</i>	5' - CCG CAA ATG CTC TCT TCT TC - 3'
				<i>Reverse</i>	5' - TTA CTG GGG AAA AGA AGT TCC A - 3'

Table S3. Monthly data for the four climatic variables obtained for the two main populations

AVERAGE RAINFALL												
	JAN	FEB	MAR	APR	MAY	JUN	JUL	AGO	SEP	OCT	NOV	DEC
RJRE	227,719	171,301	170,237	104,567	63,5877	46,0071	34,4052	38,9513	81,1135	99,7394	146,489	220,808
RSRF	154,398	183,63	122,355	176,068	175,517	149,649	145,842	118,092	166,119	239,37	150,311	156,159

AVERAGE TEMPERATURE												
	JAN	FEB	MAR	APR	MAY	JUN	JUL	AGO	SEP	OCT	NOV	DEC
RJRE	25,419	25,7323	25,3323	23,7903	21,3258	20,0226	19,6355	20,2548	20,9	22,3065	23,7258	24,5194
RSRF	25,8968	25,6452	24,6161	21,3355	17,729	15,7065	15,429	17,4581	18,5645	21,6645	23,3	24,9097

MAXIMUM TEMPERATURE AVERAGE												
	JAN	FEB	MAR	APR	MAY	JUN	JUL	AGO	SEP	OCT	NOV	DEC
RJRE	31,508	32,2999	31,039	29,8696	27,7184	26,9125	26,3829	27,1467	27,0771	28,2312	29,2128	30,6675
RSRF	32,346	31,9395	30,9485	27,6858	23,8496	21,73	21,9602	24,3679	25,3395	28,0812	29,9327	31,5312

MINIMUM TEMPERATURE AVERAGE												
	JAN	FEB	MAR	APR	MAY	JUN	JUL	AGO	SEP	OCT	NOV	DEC
RJRE	24,091	24,3656	23,8548	22,7481	20,5287	19,4603	18,6876	19,2024	19,6854	21,0757	22,1684	23,4574
RSRF	19,9283	19,8827	18,4967	15,4757	12,1709	10,9735	9,84796	11,1563	12,7821	16,0054	16,8988	19,191

CONSIDERAÇÕES FINAIS E PERSPECTIVAS

Utilizar metodologias que envolvem novas técnicas de sequenciamento e seleção de genes candidatos tem sido inovador para o estudo e o entendimento de mecanismos de adaptação e evolução de organismos. Com estas tecnologias, surgem possibilidades de acessar regiões genômicas de interesse através de uso de marcadores moleculares específicos. A utilização de marcadores moleculares SNPs permite acessar loci que estejam sob seleção, seja em relação a mecanismos de adaptação ou como ferramenta para auxiliar no entendimento da história demográfica das espécies. Considerando estas aplicações, o objetivo deste estudo foi identificar, através da seleção de genes candidatos, SNPs envolvidos com a estrutura e variação genética adaptativa em populações naturais de *E. uniflora*.

O capítulo que compõe esta dissertação sugere duas estruturas populacionais, identificando e separando dois grupos nas populações amostradas: Sudeste (RJ) e Sul (RS). Estes dois grupos também caracterizam formações vegetacionais bastante distintas, submetidas a condições heterogêneas de clima, solo e composição geomorfológica. Estes resultados corroboram estudos anteriores em *E. uniflora*, utilizando marcadores do cloroplasto, que demonstram a separação destes grupos em duas unidades genéticas bem estruturadas.

Novas inquietações surgem com a perspectiva de explorar outras análises no contexto demográfico através da identificação de SNPs em regiões intrônicas, além da análise de associações entre genótipo e clima.

Estes dados reforçam a importância da obtenção e utilização de dados genéticos em estudos de conservação aplicados a ambientes fortemente ameaçados ou degradados, como ocorre na Floresta Atlântica, local de distribuição natural de *E. uniflora* que vem se mostrando serem interessantes modelos para estudos com este enfoque.

REFERÊNCIAS INTRODUÇÃO E DISCUSSÃO

- Anderson JT, Willis JH and Mitchell-Olds T (2011) Evolutionary genetics of plant adaptation. *Trends Genet* 27:258–266. doi: 10.1016/j.tig.2011.04.001
- Angers B, Castonguay E and Massicotte R (2010) Environmentally induced phenotypes and DNA methylation: how to deal with unpredictable conditions until the next generation and after. *Mol Ecol* 19:1283–1295. doi: 10.1111/j.1365-294X.2010.04580.x
- Aragón E, Goin FJ, Aguilera YE, Woodburne MO, Carlini A and Roggiero MF (2011) Palaeogeography and palaeoenvironments of northern Patagonia from the Late Cretaceous to the Miocene: The Palaeogene Andean gap and the rise of the North Patagonian High Plateau. *Biol J Linn Soc* 103:305–315. doi: 10.1111/j.1095-8312.2011.01684.x
- Barros MJF and Morim MP (2014) *Senegalia* (Leguminosae, Mimosoideae) from the Atlantic Domain, Brazil. *Syst Bot* 39:452–477. doi: 10.1600/036364413X680807
- Behling H (1997) Late quaternary vegetation, climate and fire history of the Araucaria forest and campos region from Serra Campos Gerais, Parana State (South Brazil). *Rev Palaeobot Palynol* 97:109–121. doi: 10.1016/S0034-6667(96)00065-6
- Behling H (1998) Late Quaternary vegetational and climatic changes in Brazil. *Rev Palaeobot Palynol* 99:143–156. doi: 10.1016/S0034-6667(97)00044-4
- Behling H (2002) South and southeast Brazilian grasslands during Late Quaternary times: A synthesis. *Palaeogeogr Palaeoclimatol Palaeoecol* 177:19–27. doi: 10.1016/S0031-0182(01)00349-2
- Black IV WC, Baer CF, Antolin MF and DuTeau NM (2001) POPULATION GENOMICS : Genome-Wide Sampling of Insect Populations. *Annu Rev Entomol* 46:441–469. doi: 10.1146/annurev.ento.46.1.441
- Brookes AJ (1999) The essence of SNPs. *Gene* 234:177–186. doi: 10.1016/S0378-1119(99)00219-X
- Carnaval AC and Moritz C (2008) Historical climate modelling predicts patterns of current biodiversity in the Brazilian Atlantic forest. *J Biogeogr* 35:1187–1201. doi: 10.1111/j.1365-2699.2007.01870.x
- Chinnusamy V and Zhu J-K (2009) Epigenetic regulation of stress responses in plants. *Curr Opin Plant Biol* 12:133–139. doi: 10.1016/j.pbi.2008.12.006
- Colombo AF & Joly CA (2010) Brazilian Atlantic Forest lato sensu: the most ancient Brazilian forest, and a biodiversity hotspot, is highly threatened by climate change. *Braz J Biol* 70:697–708. doi: 10.1590/S1519-69842010000400002

- Consolini AE and Sarubbio MG (2002) Pharmacological effects of *Eugenia uniflora* (Myrtaceae) aqueous crude extract on rat's heart. *J Ethnopharmacol* 81:57–63. doi: 10.1016/S0378-8741(02)00039-9
- Da Fonseca GAB (1985) The vanishing Brazilian Atlantic forest. *Biol Conserv* 34:17–34. doi: 10.1016/0006-3207(85)90055-2
- De Mello Martins F (2011) Historical biogeography of the Brazilian Atlantic forest and the Carnaval-Moritz model of Pleistocene refugia: What do phylogeographical studies tell us? *Biol J Linn Soc* 104:499–509. doi: 10.1111/j.1095-8312.2011.01745.x
- De Wit P, Pespeni MH and Palumbi SR (2015) SNP genotyping and population genomics from expressed sequences -current advances and future possibilities. *Mol Ecol* 24:n/a–n/a. doi: 10.1111/mec.13165
- Elmer KR and Meyer A (2011) Adaptation in the age of ecological genomics: insights from parallelism and convergence. *Trends Ecol Evol* 26:298–306. doi: 10.1016/j.tree.2011.02.008
- Feder ME and M-OT (2003) Evolutionary and ecological functional genomics. *Nat Rev Genet* 4:651–657. doi: 10.1038/nrg1128
- Ferreira-Ramos R, Laborda PR, de Oliveira Santos M, Mayor MS, Mestriner M a., de Souza AP and Alzate-Marin AL (2008) Genetic analysis of forest species *Eugenia uniflora* L. through of newly developed SSR markers. *Conserv Genet* 9:1281–1285. doi: 10.1007/s10592-007-9458-0
- Fittkau EJ (1969) *Biogeography and Ecology in South-America, Volume 2.*
- Golldack D, Lüking I and Yang O (2011) Plant tolerance to drought and salinity: stress regulating transcription factors and their functional significance in the cellular transcriptional network. *Plant Cell Rep* 30:1383–1391. doi: 10.1007/s00299-011-1068-0
- Holderegger R, Kamm U and Gugerli F (2006) Adaptive vs. neutral genetic diversity: implications for landscape genetics. *Landsc Ecol* 21:797–807. doi: 10.1007/s10980-005-5245-9
- Joly C, Aidar M, Klink C, Mc Grath DG, Moreira a. G, Moutinho P, Nepstad DC, Oliveira a. a., Pott a., Rodal MJN et al. (1999) Evolution of the Brazilian phytogeography classification systems: implications for biodiversity conservation. *Cienc Cult* 51:331–348.
- Lalagüe H, Csilléry K, Oddou-Muratorio S, Safrana J, de Quattro C, Fady B, González-Martínez SC and Vendramin GG (2014) Nucleotide diversity and linkage disequilibrium at 58 stress response and phenology candidate genes in a European beech (*Fagus sylvatica* L.) population from southeastern France. *Tree Genet Genomes* 10:15–26. doi: 10.1007/s11295-013-0658-0
- Legrand, C. D., Klein, R. M., & Myrciaria P (1978) Espécies suplementares, Espécies cultivadas, Generalidades.(Fasc. Mirt.) In REITZ, R. *Flora Ilus Catarinense* 733–876.

- Lichte, M., & Behling H (1999) Dry and cold climatic conditions in the formation of the present landscape in Southeastern Brazil-An interdisciplinary approach to a controversially discussed topic. *Zeitschrift Fur Geomorphol* 43(3):341–358.
- Lowry D (2010) Landscape evolutionary genomics. *Biol. Lett.* 6:502–4
- Luikart G, England PR, Tallmon D, Jordan S and Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing. *Nat Rev Genet* 4:981–994. doi: 10.1038/nrg1226
- Madlung a. (2004) The Effect of Stress on Genome Regulation and Structure. *Ann Bot* 94:481–495. doi: 10.1093/aob/mch172
- Margis R, Felix D, Caldas JF, Salgueiro F, De Araujo DSD, Breyne P, Van Montagu M, De Oliveira D and Margis-Pinheiro M (2002) Genetic differentiation among three neighboring Brazil-cherry (*Eugenia uniflora* L.) populations within the Brazilian Atlantic rain forest. *Biodivers Conserv* 11:149–163. doi: 10.1023/A:1014028026273
- Mayle FE, Burn MJ, Power M and Urrego DH (2009) Past Climate Variability in South America and Surrounding Regions. 14:89–112. doi: 10.1007/978-90-481-2672-9
- Mazine FF (2006) Estudos taxonômicos em *Eugenia* L. (Myrtaceae), com ênfase em *Eugenia* sect. *Racemosae* O. Berg.
- Mazine FF, Souza VC, Sobral M, Forest F and Lucas E (2014) A preliminary phylogenetic analysis of *Eugenia* (Myrtaceae: Myrteae), with a focus on Neotropical species. *Kew Bull* 69:1–14. doi: 10.1007/s12225-014-9497-x
- Morrone JJ (2004) Panbiogeografía, componentes bióticos y zonas de transición. *Rev Bras Entomol* 48:149.
- Morrone JJ (2006) Biogeographic Areas and Transition Zones of Latin America and the Caribbean Islands Based on Panbiogeographic and Cladistic Analyses of the Entomofauna. *Annu Rev Entomol* 51:467–494. doi: 10.1146/annurev.ento.50.071803.130447
- Myers N, Mittermeier R a, Mittermeier CG, da Fonseca G a and Kent J (2000) Biodiversity hotspots for conservation priorities. *Nature* 403:853–8. doi: 10.1038/35002501
- Oliveira-Filho A and Fontes M (2000) Patterns of Floristic Differentiation among Atlantic Forests in Southeastern Brazil and the Influence of Climate1. *Biotropica* 32:793–810. doi: 10.1111/j.1744-7429.2000.tb00619.x
- Rattmann YD, De Souza LM, Malquevicz-Paiva SM, Dartora N, Sasaki GL, Gorin P a J and Iacomini M (2012) Analysis of flavonoids from *eugenia uniflora* leaves and its protective effect against murine sepsis. *Evidence-based Complement Altern Med.* doi: 10.1155/2012/623940
- Ribeiro J (2000) *Flora e População Ribeirinha.*

- Ribeiro MC, Metzger JP, Martensen AC, Ponzoni FJ and Hirota MM (2009) The Brazilian Atlantic Forest: How much is left, and how is the remaining forest distributed? Implications for conservation. *Biol Conserv* 142:1141–1153. doi: 10.1016/j.biocon.2009.02.021
- Rice AM, Rudh A, Ellegren H and Qvarnström A (2011) A guide to the genomics of ecological speciation in natural animal populations. *Ecol Lett* 14:9–18. doi: 10.1111/j.1461-0248.2010.01546.x
- Rotman AD (1995) Las especies argentinas del género *Eugenia* (Myrtaceae). *Bol Soc Argent Bot* 31(1-2):69–93.
- Sahu PP, Pandey G, Sharma N, Puranik S, Muthamilarasan M and Prasad M (2013) Epigenetic mechanisms of plant stress responses and adaptation. *Plant Cell Rep* 32:1151–1159. doi: 10.1007/s00299-013-1462-x
- Salgueiro F, Felix D, Caldas JF, Margis-Pinheiro M and Margis R (2004) Even population differentiation for maternal and biparental gene markers in *Eugenia uniflora*, a widely distributed species from the Brazilian coastal Atlantic rain forest. *Divers Distrib* 10:201–210. doi: <http://dx.doi.org/10.1111/j.1366-9516.2004.00078.x>
- Schapoval EES, Silveira SM, Miranda ML, Alice CB and Henriques a. T (1994) Evaluation of some pharmacological activities of *Eugenia uniflora* L. *J Ethnopharmacol* 44:137–142. doi: 10.1016/0378-8741(94)01178-8
- Schoville SD, Barreto FS, Moy GW, Wolff A and Burton RS (2012) Investigating the molecular basis of local adaptation to thermal stress: population differences in gene expression across the transcriptome of the copepod *Tigriopus californicus*.
- Sobral M (2003) A família Myrtaceae no Rio Grande do Sul.
- Sobral M, Anália M, Souza D De, Santos MF and Almeida TE De (2015) Seven new records of Myrtaceae in Brazil. 11:
- Stehmann, J. R. et al (2009) Diversidade taxonômica na Floresta Atlântica. *Plantas da Floresta Atlântica*, 3-12.
- Turchetto-Zolet a. C, Pinheiro F, Salgueiro F and Palma-Silva C (2013) Phylogeographical patterns shed light on evolutionary process in South America. *Mol Ecol* 22:1193–1213. doi: 10.1111/mec.12164
- Wilson CA (2011) Subgeneric classification in *Iris* re-examined using chloroplast sequence data. *Taxon*
- Wilson PG, Brien MMO, Heslewood MM and Quinn CJ (2005) Relationships within Myrtaceae sensu lato based on a mat K phylogeny. 251:3–19. doi: 10.1007/s00606-004-0162-y