Universidade Federal do Rio Grande do Sul

Instituto de Biociências

Programa de Pós-Graduação em Genética e Biologia Molecular

GENÉTICA DE PAISAGENS DE ESPÉCIES DA PLANÍCIE COSTEIRA DO ATLÂNTICO SUL

Gustavo Adolfo Silva-Arias

Tese submetida ao Programa de Pós-Graduação em Genética e Biologia Molecular da UFRGS como requisito parcial para obtenção do grau de Doutor em Ciências

Orientadora: Dra. Loreta Brandão de Freitas

Porto Alegre, Julho de 2016

INSTITUIÇÕES E FONTES FINANCIADORAS

Este trabalho foi realizado nas dependências do Laboratório de Evolução Molecular, Departamento de Genética – UFRGS. Parte dos experimentos de sequenciamento e genotipagem foi desenvolvida no Laboratório de Biologia Genômica e Molecular, PUCRS, sob a coordenação do Dr. Sandro L. Bonatto. O projeto foi financiado com recursos do CNPq, CAPES e PPGBM-UFRGS. A bolsa de doutorado foi concedida pelo Departamento Administrativo de Ciencia, Tecnología e Innovación (COLCIENCIAS) da Colômbia.

AGRADECIMENTOS

O sucesso na pesquisa científica sempre está ligado a uma rede de colaborações financeiras, técnicas, acadêmicas e pessoais. Com frequência a falta de alguma destas estará refletida na perda de qualidade ou impossibilidade de atingir os objetivos. Por isso que é sempre importante dar o devido reconhecimento a todos os que de alguma forma contribuíram para o desenvolvimento do trabalho. Em primeiro lugar quero agradecer a minha orientadora Loreta B. Freitas que, mesmo desde antes de começar e durante todo o doutorado me proporcionou estímulo, confiança e todos os recursos à disposição para crescer intelectualmente e por em prática todos meus conhecimentos e experiências neste e em outros projetos de nosso grupo de pesquisa. Também quero agradecer a todos os professores que participaram nos trabalhos que desenvolvi durante o doutorado. Ao professor Heinrich Hasenack que me ajudou em toda a pesquisa climática da região proporcionando-me seu conhecimento e informação. Ao professor Sandro Bonatto que colaborou enormemente na discussão de um dos artigos que compõem esta tese. Ao professor Bryan Carstens que me proveu de numerosos recursos técnicos e conceituais para desenvolver este trabalho e ademais por ter me recebido no seu laboratório durante um intercâmbio que enriqueceu imensamente minha formação e perspectiva da atividade científica. Quero agradecer a importante colaboração das colegas Maikel Reck-Kortmann e Giovanna Giudicelli que prestaram uma importantíssima ajuda na coleta de dados. Agradeço também aos colegas do laboratório Alice, Daniele, Carol, Ana Lúcia, Geraldo, Sara, Ana Laura, Lauís, Verônica, Marcelo e Michel por sua amabilidade e receptividade, pela companhia durante estes quatro anos e pelas boas discussões.

Finalmente agradeço a minha família pelo ânimo e apoio constante, por acreditar em minhas capacidades e pela constante presença na minha vida sem importar nada mais. Agradeço à Lina, o amor da minha vida que como esposa, companheira e colega se converteu no suporte mais importante para me permitir êxito como pessoa e como profissional. A minha filha Julieta que com sua alegria e espontaneidade todos os dias me lembra do significado real da vida, e sua presença me estimula a procurar sucesso e boa qualidade de vida. A minha mãe que todos os dias está ao meu lado ajudando, animando e orientando minha vida. Ao meu pai que agora desde o infinito me acompanha em todas as nossas aventuras na procura de expandir ao máximo nossas fronteiras físicas e intelectuais. E a toda minha família que sempre está na torcida e aguardando nossos esporádicos regressos à casa para compartilhar sorrisos, boas comidas e muito amor.

"Depois de marchar por sete dias através das matas, quem vai a Bauci não percebe que já chegou. As finas andas que se elevam do solo a grande distância uma da outra e que se perdem acima das nuvens sustentam a cidade. Sobe-se por escadas. Os habitantes raramente são vistos em terra: têm todo o necessário lá em cima e preferem não descer. Nenhuma parte da cidade toca o solo exceto as longas pernas de flamingo nas quais ela se apoia, e, nos dias luminosos, uma sombra diáfana e angulosa que se reflete na folhagem.

Há três hipóteses a respeito dos habitantes de Bauci: que odeiam a terra; que a respeitam a ponto de evitar qualquer contato; que a amam da forma que era antes de existirem e com binóculos e telescópios apontados para baixo não se cansam de examiná-la, folha por folha, pedra por pedra, formiga por formiga, contemplando fascinados a própria ausência".

Italo Calvino, 1972. As Cidades Invisíveis.

SUMÁRIO

RESUMO

O entendimento da contribuição diferencial de processos neutros e adaptativos envolvidos na diferenciação genética entre populações, assim como sua relação com varáveis físicas e ambientais da área de distribuição das espécies, é fundamental para melhorar o conhecimento da história evolutiva, mas também para fazer um manejo e conservação mais adequados da diversidade genética das espécies. O surgimento da Planície Costeira do Atlântico Sul foi um processo relativamente recente, que conduziu a processos de colonização e expansão dos organismos para um ambiente costeiro. Os padrões de estrutura genética gerados em processos de colonização e expansão podem ser difíceis de interpretar devido ao fato de que podem apresentar sinais sobrepostos de efeito fundador em série, isolamento por distância e isolamento por ambiente quando envolvem gradientes ecológicos na área de estudo. No presente trabalho foram conduzidas caracterizações da diversidade e estrutura genética de dois *taxa* predominantemente costeiros co-distribuídos, *Calibrachoa heterophylla* e *Petunia integrifolia* ssp. *depauperata,* em toda a amplitude da distribuição. Também foram inferidas as dinâmicas de fluxo gênico entre populações e sua relação com variáveis topográficas e climáticas reconstruídas pelo meio de um levantamento exaustivo e modelamento para a área de estudo. Processos de diferenciação genética promovidos pelo regime diferencial de chuvas nos extremos da distribuição foram inferidos para as duas espécies. Também foram identificadas populações das duas espécies apresentando alto nível de mistura de identidade genética nas localidades ao redor da Lagoa dos Patos. Isso foi associado a alta instabilidade na história geomorfológica recente desta região e dinâmicas atuais do vento que favorecem a dispersão secundária de sementes a maiores distâncias. Adicionalmente foram identificados processos espécie-específicos que se relacionaram principalmente a

fatores históricos de cada táxon. Em *P. depauperata* o efeito fundador relacionado a um processo único de colonização do ambiente costeiro determinou o nível superior de estrutura genética, enquanto que em *C. heterophylla* foi a história filogeográfica da espécie na qual a diferenciação intraespecífica é anterior à colonização da região costeira atual o fator preponderante. As diferenças de duração do ciclo de vida entre as espécies também influenciaram as dinâmicas contrastantes de fluxo gênico dos dois *taxa*, sugerindo que a colonização e adaptação local de *C. heterophylla* nas bordas da distribuição poderia ser condizente com um processo de monopolização. Em vista dos resultados obtidos neste trabalho, propõem-se o desenvolvimento de experimentos de transplante recíproco para confirmar o processo de adaptação local nas duas espécies e abordagens genômicas para identificar regiões do genoma responsáveis pelos processos de adaptação ao ambiente costeiro e de adaptação local nas margens da distribuição.

Palavras chave

Adaptação local, *Calibrachoa*, diferenciação ecológica, fluxo gênico, isolamento por ambiente, isolamento por colonização, isolamento por distância, microssatélites, *Petunia,* Solanaceae.

ABSTRACT

The understanding of differential contribution of neutral and adaptive processes to the genetic differentiation among populations, as well as its relationship to physical and environmental variables of species' distribution area, is essential to improve the knowledge of species evolutionary history, but also to direct appropriate management and conservation policies for the genetic diversity. The emergence of the South Atlantic Coastal Plain was a relatively recent event that led to colonization and expansion processes to the coastal environment. Genetic structure patterns generated in colonization and expansion processes can be difficult to interpret because the overlapping signals, which can present the founder effect in series, isolation by distance, and isolation by environment in the presence of ecological gradients in the study area. In this work characterization diversity and genetic structure were conducted to two co-distributed and predominantly coastal taxa, *Calibrachoa heterophylla* and *Petunia integrifolia* ssp. *depauperata* alongside their complete geographical range. Moreover, we also inferred dynamic of gene flow among populations and investigated the relation between topographical and climatic variables reconstructed by means of an exhaustive survey and modeling for the study area and the gene flow. Shared genetic differentiation processes promoted by differential rainfall conditions at the distribution edges were inferred. In addition, we identified populations from both species with high level of mixed genetic membership in locations around the Patos Lagoon. This was associated with a high instability in recent geomorphological history of coastal region and current wind dynamics that favor the secondary seed dispersal over longer distances. Additionally, specific species processes were identified mainly related to historical factors of each taxon. In

P. depauperata founder effects associated with unique colonization process to coastal environment determined the upper level of genetic structure, while in *C. heterophylla* the upper level of genetic structure was related to the phylogeographical history wherein the intra-specific differentiation preceded colonization to the current coastal region. The differences of the life span length of the species were also related to contrasting gene flow dynamics indicating that the colonization and local adaptation of *C. heterophylla* at the edges of the distribution could lead to monopolization process. In view of the results we propose the development of reciprocal transplant experiments to confirm the local adaptation process in both species and genomic approaches to identify regions of the genome responsible for the processes of adaptation to the coastal environment and local adaptation in distribution margins.

Key words

Calibrachoa, ecological differentiation, gene flow, isolation by environment, isolation by colonization, isolation by distance, local adaptation, microsatellite, *Petunia*, Solanaceae.

INTRODUÇÃO GERAL

Genética de paisagens: estrutura genética, fluxo gênico, ecologia de paisagens e conservação

Os estudos da genética de populações que explicitamente quantificam os efeitos da composição e configuração da paisagem sobre os processos micro-evolutivos tais como a deriva genética, fluxo gênico e seleção natural têm sido enquadrados na área interdisciplinar da genética de paisagens. Esta estuda associações genótipo-hábitat, baseadas em padrões da estrutura genética, fluxo gênico e isolamento reprodutivo contemporâneo entre indivíduos ou populações, assim como sua relação com características bióticas ou abióticas da paisagem, barreiras geográficas e diferenças no hábitat, as quais permitem gerar entendimento do movimento dos indivíduos ou gametas em relação à estrutura genética das populações, da influência do fluxo gênico para aumentar ou prevenir a diferenciação e da adaptação local, ou descrever as dinâmicas que facilitam a difusão de novas mutações benéficas (Manel et al. 2003; Holderegger and Wagner 2006; Storfer et al. 2007; Manel and Holderegger 2013; Balkenhol et al. 2016).

As análises de genética de paisagens têm implicações no entendimento de aspectos da ecologia, evolução e conservação da diversidade genética das espécies, e é por isso que permitem a identificação de unidades evolutivas significativas, assim como unidades de manejo ou de conservação. A incorporação de abordagens genômicas dentro da genética de paisagens (i.e., genômica populacional; (Black IV et al. 2001) tem permitido a identificação de '*outlier loci*', diferenciando assim os efeitos *locus*-específicos (seleção, mutação, acasalamento preferencial e

recombinação) dos efeitos genômicos (deriva genética, fluxo gênico, mudanças no tamanho populacional e endogamia) (Stinchcombe and Hoekstra 2008; Manel et al. 2010). A comparação da distribuição da diversidade genética observada em *loci* neutros e em *loci* sabidamente sob seleção permite o entendimento dos mecanismos que geram padrões espaciais e também pode ajudar na compreensão das bases genéticas da adaptação local, diferenciação adaptativa, especiação ecológica e melhorar a identificação de Unidades Evolutivas Significativas (UES). Anteriormente, esta identificação era baseada somente em diferenciação genética, sem considerar o processo envolvido no seu surgimento. Porém, com a aplicação de métodos próprios da genômica populacional, é possível conservar a diversidade genética com base no seu isolamento e também na diferenciação adaptativa (Crandall et al. 2000; Holderegger et al. 2006).

Os estudos de genética de paisagens, quando aplicados em escala da distribuição completa das espécies, precisam prestar maior atenção aos possíveis fatores históricos que podem influenciar a diferenciação genética. Por isso, as associações entre características da paisagem e padrões de estrutura genética em escala geográfica ampla acabam se assemelhando aos estudos filogeográficos, no sentido que procuram a identificação de barreiras visíveis ou crípticas (interrupções) ao fluxo gênico entre populações e também investigam processos de contato secundário entre populações previamente isoladas (Anderson et al. 2010; Dyer 2016).

Marcadores moleculares e ferramentas analíticas na genética de paisagens

Os marcadores moleculares mais utilizados em abordagens de genética de paisagens são os microssatélites (SSR) (Storfer et al. 2010). Os SSR são também os marcadores

mais empregados nas abordagens que precisam de informação em escala populacional. Estes se caracterizam por apresentar altas taxas de mutação, sendo bastante polimórficos e, por isso, permitem a identificação da subestrutura populacional em diferentes escalas espaciais (Kalia et al. 2010). Além disso, existe o conhecimento teórico de sua evolução molecular, permitindo o uso de modelos para inferir padrões de relacionamento genealógico (Schlötterer 2000; Ellegren 2004) e, no caso de microssatélites nucleares, cuja herança é codominante, permitem fazer inferências de dinâmicas micro-evolutivas (fluxo gênico e deriva genética), que podem revelar padrões de estrutura populacional e relacioná-los a aspectos históricos ou contemporâneos da interação entre populações e destas com fatores físicos do ambiente como o clima, barreiras geográficas e fragmentação do hábitat (Sork et al. 1999) ou com fatores biológicos como o sistema reprodutivo e a longevidade dos organismos (Duminil et al. 2009).

No nível de espécies, os microssatélites têm permitido inferir diversos tipos de processos evolutivos. Entre eles, destacam-se os estudos que permitem comparar padrões de diversidade e estrutura genética em hábitats conservados ou fragmentados (Dayanandan et al. 1999; Collevatti et al. 2001); compreender processos de alta diferenciação populacional sem especiação por conta de fenômenos de coesão (Barbará et al. 2007; Palma-Silva et al. 2011); identificar eventos de fluxo gênico interespecífico e hibridização (Pinheiro et al. 2010); compreender os componentes históricos e geográficos dos processos de domesticação (Matsuoka et al. 2002; Roullier et al. 2013); e estabelecer estratégias de conservação (Diniz-Filho et al. 2012). Adicionalmente, as análises de marcadores microssatélites permitem fazer comparações de padrões históricos e contemporâneos que tem servido para esclarecer os processos envolvidos na estruturação populacional (Viruel et al. 2012); obter

estimativas de fluxo gênico histórico e contemporâneo relacionados às taxas relativas de migração de pólen e sementes (Sork et al. 1999); e, em combinação com modelagens de nicho, reconstruir dinâmicas da distribuição das espécies ligadas a mudanças climáticas (Diniz-Filho et al. 2016).

Para escalas mais detalhadas, as recentes combinações de estatísticas espaciais, sistemas de informação geográfica e a obtenção de informação genética com numerosos marcadores moleculares altamente polimórficos, permitem a fusão da genética de populações e a ecologia de paisagens com o objetivo de melhorar o entendimento de aspectos biológicos e evolutivos das populações. As ferramentas de análise geográfico-espacial aplicadas a dados genéticos têm enriquecido a disciplina da genética de paisagens (Epperson 2003; Fortin 2005; Joost et al. 2007; Kozak et al. 2008; Dyer 2009; Etherington 2011; Vandergast et al. 2011). A partir do mapeamento espacial das frequências alélicas de uma ou mais populações e a posterior correlação dos padrões encontrados com características físicas e ambientais da paisagem, é possível avaliar a influência destes fatores sobre a diversidade e estruturação das populações (Diniz-Filho et al. 2009; Wei et al. 2013). Na busca de variação adaptativa, embora esteja sendo implementada com marcadores que apresentam maior quantidade de *loci* e abrangência genômica (AFLP e SNPs), alguns trabalhos baseados em EST-SSR e microssatélites anônimos têm conseguido sinais de seleção (Oetjen and Reusch 2007; Nosil et al. 2009; Ferrer et al. 2016).

Os estudos de genética de paisagens utilizam ferramentas estatísticas que medem a influência relativa da variação nas características da paisagem sobre a conectividade entre populações. As avaliações de estrutura genética através de análises de agrupamento são realizadas a partir do conjunto de dados genéticos com a finalidade de inferir populações ou grupos de populações, bem como as possíveis

barreiras ao fluxo gênico. Já na avaliação da influência da paisagem, os padrões de agrupamento podem ser visualmente sobrepostos à paisagem ao longo da distribuição da espécie resultando em análises descritivas do efeito das características da paisagem sobre a estruturação genética (Frantz et al. 2012). Os tipos de análises de agrupamento se diferenciam por dois atributos principais: (i) pela utilização ou não de modelos de genética de populações que descrevem a distribuição de alelos em populações estruturadas (e.g., equilíbrio de Hardy–Weinberg); e (ii) pela utilização ou não da informação espacial na definição dos grupos (François and Waits 2016). Adicionalmente, a estratégia de definição do número de grupos pode influenciar em grande medida as inferências sobre a estrutura populacional (Puechmaille 2016). Devido à variedade de abordagens e seu ajuste às perguntas de cada abordagem em particular, tem sido recomendado explorar várias ferramentas para fazer agrupamentos, procurar as mais adequadas, analisar o grau de confiabilidade estatística e o sentido biológico dos resultados (Manel et al. 2005; François and Durand 2010).

Em resposta à necessidade de que as análises de genética de paisagens avaliem de forma quantitativa e espacialmente explícita os efeitos do ambiente na conectividade entre as populações (Storfer et al. 2010), recentemente tem-se incrementado o uso de estatísticas que avaliam e medem a influência de variáveis espaciais e ambientais na diferenciação ou conectividade genética. As estatísticas mais utilizadas são o teste de Mantel (Mantel 1967) e a regressão entre matrizes de distância, os quais têm sido muito utilizados na genética de populações para avaliar padrões de isolamento por distância (IBD, pela sigla em inglês). Além do estabelecimento da estrutura espacial da variação genética, é imperativo estabelecer de forma inicial se um conjunto de dados apresenta IBD, porque este tipo de padrão

de diferenciação pode enviesar os resultados das análises de estruturação populacional e detecção de *loci* sob seleção (Frantz et al. 2009; Meirmans 2012). Neste mesmo sentido, o teste de Mantel parcial (Smouse et al. 1986) é usado para testar a dependência entre duas matrizes de distância (usualmente genéticas, morfológicas ou ambientais) controlando pelo efeito de uma terceira matriz (usualmente de distâncias geográficas). Numerosos estudos têm aplicado o teste de Mantel parcial para avaliar de modo espacialmente explícito se condições físicas ou ambientais têm influência sobre a variação genética, elevando este teste à condição de um dos métodos mais populares na genética de paisagens (Storfer et al. 2010). No entanto, avaliações recentes do desempenho do teste de Mantel e suas variações têm resultado em críticas relacionadas com a falta de poder estatístico (habilidade de um método de detectar um efeito quando este está presente nos dados) e à alta taxa de erro do tipo I (resultados com *p*-valores erroneamente pequenos) com o incremento da autocorrelação entre as variáveis (Balkenhol et al. 2009; Legendre and Fortin 2010; Guillot and Rousset 2013).

Porém, apesar das críticas, considera-se que o teste de Mantel e estatísticas relacionadas ainda constituem uma ferramenta útil para análises exploratórias, embora seja recomendo complementar com outras análises (Diniz-Filho et al. 2013; Cushman et al. 2013). Recentemente têm sido descritas numerosas alternativas que apresentam grande potencial para enriquecer as análises de genética de paisagens superando os problemas de multicolinearidade entre as variáveis ecológicas e espaciais (Prunier et al. 2015; Richardson et al. 2016). Entre estas, encontram-se as análises de regressão múltipla de matrizes de distâncias (Lichstein 2007; Wang 2013; Balkenhol et al. 2014); as regressões múltiplas baseadas em mapas de autovetores de Moran (ou em inglês: *Moran's Eigenvector Maps*; (Dray et al. 2006; Legendre et al. 2015); os

modelos lineares mistos generalizados (ou em inglês: *Generalized linear mixed models*), que têm ajudado a encontrar relações entre variáveis ambientais e variação genética levando em conta a autocorrelação espacial (Bolker et al. 2009; Van Strien et al. 2012); e as técnicas de ordenamento condicionado (análises de redundância ou de correspondência canônica) usadas para avaliar a contribuição relativa de um preditor ou um conjunto destes sobre a variância de uma variável dependente (Legendre and Anderson 1999; Angers et al. 1999; Balkenhol et al. 2009). Tem sido demonstrado que as análises multivariadas que incluem variáveis de paisagem conseguem explicar melhor as variações de distância genética entre as populações (Storfer et al. 2010).

Finalmente, as revisões mais recentes têm enfatizado a importância de implementar os estudos da genética de paisagens pelo meio de abordagens baseadas em hipóteses. Entre os benefícios de elaborar hipóteses *a priori* encontram-se a possibilidade de realizar estratégias de amostragem focalizadas nas variáveis de interesse, produzir estimativas mais precisas dos parâmetros, minimizar a influência de correlações espúrias e testar rigorosamente as hipóteses (Storfer et al. 2010; Richardson et al. 2016).

Planície Costeira do Atlântico Sul

Considerando que as características físicas e ambientais dos locais onde as espécies ocorrem determinam sua distribuição geográfica, o tamanho das populações, influenciam os padrões de estrutura genética e as dinâmicas de fluxo gênico, neste trabalho foi desenvolvida uma revisão detalhada das características geomorfológicas e climáticas da Planície Costeira do Atlântico Sul (PCAS) para permitir um melhor

entendimento dos processos evolutivos históricos e contemporâneos dos organismos que ocorrem nesta região.

Geomorfologia

A PCAS faz parte de uma província fisiográfica composta pelos escudos sul-rio-grandense e uruguaio (embasamento cristalino pré-cambriano) e a Bacia de Pelotas. Esta última constitui um sistema deposicional que tem acumulado mais de 10.000 m de sedimentos, principalmente continentais, desde sua formação no período Cretáceo Tardio. A seção mais jovem deste acúmulo compõe a PCAS, que se caracteriza por ser uma alongada (620 km), ampla (ca. 100 km) e contínua planície de terras baixas, localizada em uma região costeira aberta, que contém um dos registros mais completos do processo deposicional do período Quaternário das costas do Brasil (Tomazelli and Dillenburg 2007) (Figura 1). A PCAS compõe um traçado de costa regular, com orientação sudoeste (SO) - nordeste (NE), que se estende até o NE no promontório rochoso do Farol de Santa Marta (estado de Santa Catarina), até o SO no promontório rochoso de Cabo Polônio (Uruguai). A PCAS é cortada por cinco canais perenes, o canal de Rio Grande, o rio Tramandaí, o Rio Araranguá, o Rio Mampituba e o Arroio Chui.

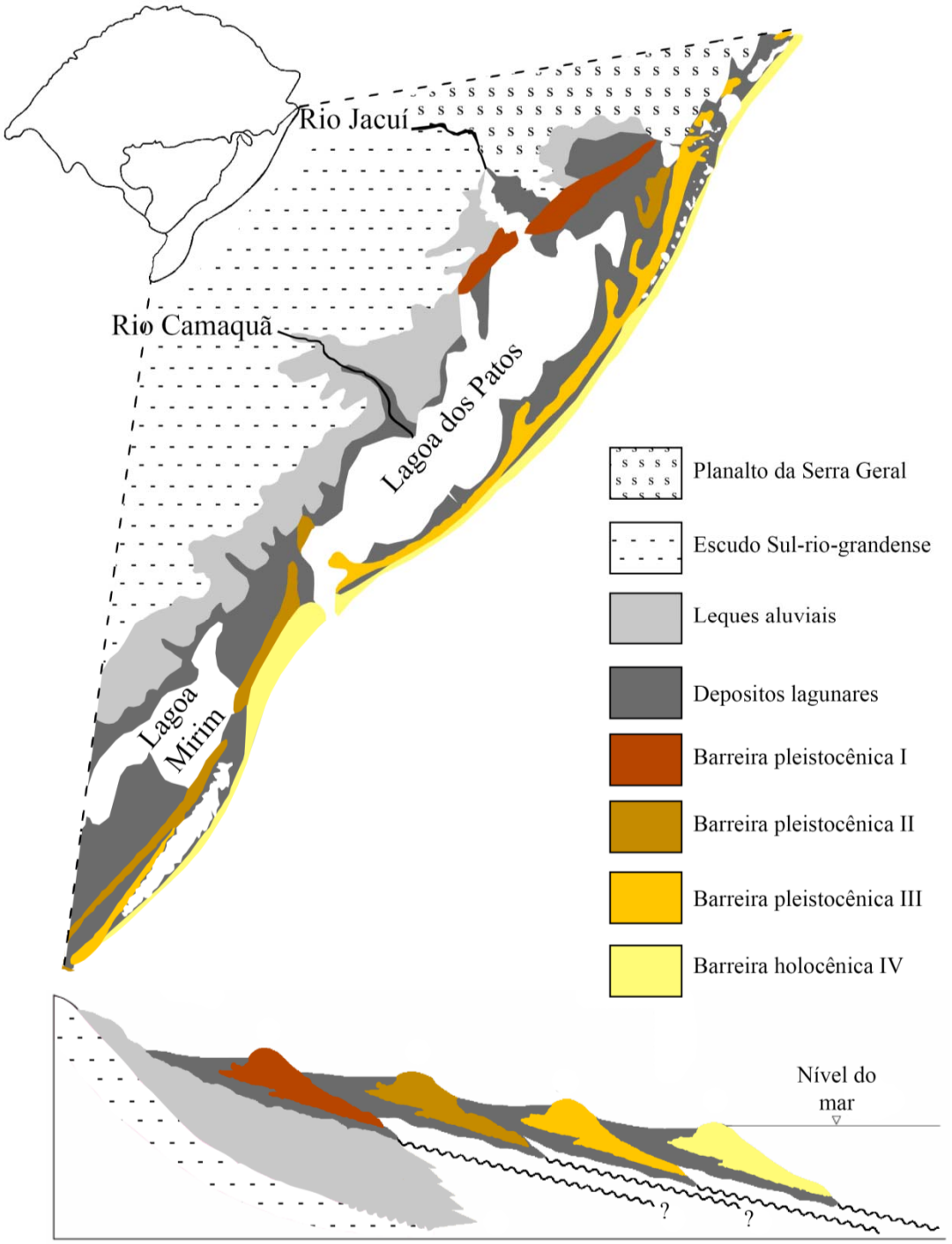


Figura 1. Representação esquemática da geomorfologia da Planície Costeira do Atlântico Sul modificado de (Tomazelli and Villwock 2005; Tomazelli and Dillenburg 2007) em vista horizontal (painel superior) e em secção transversal (painel inferior).

A geomorfologia da PCAS está caracterizada por dos sistemas deposicionais: o sistema de leques aluviais e o sistema do tipo laguna-barreira (Tomazelli and Villwock 2005); Figura 1). O sistema de leques aluviais se encontra a oeste dos grandes corpos de água da PCAS e marca a transição entre as terras altas continentais e os sistemas de tipo laguna-barreira. O sistema de leques está subdividido em dois tipos: os leques ao norte da PCAS, que recebem sedimentos do Planalto da Serra Geral provenientes da Bacia de Paraná (terras altas formadas por rochas vulcânicas e sedimentárias com origem no Paleozóico e Mesozóico, que localmente alcançam altitudes de 1000 m); e os leques de toda a margem oeste da PCAS situada ao sul da cidade de Porto Alegre (RS), que recebem sedimentos do escudo Sul-rio-grandense (escudo Pré-cambriano formado por rochas ígneas e metamórficas; (Tomazelli and Villwock 2005); Figura 1).

Atualmente os sedimentos arenosos, produto da erosão, são depositados no sistema de lagunas costeiras que dominam a maior parte da PCAS, algumas de grandes dimensões como a Lagoa dos Patos e a Lagoa de Mirim (Figura 1). Mas em estágios iniciais da formação da PCAS, os principais rios da região, como o Jacuí, Camaquã e Jaguarão, fragmentavam a costa e desaguavam diretamente no Oceano Atlântico (Weschenfelder et al. 2010; Santos-Fischer et al. 2016).

A característica geomorfológica mais relevante da PCAS é a presença de sistemas deposicionais do tipo laguna-barreira que, no flanco continental, apresentam os extensos depósitos lagunares costeiros já mencionados. O sistema apresenta quatro barreiras arenosas paralelas à linha de costa que tiveram origem nos processos de transgressão e regressão do nível do mar, relacionados com os períodos interglaciais e glaciais ocorridos no Pleistoceno e Holoceno. A formação da Laguna-Barreira I, a mais interna, é relacionada a grande transgressão marinha ocorrida há 400 mil anos

(Ka A.P.), que cobriu até a região do município de Porto Alegre, ficando emersos somente os morros graníticos que atualmente circundam a cidade. A formação da Laguna-Barreira II está relacionada com uma transgressão ocorrida há 325 Ka A.P. que invadiu uma área menor que a anterior. A formação da Laguna-Barreira III ocorreu há cerca de 120 Ka A.P. e invadiu uma área consideravelmente menor que a anterior e a qual se atribui a formação das barreiras e restingas que fecharam os grandes sistemas lagunares observados na atualidade. Finalmente, no Holoceno, há cerca de 7-8 Ka A.P., ocorreu a última transgressão que formou a Laguna-Barreira IV (Figura 1). A praia oceânica atual faz parte do sistema Laguna-Barreira IV (barreira holocênica) e é uma área que sofre alto impacto da ação marinha e eólica. A área mais próxima à costa apresenta um campo de dunas bem desenvolvido e extenso, com dunas ativas intercaladas por dunas semi-estabilizadas pela vegetação, que apresentam largura de entre 2 – 8 km (Hesp et al. 2005). Estas dunas recebem o nome de 'dunas transgressivas' porque migram com trajetória da costa para o continente.

Clima

De acordo com características climáticas, a PCAS pode ser subdividida em três setores: o Litoral Norte (de Torres até Cidreira), o Litoral Centro (de Palmares do Sul e Barra do Ribeiro até São José do Norte e Pelotas), e o Litoral Sul (de Rio Grande até o Chuí).

O clima da PCAS é classificado em temperado subtropical e a amplitude térmica é alta, indo desde temperaturas inferiores a 0 °C no inverno, até mais de 35 °C no verão. Embora de maneira geral a precipitação seja bem distribuída ao longo do ano, existem algumas diferenças ao longo da PCAS. As escarpas da Serra Geral, no Litoral Norte, geram um regime de umidade maior em relação aos outros setores que

está relacionado com a produção de precipitações orográficas com maior intensidade no verão (Carmargo 2002). Nos setores Centro e Sul da PCAS podem ocorrer períodos mais intensos de chuvas no inverno relacionados com a penetração frontal de ciclones migratórios extratropicais muito ativos nesta época do ano (Grimm et al. 1998).

O clima da região é determinado pelo Anticiclone Tropical Sul Atlântico (ATSA), o Anticiclone Polar Migratório (APM) e a Depressão Térmica Semipermanente do Chaco (depressão de Chaco). O ATSA é um centro de alta pressão composto de massas de ar quente e úmido. O APM é um centro de alta pressão alimentado por massas de ar frio, que tem comportamento migratório em direção a nordeste. Este deslocamento do APM gera a formação de dois centros de alta pressão que produzem um centro de baixa pressão no meio (frentes frias) e são acompanhados de ventos ciclônicos, instabilidades climáticas e precipitações. O ATSA ocorre predominantemente na primavera e verão e um gradiente de pressão deste com a depressão de Chaco geram ventos com sentido Leste (L) – NE ao longo de toda a PCAS. Durante o outono e inverno, o APM é mais ativo e então predominam os ventos ciclônicos em direção sul (S) – SO (Carmargo 2002; Martinho 2008).

Estas características de vento e a variação deste ao longo da PCAS representam um potencial de transporte diferencial de areia nos campos de dunas (Martinho 2008). No setor Norte da PCAS os ventos provenientes do Norte (N) – NE e S – SO são os mais importantes, produzindo um transporte de areia com direção noroeste (NO). Em Imbé e Tramandaí (setor Norte), os ventos provenientes do NE são os que apresentam maior velocidade e frequência, produzindo um transporte potencial com direção SO. Em Mostardas (setor Centro), encontra-se a localidade

com maior impacto do vento, que apresenta as maiores frequências de ventos com velocidades superiores a 17 m/s e a menor quantidade de tempos calmos; esta é também a região que apresenta uma ampla variação na direção do vento que leva a um transporte em geral com direção a oeste (O). Em Rio Grande (limite sul do setor Centro), encontra-se a localidade com os ventos mais fracos da PCAS, com direções opostas que terminam quase se anulando entre si. Nesta área, o transporte de areia é muito pequeno, com direção NO. O Chuí (setor Sul) está exposto com maior frequência a ventos provenientes do L – NE, mas os ventos provenientes do S, embora menos frequentes, têm maior velocidade e são mais determinantes, promovendo um transporte potencial com direção N – NO.

Caracterização dos modelos biológicos

Neste trabalho foram utilizadas como modelos biológicos duas espécies: *Petunia integrifolia* (Hook.) Schinz & Thell. e *Calibrachoa heterophylla* (Sendtn.) Wijsman, que ocorrem exclusiva ou predominantemente ao longo de toda a PCAS.

Os gêneros *Petunia* Juss. e *Calibrachoa* Cerv. ex LaLlave & Lex estão inclusos no clado Petuniae da família Solanaceae (Olmstead et al. 2008; Reck-Kortmann et al. 2015), apresentam distribuição subtropical Atlântica e são estritamente campestres, ocorrendo principalmente em campos com afloramentos rochosos e solos pedregosos no planalto sul-brasileiro e da região do Pampa (Stehmann et al. 2009); algumas espécies ocorrem em áreas antropizadas, sendo consideradas como pioneiras e colonizadoras de áreas abertas, como são as clareiras, bordas de floresta e mesmo beiras de estrada (Vendruscolo 2009).

***Calibrachoa* Cerv.**

As espécies de *Calibrachoa* têm hábito arbustivo e com caules lenhosos, característica esta que sugere que sejam perenes, com exceção de *C. parviflora* (Juss.) D'Arcy e *C. elegans* (Miers) Stehmann & Semir. A corola das flores deste gênero é predominantemente zigomorfa, infundibuliforme, com a base tubulosa amarela ou esbranquiçada, com prefloração conduplicada, onde as pétalas que formarão os lobos inferiores se dobram sobre as três restantes; as flores têm cinco estames arranjados em três alturas diferentes e pólen amarelo, carácter que as diferencia da maior parte das espécies do gênero *Petunia*. Não existe registro de qualquer estratégia de propagação vegetativa e o mecanismo de dispersão de sementes é por liberação direta no solo, perto da planta mãe. O hábitat preferencial das espécies de *Calibrachoa* são afloramentos rochosos, bordas de floresta e campos arenosos ou pedregosos (Stehmann 1999).

O gênero *Calibrachoa* inclui 27 espécies restritas principalmente a áreas de campos de altitude do sul do Brasil, mas também se encontram em áreas próximas aos limites com Argentina, Uruguai e Paraguai (Fregonezi et al. 2012).

***Calibrachoa heterophylla* (Sendtn.) Wijsman**

As plantas de *C. heterophylla* se caracterizam como subarbustos, fortemente lignificados na base, com muitos ramos, inicialmente eretos e logo decumbentes; apresentam braquiblastos axilares e folhas pequenas, sésseis, de lâmina linear, densamente pubérulo-glandulosas em ambas as faces. As plantas apresentam uma coloração característica esbranquiçada ou acinzentada nos ramos jovens, folhas e cálice, conferida pelos grãos de areia aderidos ao indumento. As flores são de cor

magenta ou purpúrea com fauce amarelada circundada por um anel purpúreo-escuro, características relacionadas com a melitofilia (Figura 2; (Stehmann 1999).

C. heterophylla habita dunas e campos arenosos, preferencialmente ao longo da PCAS nos estados de Santa Catarina (SC) e Rio Grande do Sul (RS), em áreas de influência marinha, lagunar ou lacustre; característica que a diferencia da maior parte das espécies do gênero e que se relaciona com sua alta afinidade por ambientes com influência do hálito marinho e alto conteúdo de sal da areia das praias (psamofilia). No entanto, três populações encontradas em praias da bacia do rio Santa Maria (região da Campanha do RS) foram determinadas como *C. heterophylla*.

O único estudo realizado para *C. heterophylla* mostrou que a espécie apresenta quatro linhagens intraespecíficas, uma delas de distribuição continental. As três linhagens restantes são encontradas apenas na PCAS e configuram uma estrutura geográfica bem estabelecida norte, centro e sul. A divergência das três linhagens costeiras foi relacionada com um processo anterior à formação da PCAS com os rios Jacuí e Jaguarão, que antes da formação da PCAS desembocavam diretamente no Oceano Atlântico, atuando como barreiras ao fluxo entre as linhagens reforçando o processo de diferenciação entre elas (Figura 3A; (Mäder et al. 2013).

Figura 2. Características morfológicas e do hábitat de *Calibrachoa heterophylla* numa localidade à beira da Lagoa do Patos próxima ao município de Pelotas, RS. Fotografias tomadas por G. Silva-Arias.

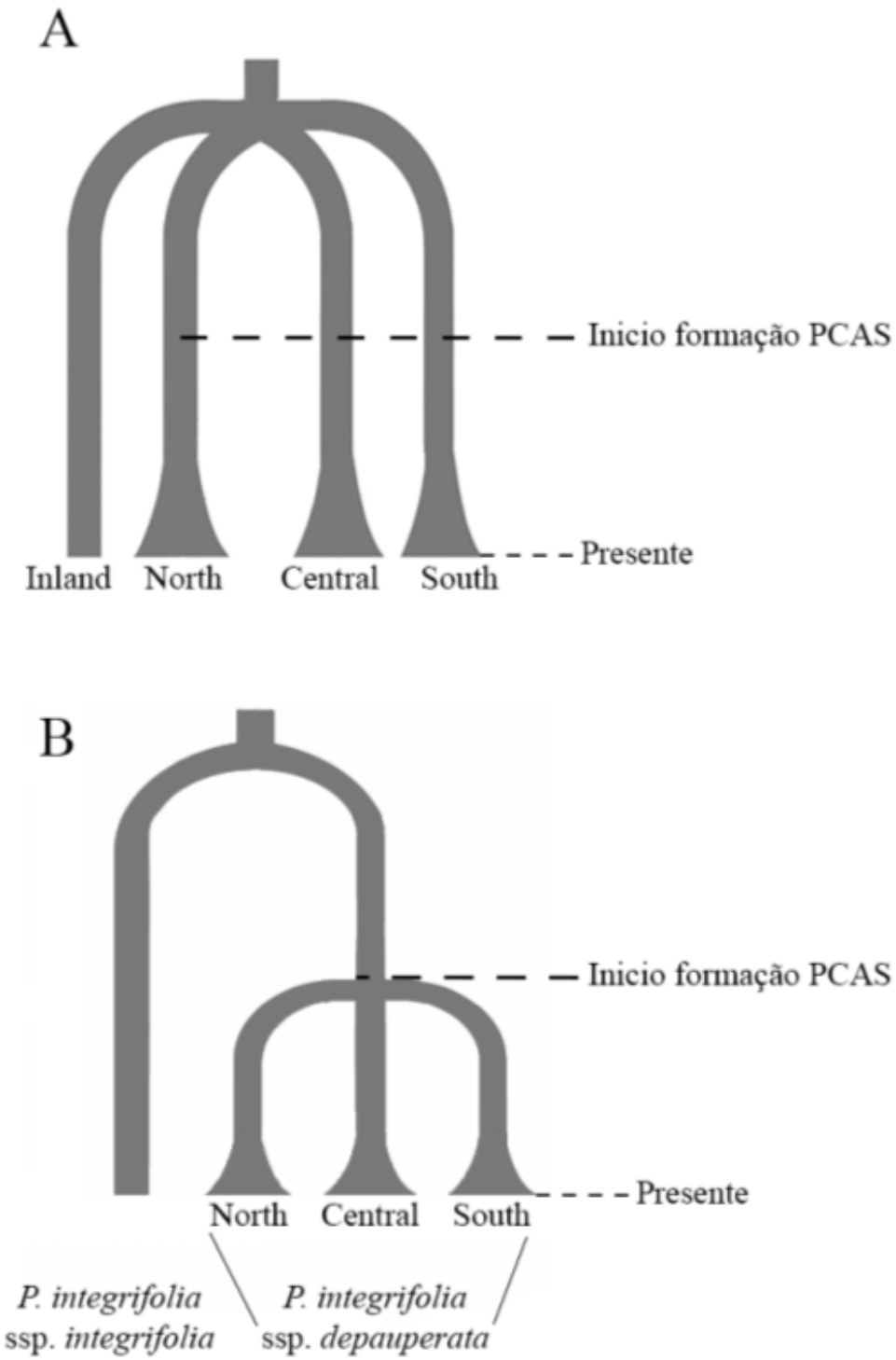


Figura 3. Modelos histórico-demográficos de *Calibrachoa heterophylla* (A) e *Petunia integrifolia* ssp. *depauperata* (B) baseados em dados de sequencias de cloroplasto publicados em (Mäder et al. 2013) e (Ramos-Fregonezi et al. 2015).

***Petunia* Juss.**

O gênero *Petunia* abrange plantas herbáceas e de crescimento monocárpico, características associadas ao caráter anual e ao ciclo de vida curto, ajustado à sazonalidade climática da região de ocorrência do gênero. A maioria das espécies apresenta tricomas simples, multicelulares, unisseriados e com ápices glandulosos, que contêm sustâncias viscosas, alcaloides ou açúcares esterificados, possivelmente relacionados à proteção contra insetos. As flores apresentam cálice fendido até próximo à base; a corola tem prefloração imbricada, é actinomorfa ou zigomorfa e de coloração purpúrea na maioria das espécies, embora algumas espécies apresentem corolas de cor vermelha ou branca. A maioria das espécies possui pólen em tons violáceos, mas tons amarelos são encontrados em *P. axillaris* (Lam.) Britton, Sterns & Poggenb., *P. exserta* Stehmann e *P. secreta* Stehmann & Semir. A síndrome floral da melitofilia é predominante no gênero, mas síndromes como ornitofilia e esfingofilia são encontradas em algumas espécies (Stehmann et al. 2009).

O gênero *Petunia* compreende 14 espécies distribuídas em regiões subtropicais e temperadas de América do Sul. Reconstruções biogeográficas propuseram que a origem do gênero *Petunia* ocorreu na região dos Pampas com posterior migração para regiões de campos de altitude (Reck-Kortmann et al. 2014).

***Petunia integrifolia* (Hook.) Schinz & Thell.**

A espécie *P. integrifolia* é dividida em duas subespécies de acordo principalmente com o ambiente onde se distribuem: *P. integrifolia* ssp. *integrifolia* (daqui para a frente chamada apenas de *P. integrifolia*), que ocorre na região dos Pampas (RS, Uruguai e Argentina), e *P. integrifolia* ssp. *depauperata* (R.E.Fr.) Stehmann (resumidamente, *P. depauperata*), que se distribui ao longo da PCAS,

atingindo o estado de Santa Catarina (Florianópolis) em seu limite norte e o litoral do Uruguai (departamento de Treinta y Tres) ao sul (Stehmann and Bohs 2007).

As plantas de *P. depauperata* apresentam hábito procumbente, com caules longos que permanecem sobre o solo sem enraizar; a forma das folhas e a densidade do indumento são variáveis; as flores são purpúreas e menores em comparação às flores das demais espécies do complexo *integrifolia* (Ando et al. 2005a); grupo de espécies morfologicamente semelhantes à *P. integrifolia*, que ao menos uma vez foram consideradas como sinônimos ou subespécies desta). É uma espécie psamófila, que cresce nos campos arenosos dos sistemas lagunares ou nas dunas e afloramentos rochosos do litoral da PCAS (Figura 4; Stehmann 1999).

P. integrifolia apresenta a segunda maior distribuição do gênero, assim como alto grau de variabilidade morfológica e ecológica. Esta diversidade tem conduzido à instabilidade taxonômica desta espécie ao longo dos anos (Ando et al. 2005b; Stehmann and Bohs 2007). *P. depauperata* é suportada por dados moleculares como uma linhagem independente de *P. integrifolia*, restrita à PCAS e que engloba também as populações anteriormente descritas como *P. littoralis* L.B. Sm. & Downs (Longo et al. 2014).

A origem de *P. depauperata* tem sido interpretada como um processo de divergência de populações periféricas a partir de uma linhagem continental que, em estágios iniciais do surgimento da PCAS há ~ 400 Ka A.P., foi isolada em ilhas temporais durante as transgressões marinhas na área conhecida como Coxilha das Lombas (Laguna-Barreira I; Figura 1) e nos morros graníticos localizados na região do município de Porto Alegre, RS (Figura 3B; (Ramos-Fregonezi et al. 2015).

Figura 4. Características morfológicas e do hábitat de *Petunia integrifolia* ssp. *depauperata* numa localidade na beira da Lagoa do Patos próximo ao município de Pelotas, RS. Fotografias tomadas por G. Silva-Arias

Considerando o conhecimento previamente desenvolvido da história genealógica de *C. heterophylla* e *P. depauperata* (Figura 3) e o forte padrão de co-ocorrência destas espécies na PCAS, surgiram novas questões de como relacionar a geomorfologia e as condições abióticas atuais da PCAS com o padrão de estruturação genética contemporânea observado nestas espécies, que consequentemente pode resultar em eventuais sinais de estruturação e adaptação local convergente. Para dar resposta a estas novas perguntas foi desenvolvida uma avaliação da diversidade e estrutura genética destas espécies por meio de uma abordagem de genética de paisagens que permitiu expandir o entendimento da história natural destas espécies, avaliar padrões gerais da colonização da PCAS e propor algumas diretrizes de conservação para a região, que se configuram na presente tese.

OBJETIVOS

Entender a influência da história demográfica, características geomorfológicas e interações ecológicas sobre os padrões de estruturação genética e fluxo gênico em espécies sob processo de colonização de ambientes costeiros. Com isto, esperava-se identificar processos de diferenciação populacional, dinâmicas preferenciais de fluxo gênico e adaptação local relacionados com diferenças nos locais de ocorrência das populações.

Objetivos específicos

- Desenvolver ou transferir marcadores genéticos microssatélites para as espécies de estudo e caracterizar sua variabilidade em populações naturais.
- Obter camadas climáticas específicas para a região de estudo adequadas para realizar correlações com dados genéticos e modelos de nicho.
- Analisar padrões de estrutura populacional e fluxo gênico nas espécies *Petunia depauperata* e *Calibrachoa heterophylla* e entender sua relação com características físicas e ecológicas da sua área de distribuição.

ESTRUTURA E ORGANIZAÇÃO DA TESE

Com o objetivo de apresentar o conjunto de resultados obtidos no desenvolvimento do projeto, esta tese está dividida em três capítulos os quais são apresentados em forma de artigos nos respectivos formatos dos periódicos aos quais foram ou serão submetidos. Finalmente as considerações finais apresentam uma discussão geral dos resultados que levaram ao alcance dos objetivos propostos.

Capítulo 1

"From inland to the coast: spatial and environmental signatures on the genetic diversity in the colonization of the South Atlantic Coastal Plain"

Autores: Gustavo A. Silva-Arias, Maikel Reck-Kortmann, Bryan C. Carstens, Heinrich Hasenack, Sandro L. Bonatto e Loreta B. Freitas.

Este manuscrito corresponde a um artigo em processo de revisão na revista *Molecular Ecology*. Neste trabalho, através da caracterização da diversidade e estrutura genética de *Petunia depauperata* em relação às variáveis climáticas e espaciais, reconstruímos os processos de colonização da PCAS por ancestrais de origem continental. Neste trabalho, encontramos que o fluxo gênico das populações do centro da distribuição em direção às populações marginais poderia ter facilitado o processo de expansão e adaptação local nos extremos da distribuição de *P. depauperata* por via do enriquecimento da diversidade genética e contribuição de alelos adaptativos.

Capítulo 2

"Novel microsatellites for *Calibrachoa heterophylla* (Solanaceae) endemic to the South Atlantic Coastal Plain of South America".

Autores: Gustavo Adolfo Silva-Arias, Geraldo Mäder, Sandro L. Bonatto e Loreta B. Freitas.

Este manuscrito corresponde a uma nota com a descrição de *primers* publicada na revista *Applications in Plant Sciences* (*American Journal of Botany*). O trabalho reporta o desenvolvimento de 16 marcadores de microssatélite específicos para *Calibrachoa heterophylla*, descreve sua variabilidade para 57 indivíduos de duas populações desta espécie e resultados da transferabilidade para 12 espécies de *Calibrachoa* representantes de todas as linhagens evolutivas até o momento identificadas para o gênero.

Capítulo 3

"New insights on climatic driven genetic differentiation in peripheral populations during a coastal colonization process"

Autores: Gustavo A. Silva-Arias, Giovanna C. Giudicelli e Loreta B. Freitas.

Este capítulo corresponde a um artigo em preparação a ser submetido à revista *Journal of Evolutionary Biology*. Este trabalho apresenta os padrões observados de diversidade e estrutura genética de *Calibrachoa heterophylla,* que podem ser explicados por um complexo grupo de processos que incluem a história demográfica da espécie, isolamento por distância, redução da diversidade durante a onda de

expansão em direção à região costeira, isolamento e diferenciação local de populações marginais possivelmente reforçado por características do ambiente.

Considerações Finais

A última seção desta tese discute de modo global os resultados mais importantes obtidos no desenvolvimento do projeto em relação às principais perguntas evolutivas envolvidas, sintetizando os resultados dos capítulos e mencionando as principais conclusões e perspectivas para o desdobramento desta tese.

CAPÍTULO 1

Artigo submetido à revista *Molecular Ecology*

From inland to the coast: spatial and environmental signatures on the genetic diversity in the colonization of the South Atlantic Coastal Plain

MOLECULAR ECOLOGY

From inland to the coast: spatial and environmental signatures on the genetic diversity in the colonization of the South Atlantic Coastal Plain

Journal:	*Molecular Ecology*
Manuscript ID	MEC-16-0332
Manuscript Type:	Original Article
Date Submitted by the Author:	22-Mar-2016
Complete List of Authors:	Silva-Arias, Gustavo; Universidade Federal do Rio Grande do Sul Instituto de Biociencias, Genetics Reck-Kortmann, Maikel; Universidade Federal do Rio Grande do Sul Instituto de Biociencias, Genetics Carstens, Bryan; State Library of Ohio Hasenack, Heinrich; Universidade Federal do Rio Grande do Sul Instituto de Biociencias, Ecology Bonatto, Sandro; Pontificia Universidade Catolica do Rio Grande do Sul, Biology Freitas, Loreta; Universidade Federal do Rio Grande do Sul, Genetica
Keywords:	Adaptation, Angiosperms, Climate Change, Landscape Genetics, Niche Modelling, Phylogeography

From inland to the coast: spatial and environmental signatures on the genetic diversity in the colonization of the South Atlantic Coastal Plain

Authors: GUSTAVO A. SILVA-ARIAS[£], MAIKEL RECK-KORTMANN[£], BRYAN C. CARSTENS[§], HEINRICH HASENACK[♯], SANDRO L. BONATTO[¶], LORETA B. FREITAS[£,*]

[£]Laboratory of Molecular Evolution, Department of Genetics, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil

[§]Department of Evolution, Ecology and Organismal Biology, Ohio State University, Columbus, OH, USA

[♯]Laboratory of Geoprocessing, Department of Ecology, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil

[¶]Laboratory of Genomic and Molecular Biology, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, RS, Brazil

Keywords: Central-marginal, isolation by distance, isolation by environment, *Petunia*, serial founder effect, source-sink dynamics

[*]Corresponding author: Loreta B. Freitas, Department of Genetics, UFRGS, P.O. Box 15053, Porto Alegre, RS 91501-970, Brazil. Fax: +55 51 3308 9823. E-mail: loreta.freitas@ufrgs.br

Running title: Genetics of coastal colonization

Abstract

The processes of colonization and range expansion to novel environments can be traceable in the genetic diversity of the organisms, but sometimes can be difficult to disentangle between them. In this work we investigate the association of spatial and climatic variables with the genetic diversity and gene flow patterns of wild *Petunia* populations involved in a process of colonization of coastal environment in a landscape genetics framework. Over 300 individuals from 17 populations were genotyped using ten microsatellite loci. Results from a variety of analyses suggest that genetic diversity is higher in populations located at the center of the species range, and that gene flow follows a source-sink dynamics between the inland and coastal founder populations to the coastal peripheral populations. We identify a sharp genetic differentiation between inland and coastal populations consistent with a process of adaptive divergence during the coast colonization. Our results show a consistent association signals between precipitation seasonality and genetic population differentiation that could be related with local adaptation processes in edge populations. We deduce that this local adaptive divergence can be facilitated by genetic enrichment by gene flow from the coastal founder populations and by rapid allele fixation processes owing to serial founder effects during the range expansion across the coast.

Introduction

Investigations into diverging lineages that occupy regions with different ecological conditions enable researchers to understand how colonization and adaptation influences spatial patterns of genetic diversity (Nosil *et al.* 2009; Ferchaud & Hansen 2016; Laurent *et al.* 2016). The recognition of isolation by distance (gradual genetic differentiation across populations as product of limited dispersal), isolation by colonization (sharp differentiation patterns result of founder effects), and isolation by environment (differentiation as product of prevented migration between populations occupying different environments) is particularly challenging due to the superimposition of their respective signal in the genetic data. For example, when an organism colonizes a region with new ecological conditions, genetic divergence can be enhanced by the reduction of gene flow related to the spatial separation as well as by selection against maladapted migrants. Local adaptation to new environmental conditions may give rise to the emergence of genetically distinct lineages by the interruption of gene flow and drift (e.g., Misiewicz & Fine 2014); however, ecological differences could do not match with the genetic differentiation in cases when selection is stronger to maintain habitat specific phenotypes in the presence of high inter-population gene flow (e.g., Hoekstra *et al.* 2005).

The complex relationship between historical and ecological processes in determining the spatial distribution of species can be assessed using a landscape genetics framework (Manel *et al.* 2003). Landscape genetics is a field that has developed rapidly due to the sophistication of molecular methods, statistical analyses, and the availability of spatial ecological data (Storfer *et al.* 2007, 2010; Manel & Holderegger 2013). Several factors, including the genetic markers, geographical range of the focal species, as well as the methods used to detect gene flow with regards to

environmental gradients, may influence the inference about the relation between space and environment on the observed patterns of genetic differentiation.

The South Atlantic Coastal Plain (SACP) is a well-characterized geomorphologic coastal formation, and constitutes the largest coastal plain in South America. It is a flat, continuous, and open coast occupied mostly by large coastal lakes. The SACP extends NE-SW for approximately 600 km and is crossed by only two perennial water channels (Tomazelli *et al.* 2000; Weschenfelder *et al.* 2010). The SACP was formed during oscillatory sea level transgressions and regressions caused by the glacial-interglacial cycles during the Pleistocene. Each transgression and regression cycle caused the formation of one sand barrier, positioned parallel to the coast (barrier-lagoon systems I to IV; (Tomazelli *et al.* 2000; Tomazelli & Dillenburg 2007). After Quaternary sea-level changes and the establishment of the current SACP, vegetation from the west could expand into this new environment and adapt to different conditions of salinity, soil composition, climate, and hydric regime (Mäder *et al.* 2013; Ramos-Fregonezi *et al.* 2015).

Coastal environments offer particular climatic and edaphic conditions, such as salt exposition, temporal flooding, and seasonal or permanent strong winds. Additionally, coast habitats present an inherent linear shape, conditioning the distribution of the taxa restricted to these regions. Global climatic changes can strongly affect the availability of suitable habitats for coastal organisms over time (Hoegh-Guldberg & Bruno 2010; Reyer *et al.* 2013). Several investigations have addressed the phylogeographical patterns of coastal organisms and their distributional and demographic responses to climate changes (Weising & Freitag 2007); they have also discussed physiological and genetic approaches to examining specific mechanisms of adaptation to saline environments (Zhu 2001; Lowry *et al.* 2009).

However, assessing the colonization process of saline environments involving sister inland-coastal lineages through the characterization of genetic diversity and gene flow in relation to spatial and ecological variables could be important to understanding the adaptation to saline environments from an evolutionary point of view.

The aim of this study was to identify genetic diversity and gene flow responses of the colonization process of a unique lineage of the *Petunia* genus (Solanaceae) restricted to the SACP through a landscape genetic assessment. We focused into the *P. integrifolia* subsp. *depauperata* (hereafter *P. depauperata*) in order to understand the process of colonizing coastal environments from a continental antecessor, and to disentangle spatial and ecological patterns in their genetic diversification within the coast. Additionally, we assessed the influence of two different sets of climatic variables on the recovering isolation by environment processes. We specifically addressed the following questions: (1) Which were the main determinants of the genetic diversity and structure in a coastal colonization process? (2) Did spatial or ecological factors constrain the inter-population gene flow?

Materials and Methods

Study system

Petunia depauperata is a diploid (2n=14), prostrate annual herb, with purplish bee-pollinated flowers. This taxon is restricted to open sandy grasslands, dunes or rocky outcrops of lakeside or marine environments from the SACP; further, *P. depauperata* is the only lineage of *Petunia* found in coastal environments. It likely originated around 400 thousand years ago (kya) from populations of its closest relative, *P. integrifolia* subsp. *integrifolia* (hereafter *P. integrifolia*), a widespread taxon in inland grasslands (Longo *et al.,* 2014; Ramos-Fregonezi *et al.,* 2015). These studies suggest

that *P. depauperata* divergence could have begun from eastern peripheral populations of *P. integrifolia* temporally isolated in emerged regions [granitic hills around the Porto Alegre municipality (Rio Grande do Sul state, Brazil; ~ 30 Lat S) and fossil dune fields; barrier-lagoon system I] during a marine transgression, which occurred around 400 kya.

The current known latitudinal distribution of *P. depauperata* follows the coastal line from the northern extreme around the city of Garopaba (Santa Catarina state, Brazil; ~ 28 Lat S) reaching the southern extreme in La Coronilla (Rocha department, Uruguay; ~34 Lat S). Populations of *P. depauperata* are found from the sea line to less than 90 km from the coast, with two populations separated from the sea by big lagoons (Fig. 1).

Sampling

We sampled 307 individuals from 17 localities (hereafter referred as populations): 236 individuals (12 populations) were identified as members of the coastal lineage (*P. depauperata*), with the remaining 71 individuals (five populations) from the inland lineage (*P. integrifolia*) based on morphological and environmental traits (Stehmann & Bohs 2007) and genetic characterization (Longo *et al.* 2014; Ramos-Fregonezi *et al.* 2015). Four of the inland lineage populations (Guaíba, Viamão, Tapes, and Arambaré; Fig. 1) were located in fossil dune fields that presented a coastal environment 400 kya. Leaves of all individuals were collected during the flowering season between 2002 and 2013 (September to February) and preserved in silica gel. The number of individuals per population varied from 14 to 33 for *P. depauperata* and five to 15 for *P. integrifolia* (Table 1). One herbarium voucher per population was taken and deposited in the ICN Herbarium, Universidade Federal do Rio Grande

do Sul, Porto Alegre, Brazil, or the BHCB Herbarium, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil.

DNA extraction and genotyping

The total DNA was extracted following the protocol of Roy *et al.* (1992). Individuals were genotyped for ten microsatellite loci (PM101, PM21, PM8, PM167, PM192, PM110, PM184, PM157, PM117, PM177); they were described and mapped in two *Petunia* hybrids (Bossolini *et al.* 2011) and successfully transferred to wild *Petunia* species (Turchetto *et al.* 2015). Selected loci were located in six of the seven *Petunia* chromosomes (Bossolini *et al.* 2011), and although some were located on the same chromosome, all were more than 40 cM apart from each other. An additional nine loci were tested and discarded due to their monomorphic status in a preliminary screening in *P. depauperata*. For details of PCR and genotyping, see Appendix S1 (Supporting information).

Genetic diversity and structure

Basic diversity statistics, such as allele numbers (A), allele richness (Ar), private alleles (pA), observed heterozygosity (H_O), and gene diversity (H_S) were calculated using the packages ADEGENET 2 (Jombart 2008; Jombart & Ahmed 2011), POPPR 2.0.2 (Kamvar *et al.* 2014, 2015), and HIERFSTAT 0.04-14 (Goudet 2005, 2014) in R 3.2.1 (R Core Team 2016) and ARLEQUIN 3.5 (Excoffier & Lischer 2010). Details can be found in Appendix S1 (Supporting information).

Genetic clustering analyses were performed with STRUCTURE 2.3.4 (Pritchard *et al.* 2000). The number of groups (K) evaluated ranged from 1 to the total number of populations (17), with ten independent runs per K-value. Each run was performed

using 2.5 x 10^5 burn-in periods and 1.0 x 10^6 Markov chain Monte Carlo repetitions after burn-in, under an admixture model, assuming correlated allele frequencies (Falush *et al.* 2003), and including a priori sampling locations as prior (LOCPRIOR) to detect weak population structure. The LOCPRIOR option is not biased toward detecting structure when it is not present and can improve the STRUCTURE results when implemented with few loci (Hubisz *et al.* 2009). All STRUCTURE runs were performed using computational resources from the Ohio Biodiversity Conservation Partnership cluster using the script STRAUTO (Chhatre 2012) to generate a batch script that automatizes multiple STRUCTURE runs in a UNIX environment. The optimal K for each analysis was chosen using the ΔK method (Evanno *et al.* 2005) implemented in STRUCTURE HARVESTER (Earl & vonHoldt 2012). CLUMPP 1.1.2 (Jakobsson & Rosenberg 2007) was used to summarize the results of the ten runs of each chosen K using Greedy's method and the G-statistic. DISTRUCT 1.1 (Rosenberg 2003) was used to generate the respective bar plots.

In addition, we implemented the free-model multivariate method, Discriminant Analysis of Principal Components (DAPC; Jombart *et al.* 2010) in the R package ADEGENET to estimate the proportion of an individual genome that originated from a given genetic group or cluster using coefficients of the alleles (loadings) in orthogonal axes, maximizing between-groups variance and minimizing within-group variance in these loadings. For this analysis, the SSR data were transformed using Principal Component Analysis and keeping all principal components (PCs). The function *find.clusters* was implemented to obtain the optimal number of clusters that maximizes the between-group variability using the lowest Bayesian information criterion (BIC) score. In addition, to avoid over-fitting we chose the optimal number of PCs in the DAPC using the function *optim.a.score*. Finally, the DAPC of the

transformed genetic data was implemented with the number of clusters and PCs were set to the optimal values.

Spatial patterns of population differentiation

Matrices of pairwise F_{ST}, $D_{PS,}$ and Nei's statistics of population genetic differentiation were calculated with R packages ADEGENET, HIERFSTAT, and MSA 4.05 (Dieringer & Schlötterer 2003), respectively. The geographic distance matrix between populations was obtained by calculating the natural logarithm of the inter-population linear Euclidean distance (EUC). To account for possible differences in the inter-population connectivity caused by the inland water bodies common in the SACP, a topographic distance matrix (TOPO) reflecting the shortest land routes connecting population pairs was generated with CIRCUITSCAPE 4.0.5 (McRae 2006) using a raster surface with all land cell values equal to one and water cell values equal to zero. Relationships between genetic (F_{ST}, $D_{PS,}$ and Nei's) and geographic (EUC and TOPO) dissimilarity matrices were analyzed using Mantel tests, and the significance was assessed by 10 000 randomizations in the R package VEGAN 2.3-0 (Oksanen *et al.* 2015). Mantel correlograms were also obtained for four geographic distances classes (0-50 km, 50-150 km, 150-300 km, and 300-650 km). Additionally, the hypothesis of isolation by distance was assessed by linear regression of linearized pairwise F_{ST}, $D_{PS,}$ and Nei's genetic distances and the logarithmic transformed EUC distance (Rousset 1997). For all these analyses, inter-population geographic distances were calculated from X and Y UTM 22S coordinates (coordinate reference EPSG: 32722) transformed from Long/Lat coordinates with RGDAL 1.0-4 R package (Bivand *et al.* 2015). Arambaré and Itaara populations were excluded from these analyses due to the small sample size and their location outside the SACP, respectively.

The scenario for the diversification of *P. depauperata* described previously predicts that genetic diversity would decrease as distance from the area of origin (the granitic hills in Porto Alegre and 'Coxilha das Lombas' fossil dune, hereafter named as the "core distribution" region) increases. Linear regressions between measures of genetic diversity (A, Ar, pA, H_O, and H_S) and geographic distance to the core distribution were implemented in R software. Tests of linear model assumptions were implemented for all data sets using GVLMA 1.0.0.2 R package (Pena & Slate 2014). Additionally, the central-marginal patterns were tested using the suitability values of population obtained from the ensemble niche modeling (Araujo & New 2007) as implemented in the R package BIOMOD2 (Thuiller *et al.* 2014). See the *Environment and population differentiation* section below for details about climatic predictors, and Appendix S1 (Supporting information) for niche modeling methodological details. Suitability values were extracted from the raster surfaces of the ensemble niche model projections over the studied area with the R package RASTER 2.4-15 (Hijmans 2015).

Spatial patterns of observed genetic structure were estimated with the spatial Principal Component Analysis (sPCA; Jombart *et al.* 2008) that incorporates spatial information while maximizing the product of spatial autocorrelation (Moran's I) and variance of each eigenvector; this produces orthogonal axes that describe spatial patterns of genetic variation. The Delaunay triangulation was used to set the network of spatial connections between populations based on UTM coordinates. A Monte Carlo simulation (global and local tests) was used with 10 000 permutations to test for non-random spatial association of population allele frequencies. As in the spatial Mantel test, Arambaré and Itaara populations were excluded from this analysis. The analysis was implemented in the ADEGENET package in R.

Test for alternative migration models

The coalescent-based program MIGRATE-N 3.2.6 (Beerli & Felsenstein 2001) was used to test the support of the genetic data to alternative gene flow scenarios for the species. For this, the populations were pooled into five groups according to the geographical distribution and genetic structure results named: 'Basal' (including the Itaara, Tapes, Arambaré, Guaíba, Viamão, and Porto Alegre populations), 'North' (formed by Garopaba and Torres populations), 'Center' (grouping Osório, Curumim, Xangri-lá, and Pinhal populations), 'South 1' (composed by Mostardas, São Lourenço do Sul, and Pelotas populations), and 'South 2' (encompassing Rio Grande and Taim populations).

Six migration models were evaluated: (1) a central-marginal model with unidirectional migration pattern from central to peripheral populations; (2) a marginal-central model with unidirectional migration pattern from peripheral to central populations; (3) a source-sink-from-basal model with unidirectional migration from basal populations to coastal populations; (4) a source-sink-from-central model with unidirectional migration from basal populations to central coastal populations, and from those to the remaining coastal populations; (5) a ring model with bidirectional asymmetric migration between adjacent populations rounding Lagoa dos Patos Lagoon and unidirectional asymmetric migration to edge populations; and (6) a north-south-barrier model with bidirectional asymmetric migration between adjacent populations (For models graphical representations see Fig. S1, Supporting information).

For each model, we ran MIGRATE-N in the CIPRES Science Gateway 3.3 (Miller *et al.* 2010) with one long chain of 2 000 000 steps, sampling at every 100th increment and a burn-in of 10 000 steps. For each model, θ_W and M were drawn from

uniform prior distributions with ranges from 0 to 5 and 0 to 100, respectively. A heating scheme was used (Metropolis-coupled Markov Chain Monte Carlo) with four parallel chains with 'temperatures' of 1.0, 1.5, 3.0, and 1 000 000. We compared the models using Bayes factors calculated from the Bézier log marginal likelihood approximation because it produces better estimates of marginal likelihood, and is less influenced by prior parameter distributions (Beerli & Palczewski 2010).

Environment and population differentiation

The widely-used WorldClim project has available climatic surfaces for the SACP, but the primary source of data used by Hijmans *et al.* (2005) have low density of weather stations in the target area, an issue that could be problematic for the spatial scale of this study. Therefore, we decided to build a new dataset of climatic surfaces specific for the SACP based on data from higher sampling of weather stations from specific Brazilian sources (Fig. S2, Supporting information) including variables not presented in the WorldClim project.

The procedure for model fitting was thin-plate splines (TPS) implemented in the R package FIELDS 8.2-1 (Nychka *et al.* 2015). The obtained models were then used to interpolate the variables to the studied area using the R package RASTER 2.4-15. We obtained eight climatic surfaces: total annual precipitation, total annual days with rain, precipitation seasonality (coefficient of variation), mean annual temperature, mean summer maximum temperature, mean winter minimum temperature, mean temperature range [mean of monthly (max temp - min temp)], and temperature seasonality (coefficient of variation). For methodological details see Appendix S1 (Supporting information).

In order to compare the performance of this new set of climatic surfaces, we replicated all environmental analyses using the 19 bioclimate raster layers set at 30 arc-second resolution obtained from the WorldClim database (Hijmans *et al.* 2005) that were clipped at same geographic extent.

The VEGAN R package was used to test for relationship between environmental and genetic (F_{ST}, D_{PS}, and Nei's) dissimilarity matrices via a Mantel test (Mantel 1967) and the significance was assessed by 10 000 randomizations. For those climatic variables that produce significant correlations, Partial Mantel tests (Smouse *et al.* 1986) were conducted to verify the correlation between environment and genetic distances controlling by EUC and TOPO spatial distances. We checked the linearity and homoscedasticity assumptions between correlated variables by checking the significance of fitted linear models variables, using the GVLMA 1.0.0.2 R package.

The use of Mantel test and its variations to test potential spatial or environmental drivers on population structure has been criticized on their statistical performance and violations of linearity and homoscedasticity assumptions (Legendre *et al.* 2015). Therefore, we also examined the relationships between genetic distance and environmental variables with a linear mixed modeling approach, based on maximum likelihood population effects modeling (MLPE; Clarke *et al.* 2002) that explicitly accounts for non-independence of values in regressions on distance matrices. The climatic variables obtained in this study and the spatial distances (EUC and TOPO) were used as explanatory variables, and genetic distances (F_{ST}, D_{PS}, Nei's) were used as dependent variables. All explanatory variables (predictor matrices) were centered on their mean values prior to analyses. MLPE models were fitted by maximizing the restricted log-likelihood method (REML) using the 'gls' function in the R package NLME 3.1-121 (Pinheiro *et al.* 2015) and utilizing the corMLPE R script

(http://github.com/nspope/corMLPE) to describe the within-group correlation structure. The model selection was conducted via corrected Akaike Information Criterion (AICc) and BIC using the R package MUMIN 1.15.1 (Barton 2013). Initially univariate models were fitted to analyze the relation between each explanatory (climatic and spatial) variable and genetic distance. After that, a subset of climatic variables that fitted the best univariate models via AICc and BIC selection were chosen to fit multivariate models including and excluding the spatial variables (EUC or TOPO). All climatic variables included in the multivariate models presented Pearson's correlation coefficients $r < 0.5$.

We also explored the relationship between climatic variables and genetic differentiation with a Principal Coordinate Analysis (PCoA) of the F_{ST} distance matrix. PCoA scatterplot was plotted with the raw values of climatic variables that showed significant relationship with genetic differentiation based on the Mantel and MLPE analyses. The PCoA was calculated with the R package LABDSV 1.7-0 (Roberts 2015) and 3D plots were obtained with the R package SCATTERPLOT3D 0.3-36 (Ligges & Mächler 2003).

Results

Genetic diversity

Basic diversity statistics are listed in Table 1. For all loci, the observed heterozygosity was lower than expected, showing a deficit of heterozygotes as compared to an idealized population in HWE. Furthermore, 13% (23 of 170) of the locus-population combinations showed departure of HWE ($P < 0.01$) (Table S2, Supporting information). A significant linkage disequilibrium signal ($P < 0.01$) was detected for some pairs of loci, but since these were not related to their putative chromosome

positions (Table S3, Supporting information) and the linkage pattern was not constant for all populations, we maintained all loci for the analyses. Positive and significant inbreeding coefficients (F_{IS}) were found for seven populations (Table 1). MICROCHECKER analysis showed that locus PM167 presents homozygote excess that may be due to null alleles or stutter peaks, however, analyses including or excluding PM167 presented same results.

Population genetic structure

The STRUCTURE HARVESTER analysis identified the best number of clusters as K=2 (Fig. S3a, Supporting information), separating the inland (*P. integrifolia*) from the coastal lineages (*P. depauperata*), with exception of the Porto Alegre population, which was grouped with the inland populations. Interestingly, although the Porto Alegre population phenotypically was identified as a coastal lineage, it was located in the fossil dunes region, together with the other inland populations (Tapes, Arambaré, Guaíba, Viamão) (Fig. S3b, Supporting information). Additional groupings between K=3 and K=7 were explored looking for genetic sub-structured patterns within coastal populations. These groupings showed a strong and consistent differentiation of a group made up of São Lourenço do Sul and Pelotas populations (both from southern region and located at the west side of the Patos Lagoon; Fig. 1). Interestingly groupings of K=4 to K=6 grouped the northernmost and southernmost edge populations; but, when K=7, the cluster of edge populations was divided in one group comprising the two southernmost coastal populations (Rio Grande and Taim), and another group encompassing the northernmost coastal population (Garopaba). Mixed membership probability was recovered for several of the coastal populations from the central region of the SACP (Osório, Xangri-lá, Curumim, and Pinhal); also the

Mostardas population, found in the southern region of the SACP (but being the closest to the central region; Fig. 1), was the most admixed population. The Torres population, located in the northern region of the SACP (but the closest to the central region; Fig 1), showed more affinity with populations from the central region (Fig. S3b, Supporting information).

The genetic structure revealed by the DAPC showed congruent results to the STRUCTURE analysis. BIC scores showed a sharp decline until K=7 (Fig. S4a, Supporting information). Genetic structure was assessed for K=2 to K=7. The K=2 bar plot showed a clear differentiation between the group encompassing the inland and fossil dunes populations and the other group formed by the remaining populations of *P. depauperata*. Also, DAPC recovered the São Lourenço do Sul and Pelotas populations. Additionally, populations from the central region of SACP presented a higher admixture of membership probability; furthermore, the Osório and Curumim populations presented several individuals with membership of the inland-fossil dunes population group (Fig. S4b, Supporting information).

Spatial genetic structure

Mantel tests that were implemented to assess the correlation between geographical and genetic inter-population distances showed marginally significant values between EUC distances and F_{ST} (Mantel's statistic r = 0.39; P = 0.025), Nei's distance (r = 0.22; P = 0.079), and D_{PS} (r = 0.24; P = 0.086). Mantel correlograms for these three genetic distances were significantly positive with EUC distance matrix only in the first distance class. Mantel correlations between TOPO and the genetic distances were not significant, but the correlograms were significantly positive in the first distance class.

All linear regressions between the genetic diversity statistics A, pA, Ar, H_O, and H_S and population distances to the core distribution region of *P. depauperata* exhibited significant negative correlations supporting a central-marginal (i.e., from species range center to margin) pattern of genetic diversity decline. The strongest correlations were observed for A (Adjusted R-squared = 0.53; p-value = 0.001), followed by Ar (Adjusted R-squared = 0.44; p-value = 0.004), H_S (Adjusted R-squared = 0.39; p-value = 0.007), pA (Adjusted R-squared = 0.2; p-value = 0.054), and H_O (Adjusted R-squared = 0.18; p-value = 0.059) (Fig. 2).

Regressions performed with suitability values (obtained through niche modeling) did not present significant correlations with distance to core nor genetic diversity values. Maps with the projection of the niche models over the studied area can be found in Figure S10 (Supporting information). Other detailed results of niche modeling can be found in Appendix S2 (Supporting information).

We found significant correlation between the logarithm of geographic distances and, linearized pairwise F_{ST} (adjusted R-squared = 0.14; $P < 0.001$), linearized D_{PS} (adjusted R-squared = 0.08; P = 0.002), and Nei's (adjusted R-squared = 0.03; P = 0.033) genetic distances (Fig. 3), supporting isolation by distance model.

In the sPCA analysis of the genetic data, the first eigenvalue (sPC1) strongly differentiated *P. depauperata* and *P. integrifolia* populations (Fig. 4a), while the sPC2, that accounts for most of the spatial correlation of our dataset, showed a pattern consistent with central-to-north and central-to-south genetic differentiation (Fig. 4b). The sPCA bar plot showed a sharp decay of explained variance after the second eigenvalue (Fig. 4). Monte-Carlo randomization test of the sPCA indicated significant global genetic structure (P = 0.009), but no significant local genetic structure (P = 0.839).

Migration models

The comparison between six historical migration models (Fig. S1, Supporting information) in MIGRATE-N suggested that the source-sink-from-basal model (i.e., unidirectional migration from 'Basal' population group to all coastal populations) is the best model to explain the genetic data (Table 2). Estimates of historical migration rates obtained from this model showed higher migration rates from 'Basal' to 'South 1' and 'South 2' population groups (Table S4, Supporting information).

Isolation by environment tests

A new set of eight climatic surfaces was generated in this study and was used for niche modeling and IBE tests (see details on climatic variables in Appendix S2, Supporting information).

Mantel tests implemented with the climatic variables obtained in this study and genetic dissimilarity matrices support the hypothesis of a relationship between genetic differentiation, and precipitation seasonality and summer mean maximum temperature values (Table 3). The results were similar for the three genetic matrices and remain significant after controlling for spatial (EUC and TOPO) distances, although correlations were higher for the TOPO distances (that incorporates landscape features).

The climatic variables that better fit MLPE models were precipitation seasonality with F_{ST} (ΔAICc>8 and ΔBIC>11); precipitation seasonality and summer mean maximum temperature with D_{PS} (ΔAICc>6 and ΔBIC >8); and precipitation seasonality with Nei's distance (ΔAICc>7 and ΔBIC >9) (Table S5, Supporting information). Given that the correlation between precipitation seasonality and summer mean maximum temperature was $r < 0.2$, these were used together in multivariate

models including each of the spatial distances (EUC and TOPO). In addition, those multivariate models were ranked with the univariate models for both climatic variables. The best fit models for the three genetic distances (F_{ST}, $D_{PS,}$ and Nei's) accounted precipitation seasonality alone as the explanatory variable, all with ΔAICc and ΔBIC > 3 (Table 4).

Visual explorations of the relation between F_{ST} genetic differentiation PCoA scatterplot and climatic variables correlated with genetic differentiation were obtained with 3D-scatterplots (Fig. 5). We found that most genetic differentiation occurs between populations of *P. integrifolia* (fossil dunes region) and the population of the northern region. At the same time these populations present the smallest and highest values of precipitation seasonality, respectively (Fig 5B). Also, the 3D-scatterplot reveal that differences related with the values of summer mean maximum temperature are related with fossil dunes and coastal populations differentiation (Fig 5C)

The exploration of the relationship between genetic population differentiation and environment, using the WorldClim data set, showed a less consistent pattern and the correlation values were mostly marginally significant. Detailed results are shown in Appendix S2 (Supporting information).

Discussion

In this study, we used the coastal populations of the plant species, *Petunia integrifolia,* as a model system to examine spatial and ecological constraints on population genetic differentiation and gene flow patterns in a coastal colonization process. Our results show several associations of genetic divergence with environmental differences between coastal and inland populations, as well as between core and edge coastal populations. Demographic processes, such as continuous

founding effects, genetic drift and core-to-margin genetic enrichment, could have facilitated local adaptation to edge conditions.

Demographical and drift processes in the coast colonization

Dramatic genetic diversity reduction is expected during colonization processes due to the population size reduction and the genetic bottleneck (Pannell & Dorken 2006). The diversity values found in this study support the presence of a colonization signature in *P. depauperata*. Despite of the fact that inland and fossil dunes populations are less numerous in our sampling, these populations present higher genetic diversity values (H_S = 0.63; sd = 0.06) than the coastal populations (H_S = 0.51; sd = 0.09) (Table 1). Additionally, the sharp genetic differentiation observed between inland-fossil dunes populations with respect to coastal populations (Figs. S3B, and S4B) could be interpreted as consequence of a founder effect in a single colonization process to the coast through the central region of SACP.

Demographic processes were also recovered in the spreading process through the coast. Serial founder effects during the range expansions (Slatkin & Excoffier 2012) following a 'propagule pool' colonization model (Slatkin 1977) explain the patterns of observed genetic diversity of coastal range expansion. The process could have involved waves of occupancy following periods of deposition of the SACP during the sea transgression-regression cycles through the last 400 000 years. Sea transgression periods could have isolated founder coastal populations in the central region of the SACP where relatively higher altitudes and more substrates were available allowing the surviving of founder populations and the accumulation of genetic diversity. Toward the edges (northwards and southwards), the substrates are more recent and their colonization could have occurred with serial founder effects originating new

populations with progressively decreasing genetic diversity leading to the central–marginal pattern of gene diversity (Fig. 2), and the fixation of common alleles in populations located towards the edge of the distribution (Appendix S3, Supporting information). Grassland-forest mosaics dynamics during the last 5 kya driven by weather changes (Macedo *et al.* 2010) could also have contributed to maintaining the signal of the serial founder effects in the *P. depauperata* genome.

Center to margins gene flow facilitate colonization and adaptation processes in the coast colonization

In addition to the neutral processes proposed above, gene flow patterns inferred from our results could be better explained by selective constraints. Despite expectations that microsatellites markers only reflect neutral differentiation, patterns of population differentiation related to selection and local adaptation could be detectable in the case of maladapted immigrants who do not successfully breed with individuals from local populations (Sexton *et al.* 2014).

Our results support a main allele moving from the fossil dunes region (identified as the origin region of all coastal lineage; Ramos-Fregonezi *et al.* 2015) to the coastal populations. Source-sink gene flow dynamic from the origin center of *P. depauperata* to the populations on the coast can facilitate the colonization process by increasing genetic variation for sustaining new populations which can persist thanks to genetic diversity enrichment from the core (Sexton *et al.* 2009; Hampe *et al.* 2013). Moreover, adaptive alleles can emerge in populations from the origin center and move to edges where local adaptation can occur (Rolland *et al.* 2015). The emergence of alleles in the origin center populations, useful for coast colonization, explains the

MIGRATE-N results that support gene flow from the 'Basal' to the other populations, as opposed to the other way round.

Lower migration rates estimated between the close 'Basal' and 'Center' population groups (Table S4, Supporting information) could suggest that the coastal environment of the 'Center' group would restrict the gene flow between them (an ecological constraint). On the other hand, higher migration rates found from the 'Basal' population group to the western-most *P. depauperata* populations ('South2' group; Table S4, Supporting information) separated from the sea by the Patos Lagoon (Figs. 1 and S1) could be explained by a relaxed selective pressure for immigrants to this region due to weaker sea influence. Higher historical gene flow from 'Basal' to south populations found in this study is congruent with the fact that some individuals from the Rio Grande population presented plastid haplotypes from the inland lineage (Ramos-Fregonezi *et al.* 2015); it also allows us to propose the possibility of a secondary contact between inland and coastal lineages isolated after their divergence.

Divergent selection driven by the environment can promote genome-wide differentiation by reducing gene flow and enhancing the stochastic effects of genetic drift (Nosil *et al.* 2009). Selective constraints related with the coastal environment could be mainly responsible for the genetic differentiation found between inland and coastal populations. Main morphological differences between *P. integrifolia* and *P. depauperata* are traits commonly associated to salinity and sand-soil adaptations (Stehmann & Bohs 2007). Niche models showed high affinity of *P. depauperata* to the coastal climatic conditions since suitability values breakdown westwards (Fig. S10, Supporting information). None of the areas proposed as the origin center of *P. depauperata* by Ramos-Fregonezi *et al.* (2015) presented high suitability values, not even the Porto Alegre region whose population present chloroplast haplotypes from

the coastal lineage. It is plausible that during the marine transgressions, Porto Alegre and fossil dunes regions had suitable conditions for the divergence and establishment of the coastal lineage, but under contemporary sea levels those areas turned to continental climatic conditions. Furthermore, higher gene flow from the inland lineage populations to Porto Alegre and fossil dunes regions can explain the fact that those populations present higher nuclear genetic affinity with populations from inland lineage than from coastal lineage (Figs. S3-S4).

Several genes were identified to be related with adaptation to salinity stress in *Petunia* x *hybrida* (Villarino *et al.* 2014) that could be related in the divergence of *P. depauperata*. Differential salt tolerance between natural populations of plants from saline and non-saline environments have been documented in *Trifolium repens* (Ab-Shukor *et al.* 1988). Reproductive isolation between inland and coastal populations were reported in *Mimulus guttatus* partially as a product of selection against migrants (Lowry *et al.* 2008). Although salt resistance could be a complex trait, genetic and physiological mechanisms for adaptive divergence in salt tolerance between coastal and inland populations were described for *M. guttatus* (Lowry *et al.* 2009). However, adaptation to coastal environments can be achieved through recruitment of different genes involved in similar processes, and this variation can be observed even at the intra-specific level (Roda *et al.* 2013). Our study and all these resources provide useful information for further characterizations of genomic features of the colonization to coastal environments using *P. depauperata* as the study system.

In addition to the salinity, differences of maximum temperature values appeared to be related with genetic differentiation in our data set (Tables 3, 4). Higher maximum temperature values were found in fossil dune region populations (Fig. 5c). Summer seasonal drought or higher evapotranspiration related by higher temperatures

could be involved in gene flow reduction between inland and coastal populations as was also detected in *M. guttatus* (Lowry *et al.* 2008). Further studies in *P. depauperata* involving genomic, transcriptomic approaches or through target genes should be implemented to improve the knowledge of the genetic bases of the colonization to coastal environments.

Divergent selection could be enhancing genetic differentiation at northern and southern edges of the distribution. Stochastic allele fixation process due to genetic drift or repetitive founder effects could decrease the appearance probability of advantageous mutations, however, the same processes could accelerate the fixation of emerging advantageous mutations (Aguilée *et al.* 2009). This process was empirically supported in *Myodes glareolus* that presents adaptation signals during the range expansion in spite of the loss of genetic diversity due to genetic drift at the distribution edges (White *et al.* 2013); supporting that natural selection is also an important force in edge populations differentiation.

Similarly, we found reduced values of genetic diversity (Table 1) and stronger genetic differentiation in edge populations (Fig. 5a). In addition, the increase of precipitation seasonality values appear to be joined to higher genetic differentiation of northernmost edge populations (Garopaba and Torres; Fig. 5b). As we found significant relationship between values of genetic and precipitation seasonality differentiation, we can infer a possible adaptive divergence driven by this climatic variable in northern populations of *P. depauperata*. Additionally, Garopaba and Torres populations are located in a region with contrastable climatic conditions to the remaining populations of *P. depauperata*. Northern SACP region presents strong influence of orographic rainfalls during the spring-summer season and present the highest values of precipitation seasonality in the SACP. Further contributions are

encouraged to test the relationship between rainfall conditions and the differentiation of northern edge population in the SACP.

Advantages of implement landscape genetic analyses using climatic data derived from local resources

The comparison of the climate related analyses outcomes when implemented with the variables obtained in this study and the WorldClim set reveal statistical power differences. When raster layers were used as explanatory variables for niche modeling, the outputs obtained with both sets of climate variables were similar. Suitability values taken from locations used for train the models did not show significant differences (Wilcoxon rank test = 1443.5, p-value = 0.34. Figs. S9), and maps of projected niche models were visually similar (Fig. S10, Supporting information). However, when variables were used independently to test relationships between them and genetic differentiation (i.e., Mantel tests and MLPE, see results) performance differences become evident. The values of the variable precipitation seasonality obtained in this study presented a consistent significant relationship with all measures of genetic dissimilarities in both methods used (Tables 3, 4, and S4). In contrast, when extracted from the same variable of the WorldClim (Bio 15) set, the results differed (Tables S5-S7).

Differences between the values of the variables obtained in this study and those of the WorldClim can be interpreted mainly as product of the higher amount of climatic stations used to acquire data for the climatic layers obtained in the present work (Fig. S1, Supporting information). This difference is more important for precipitation variables, which are more sensitive to low data density than temperature variables, due to the precipitation variables are influenced by more complex features meanwhile

the temperature variables are better modeled using just altitude and latitude as co-variables. Moreover, the spatial resolution (smaller for WorldClim dataset) as well as the rounding treatments in the generation of WorldClim variables could explain the differences between values extracted from both climate surfaces datasets.

Acknowledgments

The authors thank A.M.C. Ramos-Fregonezi, J.N. Fregonezi, and G. Mäder (PPGBM-UFRGS) for help with the fieldwork; C. Turchetto (PPGBM-UFRGS) for assistance during laboratory work; R. Hijmans (UCDAVIS) for providing the R codes for the climatic variables interpolation; R. Jaffé (USP) for help with the MLPE analysis; G. Hofmann (LABGEO-UFRGS) for help in the climatic data collection; and members of the Carstens lab (EEOB, OSU) for comments and discussion regarding the manuscript. This project was supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), and Programa de Pós Graduação em Genética e Biologia Molecular da Universidade Federal do Rio Grande do Sul (PPGBM-UFRGS). G.A.S.-A. is supported by a scholarship from the Francisco José Caldas Institute for the Development of Science and Technology (COLCIENCIAS).

This work was conducted under permit MP 2.186/16 of the Brazilian Federal Government to access plant genetic information to develop evolutionary or taxonomic studies. No specific collection permits were required because these species are not federally listed as endangered or protected, and because no collection sites were located in protected areas.

References

Ab-Shukor NA, Kay QON, Stevens DP, Skibinski DOF (1988) Salt Tolerance in Natural Populations of Trifolium repens L. *New Phytologist*, **109**, 483–490.

Aguilée R, Claessen D, Lambert A (2009) Allele fixation in a dynamic metapopulation: Founder effects vs refuge effects. *Theoretical Population Biology*, **76**, 105–117.

Araujo M, New M (2007) Ensemble forecasting of species distributions. *Trends in Ecology & Evolution*, **22**, 42–47.

Beerli P, Felsenstein J (2001) Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. *Proceedings of the National Academy of Sciences*, **98**, 4563–4568.

Beerli P, Palczewski M (2010) Unified framework to evaluate panmixia and migration direction among multiple sampling locations. *Genetics*, **185**, 313–326.

Bivand R, Keitt T, Rowlingson B (2015) *rgdal: Bindings for the Geospatial Data Abstraction Library. R package version 0.9-3. http://CRAN.R-project.org/package=rgdal*.

Bossolini E, Klahre U, Brandenburg A, Reinhardt D, Kuhlemeier C (2011) High resolution linkage maps of the model organism Petunia reveal substantial synteny decay with the related genome of tomato. *Genome*, **54**, 327–340.

Chhatre VE (2012) *StrAuto ver0.3.1: A Python utility to automate Structure analysis*.

Clarke RT, Rothery P, Raybould AF (2002) Confidence limits for regression relationships between distance matrices: Estimating gene flow with distance. *Journal of Agricultural, Biological, and Environmental Statistics*, **7**, 361–372.

Dieringer D, Schlötterer C (2003) MICROSATELLITE ANALYSER (MSA): a platform independent analysis tool for large microsatellite data sets. *Molecular Ecology Notes*, **3**, 167–169.

Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, **4**, 359–361.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564–567.

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics*, **164**, 1567–1587.

Ferchaud A-L, Hansen MM (2016) The impact of selection, gene flow and demographic history on heterogeneous genomic divergence: three-spine sticklebacks in divergent environments. *Molecular Ecology*, **25**, 238–259.

Goudet J (2005) HIERFSTAT, a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, **5**, 184–186.

Goudet J (2014) *HIERFSTAT: Estimation and tests of hierarchical F-statistics. R package version 0.04-14. http://CRAN.R-project.org/package=hierfstat*.

Hampe A, Pemonge M-H, Petit RJ (2013) Efficient mitigation of founder effects during the establishment of a leading-edge Oak population. *Proceedings of the Royal Society B: Biological Sciences*, **280**, 20131070–20131070.

Hijmans RJ (2015) *RASTER: Geographic Data Analysis and Modeling. R package version 2.3-40. http://CRAN.R-project.org/package=raster*.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965–1978.

Hoegh-Guldberg O, Bruno JF (2010) The impact of climate change on the world's marine ecosystems. *Science*, **328**, 1523–1528.

Hoekstra HE, Krenz JG, Nachman MW (2005) Local adaptation in the rock pocket mouse (*Chaetodipus intermedius*): natural selection and phylogenetic history of populations. *Heredity*, **94**, 217–228.

Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources*, **9**, 1322–1332.

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, **23**, 1801–1806.

Jombart T (2008) ADEGENET: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**, 1403–1405.

Jombart T, Ahmed I (2011) ADEGENET 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics*, **27**, 3070–3071.

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

Jombart T, Devillard S, Dufour A-B, Pontier D (2008) Revealing cryptic spatial patterns in genetic variability by a new multivariate method. *Heredity*, **101**, 92–103.

Kamvar ZN, Brooks JC, Grünwald NJ (2015) Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. *Frontiers in Genetics*, **6**.

Kamvar ZN, Tabima JF, Grünwald NJ (2014) POPPR : An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, **2**, e281.

Laurent S, Pfeifer SP, Settles ML *et al.* (2016) The population genomics of rapid adaptation: disentangling signatures of selection and demography in white sands lizards. *Molecular Ecology*, **25**, 306–323.

Legendre P, Fortin M-J, Borcard D (2015) Should the Mantel test be used in spatial analysis? *Methods in Ecology and Evolution*, **6**, 1239–1247.

Ligges U, Mächler M (2003) SCATTERPLOT3D - an R package for visualizing multivariate data. *Journal of Statistical Software*, **8**, 1–20.

Longo D, Lorenz-Lemke AP, Mäder G, Bonatto SL, Freitas LB (2014) Phylogeography of the *Petunia integrifolia* complex in southern Brazil. *Botanical Journal of the Linnean Society*, **174**, 199–213.

Lowry DB, Hall MC, Salt DE, Willis JH (2009) Genetic and physiological basis of adaptive salt tolerance divergence between coastal and inland *Mimulus guttatus*. *New Phytologist*, **183**, 776–788.

Lowry DB, Rockwood RC, Willis JH (2008) Ecological reproductive isolation of coast and inland races of *Mimulus guttatus*. *Evolution*, **62**, 2196–2214.

Macedo RB, Souza PA, Bauermann SG, Bordignon SAL (2010) Palynological analysis of a late Holocene core from Santo Antônio da Patrulha, Rio Grande do Sul, Southern Brazil. *Anais da Academia Brasileira de Ciências*, **82**, 731–745.

Mäder G, Fregonezi JN, Lorenz-Lemke AP, Bonatto SL, Freitas LB (2013) Geological and climatic changes in Quaternary shaped the evolutionary history of *Calibrachoa heterophylla*, an endemic South-Atlantic species of petunia. *BMC Evolutionary Biology*, **13**, 178.

Manel S, Holderegger R (2013) Ten years of landscape genetics. *Trends in Ecology & Evolution*, **28**, 614–621.

Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. *TRENDS in ecology and evolution*, **18**, 189–197.

Mantel N (1967) The detection of disease clustering and a generalized regression approach. *Cancer Research*, **27**, 209–220.

McRae BH (2006) Isolation by resistance. *Evolution*, **60**, 1551–1561.

Miller MA, Pfeiffer W, Schwartz T (2010) Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: *Proceedings of the Gateway Computing Environments Workshop (GCE)*, pp. 1–8. New Orleans.

Misiewicz TM, Fine PVA (2014) Evidence for ecological divergence across a mosaic of soil types in an Amazonian tropical tree: *Protium subserratum* (Burseraceae). *Molecular Ecology*, **23**, 2543–2558.

Nosil P, Funk DJ, Ortiz-Barrientos D (2009) Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, **18**, 375–402.

Nychka D, Furrer R, Sain S (2015) *FIELDS: Tools for spatial data. R package version 8.2-1. http://CRAN.R-project.org/package=fields*.

Oksanen J, Blanchet FG, Kindt R *et al.* (2015) *VEGAN: Community ecology package. R package version 2.3-0. http://CRAN.R-project.org/package=vegan*.

Pannell JR, Dorken ME (2006) Colonization as a common denominator in plant metapopulations and range expansions: effects on genetic diversity and sexual systems. *Landscape Ecology*, **21**, 837–848.

Pena EA, Slate EH (2014) *GVLMA: Global validation of linear models assumptions. R package version 1.0.0.2. http://CRAN.R-project.org/package=gvlma*.

Pinheiro J, Bates D, DebRoy S, Sarkar D, R Core Team (2015) *NLME: Linear and nonlinear mixed effects models. R package version 3.1-121. http://CRAN.R-project.org/package=nlme*.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

R Core Team (2016) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

Ramos-Fregonezi AM, Fregonezi JN, Cybis GB *et al.* (2015) Were sea level changes during the Pleistocene in the South Atlantic Coastal Plain a driver of speciation in *Petunia* (Solanaceae)? *BMC Evolutionary Biology*, **15**, 92.

Reyer CPO, Leuzinger S, Rammig A *et al.* (2013) A plant's perspective of extremes: terrestrial plant responses to changing climatic variability. *Global Change Biology*, **19**, 75–89.

Roberts DW (2015) *LABDSV: Ordination and multivariate analysis for ecology. R package version 1.7-0. http://CRAN.R-project.org/package=labdsv*.

Roda F, Liu H, Wilkinson MJ *et al.* (2013) Convergence and divergence during the adaptation to similar environments by an Australian Groundsel. *Evolution*, **67**, 2515–2529.

Rolland J, Lavergne S, Manel S (2015) Combining niche modelling and landscape genetics to study local adaptation: A novel approach illustrated using alpine plants. *Perspectives in Plant Ecology, Evolution and Systematics*, **17**, 491–499.

Rosenberg NA (2003) DISTRUCT: A program for the graphical display of population structure. *Molecular Ecology Notes*, **4**, 137–138.

Rousset F (1997) Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics*, **145**, 1219–1228.

Roy A, Frascaria N, MacKay J, Bousquet J (1992) Segregating random amplified polymorphic DNAs (RAPDs) in *Betula alleghaniensis*. *TAG Theoretical and Applied Genetics*, **85**, 173–180.

Sexton JP, Hangartner SB, Hoffmann AA (2014) Genetic isolation by environment or distance: Which pattern of gene flow is most common? *Evolution*, **68**, 1–15.

Sexton JP, McIntyre PJ, Angert AL, Rice KJ (2009) Evolution and ecology of species range limits. *Annual Review of Ecology, Evolution, and Systematics*, **40**, 415–436.

Slatkin M (1977) Gene flow and genetic drift in a species subject to frequent local extinctions. *Theoretical Population Biology*, **12**, 253–262.

Slatkin M, Excoffier L (2012) Serial founder effects during range expansion: A spatial analog of genetic drift. *Genetics*, **191**, 171–181.

Smouse PE, Long JC, Sokal RR (1986) Multiple regression and correlation extensions of the Mantel test of matrix correspondence. *Systematic Zoology*, **35**, 627.

Stehmann JR, Bohs L (2007) Nuevas combinaciones en Solanaceae. *Darwiniana*, **45**, 240–241.

Storfer A, Murphy MA, Evans JS *et al.* (2007) Putting the "landscape" in landscape genetics. *Heredity*, **98**, 128–142.

Storfer A, Murphy MA, Spear SF, Holderegger R, Waits LP (2010) Landscape genetics: where are we now? *Molecular Ecology*, **19**, 3496–3514.

Thuiller W, Georges D, Engler R (2014) *BIOMOD2: Ensemble platform for species distribution modeling. R package version 3.1-62/r677. http://R-Forge.R-project.org/projects/biomod/*.

Tomazelli LJ, Dillenburg SR (2007) Sedimentary facies and stratigraphy of a last interglacial coastal barrier in south Brazil. *Marine Geology*, **244**, 33–45.

Tomazelli LJ, Dillenburg SR, Villwock JA (2000) Late Quaternary geological history of Rio Grande do Sul coastal plain, southern Brazil. *Revista Brasileira de Geociências*, **30**, 474–476.

Turchetto C, Segatto ALA, Beduschi J, Bonatto SL, Freitas LB (2015) Genetic differentiation and hybrid identification using microsatellite markers in closely related wild species. *AoB Plants*, **7**, plv084.

Villarino GH, Bombarely A, Giovannoni JJ, Scanlon MJ, Mattson NS (2014) Transcriptomic analysis of *Petunia hybrida* in response to salt stress using high throughput RNA sequencing. *PLoS ONE*, **9**, e94651.

Weising K, Freitag H (2007) Phylogeography of halophytes from European coastal and inland habitats. *Zoologischer Anzeiger - A Journal of Comparative Zoology*, **246**, 279–292.

Weschenfelder J, Corrêa ICS, Aliotta S, Baitelli R (2010) Paleochannels related to late Quaternary sea-level changes in southern Brazil. *Brazilian Journal of Oceanography*, **58**, 35–44.

White TA, Perkins SE, Heckel G, Searle JB (2013) Adaptive evolution during an ongoing range expansion: the invasive bank vole (*Myodes glareolus*) in Ireland. *Molecular Ecology*, **22**, 2971–2985.

Zhu J-K (2001) Plant salt tolerance. *Trends in Plant Science*, **6**, 66–71.

Authors' contribution

G.A.S.-A. did laboratory work to obtain the genetic data, analyzed the data, interpreted the results, and wrote the first version of the manuscript; M.R.-K. contributed with laboratory work to obtain the genetic data; B.C.C. contributed to the analyses and manuscript writing; H.H. contributed with climatic data collection and processing; S.L.B. provided the microsatellite genotyping and help with manuscript writing; and L.B.F. designed and coordinated the study, obtained the funding, helped in collecting samples, and manuscript writing. All authors have read draft versions of the manuscript.

Data accessibility

Supporting information

Additional supporting information may be found in the online version of this article.

Appendix S1. Methodological details.
Table S1. Longitude-latitude coordinates of biological records used in the ecological niche modeling.
Figure S1. Diagrams of the six migration models tested with MIGRATE-N.
Figure S2. Geographic localization of meteorological stations that provided climatic data for the climatic surfaces obtained in this study.

Appendix S2. Results details.
Table S2. Locus-population pairs showing significant departure from Hardy-Weinberg Equilibrium.
Table S3. Summary statistics and genome localization of each of the 10 EST-SSR loci used to characterize the genetic diversity.

Table S4. Mutation-scaled effective immigration rates (M) estimated with MIGRATE-N for the best supported model (Source-sink-from-inland model; see results).
Table S5. Maximum likelihood population effects (MLPE) modeling results investigating the relationships of F_{ST}, D_{PS} and Nei's genetic differentiation statistics with the eight climatic variables obtained in this study and Euclidean or topographic spatial distances.
Table S6. Relationships tests between three population genetic differentiation statistics (F_{ST}, D_{PS}, and Nei's) and inter-population distances measured for each climatic variable obtained from WorldClim calculated with Mantel and partial Mantel tests controlling by spatial distances.
Table S7. Maximum likelihood population effects (MLPE) modeling results investigating the relationships of F_{ST}, D_{PS}, and Nei's genetic differentiation statistics with climatic variables from WorldClim and Euclidean (EUC) or topographic (TOPO) spatial distances.
Table S8. Maximum likelihood population effects (MLPE) modeling results investigating the relationships of F_{ST}, D_{PS}, and Nei's genetic differentiation statistics with single and mixed three climatic variables from WorldClim and Euclidean (EUC) or topographic (TOPO) spatial distances.
Figure S3. Bar plots obtained with STRUCTURE analyses. Populations are separated by vertical tick black lines; population names as indicated in Table 1 and Fig.1 are displayed on top and bottom of bar plots. (a) Plot of optimal K obtained with STRUCTURE HARVESTER. (b) Bar plots for the complete data set with K=2 to K=7. Different colors indicate groups and vertical bars correspond to each individual.
Figure S4. Bar plots obtained with DAPC analyses. Populations are separated by vertical tick black lines; population names as indicated in Table 1 and Fig.1 are displayed on top and bottom of bar plots. (a) Plot of optimal K obtained with *find.clusters* function of adegenet R package. (b) Bar plot for the complete data set with K=2 to K=7. Different colors indicate groups and vertical bars correspond to each individual.
Figure S5. Density plots comparing annual precipitation values. **a**. Observed values at climate stations used for fit the Tps models, fitted values with Tps model for station localities and values extracted from the same variable of the WorldClim for station localities. **b**. Extracted from the interpolated Tps model for *P. depauperata* localities and extracted from the same variable of the WorldClim for *P. depauperata* localities.
Figure S6. Density plots comparing mean annual temperature values. **a**. Observed values at climate stations used for fit the Tps models, fitted values with Tps model for station localities, and values extracted from the same variable of the WorldClim from station localities. **b**. Values extracted from the climatic surfaces generated with the Tps model for *P. depauperata* populations and extracted from the same variable of the WorldClim dataset for *P. depauperata* populations.
Figure S7. Density plots comparing precipitation seasonality values. **a**. Observed values at climate stations used for fit the Tps models, fitted values with Tps model for station localities, and values extracted from the same variable of the WorldClim from station localities. **b**. Values extracted from the climatic surfaces generated with the Tps model for *P. depauperata* populations and extracted from the same variable of the WorldClim dataset for *P. depauperata* populations.
Figure S8. Density plots comparing mean diurnal range values. **a**. Observed values at climate stations used for fit the Tps models, fitted values with Tps model for station localities and values extracted from the same variable of the WorldClim from station

localities. **b**. Values extracted from the climatic surfaces generated with the Tps model for *P. depauperata* populations and extracted from the same variable of the WorldClim dataset for *P. depauperata* populations.
Figure S9. Density plots comparing suitability values of *P. depauperata* localities used for train the models (Table S1) using the climate variables obtained in this study and the WorldClim set.
Figure S10. Ensemble niche models projections for *Petunia depauperata* achieved with climate variables obtained in this study and WorldClim variables.

Appendix S3. Bar plots of allele frequencies per population for all loci used in genetic population analyses of *Petunia depauperata*.

Tables

Table 1. Population sampling localities, sample sizes (N), geographic coordinates, and the genetic diversity measures alleles number (A), allele richness (Ar), private alleles (pA), observed heterozygosity (H_O), gene diversity (H_S) and F_{IS}.

Population name	Longitude	Latitude	N	A	Ar	pA	H_O	H_S	F_{IS}
Inland lineage									
Itaara	-53.79067	-29.55151	8	39	2.664957	-	0.6806	0.6895	0.0129
Tapes	-51.39040	-30.66296	15	45	2.416156	1	0.50295	0.64296	0.233
Arambaré	-51.49181	-30.90530	5	27	2.306548	1	0.475	0.5844	0.1872
Guaíba	-51.31768	-30.13769	23	62	2.592867	7	0.5843	0.6698	0.1277
Viamão	-51.00495	-30.38093	20	45	2.230226	1	0.4343	0.5555	0.2183*
Coastal lineage									
Porto Alegre	-51.11773	-30.07462	14	49	2.558188	3	0.5089	0.6881	0.2604*
Osório	-50.24286	-29.81331	26	48	2.273649	1	0.4944	0.5823	0.151*
Curumim	-49.94791	-29.64295	15	43	2.329053	-	0.4637	0.5938	0.2192*
Xangri-lá	-50.02934	-29.79164	20	36	1.957372	-	0.4507	0.4519	0.0028
Pinhal	-50.23041	-30.24703	19	36	1.984831	-	0.3687	0.465	0.2071*
Torres	-49.73457	-29.35748	15	36	2.059097	-	0.4157	0.49	0.1517*
Garopaba	-48.62154	-28.02139	15	25	1.743212	-	0.3833	0.3627	-0.0568
Mostardas	-50.73934	-30.93746	16	42	2.1048	1	0.4837	0.4984	0.0295
Rio Grande	-52.17406	-32.12531	33	33	1.886662	-	0.3961	0.4279	0.0742
Taim	-52.49174	-32.60474	17	34	2.057509	-	0.4506	0.4999	0.0985
São Lourenço do Sul	-51.95317	-31.37673	21	40	2.237432	-	0.5188	0.5783	0.1029
Pelotas	-52.16478	-31.70776	25	40	2.125274	-	0.4081	0.5303	0.2305*
Overall			307	93			0.4694	0.545	0.1387

*F_{IS} statistic significantly different from zero, p-value < 0.05.

Table 2. Results of the model comparisons in MIGRATE-N. Graphical model descriptions in Fig. S1 (Supporting information). The Bézier approximation scores of log marginal likelihoods (ln ML), log Bayes factors (LBF) and model probabilities are shown. LBF values < 2 indicate strong preference for the best fit model.

Model name	ln ML	LBF	Model probability
Central-marginal	-5815.13	-75.58	3.9×10^{-17}
Marginal-central	-5893.41	-232.14	3.9×10^{-51}
Source-sink-from-basal	-5777.34	0	1
Source-sink-from-central	-5893.58	-232.48	3.3×10^{-51}
Ring	-5856.46	-158.24	4.3×10^{-35}
North-south-barrier	-5819.45	-84.22	5.1×10^{-19}

Table 3. Correlations between population genetic differentiation statistics (F_{ST}, D_{PS}, and Nei's) and inter-population distances measured for each of the eight climatic variables obtained in this study calculated with Mantel tests and partial Mantel tests controlling by spatial distances.

Genetic distance	Climatic variable	Mantel R^2	Partial Mantel tests	
			Controlling by EUC	Controlling by TOPO
F_{ST}	Precipitation seasonality	0.63***	0.55**	0.61***
	Summer mean maximum temperature	0.10	-	-
	Mean temperature range	0.01	-	-
	Total annual days with rain	0.47*	0.27	0.41*
	Total annual precipitation	0.38*	0.09	0.30
	Winter mean minimum temperature	0.44*	0.22	0.37*
	Mean annual temperature	0.12	-	-
	Temperature seasonality	0.31*	-0.01	0.23
D_{PS}	Precipitation seasonality	0.51***	0.42*	0.51***
	Summer mean maximum temperature	0.35**	0.33*	0.36**
	Mean temperature range	0.26	-	-
	Total annual days with rain	0.27	-	-
	Total annual precipitation	0.23	-	-
	Winter mean minimum temperature	0.29*	0.07	0.30*
	Mean annual temperature	0.08	-	-
	Temperature seasonality	0.21	-	-
Nei	Precipitation seasonality	0.44*	0.39*	0.43*
	Summer mean maximum temperature	0.41**	0.39**	0.42***
	Mean temperature range	0.29*	0.30*	0.28*
	Total annual days with rain	0.15	-	-
	Total annual precipitation	0.13	-	-
	Winter mean minimum temperature	0.15*	-0.04	0.13
	Mean annual temperature	0.03	-	-
	Temperature seasonality	0.13	-	-

The Mantel's statistic r is reported and the significance level is indicated. *** P < 0.001; ** P < 0.005; * P < 0.05. EUC: Euclidean geographic distances; TOPO: Geographic distances accounting topographic features of study region; - : not evaluated.

Table 4. Results of the MLPE modeling investigating the relationships of F_{ST}, D_{PS}, and Nei's pairwise population genetic differentiation with two climatic variables and Euclidean (EUC) or topographic (TOPO) spatial distances.

Genetic distance	Variables	ΔAICc	weight	ΔBIC	weight
F_{ST}	Precipitation seasonality	0	0.984	0	0.995
	Precipitation seasonality + TOPO	8.854	0.012	11.234	0.004
	Precipitation seasonality + EUC	10.880	0.004	13.259	0.001
	Summer mean maximum temperature	16.478	<0.001	16.478	<0.001
	Precipitation seasonality + Summer mean maximum temperature + TOPO	20.706	<0.001	25.401	<0.001
D_{PS}	Precipitation seasonality	0	0.776	0	0.836
	Summer mean maximum temperature	3.678	0.123	3.678	0.133
	Precipitation seasonality + EUC	6.404	0.032	8.784	0.010
	Summer mean maximum temperature + EUC	6.666	0.028	9.046	0.009
	Precipitation seasonality + TOPO	7.294	0.020	9.674	0.007
Nei	Precipitation seasonality	0	0.954	0	0.985
	Precipitation seasonality + TOPO	7.054	0.028	9.433	0.009
	Precipitation seasonality + EUC	8.828	0.012	11.208	0.004
	Precipitation seasonality + Summer mean maximum temperature + TOPO	11.736	0.003	16.431	<0.001
	Summer mean maximum temperature	12.342	0.002	12.342	0.002

Figures

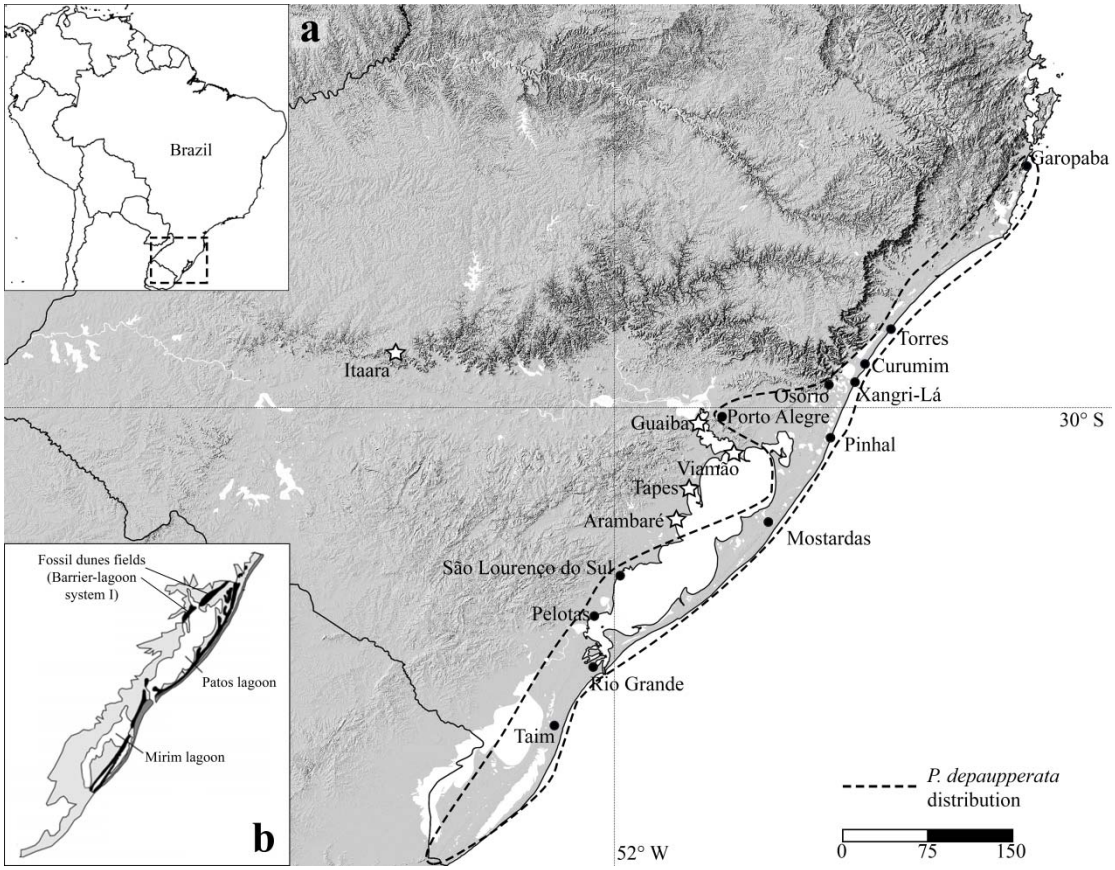


Figure 1. Localization of sampled populations included in this study. *P. integrifolia* (stars) and *P. depauperata* (circles) populations, as well as the complete known geographical distribution of *P. depauperata* are indicated. Geographical localization of the study area in South America appears at the upper-left corner. Panel **b** shows the South Atlantic Coastal Plain map modified from Tomazelli *et al.* (2000). Black stripes indicate the Pleistocene Barriers I, II, and III. Gray stripe indicate the Holocene Barrier IV. The two largest coastal lagoons of the SACP are indicated.

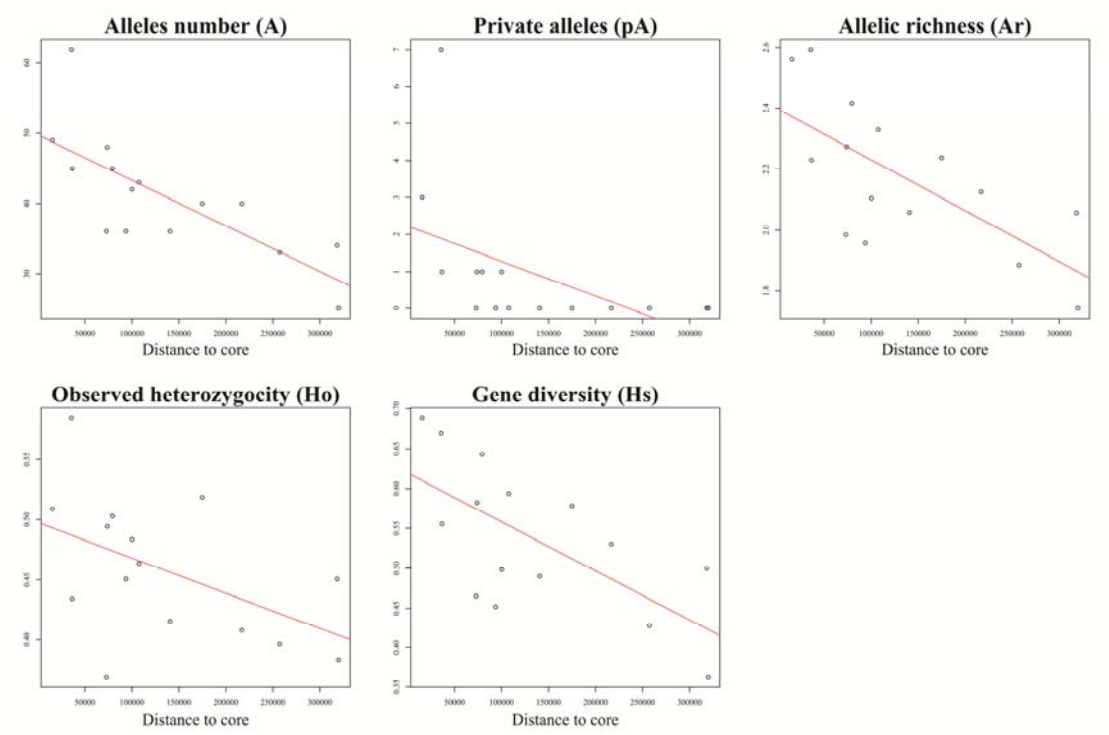


Figure 2. Tests of central-marginal decline pattern of genetic diversity. Plots of the population genetic diversity statistics vs. their distances from the core region of the distribution.

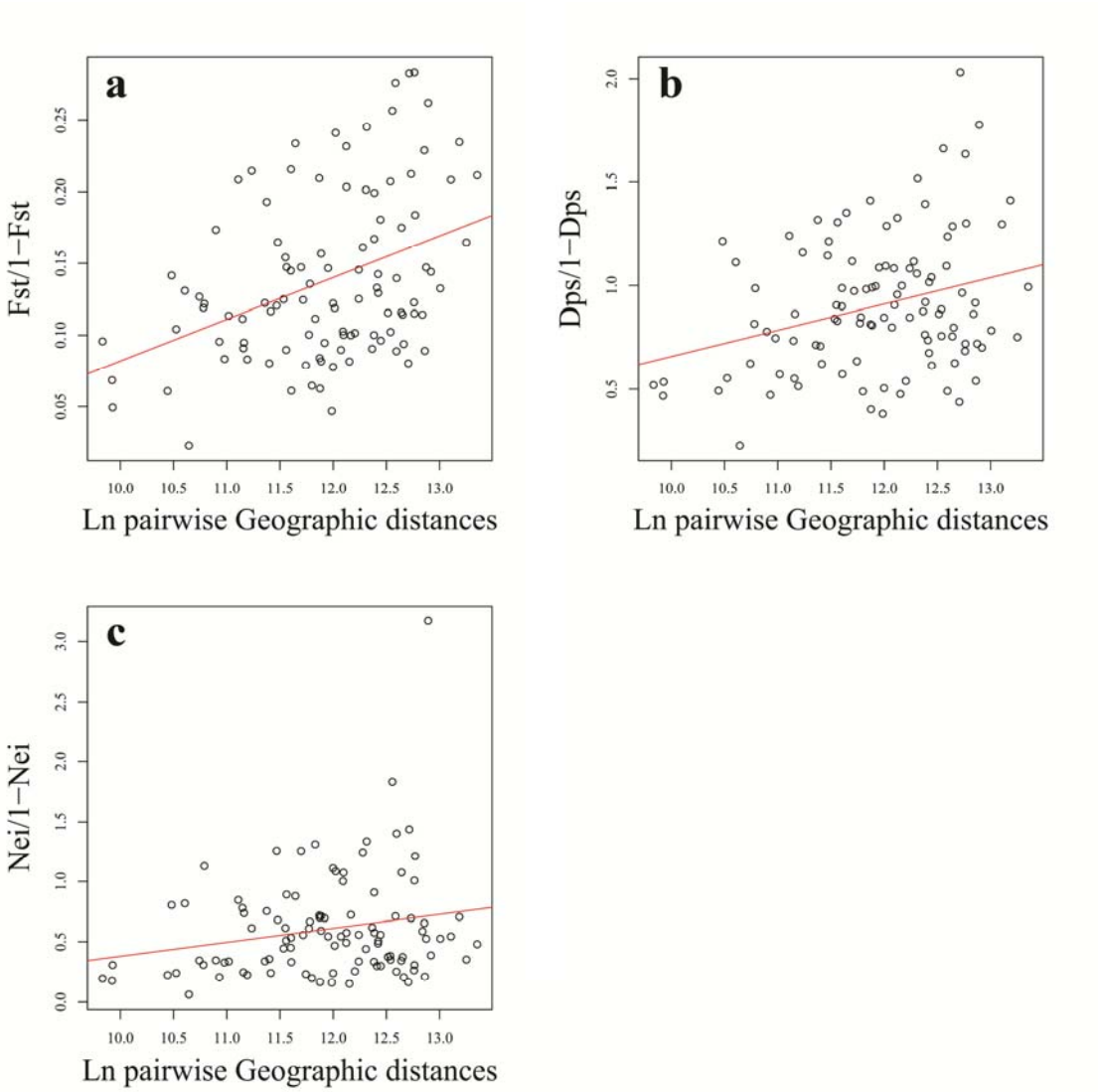


Figure 3. Isolation by distance plots. Inter-population genetic distances (**a**: F_{ST};**b**: D_{PS}; and **c**: Nei's) are plotted vs. linearized inter-population geographic distances.

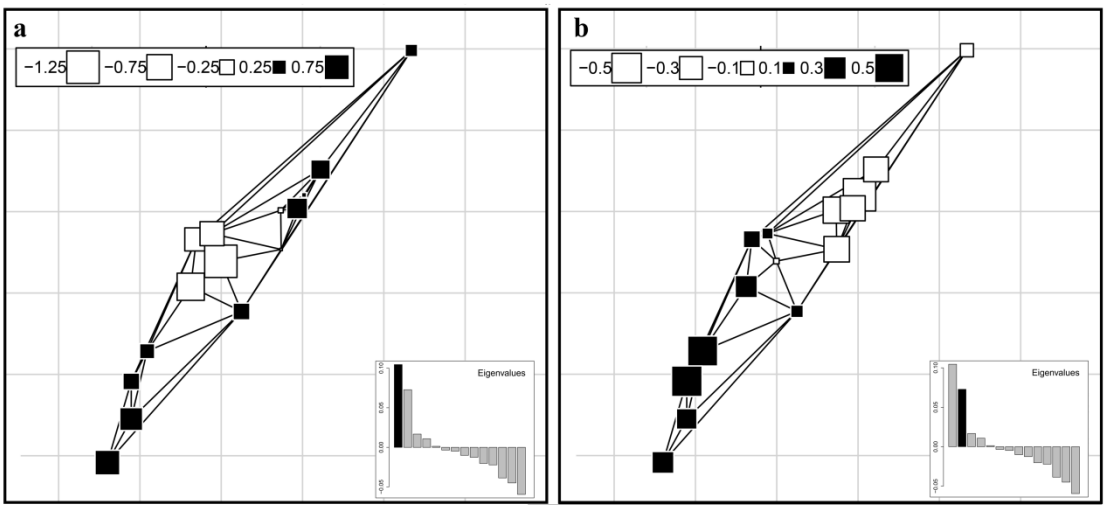


Figure 4. Graphical results of the spatial Principal Components Analysis. **a**. Synthetic map showing the inter-population differentiation pattern summarized in the sPC 1 (29% of the total variance and Moran's I = 0.36). **b**. Synthetic map showing the inter-population differentiation pattern summarized in the sPC 2 (10% of the total variance and Moran's I = 0.7).

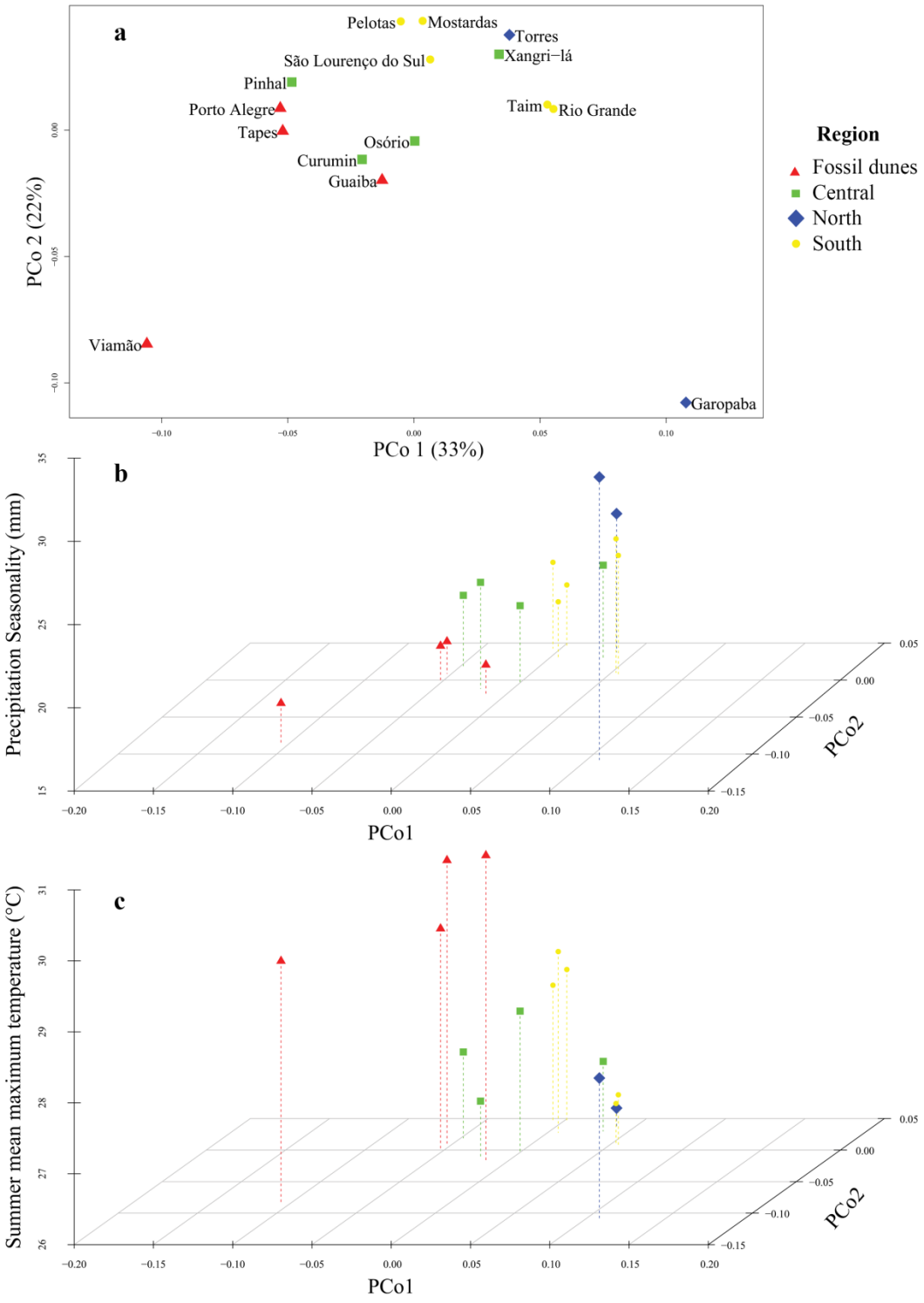


Figure 5. Principal coordinates analysis (PCoA) for the F_{ST} genetic differentiation matrix. **a**. Scatterplot of two main principal coordinates showing name and geographic group of the populations. **b**. Three dimensional PCoA scatterplot with precipitation seasonality values. **c**. Three dimensional PCoA scatterplot with summer mean maximum temperature values.

Supplementary File S1. Methodological details

DNA extraction and genotyping

We pulverized the leaves in liquid nitrogen for DNA extraction with cetyltrimethylammonium bromide (CTAB) protocol as described by Roy *et al.* (1992). The quality and quantity of the genomic DNA was evaluated by measuring the absorbance at 260 and 280 nm on a Nanodrop Spectrophotometer (NanoDrop 1000 spectrometer, Thermo Scientific Corp., USA).

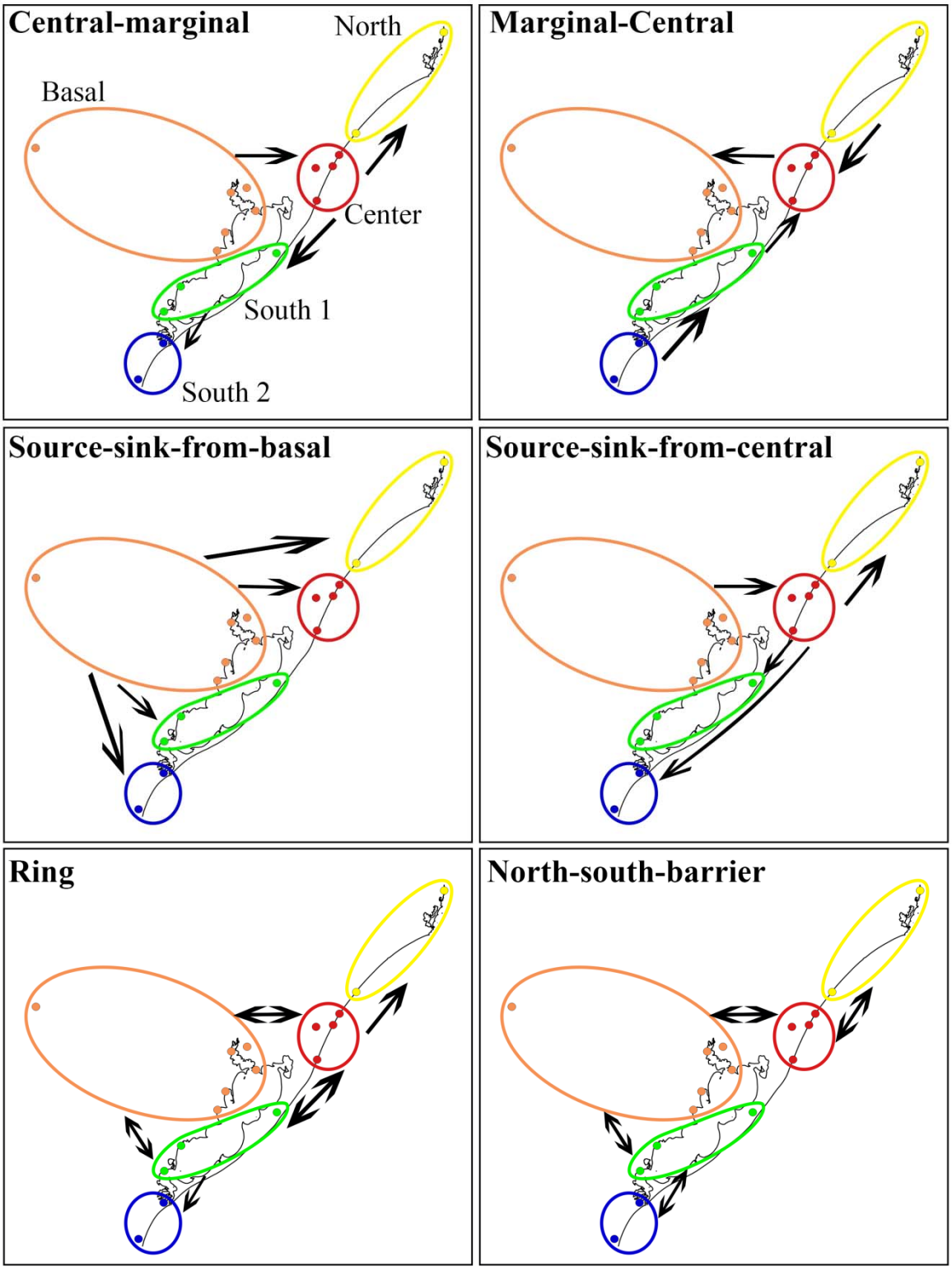
Polymerase chain reactions were conducted in a final volume of 10 μL containing ~10 ng of genomic DNA as template, 200 μM of each dNTP (Invitrogen, Carlsbad, CA, USA), 1.7 pmol of each fluorescently labelled M13(–21) primer, 3.5 pmol of reverse primer, 0.35 pmol of forward primer with a 5′-M13(–21) tail, 2.0 mM $MgCl_2$ (Invitrogen), 0.5 U of Platinum Taq DNA polymerase (Invitrogen) and 1× Platinum *Taq* reaction buffer (Invitrogen). The PCR conditions were as follow: an initial denaturation at 94 °C for 3 min; 32-35 cycles of 94 °C for 15 s, 50–54 °C for 30 s and 72 °C for 1 min; and a final extension cycle at 72 °C for 7 min. The forward primers were FAM, NED or HEX labeled. The DNA fragments were denatured and size-fractionated using capillary electrophoresis on a MegaBACE 1000 automated sequencer (GE Healthcare Biosciences, Pittsburgh, PA, USA) with a GeneTab-500 internal size ladder (GE Healthcare). The manufacturer's software was used to estimate the length of the microsatellite alleles. WE USED THE MICRO-CHECKER (Van Oosterhout *et al.*, 2004) to estimate genotyping errors due to stutter bands, allele dropout or null alleles.

Genetic diversity and structure

Loci were tested for linkage disequilibrium and deviations from Hardy–Weinberg equilibrium (HWE) were tested within each population for each locus. Significance of HWE deviations was assessed using a Markov chain method and Fisher's exact probability tests in ARLEQUIN 3.5 (Excoffier & Lischer, 2010).

For each population, we estimated genetic diversity across loci using the alleles number (A), allele richness (Ar), private alleles (pA), observed heterozygosity (H_O), gene diversity (H_S; (Nei, 1987), and inbreeding coefficient (F_{IS}). Confidence limits for F_{IS} were obtained using 1000 bootstrap resampling over loci. The analyses were implemented with packages ADEGENET 2 (Jombart, 2008; Jombart & Ahmed, 2011), POPPR 2.0.2 (Kamvar *et al.*, 2014, 2015), and HIERFSTAT 0.04-14 (Goudet, 2005, 2014) in R 3.2.1 (R Core Team, 2016).

Figure S1. Migration models tested with Migrate-n. Model names are indicated on top-left corners and population groups names are only indicated only in the 'Central-marginal' model (see Materials and methods on main text). Arrows indicate de migration direction.



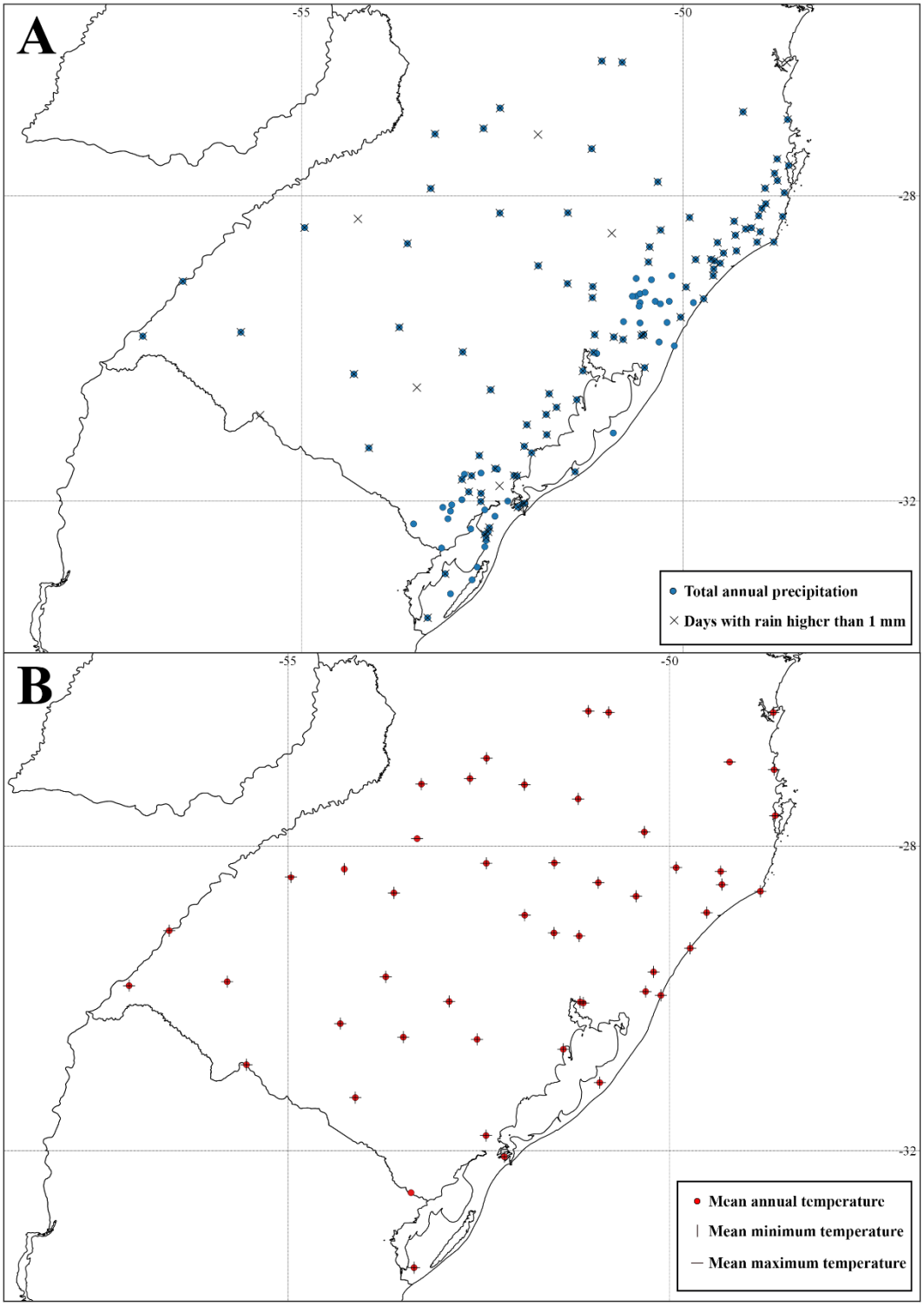
Obtaining and generation of climatic raster surfaces for SACP

The procedure implemented for model fitting was thin-plate splines (TPS). The spline model with elevation as covariate was fitted to the stations data by generalized cross-validation in the R package FIELDS 8.2-1 (Nychka *et al.*, 2015). Elevation values used in model fitting were taken from direct measures at weather stations. The use of elevation as covariate for climate interpolation assays has shown good performances in previous works (e.g. Alvarez *et al.* 2014; Pliscoff *et al.* 2014). The fitted model was then used to interpolate the variables to the studied area using the R package RASTER 2.4-15 (Hijmans, 2015). The digital elevation model Shuttle Radar Topography Mission (SRTM; (Reuter *et al.*, 2007; Jarvis *et al.*, 2008) clipped for longitude 57.1 to 48.4 W and latitude 34 to 26.1 S at 90 and 450 square meters (m^2) spatial resolutions were used as template surfaces and their elevation values as covariate for interpolation step to obtain the final climatic surfaces.

Monthly and annual data were modelled using the primary variables total precipitation (136 weather stations; Fig. S1A), days with rain higher than 1 mm (101 weather stations; Fig. S1A), mean annual temperature (50 weather stations; Fig. S1B), mean maximum temperature (48 weather stations; Fig. S1B) and mean minimum temperature (47 weather stations; Fig. S1B) obtained from online available databases of National Meteorology Institute from Brazil (INMET; http://www.inmet.gov.br/portal) and National Water Agency from Brazil (http://hidroweb.ana.gov.br). Other local data sources collected at Laboratory of Geoprocessing, Universidade Federal do Rio Grande do Sul (http://www.ecologia.ufrgs.br/labgeo/) were included. Time span of most of climatic data used were 30 years (1960-1990), but were included data with time span from 10 to 67 years.

We modeled eight variables: total annual precipitation, total annual days with rain, precipitation seasonality (coefficient of variation), mean annual temperature, mean summer maximum temperature, mean winter minimum temperature, mean temperature range [mean of monthly (max temp - min temp)], and temperature seasonality (coefficient of variation). Climatic surfaces were generated for an area larger than the extent of study area avoiding edge effects. Final climatic surfaces were keep in WGS84 system and 450 m resolution for niche modeling (see below) and projected to UTM 22S with QGIS 2.8 software (QGIS Development Team, 2015) and 90 m resolution for climatic distances calculations (see Methods section).

Figure S2. Localization of meteorological stations that provided climatic data for the climatic surfaces obtained in this study.



Ecological niche modeling and ensemble

We performed two alternative niche models in order to test the influence of two different set explanatory variables, those obtained in this study (see section above) and the complete WorldClim dataset (Hijmans *et al.*, 2005).

Species occurrence localities were taken mainly from direct field observations, but some localities derived from georeferenced herbarium specimens and SpeciesLink database were also included. All secondary occurrence points were first verified for probable errors and duplicates. Only one occurrence record per 1 km x 1km grid cell was included. A total of 67 localities covering the full extant geographic range of *P. depauperata* were included (Table S1).

The first set of explanatory variables included the eight climatic raster surfaces obtained in this study at 450 m resolution in WGS84 system. The second set included the19 bioclimate raster layers at 30 arc-second resolution (ca. 1 km at the Equator) obtained from the WorldClim database (Hijmans *et al.*, 2005). The raster layers were cropped for longitude 57.1 to 48.4 W and latitude 34 to 26.1 S with the RGDAL R package (Bivand *et al.*, 2015; R Core Team, 2016), this geographic extent include the full extant range of *P. depauperata*.

Niche models were obtained with the ensemble niche modeling framework (Araujo & New, 2007). Individual models were generated using the methods Generalized Linear Models (GLM, McCullagh & Nelder, 1983), Generalized Additive Models (GAM, Hastie & Tibshirani, 1990), Generalised Boosting Models (GBM, Death, 2007), Surface Range Envelop (SRE, Busby, 1991), Classification Tree Analysis (CTA, Breiman *et al.*, 1993), Flexible Discriminant Analysis (FDA, Hastie *et al.*, 1994), Multiple Adaptive Regression Splines (MARS, Leathwick *et al.*, 2005), Random Forests (RF, Breiman, 2001), Artificial Neural Network (ANN,

Ripley, 1996) and Maximum Entropy (MAXENT, Phillips *et al.*, 2006). Recent studies suggest that the geographic extent in which the pseudo-absences are taken have significant influences for prediction and performance of niche models (Thuiller *et al.*, 2004; VanDerWal *et al.*, 2009), following this reasoning we selected pseudo-absences as random localities throughout the full extant range of *P. integrifolia* species. We did two samplings of 10 000 random localities with 'disk' strategy, keeping 1 Km as minimum distance from presence localities. Despite Barbet-Massin *et al.* (2012) recommend to sample pseudo-absences at higher distances from presence localities, we kept 1 Km in order to account for the fine-scale of this study. We kept prevalence equal to 0.5. For each model, presence localities were divided in sets of 70% for training models and 30% for testing. For each model we performed ten runs, therefore a total of 200 models were completed for each of the four combinations of explanatory/localities sets.

Variables loadings were kept for all models calculated. To assess predictive performance of the individual and ensemble models, we measured the threshold independent statistics area under the receiver operating characteristic curve (AUC, Phillips *et al.*, 2006) and the True Skill Statistic (TSS, Allouche *et al.*, 2006).

Table S1. Longitude-Latitude coordinates of biological records used in the ecological niche modeling.

Longitude	Latitude
-53.2779	-33.6551
-53.2697	-33.6723
-53.2483	-33.656
-53.2225	-33.6371
-52.4994	-32.5168
-52.4917	-32.6047
-52.2714	-31.8668
-52.2599	-31.8795
-52.1741	-32.1253
-52.1671	-31.7078
-51.9596	-31.374
-51.4258	-31.6667
-51.0095	-31.2377
-50.9011	-31.1091
-50.7393	-30.9375
-50.7161	-30.9556
-50.5753	-30.8233
-50.5132	-29.8781
-50.4987	-29.9091
-50.4793	-29.9205
-50.4674	-29.9136

Longitude	Latitude
-50.4648	-29.8988
-50.4294	-29.8929
-50.4262	-29.9052
-50.4196	-30.5409
-50.2931	-30.359
-50.2679	-30.3407
-50.2668	-30.3355
-50.2457	-29.8323
-50.2429	-29.8133
-50.236	-30.2577
-50.2338	-30.247
-50.236	-30.2577
-50.2338	-30.247
-50.2186	-29.8879
-50.2015	-30.1741
-50.1942	-30.1495
-50.1906	-29.778
-50.1335	-30.0016
-50.0754	-29.8675
-50.0668	-29.8581
-50.042	-29.8075

Longitude	Latitude
-50.0339	-29.7916
-50.0086	-29.7422
-49.9467	-29.6398
-49.9421	-29.6295
-49.9356	-29.6247
-49.8334	-29.4746
-49.8084	-29.4413
-49.801	-29.4319
-49.7419	-29.3575
-49.7168	-29.3243
-48.6215	-28.0214
-53.5124	-33.9111
-52.3993	-32.1581
-49.8369	-29.3836
-49.7658	-29.3814
-49.6136	-29.185
-49.4178	-28.9906
-49.4305	-28.842
-49.2176	-28.8238
-48.7677	-28.4734
-48.7681	-28.4612
-48.6908	-28.2019

References

Allouche O, Tsoar A, Kadmon R (2006) Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*, **43**, 1223–1232.

Alvarez O, Guo Q, Klinger RC, Li W, Doherty P (2014) Comparison of elevation and remote sensing derived products as auxiliary data for climate surface interpolation. *International Journal of Climatology*, **34**, 2258–2268.

Araujo M, New M (2007) Ensemble forecasting of species distributions. *Trends in Ecology & Evolution*, **22**, 42–47.

Barbet-Massin M, Jiguet F, Albert CH, Thuiller W (2012) Selecting pseudo-absences for species distribution models: how, where and how many?: how to use pseudo-absences in niche modelling? *Methods in Ecology and Evolution*, **3**, 327–338.

Bivand R, Keitt T, Rowlingson B (2015) *RGDAL: Bindings for the Geospatial Data Abstraction Library. R package version 0.9-3. http://CRAN.R-project.org/package=rgdal*.

Breiman L (2001) Random forests. *Machine learning*, **45**, 5–32.

Breiman L, Friedman J, Stone CJ, Olshen RA (1993) *Classification and regression trees*. Chapman & Hall, New York, N.Y.

Busby JR (1991) BIOCLIM: a bioclimate analysis and prediction system. In: *Nature conservation: cost effective biological surveys and data analysis* (eds Margules CR, Austin MP), pp. 64–68. CSIRO, Australia.

Death G (2007) Boosted trees for ecological modeling and prediction. *Ecology*, **88**, 243–251.

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564–567.

Goudet J (2005) HIERFSTAT, a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, **5**, 184–186.

Goudet J (2014) *HIERFSTAT: Estimation and tests of hierarchical F-statistics. R package version 0.04-14. http://CRAN.R-project.org/package=hierfstat*.

Hastie TJ, Tibshirani RJ (1990) *Generalized additive models*. Chapman & Hall/CRC, London, England, 335 pp.

Hastie T, Tibshirani R, Buja A (1994) Flexible Discriminant Analysis by Optimal Scoring. *Journal of the American Statistical Association*, **89**, 1255–1270.

Hijmans RJ (2015) *RASTER: Geographic Data Analysis and Modeling. R package version 2.3-40. http://CRAN.R-project.org/package=raster*.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965–1978.

Jarvis A, Reuter HI, Nelson A, Guevara E (2008) Hole-filled seamless SRTM data V4. International Centre for Tropical Agriculture (CIAT). http://srtm.csi.cgiar.org/.

Jombart T (2008) ADEGENET: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**, 1403–1405.

Jombart T, Ahmed I (2011) ADEGENET 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics*, **27**, 3070–3071.

Kamvar ZN, Tabima JF, Grünwald NJ (2014) POPPR : An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, **2**, e281.

Kamvar ZN, Brooks JC, Grünwald NJ (2015) Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. *Frontiers in Genetics*, **6**.

Leathwick JR, Rowe D, Richardson J, Elith J, Hastie T (2005) Using multivariate adaptive regression splines to predict the distributions of New Zealand's freshwater diadromous fish. *Freshwater Biology*, **50**, 2034–2052.

McCullagh P, Nelder JA (1983) *Generalized linear models*. Chapman and Hall/CRC, London, England, 261 pp.

Nei M (1987) *Molecular evolutionary genetics*. Columbia University Press, New York, 512 pp.

Nychka D, Furrer R, Sain S (2015) *FIELDS: Tools for spatial data. R package version 8.2-1. http://CRAN.R-project.org/package=fields*.

Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, **190**, 231–259.

Pliscoff P, Luebert F, Hilger HH, Guisan A (2014) Effects of alternative sets of climatic predictors on species distribution models and associated estimates of extinction risk: A test with plants in an arid environment. *Ecological Modelling*, **288**, 166–177.

QGIS Development Team (2015) *QGIS geographic information system*. Open Source Geospatial Foundation Project. http://qgis.osgeo.org

R Core Team (2016) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

Reuter HI, Nelson A, Jarvis A (2007) An evaluation of void-filling interpolation methods for SRTM data. *International Journal of Geographical Information Science*, **21**, 983–1008.

Ripley BD (1996) *Pattern recognition and neural networks*. Cambridge University Press, Cambridge ; New York, 403 pp.

Roy A, Frascaria N, MacKay J, Bousquet J (1992) Segregating random amplified polymorphic DNAs (RAPDs) in *Betula alleghaniensis*. *TAG Theoretical and Applied Genetics*, **85**, 173–180.

Thuiller W, Brotons L, Araújo MB, Lavorel S (2004) Effects of restricting environmental range of data to project current and future species distributions. *Ecography*, **27**, 165–172.

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: Software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, **4**, 535–538.

VanDerWal J, Shoo LP, Graham C, Williams SE (2009) Selecting pseudo-absence data for presence-only distribution modeling: How far should you stray from what you know? *Ecological Modelling*, **220**, 589–594.

APPENDIX S2. Results details

Table S2. Locus-population pairs showing significant departure of Hardy-Weinberg Equilibrium.

Population	Significance level to reject HWE hypothesis		
	0.1 - 0.05	**0.05 - 0.01**	**< 0.01**
Itaara			
Tapes	PM167	PM184	PM101
Arambaré			
Guaiba			PM167, PM177
Viamão		PM101, PM167	PM177
Porto Alegre	PM157	PM184, PM117	PM167, PM177
Osório		PM101	PM177
Curumim	PM177	PM157	PM101
Xangri-lá	PM101	PM167	PM177
Pinhal	PM184		PM101, PM167, PM177
Mostardas	PM117		PM184
Torres			PM101, PM167
Garopaba	PM184	PM1PM177	PM167, PM117, PM177
Rio Grande		PM167	PM192
Taim			PM167
São Lourenço do Sul			PM21, PM167
Pelotas	PM101, PM177	PM21, PM184	PM167

Table S3. Summary statistics and genome localization of each 10 EST-SSR loci used to characterize the genetic diversity. A: allele number, H_O: observed heterozygosity, H_S: gene diversity.

Locus	Chromosome	A	Ho	Hs	F_{ST}	F_{IS}
PM101	1	5	0.1863	0.3205	0.1223	0.4187
PM21	2	7	0.5079	0.5532	0.149	0.082
PM8	4	6	0.2302	0.2271	0.2721	-0.0134
PM167	5	13	0.4966	0.7704	0.1518	0.3553
PM192	5	7	0.5709	0.5856	0.1814	0.0251
PM110	5	6	0.6075	0.6294	0.1322	0.0348
PM177	5	13	0.2847	0.4501	0.2961	0.3674
PM117	6	11	0.6613	0.7037	0.1605	0.0601
PM184	7	8	0.4942	0.5261	0.2951	0.0606
PM157	7	15	0.6539	0.6834	0.1389	0.0432

Table S4. Mutation-scaled effective immigration rates (M) estimated with MIGRATE-N for the best-supported model (Source-sink-from-basal model; see Results in main text).

Migration direction	2.50%	97.50%	mean
Basal to Central	0.07	3.47	1.79
Basal to North	2.2	8.33	5.21
Basal to South1	75.87	100	84.9
Basal to South2	29.4	87.8	53.98

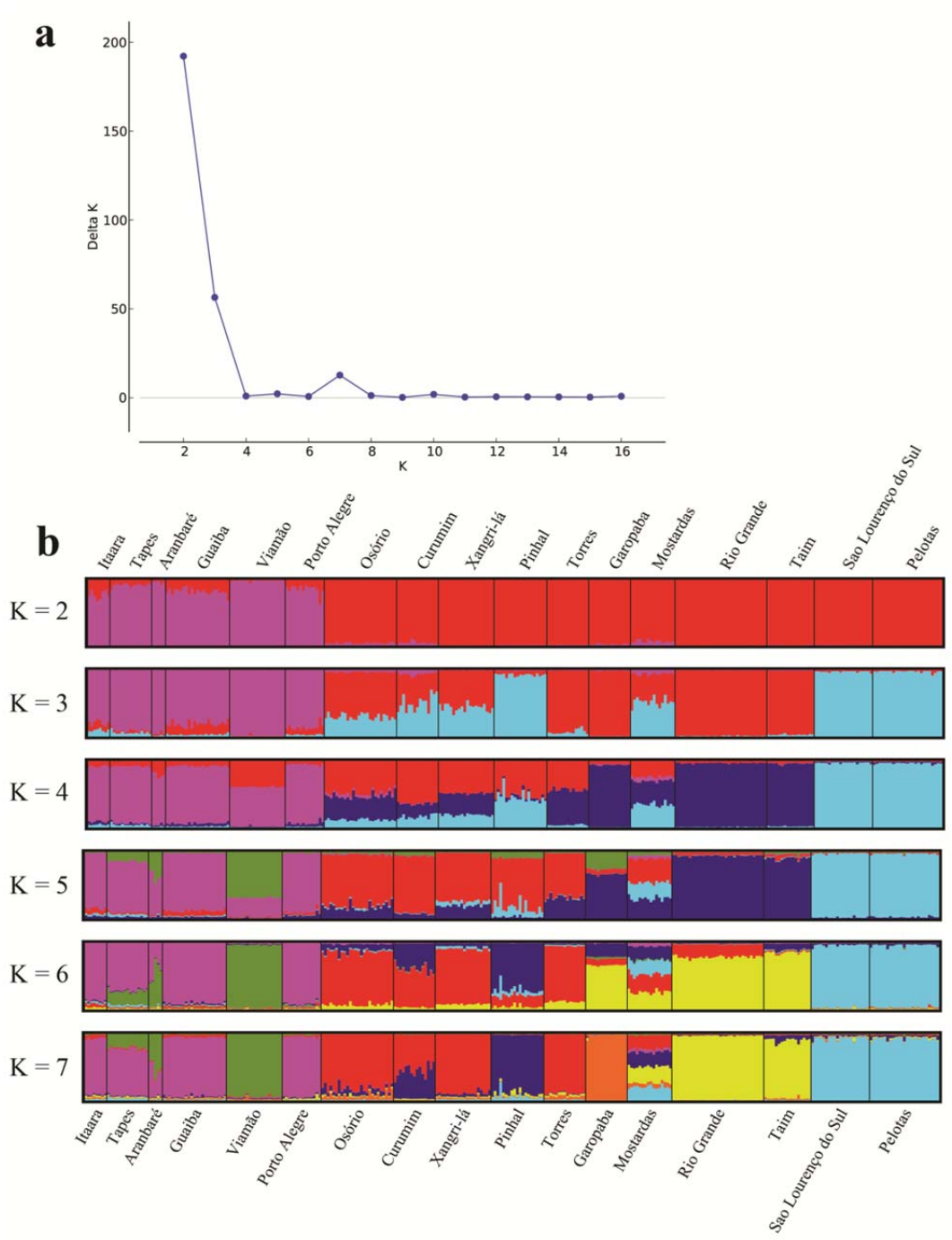


Figure S3. Barplots obtained with STRUCTURE analyses. Populations are separated by vertical tick black lines; and population names as indicated in Table 1 and Fig.1 are showed on top and bottom of barplots. (a) Plot of optimal K obtained with STRUCTURE HARVESTER. (b) Barplots for the complete data set with K=2 to K=7. Different colors indicate groups and vertical bars correspond to each individual.

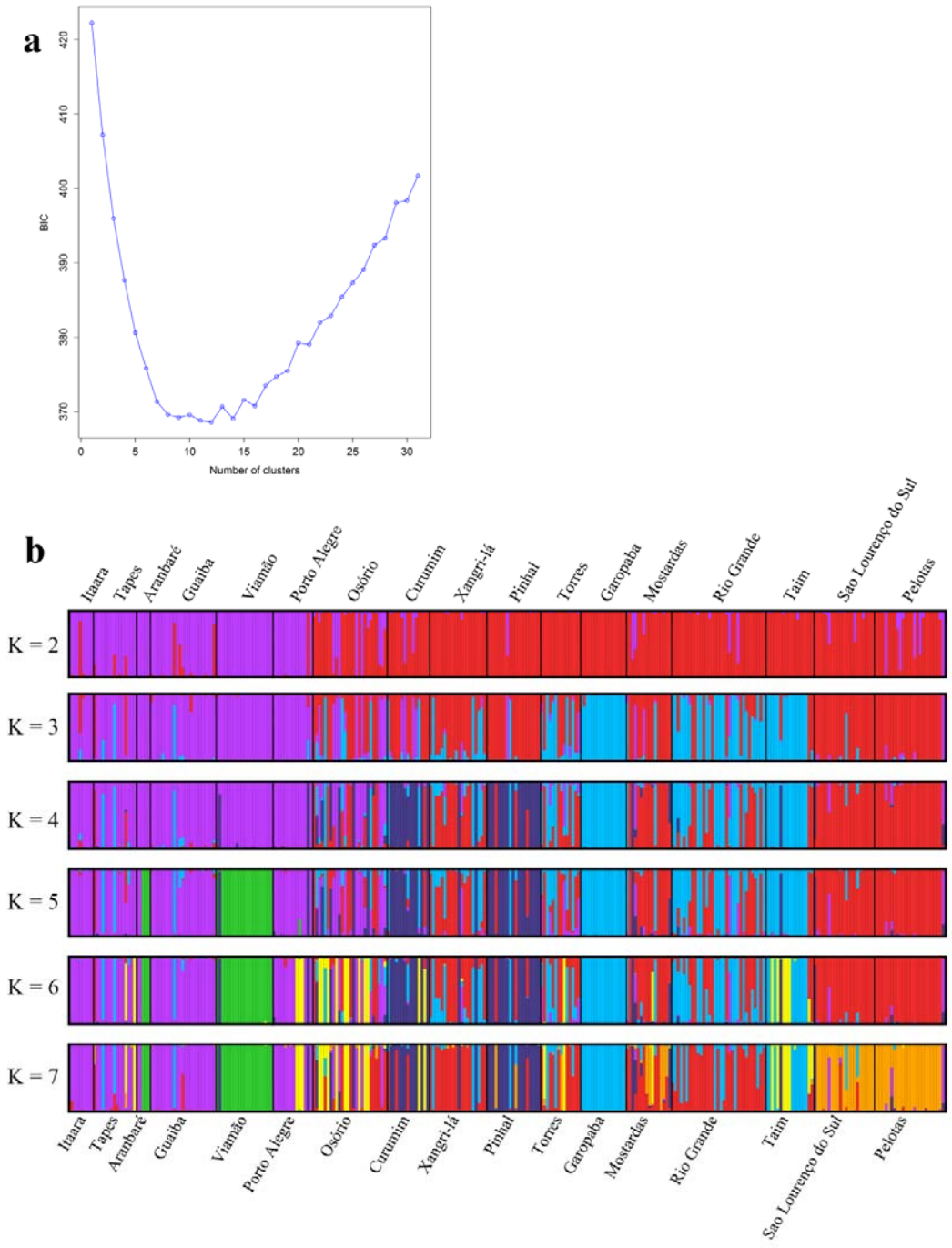


Figure S4. Barplots obtained with DAPC analyses. Populations are separated by vertical tick black lines; and population names as indicated in Table 1 and Fig.1 are showed on top and bottom of barplots. (a) Plot of optimal K obtained with *find.clusters* function of ADEGENET R package. (b) Barplot for the complete data set with K=2 to K=7. Different colors indicate groups and vertical bars correspond to each individual.

Dataset of climatic surfaces for the SACP

The new set of eight climatic variables generated in this study provided good predictability when compared with the observed data from the climate stations used to fit the models (Figs. S5A-S7A). Climatic values taken from the models fitted for mean annual temperature, annual precipitation, and mean diurnal range variables showed concordance with the values extracted from these same variables from the WorldClim set for same localities (Figs. S4A-S8A). Precipitation seasonality values showed higher discrepancy (Fig. S7A). Same comparisons were made between the climatic variables generated in this study and WorldClim set using the *P. depauperata* population localities. In this case, annual precipitation values showed high concordance (Fig. S5B); annual temperature showed a discordance of about 1°C (Fig. S6B); precipitation seasonality showed a discordance around 10% (Fig. S7B); and mean diurnal range showed higher variance for values extracted from the WorldClim variable (Fig. S8B).

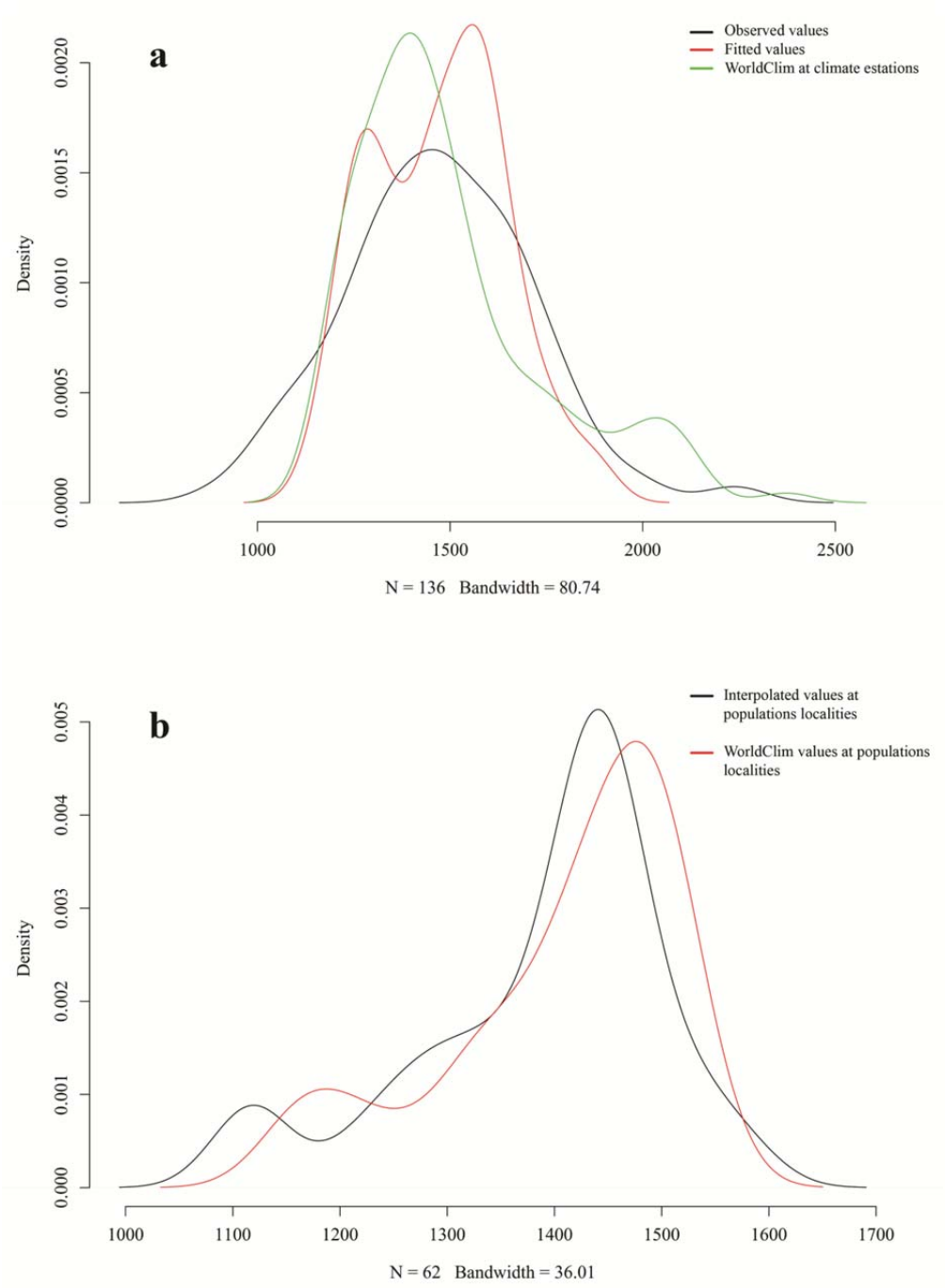


Figure S5. Density plots comparing annual precipitation values. **a**. Observed values at climate stations used for fit the Tps models, fitted values with Tps model for station localities and values extracted from the same variable of the WorldClim for station localities. **b**. Extracted from the interpolated Tps model for *P. depauperata* localities and extracted from the same variable of the WorldClim for *P. depauperata* localities

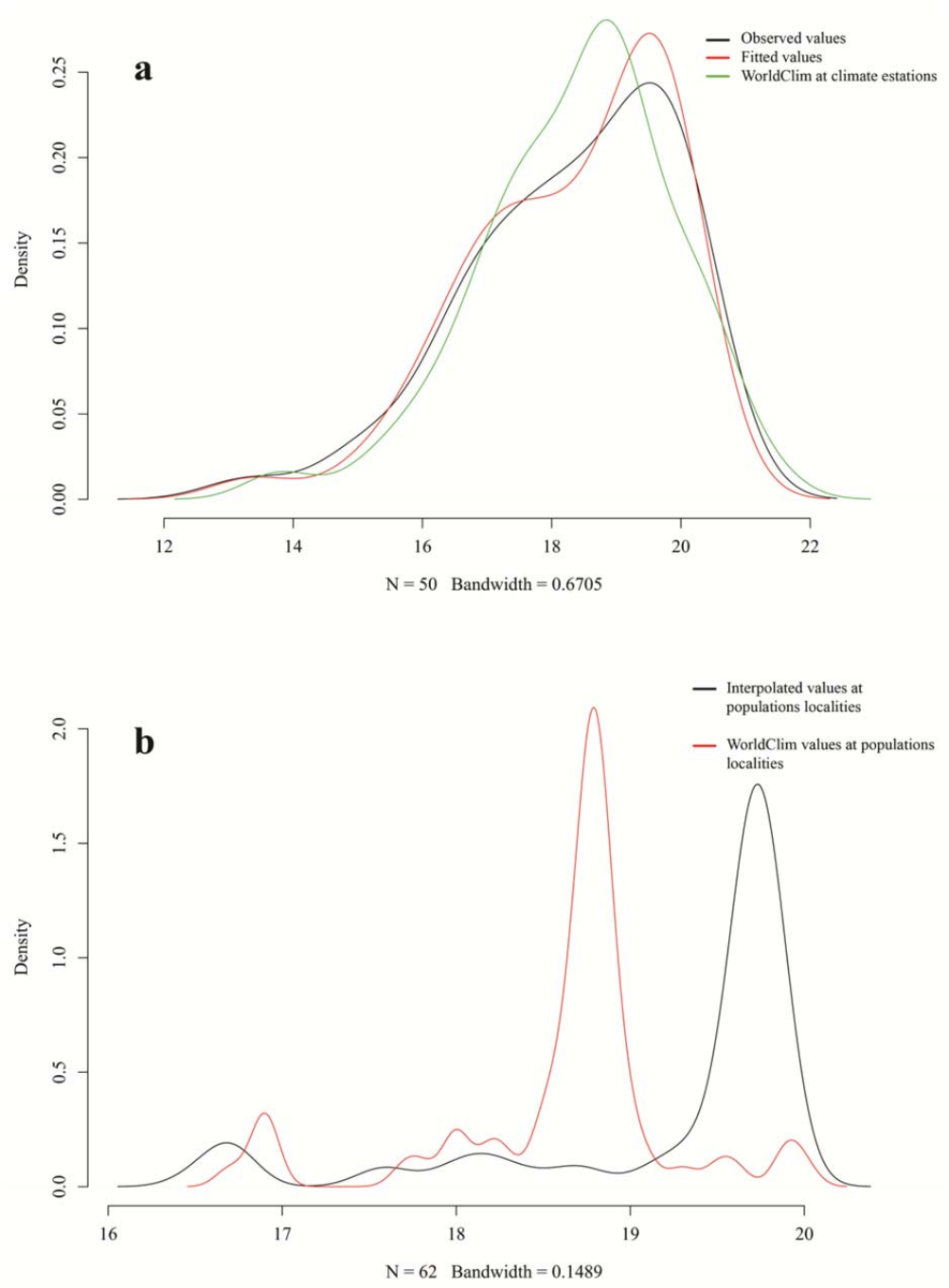


Figure S6. Density plots comparing mean annual temperature values. **a**. Observed values at climate stations used for fit the Tps models, fitted values with Tps model for station localities, and values extracted from the same variable of the WorldClim from station localities. **b**. Values extracted from the climatic surfaces generated with the Tps model for *P. depauperata* populations and extracted from the same variable of the WorldClim dataset for *P. depauperata* populations

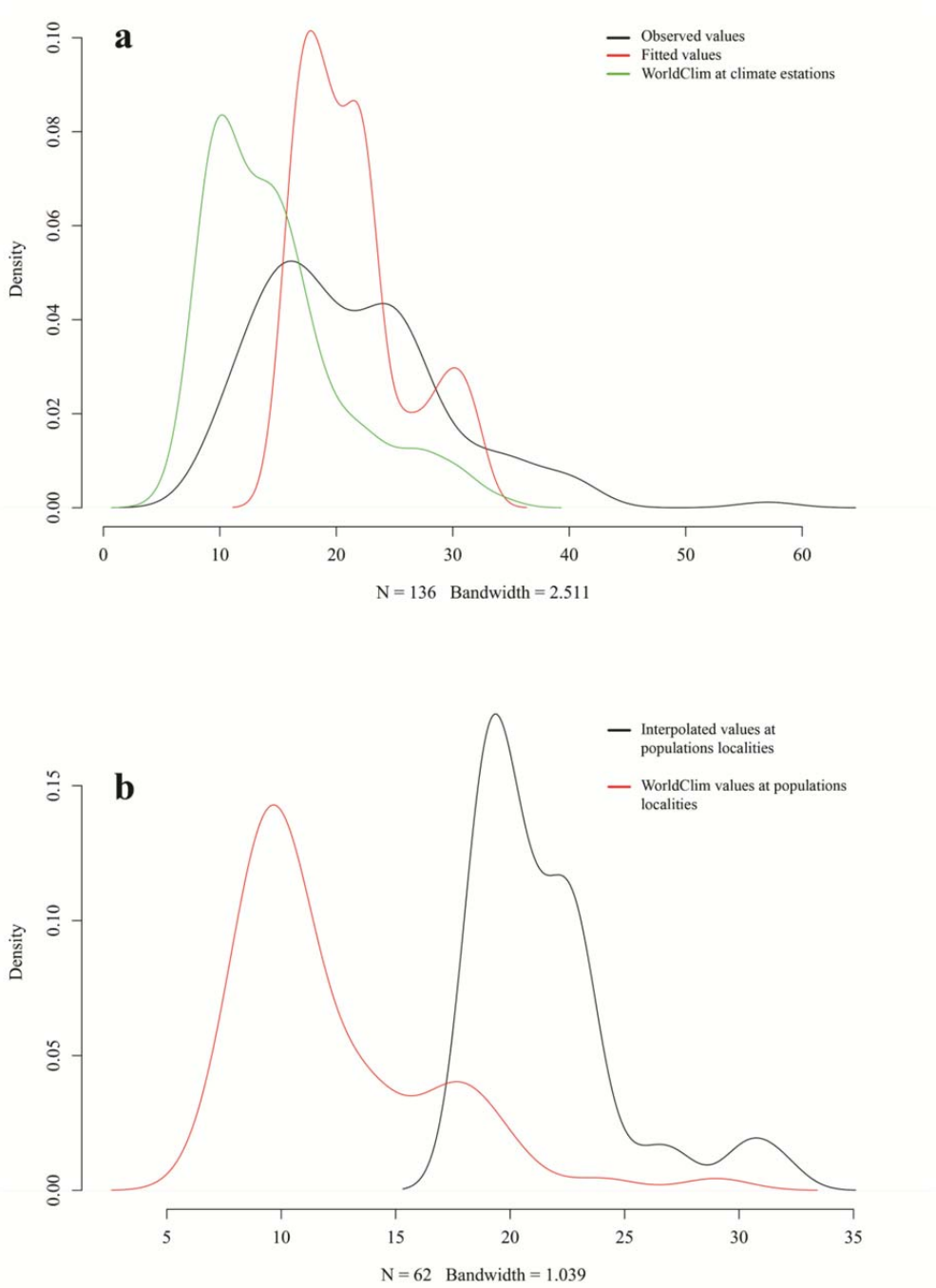


Figure S7. Density plots comparing precipitation seasonality values. **a**. Observed values at climate stations used for fit the Tps models, fitted values with Tps model for station localities, and values extracted from the same variable of the WorldClim from station localities. **b**. Values extracted from the climatic surfaces generated with the Tps model for *P. depauperata* populations and extracted from the same variable of the WorldClim dataset for *P. depauperata* populations.

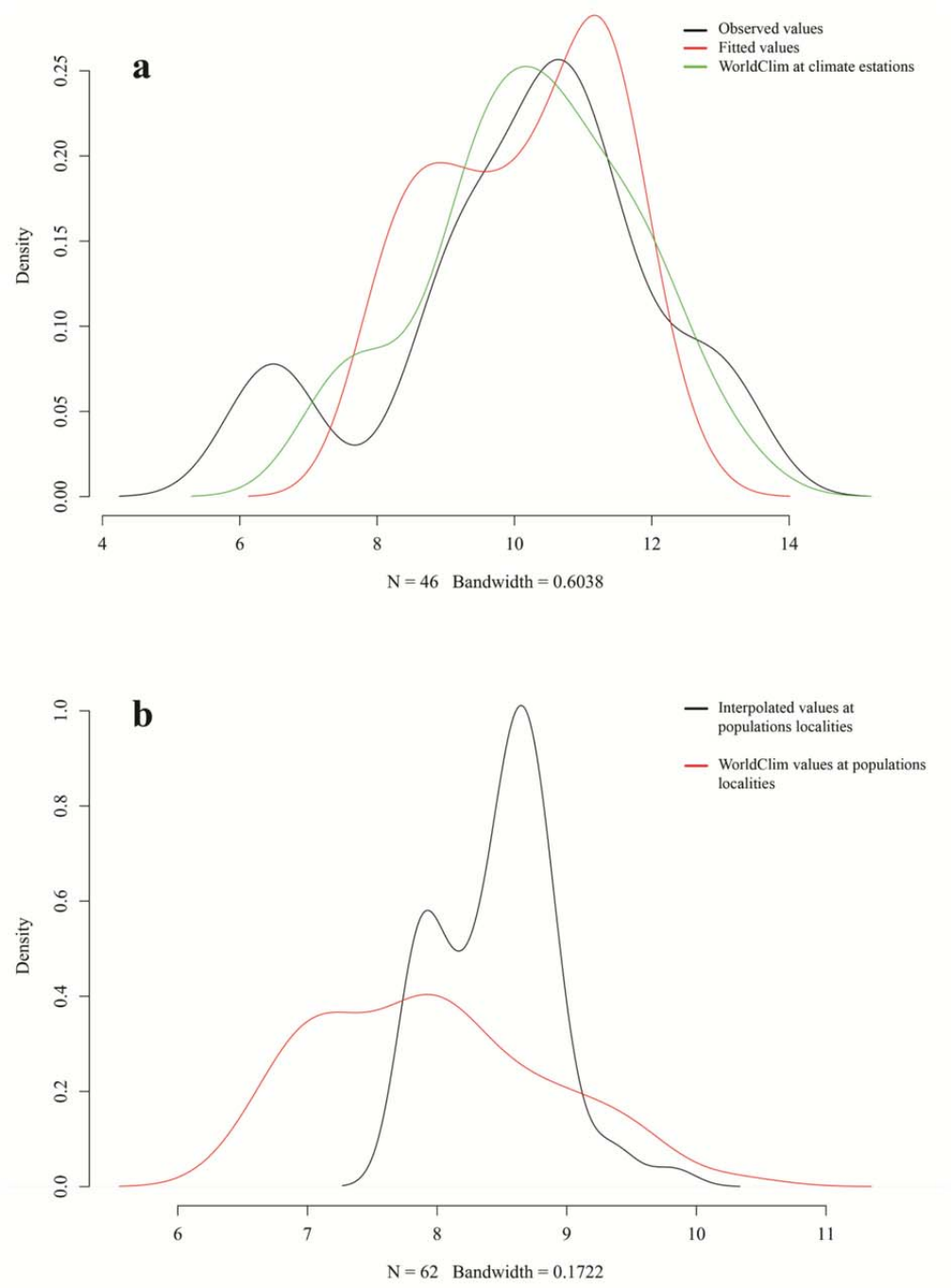


Figure S8. Density plots comparing mean diurnal range values. **a**. Observed values at climate stations used for fit the Tps models, fitted values with Tps model for station localities and values extracted from the same variable of the WorldClim from station localities. **b**. Values extracted from the climatic surfaces generated with the Tps model for *P. depauperata* populations and extracted from the same variable of the WorldClim dataset for *P. depauperata* populations.

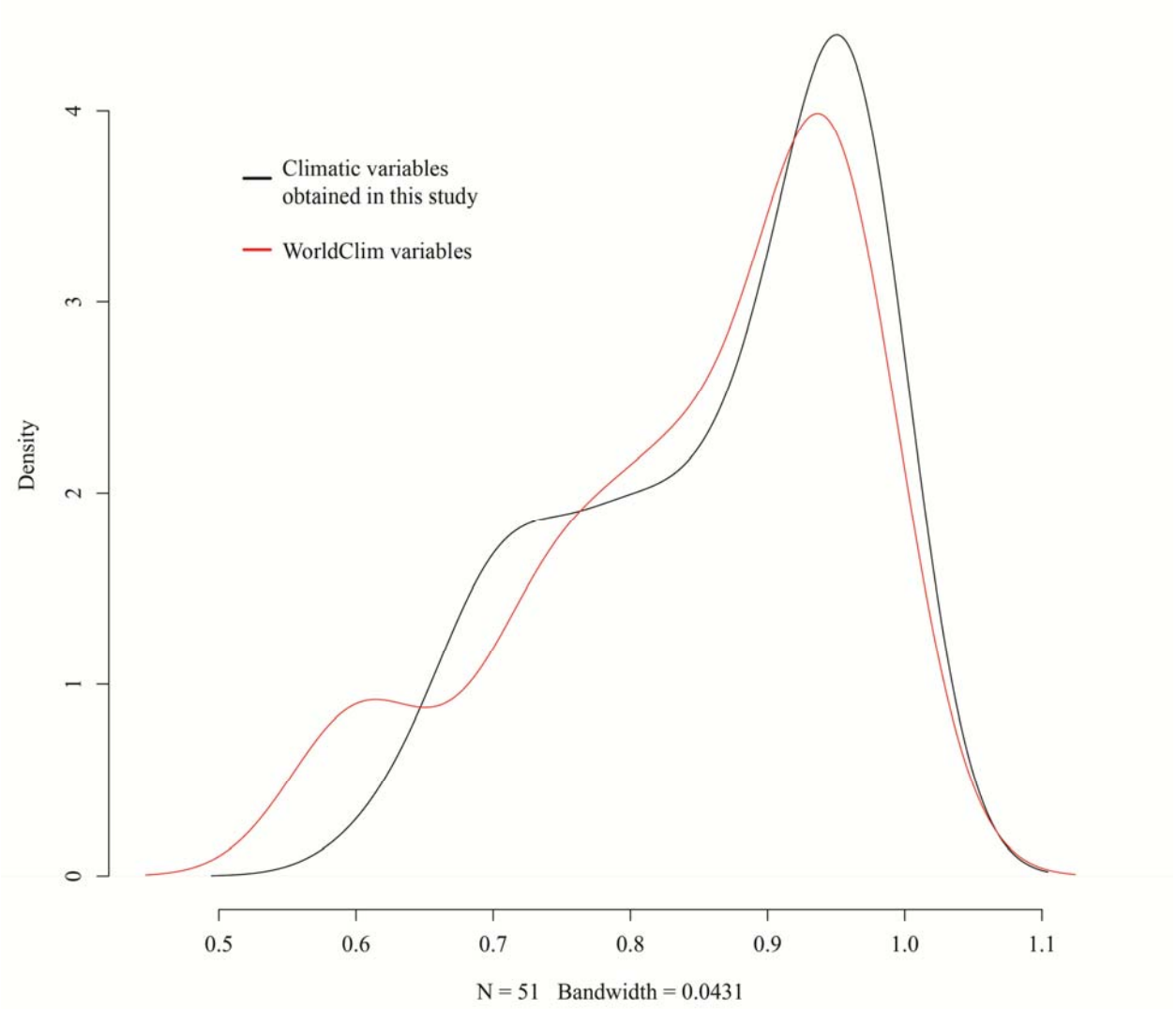


Figure S9. Density plots comparing suitability values of *P. depauperata* localities used for train the models (Table S1) using the climate variables obtained in this study and the WorldClim set.

Table S5. Maximum likelihood population effects (MLPE) modeling results investigating the relationships of F_{ST}, D_{PS} and Nei's genetic differentiation statistics with the eight climatic variables obtained in this study and Euclidean or topographic spatial distances.

Genetic distance	Variables	ΔAICc	weight	ΔBIC	weight
F_{ST}	Precipitation seasonality	0	1.000	0	1.000
	Summer mean maximum temperature	16.478	<0.001	16.478	<0.001
	Mean temperature range	21.197	<0.001	21.197	<0.001
	TOPO	23.603	<0.001	23.603	<0.001
	Total annual days with rain	26.143	<0.001	26.143	<0.001
	EUC	26.178	<0.001	26.178	<0.001
	Total annual precipitation	26.409	<0.001	26.409	<0.001
	Winter mean minimum temperature	26.795	<0.001	26.795	<0.001
	Mean annual temperature	27.703	<0.001	27.703	<0.001
	Temperature seasonality	28.378	<0.001	28.378	<0.001
D_{PS}	Precipitation seasonality	0	0.841	0	0.841
	Summer mean maximum temperature	3.678	0.134	3.678	0.134
	Mean temperature range	7.033	0.025	7.033	0.025
	EUC	20.951	<0.001	20.951	<0.001
	Total annual precipitation	21.151	<0.001	21.151	<0.001
	Winter mean minimum temperature	21.886	<0.001	21.886	<0.001
	Total annual days with rain	22.317	<0.001	22.317	<0.001
	TOPO	22.849	<0.001	22.849	<0.001
	Mean annual temperature	22.964	<0.001	22.964	<0.001
	Temperature seasonality	25.799	<0.001	25.799	<0.001
Nei's	Precipitation seasonality	0	0.998	0	0.998
	Summer mean maximum temperature	12.342	0.002	12.342	0.002
	Mean temperature range	17.165	<0.001	17.165	<0.001
	TOPO	35.937	<0.001	35.937	<0.001
	EUC	37.75	<0.001	37.747	<0.001
	Total annual days with rain	37.979	<0.001	37.979	<0.001
	Total annual precipitation	38.705	<0.001	38.705	<0.001
	Winter mean minimum temperature	38.956	<0.001	38.956	<0.001
	Temperature seasonality	39.52	<0.001	39.517	<0.001
	Mean annual temperature	39.905	<0.001	39.905	<0.001

EUC: Euclidean geographic distances; TOPO: Geographic distances accounting topographic features of study region.

Isolation by environment tests using WorldClim data set

The correlation between WorldClim variables and genetic distances showed a less consistent pattern and the correlation values were mostly marginally significant (Supplementary Tables S4 and S5). None of the WorldClim variables showed a consistent correlation with the three genetic distances after controlling by spatial distances (EUC and TOPO). F_{ST} distances were correlated with four WorldClim variables (Bio10, Bio13, Bio16, and Bio19) after controlling by EUC and TOPO distances, and Bio19 (Precipitation of Coldest Quarter) showed the higher correlation. Bio1 (Annual Mean Temperature) was the only significant variable correlated with D_{PS} after controlling by EUC and TOPO distances. Bio5 (Max Temperature of Warmest Month) was the only significant variable correlated with Nei's distance but only after controlling by TOPO distance (Table S5). MLPE analyses showed that the best-supported models including the variables Bio3 (Isothermality), Bio5, or Bio19 alone, as the variables that best explain the inter-population genetic differentiation (Tables S6 and S7).

Table S6. Relationships tests between three population genetic differentiation statistics (F_{ST}, D_{PS}, and Nei's) and inter-population distances measured for each climatic variable obtained from WorldClim calculated with Mantel and partial Mantel tests controlling by spatial distances. Variables names can be found in http://www.worldclim.org/bioclim

Genetic distance	Climatic variable	Mantel R^2	Partial Mantel tests	
			Controlling by EUC	Controlling by TOPO
F_{ST}	Bio01	0.42*	0.23	0.34
	Bio02	0.21*	0.04	0.18
	Bio03	0.11	-	-
	Bio04	0.11	-	-
	Bio05	0.001	-	-
	Bio06	0.35*	0.06	0.26
	Bio07	0.13	-	-
	Bio08	0.53*	0.40	0.47*
	Bio09	0.29	-	-
	Bio10	0.39*	0.30*	0.37*
	Bio11	0.36*	0.08	0.26
	Bio12	0.01	-	-
	Bio13	0.58*	0.48*	0.54*
	Bio14	0.04	-	-
	Bio15	0.47*	0.32	0.41
	Bio16	0.59*	0.49*	0.55*
	Bio17	0.16*	-0.11	0.09
	Bio18	0.56*	0.44	0.51*
	Bio19	0.60**	0.52*	0.56*
D_{PS}	Bio01	0.40*	0.28*	0.40*
	Bio02	0.23*	0.10	0.21*
	Bio03	0.30*	0.26	0.29*
	Bio04	-0.04	-	-
	Bio05	0.20*	0.13	0.20
	Bio06	0.17	-	-
	Bio07	0.03	-	-
	Bio08	0.29	-	-
	Bio09	0.21	-	-
	Bio10	0.28	-	-
	Bio11	0.21	-	-
	Bio12	-0.06	-	-
	Bio13	0.35	-	-

Genetic distance	Climatic variable	Mantel R^2	Partial Mantel tests	
			Controlling by EUC	Controlling by TOPO
	Bio14	-0.01	-	-
	Bio15	0.28	-	-
	Bio16	0.35	-	-
	Bio17	0.05	-	-
	Bio18	0.31	-	-
	Bio19	0.34	-	-
Nei's	Bio01	0.34	-	-
	Bio02	0.20*	0.12	0.19
	Bio03	0.23	-	-
	Bio04	-0.10	-	-
	Bio05	0.24*	0.19	0.23*
	Bio06	0.07	-	-
	Bio07	-0.01	-	-
	Bio08	0.20	-	-
	Bio09	0.18	-	-
	Bio10	0.25	-	-
	Bio11	0.11	-	-
	Bio12	-0.08	-	-
	Bio13	0.25	-	-
	Bio14	-0.07	-	-
	Bio15	0.15	-	-
	Bio16	0.24	-	-
	Bio17	-0.02	-	-
	Bio18	0.20	-	-
	Bio19	0.24	-	-

The Mantel's statistic r is reported and the significance level is indicated. *** P < 0.001; ** P < 0.005; * P < 0.05. EUC: Euclidean geographic distance; TOPO: Geographic distance accounting topographic features of studied region.

Table S7. Maximum likelihood population effects (MLPE) modeling results investigating the relationships of F_{ST}, D_{PS}, and Nei's genetic differentiation statistics with climatic variables from WorldClim and Euclidean (EUC) or topographic (TOPO)spatial distances. Variables names can be found in http://www.worldclim.org/bioclim

Genetic distance	Variables	ΔAICc	weight	ΔBIC	weight
F_{ST}	Bio19	0	0.375	0	0.375
	Bio5	0.783	0.253	0.783	0.253
	Bio3	2.652	0.099	2.652	0.099
	Bio10	3.921	0.053	3.921	0.053
	Bio18	4.066	0.049	4.066	0.049
	Bio16	4.316	0.043	4.316	0.043
	Bio13	4.454	0.040	4.454	0.040
	Bio1	6.310	0.016	6.310	0.016
	Bio17	6.557	0.014	6.557	0.014
	TOPO	6.867	0.012	6.867	0.012
	Bio9	6.984	0.011	6.984	0.011
	Bio15	7.535	0.009	7.535	0.009
	Bio2	7.650	0.008	7.650	0.008
	Bio14	8.688	0.005	8.688	0.005
	EUC	9.442	0.003	9.442	0.003
	Bio11	10.351	0.002	10.351	0.002
	Bio12	11.061	0.001	11.061	0.001
	Bio6	11.077	0.001	11.077	0.001
	Bio8	11.264	0.001	11.264	0.001
	Bio7	11.431	0.001	11.431	0.001
	Bio4	12.705	0.001	12.705	0.001
D_{PS}	Bio3	0	0.578	0	0.578
	Bio5	0.841	0.379	0.841	0.379
	Bio1	5.625	0.035	5.625	0.035
	Bio9	11.826	0.002	11.826	0.002
	Bio10	11.875	0.002	11.875	0.002
	EUC	13.147	0.001	13.147	0.001
	Bio2	13.341	0.001	13.341	0.001
	Bio11	13.970	0.001	13.970	0.001
	Bio19	14.086	0.001	14.086	0.001
	Bio15	14.093	0.001	14.093	0.001
	Bio17	14.127	<0.001	14.127	<0.001
	Bio18	14.783	<0.001	14.783	<0.001
	Bio14	15.041	<0.001	15.041	<0.001
	TOPO	15.045	<0.001	15.045	<0.001

Genetic distance	Variables	ΔAICc	weight	ΔBIC	weight
	Bio6	16.479	<0.001	16.479	<0.001
	Bio16	16.734	<0.001	16.734	<0.001
	Bio13	17.016	<0.001	17.016	<0.001
	Bio12	18.725	<0.001	18.725	<0.001
	Bio7	18.791	<0.001	18.791	<0.001
	Bio8	18.922	<0.001	18.922	<0.001
	Bio4	19.305	<0.001	19.305	<0.001
Nei's	Bio5	0	0.936	0	0.936
	Bio3	5.687	0.055	5.687	0.055
	Bio9	10.919	0.004	10.919	0.004
	Bio2	11.625	0.003	11.625	0.003
	Bio10	14.680	0.001	14.680	0.001
	Bio1	15.268	<0.001	15.268	<0.001
	Bio19	15.440	<0.001	15.440	<0.001
	Bio17	17.020	<0.001	17.020	<0.001
	TOPO	17.046	<0.001	17.046	<0.001
	Bio18	17.312	<0.001	17.312	<0.001
	Bio14	18.230	<0.001	18.230	<0.001
	Bio15	18.435	<0.001	18.435	<0.001
	Bio13	18.715	<0.001	18.715	<0.001
	EUC	18.856	<0.001	18.856	<0.001
	Bio7	19.332	<0.001	19.332	<0.001
	Bio16	19.463	<0.001	19.463	<0.001
	Bio11	20.273	<0.001	20.273	<0.001
	Bio6	20.279	<0.001	20.279	<0.001
	Bio8	20.588	<0.001	20.588	<0.001
	Bio4	20.754	<0.001	20.754	<0.001
	Bio12	21.187	<0.001	21.187	<0.001

Table S8. Maximum likelihood population effects (MLPE) modeling results investigating the relationships of F_{ST}, D_{PS}, and Nei's genetic differentiation statistics with single and mixed three climatic variables from WorldClim and Euclidean (EUC) or topographic (TOPO) spatial distances.

Genetic distance	Variables	ΔAICc	weight	ΔBIC	weight
	Bio19	0	0.506	0	0.513
	Bio5	0.783	0.342	0.783	0.347
	Bio3	2.652	0.134	2.652	0.136
	Bio19 + Bio3 + Bio5	7.646	0.011	12.122	0.001
	Bio3+ Bio5	9.742	0.004	12.340	0.001
F_{ST}	Bio5 + TOPO	10.831	0.002	13.211	0.001
	Bio5 + EUC	12.483	0.001	14.863	<0.001
	Bio19 + Bio3 + Bio5 + TOPO	18.783	<0.001	25.726	<0.001
	Bio19 + Bio3 + Bio5 + EUC	19.487	<0.001	26.430	<0.001
	Bio3+ Bio5 + TOPO	19.568	<0.001	24.263	<0.001
	Bio3+ Bio5 + EUC	20.986	<0.001	25.680	<0.001
	Bio3	0	0.496	0	0.575
	Bio5	0.841	0.326	0.841	0.377
	Bio3+ Bio5	3.050	0.108	5.430	0.038
	Bio19 + Bio3 + Bio5	4.402	0.055	9.097	0.006
	Bio5 + EUC	9.554	0.004	11.934	0.001
D_{PS}	Bio5 + TOPO	9.587	0.004	11.967	0.001
	Bio3+ Bio5 + EUC	9.934	0.003	14.629	<0.001
	Bio3+ Bio5 + TOPO	11.526	0.002	16.221	<0.001
	Bio19 + Bio3 + Bio5 + EUC	12.748	0.001	19.691	<0.001
	Bio19 + Bio3 + Bio5 + TOPO	13.936	<0.001	20.879	<0.001
	Bio19	14.086	<0.001	14.086	0.001
	Bio5	0	0.818	0	0.912
	Bio3+ Bio5	5.018	0.067	7.398	0.023
	Bio3	5.687	0.048	5.687	0.053
	Bio19 + Bio3 + Bio5	5.761	0.046	10.456	0.005
	Bio5 + TOPO	8.510	0.012	10.890	0.004
Nei	Bio5 + EUC	9.393	0.007	11.773	0.003
	Bio3+ Bio5 + TOPO	13.439	0.001	18.134	<0.001
	Bio3+ Bio5 + EUC	14.310	0.001	19.004	<0.001
	Bio19 + Bio3 + Bio5 + TOPO	14.731	0.001	21.673	<0.001
	Bio19 + Bio3 + Bio5 + EUC	15.281	<0.001	22.224	<0.001
	Bio19	15.440	<0.001	15.440	<0.001

Niche modeling results

The projected ensemble niche model for *P. depauperata* for current climate conditions achieved with the climatic variables obtained in this study is shown in Figure S10A. The ensemble distribution model summarizes 180 models with TSS ≥ 0.7. All GLM, GBM, CTA, FDA, MARS, RF, and Maxent models presented TSS ≥ 0.7 and were included in the final ensemble model (20 models each one). With ANN were obtained 16 models, GAM 13 models, and SRE 11 models with TSS ≥ 0.7. The ensemble model had a TSS = 0.97 and AUC = 0.99. In average, the most important variables for models obtained were mean diurnal range, total annual precipitation, and mean annual temperature.

The ensemble model obtained with the WorldClim variables is shown in Figure S10B. The ensemble distribution model summarizes 152 models with TSS ≥ 0.7. All GLM, GBM, and RF models presented TSS ≥ 0.7 and were included in the final ensemble model (20 models each one). CTA and FDA presented 19 models, Maxent 18 models, ANN 17 models, MARS 15 models, SRE 2 models, and GAM 1 model with TSS ≥ 0.7. The ensemble model had a TSS = 0.98 and AUC = 0.99. In average, the most important variables for models obtained were mean diurnal range, max temperature of warmest month, and temperature annual range.

The obtained niche models for current climatic conditions showed higher suitability areas close to the coastal line with a continuous decreasing to the west. The model obtained with the WorldClim variables showed the highest suitability values in the Central-North region of the SACP limited to the west by altitude (Fig. S10). Interestingly both niche models showed very low suitability values for the region of Porto Alegre and fossil dunes region (Fig. S10A-B).

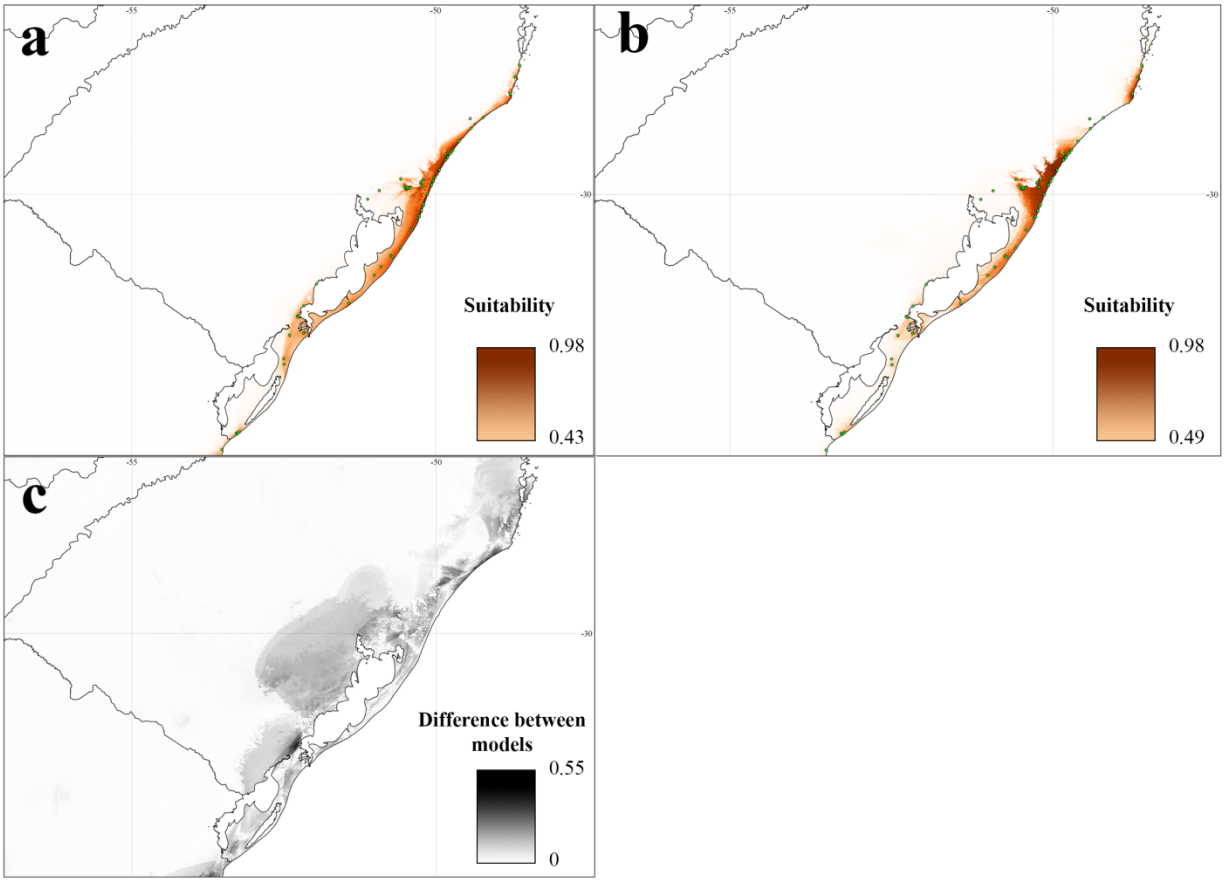
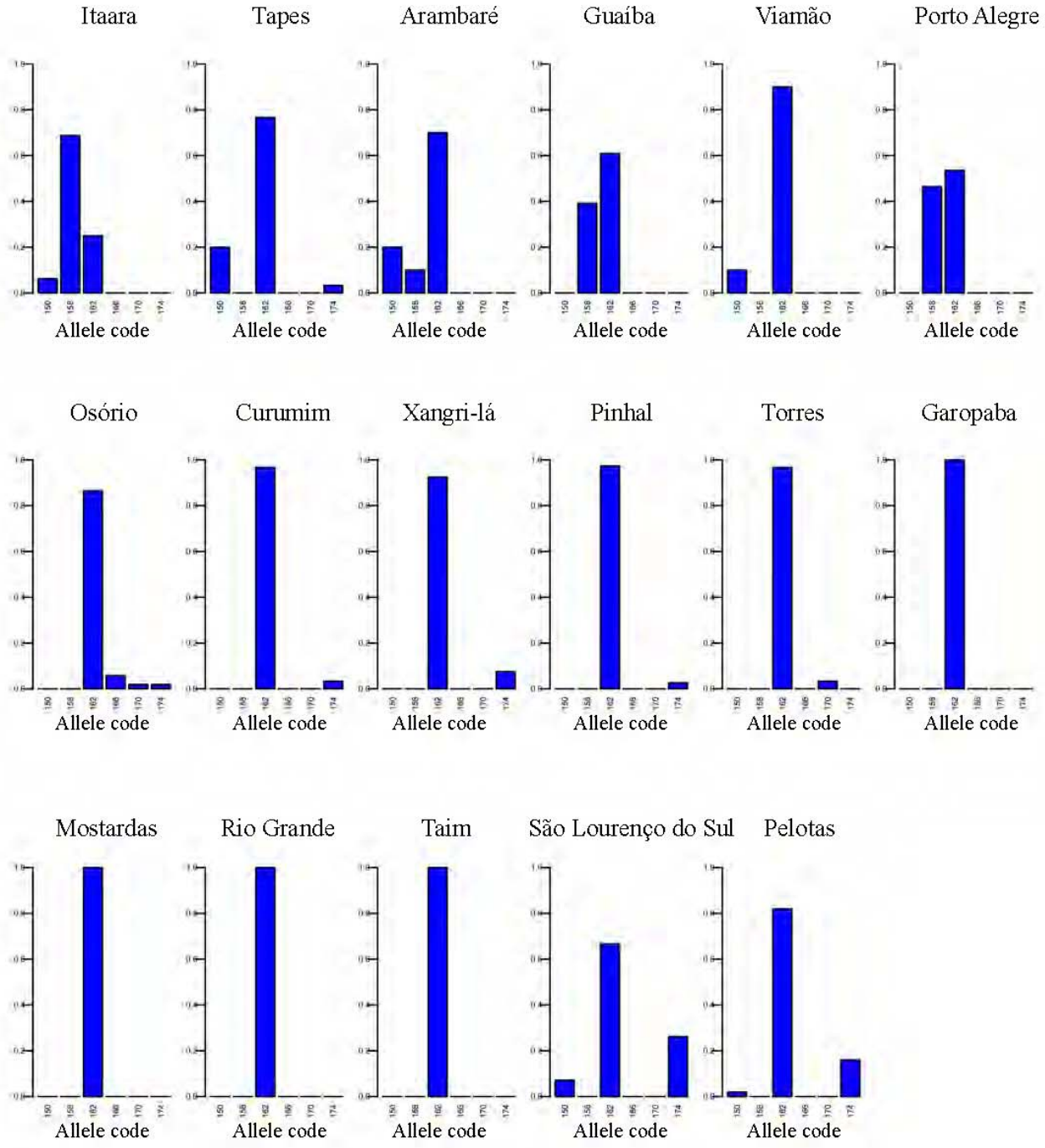


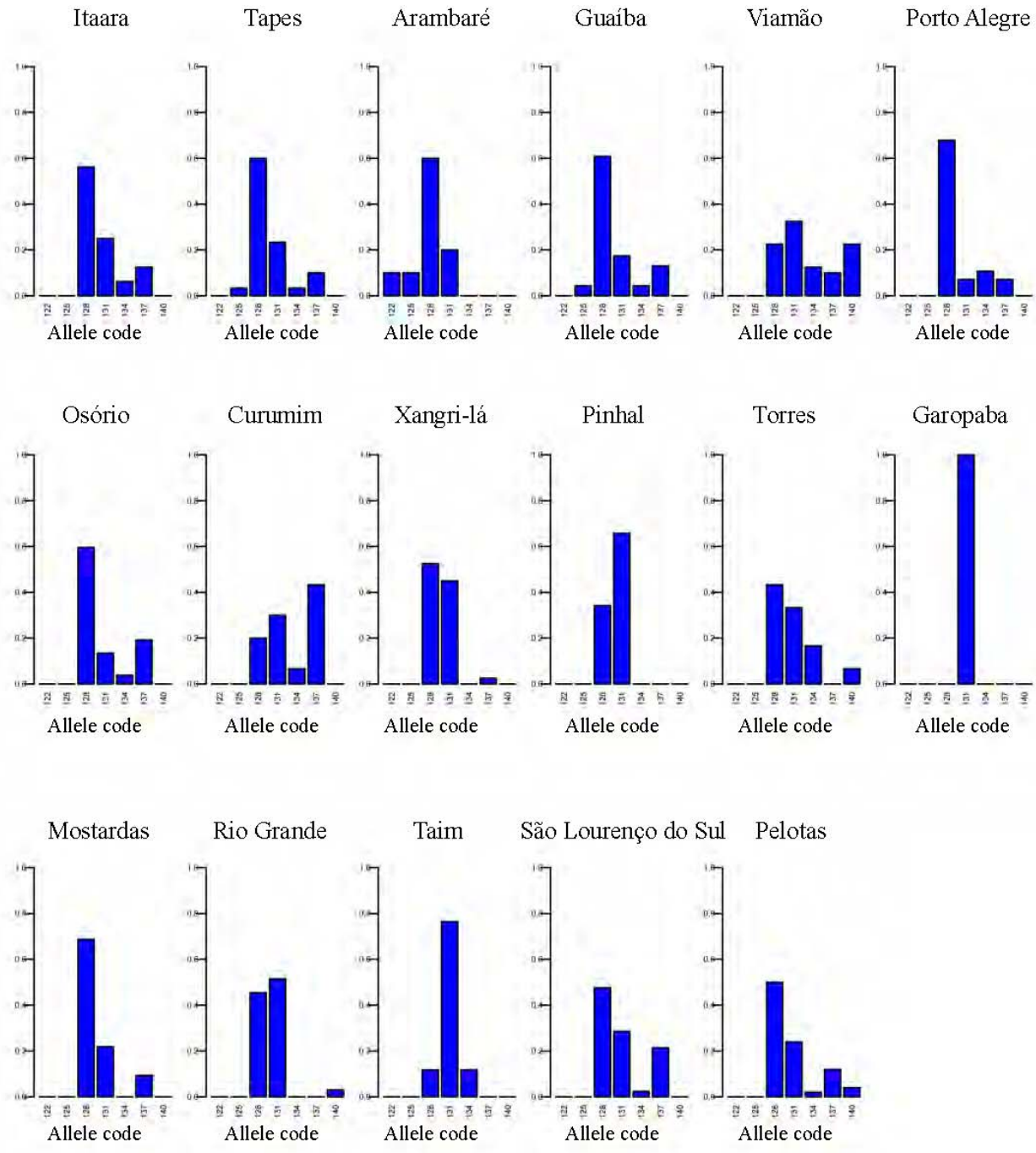
Figure S10. Ensemble niche models projections for *Petunia depauperata* achieved with climate variables obtained in this study (**a**), and WorldClim variables (**b**). Green dots represent the localities used to train the models (Table S1). Differences of suitability values between models (**c**).

APPENDIX S3. Allelic frequency per population

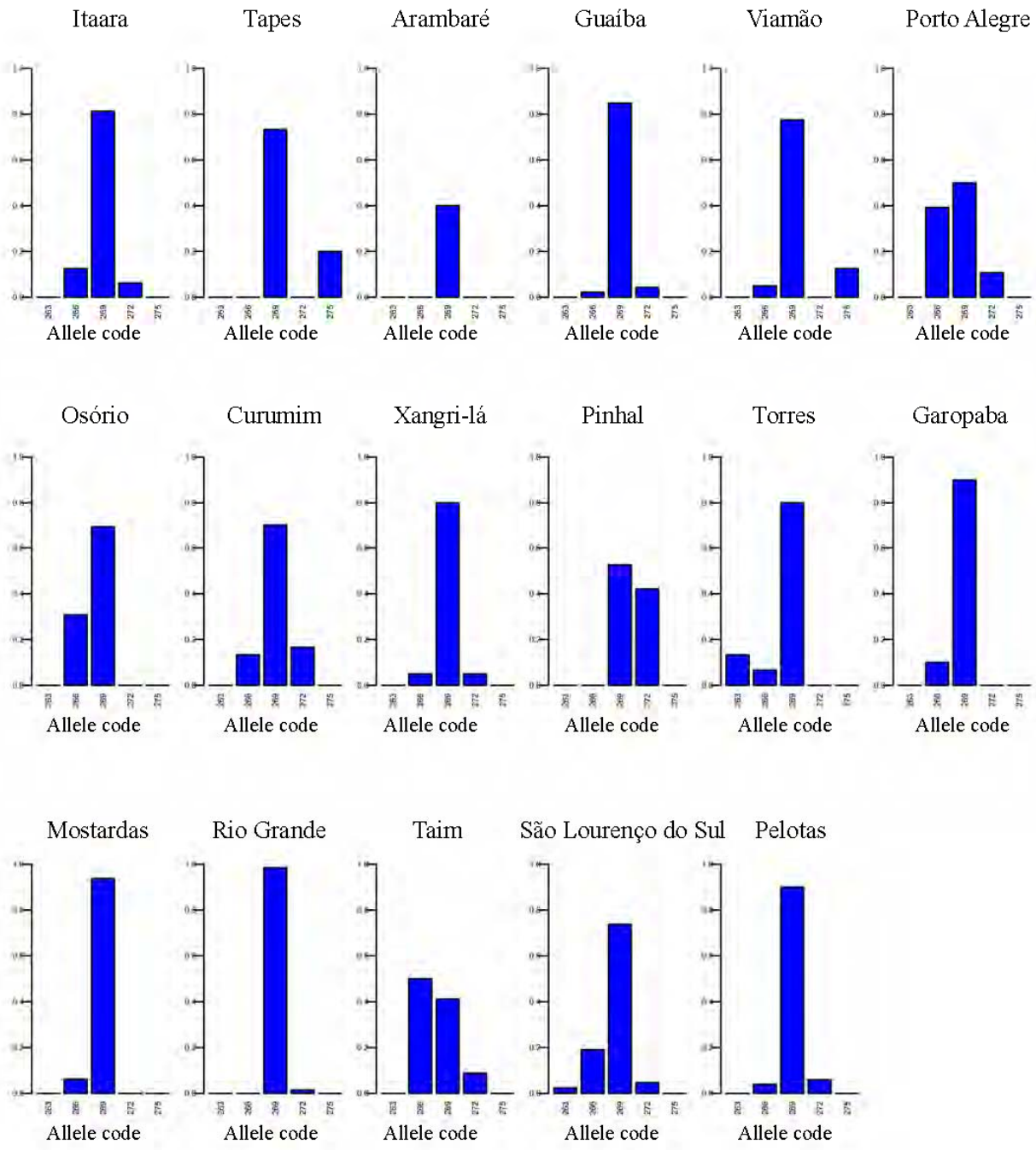
Locus PM8



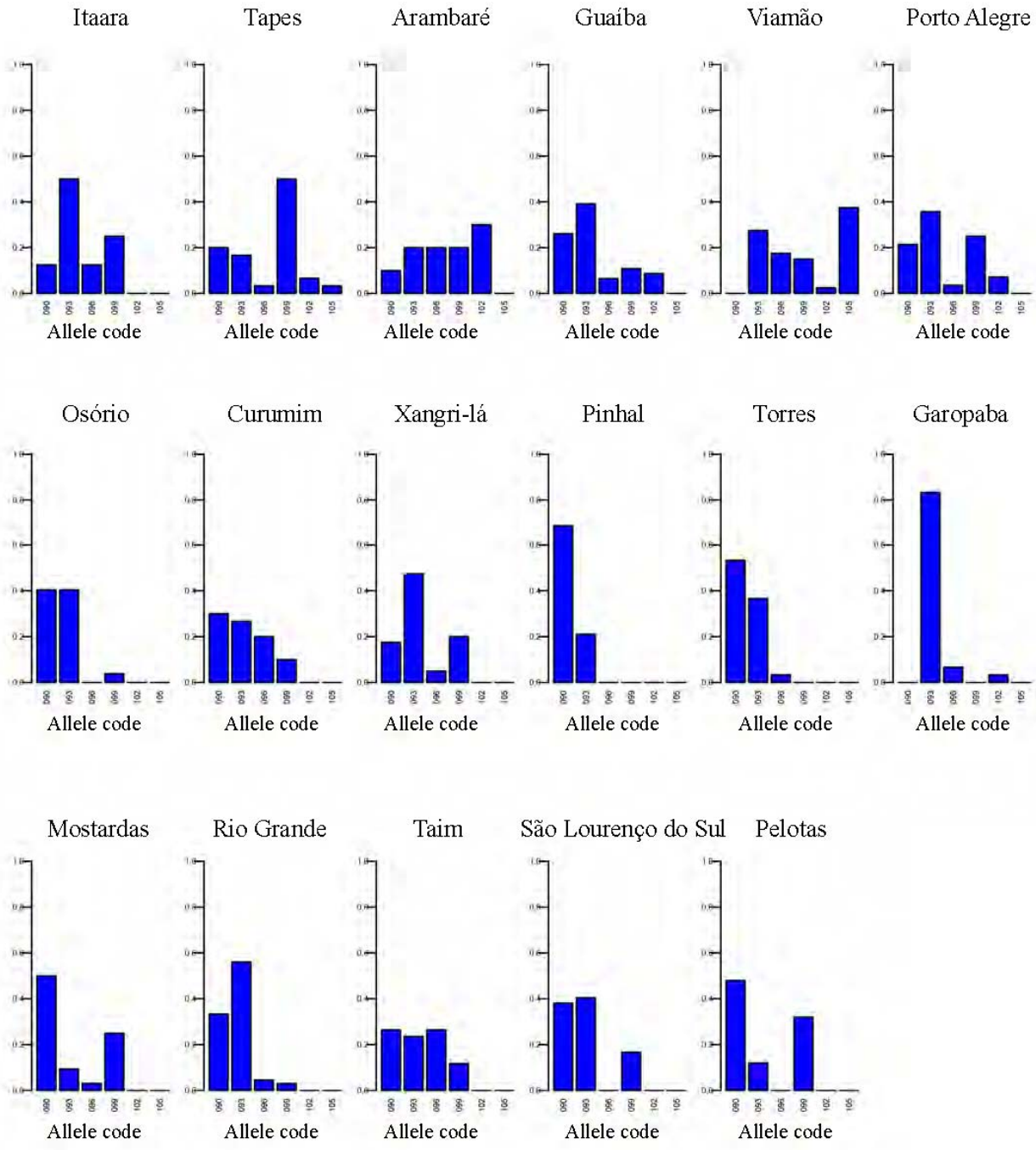
Locus PM21


Itaara
Tapes
Arambaré
Guaíba
Viamão
Porto Alegre
Osório
Curumim
Xangri-lá
Pinhal
Torres
Garopaba
Mostardas
Rio Grande
Taim
São Lourenço do Sul
Pelotas
Allele code

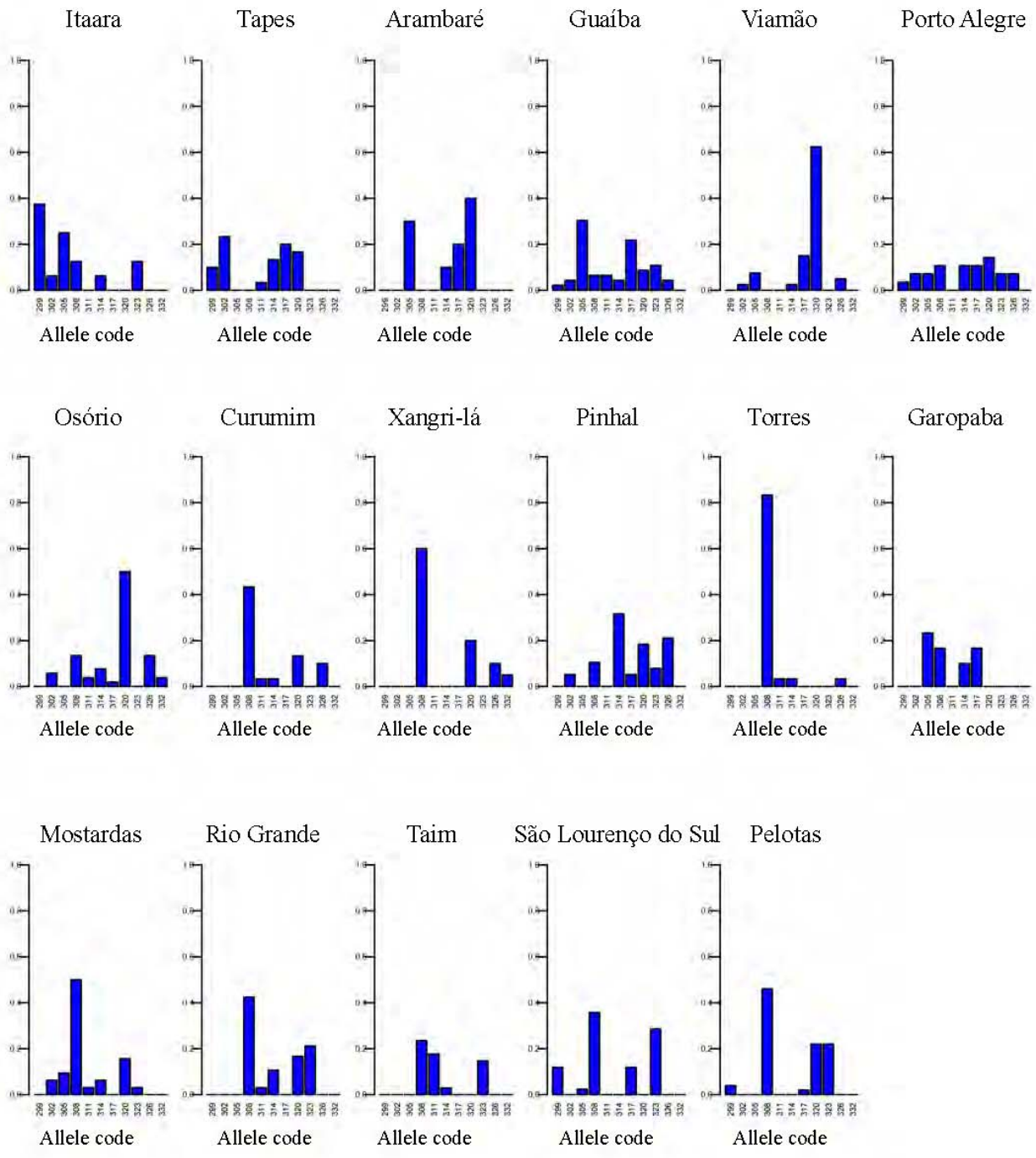
Locus PM101


Itaara
Tapes
Arambaré
Guaíba
Viamão
Porto Alegre
Osório
Curumim
Xangri-lá
Pinhal
Torres
Garopaba
Mostardas
Rio Grande
Taim
São Lourenço do Sul
Pelotas
Allele code

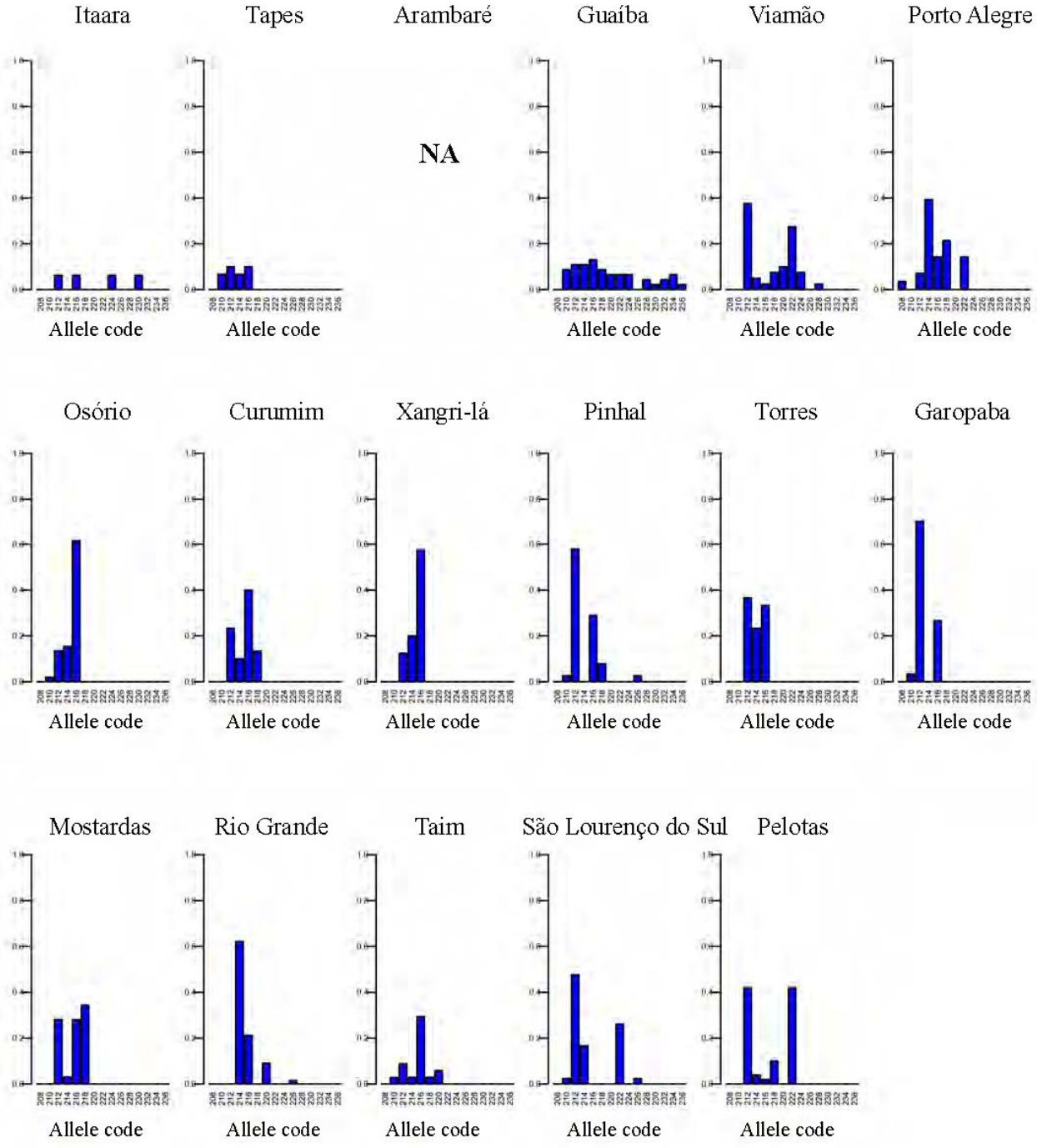
Locus PM110


Itaara
Tapes
Arambaré
Guaíba
Viamão
Porto Alegre
Osório
Curumim
Xangri-lá
Pinhal
Torres
Garopaba
Mostardas
Rio Grande
Taim
São Lourenço do Sul
Pelotas
Allele code

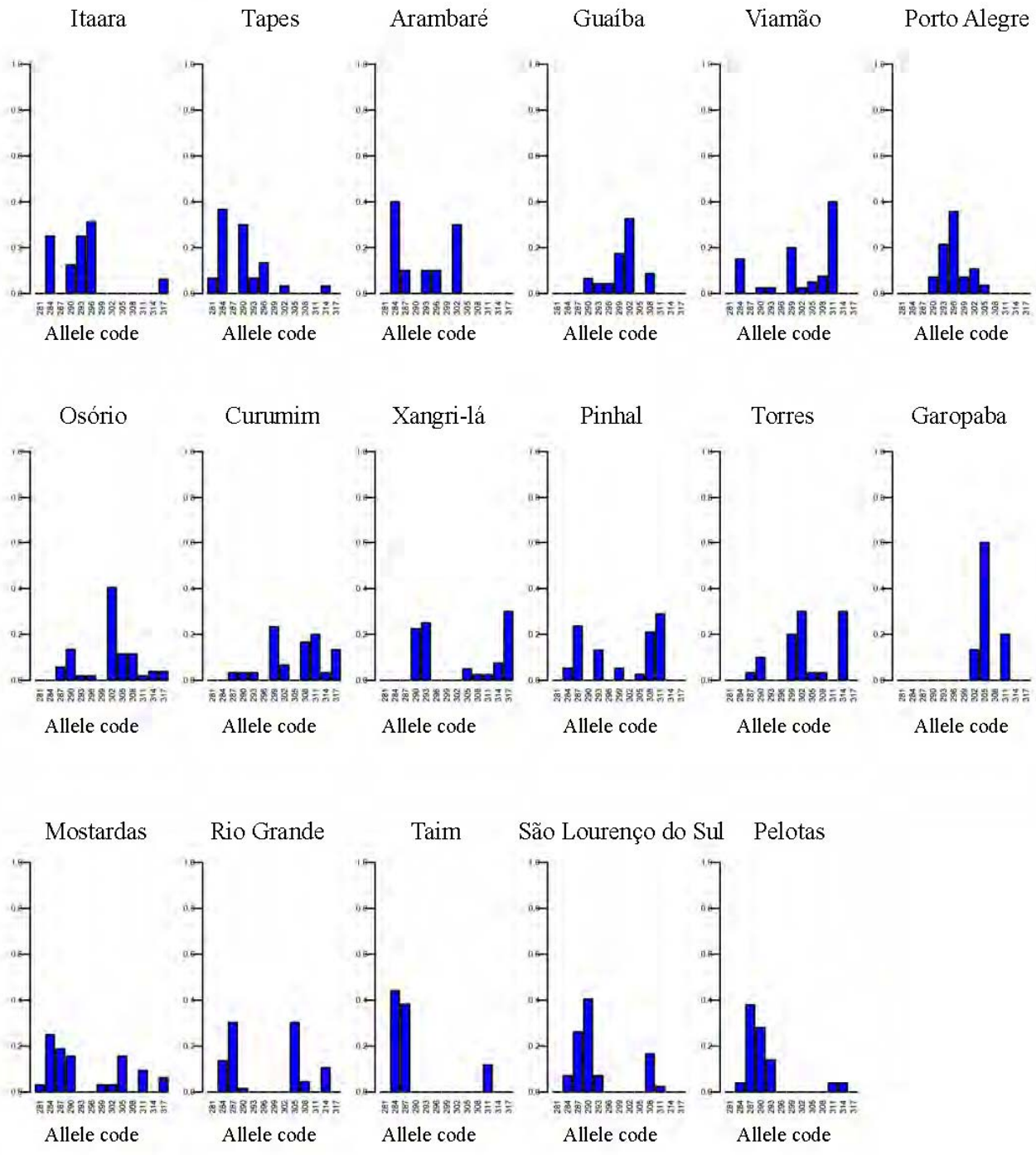
Locus PM117


Itaara
Tapes
Arambaré
Guaíba
Viamão
Porto Alegre
Osório
Curumim
Xangri-lá
Pinhal
Torres
Garopaba
Mostardas
Rio Grande
Taim
São Lourenço do Sul
Pelotas
Allele code

Locus PM157


Itaara
Tapes
Arambaré
Guaíba
Viamão
Porto Alegre
NA
Osório
Curumim
Xangri-lá
Pinhal
Torres
Garopaba
Mostardas
Rio Grande
Taim
São Lourenço do Sul
Pelotas
Allele code

Locus PM167

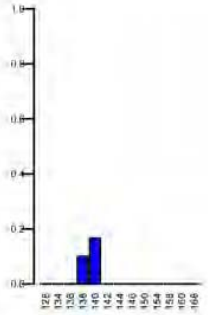

Itaara
Tapes
Arambaré
Guaíba
Viamão
Porto Alegre
Osório
Curumim
Xangri-lá
Pinhal
Torres
Garopaba
Mostardas
Rio Grande
Taim
São Lourenço do Sul
Pelotas
Allele code

Locus PM177

Itaara

NA

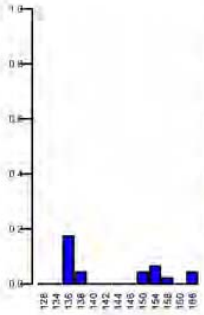
Tapes



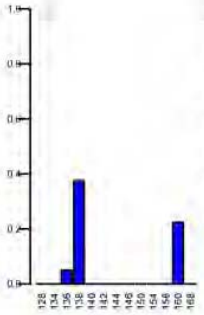
Arambaré

NA

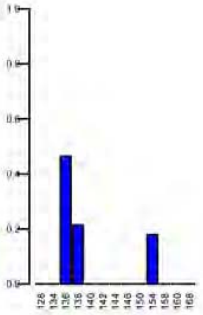
Guaíba



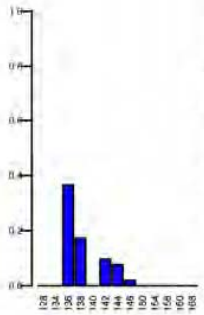
Viamão



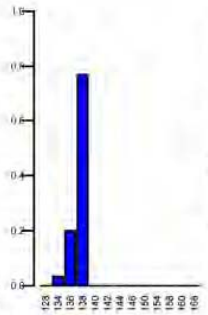
Porto Alegre



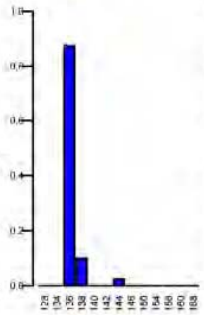
Osório



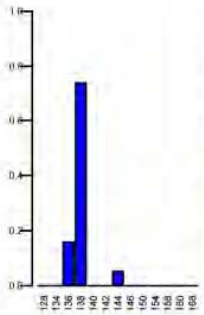
Curumim



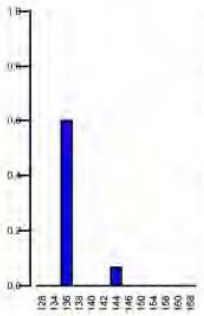
Xangri-lá



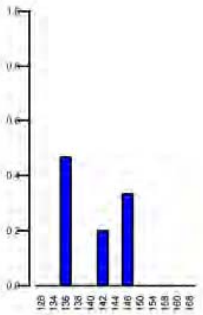
Pinhal



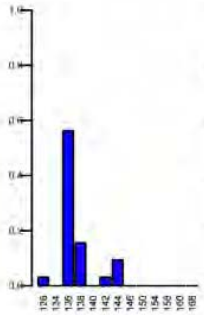
Torres



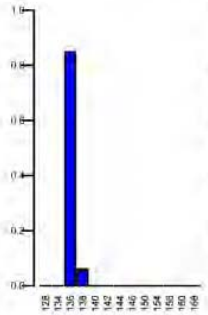
Garopaba



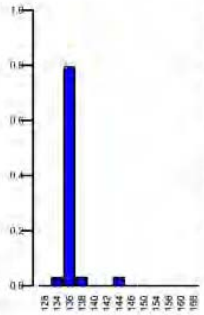
Mostardas



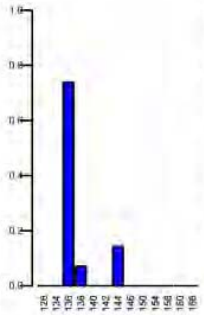
Rio Grande



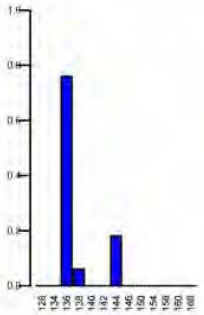
Taim



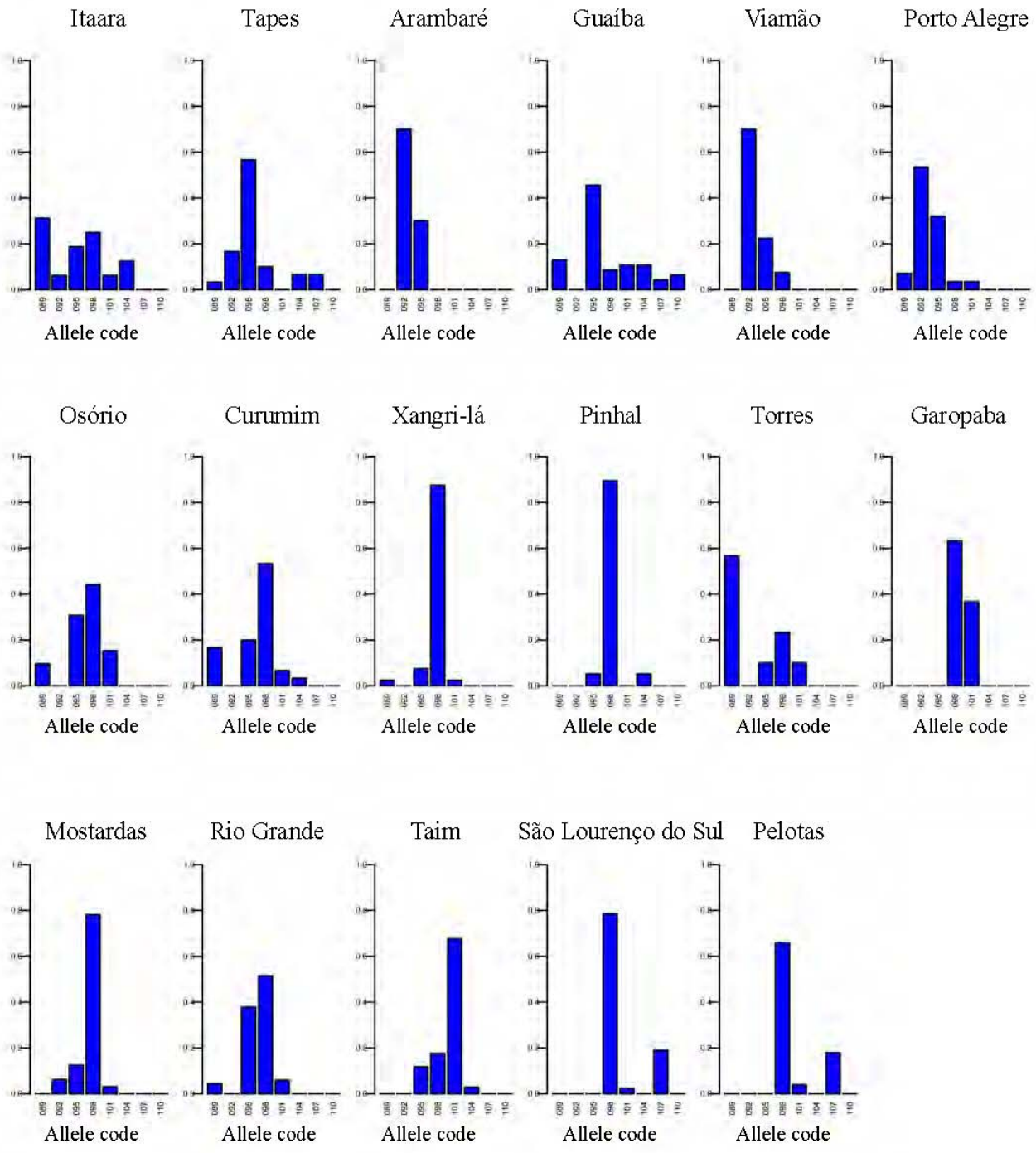
São Lourenço do Sul



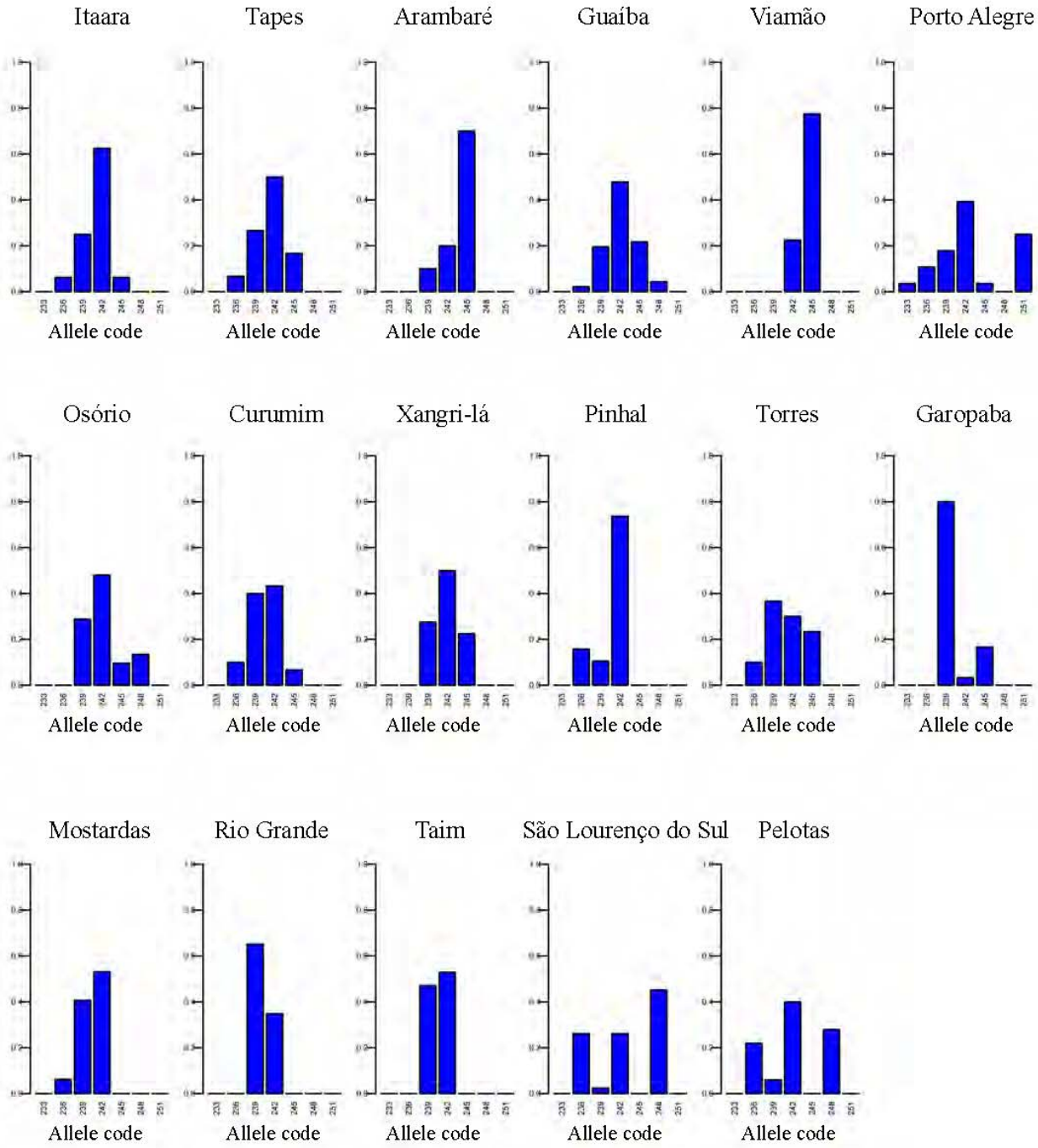
Pelotas



Locus PM184


Itaara
Tapes
Arambaré
Guaíba
Viamão
Porto Alegre
Osório
Curumim
Xangri-lá
Pinhal
Torres
Garopaba
Mostardas
Rio Grande
Taim
São Lourenço do Sul
Pelotas
Allele code

Locus PM192


Itaara
Tapes
Arambaré
Guaíba
Viamão
Porto Alegre
Osório
Curumim
Xangri-lá
Pinhal
Torres
Garopaba
Mostardas
Rio Grande
Taim
São Lourenço do Sul
Pelotas
Allele code

CAPÍTULO 2

Nota publicada na revista *Applications in Plant Sciences* (*American Journal of Botany*)

Novel microsatellites for *Calibrachoa heterophylla* (Solanaceae) endemic to the South Atlantic Coastal Plain of South America

http://dx.doi.org/10.3732/apps.1500021

Applications in Plant Sciences

Applications in Plant Sciences 2015 3(7): 1500021

Primer Note

Novel microsatellites for *Calibrachoa heterophylla* (Solanaceae) endemic to the South Atlantic Coastal Plain of South America[1]

Gustavo Adolfo Silva-Arias[2], Geraldo Mäder[2], Sandro L. Bonatto[3], and Loreta B. Freitas[2,4]

[2]Laboratory of Molecular Evolution, Department of Genetics, Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul 91501-970, Brazil; and [3]Laboratory of Genomic and Molecular Biology, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul 90610-001, Brazil

- *Premise of the study: Calibrachoa heterophylla* (Solanaceae) is a petunia species restricted to the South Atlantic Coastal Plain of South America and presents a recent history of colonization from continental to coastal environments and diversification following the formation of the Coastal Plain during the Quaternary period.
- *Methods and Results:* This study reports a suite of 16 microsatellite loci for *C. heterophylla*. The applicability of these markers was assessed by genotyping 57 individuals from two natural populations. Of the 16 described loci, 12 were found to be polymorphic. Successful cross-amplification tests were obtained using 12 *Calibrachoa* species.
- *Conclusions:* The development of microsatellite markers will be useful to recover the contemporary history of the colonization of the Coastal Plain and to provide information for the conservation of this endemic species.

Key words: *Calibrachoa heterophylla*; cross-amplification; population genetics; simple sequence repeat (SSR) markers; Solanaceae; wild petunia.

Calibrachoa heterophylla (Sendtn.) Wijsman (Solanaceae) is a wild petunia species restricted to the South Atlantic Coastal Plain (SACP) of South America (Greppi et al., 2013; Mäder et al., 2013). The species is a melittophilous shrub that inhabits dunes or sandy grasslands of lakeside or seaside environments and possesses a conspicuous phenotypic plasticity that has been related to environmental differences in geographic distribution (Mäder et al., 2013). *Calibrachoa heterophylla* has a chromosome count of $2n = 18$ (Mishiba et al., 2000). Previous phylogeographic analyses showed that *C. heterophylla* has a continental origin and the deposition of coastal plains during the Quaternary period allowed the species to colonize the SACP (Mäder et al., 2013).

New challenges have emerged, such as the need to assess the contemporary patterns of genetic structure due to the secondary contact of basal genetic lineages after colonization of the SACP; to identify differences in the interpopulation gene flow patterns related to geographical and ecological barriers along the SACP; and to propose conservation measures considering that *C. heterophylla* could undergo drastic habitat reduction due to human-induced global climatic changes. New microsatellite (simple sequence repeat [SSR]) markers will be helpful in further studies that address these questions.

This is the first time that SSR markers have been developed for the genus *Calibrachoa* Cerv. Bossolini et al. (2011) and Kriedt et al. (2011) described several primer sets for the closely related genus *Petunia* Juss., which have been useful for answering evolutionary questions (e.g., Segatto et al., 2014).

METHODS AND RESULTS

Genomic DNA was extracted from the silica-dried leaves from one individual of *C. heterophylla* (geographic coordinates: 30°25′13″S, 51°13′30″W; herbarium voucher BHCB 116994) using a cetyltrimethylammonium bromide (CTAB) protocol according to Roy et al. (1992). An enriched library methodology was used to isolate specific repeat motifs according to Beheregaray et al. (2004). For this, genomic DNA was digested with the restriction enzyme *Rsa*I, and the fragments were linked to two oligo-adapters and amplified by PCR using a thermocycler (Applied Biosystems, Foster City, California, USA). The PCR conditions were as follows: initial denaturation at 95°C for 4 min, followed by 20 cycles of 94°C for 30 s, 60°C for 1 min, and 72°C for 1 min, and a final extension cycle at 72°C for 8 min. The products were purified using the QIAquick PCR Purification Kit (QIAGEN, Hilden, Germany), enriched for three motifs [$(dAT)_8$, $(dGA)_8$, and $(dGAA)_8$], and selectively captured using streptavidin magnetic particles (Invitrogen, Carlsbad, California, USA). PCR was performed on the microsatellite-enriched eluate using one of the oligo-adapters as a primer, with an initial denaturation at 95°C for 1 min, followed by 25 cycles of 94°C for 40 s, 60°C for 1 min, and 72°C for 2 min, and a final extension cycle at 72°C for 5 min. The enriched library was purified, cloned into the pGEM-T vector (Promega Corporation, Madison, Wisconsin, USA), and transformed into competent XL1-Blue *E. coli*. A total of 188 positive clones were PCR-amplified using M13(–20) forward and M13(–40) reverse primers, with an initial denaturation at 95°C for 4 min, followed by 30 cycles of 94°C for 30 s, 52°C for 45 s, and 72°C for 1 min, and a

[1]Manuscript received 4 March 2015; revision accepted 27 April 2015.

This project was supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), and the Programa de Pós-Graduação em Genética e Biologia Molecular da Universidade Federal do Rio Grande do Sul (PPGBM-UFRGS). G.A.S.-A. was supported by a Departamento Administrativo de Ciencia y Tecnología e Innovación (COLCIENCIAS) grant.

[4]Author for correspondence: loreta.freitas@ufrgs.br

doi:10.3732/apps.1500021

TABLE 1. Characteristics of the 16 microsatellite loci developed for *Calibrachoa heterophylla*.[a]

Locus	Primer sequences (5′–3′)	Repeat motif	Allele size range (bp)	T_a (°C)	Fluorescent dye[b]	GenBank accession no.
Che18	F: TTAAGGGAGGTGTAGCCCCA	$(CTT)_{21}$	149–179	54	HEX	KP091702
	R: ACAAAGAGATACATATACGCATGTGT					
Che46	F: TCAAGATAGCACCTTGTTTGCA	$(GA)_{28}$	223–249	54	HEX	KP091703
	R: CGTGCATACATGGTTAATTGGCT					
Che59	F: TCCTCTTCTTCGCTTGCTCC	$(CT)_7$	110–120	54	FAM	KP091704
	R: ATAAATTGGGACGGACGGGG					
Che119	F: TCCAACTGCAGTCAAGCTCT	$(CT)_{12}(TC)_{15}$	162–186	54	HEX	KP091705
	R: ACTGATGACCAATAGAGGAAGAACC					
Che26	F: ACGAAGCTTGTTACCAATCTCAAAA	$(AAG)_7$	90–102	54	HEX	KP091706
	R: AGCCAAAGTAGGGACGTTGA					
Che34	F: TCTTGAAGCCAATTGGAGAATAGT	$(TC)_{10}$	216–226	50	NED	KP091707
	R: TCGATCTGTGCTGCACATCA					
Che81	F: GACTACAGATTGGTCAACTTTTGAG	$(CT)_{16}$	322–358	56	FAM	KP091708
	R: AGGAGAGGCTTCTTTGGACA					
Che82	F: AGAAAAGAGGGATGAGGAGAACT	$(GA)_8$	127–131	49	NED	KP091709
	R: CTCGTCATTTTTCCTTGTCCCA					
Che85	F: TGGTAATGGAGCACGAGGAAG	$(TG)_6(GA)_8$	297–305	49	FAM	KP091710
	R: GGCTTTCAACTTTGTTCAAAACCC					
Che72	F: GCTGAGAACCAAGGAACAGC	$(GA)_{18}$	152–164	48	FAM	KP091711
	R: TCGATCTCTCATCCCCTGCA					
Che126	F: AGAGTTGACCCAAATTTCCCT	$(GA)_{16}$	338–366	48	NED	KP091712
	R: TCCTGTCTTGCCTTGTTTTCAC					
Che33	F: CCAAAGGATGAGGCATGCATTT	$(GA)_{20}$	184–210	50	FAM	KP091713
	R: CCAAACAATGCAGATCCCAAGT					
Che12	F: AACCCCTTCCCTCCAAACAC	$(TCT)_6$	291	54	NED	KP091714
	R: ACCTGTTATGGATTTCAAATGGAGT					
Che48	F: AGCAAGCTTGTCAGACGAAGA	$(GAA)_6$	139	54	NED	KP091715
	R: TTTCATGCTGGTCCATCCCC					
Che83	F: TTGAAGAGGAGGAGGAGGAG	$(AGA)_9$	175	55	FAM	KP091716
	R: AATGAATTCCAAGATCCAAGC					
Che114	F: TCATCAGTGGGAGGTTCATCAC	$(CTT)_{23}$	277	54	HEX	KP091717
	R: ACCATCTGAAAGTGAAAGCGAAG					

Note: T_a = annealing temperature.

[a]All values are based on 57 samples from Brazilian populations representing northern and southern regions of the South Atlantic Coastal Plain ($N = 22$ and $N = 35$, respectively).

[b]Fluorescent dye used for fragment analysis.

final extension cycle at 72°C for 8 min. The PCR products were purified and sequenced with a MegaBACE 1000 automated sequencer (GE Healthcare Biosciences, Pittsburgh, Pennsylvania, USA). Forty clones possessed SSRs, of which 27 were adequate for primer design using Primer3 (Untergasser et al., 2012), with primer sizes between 18 and 25 bp, GC contents ranging from 48% to 60%, and melting temperatures varying from 55°C to 65°C.

TABLE 2. Genetic properties of the 16 microsatellite loci developed for *Calibrachoa heterophylla*.

	Santo Antônio da Patrulha[a] ($n = 22$)			Santa Vitória do Palmar[a] ($n = 35$)		
Locus	A	H_o	H_e	A	H_o	H_e
Che18	5	0.1875	0.73185	4	0.47059	0.60843
Che46	9	0.7	0.83205	12	0.47059	0.88455
Che59	2	0.42105	0.34139	3	0.5	0.60623
Che119	7	0.77273	0.77484	4	0.62857	0.5735
Che26	4	0.40909	0.56977	3	0.02941	0.37357
Che34	3	0.54545	0.42706	1	0	0
Che81	3	0.3	0.44487	5	0.23529	0.63784
Che82	1	0	0	2	0.27586	0.50817
Che85	2	0.05263	0.14936	3	0.2963	0.36618
Che72	2	0.4	0.32821	5	0.66667	0.64149
Che126	5	0.10526	0.51351	9	0.3	0.79103
Che 33	5	0.2	0.71264	5	0.22222	0.66013
Che12	1	0	0	1	0	0
Che48	1	0	0	1	0	0
Che83	1	0	0	1	0	0
Che114	1	0	0	1	0	0

Note: A = number of alleles; H_e = expected heterozygosity; H_o = observed heterozygosity; n = number of individuals sampled.

[a]Santo Antônio da Patrulha population: Geographic coordinates = 29°53′34.5″S, 50°25′45.7″W; herbarium vouchers BHCB 104866/104867. Santa Vitória do Palmar population: Geographic coordinates = 32°59′15.5″S, 52°43′56.3″W; herbarium vouchers BHCB 104907/104908. Both populations are located in the South Atlantic Coastal Plain, Brazil.

TABLE 3. Cross-amplification results for the 16 microsatellite markers developed for *Calibrachoa heterophylla* in 95 individuals of 12 *Calibrachoa* species.

Species	ID LEM[a]	Che18	Che46	Che59	Che119	Che26	Che34	Che81	Che82	Che85	Che72	Che126	Che33	Che12	Che48	Che83	Che114
C. elegans (Miers) Stehmann & Semir	C.eleg 4	0	0	1	1	1	1	0	0	1	0	1	0	1	1	0	1
	C.eleg 10	0	1	1	1	1	1	0	0	1	0	1	0	1	1	0	1
	C.eleg 20	0	0	1	1	1	1	0	0	1	0	1	0	1	1	0	1
	C.eleg 30	0	0	1	1	1	1	0	0	1	0	1	0	1	1	0	1
	C.eleg 43	0	1	1	1	1	1	0	0	1	0	0	0	1	1	0	1
	C.eleg 50	0	1	1	1	1	1	0	0	1	0	1	0	1	1	0	1
	C.eleg 60	0	1	1	1	1	1	0	0	1	0	1	0	1	1	0	1
	C.eleg 70	0	1	1	1	1	1	0	0	1	0	0	0	1	1	0	1
C. ericifolia (R. E. Fr.) Wijsman	C.eric 1	0	0	1	1	0	1	0	0	0	0	1	0	1	1	0	1
	C.eric 15	0	0	1	1	1	1	0	0	0	0	1	0	1	1	0	1
	C.eric 65	0	1	1	1	1	1	0	0	0	0	1	0	1	1	0	1
	C.eric 92	0	0	1	1	1	1	0	0	0	0	1	0	1	1	0	1
	C.eric 107	0	0	1	1	1	1	0	0	0	0	1	0	1	1	0	1
	C.eric 148	0	0	1	1	1	1	0	0	0	0	1	0	1	1	0	1
	C.eric 179	0	1	1	1	1	1	0	0	0	0	1	0	1	1	0	1
	C.eric 180	0	1	1	1	1	1	0	0	0	0	1	0	1	1	0	1
C. excellens (R. E. Fr.) Wijsman subsp. *atropurpurea* Stehmann & Semir	C.exca 1	0	1	0	0	0	0	0	0	0	0	1	0	0	1	0	1
	C.exca 7	0	1	1	0	0	0	0	0	0	0	1	0	1	1	0	1
	C.exca 9	0	1	1	0	0	0	0	0	0	0	1	0	1	1	0	1
	C.exca 20	0	1	1	0	0	0	0	0	0	0	1	0	0	1	0	1
	C.exca 24	0	1	1	0	1	0	0	0	0	0	1	0	0	1	0	1
	C.exca 25	0	1	1	0	1	0	0	0	0	0	1	0	1	1	0	1
	C.exca 40	0	1	1	0	0	0	0	0	0	0	1	0	1	1	0	1
	C.exca 50	0	0	1	0	0	0	0	0	0	0	1	0	1	1	0	1
C. excellens (R. E. Fr.) Wijsman subsp. *excellens*	C.exce 2	0	1	0	1	0	1	0	0	0	0	1	0	0	1	0	0
	C.exce 9	0	0	0	1	0	1	1	0	0	0	1	0	0	1	0	0
	C.exce 12	0	0	1	1	1	1	0	0	0	0	1	0	0	1	0	1
	C.exce 30	0	0	1	1	1	1	1	0	0	0	1	0	1	1	0	1
	C.exce 40	0	0	1	1	1	1	1	0	0	0	1	0	1	1	0	1
	C.exce 100	0	0	1	1	1	1	1	0	0	0	1	0	1	1	0	1
	C.exce 120	0	0	1	1	1	1	1	0	0	0	1	0	1	1	0	1
	C.exce 220	0	0	1	1	1	1	1	0	0	0	1	0	1	1	0	1
C. humilis (R. E. Fr.) Stehmann & Semir	C.humi 3	0	0	1	1	1	1	1	0	0	0	1	0	0	1	0	1
	C.humi 4	0	0	1	1	1	1	1	0	0	0	1	0	1	1	0	1
	C.humi 15	0	0	1	1	0	1	1	0	0	0	1	0	1	1	0	1
	C.humi 21	0	0	1	1	0	1	1	0	0	0	1	0	0	1	0	1
	C.humi 25	0	0	1	1	0	1	1	0	0	0	1	0	0	1	0	1
	C.humi 30	0	0	1	1	1	1	0	0	0	0	1	0	0	1	0	1
	C.humi 35	0	0	1	1	1	1	1	0	0	0	1	0	0	1	0	1
	C.humi 37	0	0	1	1	1	1	0	0	0	0	1	0	0	1	0	1
	C.humi 42	0	0	1	1	1	1	0	0	0	0	1	0	0	1	0	1
	C.humi 44	0	0	1	1	1	1	1	0	0	0	1	0	0	1	0	1
	C.humi 47	0	0	1	1	0	1	1	0	0	0	1	0	0	1	0	1
	C.humi 55	0	0	1	1	0	1	1	0	0	0	1	0	0	1	0	1
	C.humi 64	0	0	1	1	0	1	1	0	0	0	1	0	0	1	0	1
C. linearis (Hook.) Wijsman	C.line 5	0	0	1	1	0	0	1	0	0	0	1	0	0	1	0	1
	C.line 7	0	0	1	1	0	0	1	0	0	0	1	0	0	1	0	1
	C.line 12	0	0	1	1	1	0	1	0	0	0	1	0	0	1	0	1

TABLE 3. Continued.

Species	ID LEM[a]	Che18	Che46	Che59	Che119	Che26	Che34	Che81	Che82	Che85	Che72	Che126	Che33	Che12	Che48	Che83	Che114
C. linoides (Sendtn.) Wijsman	C.lino 1	0	1	1	1	1	1	0	0	0	0	1	0	0	1	0	1
	C.lino 13	0	1	1	1	1	1	0	0	0	0	1	0	0	1	0	1
	C.lino 20	0	1	0	1	1	1	0	0	0	0	1	0	0	1	0	1
	C.lino 33	0	0	1	1	0	1	0	0	0	0	1	0	0	1	0	1
	C.lino 52	0	1	1	1	1	1	0	0	0	0	1	0	1	1	0	1
	C.lino 72	0	0	1	1	0	1	0	0	0	0	1	0	0	1	0	1
	C.lino 189	0	1	1	1	1	1	0	0	0	0	1	0	1	1	0	1
	C.lino 205	0	1	1	1	1	1	0	0	0	0	1	0	1	1	0	1
C. ovalifolia (Miers) Stehmann & Semir	C.oval 2	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0
	C.oval 11	0	1	1	1	0	1	0	0	0	0	1	0	0	1	0	1
	C.oval 17	0	1	1	1	0	1	0	0	0	0	1	0	0	1	0	1
	C.oval 26	0	1	0	1	0	1	0	0	0	0	0	0	0	1	0	1
	C.oval 34	0	1	0	1	0	1	0	0	0	0	0	0	0	1	0	1
	C.oval 35	0	0	0	1	0	1	0	0	0	0	1	0	0	1	0	1
	C.oval 101	0	0	1	1	0	1	0	0	0	0	1	0	0	1	0	1
	C.oval 129	0	0	1	1	0	1	0	0	0	0	0	0	0	1	0	1
C. paranensis (Dusén) Wijsman	C.para 1	0	0	1	1	1	1	0	0	0	0	1	0	1	1	0	1
	C.para 28	0	0	1	1	1	1	0	0	0	0	1	0	0	1	0	1
	C.para 47	0	0	1	1	1	1	0	0	0	0	1	0	1	1	0	1
	C.para 73	0	0	1	1	1	1	0	0	0	0	1	0	1	1	0	1
	C.para 99	0	0	1	1	1	1	0	0	0	0	1	0	1	1	0	1
	C.para 125	0	0	1	1	1	1	0	0	0	0	1	0	1	1	0	1
	C.para 162	0	0	1	1	1	1	0	0	0	0	1	0	1	1	0	1
	C.para 186	0	0	1	1	1	1	0	0	0	0	1	0	0	1	0	1
C. pygmea (R. E. Fr.) Wijsman	C.pygm 1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	C.pygm 30	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	1
	C.pygm 31	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1
	C.pygm 36	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	1
	C.pygm 41	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1
	C.pygm 51	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
	C.pygm 55	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
C. serrulata (L. B. Sm. & Downs) Stehmann & Semir	C.serr 1	0	0	1	1	0	1	0	0	1	0	1	0	1	1	0	1
	C.serr 3	0	0	1	1	0	1	0	0	1	0	1	0	1	1	0	1
	C.serr 5	0	0	1	1	0	1	0	0	1	0	1	0	1	1	0	1
	C.serr 7	0	0	1	1	0	1	0	0	1	0	0	0	1	1	0	1
	C.serr 11	0	0	1	1	0	1	0	0	1	0	1	0	1	1	0	1
	C.serr 15	0	0	1	1	0	1	0	0	1	0	1	0	1	1	0	1
	C.serr 17	0	0	1	1	0	1	0	0	1	0	1	0	0	1	0	1
	C.serr 19	0	0	1	1	0	1	0	0	1	0	1	0	1	1	0	1
C. thymifolia (A. St.-Hil.) Stehmann & Semir	C.thym 1	0	0	1	1	0	1	0	0	1	0	1	0	0	1	0	1
	C.thym 3	0	0	1	1	0	1	0	0	1	0	1	0	1	1	0	1
	C.thym 5	0	0	1	1	0	1	0	0	1	0	1	0	1	1	0	1
	C.thym 36	0	0	1	1	0	1	0	0	1	0	1	0	1	1	0	1
	C.thym 54	0	0	1	1	0	1	0	0	1	0	1	0	0	1	0	1
	C.thym 56	0	0	1	1	0	1	0	0	1	0	1	0	0	1	0	1
	C.thym 58	0	0	1	1	0	1	0	0	1	0	1	0	0	1	0	1
	C.thym 59	0	0	1	1	0	1	0	0	1	0	1	0	0	1	0	1

Note: 1 = successful amplification; 0 = failed amplification.

[a] Sample identification code in the Laboratory of Molecular Evolution, Department of Genetics, Universidade Federal do Rio Grande do Sul.

The resulting markers were tested in two populations of *C. heterophylla* belonging to two different chloroplast haplogroups (Mäder et al., 2013): Santo Antônio da Patrulha (geographic coordinates 29°53′34.5″S, 50°25′45.7″W; herbarium vouchers BHCB 104866/104867) and Santa Vitória do Palmar (geographic coordinates 32°59′15.5″S, 52°43′56.3″W; herbarium vouchers BHCB 104907/104908), Rio Grande do Sul, Brazil. PCR was performed in 10-μL reactions containing ~10 ng/μL of template DNA, 200 μM each dNTP (Invitrogen), 2 pmol fluorescently labeled M13(–21) primer and reverse primer, 0.4 pmol forward primer, 2.0 mM $MgCl_2$ (Invitrogen), 0.5 units of *Taq* Platinum DNA polymerase, and 1× *Taq* Platinum reaction buffer (Invitrogen). The PCR conditions were as follows: initial denaturation at 94°C for 3 min, followed by 32 cycles of 94°C for 20 s, 53–65°C for 45 s, and 72°C for 1 min, and a final extension cycle at 72°C for 10 min. The forward primers were labeled with FAM, NED, or HEX (Table 1). The products were analyzed using a MegaBACE 1000 automated sequencer with the ET-ROX 550 size ladder (GE Healthcare Biosciences). Genotyping results were scored using GeneMarker software (version 2.4; SoftGenetics, State College, Pennsylvania, USA).

Sixteen loci with a clear and strong single band for each allele were identified and used to genotype 57 individuals from two populations of *C. heterophylla*. Twelve loci displayed polymorphism, whereas the other four loci were monomorphic (Table 1). All of the individuals presented one or two alleles (consistent with the diploid condition of *C. heterophylla*) that matched the expected sizes based on cloned sequences. In the Santo Antônio da Patrulha population, the number of alleles per locus for the 12 polymorphic loci varied from one to nine, with an average of four, and observed (H_o) and expected (H_e) heterozygosity ranged from 0 to 0.773 and 0 to 0.832, with averages of 0.341 and 0.485, respectively (Table 2). In the Santa Vitória do Palmar population, the number of alleles per locus for the 12 polymorphic loci varied from one to 12, and H_o and H_e ranged from 0 to 0.667 and 0 to 0.885, with averages of 0.341 and 0.554, respectively (Table 2). Considering both populations, the total number of alleles per locus for the 12 polymorphic loci ranged from two (Che34 and Che82) to 13 (Che46), and H_o and H_e ranged from 0.138 to 0.701 and from 0.193 to 0.899, with averages of 0.341 and 0.624, respectively (Table 2). Che18, Che46, Che126, and Che33 deviated from HWE in the two populations, and Che26, Che81, Che82, and Che85 deviated from HWE in the Santa Vitória do Palmar population ($P < 0.04$), all due to heterozygote deficiency. All analyses were conducted with Arlequin version 3.5 (Excoffier and Lischer, 2010). There are no specific studies in reproductive biology for *C. heterophylla*, but we suspect that high levels of autogamy (observed in some *Petunia* species, e.g., Turchetto et al., 2015) could be responsible for the low levels of heterozygosity in the analyzed populations. Additionally, considering that *C. heterophylla* recently colonized and is in continuing expansion over the SACP, one would expect to find populations with relatively high allelic richness and, at the same time, low heterozygosity.

Cross-amplification of all the developed loci was tested in 95 individuals of 12 *Calibrachoa* species, covering most of the geographic range and phylogenetic diversity of the genus (Table 3, Appendix 1, Appendix S1; Fregonezi et al., 2012), under the same PCR conditions used for *C. heterophylla*. Except for *C. pygmea* (R. E. Fr.) Wijsman, most of the loci showed positive amplification in the species tested, indicating that the developed markers are useful for other *Calibrachoa* species. The markers Che59, Che119, Che34, Che126, Che48, and Che114 showed the highest rates of the cross-amplification tests (Table 3, Appendix S1). The lower rates of cross-amplification for *C. pygmea* are unsurprising given that this species is classified in a different subgenus and is phylogenetically more distant to *C. heterophylla* than the remaining species included in this study (Table 3, Appendix S1; Fregonezi et al., 2012).

CONCLUSIONS

These are the first SSR markers developed for *C. heterophylla*. These loci will allow us to investigate the effects of landscape heterogeneity on the genetic structure of *C. heterophylla* populations and, combined with other analyses and species, will allow us to understand the colonization process of plant groups to the SACP. These markers may also be valuable for conservation of this endemic species.

LITERATURE CITED

Beheregaray, L. B., L. M. Moller, T. S. Schwartz, N. L. Chao, and A. Caccone. 2004. Microsatellite markers for the cardinal tetra *Paracheirodon axelrodi*, a commercially important fish from central Amazonia. *Molecular Ecology Notes* 4: 330–332.

Bossolini, E., U. Klahre, A. Brandenburg, D. Reinhardt, and C. Kuhlemeier. 2011. High resolution linkage maps of the model organism *Petunia* reveal substantial synteny decay with the related genome of tomato. *Genome* 54: 327–340.

Excoffier, L., and H. E. L. Lischer. 2010. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* 10: 564–567.

Fregonezi, J. N., L. B. de Freitas, S. L. Bonatto, J. Semir, and J. R. Stehmann. 2012. Infrageneric classification of *Calibrachoa* (Solanaceae) based on morphological and molecular evidence. *Taxon* 61: 120–130.

Greppi, J. A., J. C. Hagiwara, and J. R. Stehmann. 2013. Novedades en *Calibrachoa* (Solanaceae) y notas taxonómicas sobre el género para la Argentina. *Darwiniana, nueva serie* 1: 173–187.

Kriedt, R. A., A. M. C. Ramos-Fregonezi, L. B. Beheregaray, S. L. Bonatto, and L. B. Freitas. 2011. Isolation, characterization, and cross-amplification of microsatellite markers for the *Petunia integrifolia* (Solanaceae) complex. *American Journal of Botany* 98: e277–e279.

Mäder, G., J. N. Fregonezi, A. P. Lorenz-Lemke, S. L. Bonatto, and L. B. Freitas. 2013. Geological and climatic changes in Quaternary shaped the evolutionary history of *Calibrachoa heterophylla*, an endemic South-Atlantic species of petunia. *BMC Evolutionary Biology* 13: 178.

Mishiba, K.-I., T. Ando, M. Mii, H. Watanabe, H. Kokubun, G. Hashimoto, and E. Marchesi. 2000. Nuclear DNA content as an index character discriminating taxa in the genus *Petunia* sensu Jussieu (Solanaceae). *Annals of Botany* 85: 665–673.

Roy, A., N. Frascaria, J. MacKay, and J. Bousquet. 1992. Segregating random amplified polymorphic DNAs (RAPDs) in *Betula alleghaniensis*. *Theoretical and Applied Genetics* 85: 173–180.

Segatto, A. L. A., A. M. C. Ramos-Fregonezi, S. L. Bonatto, and L. B. Freitas. 2014. Molecular insights into the purple-flowered ancestor of garden petunias. *American Journal of Botany* 101: 119–127.

Turchetto, C., J. S. Lima, D. M. Rodrigues, S. L. Bonatto, and L. B. Freitas. 2015. Pollen dispersal and breeding structure in a hawkmoth-pollinated Pampa grasslands species *Petunia axillaris* (Solanaceae). *Annals of Botany* http://aob.oxfordjournals.org/cgi/doi/10.1093/aob/mcv025.

Untergasser, A., I. Cutcutache, T. Koressaar, J. Ye, B. C. Faircloth, M. Remm, and S. G. Rozen. 2012. Primer3—new capabilities and interfaces. *Nucleic Acids Research* 40: e115.

Appendix 1. Locality and voucher information for *Calibrachoa* species used in this study.

Species	Locality[a]	ID LEM[b]	Herbarium voucher no.[c]	Latitude	Longitude
C. elegans (Miers) Stehmann & Semir	Serra da Calçada, Brumadinho/MG, Brazil	C.eleg 4	BHCB 45365	–20.09305556	–43.98361111
	Serra da Calçada, Brumadinho/MG, Brazil	C.eleg 10	BHCB 45365	–20.09305556	–43.98361111
	Santana do Garabéu/MG, Brazil	C.eleg 20	BHCB 103115	–21.61094722	–44.12620278
	Santana do Garabéu/MG, Brazil	C.eleg 30	BHCB 103115	–21.61094722	–44.12620278
	Morro do Chapéu, Nova Lima/MG, Brazil	C.eleg 43	NA	–20.10310278	–43.94620278
	Morro do Chapéu, Nova Lima/MG, Brazil	C.eleg 50	NA	–20.10310278	–43.94620278
	Retiro das Pedras, Brumadinho/MG, Brazil	C.eleg 60	BHCB 45365	–20.09305556	–43.98333333
	Retiro das Pedras, Brumadinho/MG, Brazil	C.eleg 70	BHCB 45365	–20.09305556	–43.98333333
C. ericifolia (R. E. Fr.) Wijsman	Passo do Pupo, Ponta Grossa/PR, Brazil	C.eric 1	BHCB 96546	–25.14690737	–49.95483021
	Ponta Grossa/PR, Brazil	C.eric 15	BHCB 104025	–25.25597504	–50.15140244
	Castro, Tibagi/PR, Brazil	C.eric 65	BHCB 104026	–24.76576497	–50.1578068
	Castro, Tibagi/PR, Brazil	C.eric 92	BHCB 104026	–24.76792733	–50.15422672
	Castro, Tibagi/PR, Brazil	C.eric 107	BHCB 104027	–24.65006797	–50.22478516
	Parque Estadual do Quartelá, Tibagi/PR, Brazil	C.eric 148	BHCB 104028	–24.56777939	–50.26055384
	Parque Estadual do Quartelá, Tibagi/PR, Brazil	C.eric 179	BHCB 104028	–24.56777939	–50.26055384
	Balsa Nova/PR, Brazil	C.eric 180	BHCB 104029	–25.44122239	–49.74743217
C. excellens (R. E. Fr.) Wijsman subsp. *atropurpurea* Stehmann & Semir	Pedra da Abelha, Pedra do Segredo, Caçapava do Sul/RS, Brazil	C.exca 1	BHCB 75119	–30.53129713	–53.55028538
	Bagé, Santana do Livramento/RS, Brazil	C.exca 7	BHCB 102126	–31.21294677	–54.28641303
	Pedra do Segredo, Caçapava do Sul/RS, Brazil	C.exca 9	NA	–30.5423423	–53.55549993
	Canguçu/RS, Brazil	C.exca 20	HUEFS 79276	–31.34311667	–52.67974444
	P.E. Itapuã, Viamão/RS, Brazil	C.exca 24	NA	NA	NA
	Loreto, São Vicente do Sul/RS, Brazil	C.exca 25	BHCB 102094	–29.72229746	–54.84055541
	Galpão de Pedra, Caçapava do Sul/RS, Brazil	C.exca 40	NA	–30.547039	–53.546476
	Galpão de Pedra, Pedra do Segredo, Caçapava do Sul/RS, Brazil	C.exca 50	ICN 158617	–30.989167	–53.645278
C. excellens (R. E. Fr.) Wijsman subsp. *excellens*	Pedra do Segredo, Caçapava do Sul/RS, Brazil	C.exce 2	BHCB 75090	–30.54944504	–53.54229307
	Pantano Grande, Encruzilhada do Sul/RS, Brazil	C.exce 9	BHCB 75150	–30.35481968	–52.40705791
	Bom Jardim da Serra/SC, Brazil	C.exce 12	BHCB 80084	–28.34020411	–49.62996039
	Estrada Urubici–São Joaquim, Urubici/SC, Brazil	C.exce 30	BHCB 96689	–28.14887971	–49.71661418
	Guarapuava/PR, Brazil	C.exce 40	BHCB 96605	–25.48049958	–51.53825009
	Pinhão/PR, Brazil	C.exce 100	BHCB 96612	–25.948983	–51.61110903
	Porto Alegre/RS, Brazil	C.exce 120	NA	–30.04383263	–51.11589396
	Xaxim/SC, Brazil	C.exce 220	BHCB 128573	–26.95230679	–52.51188593
C. humilis (R. E. Fr.) Stehmann & Semir	Uruguaiana/RS, Brazil	C.humi 3	BHCB 117015	–29.41827955	–56.69409838
	Uruguaiana/RS, Brazil	C.humi 4	BHCB 117015	–29.41827955	–56.69409838
	Quaraí/RS, Brazil	C.humi 15	Ana Luíza Cazé e Geraldo Mader	–30.31516667	–56.47583333
	Três Cerros, Uruguay	C.humi 21	Greppi 1470	–29.12433333	–56.896
	Três Cerros, Uruguay	C.humi 25	Greppi 1474	–29.12433333	–56.896
	Três Cerros, Uruguay	C.humi 30	Greppi 1479	–29.12433333	–56.896
	Três Cerros, Uruguay	C.humi 35	Greppi 1484	–29.12433333	–56.896
	Três Cerros, Uruguay	C.humi 37	Greppi 1486	–29.12433333	–56.896
	Três Cerros, Uruguay	C.humi 42	Greppi 1491	–29.12433333	–56.896
	Três Cerros, Uruguay	C.humi 44	Greppi 1493	–29.12433333	–56.896
	Mercedes-Corrientes, Argentina	C.humi 47	Greppi 1496	–29.40433333	–57.96833333
	Mercedes-Corrientes, Argentina	C.humi 55	Greppi 1510	–29.40433333	–57.96833333
	Monte Caseros, Argentina	C.humi 64	Greppi 1519	–30.250467	–57.638451
C. linearis (Hook.) Wijsman	Colon. Balneario San Jose, Argentina	C.line 5	Greppi 1023	–32.18333333	–58.16666667
	Corrientes, Argentina	C.line 7	NA	–29.21283333	–59.22025
	Corrientes, Argentina	C.line 12	NA	–29.57605556	–59.33047222

Applications in Plant Sciences 2015 3(7): 1500021
doi:10.3732/apps.1500021

Appendix 1. Continued.

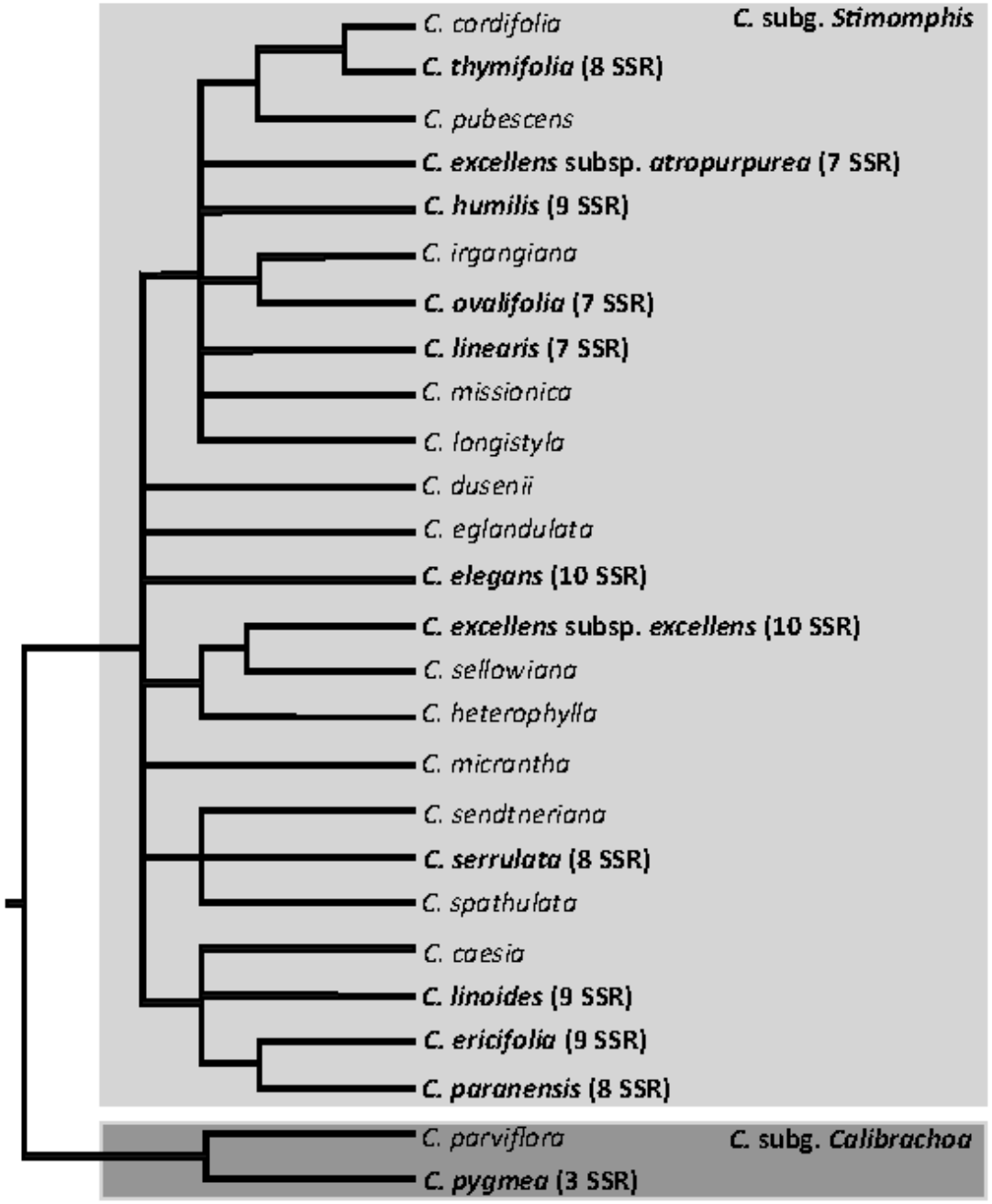
Species	Locality[a]	ID LEM[b]	Herbarium voucher no.[c]	Latitude	Longitude
C. linoides (Sendtn.) Wijsman	São Bento do Sul, Campo Alegre/SC, Brazil	C.lino 1	João Renato Stehmann 3313 (BHCB)	–26.20806629	–49.30657214
	Rio do Sul/SC, Brazil	C.lino 13	HUEFS 75248	–27.31232411	–50.38754739
	Otacílio Costa, Petrolândia/SC, Brazil	C.lino 20	João Renato Stehmann 3322 (BHCB)	–27.57700584	–50.00116462
	Paraíso da Serra/SC, Brazil	C.lino 33	João Renato Stehmann 3330 (BHCB)	–27.81610248	–49.54344749
	Bituruna/PR, Brazil	C.lino 52	ICN 161236	–26.16983668	–51.59984694
	General Carneiro, Matos Costa/SC, Brazil	C.lino 72	ICN 160338	–26.46892235	–51.17974578
	Monte Verde/MG, Brazil	C.lino 189	NA	–22.86653442	–46.04815029
	Monte Verde/MG, Brazil	C.lino 205	NA	–22.86827316	–46.04316213
C. ovalifolia (Miers) Stehmann & Semir	Pedra da Abelha, Pedra do Segredo, Caçapava do Sul/RS, Brazil	C.oval 2	BHCB 75116	–30.53078785	–53.54969655
	Morro Grande, Cachoeira do Sul/RS, Brazil	C.oval 11	BHCB 75136	–30.3199487	–52.92622084
	Santana do Livramento, Quaraí/RS, Brazil	C.oval 17	BHCB 79868	–30.80620235	–55.61801598
	Cerro do Jarau, Quaraí–Uruguaiana/RS, Brazil	C.oval 26	BHCB 79879	–30.19979135	–56.49454328
	Pedra do Segredo, Caçapava do Sul/RS, Brazil	C.oval 34	BHCB 75123	–30.5423423	–53.55549993
	Pedra do Segredo, Caçapava do Sul/RS, Brazil	C.oval 35	BHCB 75123	–30.5423423	–53.55549993
	S. Martinho da Serra, Júlio de Castilho/RS, Brazil	C.oval 101	BHCB 117004	–29.32271143	–53.67828031
	Morro da Cruz, Caçapava do Sul/RS, Brazil	C.oval 129	BHCB 75123	–30.896659	–53.421003
C. paranensis (Dusén) Wijsman	Serra de São Luis do Purunã, Campo Largo-Ponta Grossa/PR, Brazil	C.para 1	ICN 189546	–25.45953736	–49.64560739
	Serra de São Luis do Purunã, Balsa Nova/PR, Brazil	C.para 28	ICN 161233	–25.46811154	–49.65597966
	Serra de São Luis do Purunã, Balsa Nova/PR, Brazil	C.para 47	João Renato Stehmann 4211 (BHCB)	–25.46813099	–49.63903748
	Serra de São Luis do Purunã, Balsa Nova/PR, Brazil	C.para 73	João Renato Stehmann 4215 (BHCB)	–25.46581876	–49.72120914
	Passo do Pupo, Ponta Grossa/PR, Brazil	C.para 99	João Renato Stehmann 4222 (BHCB)	–25.14690737	–49.95483021
	Reserva Natural Buraco do Padre, Ponta Grossa/PR, Brazil	C.para 125	João Renato Stehmann 4227 (BHCB)	–25.17168009	–49.96811494
	Guarapuava, Laranjeiras do Sul/PR, Brazil	C.para 162	João Renato Stehmann 4259 (BHCB)	–25.41794392	–51.67142328
	Distrito de São Luiz do Purunã, Balsa Nova/PR, Brazil	C.para 186	BHCB 104021	–25.46323688	–49.71047686
C. pygmea (R. E. Fr.) Wijsman	Quaraí–Uruguaiana/RS, Brazil	C.pygm 1	BHCB 79881	–30.15518605	–56.43883799
	Alegrete/RS, Brazil	C.pygm 30	BHCB 102099	–29.7880923	–55.72627732
	Alegrete/RS, Brazil	C.pygm 31	BHCB 102099	–29.7880923	–55.72627732
	Alegrete/RS, Brazil	C.pygm 36	BHCB 102099	–29.7880923	–55.72627732
	Alegrete/RS, Brazil	C.pygm 41	BHCB 102099	–29.7880923	–55.72627732
	Alegrete/RS, Brazil	C.pygm 51	BHCB 102099	–29.7880923	–55.72627732
	Alegrete/RS, Brazil	C.pygm 55	BHCB 102099	–29.7880923	–55.72627732
C. serrulata (L. B. Sm. & Downs) Stehmann & Semir	Cânion da Serra do Rio do Rastro, Lauro Muller/SC, Brazil	C.serr 1	BHCB 99775	–28.38594625	–49.54480067
	Cânion da Serra do Rio do Rastro, Lauro Muller/SC, Brazil	C.serr 3	BHCB 02009	–28.3877839	–49.54396608
	Cânion da Serra do Rio do Rastro, Lauro Muller/SC, Brazil	C.serr 5	BHCB 02009	–28.3877839	–49.54396608
	Cânion da Serra do Rio do Rastro, Lauro Muller/SC, Brazil	C.serr 7	BHCB 02009	–28.3877839	–49.54396608
	Cânion da Serra do Rio do Rastro, Lauro Muller/SC, Brazil	C.serr 11	BHCB 02009	–28.3877839	–49.54396608
	Bom Jardim da Serra/SC, Brazil	C.serr 15	BHCB 116965	–28.36292275	–49.55145128
	Bom Jardim da Serra/SC, Brazil	C.serr 17	BHCB 116965	–28.36292275	–49.55145128
	Bom Jardim da Serra/SC, Brazil	C.serr 19	BHCB 116964	–28.35748858	–49.55059298
C. thymifolia (A. St.-Hil.) Stehmann & Semir	Federación Santa Ana, playa de Sta. Ana, Argentina	C.thym 1	Greppi 1003	–30.91305556	–57.91805556
	Entre Rios, Sta. Ana, playa arenosa sobre Rio Uruguay, Argentina	C.thym 3	Greppi 1017	–30.9	–57.91666667
	Santa Ana, Entre Rios, Argentina	C.thym 5	Greppi 1416	–30.899868	–57.931123
	Paso de Los Libres, Argentina	C.thym 36	Greppi 1447	–29.714594	–57.097374
	Colon, Entre Rios, Argentina	C.thym 54	Greppi 1502	–32.184	–58.17416667
	San Roque, Corrientes, Argentina	C.thym 56	Greppi 1504	–28.90638889	–58.66166667
	Concordia, Entre Ríos, Argentina	C.thym 58	Greppi 1506	–31.40183333	–58.10166667
	Concordia, Entre Ríos, Argentina	C.thym 59	Greppi 1507	–31.40183333	–58.10166667

[a] MG = Minas Gerais state; PR = Paraná state; RS = Rio Grande do Sul state; SC = Santa Catarina state.

[b] Sample identification code in the Laboratory of Molecular Evolution, Department of Genetics, Universidade Federal do Rio Grande do Sul.

[c] Herbarium acronyms are per Index Herbariorum (http://sweetgum.nybg.org/ih/); NA = not available. Vouchers identified as "Greppi" are greenhouse-cultivated samples provided by Julián Alejandro Greppi (Instituto Nacional de Tecnología Agropecuaria [INTA], Argentina).

Appendix S1. Phylogenetic tree modified from Fregonezi et al. (2012) highlighting (in boldface) *Calibrachoa* species chosen for cross-amplification tests of microsatellite markers developed for *C. heterophylla*. The number of markers that showed positive amplification is indicated in parentheses.



CAPÍTULO 3

Artigo em preparação para ser submetido à revista *Journal of Evolutionary Biology*

New insights on climatic driven genetic differentiation in peripheral populations during a coastal colonization process

New insights on climatic driven genetic differentiation in peripheral populations during a coastal colonization process

Running title: Isolation by environment in marginal populations

GUSTAVO A. SILVA-ARIAS, GIOVANNA C. GIUDICELLI, and LORETA B. FREITAS*

Laboratory of Molecular Evolution, Department of Genetics, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil

*Corresponding author: Loreta B. Freitas, tel: +55 51 33086731, fax: +55 51 3308 9823, e-mail: loreta.freitas@ufrgs.br

Type of paper: Research Article

Abstract

Colonization and spreading process to new environments generate several signals in the genome of species that become key information to identify evolutionary processes responsible for population establishment and adaptation to differential conditions, as well as to address conservation efforts. In this work, we characterize the genetic diversity, structure, and gene flow patterns of the small shrub *Calibrachoa heterophylla* and associate those metrics with spatial and climatic variables. Using 10 microsatellite loci we genotyped 253 individuals from 15 populations, covering the entire geographic range of the species. Our main results showed higher genetic diversity and effective population size in populations located more distantly from the seashore than those closer to the sea. Interestingly, we found isolation by distance pattern linked with precipitation seasonality that could be related with local adaptation processes in edge coastal populations. This pattern was congruent with other co-distributed species. Gene flow estimations suggested that marginal populations had restricted immigration rates, which could enhance local adaptation process by preventing gene swamping. Populations located close to the Patos Lagoon presented high admixture pattern probably related to the historical and current environmental dynamic features of the region.

Keywords

Calibrachoa heterophylla; gene flow; isolation-by-distance; isolation-by-environment; landscape genetics; South Atlantic Coastal Plain

Introduction

The gene flow has been widely viewed as a constraining evolutionary force because the homogenizing or cohesive effect that it exerts on conspecific populations, opposing to divergent selection and drift that prevent genetic differentiation, local adaptation, and ultimately speciation processes. However, the importance of the gene flow has been discussed as evolutionary promoter that allows the spread of advantageous mutations improving the population ability to explore the adaptive landscape. In addition, has been showed that gene flow helps to reload the genetic variance and purging deleterious mutations mitigating the effect of drift, especially in marginal populations (Slatkin, 1987; Rieseberg & Burke, 2001; Morjan & Rieseberg, 2004; Alleaume-Benharira *et al.*, 2006).

The dynamic of gene flow among populations is conditioned by several factors such species dispersal ability, geographic distance among individuals and populations, physical landscape features, and ecological (biotic and abiotic) factors that differentially delimit suitable portions of the landscape facilitating or constraining the allele movement and establishment (Sexton *et al.*, 2014; Wang & Bradburd, 2014; Cushman *et al.*, 2016). Moreover, identification of asymmetric gene flow patterns in which some populations preferentially act as sources (provide more emigrants than they receive immigrants) whereas others are prone to function as sinks (get many immigrants, but release few emigrants) is essential for the understanding evolutionary processes. Differential gene flow patterns among populations can lead to differential genetic structure patterns or local adaptation processes, and could be pivotal for conservation efforts to identify priority populations in order to keep the persistence of migrants supply (Furrer & Pasinelli, 2015). Both, landscape features (like habitat quality) and populations characteristics (as population size and position within species

geographic range) can determine the source-sink dynamics (Holderegger & Gugerli, 2012; Heinrichs *et al.*, 2016; Sundqvist *et al.*, 2016).

Study area

The South Atlantic coastal Plain (SACP) is a flat, continuous, open region, occupied mostly by large coastal lakes, and constitutes the largest coastal region in South America. The SACP extends NE-SW for approximately 600 km and is crossed just by two perennial water channels (Tomazelli *et al.*, 2000; Weschenfelder *et al.*, 2010). The SACP gradually arose during sea level transgressions and regressions processes caused by glacial-interglacial cycles during the last 400 thousand years. The largest transgression and regression cycles let to the formation of four main sand barriers positioned parallel to the coast line (barrier-lagoon systems I to IV; Tomazelli *et al.*, 2000; Tomazelli & Dillenburg, 2007).

Study system

Calibrachoa heterophylla (Solanaceae) is a diploid (2n=18), perennial, semi-prostrated small shrub, with woody stem base, purplish mellittophilous flowers, with capsular fruits that produce dozen tiny seeds (< 1.4 mm) with no dispersal mechanisms. The species inhabits open sandy grasslands, dunes or rocky outcrops of lakeside or marine environments from the SACP. This is the only species of the genus *Calibrachoa* found in coastal environments. The current known distribution of *C. heterophylla* follows the SACP coastal line from the city of Laguna (Santa Catarina state, Brazil; ~ 28 Lat S) to Cape Polónio (Uruguay; ~34 Lat S). Longitudinally, populations of *C. heterophylla* occur from the seashore to less than 90 km from the coast, with four populations separated from the sea by big lagoons, and other three

populations occurring outside of SACP restricted to a single inland locality of sandbanks environment alongside the Santa Maria River (Rio Grande do Sul state, Brazil; ~ 55 Long W; Fig. 1).

Calibrachoa heterophylla presents a phylogeographical structure with an inland and three coastal intra-specific plastid DNA (cpDNA) lineages likely resulted from isolation processes before the formation of the current continuous topography of the SACP drove by river channels as paleo-barriers that dissected the species ancestral area. Furthermore, the spread process into the SACP to the present-day distribution could have occurred with a sudden population expansion process following the last marine regression (Mäder *et al.*, 2013).

Considering the intra-specific diversification history of *C. heterophylla*, the current geographic distribution, and the present-day physical features of the SACP, we expected that the contemporary patterns of population genetic structure and gene flow could be shaped by processes of secondary contact between previously isolated lineages, as well as be influenced by differences on environmental conditions, or topographic features alongside SACP. In this work, we specifically aim to characterize the inter-population gene flow in terms of geographical proximity, directionality, and intensity, and to test whether differential environmental or topographical conditions of the SACP drove the genetic structure and gene flow patterns of *C. heterophylla*. With this research we expect to contribute to the assembling of a comprehensive picture about the eco-evolutionary patterns involved in the colonization of coastal environments.

Materials and Methods

Sample collections

Throughout all SACP we sampled a total of 253 individuals from 13 localities (hereafter referred as populations; Fig. 1). In addition, three localities (Cacequi 1, Cacequi 2, and São Francisco de Assis; Fig. 1) were sampled in sandy soils along riverbanks of a region about 500 km away from the SACP. The sampled populations represent the entire *C. heterophylla* known distribution. Despite the herbarium records of *C. heterophylla* in Uruguay, the species was not found southern to Santa Vitória do Palmar (Fig. 1).

We collected leaves of all individuals found in each locality during the flowering season between 2002 and 2013 (September to February) and preserved it in silica gel. The number of individuals per population varied from 3 to 41 (Table 1). We collected one herbarium voucher per population and deposited in the ICN Herbarium (Universidade Federal do Rio Grande do Sul) or the BHCB Herbarium (Universidade Federal de Minas Gerais).

Laboratory procedures

The total DNA was extracted following the protocol of Roy *et al.* (1992). Individuals were genotyped for ten anonymous microsatellite loci (Che18, Che59, Che119, Che26, Che34, Che81, Che82, Che85, Che72, and Che126) specifically developed for *C. heterophylla*, following standard protocols for PCR and genotyping procedures as described in Silva-Arias *et al.* (2015).

Characterization of the genetic diversity

Tests for linkage disequilibrium for all locus pairs were performed and deviations from Hardy–Weinberg equilibrium (HWE) were tested within each population for each locus. Significance of HWE deviations was assessed using 10^6 Markov chain

steps and Fisher's exact probability tests in ARLEQUIN 3.5 (Excoffier & Lischer, 2010). The genetic diversity within populations was evaluated by calculating the following statistics: average rarefied allelic richness, private alleles, observed heterozygosity (H_O), expected heterozygosity (H_E), the Garza-Williamson (G-W) index, and inbreeding coefficient (F_{IS}; with confidence limits from 1000 bootstrap resampling over loci) using the packages POPPR 2.0.2 (Kamvar *et al.*, 2014, 2015), HIERFSTAT 0.04-14 (Goudet, 2005, 2014) in R 3.3.0 (R Core Team, 2016), and ARLEQUIN.

Genetic structure

We assessed the genetic structure by means of two model-based clustering methods and three exploratory data analyses (François & Waits, 2016). The model-based clustering methods used were STRUCTURE 2.3.4 (Pritchard *et al.*, 2000) and the spatial Bayesian clustering program TESS 2.3. Those provide estimates for the number of genetic clusters (K) in HWE equilibrium, individual assignment probabilities, and compute the proportion of the genome of each individual that can be assigned to the inferred clusters.

For STRUCTURE analysis the number of clusters evaluated ranged from 1 to the total number of populations (15), with ten independent runs per K-value. Each run was performed using 2.5×10^5 burn-in period and 1×10^6 Markov chain Monte Carlo repeats after burn-in, under an admixture model, assuming correlated allele frequencies (Falush *et al.*, 2003), and including a priori sampling locations as prior (*locprior*) to detect weak population structure. The *locprior* option is not biased toward detecting structure when it is not present and can improve the STRUCTURE results when implemented with few loci (Hubisz *et al.*, 2009). To obtain the K

value(s) that best explain the structure in our genetic dataset we assessed the measures of ΔK method (Evanno *et al.*, 2005) useful to recover the broadest level of genetic structure, and the posterior probability (PP) of each K value calculated following Bayes rule that is prone to recover finer levels of genetic structure. In addition we assessed the threshold based corrections of best K values proposed by Puechmaille (2016) that could be useful to reduce the appearance of spurious clusters. Those K values were obtained by processing the STRUCTURE outputs with the function *structure_analysis* of the KESTIMATOR 1-12 script (Puechmaille, 2016) in R. In addition, to test for spurious clustering results related with the uneven population sampling (Kalinowski, 2011) we did an independent run with exactly same parameters for a random subsampled dataset with maximum size of ten individuals per population. The reduced matrix was generated with the function *subsampFSTAT* of KESTIMATOR.

TESS implements a spatial assignment approach to group individuals into clusters accounting for samples geographic locations, giving to those spatially closer samples in the connection network higher probabilities of belonging to the same genetic cluster than those that are more distant. We did TESS runs of 100 000 generations, with burn-in of 50 000, using the conditional autoregressive (CAR) admixture model, and starting from a neighbour-joining tree. We ran 20 iterations for each value of maxK ranging from 2 to 15. Because we recorded only one coordinate per population, we added small perturbations to the original spatial coordinates using the *generate spatial coordinates for individuals* option with a standard deviation equal to 0.2 to obtain single different coordinates for each individual. Convergence was assessed by the inspection of post-run log-likelihood plots, and support for alternative K values was assessed by inspection of the statistical measure of the model prediction

capability Deviance Information Criterion (DIC; Spiegelhalter *et al.*, 2002). We computed and plotted the average DIC values to detect maxK value at the beginning of a plateau. Replicated runs of best K results for both STRUCTURE and TESS were respectively summarized and plotted with the POPHELPER (Francis, 2016) R package.

The exploratory methods to detect the genetic structure summarize multilocus genotype data sets within a reduced number of synthetic variables that are then used to delineate genetic clusters without any assumptions about the processes that generate the data (François & Waits, 2016). We implemented the exploratory data analyses multivariate method Discriminant Analysis of Principal Components (DAPC; Jombart *et al.*, 2010), the spatial Principal Component Analysis (sPCA; Jombart *et al.* 2008), both implemented in the R package ADEGENET 2.0.1 (Jombart, 2008), and the spatial Factor Analysis (spFA; Frichot *et al.*, 2012).

DAPC estimates the proportion of an individual genome to be originated from a given genetic group or cluster using coefficients of the alleles (loadings) in orthogonal axes, maximizing between-groups variance and minimizing within-group variance in these loadings. For this analysis, the SSR data were first transformed using Principal Component Analysis and keeping all principal components (PCs). After that, the function *find.clusters* was implemented to obtain the optimal number of clusters that maximizes the between-group variability using the lowest Bayesian Information Criterion (BIC) score. Next, to avoid overfitting an optimal reduced number of PCs to retain given the best number of clusters were set for the DAPC using the function *optim.a.score*. Finally, the DAPC of the transformed genetic data was implemented with the number of clusters and PCs to the optimal values and plotted a scatter plot of the two first components with the function *s.class*.

sPCA of genetic structure incorporates spatial information maximizing the product of spatial autocorrelation (Moran's I) and variance of each eigenvector. This produces orthogonal axes that describe spatial patterns of genetic variation. The spatial information is included in the analysis by means of a spatial weighting matrix derived from a connection network. In order to test the effect of the neighbouring definition on the results, we ran the sPCA using six different connection networks available in the function *chooseCN*. For these analyses we used the same perturbed coordinates used in TESS analysis. Monte Carlo simulations (global and local tests) were used with 10 000 permutations to test for non-random spatial association of population allele frequencies for all sPCA implemented.

spFA uses the geographical information in an explicit way to infer the population genetic structure breaking the geographical correlation between proximal individuals. The main purpose of this analysis is to obtain synthetized factors of genetic variation accounting for the statistical artefacts generated by the decay of genetic similarity with geographical distance (IBD) using an inverse spatial correlation matrix as weights to remove spatial correlation. The main expected outcome is that important structure features likely to be masked by smooth spatial patterns could be better identified. The analysis was implemented with the *spfa* code in R available at http://membres-timc.imag.fr/Olivier.Francois/spfa.R. For this analysis, we used the same perturbed coordinates generated in TESS and the resulting factors were plotted in a scatter plot. We ran the spFA analysis using number of factors K = 2 to K = 8. Accounting that different values of the scale parameter (i.e., the scale at which the IBD effect is apparent) are prone to led to different interpretations of the genetic structure (Frichot *et al.*, 2012), values of 0.1, 0.5, 1, and 10 for the scale parameter were explored.

Historical and contemporary gene flow estimations

Contemporary asymmetric migration rates were estimated using a Bayesian approach implemented in BAYESASS 3.0 (Wilson & Rannala, 2003). We ran 10^8 iterations and a burn-in of 10^7. Following the BAYESASS manual the adjustment for mixing of continuous parameters were optimized to obtain acceptance rates around 40% to adequately explore the state space, for that mixing parameters of allele frequencies, inbreeding coefficients, and migration rates were set to 0.80, 0.95, and 0.30, respectively. Convergence was assessed by examining the plots of the log-probability and the effective sample sizes for each run using TRACER 1.6 (Rambaut *et al.*, 2014) and looking for consistency of the migration estimates between four independent runs initialized with different seed numbers.

Historical gene flow patterns were assessed testing the support of our genetic dataset to four alternative gene flow scenarios using Bayes factors calculated from the marginal likelihood approximations of each migration model (Beerli & Palczewski, 2010). The coalescent-based program MIGRATE-N 3.2.6 (Beerli & Felsenstein, 2001) was used to estimate the mutation scaled population size (Θ) and a mutation scaled migration (M) parameters. The Bézier log marginal likelihood approximations were used to assess the support of each proposed migration model.

For every models, the populations were pooled into four groups according to the geographical distribution and genetic structure results (Figs. 1, 2, 3) named as: 'Inland' (including São Francisco de Assis, Cacequi 1, and Cacequi 2 populations); 'West' (including Arambaré, Barra do Ribeiro, and Pelotas populations); 'North' (including Santo Antônio da Patrulha, Torres, and Laguna populations); and 'South'

(including Mostardas 1, Mostardas 2, São José do Norte 1, São José do Norte 2, Rio Grande, and Santa Vitória do Palmar populations).

Four migration models were evaluated named as: (1) *source-sink from inland* with unidirectional migration from 'Inland' population group to the remaining groups; (2) *source-sink from west* with unidirectional migration from 'West' population group to the remaining groups; (3) *step stone from inland* with unidirectional migration from 'Inland' to 'West' and from 'West' to 'North' and 'South'; and (4) *step stone from coast* with unidirectional migration from 'North' to 'West', 'South' to 'West' and from 'West' to 'Inland' (for models graphical representations see Fig. S1).

For each model, we ran a bayesian inference of MIGRATE-N in the CIPRES Science Gateway 3.3 (Miller *et al.*, 2010) with one long chain of 5×10^6 steps, sampling at every 100^{th} increment, and a burn-in of 3×10^4 steps. We used uniform priors and slice sampling for both Θ_W and M ranging from 0 to 20 (mean = 10, delta = 0.5). A heating scheme was used (Metropolis-coupled Markov Chain Monte Carlo) with four parallel chains with temperatures of 1, 1.5, 3, and 10^6.

Space, environment, and genetic differentiation

Pairwise F_{ST} (Weir & Cockerham, 1984) matrix of genetic differentiation was calculated with SPAGEDI 1.5 (Hardy & Vekemans, 2002). The geographic inter-population distance matrix was obtained by calculating the linear Euclidean distance between X and Y UTM 22S (reference EPSG: 32722) populations' coordinates transformed from Long/Lat coordinates with RGDAL 1.0-4 (Bivand *et al.*, 2015) R package.

It has been shown that patterns of spatial autocorrelation due to isolation by distance (IBD) provoke bias for several genetic structure tests (Frantz *et al.*, 2009;

Meirmans, 2012). Therefore, the hypothesis of IBD was assessed by linear regression of linearized pairwise F_{ST} genetic distances and logarithmic transformed geographic distances (Rousset, 1997) and was analysed using Mantel tests, which significance was assessed through 10 000 randomizations in the R package VEGAN 2.3-0 (Oksanen *et al.*, 2015). In order to test whether the three highly distant populations located outside of the SACP could bias the distance-genetic models, both linear regressions and Mantel tests were also implemented excluding the populations Cacequi 1, Cacequi 2, and São Francisco de Assis located outside of the SACP (Fig. 1).

We also tested isolation by environment (IBE) models in order to examine whether differences in climatic conditions explain the observed inter-population genetic differentiation in *C. heterophylla*. Climatic dissimilarity matrices between each pair of studied populations were calculated for the climatic variables total annual precipitation, total annual days with rain, precipitation seasonality, mean annual temperature, mean summer maximum temperature, mean winter minimum temperature, mean temperature range, and temperature seasonality. The values of those climatic variables were taken from raster layers specifically developed for the SACP derived from a high dense sampling of climatic stations through the region, geostatistical modelling and spatial interpolation (Silva-Arias *et al.*, submitted).

Moreover, isolation by resistance (IBR) models based on circuit theory were assessed to test for possible models of inter-population differentiation linked to landscape discontinuities alongside the SACP. Using raster grids we outlined three different connectivity models (Fig. S2): (1) the *continuous* (or null) model wherein no landscape discontinuity affects the inter-population connectivity. For that we created a raster grid with all cells values equal to 1, including all cells on freshwater surfaces. This model is expected to resemble a Euclidean geographical distance but is more

properly for comparisons with models based on circuit theory. (2) The *topographic* model proposes that freshwater bodies widespread in the SACP restrict the connectivity between populations. For that, we created a raster grid with all land cells values equal to 1, and cells within freshwater surfaces as complete barriers (*nodata*). And (3) the *environmental* model suggests that connectivity between populations is enhanced with the species' habitat suitability. For that we obtained a raster grid with a projection of an ensemble niche model (Araujo & New, 2007) obtained for *C. heterophylla* implemented in the R package BIOMOD2 (Thuiller *et al.*, 2014) using as explanatory variables the same eight climatic surfaces assessed in the IBE tests (see Appendix S1 for details). We generated pairwise connectivity matrices with CIRCUITSCAPE 4.0.5 (McRae, 2006) considering an eight-neighbour cell connection scheme, using as nodes the Long/Lat coordinates of each population and raster resolution of ~ 0.5 km.

The relationships between genetic differentiation (F_{ST}) and matrices of geographical distance (IBD), environmental dissimilarity (IBE), and landscape resistance (IBR) were examined using multiple matrix regressions with randomization (MMRR) through the MMRR script (Wang, 2013) implemented in R.

Results

Genetic diversity

We found 140 alleles considering the 10 microsatellite loci, with an average of 14 alleles per locus, ranging from 7 (Che59) to 17 (Che81). All loci showed higher H_E than H_O (Fig. S3). Furthermore, 25% of the locus-population combinations showed departure of HWE ($P < 0.05$). A significant linkage disequilibrium signal ($P < 0.01$) was detected for several pairs of loci, but since the linkage pattern was not consistent

across populations for any loci pair, we assumed linkage equilibrium and maintained all loci for the downstream analyses. MICROCHECKER analysis showed no evidence of null alleles, scoring errors, or stutter peaks for any locus.

Genetic diversity statistics estimated for each population over all loci presented higher values in populations located outside of SACP, as well as those located at the west side of the Patos Lagoon (Fig. 1). On the other hand, lower genetic diversity values were recovered for coastal populations located at the northern and southern distribution edges of the species distribution at the SACP (Laguna and Santa Vitória do Palmar populations, respectively; Fig. 1). Average values across loci of H_O ranged from 0.72 (São Francisco de Assis) to 0.31 (Laguna) and H_E from 0.74 (Cacequi 2) to 0.48 (São José do Norte 1). 22% of recovered alleles were exclusive, Barra do Ribeiro population showed the highest amount of private alleles (8), whereas Pelotas, Mostardas 1, São José do Norte 1, and São José do Norte 2 populations presented no private alleles. G-W values ranged from 0.39 (Cacequi 1) to 0.83 (Torres). Positive and significant inbreeding coefficients (F_{IS}) were found for five populations (Table 1) all of them located at the borders (northern and southern) of species distribution in the SACP (Fig. 1).

Genetic structure

The best K inferred with the ΔK method for STRUCTURE analysis for the complete data set, with and without the threshold based corrections, detected consistently K = 4 as the best grouping pattern, while for the subsampled dataset was K = 3 (Fig. S4). Meanwhile, the PP K-estimator method as expected returned higher number of groups, but with a no consistent pattern. For the complete data set the PP K estimator proposed a best K = 14 and the corrected values ranged from 8 to 12 depending of the

threshold level. For the subsampled dataset we obtained a best K = 13 and the corrected values ranged from 6 to 12 (Fig. S4). On the other hand, the curve of average DIC values for the repeated TESS analyses showed a plateau after maxK = 8, however runs of maxK = 2, 3, and 4 showed the lowest standard deviations (Fig. S5). Accounting for these results we analysed the barplots of membership probability of each individual for K = 2 to K = 8.

The recovered population structure had a strong geographic signal. Genetic clustering obtained with STRUCTURE and TESS showed similar results. For both approaches, assignment probabilities obtained for K = 2 showed that the main differentiation delimited one cluster composed by populations located at northern SACP, and a second group composed by remain populations (Figs. 2 and 3). When K = 3, one group still clustering the northern coastal populations, the second group encompassed the southern coastal populations, whereas the third group brought together all populations located on west side of the Patos Lagoon and the three inland populations (Figs. 2 and 3). Higher values of K showed differences between the two Bayesian clustering methods in the order that groups were incorporated, however a common clustering pattern was recovered for both approaches for K = 8 (Figs. 2 and 3), showing one group composed by the three inland populations; other two groups composed solely by one population each one (Barra do Ribeiro and Pelotas populations, respectively, both located west of Patos Lagoon); the forth cluster grouped two populations from the northern SACP (Santo Antonio da Patrulha and Torres populations); individuals from Laguna population (the northernmost occurrence site) were grouped in a fifth cluster; the sixth group was composed by São José do Norte 1 and São José do Norte 2 populations; and individuals from the southernmost population (Santa Vitória do Palmar) kept a separated group. The eighth

group was not preferentially linked to any population, instead showed low membership probabilities in several populations. We found higher admixed membership for populations located in geographical transition regions. The most outstanding case were Mostardas 1 and Mostardas 2 populations that are located between north and south SACP regions and all individuals from these populations showed membership probability assigned for all groups except for that from the inland populations (Figs. 2 and 3). Rio Grande population also showed high admixture membership and achieved some discrepancy between Bayesian assignment tests, while STRUCTURE suggested that present around of 60% of membership for a group mostly shared with Torres population, TESS supported a higher membership probability preferentially for a single group shared with Santa Vitória do Palmar.

Exploratory analyses let us to identify several common patterns obtained with Bayesian clustering analyses. Although the lowest BIC score was found for K = 8 (Fig. S6), DAPC scatterplot of the two main discriminant components showed a main differentiation in three groups delimiting northern and southern coastal populations in two different groups, and a third group compose by inland, Patos Lagoon western side, and the two Mostardas populations (Fig. 1). sPCA recovered a main differentiation of the coastal northern and southern edge populations on first sPC axis. Torres and Santo Antônio de Patrulha populations from northern SACP, as well as Santa Vitória do Palmar from southern SACP were opposed on the first sPC axis, while the remain populations showing a gradient of differentiation on the second sPC axis (Fig. 1). spFA showed similar result and this analysis appeared more sensitive to the number of factors (K) than to the variation in the parameter scale. Lower K values collapsed the most populations producing a single group in the middle of the two factors and just differentiating the populations from the edges (northern and southern)

as ‘satellites’ around of the main collapsed group in the spFA scatterplot (not shown). With the increase of the number of factors up to K = 5 was possible to detect the continuous arrange of differentiation in the factor 2 similar to that observed in sPCA. No different results were found increasing the number of factors above of K = 5.

Migration rate estimations

BAYESASS estimations showed a main asymmetric pattern of contemporary gene flow wherein the São Francisco de Assis, Arambaré, and Mostardas 1 populations mainly functioning as receptors from remaining populations, while the Santo Antônio da Patrulha and Santa Victoria do Palmar populations presented the lowest immigration estimates (Fig. 4). Highest mean migration estimates were obtained between geographically closer populations, but with asymmetric patterns. From São José do Norte 2 to São José do Norte 1 (m = 0.155) and in the opposite way (m = 0.013); from Cacequi 1 to Cacequi 2 (m = 0.123) and in the opposite way (m = 0.014); and from Mostardas 2 to Arambaré (m = 0.077) and in the opposite way (m = 0.012). Lowest estimates were obtained for migration from all populations to Santa Vitória do Palmar (m = 0.006), and to Santo Antônio da Patrulha (m = 0.006 - 0.011; Fig. 4). Migration estimates obtained from independent runs of BAYESASS showed similar results (Fig. S7). Moreover, trace profile plots for log-probability post-burn-in values of all independent runs reached a plateau, and all migration estimates showed effective sample sizes (ESS) higher than 400, except for migration between populations 2 → 1, 6 → 1, and 8 → 1 (see Fig. 1 or Table 1 for population numbers) that presented ESS between 50 - 100.

The model-based coalescent approach implemented in MIGRATE-N supported that the model *step stone from coast* was the most likely historical migration pattern

between population groups (Table 2, Fig. S1). Parameter estimation taken from the best-supported model showed that 'Inland' group presented the highest mean mutation scaled population size (Θ), which was around eight times higher than the Θ estimated for 'West' and 'North' groups, and around of 20 times higher than the Θ estimated for 'South' group that presented the lowest value (Table 2). The highest mean mutation scaled migration rate was recovered from 'West' to 'Inland' group that was twice higher than the 'North' to 'West' and five times when compared to the 'South' to 'West' values (Table 2).

Isolation by distance, environment and resistance tests

Measures of population differentiation F_{ST} ranged from 0.01 between Mostardas 1 and Mostardas 2 populations to 0.54 between Laguna and São José do Norte 1 populations (Fig. 5). Linear regressions showed significant positive relationship between linearized geographic and genetic distances (R-squared = 0.14; $P < 0.001$) supporting a pattern of IBD for our data set (Fig. 6). Mantel test also supported the relationship between the geographic and the genetic distance matrices (Mantel´s r = 0.38, $P < 0.001$). Spatial tests of genetic differentiation including only the 12 populations located in the SACP also resulted in positive and significant relationships with an adjusted R-squared = 0.192 ($P < 0.001$), and Mantel's r = 0.4521 ($P = 0.003$).

Analyses testing the relationship among geographic, climatic variables, and genetic differentiation by means of MMRR approach showed significant association of genetic dissimilarity with Euclidean distances (R-squared = 0.16, $\beta = 2.4 \times 10^{-7}$, $P = 0.007$) when regressed with F_{ST} as the unique explanatory variable. MMRR models implemented with each of the assessed climatic variables and Euclidean distances as explanatory variables showed only significant relationship for precipitation

seasonality (R-squared = 0.32, P < 0.006; $\beta_{precseason}$ = 0.01, P = 0.038; and $\beta_{Euc} = 2.8 \times 10^{-7}$, P = 0.084, respectively). IBR tests comparing three landscape models based on circuit theory showed that the *continuous* model considering equal resistance along all species' geographic range best explained the genetic differentiation estimated with F_{ST} (R-squared = 0.35, β = 0.12, P = 0.002).

Discussion

In this study we estimated measures of population genetic diversity and structure with the purpose of detect geographical and environmental influences on the inter-population gene flow, and their relation with the colonization process in a recently deposited coastal plain using as biological model the small shrub species *Calibrachoa heterophylla*. Our results support that distance is the main factor shaping the inter-population genetic differentiation as well as an IBE process linked to differential conditions in precipitation seasonality in marginal coastal populations. Remarkably, this IBE pattern recovered for *C. heterophylla* seems to depict a congruent evolutionary picture previously reported for another co-distributed coastal plant lineage (*Petunia integrifolia* ssp. *depauperata*; Silva-Arias *et al.,* submitted) and could represent an emerging community trend in the colonization and spreading processes in the SACP.

Furthermore, our genetic data set provides consistent evidence of an asymmetric gene flow dynamic wherein populations located on northern and southern edges of the species distribution present restricted historical and contemporary immigration rates.

Insights of local adaptation process on edge coastal populations

We observed differential conditions in precipitation seasonality for northern and southern extremes of SACP (Fig. 5). Join to that, significant positive relationship verified for precipitation seasonality and genetic differentiation matrices, and the fact that immigration for edge SACP populations fall within the lowest estimates (Fig. 4, Table 2) let us to consider the possibility of a genetic divergence process driven by differential rainfall regimes alongside SACP. Ecologically based divergence driven by environmental conditions can promote selection against immigrants that lead, in a genome-wide context, a reduction of gene flow, reproductive isolation, and enhances the stochastic effects of genetic drift (Hendry, 2004; Nosil *et al.*, 2005, 2009).

Limited levels migration rates from the central to the marginal distribution could enhance the establishment of locally adaptive alleles in peripheral populations by preventing gene swamping (Alleaume-Benharira *et al.*, 2006). Even so, several studies have showed many beneficial effects of migration enhancing local adaptation processes. Within them we highlight the increasing of genetic variation diminishing deleterious effects generated by inbreeding and drift (Sexton *et al.*, 2009), and the enrichment of adaptive alleles emerged in central populations and their subsequent movement by gene flow to the edges where local adaptation can occur (Rolland *et al.*, 2015). In addition, intermediate slope environmental gradients have been proposed to facilitate local adaptation at the distribution edges in the presence of gene flow (Doebeli & Dieckmann, 2003; Bridle *et al.*, 2009).

Differential rainfall conditions between populations during the time span of germination until reproduction (spring-summer seasons for *C. heterophylla*) could affect several plant traits that directly interact with the immigrant or hybrid survival and reproduction, such as seedling recruitment and flowering time. We now encourage reciprocal transplant experiments in order to verify local adaptation

processes in *C. heterophylla* alongside their geographical distribution by means of specialized approaches (Blanquart *et al.*, 2013), and given the case, experiments to test the possible relationship with differential precipitation seasonality regimes.

Patos Lagoon development let a secondary contact of previously diverged lineages

One of the most outstanding results is that populations located close around of the Patos Lagoon (i.e., Arambaré, Mostardas 1, Mostardas 2, São José do Norte 1, São José do Norte 2, and Rio Grande; Fig. 1) were those that presented the highest levels of genetic admixture (Figs. 2 and 3) and lowest F_{ST} values among the coastal populations (Fig. 5). These results could be related with a recent dynamic geomorphological history of SAPC region. During the most part of the Quaternary Period, the Patos Lagoon area was mainly crossed by two rivers (Jacuí and Camaquã rivers), as well as several channels corresponding to their temporal dynamic delta systems (Weschenfelder *et al.*, 2014). After the two most recent ocean regressive-transgressive events that gave rise to the youngest SACP strips (barrier systems III and IV), a complete closure by sands of the inlets and the complete isolation of Patos Lagoon from the sea process was concluded (Santos-Fischer *et al.*, 2016).

This recent dynamic could be an obstacle for the population establishment in the east side of the Patos Lagoon, contrary to the northern and southern SACP regions where previously well established barrier systems I and II (cf. Fig. 1 in Tomazelli & Dillenburg, 2007) could let the establishment and differentiation of main coastal lineages, that later spread and could experience a secondary contact on east side of the Patos Lagoon generating the current patterns of genetic admixture in this region. This recent admixture processes can also explain the lack of private alleles in Pelotas, São José do Norte 1, São José do Norte 2, and Mostardas 1 populations (Fig. 1).

Moreover, east side of Patos Lagoon region presents the strongest wind influence within the SACP during most part of the year (Martinho *et al.*, 2010). This continuous dynamic could increase the secondary seed dispersal alongside that region, generating in consequence higher admixture patterns, since the seeds in *Calibrachoa* are light and disperses through barochoric system (van der Pijl, 1982).

Despite of the differences in the intra-specific divergence history recovered in phylogeographical approaches for *C. heterophylla* (Mäder *et al.*, 2013) and the co-distributed species *P. integrifolia* ssp. *depauperata* (Ramos-Fregonezi *et al.*, 2015), both species shared same pattern of high genetic admixture for microsatellite data in populations located east of Patos Lagoon (Silva-Arias *et al.,* submitted). This congruent pattern supports that the above-discussed current and historical dynamics in the environmental conditions of the SACP could be the main drivers of the observed admixture patterns in the genetic structure of the species in this region.

Genetic structure of Calibrachoa heterophylla

A hierarchical pattern of genetic structure related with historical and geographical features, as well as population specific processes, emerged from our results of population structure analyses (Figs. 1, 2, and 3). The broadest pattern presents three well-supported groups, northern, southern, and west Patos Lagoon populations, the last group also including inland populations. This main clustering pattern is historically congruent because mirrors the phylogeographical structure of *C. heterophylla* recovered by cpDNA data (Mäder *et al.*, 2013). To find historical signals of genetic structure with molecular markers able to reflect contemporary patterns (e.g., microsatellites) is expected for studies involving the entire geographic range of the species and for this reason have been suggested to account for historical patterns

in landscape genetic approaches (Anderson *et al.*, 2010). As previously discussed, populations located between the northern and southern regions (i.e., Mostardas), or between west Patos Lagoon and southern regions (i.e., Pelotas) showed higher admixture values possibly reflecting recent gene flow dynamic between previously differentiated groups.

A geographical frame in the genetic structure was evident in several levels of the observed genetic structure. The northern, southern, and central groups showed a main latitudinal pattern. In finer scale, we found that individuals from peripheral populations tend to make up separated groups with membership probability higher that 80%, as we can observe in Santa Vitória do Palmar population, in the three inland populations (São Francisco de Assis, Cacequi 1, and Cacequi 2), and in Laguna locality (Figs. 1, 2, and 3). The assignment of individuals from these populations in completely separated groups can be mainly explained by genetic drift due to a stronger geographical isolation (Whigham *et al.*, 2008).

Few departures from the geographical frame in the genetic structure could be observed in some distant populations that the assignment tests suggested to share partial genetic identity such Laguna-São José do Norte and Torres-Rio Grande population pairs (Figs. 2 and 3). In both cases populations from northern and southern SACP regions are involved. We infer that partial genetic affinity between these spatially distant populations could be better explained as a process of common allele fixation during a expansion wave (Excoffier & Ray, 2008), persistent ancestral variation after population divergence, or homoplasy (Grimaldi & Crouau-Roy, 1997; Oppen *et al.*, 2000; Schaal & Olsen, 2000) instead of long distance gene flow. Higher F_{ST} measures (Fig. 5) and lower migration rates (Fig. 4) between these populations support this suggestion.

Pelotas population showed several contrasting results of genetic clustering depending of the methodology and structure level inferred. Exploratory and STRUCTURE analyses grouped Pelotas with the inland populations (Figs. 1 and 2), while TESS suggested that Pelotas could more related with the southern population group (Fig. 3). For finer population structure patterns (i.e. K=5-8) both STRUCTURE and TESS suggested higher affinity of Pelotas population with São José do Norte or separated a group lonely for Pelotas with almost 100% of membership probability for all individuals (Figs. 2 and 3). These results could be related with both the geographic position of Pelotas population and environmental dynamic features. Regarding to the geographic position, Pelotas population is located close to southern coastal populations such Rio Grande and São José do Norte, but separated from the seashore by the Patos Lagoon. These features could shape a population subject to receive migrants from coastal population but with an environment with more continental affinity. Nevertheless, inter-annual rainfall conditions, as well as continental scale periodic climatic fluctuations, such as El Niño, can both affect the fluvial discharge and the wind action responsible for the salinization and desalination processes of the Patos Lagoon (Möller *et al.*, 1996, 2001; Möller & Castaing, 1999). These environmental dynamic conditions should continuously change the establishment or survival rates of individuals with coastal or continental gene pools probably leading to the differential genetic observed profile.

Acknowledgments

The authors thank to G. Mäder for help in fieldwork. This project was supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), and

Programa de Pós-graduação em Genética e Biologia Molecular da Universidade Federal do Rio Grande do Sul (PPGBM-UFRGS). G.A.S.-A. was supported by a scholarship from the Francisco José Caldas Institute for the Development of Science and Technology (COLCIENCIAS). This work was conducted under permit MP 2.186/16 of the Brazilian Federal Government to access plant genetic information to develop evolutionary or taxonomic studies. No specific collection permits were required because this species are not federally listed as endangered or protected and because no collection sites occurred in protected areas.

References

Alleaume-Benharira, M., Pen, I.R. & Ronce, O. 2006. Geographical patterns of adaptation within a species' range: interactions between drift and gene flow. *J. Evol. Biol.* **19**: 203–215.

Anderson, C.D., Epperson, B.K., Fortin, M.-J., Holderegger, R., James, P.M.A., Rosenberg, M.S., *et al.* 2010. Considering spatial and temporal scale in landscape-genetic studies of gene flow. *Mol. Ecol.* **19**: 3565–3575.

Araujo, M. & New, M. 2007. Ensemble forecasting of species distributions. *Trends Ecol. Evol.* **22**: 42–47.

Beerli, P. & Felsenstein, J. 2001. Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. *Proc. Natl. Acad. Sci.* **98**: 4563–4568.

Beerli, P. & Palczewski, M. 2010. Unified framework to evaluate panmixia and migration direction among multiple sampling locations. *Genetics* **185**: 313–326.

Bivand, R., Keitt, T. & Rowlingson, B. 2015. *RGDAL: Bindings for the Geospatial Data Abstraction Library. R package version 0.9-3. http://CRAN.R-project.org/package=rgdal*.

Blanquart, F., Kaltz, O., Nuismer, S.L. & Gandon, S. 2013. A practical guide to measuring local adaptation. *Ecol. Lett.* **16**: 1195–1205.

Bridle, J.R., Gavaz, S. & Kennington, W.J. 2009. Testing limits to adaptation along altitudinal gradients in rainforest *Drosophila*. *Proc. R. Soc. B Biol. Sci.* **276**: 1507–1515.

Cushman, S.A., H. McRae, B. & McGarigal, K. 2016. Basics of landscape ecology: An introduction to landscapes and population processes for landscape geneticists. In:

Landscape genetics: concepts, methods, applications (N. Balkenhol et al., eds), pp. 9–34. John Wiley & Sons, Ltd, Chichester, UK.

Doebeli, M. & Dieckmann, U. 2003. Speciation along environmental gradients. *Nature* **421**: 259–264.

Evanno, G., Regnaut, S. & Goudet, J. 2005. Detecting the number of clusters of individuals using the software structure: A simulation study. *Mol. Ecol.* **14**: 2611–2620.

Excoffier, L. & Lischer, H.E.L. 2010. ARLEQUIN suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* **10**: 564–567.

Excoffier, L. & Ray, N. 2008. Surfing during population expansions promotes genetic revolutions and structuration. *Trends Ecol. Evol.* **23**: 347–351.

Falush, D., Stephens, M. & Pritchard, J.K. 2003. Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics* **164**: 1567–1587.

Francis, R.M. 2016. POPHELPER: An R package and web app to analyse and visualise population structure. *Mol. Ecol. Resour.* n/a-n/a.

François, O. & Waits, L.P. 2016. Clustering and assignment methods in landscape genetics. In: *Landscape genetics: concepts, methods, applications* (N. Balkenhol et al., eds). John Wiley & Sons, Inc.

Frantz, A.C., Cellina, S., Krier, A., Schley, L. & Burke, T. 2009. Using spatial Bayesian methods to determine the genetic structure of a continuously distributed population: clusters or isolation by distance? *J. Appl. Ecol.* **46**: 493–505.

Frichot, E., Schoville, S., Bouchard, G. & François, O. 2012. Correcting principal component maps for effects of spatial autocorrelation in population genetic data. *Front. Genet.* **3**: 254.

Furrer, R.D. & Pasinelli, G. 2015. Empirical evidence for source-sink populations: a review on occurrence, assessments and implications: Source-sink dynamics in animals. *Biol. Rev.*, doi: 10.1111/brv.12195.

Goudet, J. 2005. HIERFSTAT, a package for R to compute and test hierarchical F-statistics. *Mol. Ecol. Notes* **5**: 184–186.

Goudet, J. 2014. *HIERFSTAT: Estimation and tests of hierarchical F-statistics. R package version 0.04-14. http://CRAN.R-project.org/package=hierfstat*.

Grimaldi, M.-C. & Crouau-Roy, B. 1997. Microsatellite allelic homoplasy due to variable flanking sequences. *J. Mol. Evol.* **44**: 336–340.

Hardy, O.J. & Vekemans, X. 2002. SPAGEDI: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Mol. Ecol. Notes* **2**: 618–620.

Heinrichs, J.A., Lawler, J.J. & Schumaker, N.H. 2016. Intrinsic and extrinsic drivers of source-sink dynamics. *Ecol. Evol.* **6**: 892–904.

Hendry, A.P. 2004. Selection against migrants contributes to the rapid evolution of ecologically dependent reproductive isolation. *Evol. Ecol. Res.* **6**: 1219–1236.

Holderegger, R. & Gugerli, F. 2012. Where do you come from, where do you go? Directional migration rates in landscape genetics. *Mol. Ecol.* **21**: 5640–5642.

Hubisz, M.J., Falush, D., Stephens, M. & Pritchard, J.K. 2009. Inferring weak population structure with the assistance of sample group information. *Mol. Ecol. Resour.* **9**: 1322–1332.

Jombart, T. 2008. ADEGENET: A R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**: 1403–1405.

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

Jombart, T., Devillard, S., Dufour, A.-B. & Pontier, D. 2008. Revealing cryptic spatial patterns in genetic variability by a new multivariate method. *Heredity* **101**: 92–103.

Kalinowski, S.T. 2011. The computer program STRUCTURE does not reliably identify the main genetic clusters within species: simulations and implications for human population structure. *Heredity* **106**: 625–632.

Kamvar, Z.N., Brooks, J.C. & Grünwald, N.J. 2015. Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. *Front. Genet.* **6**.

Kamvar, Z.N., Tabima, J.F. & Grünwald, N.J. 2014. POPPR: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* **2**: e281.

Mäder, G., Fregonezi, J.N., Lorenz-Lemke, A.P., Bonatto, S.L. & Freitas, L.B. 2013. Geological and climatic changes in Quaternary shaped the evolutionary history of *Calibrachoa heterophylla*, an endemic South-Atlantic species of petunia. *BMC Evol. Biol.* **13**: 178.

Martinho, C.T., Hesp, P.A. & Dillenburg, S.R. 2010. Morphological and temporal variations of transgressive dunefields of the northern and mid-littoral Rio Grande do Sul coast, Southern Brazil. *Geomorphology* **117**: 14–32.

McRae, B.H. 2006. Isolation by resistance. *Evolution* **60**: 1551–1561.

Meirmans, P.G. 2012. The trouble with isolation by distance. *Mol. Ecol.* **21**: 2839–2846.

Miller, M.A., Pfeiffer, W. & Schwartz, T. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: *Proceedings of the Gateway Computing Environments Workshop (GCE)*, pp. 1–8. New Orleans.

Möller, O.O. & Castaing, P. 1999. Hydrographical characteristics of the estuarine area of Patos Lagoon (30°S, Brazil). In: *Estuaries of South America* (G. M. E. Perillo et al., eds), pp. 83–100. Springer Berlin Heidelberg, Berlin, Heidelberg.

Möller, O.O., Castaing, P., Salomon, J.-C. & Lazure, P. 2001. The influence of local and non-local forcing effects on the subtidal circulation of Patos Lagoon. *Estuaries* **24**: 297–311.

Möller, O.O., Lorenzzentti, J.A., Stech, J. & Math, M.M. 1996. The Patos Lagoon summertime circulation and dynamics. *Cont. Shelf Res.* **16**: 335–351.

Morjan, C.L. & Rieseberg, L.H. 2004. How species evolve collectively: implications of gene flow and selection for the spread of advantageous alleles. *Mol. Ecol.* **13**: 1341–1356.

Nosil, P., Funk, D.J. & Ortiz-Barrientos, D. 2009. Divergent selection and heterogeneous genomic divergence. *Mol. Ecol.* **18**: 375–402.

Nosil, P., Vines, T.H. & Funk, D.J. 2005. Reproductive isolation caused by natural selection against immigrants from divergent habitats. *Evolution* **59**: 705–719.

Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., *et al.* 2015. *VEGAN: Community ecology package. R package version 2.3-0. http://CRAN.R-project.org/package=vegan*.

Oppen, M.J.H. van, Rico, C., Turner, G.F. & Hewitt, G.M. 2000. Extensive homoplasy, nonstepwise mutations, and shared ancestral polymorphism at a complex microsatellite locus in lake Malawi cichlids. *Mol. Biol. Evol.* **17**: 489–498.

Pritchard, J.K., Stephens, M. & Donnelly, P. 2000. Inference of population structure using multilocus genotype data. *Genetics* **155**: 945–959.

Puechmaille, S.J. 2016. The program STRUCTURE does not reliably recover the correct population structure when sampling is uneven: sub-sampling and new estimators alleviate the problem. *Mol. Ecol. Resour.* **16**: 608–627.

R Core Team. 2016. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

Rambaut, A., Suchard, M.A., Xie, D. & Drummond, A.J. 2014. *TRACER v1.6, Available from http://beast.bio.ed.ac.uk/Tracer*.

Ramos-Fregonezi, A.M., Fregonezi, J.N., Cybis, G.B., Fagundes, N.J., Bonatto, S.L. & Freitas, L.B. 2015. Were sea level changes during the Pleistocene in the South

Atlantic Coastal Plain a driver of speciation in *Petunia* (Solanaceae)? *BMC Evol. Biol.* **15**: 92.

Rieseberg, L.H. & Burke, J.M. 2001. The biological reality of species: gene flow, selection, and collective evolution. *Taxon* **50**: 47–67.

Rolland, J., Lavergne, S. & Manel, S. 2015. Combining niche modelling and landscape genetics to study local adaptation: A novel approach illustrated using alpine plants. *Perspect. Plant Ecol. Evol. Syst.* **17**: 491–499.

Rousset, F. 1997. Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics* **145**: 1219–1228.

Roy, A., Frascaria, N., MacKay, J. & Bousquet, J. 1992. Segregating random amplified polymorphic DNAs (RAPDs) in *Betula alleghaniensis*. *Theor. Appl. Genet.* **85**: 173–180.

Santos-Fischer, C.B. dos, Corrêa, I.C.S., Weschenfelder, J., Torgan, L.C. & Stone, J.R. 2016. Paleoenvironmental insights into the Quaternary evolution of the southern Brazilian coast based on fossil and modern diatom assemblages. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **446**: 108–124.

Schaal, B.A. & Olsen, K.M. 2000. Gene genealogies and population variation in plants. *Proc. Natl. Acad. Sci.* **97**: 7024–7029.

Sexton, J.P., Hangartner, S.B. & Hoffmann, A.A. 2014. Genetic isolation by environment or distance: Which pattern of gene flow is most common? *Evolution* **68**: 1–15.

Sexton, J.P., McIntyre, P.J., Angert, A.L. & Rice, K.J. 2009. Evolution and ecology of species range limits. *Annu. Rev. Ecol. Evol. Syst.* **40**: 415–436.

Silva-Arias, G.A., Mäder, G., Bonatto, S.L. & Freitas, L.B. 2015. Novel Microsatellites for *Calibrachoa heterophylla* (Solanaceae) Endemic to the South Atlantic Coastal Plain of South America. *Appl. Plant Sci.* **3**: 1500021.

Slatkin, M. 1987. Gene flow and the geographic structure of natural populations. *Science* **236**: 787–792.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P. & van der Linde, A. 2002. Bayesian measures of model complexity and fit. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **64**: 583–639.

Sundqvist, L., Keenan, K., Zackrisson, M., Prodöhl, P. & Kleinhans, D. 2016. Directional genetic differentiation and relative migration. *Ecol. Evol.*, doi: 10.1002/ece3.2096.

Thuiller, W., Georges, D. & Engler, R. 2014. *BIOMOD2: Ensemble platform for species distribution modeling. R package version 3.1-62/r677. http://R-Forge.R-project.org/projects/biomod/*.

Tomazelli, L.J. & Dillenburg, S.R. 2007. Sedimentary facies and stratigraphy of a last interglacial coastal barrier in south Brazil. *Mar. Geol.* **244**: 33–45.

Tomazelli, L.J., Dillenburg, S.R. & Villwock, J.A. 2000. Late Quaternary geological history of Rio Grande do Sul coastal plain, southern Brazil. *Rev. Bras. Geociências* **30**: 474–476.

van der Pijl, L. 1982. *Principles of dispersal in higher plants*. Springer Berlin Heidelberg, Berlin, Heidelberg.

Wang, I.J. 2013. Examining the full effects of landscape heterogeneity on spatial genetic variation: A multiple matrix regression approach for quantifying geographic and ecological isolation: special section. *Evolution* **67**: 3403–3411.

Wang, I.J. & Bradburd, G.S. 2014. Isolation by environment. *Mol. Ecol.* **23**: 5649–5662.

Weir, B.S. & Cockerham, C.C. 1984. Estimating F-statistics for the analysis of population structure. *Evolution* **38**: 1358–1370.

Weschenfelder, J., Baitelli, R., Corrêa, I.C.S., Bortolin, E.C. & dos Santos, C.B. 2014. Quaternary incised valleys in southern Brazil coastal zone. *J. South Am. Earth Sci.* **55**: 83–93.

Weschenfelder, J., Corrêa, I.C.S., Aliotta, S. & Baitelli, R. 2010. Paleochannels related to late Quaternary sea-level changes in southern Brazil. *Braz. J. Oceanogr.* **58**: 35–44.

Whigham, P.A., Dick, G.C. & Spencer, H.G. 2008. Genetic drift on networks: Ploidy and the time to fixation. *Theor. Popul. Biol.* **74**: 283–290.

Wilson, G.A. & Rannala, B. 2003. Bayesian inference of recent migration rates using multilocus genotypes. *Genetics* **163**: 1177–1191.

TABLES

Table 1. Names, geographical coordinates, sample size, and F_{IS} inbreeding coefficients of sampled populations of *Calibrachoa heterophylla*. Population IDs are the same that in Figure 1.

ID Population	Sample size	Locality name	Longitude	Latitude	F_{IS}
1	4	São Francisco de Assis	-55.10077	-29.58307	-0.16
2	10	Cacequi 1	-54.85375	-29.8947	-0.01
3	9	Cacequi 2	-54.90852	-29.85478	0
4	4	Arambaré	-51.49195	-30.90082	0.12
5	27	Barra do Ribeiro	-51.20255	-30.40754	0.08
6	23	Pelotas	-52.16478	-31.70757	0.07
7	3	Mostardas 1	-50.73934	-30.93746	0.35
8	13	Mostardas 2	-50.90112	-31.10909	0.08
9	37	Santo Antônio da Patrulha	-50.42936	-29.89291	0.10*
10	10	Laguna	-48.76501	-28.45991	0.34**
11	23	Torres	-49.79809	-29.43227	0.17**
12	26	Rio Grande	-52.54661	-32.52396	0.27***
13	12	São José do Norte 1	-51.42576	-31.66673	-0.01
14	11	São José do Norte 2	-52.03612	-32.02393	-0.05
15	41	Santa Vitória do Palmar	-52.7323	-32.98765	0.25***

*P-value < 0.05; ** P-value < 0.01; *** P-value < 0.001

Table 2. Model support statistics (upper panel) and parameter estimations taken from the best supported model (lower panel). Bayes factor values < 2 indicate strong preference for the model with highest probability. Θ = mutation scaled population size; M = mutation scaled migration rate; CI = confidende interval. For graphical model descriptions see Fig. S1.

Model name	Ln marginal Likelihood	Log Bayes Factor	Model probability
Step stone from coast	-5900.7	0	1.0
Source-sink from west	-5940.5	-79.6	5.1E-18
Source-sink from inland	-5941.5	-81.6	1.9E-18
Step stone from inland	-5944.9	-88.4	6.5E-20

Parameter	Median	Mean	95 % CI
Θ 'Inland'	4.11	4.25	2.55 - 6.28
Θ 'West'	0.55	0.56	0.12 - 0.96
Θ 'North'	0.51	0.52	0.05 - 0.95
Θ 'South'	0.27	0.22	0 - 0.55
M 'West' -> 'Inland'	0.41	0.45	0 - 1.16
M 'North' -> 'West'	0.3	0.26	0 - 0.6
M 'South' -> 'West'	0.19	0.08	0 - 0.41

FIGURES

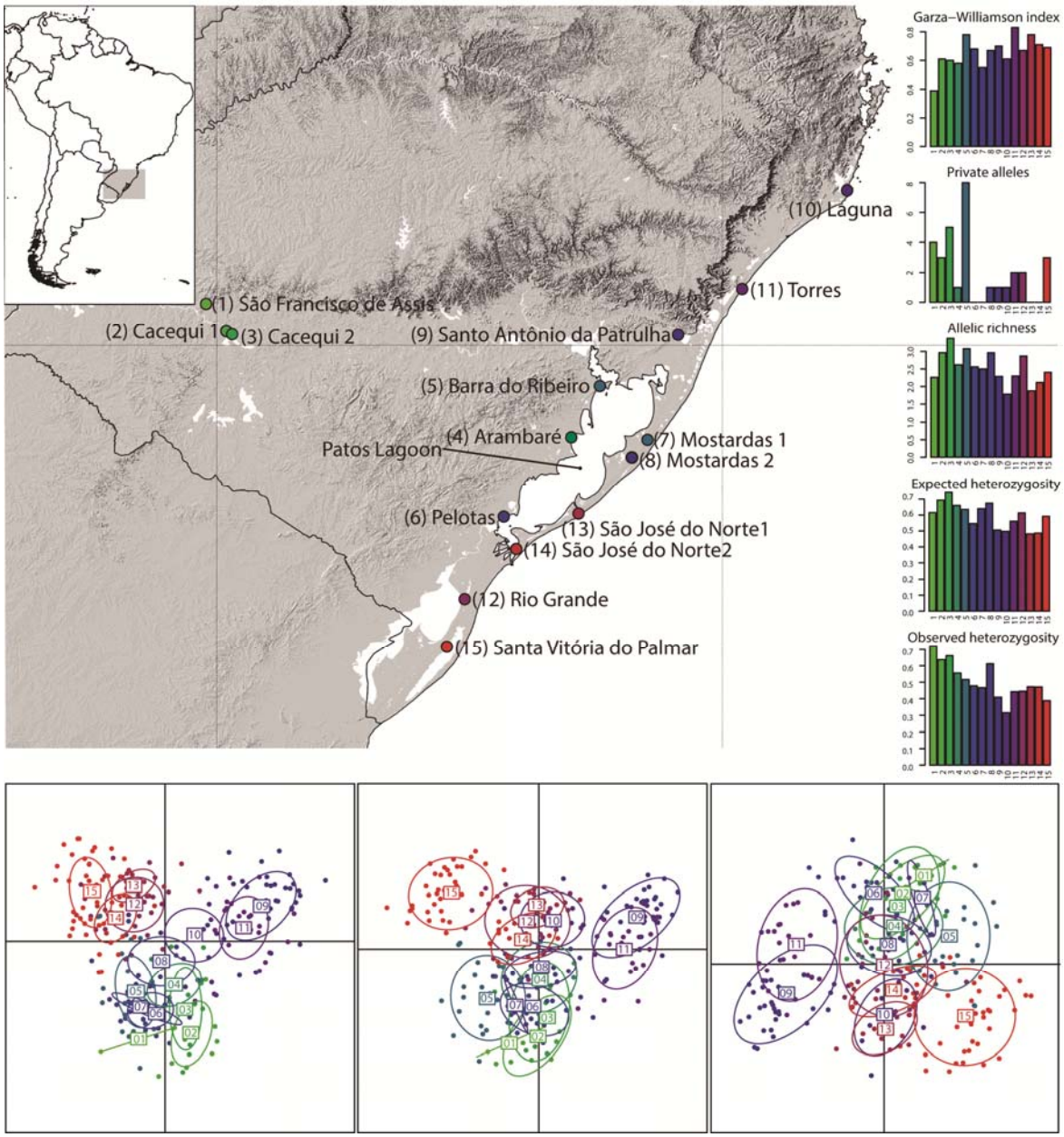


Figure 1. Geographic localization of *Calibrachoa heterophylla* sampled populations. Right panel shows a graphical representation of the mean genetic diversity statistics calculates for each population across all loci. Bottom panel shows the scatterplots of the three exploratory genetic structure analyses implemented (from left to right: DAPC, sPCA, and spFA). Populations' umbers and colours are in correspondence on each panel.

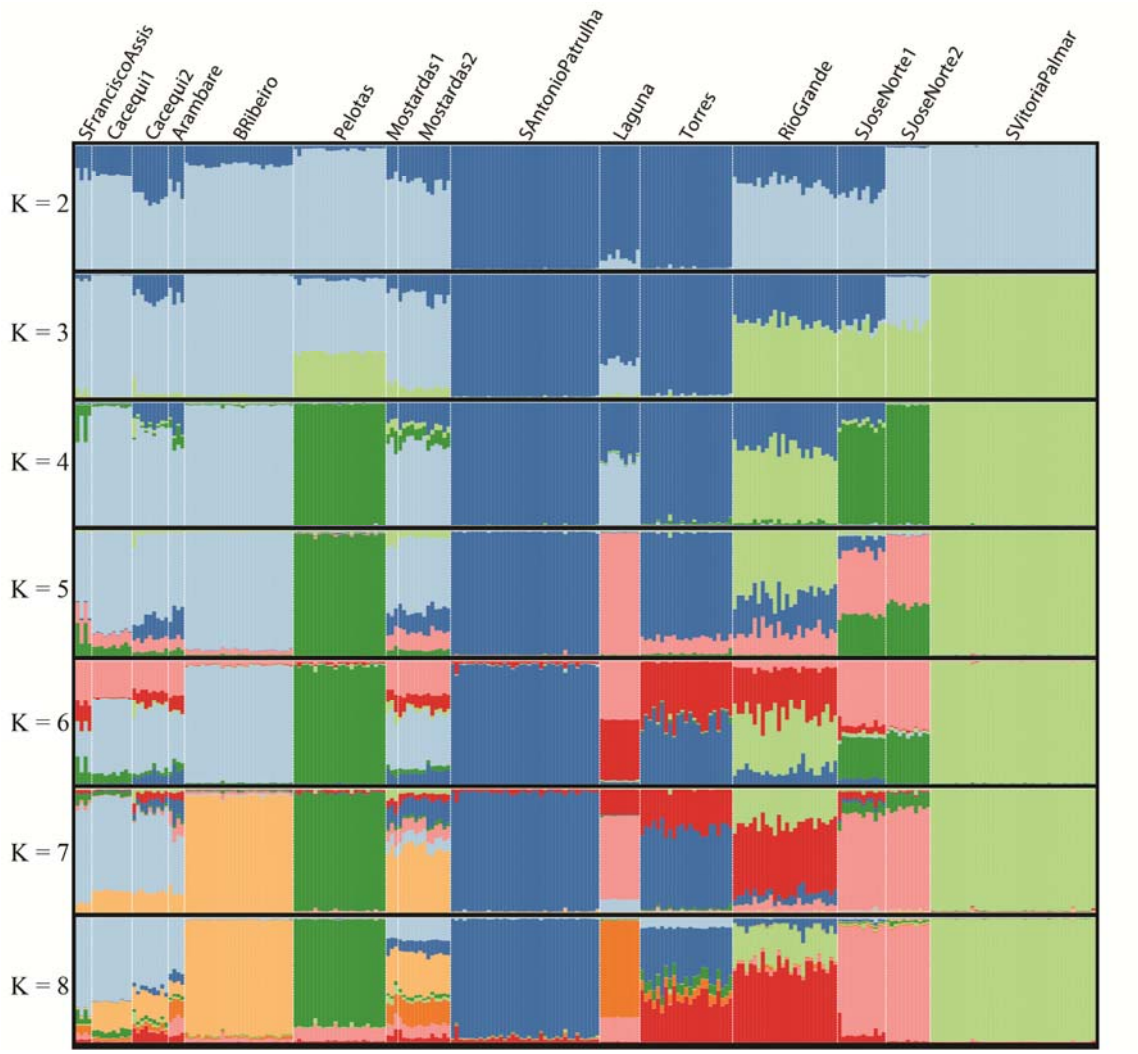


Figure 2. Barplots showing the individual membership to each identified genetic cluster with STRUCTURE. Populations are separated by white dashed lines and named on the top side of the figure.

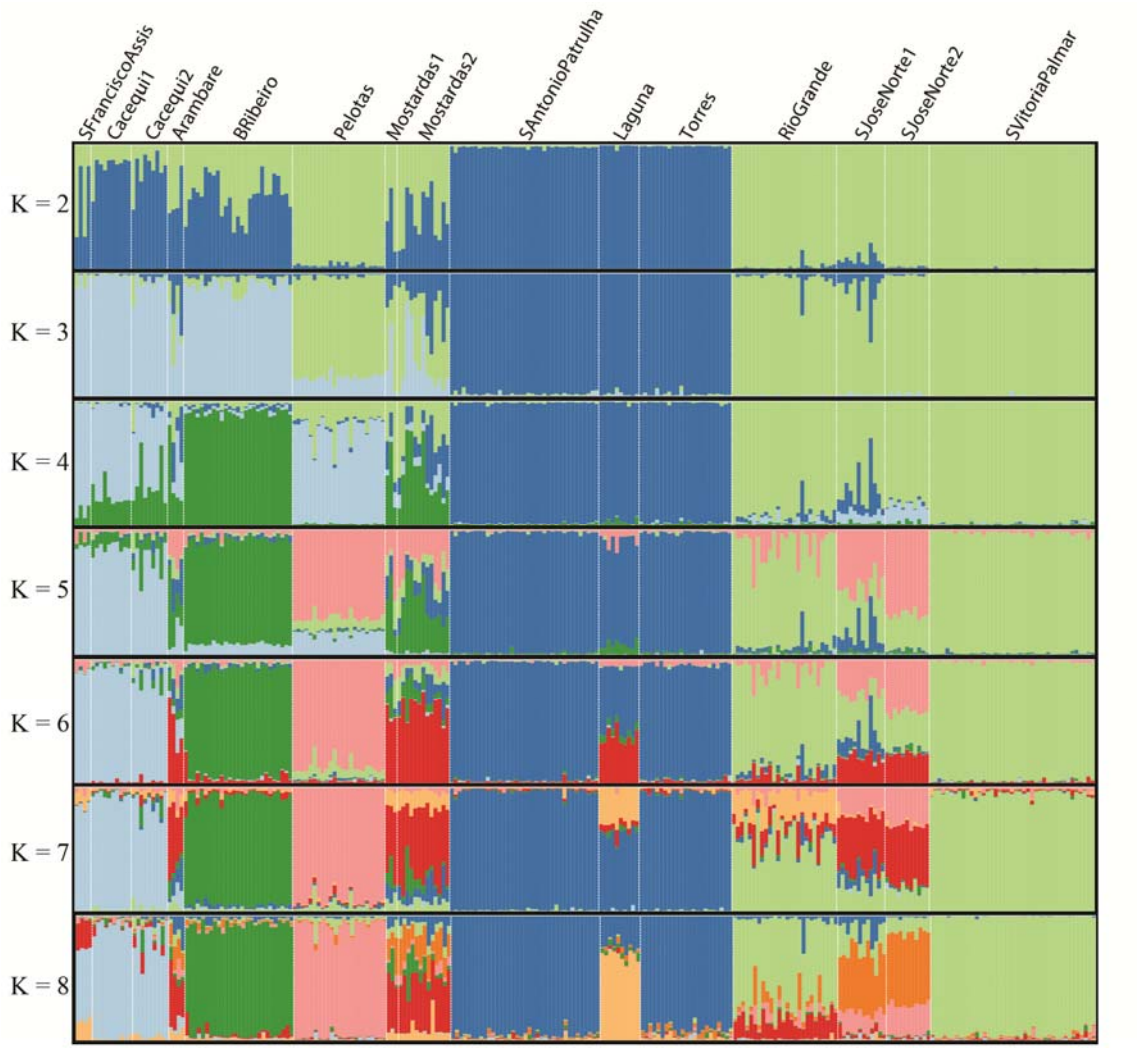


Figure 3. Barplots showing the individual membership to each identified genetic cluster with TESS. Populations are separated by white dashed lines and names are indicated on the top side of the figure.

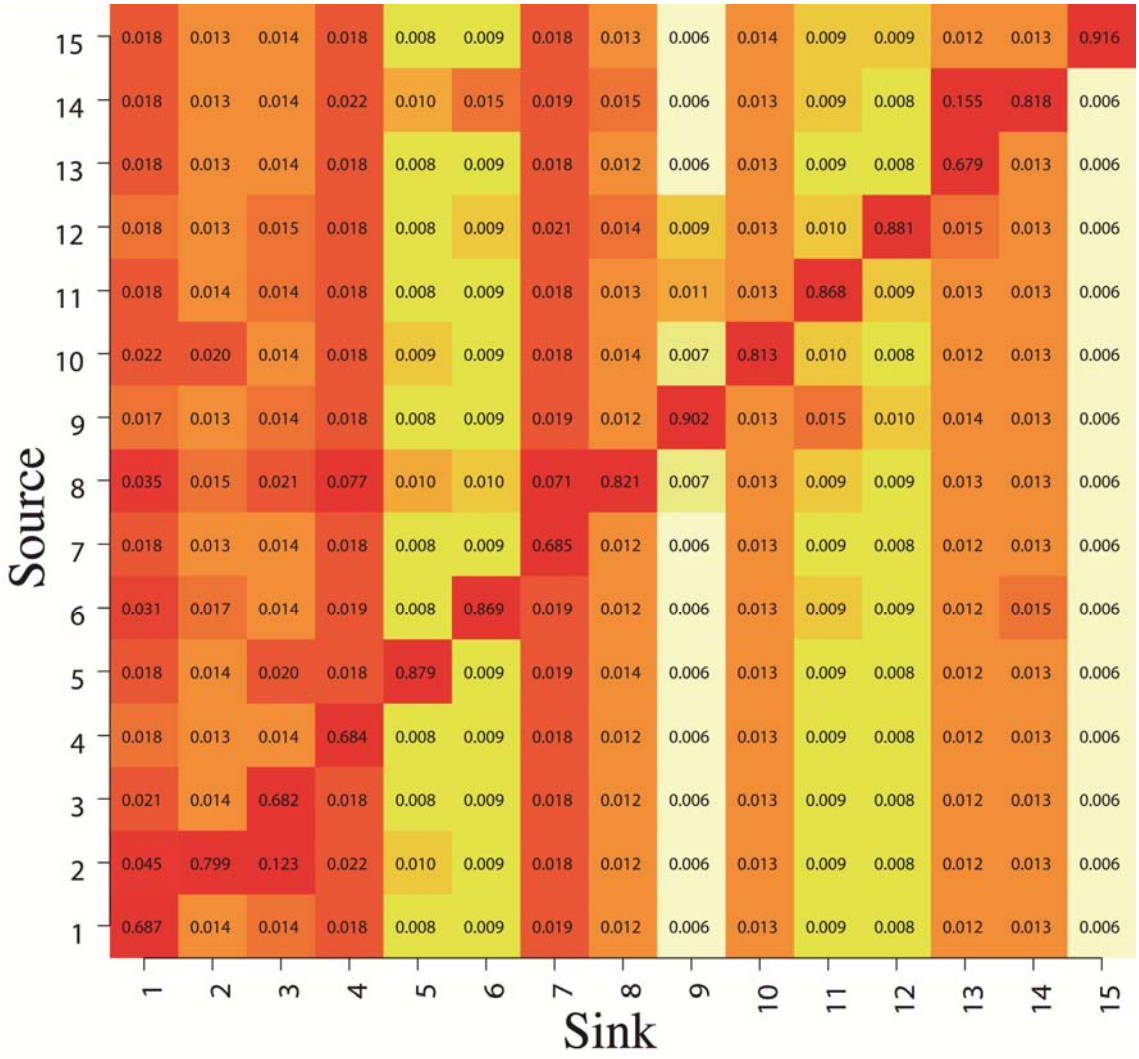


Figure 4. Estimation of contemporary migration rates with BAYESASS between each pair of sampled populations of *Calibrachoa heterophylla*. Pairwise migration estimations are represented from red (higher values) to white (lower estimates).

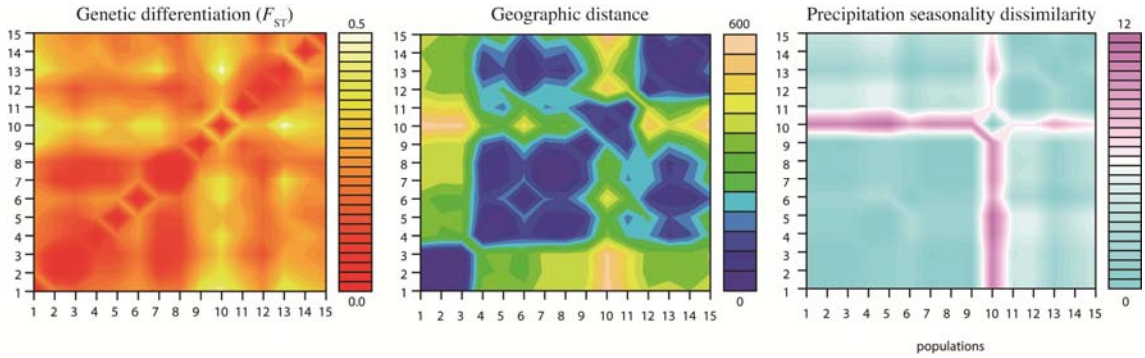


Figure 5. Coloured contour plots representing the inter-population genetic, geographic, and environmental dissimilarity matrices. Populations are numbered as in Figure 1 and Table 1.

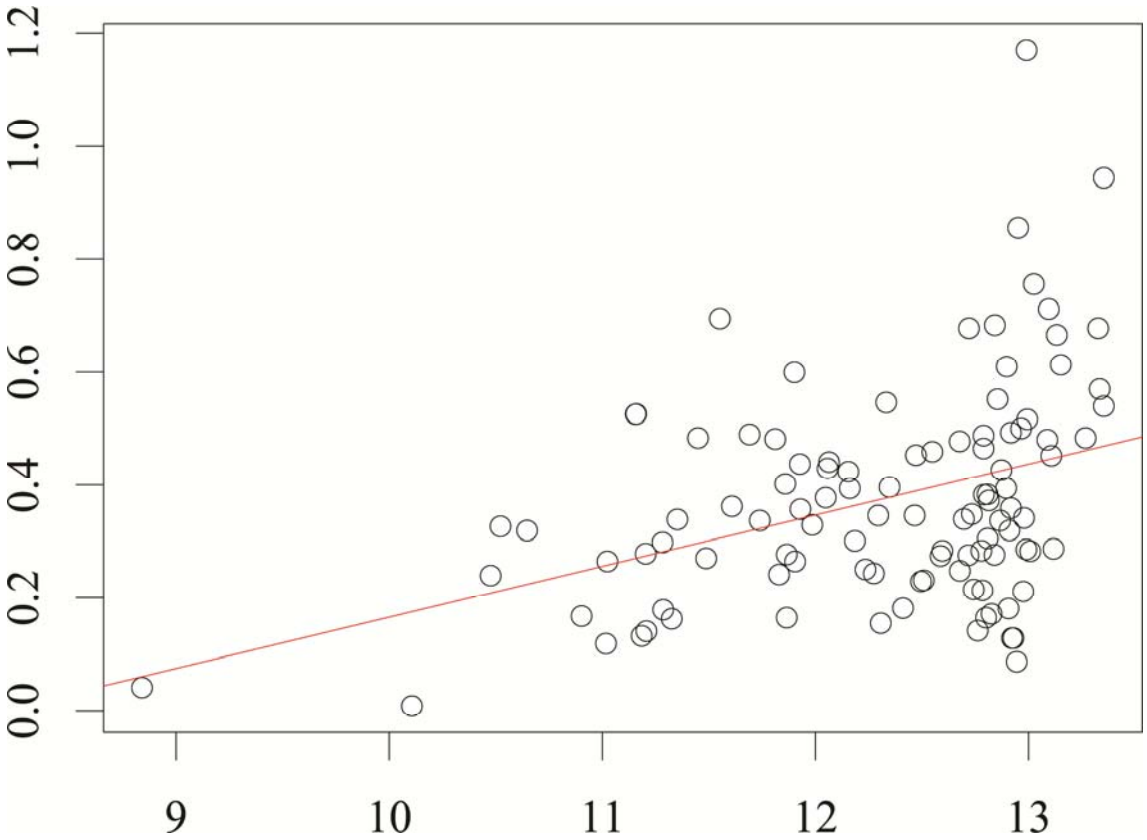


Figure 6. Plot representing the relationship between the natural logarithm of geographic distances (X-axis) and linearized genetic inter-population distances (Y-axis).

SUPPLEMENTARY DATA

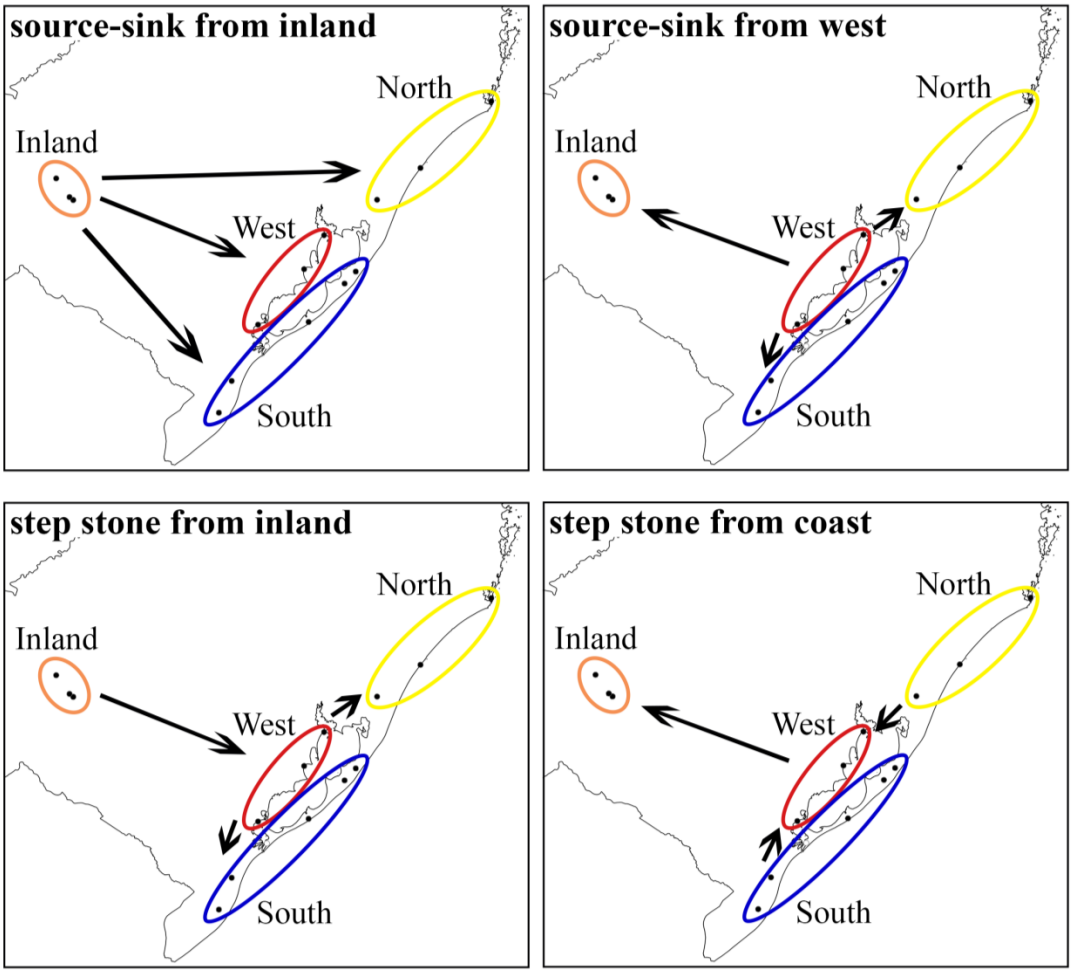


Figure S1. Maps showing the coalescent migration models tested in MIGRATE-N for *Calibrachoa heterophylla*.

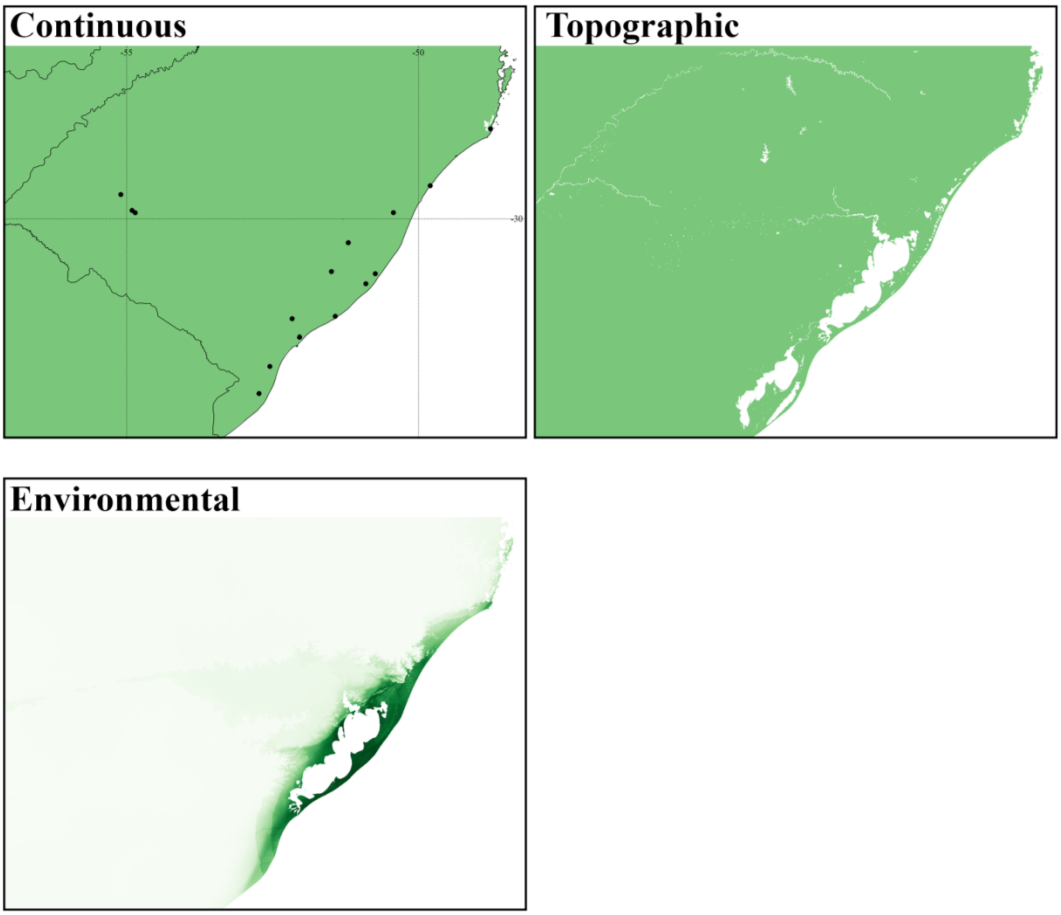


Figure S2. Maps showing the raster layers used to obtain the connectivity values with CIRCUITSCAPE used to test the isolation by resistance models.

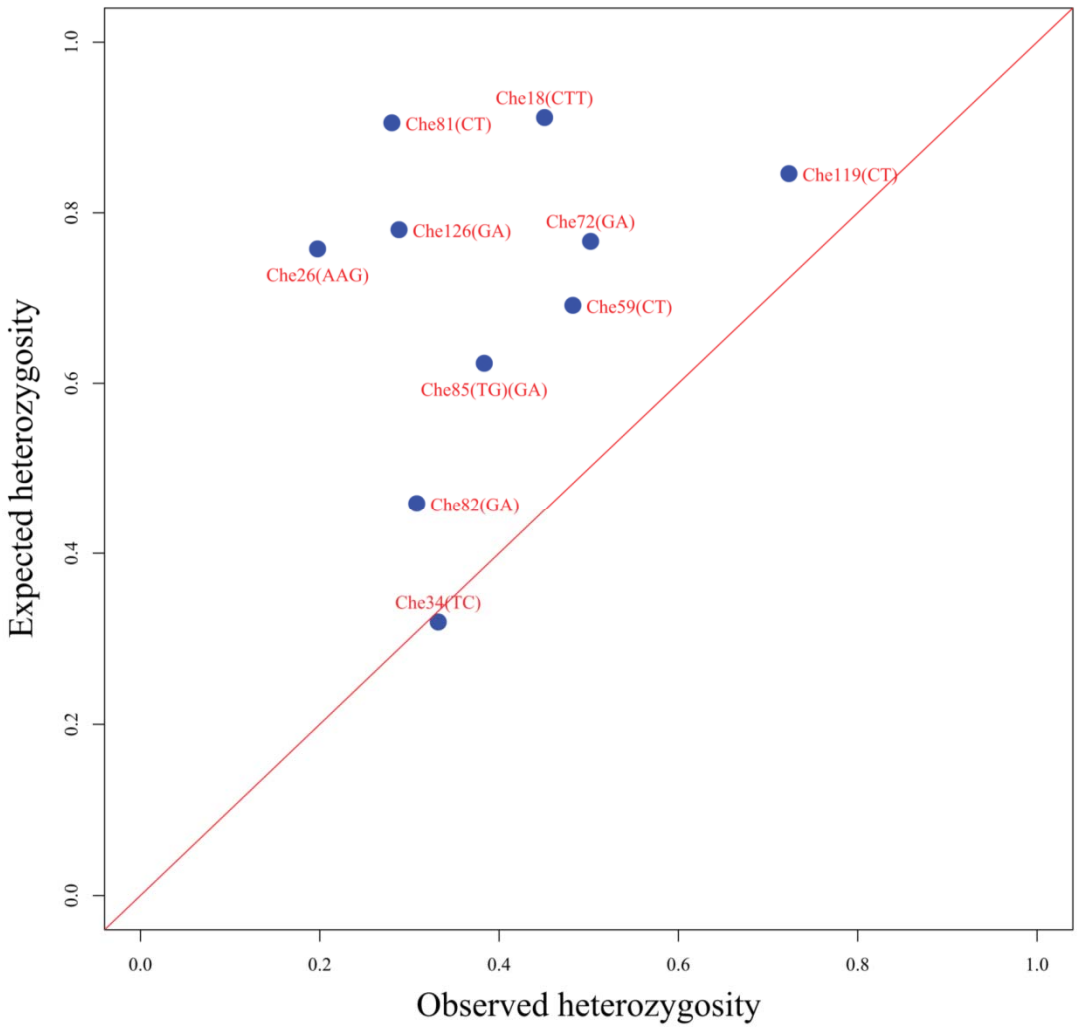


Figure S3. Plot of observed vs. expected heterozygosity for each locus.

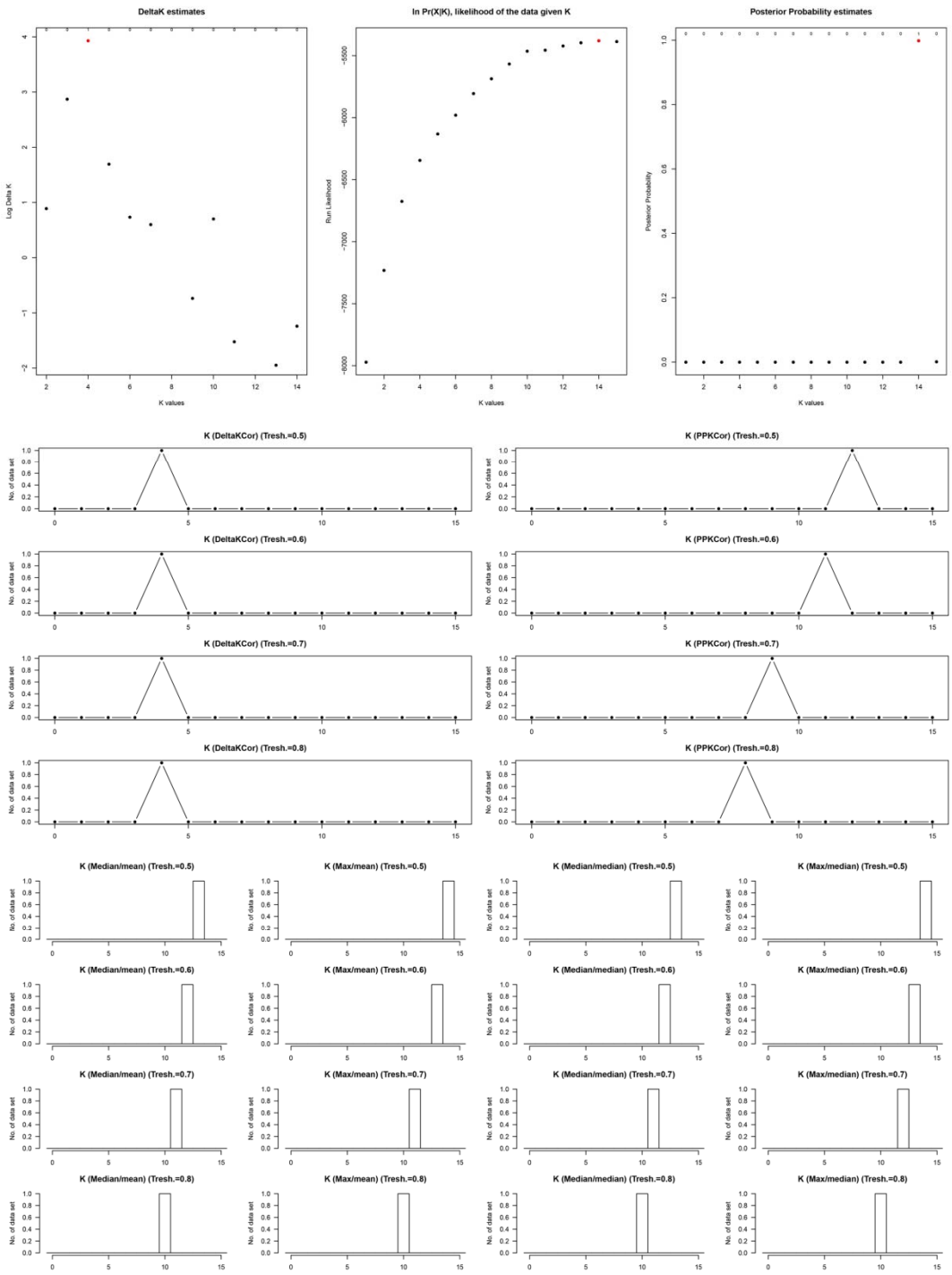


Figure S4. Plots of best K estimates for STRUCTURE results ontained with KSTIMATOR.

Figure S4. Plots of best K estimates for STRUCTURE results obtained with KSTIMATOR.

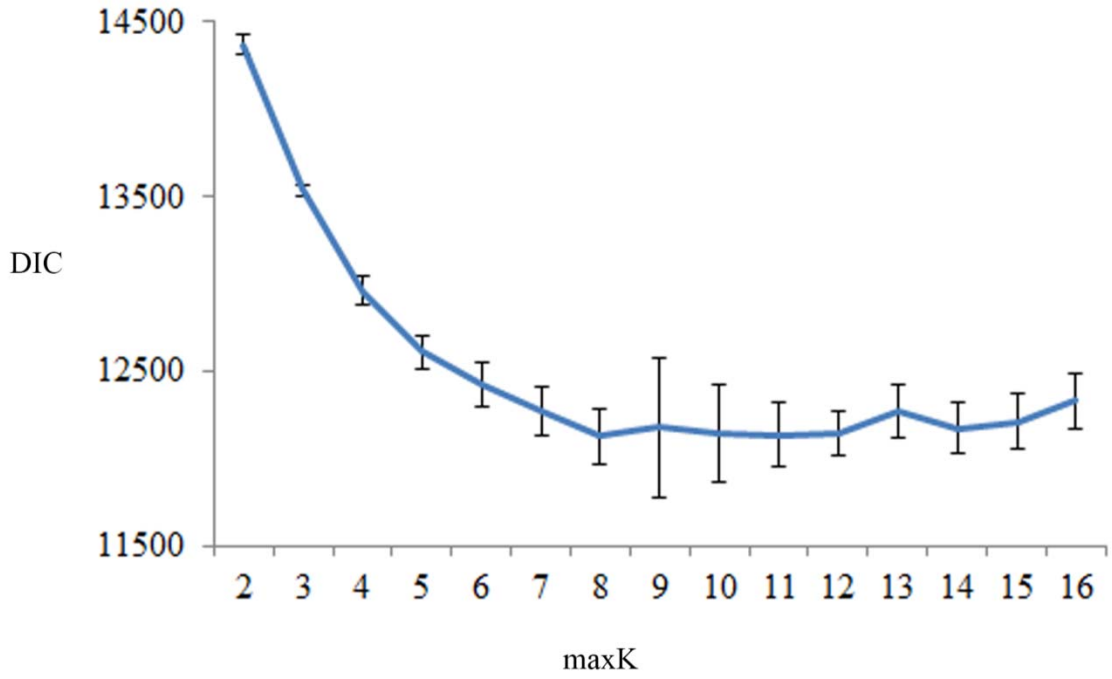


Figure S5. Plots of mean and standard deviation measures of the Devienace Information Criterion (DIC) statistics obtained for each maxK value assessed with TESS.

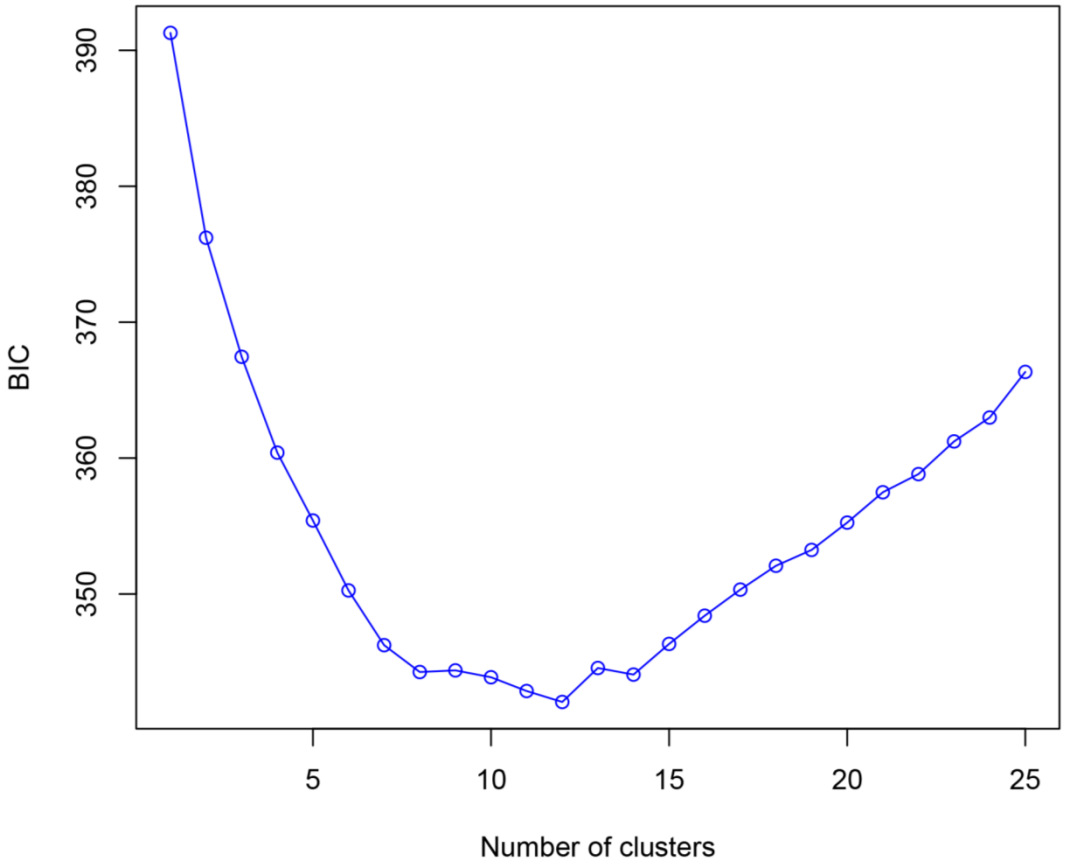


Figure S6. Plot of the Bayesian Information Criterion (BIC) statistics obtained for each K value assessed with the multivariate method Discriminant Analysis of Principal Components.

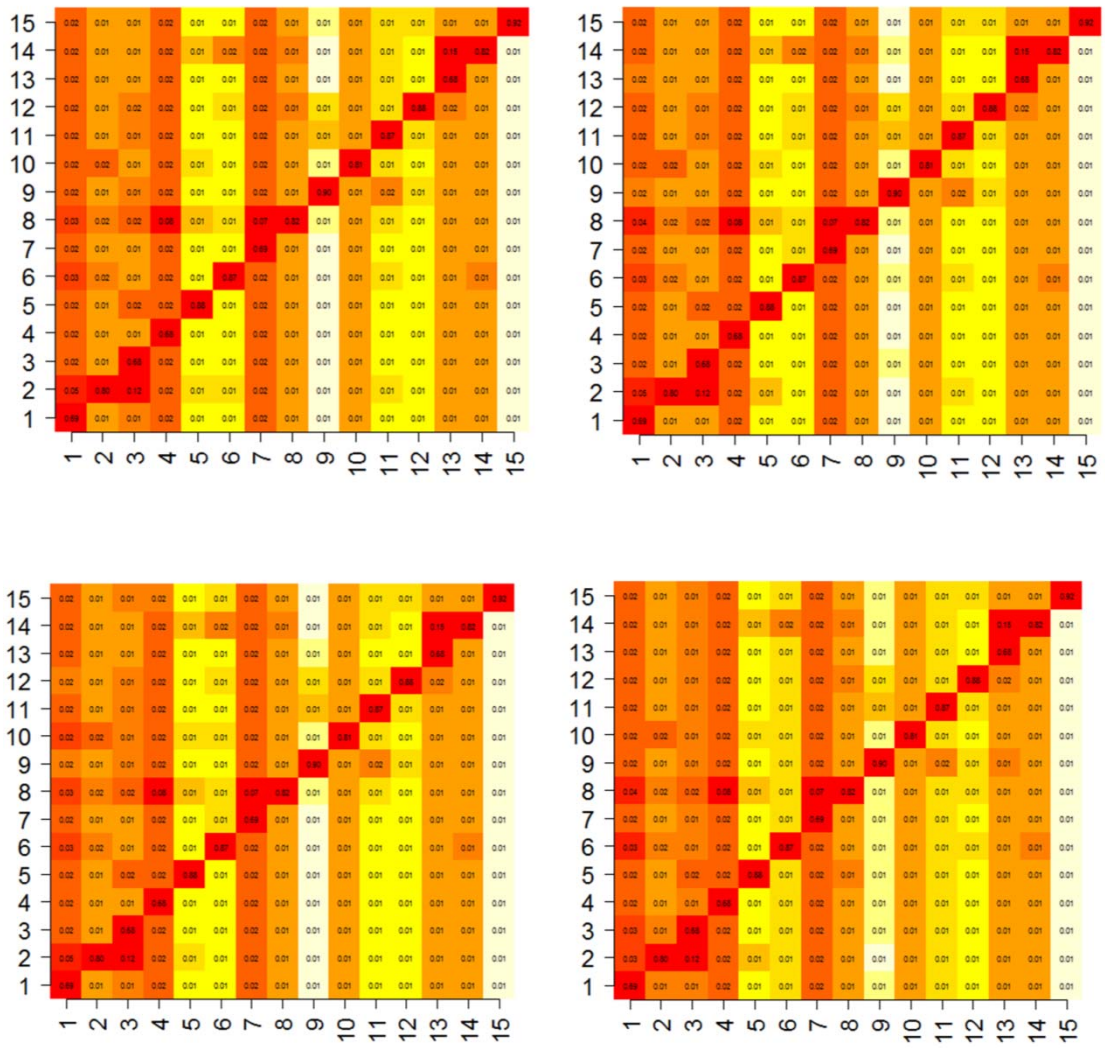


Figure S7. Matrices of migration estimates obtained with each of the four replicated runs of BAYESASS.

Ecological niche modelling and ensemble of *Calibrachoa heterophylla*

Species occurrence localities were taken from direct field observations. All occurrence points were first verified for duplicates. Only one occurrence record per grid cell was included. We included total of 27 localities covering the full extent geographic range of *C. heterophylla* (Table S1).

Table S1. Longitude-Latitude coordinates of biological records used in the ecological niche modelling.

Longitude	Latitude	Longitude	Latitude
-49.7981	-29.4323	-51.1461	-30.4753
-55.1008	-29.5831	-51.1468	-30.476
-50.4196	-30.5409	-51.1476	-30.4749
-50.7393	-30.9375	-51.1467	-30.4806
-50.9011	-31.1091	-51.2026	-30.4075
-52.5466	-32.524	-51.2251	-30.4203
-50.4294	-29.8929	-51.2312	-30.4237
-50.4262	-29.9052	-54.8538	-29.8947
-51.4258	-31.6667	-54.9085	-29.8548
-51.4243	-31.6679	-51.492	-30.9008
-52.01	-32.0465	-48.765	-28.4599
-52.7323	-32.9876	-52.0361	-32.0239
-51.1446	-30.4735	-52.1648	-31.7078
-51.1456	-30.4759		

The set of explanatory variables included five of the eight climatic raster surfaces (Fig. S8) obtained in Silva-Arias *et al.* (submitted) that showed the lowest pairwise spearman correlation values measured with values extracted from occurrence localities. Raster layers were set to 450 m resolution, WGS84 system, and covering the full geographic extent range of *C. heterophylla*.

Niche models were obtained with the ensemble niche modelling framework (Araujo & New, 2007). Individual models were generated using the methods

Generalized Linear Models (GLM; McCullagh & Nelder, 1983), Generalized Boosting Models (GBM; Death, 2007), Surface Range Envelop (SRE; Busby, 1991), Classification Tree Analysis (CTA; Breiman *et al.*, 1993), Multiple Adaptive Regression Splines (MARS; Leathwick *et al.*, 2005), and Artificial Neural Network (ANN; Ripley, 1996). Recent studies suggest that the geographic extent in which the pseudo-absences are taken have significant influences for prediction and performance of niche models (Thuiller *et al.*, 2004; VanDerWal *et al.*, 2009), following this reasoning we selected pseudo-absences as random localities throughout the full extant range of *C. heterophylla* . We did a sample of 10 000 random localities with 'disk' strategy, keeping 1 km as minimum distance from presence localities. Despite Barbet-Massin *et al.* (2012) recommend to sample pseudo-absences at higher distances from presence localities, we kept 1 km in order to account for the fine-scale of this study. We kept prevalence equal to 0.5. For each model, presence localities were divided in sets of 70% for training models and 30% for testing. For each model we performed ten runs.

Variables loadings were kept for all models calculated. To assess predictive performance of the individual and ensemble models, we measured the threshold independent statistics area under the receiver operating characteristic curve (AUC; Phillips *et al.*, 2006) and the True Skill Statistic (TSS; Allouche *et al.*, 2006).

The projected ensemble niche model for *C. heterophylla* for current climate conditions achieved with the climatic variables obtained in this study is shown in Figure S2. The ensemble niche model summarizes 38 models with TSS $\geq$ 0.7 (10 for GLM, 7 for GBM, 4 for CTA, 5 for MARS, 8for ANN, and 4 for SRE). The mean weighted by TSS ensemble model had a TSS = 0.98 and AUC = 0.99. Overall, the

most important variables for models were total annual precipitation and mean annual temperature (Fig. S8).

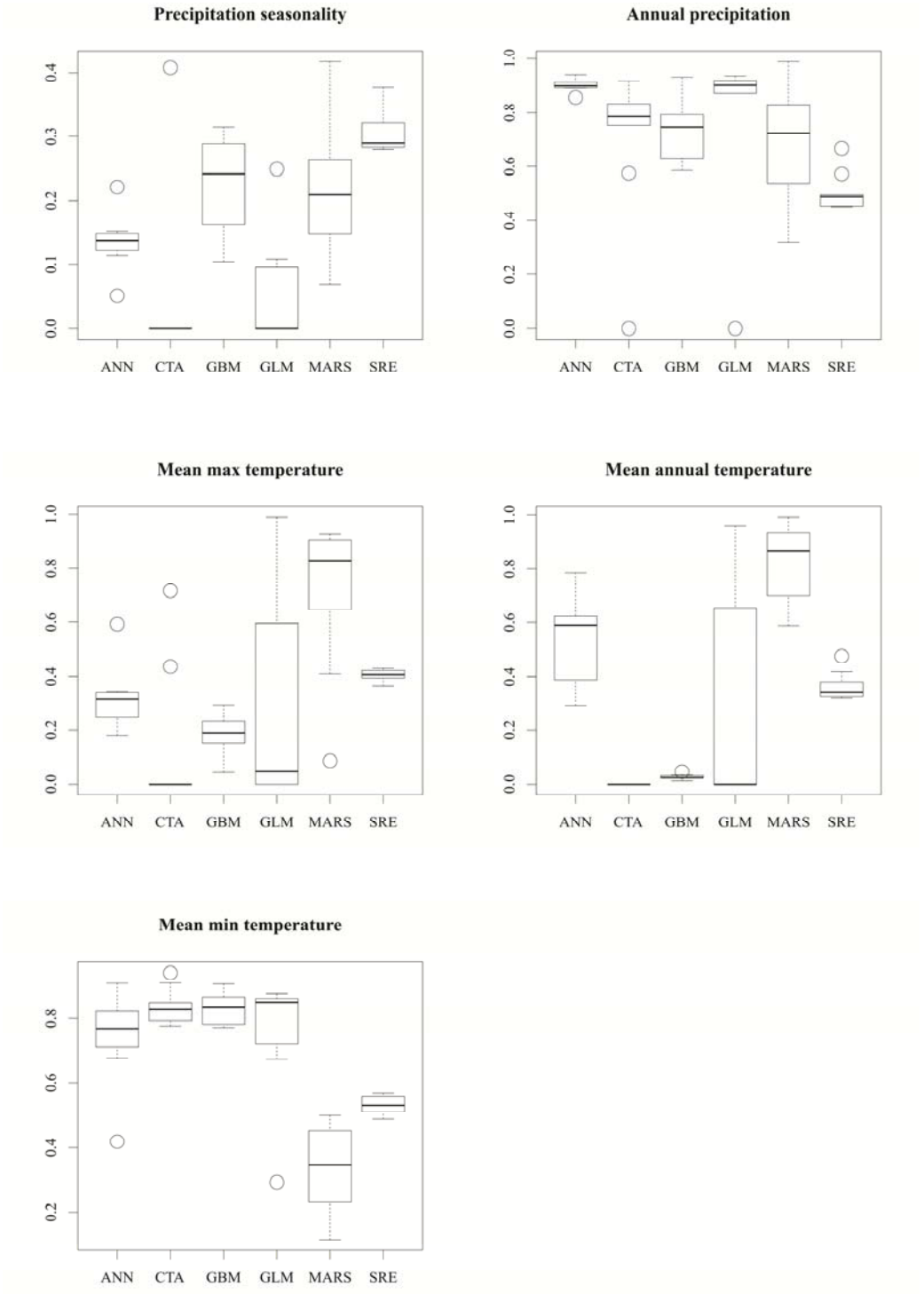


Figure S8. Summary of variable importance measures of each model.

References

Allouche, O., Tsoar, A. & Kadmon, R. 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *J. Appl. Ecol.* **43**: 1223–1232.

Araujo, M. & New, M. 2007. Ensemble forecasting of species distributions. *Trends Ecol. Evol.* **22**: 42–47.

Barbet-Massin, M., Jiguet, F., Albert, C.H. & Thuiller, W. 2012. Selecting pseudo-absences for species distribution models: how, where and how many?: how to use pseudo-absences in niche modelling? *Methods Ecol. Evol.* **3**: 327–338.

Breiman, L., Friedman, J., Stone, C.J. & Olshen, R.A. 1993. *Classification and regression trees*. Chapman & Hall, New York, N.Y.

Busby, J.R. 1991. BIOCLIM: a bioclimate analysis and prediction system. In: *Nature conservation: cost effective biological surveys and data analysis* (C. R. Margules & M. P. Austin, eds), pp. 64–68. CSIRO, Australia.

Death, G. 2007. Boosted trees for ecological modeling and prediction. *Ecology* **88**: 243–251.

Leathwick, J.R., Rowe, D., Richardson, J., Elith, J. & Hastie, T. 2005. Using multivariate adaptive regression splines to predict the distributions of New Zealand's freshwater diadromous fish. *Freshw. Biol.* **50**: 2034–2052.

McCullagh, P. & Nelder, J.A. 1983. *Generalized linear models*. Chapman and Hall/CRC, London, England.

Phillips, S.J., Anderson, R.P. & Schapire, R.E. 2006. Maximum entropy modeling of species geographic distributions. *Ecol. Model.* **190**: 231–259.

Ripley, B.D. 1996. *Pattern recognition and neural networks*. Cambridge University Press, Cambridge ; New York.

Thuiller, W., Brotons, L., Araújo, M.B. & Lavorel, S. 2004. Effects of restricting environmental range of data to project current and future species distributions. *Ecography* **27**: 165–172.

VanDerWal, J., Shoo, L.P., Graham, C. & Williams, S.E. 2009. Selecting pseudo-absence data for presence-only distribution modeling: How far should you stray from what you know? *Ecol. Model.* **220**: 589–594.

CONSIDERAÇOES FINAIS

O descobrimento dos fatores que determinam a estruturação genética e o fluxo gênico entre as populações é essencial para o entendimento dos processos históricos que originaram a diversidade genética e a distribuição geográfica das espécies, dos mecanismos de adaptação local, do isolamento reprodutivo e da relação entre estes processos e as características físicas e ecológicas dos locais de ocorrência dos organismos. O processo de diferenciação entre populações (ou grupos de populações) pode ser condicionado por processos neutros ou adaptativos, mas na natureza pode estar vinculado a um complexo de processos ao longo da história como é o caso de espécies envolvidas na colonização e expansão em novos ambientes (Orsini et al. 2013).

No presente trabalho foram analisadas em conjunto as características da estrutura genética e do fluxo gênico em relação às condições físicas e ambientais de duas espécies de plantas envolvidas em processos paralelos de colonização da Planície Costeira do Atlântico Sul (PCAS). Sinais de efeito fundador, isolamento por distância, isolamento por ambiente e contato secundário entre linhagens intraespecíficas previamente isoladas foram inferidos a partir da caracterização da diversidade genética com marcadores do tipo microssatélites em *Calibrachoa heterophylla* e *Petunia depauperata*. A pesar de que as duas espécies apresentam histórias demográficas basais diferentes (Mäder *et al.* 2013; Ramos-Fregonezi *et al.* 2015), alguns processos contemporâneos mostraram concordância entre elas.

Para as duas espécies foi encontrado um padrão consistente de isolamento por ambiente relacionado com a variável de sazonalidade da precipitação. Explorações das diferenças desta variável na área de estudo mostraram que nos extremos norte e

sul da PCAS ocorrem a maiores variações nas chuvas ao longo do ano (Carmargo 2002). Esta variação é mais forte na região norte e está influenciada principalmente pela proximidade com as escarpas da Serra Geral que geram precipitações com maior intensidade no verão. Congruentemente, populações das duas espécies localizadas no extremo norte da distribuição apresentaram os maiores valores nas estatísticas de distância genética.

A forte diferenciação das populações dos extremos da distribuição, especialmente aquelas do extremo norte, também pode ter implicações históricas. A proximidade das escarpas da Serra Geral com a linha da costa na região norte tem uma influência grande já que os hábitats rochosos desta região onde se encontram as populações de *C. heterophylla* e *P. depauperata* têm estado disponíveis por mais tempo e têm sido mais estáveis do que os hábitats dos campos arenosos das regiões centro e sul da PCAS, que surgiram mais tardiamente durante o Pleistoceno ou no Holoceno. Isso possivelmente contribuiu para que, uma vez estabelecidas as populações na região norte, estas tivessem mais tempo para gerar diferenciação neutra ou adaptativa.

Adicionalmente, as dinâmicas de fluxo gênico que envolvem as populações marginais de *C. heterophylla* e *P. depauperata* tiveram resultados contrastantes, os quais poderiam estar relacionados com as diferenças na duração do ciclo de vida destas espécies. Embora as duas espécies possam estar apresentando processos de isolamento por ambiente nas populações marginais, particularmente *C. heterophylla* pode ter um efeito de prioridade ou monopolização (De Meester et al. 2016) mais forte, já que a longevidade maior dos indivíduos desta espécie limitaria o estabelecimento de novos imigrantes fazendo com que o *pool* gênico dos indivíduos fundadores perdurasse por mais tempo na população. Os baixos valores de tamanho

efetivo populacional e imigração restrita encontrados nas populações marginais de *C. heterophylla* apoiam estas ideias (ver Capítulo 3). Em contraste, *P. depauperata* apresenta ciclo de vida anual o que permite que a perda de diversidade relacionada com fenômenos de efeito fundador ou alelos surfistas em processos de expansão possa ser mitigada com maior rapidez pela recombinação e o fluxo gênico que, embora mais restrito em populações marginais, pode acabar sendo significativo devido ao tamanho efetivo reduzido das populações recentemente estabelecidas (Alleaume-Benharira et al. 2006; Kawecki 2008).

A evidência conjunta até agora disponível aponta para uma história evolutiva de *C. heterophylla* e *P. depauperata* moldada por vários processos demográficos e adaptativos. A partir dos resultados obtidos nos trabalhos aqui apresentados se recomenda desenvolver novos estudos reunindo aproximações mecanicistas (avaliações morfológicas, genéticas e de aptidão diferenciada) e genômicas (varreduras genômicas na procura de sinais de divergência e seleção) para confirmar os descobrimentos apresentados e obter um completo entendimento dos processos evolutivos envolvidos nestes sistemas (Byers et al. 2016).

Os avanços recentes nas caracterizações genômicas, o conhecimento até agora adquirido das espécies do grupo, e a grande disponibilidade de informação genética que perfila *Petunia* como uma planta modelo (Vandenbussche et al. 2016), permitem o desenvolvimento de novas abordagens integradas para entender processos de adaptação local, entre outros, e tornarão as populações naturais das espécies de *Petunia* um laboratório de estudo de evolução em tempo real.

REFERÊNCIAS DA INTRODUÇÃO E CONSIDERAÇÕES FINAIS

Alleaume-Benharira M, Pen IR and Ronce O (2006) Geographical patterns of adaptation within a species' range: interactions between drift and gene flow. J Evol Biol 19:203–215. doi: 10.1111/j.1420-9101.2005.00976.x

Anderson CD, Epperson BK, Fortin M-J, Holderegger R, James PMA, Rosenberg MS, Scribner KT and Spear S (2010) Considering spatial and temporal scale in landscape-genetic studies of gene flow. Mol Ecol 19:3565–3575. doi: 10.1111/j.1365-294X.2010.04757.x

Ando T, Ishikawa N, Watanabe H, Kokubun H, Yanagisawa Y, Hashimoto G, Marchesi E and Suárez E (2005a) A morphological study of the *Petunia integrifolia* complex (Solanaceae). Ann Bot 96:887–900. doi: 10.1093/aob/mci241

Ando T, Kokubun H, Watanabe H, Tanaka N, Yukawa T, Hashimoto G, Marchesi E, Suárez E and Basualdo IL (2005b) Phylogenetic Analysis of *Petunia* sensu Jussieu (Solanaceae) using Chloroplast DNA RFLP. Ann Bot 96:289–297. doi: 10.1093/aob/mci177

Angers B, Magnan P, Plante M and Bernatchez L (1999) Canonical correspondence analysis for estimating spatial and environmental effects on microsatellite gene diversity in brook charr (*Salvelinus fontinalis*). Mol Ecol 8:1043–1053. doi: 10.1046/j.1365-294x.1999.00669.x

Balkenhol N, Cushman SA, Storfer A and Waits LP (2016) Introduction to Landscape Genetics - Concepts, Methods, Applications. In: Balkenhol N, Cushman SA, Storfer AT and Waits LP (eds) Landscape genetics: concepts, methods, applications. John Wiley & Sons, Ltd, Chichester, UK, pp 1–8

Balkenhol N, Holbrook JD, Onorato D, Zager P, White C and Waits LP (2014) A multi-method approach for analyzing hierarchical genetic structures: a case study with cougars *Puma concolor*. Ecography 37:552–563. doi: 10.1111/j.1600-0587.2013.00462.x

Balkenhol N, Waits LP and Dezzani RJ (2009) Statistical approaches in landscape genetics: an evaluation of methods for linking landscape and genetic data. Ecography 32:818–830. doi: 10.1111/j.1600-0587.2009.05807.x

Barbará T, Martinelli G, Fay MF, Mayo SJ and Lexer C (2007) Population differentiation and species cohesion in two closely related plants adapted to neotropical high-altitude ´inselbergs´, *Alcantarea imperialis* and *Alcantarea geniculata* (Bromeliaceae). Mol Ecol 16:1981–1992. doi: 10.1111/j.1365-294X.2007.03272.x

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.

Bolker BM, Brooks ME, Clark CJ, Geange SW, Poulsen JR, Stevens MHH and White J-SS (2009) Generalized linear mixed models: a practical guide for ecology and evolution. Trends Ecol Evol 24:127–135. doi: 10.1016/j.tree.2008.10.008

Byers KJRP, Xu S and Schlüter PM (2016) Molecular mechanisms of adaptation and speciation: Why do we need an integrative approach? Mol Ecol. doi: 10.1111/mec.13678

Carmargo OA (2002) Atlas eólico: Rio Grande do Sul. Porto Alegre, RS

Collevatti RG, Grattapaglia D and Hay JD (2001) Population genetic structure of the endangered tropical tree species *Caryocar brasiliense*, based on variability at microsatellite loci. Mol Ecol 10:349–356. doi: 10.1046/j.1365-294x.2001.01226.x

Crandall KA, Bininda-Emonds ORP, Mace GM and Wayne RK (2000) Considering evolutionary processes in conservation biology. Trends Ecol Evol 15:290–295.

Cushman S, Wasserman T, Landguth E and Shirk A (2013) Re-evaluating causal modeling with Mantel tests in landscape genetics. Diversity 5:51–72. doi: 10.3390/d5010051

Dayanandan S, Dole J, Bawa K and Kesseli R (1999) Population structure delineated with microsatellite markers in fragmented populations of a tropical tree, *Carapa guianensis* (Meliaceae). Mol Ecol 8:1585–1592. doi: 10.1046/j.1365-294x.1999.00735.x

De Meester L, Vanoverbeke J, Kilsdonk LJ and Urban MC (2016) Evolving perspectives on monopolization and priority effects. Trends Ecol Evol 31:136–146. doi: 10.1016/j.tree.2015.12.009

Diniz-Filho JAF, Barbosa ACOF, Collevatti RG, Chaves LJ, Terribile LC, Lima-Ribeiro MS and Telles MPC (2016) Spatial autocorrelation analysis and ecological niche modelling allows inference of range dynamics driving the population genetic structure of a Neotropical savanna tree. J Biogeogr 43:167–177. doi: 10.1111/jbi.12622

Diniz-Filho JAF, Melo DB, Oliveira G, Collevatti RG, Soares TN, Nabout JC, Lima J de S, Dobrovolski R, Chaves LJ, Naves RV et al. (2012) Planning for optimal conservation of geographical genetic variability within species. Conserv Genet 13:1085–1093. doi: 10.1007/s10592-012-0356-8

Diniz-Filho JAF, Nabout JC, Bini LM, Soares TN, Campos Telles MP, Marco P and Collevatti RG (2009) Niche modelling and landscape genetics of *Caryocar brasiliense* ("Pequi" tree: Caryocaraceae) in Brazilian Cerrado: an integrative approach for evaluating central–peripheral population patterns. Tree Genet Genomes 5:617–627. doi: 10.1007/s11295-009-0214-0

Diniz-Filho JAF, Soares TN, Lima JS, Dobrovolski R, Landeiro VL, Telles MP de C, Rangel TF and Bini LM (2013) Mantel test in population genetics. Genet Mol Biol 36:475–485. doi: 10.1590/S1415-47572013000400002

Dray S, Legendre P and Peres-Neto PR (2006) Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). Ecol Model 196:483–493. doi: 10.1016/j.ecolmodel.2006.02.015

Duminil J, Hardy OJ and Petit RJ (2009) Plant traits correlated with generation time directly affect inbreeding depression and mating system and indirectly genetic structure. BMC Evol Biol 9:177. doi: 10.1186/1471-2148-9-177

Dyer RJ (2016) Landscapes and plant population genetics. In: Balkenhol N, Cushman SA, Storfer AT and Waits LP (eds) Landscape genetics: concepts, methods, applications. John Wiley & Sons, Ltd, Chichester, UK, pp 181–198

Dyer RJ (2009) GENETICSTUDIO: a suite of programs for spatial analysis of genetic-marker data. Mol Ecol Resour 9:110–113. doi: 10.1111/j.1755-0998.2008.02384.x

Ellegren H (2004) Microsatellites: simple sequences with complex evolution. Nat Rev Genet 5:435–445. doi: 10.1038/nrg1348

Epperson BK (2003) Geographical genetics. Princeton University Press, Princeton, N.J

Etherington TR (2011) Python based GIS tools for landscape genetics: visualising genetic relatedness and measuring landscape connectivity. Methods Ecol Evol 2:52–55. doi: 10.1111/j.2041-210X.2010.00048.x

Ferrer ES, García-Navas V, Bueno-Enciso J, Barrientos R, Serrano-Davies E, Cáliz-Campal C, Sanz JJ and Ortego J (2016) The influence of landscape configuration and environment on population genetic structure in a sedentary passerine: insights from loci located in different genomic regions. J Evol Biol 29:205–219. doi: 10.1111/jeb.12776

Fortin M-J (2005) Spatial analysis: a guide for ecologists. Cambridge University Press, Cambridge, N.Y

François O and Durand E (2010) Spatially explicit Bayesian clustering models in population genetics. Mol Ecol Resour 10:773–784. doi: 10.1111/j.1755-0998.2010.02868.x

François O and Waits LP (2016) Clustering and assignment methods in landscape genetics. In: Balkenhol N, Cushman SA, Storfer AT and Waits LP (eds) Landscape genetics: concepts, methods, applications. John Wiley & Sons, Ltd, Chichester, UK, pp 114–128

Frantz AC, Bertouille S, Eloy MC, Licoppe A, Chaumont F and Flamand MC (2012) Comparative landscape genetic analyses show a Belgian motorway to be a gene flow barrier for red deer (*Cervus elaphus*), but not wild boars (*Sus scrofa*). Mol Ecol 21:3445–3457. doi: 10.1111/j.1365-294X.2012.05623.x

Frantz AC, Cellina S, Krier A, Schley L and Burke T (2009) Using spatial Bayesian methods to determine the genetic structure of a continuously distributed population:

clusters or isolation by distance? J Appl Ecol 46:493–505. doi: 10.1111/j.1365-2664.2008.01606.x

Fregonezi JN, Freitas LB, Bonatto SL, Semir J and Stehmann JR (2012) Infrageneric classification of *Calibrachoa* (Solanaceae) based on morphological and molecular evidence. Taxon 61:120–130.

Grimm AM, Ferraz SET and Gomes J (1998) Precipitation anomalies in southern Brazil associated with El Niño and La Niña events. J Clim 11:2863–2880.

Guillot G and Rousset F (2013) Dismantling the Mantel tests. Methods Ecol Evol 4:336–344. doi: 10.1111/2041-210x.12018

Hesp PA, Dillenburg SR, Barboza EG, Tomazelli LJ, Ayup-Zouain RN, Esteves LS, Gruber NLS, Tabajara LLCA and Clerot LCP (2005) Beach ridges, foredunes or transgressive dunefields? Definitions and an examination of the Torres to Tramandaí barrier system, Southern Brazil. An Acad Bras Ciênc 77:493–508.

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

Holderegger R and Wagner HH (2006) A brief guide to landscape genetics. Landsc Ecol 21:793–796. doi: 10.1007/s10980-005-6058-6

Joost S, Bonin A, Bruford MW, DespréS L, Conord C, Erhardt G and Taberlet P (2007) A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation. Mol Ecol 16:3955–3969. doi: 10.1111/j.1365-294X.2007.03442.x

Kalia RK, Rai MK, Kalia S, Singh R and Dhawan AK (2010) Microsatellite markers: an overview of the recent progress in plants. Euphytica 177:309–334. doi: 10.1007/s10681-010-0286-9

Kawecki TJ (2008) Adaptation to marginal habitats. Annu Rev Ecol Evol Syst 39:321–342. doi: 10.1146/annurev.ecolsys.38.091206.095622

Kozak KH, Graham CH and Wiens JJ (2008) Integrating GIS-based environmental data into evolutionary biology. Trends Ecol Evol 23:141–148. doi: 10.1016/j.tree.2008.02.001

Legendre P and Anderson MJ (1999) Distance-based redundancy analysis: Testing multispecies responses in multifactorial ecological experiments. Ecol Monogr 69:1–24. doi: 10.1890/0012-9615(1999)069[0001:DBRATM]2.0.CO;2

Legendre P and Fortin M-J (2010) Comparison of the Mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. Mol Ecol Resour 10:831–844. doi: 10.1111/j.1755-0998.2010.02866.x

Legendre P, Fortin M-J and Borcard D (2015) Should the Mantel test be used in spatial analysis? Methods Ecol Evol 6:1239–1247. doi: 10.1111/2041-210X.12425

Lichstein JW (2007) Multiple regression on distance matrices: a multivariate spatial analysis tool. Plant Ecol 188:117–131. doi: 10.1007/s11258-006-9126-3

Longo D, Lorenz-Lemke AP, Mäder G, Bonatto SL and Freitas LB (2014) Phylogeography of the *Petunia integrifolia* complex in southern Brazil. Bot J Linn Soc 174:199–213. doi: 10.1111/boj.12115

Mäder G, Fregonezi JN, Lorenz-Lemke AP, Bonatto SL and Freitas LB (2013) Geological and climatic changes in Quaternary shaped the evolutionary history of *Calibrachoa heterophylla*, an endemic South-Atlantic species of petunia. BMC Evol Biol 13:178. doi: 10.1186/1471-2148-13-178

Manel S, Gaggiotti O and Waples R (2005) Assignment methods: matching biological questions with appropriate techniques. Trends Ecol Evol 20:136–142. doi: 10.1016/j.tree.2004.12.004

Manel S and Holderegger R (2013) Ten years of landscape genetics. Trends Ecol Evol 28:614–621. doi: 10.1016/j.tree.2013.05.012

Manel S, Joost S, Epperson BK, Holderegger R, Storfer A, Rosenberg MS, Scribner KT, Bonin A and Fortin M-J (2010) Perspectives on the use of landscape genetics to detect genetic adaptive variation in the field. Mol Ecol 19:3760–3772. doi: 10.1111/j.1365-294X.2010.04717.x

Manel S, Schwartz MK, Luikart G and Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. TRENDS Ecol Evol 18:189–197.

Mantel N (1967) The detection of disease clustering and a generalized regression approach. Cancer Res 27:209–220.

Martinho CT (2008) Morfodinâmica e evolução de campos de dunas transgressivos quaternários do litoral do Rio Grande do Sul. Tese de Doutorado, Universidade Federal do Rio Grande do Sul

Matsuoka Y, Vigouroux Y, Goodman MM, Sanchez J, Buckler E and Doebley J (2002) A single domestication for maize shown by multilocus microsatellite genotyping. PNAS 99:6080–6084.

Meirmans PG (2012) The trouble with isolation by distance. Mol Ecol 21:2839–2846. doi: 10.1111/j.1365-294X.2012.05578.x

Nosil P, Funk DJ and Ortiz-Barrientos D (2009) Divergent selection and heterogeneous genomic divergence. Mol Ecol 18:375–402. doi: 10.1111/j.1365-294X.2008.03946.x

Oetjen K and Reusch TBH (2007) Genome scans detect consistent divergent selection among subtidal vs. intertidal populations of the marine angiosperm *Zostera marina*. Mol Ecol 16:5156–5167. doi: 10.1111/j.1365-294X.2007.03577.x

Olmstead RG, Bohs L, Migid HA, Santiago-Valentin E, Garcia VF and Collier SM (2008) A molecular phylogeny of the Solanaceae. Taxon 57:1159–1181.

Orsini L, Vanoverbeke J, Swillen I, Mergeay J and De Meester L (2013) Drivers of population genetic differentiation in the wild: isolation by dispersal limitation, isolation by adaptation and isolation by colonization. Mol Ecol 22:5983–5999. doi: 10.1111/mec.12561

Palma-Silva C, Wendt T, Pinheiro F, Barbará T, Fay MF, Cozzolino S and Lexer C (2011) Sympatric bromeliad species (*Pitcairnia* spp.) facilitate tests of mechanisms involved in species cohesion and reproductive isolation in Neotropical inselbergs. Mol Ecol 20:3185–3201. doi: 10.1111/j.1365-294X.2011.05143.x

Pinheiro F, De Barros F, Palma-Silva C, Meyer D, Fay MF, Suzuki RM, Lexer C and Cozzolino S (2010) Hybridization and introgression across different ploidy levels in the Neotropical orchids *Epidendrum fulgens* and *E. puniceoluteum* (Orchidaceae). Mol Ecol 19:3981–3994. doi: 10.1111/j.1365-294X.2010.04780.x

Prunier JG, Colyn M, Legendre X, Nimon KF and Flamand MC (2015) Multicollinearity in spatial genetics: separating the wheat from the chaff using commonality analyses. Mol Ecol 24:263–283. doi: 10.1111/mec.13029

Puechmaille SJ (2016) The program STRUCTURE does not reliably recover the correct population structure when sampling is uneven: sub-sampling and new estimators alleviate the problem. Mol Ecol Resour 16:608–627. doi: 10.1111/1755-0998.12512

Ramos-Fregonezi AM, Fregonezi JN, Cybis GB, Fagundes NJ, Bonatto SL and Freitas LB (2015) Were sea level changes during the Pleistocene in the South Atlantic Coastal Plain a driver of speciation in *Petunia* (Solanaceae)? BMC Evol Biol 15:92. doi: 10.1186/s12862-015-0363-8

Reck-Kortmann M, Silva-Arias GA, Segatto ALA, Mäder G, Bonatto SL and de Freitas LB (2014) Multilocus phylogeny reconstruction: New insights into the evolutionary history of the genus *Petunia*. Mol Phylogenet Evol 81:19–28. doi: 10.1016/j.ympev.2014.08.022

Reck-Kortmann M, Silva-Arias GA, Stehmann JR, Greppi JA and Freitas LB (2015) Phylogenetic relationships of *Petunia patagonica* (Solanaceae) revealed by molecular and biogeographical evidence. Phytotaxa 222:17. doi: 10.11646/phytotaxa.222.1.2

Richardson JL, Brady SP, Wang IJ and Spear SF (2016) Navigating the pitfalls and promise of landscape genetics. Mol Ecol 25:849–863. doi: 10.1111/mec.13527

Roullier C, Duputié A, Wennekes P, Benoit L, Fernández Bringas VM, Rossel G, Tay D, McKey D and Lebot V (2013) Disentangling the origins of cultivated sweet potato

(*Ipomoea batatas* (L.) Lam.). PLoS ONE 8:e62707. doi: 10.1371/journal.pone.0062707

Santos-Fischer CB dos, Corrêa ICS, Weschenfelder J, Torgan LC and Stone JR (2016) Paleoenvironmental insights into the Quaternary evolution of the southern Brazilian coast based on fossil and modern diatom assemblages. Palaeogeogr Palaeoclimatol Palaeoecol 446:108–124. doi: 10.1016/j.palaeo.2016.01.018

Schlötterer C (2000) Evolutionary dynamics of microsatellite DNA. Chromosoma 109:365–371. doi: 10.1007/s004120000089

Smouse PE, Long JC and Sokal RR (1986) Multiple regression and correlation extensions of the Mantel test of matrix correspondence. Syst Zool 35:627. doi: 10.2307/2413122

Sork VL, Nason J, Campbell DR and Fernandez JF (1999) Landscape approaches to historical and contemporary gene flow in plants. Trends Ecol Evol 14:219–224. doi: 10.1016/S0169-5347(98)01585-7

Stehmann JR (1999) Estudos taxonomicos na tribo Nicotianeae G. Don (Solanaceae) : revisão de *Petunia* Jussieu, das especies brasileiras de *Calibrachoa* La Llave & Lexarza e o estabelecimento do novo genero *Petuniopsis* Stehmann & Semir. Tese de Doutorado, Universidade Estadual de Campinas

Stehmann JR and Bohs L (2007) Nuevas combinaciones en Solanaceae. Darwiniana 45:240–241.

Stehmann JR, Lorenz-Lemke AP, Freitas LB and Semir J (2009) The genus *Petunia*. In: Gerats T and Strommer J (eds) *Petunia*: Evolutionary, developmental and physiological genetics. Springer Science Business Media, New York, pp 1–28

Stinchcombe JR and Hoekstra HE (2008) Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. Heredity 100:158–170. doi: 10.1038/sj.hdy.6800937

Storfer A, Murphy MA, Evans JS, Goldberg CS, Robinson S, Spear SF, Dezzani R, Delmelle E, Vierling L and Waits LP (2007) Putting the "landscape" in landscape genetics. Heredity 98:128–142. doi: 10.1038/sj.hdy.6800917

Storfer A, Murphy MA, Spear SF, Holderegger R and Waits LP (2010) Landscape genetics: where are we now? Mol Ecol 19:3496–3514. doi: 10.1111/j.1365-294X.2010.04691.x

Tomazelli LJ and Dillenburg SR (2007) Sedimentary facies and stratigraphy of a last interglacial coastal barrier in south Brazil. Mar Geol 244:33–45. doi: 10.1016/j.margeo.2007.06.002

Tomazelli LJ and Villwock JA (2005) Mapeamento geológico de planícies costeiras: o exemplo da costa do Rio Grande do Sul. Gravel 3:109–115.

Van Strien MJ, Keller D and Holderegger R (2012) A new analytical approach to landscape genetic modelling: least-cost transect analysis and linear mixed models. Mol Ecol 21:4010–4023. doi: 10.1111/j.1365-294X.2012.05687.x

Vandenbussche M, Chambrier P, Rodrigues Bento S and Morel P (2016) *Petunia*, your next supermodel? Front Plant Sci 7:72. doi: 10.3389/fpls.2016.00072

Vandergast AG, Perry WM, Lugo RV and Hathaway SA (2011) Genetic landscapes GIS Toolbox: tools to map patterns of genetic divergence and diversity. Mol Ecol Resour 11:158–161. doi: 10.1111/j.1755-0998.2010.02904.x

Vendruscolo GS (2009) Diversidade e distribuição de Solanaceae em formações vegetais altomontanas no sul do Brasil. Tese de Doutorado, Universidade Federal do Rio Grande do Sul

Viruel J, Catalán P and Segarra-Moragues JG (2012) Disrupted phylogeographical microsatellite and chloroplast DNA patterns indicate a vicariance rather than long-distance dispersal origin for the disjunct distribution of the Chilean endemic *Dioscorea biloba* (Dioscoreaceae) around the Atacama Desert. J Biogeogr 39:1073–1085. doi: 10.1111/j.1365-2699.2011.02658.x

Wang IJ (2013) Examining the full effects of landscape heterogeneity on spatial genetic variation: A multiple matrix regression approach for quantifying geographic and ecological isolation: special section. Evolution 67:3403–3411. doi: 10.1111/evo.12134

Wei X, Meng H and Jiang M (2013) Landscape genetic structure of a streamside tree species *Euptelea pleiospermum* (Eupteleaceae): Contrasting roles of river valley and mountain ridge. PLoS ONE 8:e66928. doi: 10.1371/journal.pone.0066928

Weschenfelder J, Corrêa ICS, Aliotta S and Baitelli R (2010) Paleochannels related to late Quaternary sea-level changes in southern Brazil. Braz J Oceanogr 58:35–44.