Research article

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

Gustavo A. Silva-Arias\textsuperscript{a}, Maikel Reck-Kortmann\textsuperscript{a}, Bryan C. Carstens\textsuperscript{b}, Heinrich Hasenack\textsuperscript{c}, Sandro L. Bonatto\textsuperscript{d}, Loreta B. Freitas\textsuperscript{a,⁎}

\textsuperscript{a} Laboratory of Molecular Evolution, Department of Genetics, Universidade Federal do Rio Grande do Sul, P.O. Box 15053, 91501-970, Porto Alegre, RS, Brazil
\textsuperscript{b} Department of Evolution, Ecology and Organismal Biology, Ohio State University, 318W. 12th Avenue, 43210, Columbus, OH, USA
\textsuperscript{c} Laboratory of Geoprocessing, Department of Ecology, Universidade Federal do Rio Grande do Sul, P.O. Box 15007, 91501-970, Porto Alegre, RS, Brazil
\textsuperscript{d} Laboratory of Genomic and Molecular Biology, Pontifícia Universidade Católica do Rio Grande do Sul, Av. Ipiranga 6661, 90619-900, Porto Alegre, RS, Brazil

A R T I C L E   I N F O

Keywords:
Central-marginal
Gene flow
Isolation by environment
Petunia
Serial founder effect

A B S T R A C T

The process of colonization and range expansion to novel environments involves both demographic and selective processes that can be traceable in the genetic diversity of the organisms. Nevertheless, it can be difficult to disentangle the signatures of the demographic and selective in the current genetic diversity of the populations. In this work, we use a landscape genetics framework to investigate the association of spatial and climatic variables with the genetic diversity and differentiation patterns of wild Petunia populations involved in a process of coastal colonization during the last 400 thousand years. Over 300 individuals from 17 populations were genotyped using ten microsatellite loci. Our results suggest that the genetic diversity is higher in populations located at the center of the species range with a decline toward the edges, and that gene flow follows an inland-to-coastal and central-to-peripheral dynamic that parallels the colonization history of this coastal lineage. We identify high levels of genetic differentiation between inland and coastal populations. As part of the inland-coastal genetic differentiation, we find signals of isolation by environment associated with differences in extreme temperature regimes. The differentiation of edge populations along the coast is associated with precipitation seasonality. Our results are robust to controls for spatial and historical divergence in the observed genetic differentiation. We conclude that the ecological differentiation process during the colonization of the South Atlantic Coastal Plain (SACP) was likely facilitated by genetic enrichment resulting from gene flow from central to marginal populations, as well as by rapid allele fixation resulting from serial founder effects during the range expansion along the coast.

1. Introduction

Investigations of diverging lineages that occupy regions with different ecological conditions enable researchers to understand how biotic or abiotic constraints can shape patterns of genetic differentiation across the geographic range of a species (Orsini et al., 2013). The distinction between isolation by distance (gradual genetic differentiation across populations as a product of limited dispersal), isolation by colonization (clear differentiation patterns resulting from founder effects), and isolation by environment (differentiation resulting from restricted gene flow between populations inhabiting different environments by selection against dispersers or by preference to remain in a particular environment due to local adaptation) is particularly challenging due to the superimposition of their respective signals in the genetic data (Perchaud and Hansen, 2016; Laurent et al., 2016). For example, when a species 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.

The complex interplay of spatial and ecological processes in determining the spatial distribution of a species can be assessed using a landscape genetics framework (Balkenhol et al., 2016; Manel et al., 2003). Landscape genetics is a field that has been developing rapidly due to the increasing sophistication of molecular methods and statistical analyses, as well as the growing availability of spatial ecological data (Manel and Holderegger, 2013; Storfer et al., 2010, 2007). Several factors may have an impact on the interpretation of the influence of space and environment on the genetic differentiation. These factors include the type of genetic marker used, the geographical range of the focal species, and the methods used to detect gene flow with respect to environmental gradients.
The South Atlantic Coastal Plain (SACP; Fig. 1; Fig. A.1, Appendix A) is a flat and continuous open coastal formation occupied mostly by large coastal lakes. Extending over approximately 600 km NE-SW in southern Brazil and Uruguay, SACP is crossed by only two perennial water channels and it is the largest coastal plain in South America (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 over the past 400,000 years. Each transgression-regression cycle led to the formation of a sand barrier running parallel to the coast (barrier-lagoon systems I–IV, Tomazelli et al., 2000; Weschenfelder et al., 2010; Fig. A.1, Appendix A). Due to this well-characterized history of cycles of exposure and flooding during the glacial and interglacial periods, it can be assumed that most organisms inhabiting this coastal plain, colonized it, and are expanding into and adapting to this new environment during the Pleistocene-Holocene period. The SACP therefore represents a promising model for the understanding of recent colonization processes.

Coastal environments offer particular climatic and edaphic conditions, such as salt exposition, temporal flooding, and seasonal or permanent strong winds. Additionally, the shape of coastal habitats is inherently linear, thereby conditioning the distribution of the taxa restricted to these regions. Moreover, global climatic changes can strongly affect the availability of suitable habitats for coastal organisms over time (Hoegh-Guldberg and Bruno, 2010; Reyer et al., 2013). Several studies have addressed the phylogeographical patterns of coastal organisms (Lopes et al., 2013; Mäder et al., 2013; Ramos-Fregonezi et al., 2015) and their distributional and demographic responses to climate change (Weising and Freitag, 2007; Turchetto-Zolet et al., 2016). Others have discussed physiological and molecular approaches to examine specific mechanisms of adaptation to saline environments (Lowry et al., 2009; Zhu, 2001). The study of the colonization 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 understand the adaptation to saline environments from an evolutionary point of view.

The aim of this study was to identify genetic signatures related to the recent coastal colonization process of a lineage of the Petunia genus (Solanaceae) to the SACP from a continental ancestor. To achieve this, we characterized the genetic diversity and structure of coastal populations and used a landscape genetics approach to examine alternative spatial and ecological drivers of genetic differentiation and gene flow in coastal populations. We specifically addressed the following questions: (1) Do historical and current gene flow explain current patterns of genetic diversity and structure? (2) Do environmental conditions constrain the gene flow between SACP populations?

2. Materials and methods

Petunia integrifolia subsp. depauperata (hereafter P. depauperata) is a diploid (2n = 14), prostrate annual herb, with purplish flowers pollinated by the solitary small bee species Callonychium petuniae (Wittmann et al., 1990), and also likely by other species of the genus Leioproctus and Calliopsis (Ando et al., 2001; Gütbitz et al., 2009), and dry dehiscent capsules that produce hundreds of tiny seeds with no dispersion mechanism. Ando et al. (2001) analyzed one population from São Lourenço do Sul and concluded that this taxon is self-incompatible. According to the strong wind potential in the SACP. There are no studies about gene flow in P. depauperata, but P. axillaris (a hawkmoth pollinated species) have been found pollen dispersal distances until 1013 m (Turchetto et al., 2015a).

P. depauperata is restricted to open sandy grasslands, dunes and rocky outcrops of lakeside or marine environments along the SACP. The P. depauperata lineage likely diverged around 400 thousand years ago (kya) from populations of its closest relative, P. integrifolia subsp. integrifolia (hereafter P. integrifolia) (Ramos-Fregonezi et al., 2015), a taxon that is widespread in inland subtropical grasslands and belongs to the P. integrifolia complex (Longo et al., 2014) that in previous
biogeographical reconstruction showed continental origin (Reck-Kortmann et al., 2014).

Previous phylogeographical assessments suggested that P. depauperata divergence could have begun in eastern peripheral populations of P. integrifolia since the divergence age of these two sister lineages is compatible with the estimate of the SACP origin (Ramos-Fregonezi et al., 2015). During a marine transgression around 400,000 years, several populations at the eastern extreme of the P. integrifolia distribution would have been temporarily isolated in non-submerged regions, such as the granitic hills around Porto Alegre. Afterward, during the subsequent marine regression, the first coastal populations may have established themselves in the fossil dune fields of barrier-lagoon system I (Fig. A.1, Appendix A). The divergence ages of the haplogroups of P. depauperata are compatible with the estimates of the ages of the depositional systems SACP, which supports that the diversification of these lineage occurred when the habitat was available for plant colonization (Ramos-Fregonezi et al., 2015).

The current known latitudinal distribution of P. depauperata follows the coastal line from a northern extreme around the city of Garopaba (Santa Catarina Brazilian state; ∼ 28 Lat S) to a southern extreme in La Corolla in Ronaldo (Rocha Uruguayan department; ∼ 34 Lat S). Populations of P. depauperata are found between the sea line and less than 90 km from the coast, with a few populations separated from the sea by big lagoons (Fig. 1).

2.2. Sampling

We sampled 307 individuals from 17 localities (hereafter referred to as populations): 236 individuals (12 populations) were identified as members of the coastal lineage (P. depauperata), with the remaining 71 individuals (five populations) classified as the inland lineage (P. integrifolia) based on morphological and environmental traits (Steinhann and Boys, 2007), and on plastid 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 sampled in fossil dune fields that were part of a coastal environment 400 kya. Leaves of all individuals were collected during the flowering seasons of 2002–2013 (September–February), preserved in silica gel until completely dried, ground and kept at ∼ 20 °C. The number of individuals per population varied from five to 33 (Table 1). For each population, one herbarium voucher was taken and deposited either in the ICN Herbarium, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil, or in the BHCB Herbarium, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil.

2.3. 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), which had been described and mapped in two Petunia hybrids (Bossoni et al., 2011), before being successfully transferred to wild Petunia species (Turchetto et al., 2013). The selected loci were located on six of the seven Petunia chromosomes (Bossoni 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 because they were found to be monomorphic in a preliminary screening of P. depauperata. For details of PCR and genotyping, see Appendix A.

2.4. Genetic diversity and structure

Basic diversity statistics, such as allele number (A), allelic richness (Ar, implementing rarefaction to account for different sample sizes), private allelic richness (pAr, with rarefaction), observed heterozygosity (H0), gene diversity (Hs), and inbreeding coefficient (FIS) were calculated using the packages ADEGENET 2 (Jombart, 2008; Jombart and Ahmed, 2011), POPPR 2.0.2 (Kamvar et al., 2015, 2014), and HIERFSTAT 0.04-14 (Goudet, 2014, 2005) for R 3.3.2 (R Core Team, 2016), ARLEQUIN 3.5 (Excoffier and Lischer, 2010), and HP-RARE (Kalinowski, 2005). Details can be found in Appendix A.

Genetic clustering analyses were performed using 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 × 10^5 burn-in periods and 1.0 × 10^6 Markov chain Monte Carlo repetitions after burn-in, under an admixture model, assuming correlated allele frequencies (Falush et al., 2003), and including the sampling locations

Table 1

<table>
<thead>
<tr>
<th>Population name</th>
<th>Longitude</th>
<th>Latitude</th>
<th>N</th>
<th>A</th>
<th>Ar</th>
<th>pAr</th>
<th>H0</th>
<th>Hs</th>
<th>Fis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inland lineage</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Itaara</td>
<td>−53.7967</td>
<td>−29.55151</td>
<td>8</td>
<td>39</td>
<td>3.49</td>
<td>0.09</td>
<td>0.68</td>
<td>0.69</td>
<td>0.11</td>
</tr>
<tr>
<td>Tapes</td>
<td>−51.3940</td>
<td>−30.66296</td>
<td>15</td>
<td>45</td>
<td>3.59</td>
<td>0.15</td>
<td>0.50</td>
<td>0.64</td>
<td>0.23</td>
</tr>
<tr>
<td>Arambarê</td>
<td>−51.49181</td>
<td>−30.90530</td>
<td>5</td>
<td>27</td>
<td>3.00</td>
<td>0.15</td>
<td>0.47</td>
<td>0.58</td>
<td>0.18</td>
</tr>
<tr>
<td>Guabiâ</td>
<td>−51.31768</td>
<td>−30.13769</td>
<td>23</td>
<td>62</td>
<td>4.26</td>
<td>0.35</td>
<td>0.58</td>
<td>0.67</td>
<td>0.12</td>
</tr>
<tr>
<td>Viamão</td>
<td>−51.00495</td>
<td>−30.38093</td>
<td>20</td>
<td>45</td>
<td>3.43</td>
<td>0.22</td>
<td>0.43</td>
<td>0.56</td>
<td>0.21</td>
</tr>
<tr>
<td>Coastal lineage</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Porto Alegre</td>
<td>−51.11773</td>
<td>−30.07462</td>
<td>14</td>
<td>49</td>
<td>4.06</td>
<td>0.22</td>
<td>0.51</td>
<td>0.69</td>
<td>0.26</td>
</tr>
<tr>
<td>Osório</td>
<td>−50.24286</td>
<td>−29.61331</td>
<td>26</td>
<td>48</td>
<td>3.53</td>
<td>0.14</td>
<td>0.49</td>
<td>0.58</td>
<td>0.15</td>
</tr>
<tr>
<td>Curumim</td>
<td>−49.94791</td>
<td>−29.64295</td>
<td>15</td>
<td>43</td>
<td>3.58</td>
<td>0.02</td>
<td>0.46</td>
<td>0.59</td>
<td>0.21</td>
</tr>
<tr>
<td>Xangri-lá</td>
<td>−50.02934</td>
<td>−29.79164</td>
<td>20</td>
<td>36</td>
<td>2.87</td>
<td>0.03</td>
<td>0.45</td>
<td>0.45</td>
<td>0.00</td>
</tr>
<tr>
<td>Pinhal</td>
<td>−50.23041</td>
<td>−30.24703</td>
<td>19</td>
<td>36</td>
<td>2.86</td>
<td>0.02</td>
<td>0.37</td>
<td>0.46</td>
<td>0.20</td>
</tr>
<tr>
<td>Mostardas</td>
<td>−50.73934</td>
<td>−30.93746</td>
<td>16</td>
<td>42</td>
<td>3.22</td>
<td>0.05</td>
<td>0.48</td>
<td>0.50</td>
<td>0.02</td>
</tr>
<tr>
<td>Torres</td>
<td>−49.73457</td>
<td>−29.35748</td>
<td>15</td>
<td>36</td>
<td>3.06</td>
<td>0.09</td>
<td>0.42</td>
<td>0.49</td>
<td>0.15</td>
</tr>
<tr>
<td>Garopaba</td>
<td>−48.62154</td>
<td>−28.02319</td>
<td>15</td>
<td>25</td>
<td>2.32</td>
<td>0.07</td>
<td>0.38</td>
<td>0.36</td>
<td>0.05</td>
</tr>
<tr>
<td>Rio Grande</td>
<td>−52.17406</td>
<td>−32.12531</td>
<td>33</td>
<td>33</td>
<td>2.60</td>
<td>0.01</td>
<td>0.40</td>
<td>0.43</td>
<td>0.07</td>
</tr>
<tr>
<td>Taim</td>
<td>−52.49174</td>
<td>−32.60474</td>
<td>17</td>
<td>34</td>
<td>2.95</td>
<td>0.03</td>
<td>0.45</td>
<td>0.50</td>
<td>0.09</td>
</tr>
<tr>
<td>São Lourenzo do Sul</td>
<td>−51.95317</td>
<td>−31.37673</td>
<td>21</td>
<td>40</td>
<td>3.19</td>
<td>0.02</td>
<td>0.52</td>
<td>0.58</td>
<td>0.10</td>
</tr>
<tr>
<td>Pelotas</td>
<td>−52.16478</td>
<td>−31.70776</td>
<td>25</td>
<td>40</td>
<td>3.17</td>
<td>0.13</td>
<td>0.41</td>
<td>0.53</td>
<td>0.23</td>
</tr>
<tr>
<td>Overall</td>
<td>307</td>
<td>93</td>
<td>469</td>
<td>0.54</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Superscripts in the population names indicate the groups depicted in Fig. A.3 (Appendix A) used for the comparison of historical gene flow models in MIGRATE-N (see Table 3);

i = inland, w = west, c = central, n = north, s = south, and sw = south west.

* Fis statistic significantly different from zero, p-value < 0.05.
as a 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 using few loci (Hubisz et al., 2009). The optimal K for each analysis was chosen using the AK method (Evanno et al., 2005) implemented in STRUCTURE HARVESTER (Earl and vonHoldt, 2012). CLUMPP 1.1.2 (Jakobsson and Rosenberg, 2007) was used to summarize the results of the ten runs of each chosen K based on Greedy’s method and the G-statistic. DISTRACT 1.1 (Rosenberg, 2003) was used to generate the respective bar plots.

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

2.5. Central-marginal tests of genetic diversity

Based on our current knowledge of the history of diversification of *P. depauperata*, the coastal lineages is thought to have originated in the granitic hills of Porto Alegre and on the ‘Coxilha das Lombas’ fossil dune fields. From there, the species is thought to have spread both to the north and to the south. This scenario leads to the hypothesis of decreasing genetic diversity with increasing distance from the region of origin, which we tested for here. Linear regressions between measures of genetic diversity (A, Ar, pAr, H0, and H1) and geographic distance to the origin were implemented in R, and tests of linear model assumptions were carried out using the GVLMA 1.0-0.2 package (Pena and Slate, 2014) in R.

2.6. Environmental variables

We built a new dataset of climatic surfaces specific for the SACP based on data taken from a dense sampling of weather stations in the studied area (Fig. A.2, Appendix A). The climatic surfaces were obtained through the interpolation of thin-plate splines (TPS) models for eight climatic variables using the R packages FIELDS 8.2-1 (Nychka et al., 2015) and RASTER 2.4-15 (Hijmans, 2015). We obtained the following climatic surfaces: total annual precipitation, total annual days with rain, precipitation seasonality (coefficient of variation), mean annual temperature, mean maximum summer temperature, mean minimum winter temperature, mean temperature range [mean of monthly (max temp–min temp)], and temperature seasonality (coefficient of variation). For methodological details, see Appendix A. We chose to generate a local data set for our landscape genetic analyses instead of relying on the widely-used WorldClim climatic surfaces, since the density of weather stations in the target area contributing primary data to Hijmans et al. (2005) was very low, an issue that could be problematic for the spatial scale of this study.

2.7. Landscape genetics analyses

We estimated the inter-population genetic differentiation by calculating matrices of pairwise linearized $F_{ST}$ ($F_{ST}/1- F_{ST}$; Rousset, 1997) using SPAGeDi (Hardy and Vekemans, 2002). We also obtained the conditional genetic distance matrix (cGD; Dyer and Nason, 2004) of the shortest paths between population pairs inferred from a population graph representing genetic correlation that better capture the information on gene flow between populations. cGD distances were obtained using the packages GENETICSTUDIO and POPGRAPH (Dyer, 2014, 2009) in R.

The hypothesis of isolation by distance (IBD) was assessed by linear regression of the calculated genetic distances and the geographic distance (logarithm transformed when regressed with $F_{ST}$) using a multiple regression on distance matrices (MRDM; Legendre et al., 1994) implemented in the R library ECODIST (Goslee and Urban, 2007). Inter-population geographic distances were calculated from UTM 22S coordinates (coordinate reference EPSG: 32722) transformed from Long/Lat coordinates with RGDAL 1.0-4 R package (Bivand et al., 2015).

Based on our new set of local climatic surfaces (see Environmental variables section above), we explored isolation by environment (IBE) patterns in the genetic differentiation by means of MRDM analysis. Has been suggested to take historical patterns into account when using current genetic differentiation measures in landscape genetic approaches (Anderson et al., 2010; Epps and Keyghobadi, 2015). Therefore, we implemented our IBE tests using adjusted inter-population genetic distance matrices that account for possible effects of historical differentiation processes. For that, following Dyer et al. (2010), we built the adjusted genetic distance matrices with the residuals taken from regression analysis of the observed genetic distance matrices ($F_{ST}$ and cGD) on a binary matrix representing the phylogeographic patterns recovered from Ramos-Fregonezi et al. (2015). For the phylogeographic matrix, a value of 0 denoted populations from the same haplogroup (i.e. connected populations), and a value of 1 denoted populations from different haplogroups (i.e. unconnected populations) (Fig. A.3, Appendix A). In addition, to account for the spatial autocorrelation of genetic and climatic data, the matrix of Euclidean geographic distances was also included in the regression.

Previously to analyses, all matrices were standardized to a mean of 0 and a variance of 1. Statistical significance to all implemented MRDM analyses was assessed through 10,000 randomizations. The Itaara population was excluded from all spatial and environmental analyses because it is most distant from the remaining populations; it is also located outside the SACP.

To complement the IBE tests described above, we explored the influence of the climatic variables on the functional inter-population connectivity using a circuit theory approach (McRae, 2005). To that end, climatic layers were converted into landscape resistance surfaces by means of an empirical optimization procedure that avoids subjectivity with costs assigned by expert opinion (Richardson et al., 2016). We used the nonlinear optimization algorithm implemented in the R package RESISTANCEGA (Peterman, 2014). Briefly, the parameter space for the determination of the optimal resistance values is extensively explored through monomolecular and Ricker functions. Each iteration uses CIRCUITSCAPE (McRae and Beier, 2007) to calculate inter-population pairwise resistance distances. A genetic algorithm is used to maximize the fit of resistance values to the pairwise genetic distances, minimizing the AICc values of linear mixed – maximum likelihood population effects models (MLPE; Clarke et al., 2002). Final optimized resistance MLPE models for each assessed climatic variable were ranked by ΔAICc to identify the climatic variables that best explain the observed patterns of genetic differentiation.

2.8. Test for alternative historical migration models

The coalescent-based program MIGRATE-N 3.2.6 (Beerli and Felsenstein, 2001) was used to test the support provided by the genetic data to alternative gene flow scenarios for the studied populations. For this, the populations were pooled into six groups according to their geographical distribution and genetic structure. The population groups were named: ’Inland’ (population from Itaara), ’Basal’ (populations from Tapes, Arambarê, Guaíba, Viamão, and Porto Alegre), ’North’ (populations from Garopaba and Torres), ’Center’ (populations from Osório, Curumim, Xangri-lá, Pinhal, and Mostardas), ’South 1’ (populations from São Lourenço do Sul and Pelotas), and ’South 2’ (populations from Rio Grande and Taim) (Fig. A.4, Appendix A).

Seven unidirectional migration models were evaluated to test
different gene flow scenarios, including patterns of central-marginal, marginal-central, different source-sink dynamics, and restricted migration between nearby populations. The models are graphically represented in Fig. A.4 (Appendix A). For each model, we ran three independent runs of 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 with a burn-in of 10,000 steps. For each model, the mutation scaled effective population size (θW) and the migration rate (M) were drawn from uniform prior distributions ranging from 0 to 10, and from 0 to 50, 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 this produces better marginal likelihood estimates being less influenced by prior parameter distributions (Beerli and Palczewski, 2010).

3. Results

3.1. 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 a departure from HWE (P < 0.01) (Table B.1, Appendix B). A significant signal of linkage disequilibrium (P < 0.01) was detected for several pairs of loci, but since these were not related to their putative chromosome positions (Table B.2, Appendix B), and since the pattern of linkage disequilibrium was not constant for all populations, we maintained all loci for the downstream analyses. Positive and significant inbreeding coefficients (FIS) were found for seven populations (Table 1). The MICROCHECKER analysis showed a homozygote excess for locus PM167; this may be attributed to null alleles or stutter peaks. Analyses including or excluding PM167 gave the same results.

3.2. Genetic structure

The genetic structure revealed by the STRUCTURE and DAPC analyses showed congruent results. Nevertheless, STRUCTURE identified the optimal number of clusters as K = 2, whereas BIC scores of DAPC showed a sharp decline between K = 2 and K = 7. We therefore opted to explore genetic structure based on all K values between 2 and 7 (Figs. B.1 and B.2; Appendix B).

For K = 2, we found a cluster containing the inland population (Itaara), four populations from the fossil dune fields (Guaíba, Viamão, Tapes, and Arambaré), and the Porto Alegre population. The second cluster grouped all coastal populations (located east of the Patos Lagoon) and São Lourenço do Sul and Pelotas populations (located south west of the Patos Lagoon). This clustering mainly parallels the differentiation between the inland (P. integrifolia) and coastal (P. depauperata) lineages, with exception of the Porto Alegre population, which was grouped with the inland populations in our analysis (Figs. B.1 and B.2; Appendix B).

Additional groupings ranging from K = 3 to 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 on the west of the Patos Lagoon). This clustering mainly parallels the differentiation between the inland (P. integrifolia) and coastal (P. depauperata) lineages, with exception of the Porto Alegre population, which was grouped with the inland populations in our analysis (Figs. B.1 and B.2; Appendix B).
the Mostardas population. The Torres population, located between the central and northern regions of the SACP, showed more affinity with populations from the central region (Figs. B.1 and B.2; Appendix B).

3.3. Central-marginal pattern of genetic diversity

All linear regressions between the genetic diversity statistics A, Ar, pAr, Hs, and Hc and population distances to the region of origin of *P. depauperata* showed a significant negative relationship, supporting a central-marginal pattern of genetic diversity decline. The strongest correlations were observed for Ar (Adjusted $R^2 = 0.49$; p-value = 0.001), followed by A (Adjusted $R^2 = 0.40$; p-value = 0.005), Hs (Adjusted $R^2 = 0.41$; p-value = 0.005), pAr (Adjusted $R^2 = 0.25$; p-value = 0.029) and Hc (Adjusted $R^2 = 0.20$; p-value = 0.048) (Fig. 2).

3.4. Contemporary genetic differentiation of *P. depauperata*

The global $F_{ST}$ was 0.19. Pairwise population differentiation measures ranged from 0.02 (between Pelotas and São Lourenço do Sul) to 0.67 (between Arambaré and Garopaba; Fig. 3A), cGD distances ranged from 0.02 (between Pelotas and São Lourenço do Sul) to 52.5 (between Viamão and Mostardas; Fig. 3B). Broadly, inter-population $F_{ST}$ values showed the highest level of differentiation for southern and northern edge populations, while cGD values resulted in a pattern of strong differentiation between inland and coastal populations, in line with the results of the structure analysis for K = 2. The population network topology showed the same grouping of inland and coastal populations, and a geographic pattern except for the marginal coastal populations (Garopaba and Taim; Fig. 3C).

MRDM analyses showed significant linear relationships between geographic distance and matrices of genetic differentiation as measured by $F_{ST}$ ($R^2 = 0.24$, $\beta_{geo} = 1.7$, $P < 0.001$) and cGD ($R^2 = 0.03$, $\beta_{geo} = 0.13$, $P = 0.05$). These results support an IBD pattern for both measures of genetic differentiation, but more strongly captured by the $F_{ST}$. The regression analysis using the phylogeographic matrix showed significant coefficients for both $F_{ST}$ ($R^2 = 0.06$, $\beta_{phylo} = 0.44$, $P = 0.007$) and cGD ($R^2 = 0.07$, $\beta_{phylo} = 0.41$, $P = 0.006$) matrices, suggesting that the consideration of historical processes is important because it explains a significant portion of the existing genetic covariance.

MRDM analyses testing for IBE and accounting for the phylogeographic history and IBD, showed a significant relationship between several climatic variables and the measures of genetic differentiation. While we found a significant relationship with precipitation seasonality for the $F_{ST}$ ($R^2 = 0.34$, $\beta = 0.41$, $P < 0.001$), for the cGD, we found a significant relationship with mean maximum summer temperature ($R^2 = 0.04$, $\beta = 0.16$, $P = 0.03$) and mean temperature range ($R^2 = 0.04$, $\beta = 0.16$, $P = 0.04$; Table 2).

In line with the tests for IBE, the optimization of the resistance layers showed different best predictors for the $F_{ST}$ and cGD differentiation measures. For the $F_{ST}$, the variables resulting in the largest association coefficient and the lowest AICc values in the optimized models were precipitation seasonality and mean temperature range. For cGD, the optimization procedure found the largest association coefficient and lowest AICc value for the mean maximum summer temperature (Table 3; Appendix B).

3.5. Migration models

The comparison of the seven historical migration models (Fig. A.4, Appendix A) in MIGRATE-N suggested that our genetic data are best explained by models representing inland-to-coast and central-to-marginal migration patterns (Table 3). Migration rates estimated from the best supported model were highest from ‘Central’ to ‘South 1’ (median = 41.9; CI 95% 36.1–48.7) and from ‘Central’ to ‘South 2’ (median = 30.3; CI 95% 18.7–41.5). ‘Central’ to ‘North’ showed strongly lower migration values (median = 4.1; CI 95% 0.3–2.5).
Migration from 'Inland' to 'West' (median = 1.4; CI 95% 0.2–2.4) as well as from 'West' to 'Central' (median = 1.2; CI 95% 0.3–2.1) was also lower.

4. Discussion

In this study, we assessed the genetic diversity and structure of coastal populations to infer contemporary patterns of gene flow and their relationship with spatial and ecological variables. Our results support both historical and contemporary processes as the source of current patterns of genetic differentiation. The results also support associations with current environmental conditions along two main axes of differentiation within the SACP: (1) an inland/coastal differentiation is supported by the uppermost hierarchy of genetic structure (K = 2) and the conditional genetic distance (cGD) obtained from population network topology; it was found to be associated with differences in maximum temperature regimes; (2) a central/marginal differentiation pattern within the coastal populations that is supported by the lowest hierarchies of genetic structure (K = 6 and K = 7) and by the FSG; it was found to be related to differences in the precipitation seasonality at the edges of the distribution along the coast.

4.1. Inland-coastal genetic differentiation

Several features linked to the colonization of the SACP by *P. depauperata* are likely to have contributed to the current patterns of genetic structure in SACP populations of this species. Most prominently, features related to the distance from the shoreline could be associated with the high levels of differentiation between inland and coastal populations, which were consistently detected by our genetic structure analyses (Figs. B.1 and B.2; Appendix B).

Both historical and contemporary processes might be related to the strong differentiation between inland and coastal populations. Despite the short geographical distance between some inland and coastal populations (i.e., 'West' and 'Central' population groups in Fig. A4; Appendix A), there might be important barriers to gene flow between them. The interaction between demographic processes (such as the founder effect) and selective forces (such as ecological divergence) can reduce gene flow and enhance the stochastic effect of genetic drift (Nosil et al., 2009) and thereby result in a genome-wide differentiation. This process, which Orsini et al. (2013) named as isolation by colonization, could have played a role in the divergence of *P. depauperata*.

Colonization implies an important reduction in population size and a dramatic reduction in genetic diversity or founder effect, so it is therefore expected (Pannell and Dorken, 2006). The diversity values found in this study support a genetic signature of colonization in *P. depauperata*. Even though we included fewer inland populations with fewer samples, we found a higher genetic diversity values for these populations than for the coastal populations (Table 1, Fig. 2). Additionally, the high level of genetic differentiation observed between inland-fossil dunes populations on the one hand, and coastal populations on the other (Figs. B.1 and B.2; Appendix B) might be the result of a founder effect as part of the coastal colonization by populations from the inland.

In the specific case of Porto Alegre population that, despite been considered with the *P. depauperata* subspecies (and present chloroplast haplotypes from the coastal lineage; Ramos-Fregonezi et al., 2015), was grouped with the populations from inland lineage. It is plausible that during the marine transgressions, Porto Alegre and fossil dunes regions (Fig. 1, Fig. A.1, Appendix A) had suitable conditions for the divergence and establishment of the coastal lineage, but under contemporary sea levels those areas turned into continental climatic conditions. Therefore, higher gene flow from the inland lineage populations (*P. integrifolia* subspecies) 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, but some of them (as the case of Porto Alegre population) currently preserving the chloroplast of the coastal lineage since plastid markers present reduced gene flow and mutation rates (Lorenz-Lemke et al., 2010).

Our results showed IBE signals for two climatic variables summarizing extreme temperature regimes (summer mean maximum temperature and mean temperature range), which could be related to the genetic differentiation between inland and coastal populations (Fig. 4). Summer seasonal drought or higher evaporations related to higher temperatures could be involved in the reduction of gene flow between inland and coastal populations in *P. depauperata*, as has been described for *Mimulus guttatus* (Lowry et al., 2008). Likewise, differences in salinity could be related to the maintenance of the genetic differentiation between inland and coast. The main morphological differences between *P. integrifolia* and *P. depauperata* lineages are traits commonly associated to salinity and sand-soil adaptations (Stehmann and Bohs, 2007). Differential salt tolerance between natural populations of plants from saline and non-saline environments has been
documented in *Trifolium repens* (Ab-Shukor et al., 1988) and *M. guttatus*, partially as a product of selection against migrants (Lowry et al., 2009, 2008). Adaptation to coastal environments can be achieved through the recruitment of different genes involved in similar processes, and this variation can be observed even at the intra-specific level (Roda et al., 2013).

4.2. Northward and southward spread of *P. depauperata* within the SACP

We found a consistent pattern of declining genetic diversity towards the northern and southern edges of the *P. depauperata* range (Fig. 2). In addition, marginal coastal populations showed higher levels of genetic differentiation (Fig. 3) and formed differentiated groups in the clustering analyses (Figs. B.1 and B.2; Appendix B). These patterns could have been shaped via demographic processes as part of the species’ expansion along the coast. Serial founder effects during range expansion (Slatkin and Excoffier, 2012) following a ‘propagule pool’ colonization model (Slatkin, 1977) explain the observed patterns in *P. depauperata* well. The process could have involved waves of occupancy following defined periods of substrate deposition on the SACP during the sea transgression-regression cycles over the past 400,000 years (Fig. A.1, Appendix A). During periods of sea transgression, the eastern-most populations could have been isolated in the central region of the SACP. The increased altitude and substrate availability of the central SACP would have allowed the survival of founder populations and the accumulation of genetic diversity. Substrate was progressively deposited, allowing subsequent colonization waves and thereby creating a serial founder effect, giving rise to new populations with a progressively lower genetic diversity (Fig. 2).

This scenario is congruent with the MIGRATE-N results supporting a model of historical gene flow from inland-to-coastal and central-to-marginal populations. This gene flow dynamic could have facilitated
the colonization process by increasing genetic variation and thereby allowing new populations to persist thanks to genetic diversity enrichment from the core (Hampe et al., 2013; Sexton et al., 2009).

In addition, we found a consistent signal of IBE related to the differentiation of marginal populations on the coast and precipitation seasonality. A possible local adaptation process could be only detected with our genetic data set in the case of maladapted immigrants do not successfully breed with individuals from ecologically diverged populations causing genome-wide differentiation (Sexton et al., 2014). The occurrence of adaptive divergence is not too surprising, at least in the northernmost populations, due to the contrasting climatic conditions compared to the other populations of P. depauperata. The northern SACP is under the strong influence of orographic rainfalls during the spring and summer seasons, the time of flowering and seed dispersion.

It has been thought that selective processes are less likely to take place in smaller, marginal populations due to the more pronounced effect of genetic drift, which counteracts selection by eliminating possible advantageous mutations, and due to the swamping effect of the gene flow from larger populations (Willi et al., 2006). However, recent evidence has shown that selective processes can be important in marginal populations even in the presence of gene flow, since adaptive alleles can emerge in central populations and move to the edges, where local adaptation can occur (Rolland et al., 2015; Tigano and Friesen, 2016). Likewise, the fixation of emerging advantageous mutations could be accelerated in marginal founder populations (Aguilée et al., 2009). Evidence for this last process has been found in Myodes glareolus, where signals of adaptation during range expansion have been detected despite the loss of genetic diversity through genetic drift at the edges of the distribution (White et al., 2013). Several studies in P. axillaris suggested shifts from self-incompatibility to self-compatibility in marginal populations (Kokubun et al., 2006; Turchetto et al., 2015a). That shift can be advantageous for preventing local extinction in low-density populations as could be found in founder populations and at edges of the species distribution. In addition, selfing can increase the fixation rate of advantageous mutations and then enhance local adaptation processes to marginal sub-optimal environments. Whether this kind process in marginal populations is also occurring in P. depauperata, it could be an important factor for both their origin (i.e. coastal colonization) and the differentiation processes towards the edges (northern and southern populations).

Further studies in P. depauperata involving genomics, transcriptomics or target genes should be implemented to improve our knowledge of the genetic bases of the colonization of coastal environments, as well as to test the relationship between climatic conditions and the genetic differentiation between populations of the SACP. This will be helped by newly available resources for Petunia genomics, including the recently published whole-genome sequence of two Petunia species (Bombarely et al., 2016) and the transcriptome analysis of salinity stress genes in Petunia hybridra (Villarino et al., 2014).

5. Conclusions

Our results consistently support relationships between extreme temperature regimes and differences in precipitation seasonality with current genetic differentiation as part of the colonization of a coastal region by Petunia depauperata. Those findings let us to suggest that differential climatic conditions constrain gene flow dynamics and thereby likely promote ecological divergence within the SACP. The observed genetic diversity supports a central-marginal pattern of diversity decline that is consistent with a gradual colonization process from the central region of the coastal plain towards the northern and southern edges of the current distribution of P. depauperata. This study allows new insights about a possible ecological differentiation processes during a coastal colonization by a plant species. We suggest a detailed investigation of the processes discussed in further studies.

Funding

This work 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 these species are not federally listed as endangered or protected, and because no collection sites were in protected areas.

Data accessibility

All original data are available under request to the first author (G.A. Silva-Arias; e-mail: gasilvaarias@gmail.com) or to the corresponding author (L.B. Freitas).

Author contributions

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

Acknowledgments

The authors thank A.M.C. Ramos-Fregonezi, J.N. Fregonezi, and G. Mäder (UFRGS) for help with the fieldwork; C. Turchetto (UFRGS) for assistance during laboratory work; R. Hijmans (UCDAVIS) for providing the R codes for the climatic variable interpolation; G. Hofmann (UFRGS) for help with the climatic data collection; and N.J.R. Fagundes (UFRGS), E. Eizirik (PUCRS), M.B. Lovato (UFMG), and members of the Carstens lab (OSU) for comments and discussion regarding the first version of this manuscript. B. Kremeyer, Ph.D., provided language help and proof reading. We also thank two anonymous reviewers for comments and suggestions that improved the manuscript.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jpees.2017.06.006.

References


