Research Article

Landscape and climatic features drive genetic differentiation processes in a South American coastal plant

GUSTAVO A. SILVA-ARIAS[1,2,*], LINA CABALLERO-VILLALOBOS[2], GIOVANNA C. GIUDICELLI[2], LORETA B. FREITAS[2,*]

[1]*Professorship for Population Genetics, Department of Life Science Systems, Technical University of Munich Freising, Germany.*

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

[*]Correspondence. E-mail: gustavo.silva@tum.de; loreta.freitas@ufrgs.br

Genetic differentiation along South America's coastline

ABSTRACT

- **Background*:*** Historical and ecological processes shape patterns of genetic diversity in plant species. Colonization to new environments and geographical landscape features determine, amongst other factors, genetic diversity within- and differentiation between-populations. We analyse the genetic diversity and population structure of *Calibrachoa heterophylla* to infer the influence of abiotic landscape features on the level of gene flow in this coastal species of the South Atlantic Coastal Plain.
- **Results*:*** The *C. heterophylla* populations located on early-deposited coastal plain regions show higher genetic diversity than those closer to the sea. The genetic differentiation follows a pattern of isolation-by-distance. Landscape features, such as water bodies and wind corridors, and geographical distances equally explain the observed genetic differentiation, whereas the precipitation seasonality exhibits a strong signal for isolation-by-environment in marginal populations. The estimated levels of gene flow suggest that marginal populations had restricted immigration rates enhancing differentiation.
- **Conclusions*:*** Topographical features related to coastal plain deposition history influence population differentiation in *C. heterophylla*. Gene flow is mainly restricted to nearby populations and facilitated by wind fields, albeit without any apparent influence of large water bodies. Furthermore, differential rainfall regimes in marginal populations seem to promote genetic differentiation.

Keywords: *Calibrachoa heterophylla,* colonization, gene flow, genetic structure, landscape genetics, Solanaceae, South Atlantic Coastal Plain.

BACKGROUND

Coastal areas in South America constitute distinct landscapes with unique abiotic and biotic compositions. Many geomorphological, climate, oceanographic features, and colonization events from the surrounding biomes shape the ecosystems of these areas [1–6]. Therefore, South American coastal flora shows a peculiar diversity of species, ecosystems, and various biogeographical processes shaping population demography [7, 8]. Although studies on plant diversification in South America have received increased attention, analyses focusing on the colonization of coastal areas, migration and gene flow between populations, and recent speciation events are still scarce [9–11]. There is indeed a lack of studies assessing the genetic diversity of wild plants from open areas such as sand dune and grassland plant communities from South-American coastal environments [12–14]. There are few works linking landscape genetics and phylogeography for coastal plant species in South America. Moreover, several gaps remain for fully interpreting population structure on spatially correlated genetic differentiation [15]. Disentangling the factors influencing gene flow is also important for understanding evolutionary dynamics at the extremes of species distribution [16] where processes such as local adaptation or peripatric speciation occur.

The species' geographical distribution and genetic (nucleotide) diversity result from historical and contemporary processes acting together with ecological factors [17–19]. Colonization of new habitats and subsequent genetic isolation are critical events in the eco-evolutionary dynamics of coastal plant populations [20]. It is possible to reconstruct such events because the spread to new environments generates footprints on the genetic diversity and genetic spatial structure of populations [21]. Coastal regions have common environmental characteristics, such

as intrinsic linear distributions, high salinity, wind strength, and tidal influence, which make these regions exciting models for studying genetic differentiation in response to climatic changes, changes in physical barriers, and in ecological features [22–24].

Calibrachoa heterophylla is a perennial nightshade shrub growing in dunes and sandy grasslands predominantly along the South Atlantic Coastal Plain (SACP). Previously phylogeographical assessment based on plastid markers supports that the species likely originated and subsequently diversified into four intraspecific lineages between 1 - 0.85 Mya [14]. These lineages have remained isolated by riverine barriers until their recent expansions following the formation of the SACP (400 – 7 kya), which determines their current geographical range. Currently, the species shows a continuous distribution along the SACP with a strong spatial genetic structure on the plastid markers albeit without conspicuous geographical barriers separating the intraspecific lineages [14]. This raises the questions whether nuclear polymorphism resembles the observed patterns in the chloroplast and if contemporary landscape features affect and shape the genetic structure of the species. We addressed these questions analysing a new set of polymorphic nuclear microsatellite markers and a comprehensive set of topographical and environmental variables in a spatial explicit framework.

The SACP is a flat, continuous, and open region constituting the most extensive coastal region in South America. The region extends NE-SW for approximately 600 km, is occupied mostly by large coastal lakes, and is crossed by two perennial water channels [25, 26]. This coastal formation gradually arose during sea-level transgressions and regressions caused by glacial-interglacial cycles during the last 400 ky. The most substantial transgression and regression cycles led to the formation of four main sand barriers that are positioned parallel to the coastline

(barrier-lagoon systems I to IV; [25, 27]). Harsh environmental features such as strong spring-summer sea breezes from the Northeast and high insolation (solar irradiance) strongly influence the SACP [28] and consequently define suitable habitats for endemic plant species (e.g., [13, 14, 29]).

Currently, there are few studies assessing the genetic diversity of wild plants from the South American coastal environments (see [30, 31] for few examples). Even less works explicitly evaluate the relative influence of physical landscape (i.e., distance and topographical features) and environment on the genetic differentiation [13, 32]. Therefore, a landscape genetics assessment for *C. heterophylla* is useful to bring new insights to understand the evolution of coastal plants whereascomplementing the historical divergence processes described in *Petunia integrifolia* and *C. heterophylla* [13, 14, 29]. We hypothesize that the SACP colonization process altered the historical pattern of genetic structure shaped before the SACP deposition because the lineages came into contact due to the absence of strong geographic barriers along the SACP. Moreover, we expect that the geographical distance and differential features of the physical environment along the SACP, such as the age of the barrier-lagoon deposition, the presence of big water bodies, wind strength, and climatic gradients shaped local patterns of population admixture.

This study aims to understand the forces responsible for the current patterns of genetic structure in the coastal nightshade *C. heterophylla*. Based on an evaluation of relevant environmental and topographical features of the SACP and the analysis of polymorphic microsatellite markers, we (I) identify and infer the parameters of contemporary and historical factors promoting genetic divergence (colonization process, demographic process, rates of gene flow) and (II) assess which topographical and climatic factors determined the population differentiation and gene

flow during the recent colonization of the SACP. We discuss the results in the light of relevant drivers of genetic diversification already identified for SACP species to find general scenarios shaping evolutionary trajectories of coastal plants in South America.

RESULTS

Genetic diversity

We found 140 alleles across ten microsatellite loci. The mean number of alleles per locus was 14, ranging from seven (Che59) to 17 (Che81). All loci showed higher H_e than H_o (Fig. S1) with 25% of the locus-population combinations showing a departure of HWE ($P < 0.05$). We detected a significant linkage disequilibrium signal ($P < 0.01$) for several loci pairs, however as the linkage pattern was not consistent across populations for any loci pair, we assumed linkage equilibrium and maintained all loci in the analyses. MICRO-CHECKER analysis did not show evidence of null alleles, scoring errors, or stutter peaks for any locus.

Populations located outside of SACP (I1-3) and those collected around the Patos Lagoon (W1-2 and S1-2) showed higher genetic diversity (Fig. 1B). In contrast, the coastal populations located at the northern and southern edges of species distribution in SACP (N1 and S6) showed lower genetic diversity values. Average H_o values across loci ranged from 0.72 (I1) to 0.31 (N1) and for H_e from 0.74 (I3) to 0.48 (S3). We found 22% of the alleles restricted to a single population, with W1 showing the highest number of private alleles (eight), whereas W3, S1, S3, and S4 populations had no private alleles. Garza-Williamson values ranged from 0.39 (I2) to 0.83 (N2). We found positive and significant F_{IS} values for five populations (Table 1),

all of them located at the borders (northern and southern) of species' distribution inside the SACP (Fig. 1).

Genetic structure

The recovered population structure showed a concordant geographic signal for marginal populations and higher admixed membership for populations located in geographical transitional regions (Fig. 2; Fig. S3). The best K = 4 was inferred from the ΔK method in STRUCTURE (Fig. S2A), whereas the DIC values from TESS showed the lowest standard for K = 2-4 runs and reach a plateau after maxK = 8 (Fig. S2B). DAPC showed the lowest BIC score for K = 8 (Fig. S4). Results obtained with all approaches showed consistent clustering of three to four well-differentiated groups.

Populations from the North of the SACP (N1-3) became the most differentiated group supported by the K = 2 clustering of STRUCTURE (Fig. 2) and TESS (Fig. S3) analyses, and the two main axes of both DAPC and sPCA (Fig. 1C,D). Considering three clusters, all approaches consistently recovered one group for the northern coastal populations (N1-3), a second group for the southern coastal populations (S3-6), and the third cluster for the three inland populations (I1-3) and the populations from the West side of the Patos Lagoon (W1-3). The two remaining populations (S1-2) showed a higher affinity with the Inland-West group in the exploratory analyses (Fig. 1C-D) and highly admixed memberships in the Bayesian clustering methods (Fig. 2, S3).

Migration rates

The mean migration rate estimated with BAYESASS was 0.015. However, only four population pairs showed higher posterior effective migration rates and confidence intervals above zero. Among them, the most outstanding was S2 -> W2 ($N_m \approx 0.08$; 95% CI 0.01 - 0.14), supporting migration between populations separated by the Patos Lagoon. The other three cases involved neighbour populations, I2 -> I3 ($N_m \approx 0.12$; 95% CI 0.05-0.19), S2 -> S1 ($N_m \approx 0.07$; 95% CI 0.01-0.13), and S4 -> S3 ($N_m \approx 0.16$; 95% CI 0.09-0.22). Migration estimates obtained from independent runs of BAYESASS showed similar values (Table S1).

The model-based coalescent approach implemented in MIGRATE-N supported the *step-stone from coast* as the most likely historical migration model between population groups (Table 2; Fig. S5D). Parameter estimation showed that the 'Inland' group had the highest mean θ, which was around eight times higher than the θ estimated for 'West' and 'North' groups, and around 20 times higher than the θ estimated for the 'South' group (Table 2). Migration from 'West' to 'Inland' showed the highest mean M being two times higher than the 'North' to 'West' and five times higher than the 'South' to 'West' values (Table 2). We verified that all estimated parameter estimation procedures did reach convergence (effective sample > 10 000 and posterior estimates showed unimodal distribution; Additional file 1).

Isolation-by-distance, isolation-by-environment, and resistance tests

Measures of population differentiation F_{ST} ranged from 0.01 (S1 - S2 populations) to 0.54 (N1 - S3 populations; Fig. 3A). The Mantel test also supported a positive correlation between the genetic and geographical distance matrices (Mantel's r = 0.38, P < 0.001). We then explored relevant landscape and climatic features

throughout the SACP as potential determinants of genetic differentiation based on the MMRR approach.

Further IBD tests assessing topographic cost distances models showed that the *continuous model* (landscape matrix with no topographic discontinuities) explained slightly better the genetic differentiation than the *water bodies model* (landscape matrix with water bodies as full barriers to population connectivity) (R^2 = 0.16, β = 0.022, P = 0.006 and R^2 = 0.14, β = 0.02, P = 0.011; respectively; Fig. S6). The relationship among climate variables and genetic differentiation including geographical distance showed significant association only for precipitation seasonality (Fig. 3C; R^2 = 0.35, P = 0.003; $\beta_{precseason}$ = 7.5 × 10-3, P = 0.05; β_{Euc} = 1.1 × 10-7, P = 0.02, respectively).

The "windscape" connectivity matrix accounting for both wind strength and direction measures (Figs. 4B, C) showed a North-to-South asymmetric step-stone pattern where marginal populations resulted strongly isolated and showing more intense wind influence at the central part of the SACP. Moreover, populations located at the West side of the Patos Lagoon became receptors from coastline populations. The coast distance wind matrix showed significant correlation with the F_{ST} genetic distance matrix (R^2 = 0.19, β = 0.001, P = 0.0037; Fig. 4A).

DISCUSSION

In this study we analyse the genetic diversity and structure of *Calibrachoa heterophylla* to infer the influence of topographical and environmental features on the population differentiation during the recent colonization of a coastal region in South America. The results support both contemporary and historical factors promoting genetic divergence throughout populations of a coastal plant species. Here, we

provide consistent evidence for limited and asymmetric gene flow, mainly restricted by geographical distance. The populations from northern and southern edges of the species distribution show negligible historical and contemporary immigration rates related to historical and geographical isolation. We also found that one of the most outstanding topographical feature in the SACP, namely the large water bodies, does not constrain *C. heterophylla* populations' gene flow. Gene flow seems promoted by the wind, at least between adjacent populations from the central portion of the SACP. Our results highlight the importance of considering both the physical landscape (contemporary) and phylogeographical context (historical) processes for complete interpretation of genetic differentiation processes.

Role of historical, spatial, and environmental factors on the genetic differentiation in *Calibrachoa heterophylla*

There is a hierarchical pattern of genetic structure related to both historical and contemporary landscape features. The main clustering pattern mirrors the phylogeographical structure of *C. heterophylla* previously recovered with plastid markers [14]. The retention of historical signals of genetic structure in highly variable markers, such as microsatellites, is expected for studies involving the entire geographic range of species, reinforcing the importance of considering the historical patterns for interpreting landscape genetic analyses [33]. Moreover, northernmost populations from the 'South' group (S1 and S2) or the southernmost or the 'West' group (W3), given their intermediary location, display higher admixture values supporting secondary gene flow between previously differentiated intraspecific lineages (Fig. 1).

The influence of geographical distance on the genetic structure is evident in the genetic structure of *C. heterophylla*. As expected, the effect of geographical isolation is stronger in peripheral populations such as S6, I1-3, and N1. Therefore, genetic drift due to long-term geographical isolation mainly explains the strong differentiation at the edges. The immigration of populations at the SACP edges falls within the lowest estimates (Fig. S3). However, differential conditions at the edge of the distribution could also be involved. For example, the northern portion of the SACP presents significant differences in precipitation seasonality because of the influence of orographic rainfalls during the spring and summer seasons. This environmental feature significantly correlates with high genetic differentiation in northern populations (Fig. 3). These results point to a genetic divergence process enhanced by local adaptation. Ecological differentiation can promote selection against immigrants (maladaptive gene flow), leading to reduced gene flow, reproductive isolation, and enhancing the stochastic effects of genetic drift [34–36]. As this pattern is also seen in co-distributed coastal populations of *Petunia integrifolia* [13], further research is worthwhile to uncover potential convergence local adaptation processes related to precipitation differences.

Environmental and geomorphological processes around the Patos Lagoon led to a secondary contact between previously diverged lineages

Intricate spatial and environmental influences on the genetic structure is exemplified through the discordant clustering patterns of population W3 with ‘South’ and ‘West’ groups depending on the approach. This feature reflects the intermediary geographic location between those two regions but also the fact that this zone shows fluctuating inland and coast environmental conditions. Both inter-annual rainfall differences and

long-term climatic fluctuations, such as El Niño phenomenon, affect the fluvial discharge and wind currents responsible for the salinization and desalination processes in the Patos Lagoon [37, 38]. This environmental dynamic could periodically change the individuals' establishment or survival rates of either coastal or continental gene pools probably leading to a mixed genetic pool in this region.

The populations W2-3 and S1-5, all located around the Patos Lagoon (Fig. 1A), show high levels of genetic admixture (Fig. 2; Fig. S1) and the lowest F_{ST} values (Fig. 3A). These results are consistent with the recent geomorphological history of the SAPC. During most of the Quaternary Period, two rivers (Jacuí and Camaquã), including several channels corresponding to their dynamic delta systems, maintained distinct inlets on the Patos Lagoon area [14, 39]. Only after the formation of the barrier systems III and IV (the most recent and closer to the shoreline coastal strips) between 12-7 kya, the Patos Lagoon reached its current conformation, and the current continuous coastline was established [40]. In contrast, the northern and southern regions, corresponding to the older barrier systems I and II (cf. Fig. 1 in [27], let to earlier expansion and differentiation of the coastal lineages that, later, spread and experienced a secondary contact on the East side of the Patos Lagoon generating the current patterns of genetic admixture in this region. The recent admixture processes are also supported by the lack of private alleles in W3, S1, S3, and S4 populations (Fig. 1B). This geomorphological history seems to determine common patterns among co-distributed species from the region. Despite the differences in divergence times of the intraspecific lineages of *C. heterophylla* (earlier) and the coastal populations of *P. integrifolia* (recent) [14, 29], these co-distributed taxa share the patterns of high genetic admixture in populations located at the East side of Patos Lagoon [13].

The East side of Patos Lagoon (seashore side) undergoes the strongest wind influence within the SACP (Fig. 4B; [41]), potentially increasing secondary seed dispersal alongside the region generating, in consequence, higher admixture rates. The gene flow estimations among *C. heterophylla* populations support an asymmetric migration from coastal to inland locations, even at long distances crossing the coastal lakes (Figs. 1 and 4). Although wind can significantly influence coastal environments, and it shapes large-scale population differentiation, gene flow, and genetic diversity [42], the influence of wind variables is poorly explored in landscape genetics approaches. Wind conditions also affect population dynamics of Tuco-Tuco rodents (*Ctenomys* sp.) in the SACP [43]. This convergent factor between co-distributed taxa supports that the current dynamics in the topographical and environmental conditions in SACP play a role in the structuration at the community level.

Our findings expand the knowledge of genetic differentiation and diversification processes across coastal areas. According to Wieringa et al. [44], our study highlights multiple processes likely influencing genetic structure. For example, *C. heterophylla* and the coastal lineage of *P. integrifolia* have strong differences in the divergence times and intraspecific differentiation but convergent contemporary distribution and genetic structure. Our results suggest that a complex mixture of features related to physical barriers, geographic distance, and environment along the SACP shape shared contemporary genetic differentiation patterns on the region's species. Therefore, considering historical and recent diversification processes is crucial to interpret either shared or idiosyncratic patterns in contemporary genetic structure. We strongly encourage new research into the environmental factors driving genetic structure within and among populations on plant species distributed along different coastal regions from South America.

CONCLUSIONS

Calibrachoa heterophylla recently colonized the SACP leading to a typical linear distribution shape for coastal species. The species shows limited and asymmetric gene flow patterns, mainly influenced by geographical distance and wind. The presence of big water bodies, which constitutes the most outstanding topographical feature in the SACP, does not constrain inter-population gene flow. Negligible historical and contemporary immigration rates in marginal populations coupled to contrasting precipitation conditions could promote genetic differentiation in the northern and south marginal populations. Recent admixture from previously differentiated populations and higher gene flow explain the genetic diversity in the most recently formed coastal areas and more substantial wind influence region of the SACP. Our results highlight the need to integrate both phylogeographic and landscape genetic approaches to disentangle processes affecting the genetic differentiation of coastal plant species.

METHODS

Study system

The species of *Calibrachoa* (Solanaceae) occur in subtropical and temperate grasslands in southern Brazil, northeast Argentina, and Uruguay. The genus encompasses ca. 30 species, among which *C. heterophylla* is the only species that colonized coastal environments [45]. This species is diploid (2n = 18), semi-prostrated, and displays purplish bee-pollinated flowers; the fruits are capsules and produce dozens of tiny seeds (< 1.4 mm) with no dispersal mechanisms. *C. heterophylla* occupies open sandy grasslands, dunes, or rocky outcrops in lakeside or marine environments from ~ 28 Lat S to 32 Lat S in the SACP [14]. Longitudinally, populations of *C. heterophylla* occur from the seashore to less than 90 km from the coast, with few populations separated from the seashore by big lagoons. Just one disjointed and small population group occurs outside SACP, restricted to the sandbanks alongside the Santa Maria River basin, ~ 55 Long W (Fig. 1A).

Sample collections

For this study, we used all samples included in Mäder et al. [14] plus additional samples from the wild for a total sampling of 253 individuals from 15 locations (hereafter populations; Fig. 1A) that covered the entire species' distribution. We collected leaves of all individuals found in each locality and preserved them in silica gel. The number of individuals per population varied from three to 41 (Table 1). We also sampled complete branches for herbarium specimens from those localities subsequently deposited in BHCB (Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil) and ICN (Universidade Federal do Rio Grande do Sul, RS, Brazil). Plant identification was performed by G. Mäder, J. Fregonezi, or G. Silva-

Arias and then confirmed by the group expert J. R. Stehmann. We required no specific permits since collection localities correspond to neither private properties nor protected areas. Also, the field studies did not involve endangered or protected species. This work was conducted under MP 2.186-16 of the Brazilian Federal Government.

Laboratory procedures and genotyping

The total DNA was extracted following a CTAB-based protocol [46] and amplified for ten anonymous microsatellite loci developed for *C. heterophylla* (Che18, Che59, Che119, Che26, Che34, Che81, Che82, Che85, Che72, and Che126) following optimized protocols for PCR and genotyping procedures [47]. We used MICRO-CHECKER [48] to estimate genotyping errors due to stutter bands, allele dropout, or null alleles.

Characterization of the genetic diversity

We performed tests for linkage disequilibrium and deviations from HWE within each population for each locus. We assessed the significance of HWE deviations using 10^6 Markov chain steps and Fisher's exact probability tests in ARLEQUIN v.3.5 [49]. We estimated the genetic diversity based on average rarefied allelic richness, private alleles, H_o, H_e, the G-W index, and F_{IS} (with confidence limits from 1000 bootstrap resampling over loci) using the POPPR v.2.8.5 [50] and HIERFSTAT v.0.04-22 [51] packages in R v.3.6.3 package [52], and ARLEQUIN.

Population genetic structure

We assessed the genetic structure employing two model-based clustering methods and two exploratory data analyses [53]. The model-based clustering methods used are STRUCTURE v.2.3.4 [54] and the spatial Bayesian clustering program TESS v.2.3 [55, 56]. These analyses provide estimates for the K ancestral clusters assuming HWE equilibrium, individual assignment probabilities and compute the proportion of each individual's genome 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. We performed each run using 2.5×10^5 burn-in periods and 1.0×10^6 Markov chain Monte Carlo repetitions after the burn-in, under an admixture model, assuming correlated allele frequencies [57], 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 [58]. To obtain the K value that better explains the structure based on the genetic dataset, we assessed the measures of the ΔK method [59] that is useful to recover the hierarchical highest level of genetic structure.

TESS implements a spatial assignment approach to group individuals into clusters accounting for samples' geographical locations, giving them higher probabilities of belonging to the same genetic cluster to those that are spatially closer in the connection network. For TESS, we ran100 000 generations, with 50 000 generations as the burn-in, using the CAR admixture model, and starting from a neighbor-joining tree. We ran 20 iterations for each value of maxK ranging from 2 to 15. We added a small perturbation to the original population coordinates with a standard deviation equal to 0.2 to obtain single different coordinates for each individual. We assessed the convergence inspecting the post-run log-likelihood plots

and obtained the support for alternative K values inspecting the statistical measure of the model prediction capability from DIC [60]. We computed and plotted the average of DIC values to detect maxK value at the beginning of a plateau. Replicated runs of best K results for STRUCTURE and TESS were summarized and plotted with the POPHELPER [61] R package.

Additionally, to detect genetic structure, we implemented the exploratory multivariate methods Discriminant analysis of principal components (DAPC [62]) and the spatial Principal Components Analysis (sPCA [63]) implemented in the ADEGENET v.2.1.3 [64] R package. For the DAPC analysis, the SSR data were first transformed using PCA and keeping all PCs. The number of clusters that maximizes the between-group variability using the BIC score was optimized using the function *find.clusters*. To avoid overfitting, we set an optimal reduced number of PCs using the function *optim.a.score*.

The sPCA incorporates spatial information to maximize the product of spatial autocorrelation (Moran's I) and the variance for each eigenvector, producing orthogonal axes that describe spatial patterns of genetic variation. The spatial information is included in the analysis using a spatial weighting matrix derived from a connection network. To test the effect of the neighbors definition on the results, we ran the sPCA using six different connection networks available in the function *chooseCN*. For this analysis, 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 implemented sPCA. Clustering patterns recovered with the DAPC and sPCA were visualized in scatter plots obtained with the function *s.class* in R.

Historical and contemporary gene flow estimations

Contemporary asymmetric migration rates were estimated using the Bayesian approach implemented in BayesAss v.3.0 [65]. We ran 10^8 iterations and a burn-in of 10^7. We adjusted the mixing of allele frequencies, inbreeding coefficients, and migration rates parameters to 0.6, 0.6, and 0.3, respectively, to obtain acceptance rates of around 40%. We assessed convergence by examining the log-probability plots and the effective sample sizes for each run using Tracer v.1.6 [66] and looking for consistency of the migration estimates among three independent runs with different initial seed numbers.

We assessed historical gene flow by testing the support of four alternative scenarios given our genetic dataset using Bayes factors calculated from the Bézier log-marginal likelihood approximations [67]. We used the coalescent-based Migrate-N v.3.2.6 [68] software to estimate the mutation-scaled effective population size (θ) and the mutation scaled migration rate (M) parameters. We pooled the populations into four groups for all models according to the geographical distribution and genetic structure (see Results Figs. 1A and 2). The 'Inland' group included the I1-3 populations; the 'West' group encompassed the W1-3; 'North' included the N1-3 populations, and the 'South' group clustered the S1-6 populations.

We evaluated four migration models: (1) *source-sink from inland* with unidirectional migration from 'Inland' group to the remaining groups; (2) *source-sink from the West* with unidirectional migration from 'West' 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, from South to West, and from West to Inland (Fig. S5).

We ran the MIGRATE-N Bayesian inference in the Cipres Science Gateway v.3.3 [69], with one long chain of 5×10^6 steps, sampling at every 100th increment, and a burn-in of 3×10^4 steps. We used uniform priors and slice sampling for both θ and M ranging from 0 to 20 (mean = 10, delta = 0.5). We used a heating scheme MCMCMC with four parallel chains and temperatures of 1, 1.5, 3, and 10^6.

Space, topography, environment, and genetic differentiation

Spatial correlation patterns under IBD generate bias in several genetic structure tests [15, 70, 71]. Therefore, we assessed the IBD through linear regression of linearized pairwise F_{ST} genetic distances and log-transformed geographical distances [72] using a Mantel test, assessing the significance with 10 000 randomizations in VEGAN v.2.5-6 [73] R package. Pairwise F_{ST} [74] matrix was calculated with the HIERFSTAT package and geographical inter-population distance matrix by calculating the linear Euclidean distance between X and Y UTM 22S (reference EPSG: 32722) populations' coordinates transformed from Long/Lat coordinates with RGDAL v.1.0-4 [75] R package.

We tested IBE models to examine whether differences in climatic conditions explain inter-population genetic differentiation in *C. heterophylla*. Pairwise climatic dissimilarity matrices were obtained for the following bioclimatic 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. Climatic data derive from raster layers specifically developed for the SACP obtained from a high-density sampling of climate stations throughout the region, geostatistical modeling, and spatial interpolation, as described in Silva-Arias et al. [13].

We also included a wind connectivity matrix in the IBE tests to evaluate the influence of strong winds in the SACP on the population migration rate. We calculated surface wind direction and speed data for the Southern Hemisphere's spring months (September to November) 2011–2016 sampled every three hours. We downloaded the data from the Global Forecasting System using the RWIND v.1.1.5 [76] R package. We transformed direction and speed values into raster layers for each sampled time using the *wind2raster* function to obtain transition layers using the function *flow.dispersion*. Finally, we calculated pairwise cost distance matrices with the function *costDistance* in GDISTANCE v.1.3-1 [77] R package. We then averaged the matrices for the all-time series. We plotted the final matrix with the QGRAPH v.1.6.5 [78] R package.

We extended the IBD analyses using raster grids to test for possible models of inter-population differentiation linked to landscape discontinuities alongside the SACP. We outlined two coast distance models (Fig. S6): (1) the *continuous* (or null) model wherein no landscape discontinuity affects the interpopulation connectivity. 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 it is more appropriate for comparisons with models based on circuit theory; and (2) the *water bodies* model, representing the widespread freshwater bodies in the SACP as connectivity barriers 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 (*no data*). We generated pairwise cost distance matrices using the function *transition* in GDISTANCE package considering an eight-neighbors cell connection scheme, Long/Lat coordinates per population as nodes, and raster resolution of 0.09 degrees (~ 10 km).

We examined the relationships between F_{ST} and geographical or topographical distances (IBD) and environmental dissimilarity (IBE) using MMRR; [79] implemented in R.

Supplementary Information

Fig. S1: Comparison observed *vs*. expected heterozygosity per microsatellite locus.

Fig. S2: Evaluation of the best number of clusters for STRUCTURE and TESS analyses.

Fig. S3: Bar plots of the individual membership obtained with STRUCTURE.

Fig. S4**:** Bayesian information criterion assessed with Discriminant Analyses of Principal Components.

Fig. S5: Graphical representation of the coalescent migration models.

Fig. S6: Graphical representation of the raster layers for topographic tests.

Table S1: Migration estimates obtained with three independent runs of BAYESASS.

Additional file 1. MIGRATE-N detailed output for the best-supported model showing parameter estimation values and convergence statistics.

Abbreviations

BIC - Bayesian Information Criterion; CAR - conditional autoregressive; Che - *Calibrachoa heterophylla;* CI - confidence interval; cpDNA - plastid DNA; CTAB - cetyl-tetramethyl ammonium bromide; DAPC - Discriminant Analysis of Principal Components; DIC - deviance information criterion; F_{IS} - inbreeding coefficient; F_{ST} - genetic differentiation index; G-W - Garza-Williamson index; H_e - expected heterozygosity; H_o - observed heterozygosity; HWE Hardy-Weinberg equilibrium; IBD - isolation-by-distance; IBE - isolation-by-environment; K - number of genetic

clusters; km - kilometre; ky - thousand years; kya - thousand years ago; Lat - latitude; Long - longitude; M - mutation-scaled migration parameter; MCMCMC - Metropolis-coupled Markov Chain Monte Carlo; MMRR - multiple matrix regressions with randomization; NE- northeast; N_m – Effective migration rate; PCs - principal components; PCR - polymerase chain reaction; R^2 - coefficient of determination; r - correlation coefficient; SACP - South Atlantic coastal Plain; sPCA - spatial Principal Component Analysis; sPC - spatial Principal Components; S - south; SW - southwest; SSR - microsatellite locus; subsp. - subspecies; UTM - Universal Transverse Mercator coordinate system; W - west; θ - mutation-scaled population size.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

All authors reviewed the manuscript and agreed with this publication.

Availability of data and materials

All data generated or analyzed during this study are included in this published article [and its supplementary information files].

Competing interests

The authors declare that they have no competing interests

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. Silva-Arias was supported by a fellowship from the Departamento Administrativo de Ciencia y Tecnología e Innovación (COLCIENCIAS) and the TUM University Foundation Fellowship (TUFF).

The funding bodies played no role in the design of the study and collection, analysis, and interpretation of the data, and the writing of the manuscript.

Authors' contributions

G.A.S.-A. and L.B.F. designed the study; G.A.S.-A. performed laboratory experiments, analyzed data, and interpreted the results; G.A.S.-A. and L.C.-V. discussed the results and wrote the manuscript; G.C.G. supported laboratory genotyping; L.B.F. supervised the project and provided resources for the research.

Acknowledgements

The authors acknowledge G. Mäder and J. Fregonezi for help in fieldwork, J.R. Stehmann for plant identification, and the Technical University of Munich Publishing Fund for covering the costs for the publication of this article.

REFERENCES

1. Hulton NRJ, Purves RS, McCulloch RD, Sugden DE, Bentley MJ. The Last Glacial Maximum and deglaciation in southern South America. Quaternary Science Reviews. 2002;21:233–41.

2. Scarano FR. Structure, Function and Floristic Relationships of Plant Communities in Stressful Habitats Marginal to the Brazilian Atlantic Rainforest. Annals of Botany. 2002;90:517–24.

3. Behling H. Late glacial and Holocene vegetation, climate and fire history inferred from Lagoa Nova in the southeastern Brazilian lowland. Vegetation History and Archaeobotany. 2003;12:263–70.

4. Carnaval AC, Moritz C. Historical climate modelling predicts patterns of current biodiversity in the Brazilian Atlantic forest. Journal of Biogeography. 2008;35:1187–201.

5. Saillard M, Hall SR, Audin L, Farber DL, Hérail G, Martinod J, et al. Non-steady long-term uplift rates and Pleistocene marine terrace development along the Andean margin of Chile (31°S) inferred from 10Be dating. Earth and Planetary Science Letters. 2009;277:50–63.

6. Miloslavich P, Klein E, Díaz JM, Hernández CE, Bigatti G, Campos L, et al. Marine Biodiversity in the Atlantic and Pacific Coasts of South America: Knowledge and Gaps. PLoS ONE. 2011;6:e14631.

7. Silva GAR, Antonelli A, Lendel A, Moraes E de M, Manfrin MH. The impact of early Quaternary climate change on the diversification and population dynamics of a South American cactus species. J Biogeogr. 2018;45:76–88.

8. Massante JC, Gerhold P. Environment and evolutionary history depict phylogenetic alpha and beta diversity in the Atlantic coastal white-sand woodlands. J Veg Sci. 2020;31:634–45.

9. Sérsic AN, Cosacov A, Cocucci AA, Johnson LA, Pozner R, Avila LJ, et al. Emerging phylogeographical patterns of plants and terrestrial vertebrates from Patagonia. Biological Journal of the Linnean Society. 2011;103:475–94.

10. Turchetto-Zolet AC, Pinheiro F, Salgueiro F, Palma-Silva C. Phylogeographical patterns shed light on evolutionary process in South America. Molecular Ecology. 2013;22:1193–213.

11. Leal BSS, Palma da Silva C, Pinheiro F. Phylogeographic Studies Depict the Role of Space and Time Scales of Plant Speciation in a Highly Diverse Neotropical Region. Critical Reviews in Plant Sciences. 2016;35:215–30.

12. Pinheiro F (biopinheiro@yahoo. com. br), de Barros F, Palma-Silva C, Fay MF, Lexer C, Cozzolino S. Phylogeography and genetic differentiation along the distributional range of the orchid Epidendrum fulgens: a Neotropical coastal species not restricted to glacial refugia. Journal of Biogeography. 2011;38:1923–35.

13. Silva-Arias GA, Reck-Kortmann M, Carstens BC, Hasenack H, Bonatto SL, Freitas LB. From inland to the coast: Spatial and environmental signatures on the genetic diversity in the colonization of the South Atlantic Coastal Plain. Perspectives in Plant Ecology, Evolution and Systematics. 2017;28:47–57.

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

15. Perez MF, Franco FF, Bombonato JR, Bonatelli IAS, Khan G, Romeiro-Brito M, et al. Assessing population structure in the face of isolation by distance: Are we neglecting the problem? Diversity and Distributions. 2018;24:1883–9.

16. Kottler EJ, Dickman EE, Sexton JP, Emery NC, Franks SJ. Draining the Swamping Hypothesis: Little Evidence that Gene Flow Reduces Fitness at Range Edges. Trends in Ecology & Evolution. 2021;36:533–44.

17. Loveless MD, Hamrick JL. Ecological determinants of genetic structure in plant populations. Annu Rev Ecol Syst. 1984;15:65–95.

18. Huang W, Zhao X, Zhao X, Li Y, Lian J. Effects of environmental factors on genetic diversity of *Caragana microphylla* in Horqin Sandy Land, northeast China. Ecol Evol. 2016;6:8256–66.

19. Schierenbeck KA. Population-level genetic variation and climate change in a biodiversity hotspot. Ann Bot. 2017;119:215–28.

20. Thompson JD. Population differentiation in Mediterranean plants: insights into colonization history and the evolution and conservation of endemic species. Heredity. 1999;82:229–36.

21. Excoffier L, Foll M, Petit RJ. Genetic Consequences of Range Expansions. Annu Rev Ecol Evol Syst. 2009;40:481–501.

22. Kadereit J, Westberg E. Determinants of phylogeographic structure: a comparative study of seven coastal flowering plant species across their European range. Watsonia. 2007;26:229–38.

23. Escudero M, Vargas P, Arens P, Ouborg NJ, Luceño M. The east-west-north colonization history of the Mediterranean and Europe by the coastal plant Carex extensa (Cyperaceae): COLONIZATION HISTORY OF CAREX EXTENSA. Molecular Ecology. 2010;19:352–70.

24. Sork VL. Gene flow and natural selection shape spatial patterns of genes in tree populations: implications for evolutionary processes and applications. Evol Appl. 2016;9:291–310.

25. Tomazelli LJ, Dillenburg SR, Villwock JA. Late Quaternary geological history of Rio Grande do Sul coastal plain, southern Brazil. Revista Brasileira de Geociências. 2000;30:474–6.

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

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

28. Dillenburg SR, Barboza EG, Tomazelli LJ, Ayup-Zouain RN, Hesp PA, Clerot LCP. The Holocene Coastal Barriers of Rio Grande do Sul. In: Dillenburg SR, Hesp PA, editors. Geology and Geomorphology of Holocene Coastal Barriers of Brazil. Berlin, Heidelberg: Springer; 2009. p. 53–91. doi:10.1007/978-3-540-44771-9_3.

29. Ramos-Fregonezi AM, Fregonezi JN, Cybis GB, Fagundes NJ, Bonatto SL, Freitas LB. Were sea level changes during the Pleistocene in the South Atlantic Coastal Plain a driver of speciation in *Petunia* (Solanaceae)? BMC Evolutionary Biology. 2015;15:92.

30. Baranzelli MC, Johnson LA, Cosacov A, Sérsic AN. Historical and ecological divergence among populations of *Monttea chilensis* (Plantaginaceae), an endemic endangered shrub bordering the Atacama Desert, Chile. Evol Ecol. 2014;28:751–74.

31. Meireles JE, Manos PS. Pervasive migration across rainforest and sandy coastal plain Aechmea nudicaulis (Bromeliaceae) populations despite contrasting environmental conditions. Molecular Ecology. 2018;27:1261–72.

32. Arjona Y, Fernández-López J, Navascués M, Alvarez N, Nogales M, Vargas P. Linking seascape with landscape genetics: Oceanic currents favour colonization across the Galápagos Islands by a coastal plant. Journal of Biogeography. 2020;47:2622–33.

33. Anderson CD, Epperson BK, Fortin M-J, Holderegger R, James PMA, Rosenberg MS, et al. Considering spatial and temporal scale in landscape-genetic studies of gene flow. Molecular Ecology. 2010;19:3565–75.

34. Nosil P, Funk DJ, Ortiz-Barrientos D. Divergent selection and heterogeneous genomic divergence. Molecular Ecology. 2009;18:375–402.

35. Hendry AP. Selection against migrants contributes to the rapid evolution of ecologically dependent reproductive isolation. Evolutionary Ecology Research. 2004;6:1219–36.

36. Nosil P, Vines TH, Funk DJ. Reproductive isolation caused by natural selection against immigrants from divergent habitats. Evolution. 2005;59:705–19.

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

38. Möller OO, Lorenzzentti JA, Stech JoséL, Math MM. The Patos Lagoon summertime circulation and dynamics. Continental Shelf Research. 1996;16:335–51.

39. Weschenfelder J, Baitelli R, Corrêa ICS, Bortolin EC, dos Santos CB. Quaternary incised valleys in southern Brazil coastal zone. Journal of South American Earth Sciences. 2014;55:83–93.

40. Santos-Fischer CB dos, Corrêa ICS, Weschenfelder J, Torgan LC, Stone JR. Paleoenvironmental insights into the Quaternary evolution of the southern Brazilian coast based on fossil and modern diatom assemblages. Palaeogeography, Palaeoclimatology, Palaeoecology. 2016;446:108–24.

41. Martinho CT, Hesp PA, Dillenburg SR. Morphological and temporal variations of transgressive dunefields of the northern and mid-littoral Rio Grande do Sul coast, Southern Brazil. Geomorphology. 2010;117:14–32.

42. Kling MM, Ackerly DD. Global wind patterns shape genetic differentiation, asymmetric gene flow, and genetic diversity in trees. PNAS. 2021;118. doi:10.1073/pnas.2017317118.

43. Garcias FM, Stolz JFB, Fernández GP, Kubiak BB, Bastazini VAG, Freitas TRO. Environmental predictors of demography in the tuco-tuco of the dunes (*Ctenomys flamarioni*). Mastozool Neotrop. 2018;25:293–304.

44. Wieringa JG, Boot MR, Dantas-Queiroz MV, Duckett D, Fonseca EM, Glon H, et al. Does habitat stability structure intraspecific genetic diversity? It's complicated... Frontiers of Biogeography. 2020;12. doi:10.21425/F5FBG45377.

45. Mäder G, Freitas LB. Biogeographical, ecological, and phylogenetic analyses clarifying the evolutionary history of *Calibrachoa* in South American grasslands. Molecular Phylogenetics and Evolution. 2019;141:106614.

46. Roy A, Frascaria N, MacKay J, Bousquet J. Segregating random amplified polymorphic DNAs (RAPDs) in *Betula alleghaniensis*. TAG Theoretical and Applied Genetics. 1992;85:173–80.

47. Silva-Arias GA, Mäder G, Bonatto SL, Freitas LB. Novel Microsatellites for *Calibrachoa heterophylla* (Solanaceae) Endemic to the South Atlantic Coastal Plain of South America. Applications in Plant Sciences. 2015;3:1500021.

48. Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P. MICRO-CHECKER: Software for identifying and correcting genotyping errors in microsatellite data. Molecular Ecology Notes. 2004;4:535–8.

49. Excoffier L, Lischer HEL. ARLEQUIN suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology Resources. 2010;10:564–7.

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

51. Goudet J, Jombart T. HIERFSTAT: Estimation and tests of hierarchical F-statistics. R package version 0.5-7. http://CRAN.R-project.org/package=hierfstat. 2020. http://CRAN.R-project.org/package=hierfstat. Accessed 6 Jul 2015.

52. R Core Team. R: A language and environment for statistical computing. Vienna, Austria. http://www.R-project.org/: R Foundation for Statistical Computing; 2020. http://www.R-project.org/.

53. François O, Waits LP. Clustering and assignment methods in landscape genetics. In: Balkenhol N, Cushman SA, Storfer AT, Waits LP, editors. Landscape genetics: concepts, methods, applications. Chichester, UK: John Wiley & Sons, Ltd; 2016. p. 114–28. http://doi.wiley.com/10.1002/9781118525258.ch07. Accessed 15 May 2016.

54. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000;155:945–59.

55. Durand E, Jay F, Gaggiotti OE, François O. Spatial Inference of Admixture Proportions and Secondary Contact Zones. Molecular Biology and Evolution. 2009;26:1963–73.

56. Chen C, Durand E, Forbes F, François O. Bayesian clustering algorithms ascertaining spatial population structure: a new computer program and a comparison study. Molecular Ecology Notes. 2007;7:747–56.

57. Falush D, Stephens M, Pritchard JK. Inference of population structure using multilocus genotype data: dominant markers and null alleles. Molecular Ecology Notes. 2007;7:574–8.

58. Hubisz MJ, Falush D, Stephens M, Pritchard JK. Inferring weak population structure with the assistance of sample group information. Molecular Ecology Resources. 2009;9:1322–32.

59. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. Molecular Ecology. 2005;14:2611–20.

60. Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2002;64:583–639.

61. Francis RM. POPHELPER: An R package and web app to analyse and visualise population structure. Molecular Ecology Resources. 2016;:n/a-n/a.

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

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

64. Jombart T. ADEGENET: A R package for the multivariate analysis of genetic markers. Bioinformatics. 2008;24:1403–5.

65. Wilson GA, Rannala B. Bayesian inference of recent migration rates using multilocus genotypes. Genetics. 2003;163:1177–91.

66. Rambaut A, Suchard MA, Xie D, Drummond AJ. TRACER v1.6, Available from http://beast.bio.ed.ac.uk/Tracer. 2014. http://beast.bio.ed.ac.uk/Tracer.

67. Beerli P, Palczewski M. Unified framework to evaluate panmixia and migration direction among multiple sampling locations. Genetics. 2010;185:313–26.

68. Beerli P, Felsenstein J. 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. 2001;98:4563–8.

69. Miller MA, Pfeiffer W, Schwartz T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: Proceedings of the Gateway Computing Environments Workshop (GCE). New Orleans; 2010. p. 1–8. http://www.phylo.org/sub_sections/portal/sc2010_paper.pdf.

70. Frantz AC, Cellina S, Krier A, Schley L, Burke T. Using spatial Bayesian methods to determine the genetic structure of a continuously distributed population: clusters or isolation by distance? Journal of Applied Ecology. 2009;46:493–505.

71. Meirmans PG. The trouble with isolation by distance. Molecular Ecology. 2012;21:2839–46.

72. Rousset F. Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. Genetics. 1997;145:1219–28.

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

74. Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. Evolution. 1984;38:1358–70.

75. Bivand R, Keitt T, Rowlingson B. RGDAL: Bindings for the geospatial data abstraction library. R package version 0.9-3. http://CRAN.R-project.org/package=rgdal. 2015. http://CRAN.R-project.org/package=rgdal.

76. Fernández-López J, Schliep K. RWIND: download, edit and include wind data in ecological and evolutionary analysis. Ecography. 2019;42:804–10.

77. Etten J van. R Package gdistance: Distances and Routes on Geographical Grids. Journal of Statistical Software. 2017;76:1–21.

78. Epskamp S, Cramer AOJ, Waldorp LJ, Schmittmann VD, Borsboom D. QGRAPH: Network visualizations of relationships in psychometric data. Journal of Statistical Software. 2012;48:1–18.

79. Wang IJ. 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. 2013;67:3403–11.

FIGURE LEGENDS

Figure 1. (A) Locations of *Calibrachoa heterophylla* populations. (B) Graphical representation of the mean genetic diversity statistics estimated for each population across all microsatellite loci. (C) Scatterplot of the DAPC analysis. (D) Scatterplot of the sPCA. Populations' numbers and colors follow each panel.

Figure 2. Bar plots of the individual membership for K = 2-8 genetic clusters as estimated with TESS. Populations are separated by white dashed lines and named on the top side.

Figure 3. Interpopulation genetic, geographical, and environmental dissimilarity matrices. (A) Genetic differentiation index F_{ST}; (B) Linear geographical distance; (C) Precipitation seasonality dissimilarity. A continuous color scale represents the values.

Figure 4. Spring season wind fields in the South Atlantic Coastal Plain and inference of wind in the population connectivity and gene flow. (A) Correlation between geographical distance (red) and wind connectivity (blue) cost distance matrices with the F_{ST} genetic distance matrix; (B) Mean values 2011-2016 of wind speed (blue scale m/s) and direction (arrows) for the spring season (September-November); (C) Inter-population wind connectivity network.

SUPPLEMENTARY MATERIAL LEGENDS

Figure S1. The plot of observed *vs*. expected heterozygosity for each locus.

Figure S2. (A) Plots of the best K estimates for STRUCTURE results. (B) Plots of the mean and standard deviation of the deviance information criterion (DIC) obtained for each maxK assessed with TESS.

Figure S3. Bar plots of the individual membership for each genetic cluster obtained with STRUCTURE. White dashed lines separate populations, and names are indicated on the figure top side.

Figure S4. The plot of Bayesian information criterion (BIC) values obtained for each K number assessed using the multivariate method Discriminant Analyses of Principal Components.

Figure S5. Graphical representation of the four coalescent migration models tested in MIGRATE-N for *Calibrachoa heterophylla*. (A) Source-sink from inland; (B) Source-sink from the west; (C) Step-stone from inland; (D) Step-stone from coast.

Figure S6. Graphical representation of the raster layers used to calculate the connectivity values in topographic tests (A) Continuous model; (B) Water bodies model.

Table S1. Migration estimates obtained with three independent runs of BAYESASS. The values indicate the estimated posterior mean effective migration rate per generation [the fraction of individuals in population *i* (rows) that are migrants derived from population *j (*columns)], and the numbers in parentheses show the standard deviation. Bold values indicate the diagonal (intra-population estimates), and red

values indicate the highest migration estimates (those with above zero 95% confidence intervals).

Additional file 1. MIGRATE-N detailed output for the best-supported model showing parameter estimation values and convergence statistics.

TABLES

Table 1. Sampling information and inbreeding coefficients for *Calibrachoa heterophylla* populations. Population ID codes follow Fig. 1A.

Pop ID	N	Locality	Long	Lat	F_{IS}
I1	4	São Francisco de Assis	-55.10077	-29.58307	-0.16
I2	10	Cacequi 1	-54.85375	-29.8947	-0.01
I3	9	Cacequi 2	-54.90852	-29.85478	0
W1	27	Barra do Ribeiro	-51.20255	-30.40754	0.08
W2	4	Arambaré	-51.49195	-30.90082	0.12
W3	23	Pelotas	-52.16478	-31.70757	0.07
N1	10	Laguna	-48.76501	-28.45991	0.34**
N2	23	Torres	-49.79809	-29.43227	0.17**
N3	37	S Antônio da Patrulha	-50.42936	-29.89291	0.10*
S1	3	Mostardas 1	-50.73934	-30.93746	0.35
S2	13	Mostardas 2	-50.90112	-31.10909	0.08
S3	12	S José do Norte 1	-51.42576	-31.66673	-0.01
S4	11	S José do Norte 2	-52.03612	-32.02393	-0.05
S5	26	Rio Grande	-52.54661	-32.52396	0.27***
S6	41	S Vitória do Palmar	-52.7323	-32.98765	0.25***

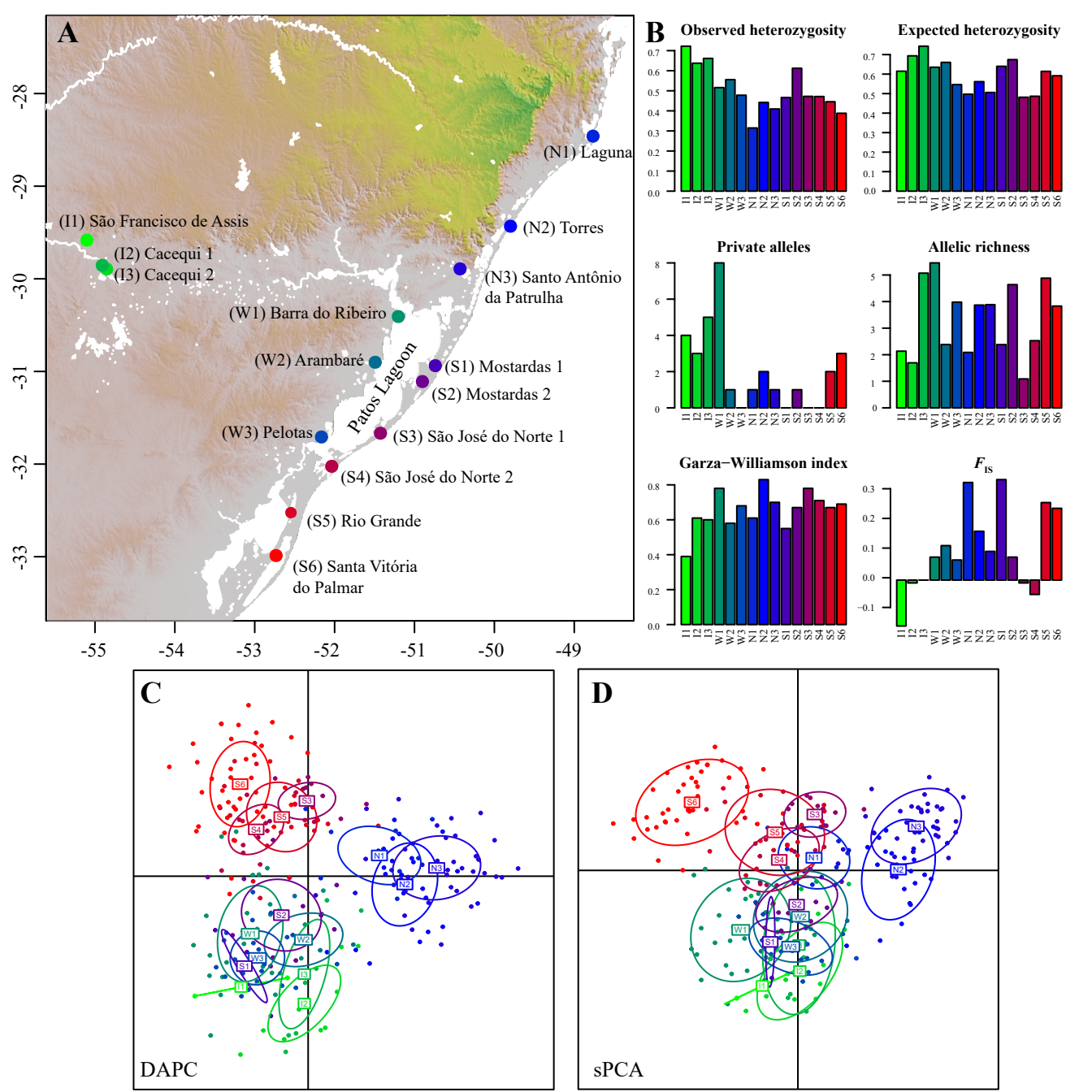
*P-value < 0.05; ** P-value < 0.01; *** P-value < 0.001

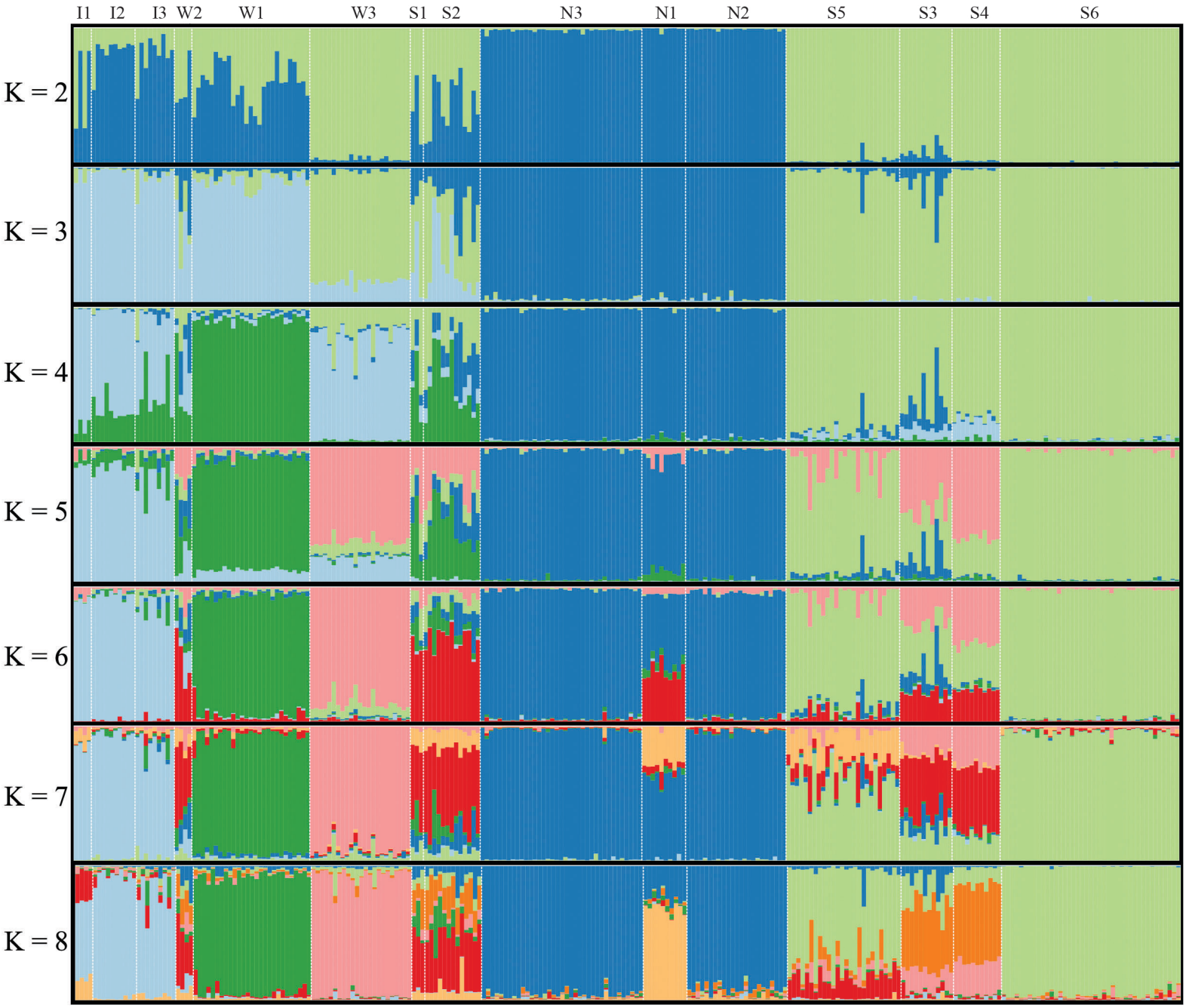
Table 2. Model support statistics (upper panel) and parameter estimations taken from the best-supported model (lower panel).

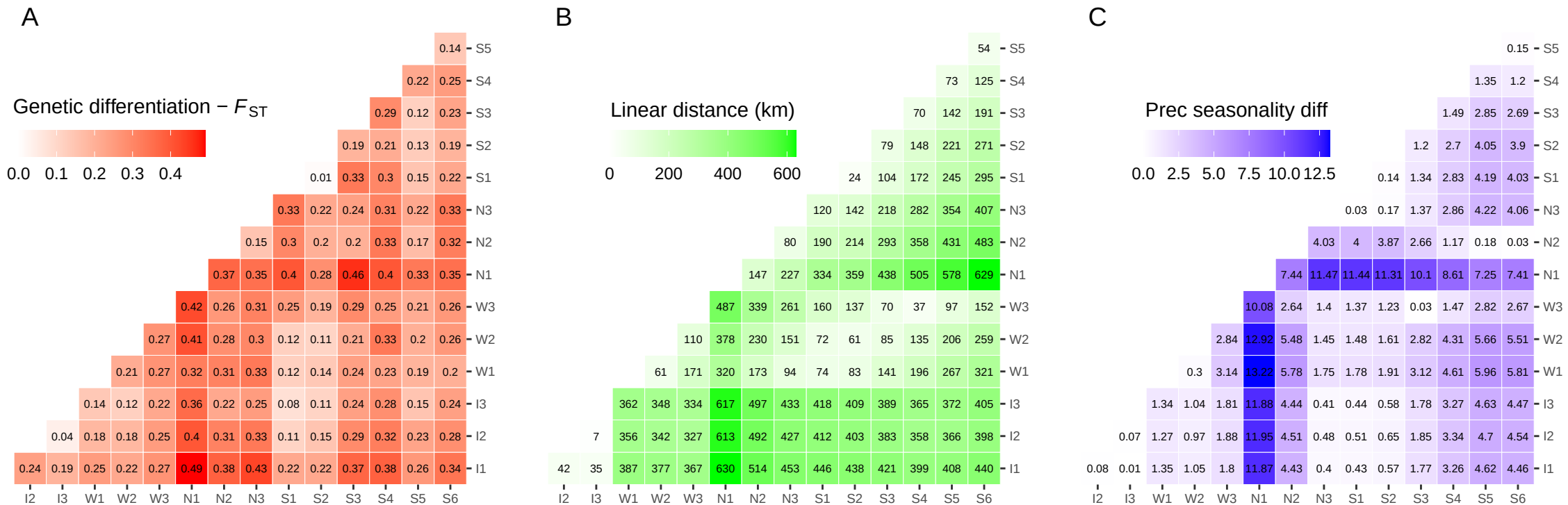
Model name	Ln marginal Likelihood	Log Bayes Factor	Model probability
Step stone from coast	-5900.7	0	1.0
Source-sink from west	-5940.5	-79.6	5.1E-18
Source-sink from inland	-5941.5	-81.6	1.9E-18
Step stone from inland	-5944.9	-88.4	6.5E-20

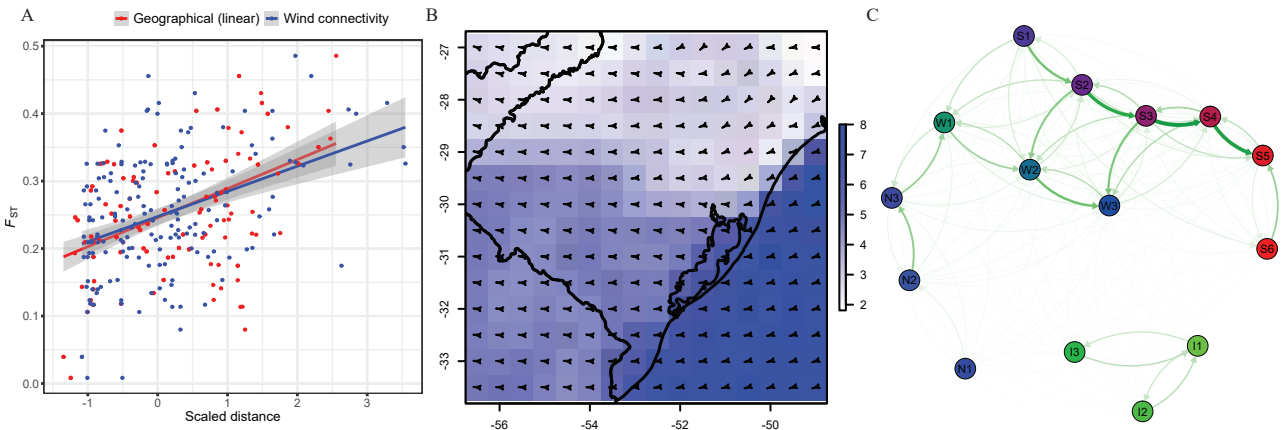
Parameter	Median	Mean	95 % CI
θ 'Inland'	4.11	4.25	2.55 - 6.28
θ 'West'	0.55	0.56	0.12 - 0.96
θ 'North'	0.51	0.52	0.05 - 0.95
θ 'South'	0.27	0.22	0 - 0.55
M 'West' -> 'Inland'	0.41	0.45	0 - 1.16
M 'North' -> 'West'	0.3	0.26	0 - 0.6
M 'South' -> 'West'	0.19	0.08	0 - 0.41

Bayes factor values < 2 indicate a strong preference for the model with the highest probability. Θ = mutation scaled population size; M = mutation scaled migration rate; CI = confidence interval. For graphical model descriptions, see Fig. S5.


A
(N1) Laguna
(N2) Torres
(N3) Santo Antônio da Patrulha
(I1) São Francisco de Assis
(I2) Cacequi 1
(I3) Cacequi 2
(W1) Barra do Ribeiro
(W2) Arambaré
(W3) Pelotas
Patos Lagoon
(S1) Mostardas 1
(S2) Mostardas 2
(S3) São José do Norte 1
(S4) São José do Norte 2
(S5) Rio Grande
(S6) Santa Vitória do Palmar
-28
-29
-30
-31
-32
-33
-55
-54
-53
-52
-51
-50
-49
B
Observed heterozygosity
Expected heterozygosity
Private alleles
Allelic richness
Garza−Williamson index
F_{IS}
I1 I2 I3 W1 W2 W3 N1 N2 N3 S1 S2 S3 S4 S5 S6
C
DAPC
D
sPCA


I1
I2
I3
W2
W1
W3
S1
S2
N3
N1
N2
S5
S3
S4
S6
K = 2
K = 3
K = 4
K = 5
K = 6
K = 7
K = 8


A
Genetic differentiation − F_{ST}
0.0 0.1 0.2 0.3 0.4
B
Linear distance (km)
0 200 400 600
C
Prec seasonality diff
0.0 2.5 5.0 7.5 10.0 12.5
I2 I3 W1 W2 W3 N1 N2 N3 S1 S2 S3 S4 S5 S6
S5 S4 S3 S2 S1 N3 N2 N1 W3 W2 W1 I3 I2 I1


A
Geographical (linear)
Wind connectivity
F_{ST}
Scaled distance
B
C
S1
S2
S3
S4
S5
S6
W1
W2
W3
N1
N2
N3
I1
I2
I3