the **genetics**society

Check for updates

ARTICLE

A historical stepping-stone path for an island-colonizing cactus across a submerged "bridge" archipelago

Fernando Faria Franco [1,7 ✉], Danilo Trabuco Amaral[2,3,7], Isabel A. S. Bonatelli[4], Jared B. Meek [5], Evandro Marsola Moraes[1], Daniela Cristina Zappi [6], Nigel Paul Taylor[1] and Deren A. R. Eaton[5]

Here we use population genomic data (ddRAD-Seq) and ecological niche modeling to test biogeographic hypotheses for the divergence of the island-endemic cactus species *Cereus insularis* Hemsl. (Cereeae; Cactaceae) from its sister species *C. fernambucensis* Lem. The *Cereus insularis* grows in the Fernando de Noronha Islands (FNI), a Neotropical archipelago located 350 km off the Brazilian Atlantic Forest (BAF) coast. Phylogeographic reconstructions support a northward expansion by the common ancestor of *C. insularis* and *C. fernambucensis* along the mainland BAF coast, with *C. insularis* diverging from the widespread mainland taxon *C. fernambucensis* after colonizing FNI in the late Pleistocene. The morphologically distinct *C. insularis* is monophyletic and nested within *C. fernambucensis*, as expected from a progenitor-derivative speciation model. We tested alternative biogeographic and demographic hypotheses for the colonization of the FNI using Approximate Bayesian Computation. We found the greatest support for a stepping-stone path that emerged during periods of decreased sea level (the "bridge" hypothesis), in congruence with historical ecological niche modeling that shows highly suitable habitats on stepping-stone islands during glacial periods. The outlier analyses reveal signatures of selection in *C. insularis*, suggesting a putative role of adaptation driving rapid anagenic differentiation of this species in FNI.

INTRODUCTION

Islands are natural laboratories for the study of evolution, as they often contain high levels of biological diversity and endemism and can be modeled as the outcome of a few processes, such as dispersal, speciation, and extinction (MacArthur and Wilson 1967; Losos and Ricklefs 2009; Whittaker et al. 2017). However, many questions about the origins of island diversity remain to be answered, such as the processes of speciation between island and mainland populations and the role of climate change in shaping population connectivity (Fernández-Palacios et al. 2016; Patiño et al. 2017; Flantua et al. 2020). Recent advances in statistical phylogeographic methods, as well as ecological niche modeling, have now the potential to address such questions (Rijsdijk et al. 2014; Papadopoulou and Knowles 2015; Weigelt et al. 2016; García-Verdugo et al. 2017).

The emergence of new species resulting from in situ speciation after the colonization from a larger mainland ancestral population offers a plausible explanation for the rich diversity and unique species found on islands (Whittaker et al. 2017). This mode of speciation, termed peripatric speciation (Mayr 1954) or peripheral isolation speciation (Frey 1993), is a form of allopatric speciation in which the new species is derived from an isolated and small peripheral population. Quantum speciation (Grant 1971) describes

a similar process (Futuyma 1989; Grant 1989) but emphasizes the expectation that the rapid divergence of isolated peripheral populations is enhanced by genetic drift and inbreeding in the founder populations (Gottlieb 2004). In either model, a peripheral population "buds off" from an ancestral widespread species, resulting in a "progenitor-derivative" pattern (Mayr 1954; Gottlieb 2004; Grossenbacher et al. 2014) associated with anagenetic differentiation of the derivative species while progenitors tend to retain the ancestral type. In the context of island biogeography, this process was named "anacladogenesis" (Stuessy et al. 1990) and "anagenic speciation" (Stuessy et al. 2006; Takayama et al. 2015), although the use of the terms cladogenesis and anagenesis in this context may have deviated from their original definitions (Emerson and Patiño 2018). Here, we adopt the more general term "progenitor-derivative speciation model" (sensu Crawford 2010) to describe such a pattern, as this embraces, at least in geographic and phylogenetic terms, the assumptions of both peripatric, quantum, and anagenic speciation models.

Although phylogenetic patterns consistent with a progenitor-derivative speciation model have been frequently observed, the role of adaptation in driving phenotypic divergence and isolation between mainland and island species remains poorly understood. Several hypotheses have been proposed to describe potential

[1]Departamento de Biologia. Centro de Ciências Humanas e Biológicas, Universidade Federal de São Carlos (UFSCar), Sorocaba, Brazil. [2]Centro de Ciências Naturais e Humanas, Universidade Federal do ABC (UFABC), Santo André, São Paulo, Brazil. [3]Programa de Pós Graduação em Biologia Comparada. Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo (USP), Ribeirão Preto, Brazil. [4]Instituto de Ciências Ambientais, Químicas e Farmacêuticas. Departamento de Ecologia e Biologia Evolutiva, Universidade Federal de São Paulo, Diadema, SP, Brazil. [5]Department of Ecology, Evolution and Environmental Biology, Columbia University, New York, NY 10027, USA. [6]Programa de Pós Graduação em Botânica, Instituto de Ciências Biológicas, Universidade de Brasília, PO Box 04457, Brasília, DF 70910970, Brazil. [7]These authors contributed equally: Fernando Faria Franco, Danilo Trabuco Amaral. Associate editor: Dario Grattaplaglia. ✉email: franco@ufscar.br
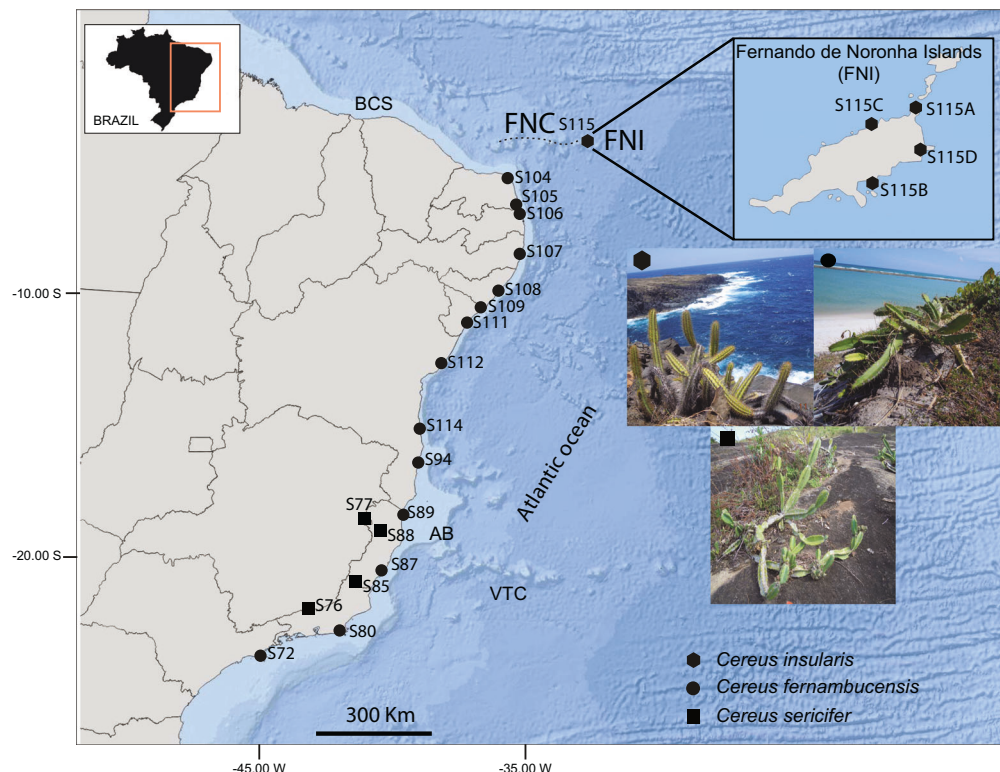
Fig. 1 Map of eastern Brazil highlighting the study area and the locations of *Cereus* sampled during this study (details in Table S1). FNI, Fernando de Noronha Islands; FNC, Fernando de Noronha Chain; BCS, Brazilian Continental Shelf; AB, Abrolhos Bank; VTC, Vitória-Trindade Chain. The dotted line highlights the seamounts of FNC. Photos: Franco, FF.

effects of genetic drift, selection, or both during the slow or rapid differentiation of founder populations (e.g., genetic revolution, founder-flush, genetic transilience; Mayr 1954; Carson 1968; Grant 1971; Templeton 1981); but several of these have been criticized for lacking a clear test framework (e.g., Barton and Charlesworth 1984; Patiño et al. 2017). The role of adaptation during speciation has become widely accepted (Nosil 2008; Anacker and Strauss 2014; Sendell-Price et al. 2021). However, non-ecological specia-tion can occur through mutation-order fixation under genetic drift or purifying selection, especially in small populations (Schluter 2009; Maya-Lastra and Eaton 2021). Thus, examining the popula-tion demographic changes associated with island colonization can reveal insights into the process of speciation. The progenitor-derivative speciation model associated with island biogeography may be associated with several geographic, phylogenetic, and adaptive predictions, which can be tested within a phylogeo-graphic framework. These predictions include: (1) the derivative species may rapidly become monophyletic while the progenitor species may appear paraphyletic for an extended period (Mayr 1954; Rieseberg and Brouillet 1994; Gottlieb 2004; Anacker and Strauss 2014; Grossenbacher et al. 2014; Bock et al. 2023); (2) the derivative species may exhibit a smaller (Frankham et al. 1999; Losos and Glor 2003) or similar (García-Verdugo et al. 2015) effective population size than the progenitor species, but this will depend on the initial process of colonization (single or multiple), habitat availability, dispersal capacity (cost of dispersal) and reproductive systems (Stuessy et al. 2014; Alsos et al. 2015; García-Verdugo et al. 2017; García-Verdugo et al. 2019; Kimura et al. 2022); and (3) rapid adaptation by the derivative species to a newly colonized environment may leave genomic signatures of natural selection (Mayr 1954; Carson 1968; Templeton 2008; Sendell-Price et al. 2021).

Cereus insularis Hemsl. (Cereeae; Cactaceae) is an insular cactus species endemic to Fernando de Noronha islands (FNI), a small archipelago in the southwestern Atlantic Ocean 350 km from mainland Brazil (Fig. 1). Previous studies have shown *C. insularis* to be peripherally associated with a paraphyletic sister species, *C. fernambucensis* Lem., consistent with a progenitor-derivative phylogenetic pattern typical of a peripatric origin (Franco et al. 2017; Bombonato et al. 2020). Both species are pollinated by hawkmoths and have zoochorous fruits that may be dispersed by frugivorous vertebrates (Taylor and Zappi 2004). Birds are considered to be the main agents of seed dispersal in *Cereus* (Gomes et al. 2014). Franco et al. (2017) argue that the colonization of FNI by the *C. insularis* ancestral population occurred via a putative seamount stepping-stone "bridge" (henceforward called the "bridge" hypothesis) during Pleistocene glacial periods when the sea level decreased and likely exposed some seamounts of the Fernando de Noronha Chain (FNC; Fig. 1; Vital 2014). In agreement with this hypothesis, similar levels of microsatellite genetic diversity have been observed between *C. insularis* and mainland *C. fernambucensis*, supporting a long-term gene flow after the colonization of FNI; however, this hypothesis has yet to be validated using more comprehensive genome-scale datasets. Moreover, suitable ecological conditions for *C. insularis* during the Last Glacial Maximum (LGM) have been detected among some emerging seamounts of FNC, further corroborating the "bridge" hypothesis (Franco et al. 2022). However, this work did not include a broader sampling of its close relatives.

This is the first study to combine ecological niche modeling with genomic analyses of island and mainland populations in this clade. We infer how populations are structured in space, the timing of divergence, the effects of climate change on geographic distributions, and evidence of selection in the new insular environment. Our main objective is to investigate the "bridge" hypothesis concerning the microevolutionary processes under-lying the origin and differentiation of *C. insularis* from its mainland relatives. Signatures of foundation followed by long-term gene

Table 1. The hypotheses tested.

Hypotheses	Mechanism	Prediction	Analyses	Acceptance
'Bridge' hypothesis	Some seamounts of FNC had suitable conditions for the occurrence of our ingroup during periods of sea level decrease. The colonization was due to a long-lasting contact promoted by a stepping-stone bridge of emerging seamounts between FNI and the mainland.	Suitable areas below the present-day sea level will be available during periods of sea level as low as LGM. Signature of foundation followed by a gradual decline of gene flow with mainland populations along time	Paleomodeling (MaxEnt, Random Forest, ANN) Quantification of the transient-land bridges in FNC (PleistoDist R package) Coalescent Analyses (ABC) 'Model 1' (Fig. 2)	Yes
Long-distance dispersal	Long-distance dispersal of few mainland individuals	severe short-term colonization of FNI, according to the classical founder event model Absence of suitable areas below the present-day sea level during periods of sea level as low as LGM	Coalescent Analyses (ABC) 'Model 2' (Fig. 2) Paleomodeling (MaxEnt, Random Forest, ANN)	No
Rapid adaptive divergence of C. insularis after the colonization of FNI	The new insular environment may play selective pressures in the colonizing population.	Signatures of divergent selection between C. insularis and C. fernambucensis	Outlier loci detection (BayeScan, PCAdapt)	Yes

Predictions are presented for each mechanism, stating expected results, the main analyses involved, and whether our results supported the hypothesis. FNC, Fernando de Noronha Chain; FNI, Fernando de Noronha Islands.

flow with mainland populations along Pleistocene, i.e., multiple colonization, would be consistent with this hypothesis as well as the identification of transient-land bridges of the FNC due to eustatic Pleistocene sea-level oscillation. The acceptance of this hypothesis would align with dynamic theory of island biogeography, which embraces the potential impact of sea level fluctuations on the shape, spatial arrangements, and connectedness of archipelagos and island-mainland systems throughout the Quaternary (Brown et al. 2013; Rijsdijk et al. 2014; Papadopoulou and Knowles 2015; Fernández-Palacios et al. 2016; Weigelt et al. 2016). The alternative to the 'bridge hypothesis' would be a long-distance dispersal, rendering signatures of a severe short-term colonization of FNI, following a classic founder event model and absence of climatic suitable conditions to the cactus occupying FNC submerged seamounts during the Pleistocene (Table 1). Finally, we also investigate if the new insular environment may play selective pressures in the colonizing population. Rapid adaptive divergence of *C. insularis* lineage after the colonization of FNI would be expected under this scenario (Table 1).

MATERIALS AND METHODS
Sampling, ddRAD Library Preparation and Sequencing
We sampled 50 specimens from 22 locations, comprising 14 with *C. fernambucensis*, four of *C. sericifer* F. Ritter (*C. fernambucensis* subsp. *sericifer*, Taylor and Zappi 2004), and four of *C. insularis*, covering the entire documented distribution of our ingroup (Fig. 1, Table S1). As FNI is categorized as a strictly protected area, we performed the sampling in accordance with Brazilian law through special permits provided to F.F.F. (permit from Chico Mendes Biodiversity Conservation Institute, SISBIO n° 38141-1). The final sampling size, varying between 2 and 8 individuals per location, has been shown to be appropriate for estimating population parameters for large SNP datasets (Nazareno et al. 2017). Seventeen samples were obtained from a previous ddRAD sequencing dataset generated in a recent study (Bombonato et al. 2020; SRA database access: PRJNA615034). The remaining were obtained from new ddRAD libraries (NCBI Sequence Read Archive (SRA) database access: PRJNA785598). We extracted high-quality genomic DNA for these samples from root tissues using the DNeasy Plant Mini Kit (Qiagen), and ddRAD libraries were prepared using the same enzymes (*Eco*RI and *HPA*II), size selection (400–500 bp), and protocol as Bombonato et al. (2020). Massive library sequencing was performed using paired-end Illumina HiSeq 2500 Rapid Run Mode (2 × 125 bp). Sequencing was conducted at BPI Biotecnologia, Pesquisa e Inovação Facility (Botucatu, São Paulo, Brazil).

Genomic data preparation, assembly, and datasets
After trimming adapters from the raw data, the quality of sequences was checked using FastQC 0.11.2 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc), visualized in MultiQC 1.0 (https://github.com/ewels/MultiQC) and filtered with SeqyClean 1.9.12 (Zhbannikov et al. 2017) setting the minimum quality (Phred Score ≥ 20), minimum size (>65 bp), and presence of Illumina contaminants (UniVec.fas). Illumina sequencing returned high coverage RAD-seq data across samples (mean = 1.25 M, sd = 654 K). Less than 1% of reads were excluded after trimming adapters and low-quality bases. Then, the data were demultiplexed and assembled using the de novo assembly pipeline in ipyrad v.0.9.57 (Eaton and Overcast 2020) using default assembly parameters except setting datatype = "pairddrad", mindepth_majrule = 2, and max_SNPs_locus = 0.25.

The assemble clustering identified homologous reads within each sample (mean = 80 K, sd = 28 K, mean_depth = 11.7) with relatively high coverage after excluding clusters below our minimum threshold (mean = 33 K, sd = 9 K, mean_depth = 25.3). Consensus diploid allele calls were made for each sample after filtering for paralogs based on excess heterozygosity or miscalled bases (mean = 31 K, sd = 8 K). Following initial exploratory analyses, two samples were dropped from the dataset due to low coverage (2B0, 4R0) and one as a potential contaminant (2E0). The final dataset was composed of 47 samples.

Clustering across samples identified 137,776 putative orthologous RAD loci, which, after filtering for minimum coverage and removing putative paralogs, was reduced to 56,599 loci. This dataset represents loci shared across at least four samples. The average number of loci per sample in this dataset was 14,122 (sd = 3210). This contains 355,688 SNPs in a total

concatenated matrix of 11,841,746 bp with 76.2% missing data. Downstream analyses of this dataset were performed using the *ipyrad.analysis* toolkit (Eaton and Overcast 2020; https://ipyrad.readthedocs.io/en/latest/API-analysis/index.html), which allows applying additional filters to reduce missing data for each analysis: 10,584 SNPs (522,328 bp, 25.6% missing data) for maximum likelihood phylogenetics; and 958 neutral loci (8% missing data) for clustering (PCA, STRUCTURE, AMOVA), coalescent-based and SNP-based approaches (ABC, BPP, SNaQ, and Spatial Diffusion). The latter dataset was submitted to a natural selection test to ensure the use of neutral loci in the downstream demographic, gene flow, and structure analyses (Note S1, Supporting Information). To test divergent selection on FNI, we perform outlier loci locus analysis with 5068 unlinked SNPs (~30% missing data), filtering from the island and mainland source population (details below).

Circumscription of genetic groups

To investigate the genetic circumscription of population groups, we first performed principal component analysis (PCA) (Pedregosa et al. 2011) and STRUCTURE (Pritchard et al. 2000) analyses implemented in *ipyrad-analysis*. The SNP dataset was filtered to allow at most 25% missing data across all samples (mincov = 0.75) and to require at least two samples with data from each of the six population groups identified (see results). A minor allele frequency cutoff of 0.10 was also applied (minmaf = 0.10). For PCA and STRUCTURE, a single SNP was randomly sampled from each variable RAD locus to approximate unlinked data. This process was repeated over ten replicates per K value in STRUCTURE and 20 replicates in PCA to examine variation among subsampled SNP datasets. The STRUCTURE analysis tested K values between 2 and 7 and was run for 500 K generations following a 100 K generation burn-in. The most likely number of genetic groups (K) was determined based on ΔK statistics (Evanno et al. 2005). Replicate analyses were combined using CLUMPP (Jakobsson and Rosenberg 2007). The remaining missing values after filtering (16.5%) were imputed prior to PCA using the "sampled" imputation method in ipyrad-analysis, which samples alleles by their frequency in each population. Finally, we performed global and hierarchical analyses of molecular variance (AMOVA) using the R packages *adegenet* (Jombart 2008) and *poppr* (Kamvar et al. 2014) to quantify genetic variation among groups obtained in the clustering analyses.

Phylogeny, divergence time estimates, and spatial diffusion analyses

The supermatrix alignment (522,328 bp; 10,584 SNPs; 25.6% missing data) was input to RAxML v.8.2.12 (Stamatakis 2014) to infer a maximum likelihood phylogeny under the GTRGAMMA model following Bombonato et al. (2020), using the "-f a" algorithm (i.e., rapid bootstrap) with starting searches from 10 random starting trees and 100 bootstrap replicates.

A quartet supertree was also inferred from SNP data using the program *tetrad* (Eaton et al. 2017). This method implements the *SVDquartets* algorithm (Chifman and Kubatko 2014) to infer a tree for every quartet of samples based on phylogenetic invariants and then uses the *maxcut* algorithm (Snirand Rao 2012) to join the quartet trees into a supertree. The final product is a consistent estimator of the species tree topology under the multispecies coalescence (Chifman and Kubatko 2014). *Tetrad* is optimized for RAD-seq data by maximizing the sampling of unlinked SNP data shared among each set of four samples regardless of missing data with other samples. This dataset also includes an average of 958 SNPs per quartet.

A time-calibrated species tree was also inferred from multi-locus data under the multispecies coalescent using *BPP* v.4.1.4 by assuming a molecular clock. This analysis implemented the BPP algorithm A00, taking a fixed species tree topology representing the backbone topology for the six lineages based on population clustering results to estimate model parameters. To translate estimated parameters of population genetic diversity (theta) and divergence times (tau) into effective population sizes (Ne) and time in years, respectively, we assumed a mean mutation rate of 7e-9/site/generation (Ossowski et al. 2010) and a 95% CI of (5e-9, 9e-9) that accommodate many rates described for seed plants (de la Torre et al. 2017), and mean generation time (g) of 7.5 years (95% CI: 5–9) based on observations by the author N.T. (Franco et al. 2022). Priors were inverse gamma distributed and weakly informative, theta = IG (3, 0.01) and tau = IG (3, 0.03). Loci were filtered to randomly select 1000 loci with no missing data for at least two samples from each lineage. BPP was run for 500 K generations, sampling every five generations after a 10 K generation burn-in.

To generate posterior probabilities for ancestral geographic areas, we used a spatial diffusion migration model in BEAST2 (Bouckaert et al. 2014) to conduct a spherical continuous phylogeography analysis. This analysis was run on the concatenated sequence alignment and applied a Yule tree prior and GTRGAMMA substitution model. We performed two independent runs of 50 million generations, sampled every 5000 steps, and combined them using LogCombiner v.2.3.2 (Bouckaert et al. 2014) after removing 25% as burn-in. The ESS values (>200) were analyzed in TRACER 1.6 (http://beast.bio.ed.ac.uk/Tracer). Spatial diffusion was displayed in Google Earth (https://earth.google.com/) based on the maximum clade credibility (MCC) tree using SPREAD v.1.07 (Bielejec et al. 2011).

Gene flow and introgression

To investigate the possible events of introgression and reticulation among the different lineages of *C. fernambucensis* and *C. insularis*, we used SNaQ (Solís-Lemus and Ané 2016) implemented in the Julia package PhyloNetworks v. 3.0.0 (Solís-Lemus et al. 2017). This method implements a maximum pseudolikelihood approach to infer a phylogenetic network from distributions of gene tree discordance. We inferred gene trees for 958 neutral loci using IQTREE2 v.2.0.3 (Minh et al. 2020) and a species tree using ASTRAL-III v.5.7.5 (Zhang et al. 2018), in which samples were collapsed into six lineages based on population clustering results. We estimated the quartet concordance factors (CF), as described by Solís-Lemus and Ané (2016), and used the species tree as the initial tree to infer additional network edges. We estimated a network from 0 to 10 reticulations with the best pseudolikelihood after 30 runs using the SNaQ function for each replication. We carried out ten replicate runs at each step to investigate pseudo-convergence.

Ecological Niche Modeling and effects of Pleistocene sea-level change on FNI

To investigate transient-land bridges of the FNC, we first investigate niche availability of island habitats using ecological niche modeling of climatic scenarios covering all available periods between the present and the Last Glacial Maximum, thus complementing the previous analyses presented in Franco et al. (2022); (present-day, LGM, LIG). The occurrence records (longitude/latitude) of *C. fernambucensis* and *C. insularis* were obtained from the online databases of the Global Biodiversity Information Facility (GBIF, http://www.gbif.org/) and complemented by our field registers. We checked and removed duplicate data, unlikely/conflicting, impossible (e.g., points in the ocean), inappropriate (e.g., within the city) points, and reduced spatial autocorrelation (Varela et al. 2014). The final filtered dataset includes 136 observations (Fig. S2). We used here the current climate scenario with the same spatial resolution (2.5 min), which was obtained from the most recent version of the 19 bioclimatic variables of WorldClim (Version 1.4; Hijmans et al. 2005) and nine past climate scenarios: (1) late-Holocene (lh; from 0.3 ka to 4.2 ka; Fordham et al. 2017); (2) mid-Holocene (mh; from 4.2 ka to 8.3 ka; Fordham et al. 2017); (3) early-Holocene (eh; from 8.3 ka to 11.7 ka; Fordham et al. 2017); (4) Younger Dryas Stadial (yds; from 11.7 ka to 12.9 ka; Fordham et al. 2017); (5) Bølling-Allerød (ba; from 12.9 ka to 14.7 ka; Fordham et al. 2017); (6) Heinrich Stadial 1 (hs; from 14.7 ka to 17 ka; Fordham et al. 2017); (7) the Last Glacial Maximum (LGM, ca. 21 Ka), and (8) the Last Interglacial (LIG, ca. 130 Ka; Otto-Bliesner et al. 2006). Highly correlated climate variables (Pearson's correlation > 0.75, Table S2) were removed using the functions present in the *RStoolbox* R package (Leutner and Horning 2017). The analysis was carried out using the *biomod2* R package (Thuiller et al. 2016), exploring three machine learning algorithms and presence-background methods, Artificial Neural Network (ANN), Random Forest (RF), and MaxEnt, with 50 replicates and using jackknife method, which approximate optimal model complexity and enhance predictions for small datasets and geographic range (Phillips et al. 2006; Shcheglovitova and Anderson 2013; Weber et al. 2017). The accessible area (M) has been defined to monitor the entire set of cells the species explores during the current period (Cooper and Soberón 2018). We used the results of the independent approaches to ensemble a final raster file. The area under the curve (AUC) was applied to evaluate model performance.

Additionally, to quantify the historical effects of eustatic Pleistocene sea-level in the transient-land bridges of the FNC, we used the *PleistoDist* R package, which enables visualization and quantifying Pleistocene sea-level effects on islands, testing various hypotheses of inter-island dispersal and community assembly (Tan et al. 2022). This software estimates the area and inter-island distance between the island and mainland in two sea-level time slices: the LGM and present-day (interval of 3 kya). This package
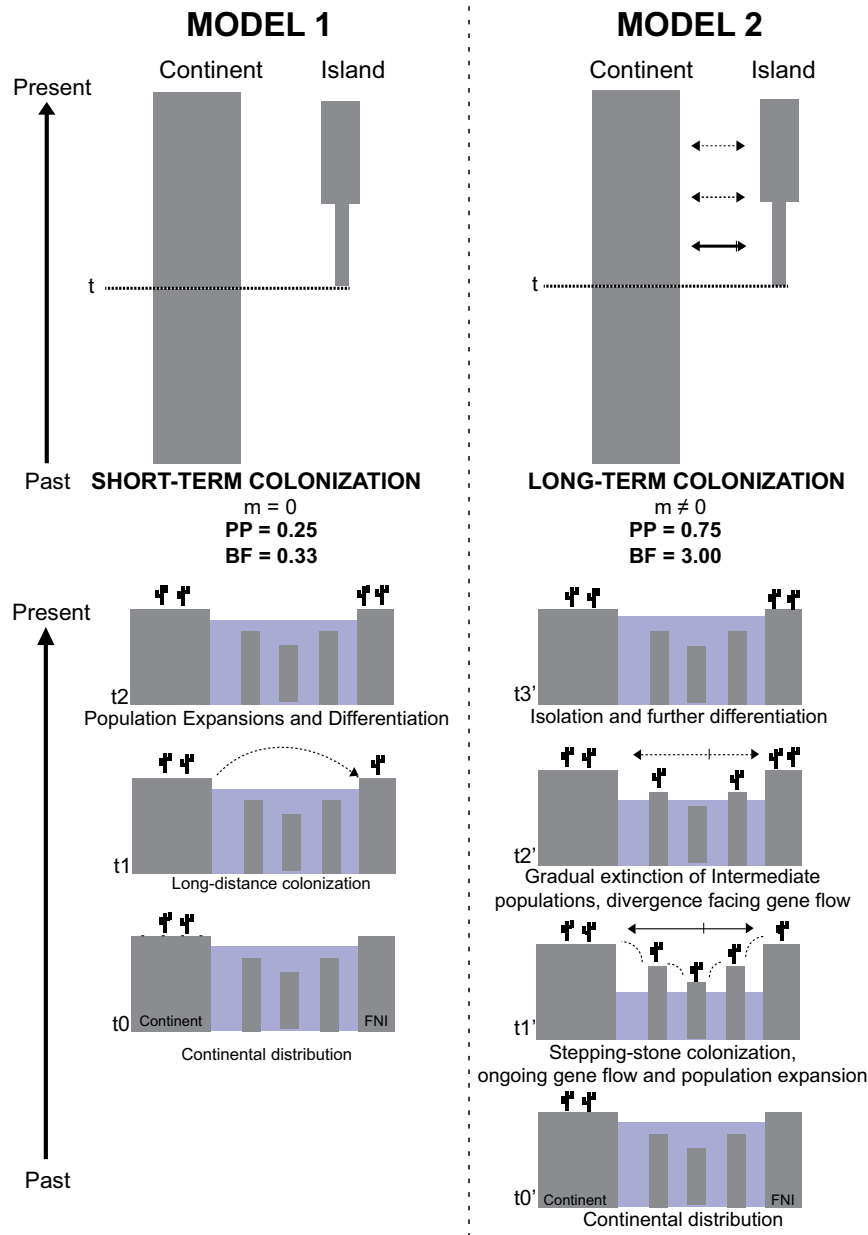
Fig. 2 Demographic models compared to test alternative scenarios for colonization of FNI by *C. insularis*. Arrows represent the probability of gene flow in models with a non-zero migration rate (m ≠ 0). Below each model, a schematic demonstrates a putative geographic rationale. Model 2 (outlined by a dotted rectangle) is the best fit for the data. pp = posterior probability for each model; BF: Bayes Factor.

models historical sea-level changes by decomposing Pleistocene levels into discrete intervals, providing flexibility in calculations. We use the bathymetry data obtained from the General Bathymetric Chart of the Oceans (GEBCO; https://www.gebco.net), points of the islands, and ten discrete intervals determined based on decomposing eustatic Pleistocene sea-levels proposed by Bintanja and van de Wal (2008).

Diversification history between insular and continental populations

We implemented a model testing framework using Approximate Bayesian Computation (ABC) to compare alternative demographic models of isolation and migration between *C. insularis* and the closest mainland *C. fernambucensis* population (Fig. 2). This includes the following two scenarios with ancestral continental distribution (t0 in Fig. 2): (1) "short-term colonization model" (STC) - the insular population is founded by few individuals from the continent (t1 in Fig. 2), followed by isolation at time τ, and exponential growth until it reaches its present effective size (t2 in Fig.

2); (2) "long-term colonization model" (LTC) - ongoing migration after the colonization (M1 = 1.0 to 100.0 migrants per generation) until time τ (t'1 in Fig. 2), when migration decreases (M2 = 0.0–1.0 migrants per generation) simulating extinction of intermediary populations as a response to sea level increase (t'2 in Fig. 2), while the insular population also experiences exponential growth from a founding population (t'3 in Fig. 2). Priors of population sizes and divergence times in demographic models were set to match the 95% Highest Posterior Density (HPD) from BPP results (Fig. S1).

We simulated 100,000 data sets of 46 summary statistics (SuSt, see GitHub/Gehara/PipeMaster for a detailed description of the SuSt) per model using 958 neutral loci and the *sim.msABC.sumstat* function of the R package PipeMaster (Gehara et al. 2017). For model inference, we summarized the information of the SuSt using PCA and incorporated >98% of the variation in the original data. We used the R package *abc* (Csilléry et al. 2012) to estimate the best-fitting demographic model based on a simple rejection method for each model with a threshold level of 0.0005, resulting in 100 simulations retained in the posterior. Posterior distributions for Ne and migration times for the selected model were also
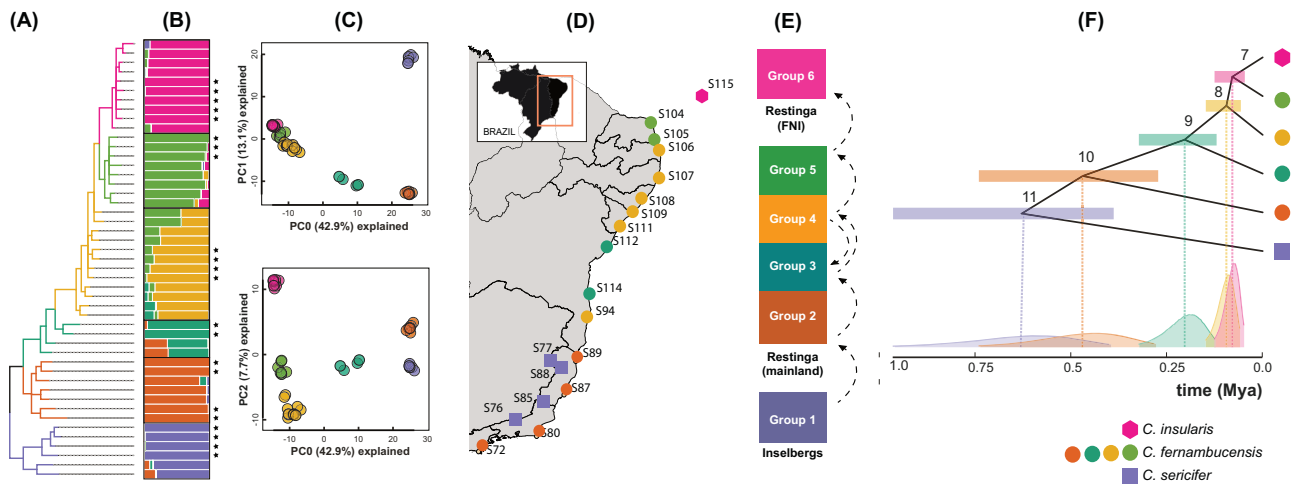
Fig. 3 Phylogeography and population structure. A Maximum Likelihood tree based on the concatenated supermatrix (522,328 bp, 10,584 SNPs; 25.6% missing data) generated using RaxML. **B, C** STRUCTURE barplot (k = 6) and PCA scatterplots showing population structure. **D** Population sampling locations are colored by the majority population cluster assignment. **E** Summary of Bayesian spatiotemporal diffusion (spherical continuous phylogeography analysis; see Fig. S4) for the six discrete genetic clusters, showing a northward range expansion along restinga forest, eventually reaching Fernando de Noronha Island (FNI). **F** Bayesian estimation of node ages under the multispecies coalescent model and a relaxed molecular clock in BPP. Asterisks next to taxon names in (**B**) indicate non-admixed individuals selected for divergence time estimation.

estimated using the 'abc' function in the abc package based on the neural networks (NN) method. As the performance of this method tends to improve with a greater number of retained simulations, we opted for a lower threshold level of 0.01, retaining 1200 simulations in the posterior.

Outlier loci detection

To investigate signatures of natural selection during FNI colonization, we used the 5068 biallelic unlinked SNPs (~30% of missing data), filtering from the FNI samples and the sister mainland *C. fernambucensis* population (group 5 and 6; see Fig. 3) and employed two scan methods based on locus-specific F_{ST} using the software BAYESCAN v2.1 (Foll and Gaggiotti 2008) and PCAdapt (Luu et al. 2017). For BAYESCAN, we implemented a prior odds of 100 and a false discovery rate (FDR) of 0.05. We conducted 20 pilot runs of 5000 iterations each, resulting in 100,000 iterations and a burn-in of 50,000, taking 5000 MCMC samples with a thinning interval of 10 steps. We used the *CODA* R-package to diagnose the log-likelihoods and F_{ST} values for the analysis. Considering that BAYESCAN is prone to Type I error, especially in complex population structures (Lotterhos and Whitlock 2014), we also applied a principal components analysis method using the *PCAdapt* R package (Luu et al. 2017), with an adjusted $p < 0.05$. SNPs commonly identified as outliers in both BAYESCAN and PCAdapt were classified as putative loci under selection. These loci were annotated using UniProtKB and KEGG Orthology (KO) databases.

RESULTS

Circumscription of genetic groups and evolutionary relationships

Phylogenetic results obtained by RAxML and tetrad (Fig. S3) recovered the same backbone topology with slight variations among relationships of samples within clades. This topology shows that (i) the branch of non-coastal populations associated with inselbergs (*C. sericifer*) is reciprocally monophyletic to a clade composing populations with a coastal distribution and associated with restinga forest (*C. fernambucensis* and *C. insularis*); (ii) *C. insularis* is a monophyletic group sister to a northernmost lineage of *C. fernambucensis;* and (iii) *C. fernambucensis* appears paraphyletic with respect to *C. insularis* (Fig. 3A). The Bayesian spatial diffusion ancestral phylogeographic reconstruction further supports these results, as the initial dispersal events occur from the inland countryside to coastal regions, followed by a south-to-north population expansion (Fig. 3E; Fig. S4).

We identified K = 4 and K = 6 as the best-fitting models using STRUCTURE (Fig. S5). We further confirmed K = 6 as a good fit by

performing a hierarchical AMOVA contrasting the four versus six groups, in which the K = 6 scenario shows the highest variation among groups (Φct = 0.45, $p < 0.05$; Table S3). The distribution of samples into discrete clusters in genetic PCA space also supports six groups (Fig. 3C). When aligned with our phylogenetic results (Fig. 3A, B), the K = 6 model reveals an admixture pattern between intermediate populations along the coast. The standard diversity indices for each population group are presented in Table 2.

The independent runs in PhyloNetwork with the same priors generated slightly distinctive results, but two common patterns were found in most analyses: introgression from group 3 to the ancestral node of groups 6, 5, and 4 and introgression from group 4 to group 6 (Fig. S6). These results are consistent with the admixture observed in STRUCTURE results and reinforce the presence of soft barriers along coastal restinga.

Demographic model fitting and divergence time estimate

The demographic model fit by BPP converged with ESS values > 200 for all parameters. Estimates of theta and tau were converted to Ne and divergence times in millions of years (Fig. S1; Table 2). The results support a northern expansion model for *C. fernambucensis*, where the southernmost population exhibits a similar effective population size as the clade ancestor. In contrast, each population to the north exhibits sequentially decreasing mean effective population sizes with increased distances (Fig. S1). The lowest mean effective population size is found in *C. insularis*, at less than one-quarter the value of the ancestral southernmost *C. fernambucensis* population (Table 2). Average levels of genetic diversity (He and π) show a similar south-to-north pattern (Table 2). The estimated crown age of our ingroup was 0.628 Mya (HPD 95%: 0.389–0.964 Mya), while the divergence of the island population *C. insularis* from the mainland occurred at 0.079 Mya (0.047–0.125 Mya) (Table 2; Fig. 3f). These divergence estimates are consistent with previous results based on more distant secondary calibrations within the genus *Cereus*, where the crown age of *C. insularis + C. fernambucensis* was 0.960 Mya (HPD 95%: 0.832–1.791) and the divergence of *C. insularis* from the mainland was 0.080 Mya (HPD 95%: 0.63–0.94) (Amaral et al. 2021).

Ecological Niche Modeling and FNC surface area along the Pleistocene

Overall, our modeling precisely predicts the present-day occurrence of both species (Fig. 4). The predictions of past paleoclimates

Table 2. Population genetic and demographic parameters inferred for five population groups of *Cereus fernambucensis* and one group of *C. insularis*.

Population group	Ho[a]	He[a]	π[¶]	Ne (95% HPD)	Stem Age (years)
Group 6	0.03	0.04	0.441 (0.69[b])	6287 (4100–9162)	79,088 (46,570–125,698)
Group 5	0.03	0.06	0.585 (1.23[b])	7342 (4660–10,987)	79,088 (46,570–125,698)
Group 4	0.04	0.06	0.665 (0.83[b])	6671 (4180–10,124)	94,650 (56,873–148,359)
Group 3	0.06	0.10	0.933 (1.83[b])	25,360 (15,709–38,733)	203,495 (119,953–323,196)
Group 2	0.05	0.10	0.943 (1.54[b])	43,675 (28,235–62,714)	469,877 (272,844- 741,054)
Group 1	0.04	0.10	0.815 (1.06[b])	26,051 (18,073–35,904)	628,337 (388,709–963,951)[c]

[a]Calculated using the *adegenet* R package (Jombart 2008).
[b]variance values.
[c]Crown age of the study ingroup.

show remarkable impacts on *C. fernambucensis* and *C. insularis*, both of which show suitable climatic conditions to inhabit FNI in LIG, as well as in some emerging sea mountains of the Fernando de Noronha chain during LGM and subsequent periods (Fig. 4; Fig. S7), suggesting that the present-day suitability for the occurrence of *C. insularis* only in FNI was reached in the mid-Holocene, when Rocas Atoll became unsuitable (8.4–4.2 ka; Fig. 4), complementing the Franco et al. (2022) finds. The surface area and inter-distance between islands in the FNC archipelago in the LGM and present-day are presented in Table S4. The exposition of the transient-land bridges and the Brazilian continental shelf during marine regressions likely allowed an eastern expansion of the mainland *C. fernambucensis* during the late Pleistocene (Fig. 4; Fig. S8).

Testing explicit scenarios for Fernando de Noronha Island colonization

ABC analysis recovered the highest posterior probability (PP: 0.75) for the "long-term colonization" model (Model 2; Fig. 2), supporting the scenario of colonization followed by long-term gene flow with mainland populations after the colonization (multiple colonizations). Our simulation analyses using ABC detected gene flow levels from about 79 ky up to 53 ky (Table S5), a timeframe that coincides with the *C. insularis* TMRCA (Time to the Most Recent Common Ancestor) and with the occurrence of sea levels low enough to allow the formation, at least in part, of the transient-land bridge of FNC (Fig. 5). This result suggests that island and continent populations maintained some genetic connectivity even after island colonization. This scenario could be interpreted as a long-term colonization of FNI and congruent with multiple colorizations within a model of divergence with gene flow.

Outlier loci under selection

We performed the FST outlier locus analysis in a pairwise comparison between the source mainland population (group 5) and FNI (group 6), as identified by previous analyses (Fig. 3). We recovered 375 and 59 candidate outliers loci using the PCAdapt and BayeScan software, respectively (Fig. S9). All outliers identified with BayeScan were also recovered by PCAdapt analysis and, thus, classified as putative loci under selection. Twenty-three of these loci are functional annotated, of which 15 presents metabolic pathways annotated in the KEGG database (Table S6).

DISCUSSION

By integrating genome-wide phylogeographic analyses with historical ecological niche modeling, we were able to test explicit hypotheses about the origins of *C. insularis* on the Fernando de Noronha islands. Our results support a progenitor-derivative pattern for the *C. fernambucensis - C. insularis* system, where the long-term colonization (i.e., multiple colonizations) of FNI occurred through a stepping-stone path during the late Pleistocene across a currently submerged landscape, supporting previous findings

(Franco et al. 2017; Amaral et al. 2021). The serial pattern of northward range expansion along the restinga reported by Franco et al. (2017) was corroborated here (Fig, S4), but the finer resolution provided by RAD-seq recovered new evidence of admixture among neighboring populations (Fig. 3; Fig, S6). These admixture events were inferred especially among the restinga populations, suggesting relatively soft barriers to gene flow along the Brazilian coast. This highlights the restinga forest as a more connected landscape (Pinheiro et al. 2011; Marques et al. 2015) and that the geographic barriers previously proposed for *C. fernambucensis*, such as the region of the 'Cabralia fault' (Franco et al. 2017), may be relatively permeable to gene flow (i.e., a soft barrier).

Considering the growing consensus on the impact of Quaternary climate on Brazilian Atlantic Forest biogeography (Peres et al. 2020), these climatic oscillations must be analyzed not only in terms of their impact on the distribution of this domain but also on the impact on the eustatic sea-levels affecting the distribution of a variety of species and populations of BAF (Ramos-Fregonezi et al. 2015; Leite et al. 2016). Here, we observe a shift in the distribution of *C. fernambucensis* since the LGM to the early Holocene, with incursions toward the Brazilian continental shelf. This result indicates that Restinga forest may have had a distribution reaching further to the east in the recent past, especially in the region of the Abrolhos bank toward the south, where the Brazilian shelf is enlarged (Fig. 4; Fig. S8), agreeing with previous finds for BAF (Leite et al. 2016; Ledru and Araújo 2023).

Fernando de Noronha islands (FNI) colonization: the 'bridge' hypothesis

Similar to other oceanic archipelagos, conjectures on the origins of biodiversity on the FNI have long been debated. For example, the 19th-century American geologist John C. Branner commented:

"It has been suggested also that the islands of Fernando may have been joined to the Brazilian mainland at one time, and that by the cutting away of the isthmus joining the two, the island was thus left with the fauna found by its discoverers. The form of the ocean's bottom between the mainland and the island puts this hypothesis out of question. It was formerly supposed that Fernando had once been the northeastern point of Brazil, but the deep-sea soundings by the Challenger expedition show that this is not true, and that Fernando is separated from the mainland by a trough more than 12,000 feet deep. In the light of these facts the question remains: where did these animals come from, and how did they get here?" (Branner 1888)

Indeed, except for migratory birds, the question posed by Dr. Branner (and likely many others) can be extended to the entire terrestrial FNI biota. It is now well-established that FNI originated from volcanic activity and has never been structurally connected to the mainland (Mohriak 2020). So how did the biota manage to disperse there? Our results support a "bridge hypothesis" as the
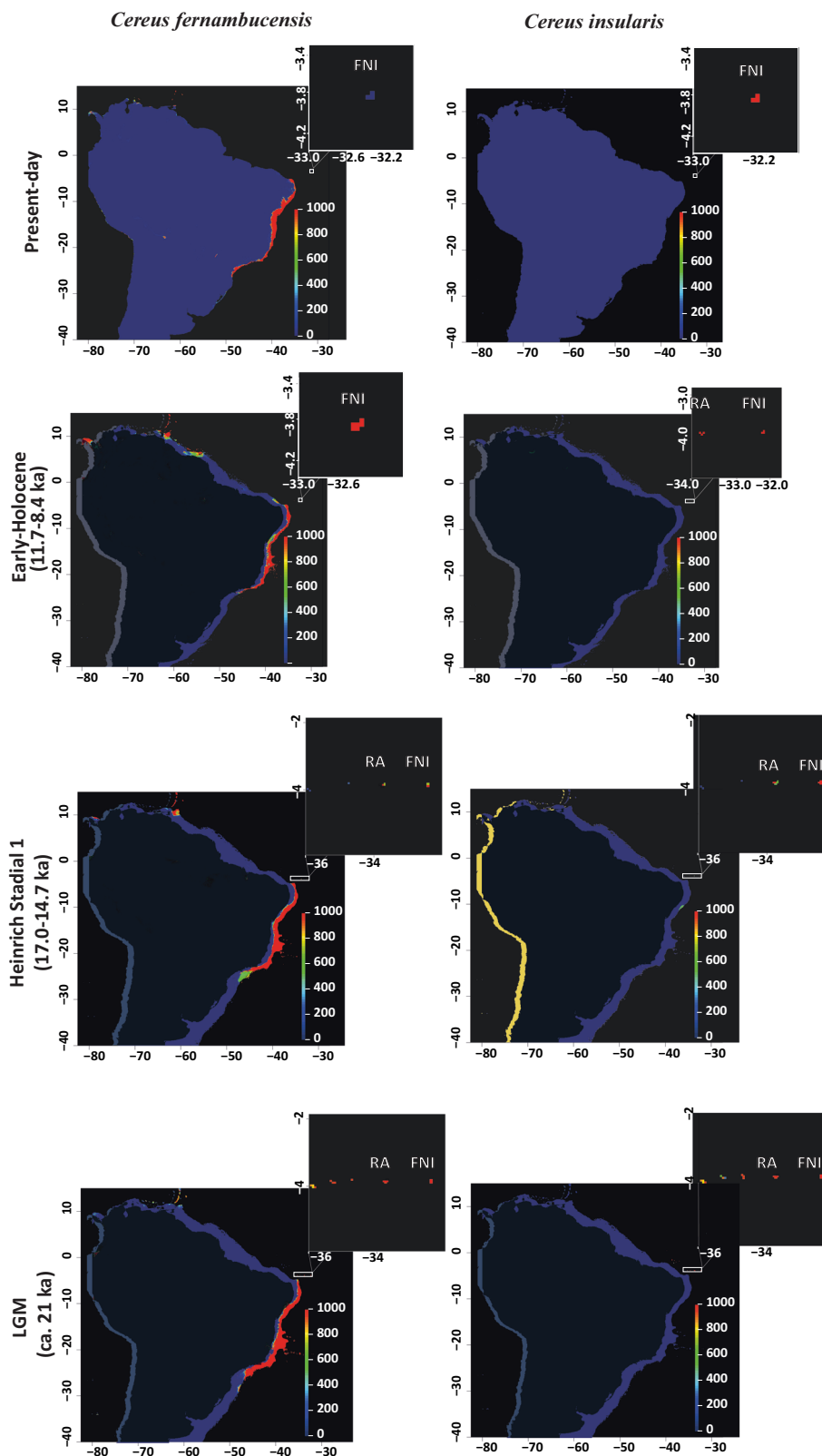
Fig. 4 Highlights of the Ecological niche modeling (ENM) of *Cereus fernambucensis* and *C. insularis* (first and second columns, respectively). All periods modeled can be found in Figure S7. Warmer colors on the SDM maps indicate more suitable areas for species occurrence, highlighting the Fernando de Noronha Chain (FNC), including the Fernando de Noronha Islands (FNI) and Rocas Atoll (RA).
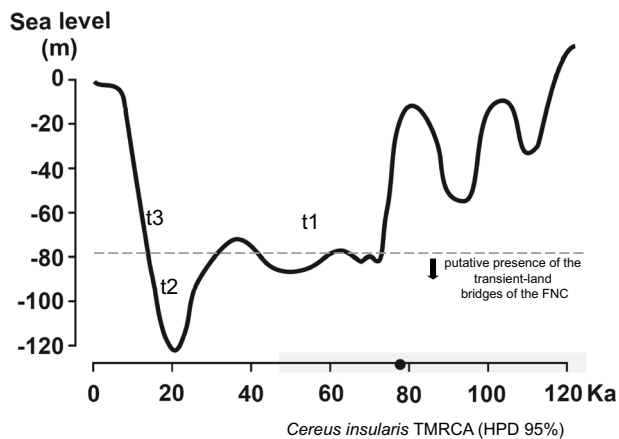
Fig. 5 Global sea-level change during the last 120 ka (adapted from Rijsdijk et al. 2014). The gray bar represents the 95% Highest Posterior Density (HPD) interval (46 Ka–125 Ka), and the black dot is the mean for the stem age of *C. insularis* (79 Ka). The dashed line represents the hypothesized sea level threshold (−80 m) associated with the presence or absence of the transient land bridge in the Fernando de Noronha Chain (FNC), as determined by our paleomodeling (Fig. 4) and assuming the same impact along the late Pleistocene. t1, putative period of long-term colonization followed by divergence facing recurrent gene flow; t2, putative period of maximum extension of the transient bridge in the LGM followed by gradual disconnection (Fig. 4); t3, isolation and further differentiation. TMRCA, Time to the Most Recent Common Ancestor.

most likely scenario for a cactus taxon, according to which, during periods of low sea level in the Pleistocene, regions of the island chain were exposed and sufficiently close to the mainland and to each other, to allow effective dispersal across a series of stepping-stones. First, this hypothesis is supported by discovering a "stepping-stone path" of exposed land with suitable niche space across the FNC island chain in times of reduced sea levels (Fig. 4; Table S4). Second, the estimated divergence time of *C. insularis* from the mainland (79 Ka, HPD: 47–126 Ka) overlaps with periods in which the sea level was as low as it was estimated to be at ~15 Ka (Fig. 5; Rijsdijk et al. 2014; Lambeck et al. 2014), when ENM models show high niche suitability on FNC seamounts (Fig. 4). Although interpolation methods are now available to simulate paleoclimates between LGM and LIG to test this statement (e.g., Oscillayers; Gamisch 2019), they are unable to depict events of sea-level oscillations in FNC due to database limitations (results not shown). Nonetheless, it is reasonable to assume that similar expositions of FNC were recurrent in the late Pleistocene, even before LGM (Fig. 5). Third, according to our ABC analysis, the best fitting demographic model is a "long-term colonization" scenario, where FNI colonization was followed by a gradual extinction of intermediate populations, leading to reduced genetic connectivity to the mainland (Fig. 2). While we have not identified gene flow extending to the lowest sea levels (LGM), the estimated period of gene flow (Table S5) coincides with a timeframe in which the transilient bridge would have probably existed (Fig. 5), supporting the "bridge" hypothesis. Finally, the genetic diversity of *C. insularis* is nearly identical to the northernmost mainland populations of *C. fernambucensis* (Table 2), as expected if connectivity allowed for gene flow after multiple colonization (e.g., García-Verdugo et al. 2015; 2017; Kimura et al. 2022), in contrast to a severe founder-event, or short-term colonization of FNI (Franco et al. 2022).

Even under the bridge stepping-stone path, a dispersal to FNC may have represented a major biogeographic event, given that the Euclidean distance from the continent to one of the nearest putative intermediate islands is 80–100 Km (Table S4). However, according to island biogeography theory, the FNC system meets all relevant criteria that contribute to the likelihood of such an

event (Stuessy et al. 2014; Flantua et al. 2020). For example, (1) we predict the existence of intermediate populations (across the island chain) which reduce the "distance to the nearest species source" for each island (MacArthur and Wilson 1967); (2) the large "number of stepping-stones" in FNC also reduces the longest distance among all pairs of stepping stones (Diver 2008; Mendez-Castro et al. 2020); and finally (3) we predict much shorter distances between islands and mainland during periods of low sea level. This last component comes from the greater area of exposed islands [for example, seamounts such as 'Alto Fundo Drina' guyot (see Figure 17 in Mohriak 2020)], providing additional intermediate stepping-stone paths, and also from the intrusion of continental populations to the northeastern Brazilian shelf, which reduces the linear distance between continent and islands. The likely shortened linear distance would lead to a reduced "target effect", where smaller and more isolated islands are less likely to be colonized (MacArthur and Wilson 1967; Stuessy et al. 2014; Mendez-Castro et al. 2020). The impacts of sea level on both the size and connectedness of islands across the FNC makes the dispersal of a cactus species across this region more likely through seed dispersal by birds or storm-induced floating stems and fruits (e.g., Taylor et al. 2021; Taylor 2021).

In addition to FNI, the Rocas Atoll ("RA" in Fig. 4), which lies between FNI and the mainland, was also consistently recovered as a suitable area for *C. insularis* from the LGM to the early Holocene, suggesting it could have played a key role as part of a stepping-stone path. A puzzling question, then, is: why no cactus population exists on Rocas Atoll in the present day? A likely explanation may be the rapid rate at which sea level changes can impact oceanic islands. Currently, the volcanic edifice of Rocas Atoll is fully submerged and fringed by carbonate reefs (Mohriak 2020), extending on 7.2 Km² of the exposed area above coralline formation (Soares et al. 2011). The sand patches ('Farol' and 'Cemiterio') found in this Atoll have less than 0.5 Km² and a maximum altitude of 3 m, harboring an impoverished flora restricted to some herbs. These sand patches are likely unsuitable for the long-term persistence of cactus populations at the present day, as they would be susceptible to periodic flooding and other marine weather.

To the best of our knowledge, this is the first work to explicitly test a hypothesis of colonization of a true-island system associated with the BAF domain. Our hypothesis is also wide enough to explain the colonization of other FNI terrestrial flora and fauna, which may be the target of future studies. Likewise, the same hypothesis testing framework may be applied to other archipelagos with similar spatial configurations (i.e., connected with the mainland by a chain of seamounts). This is the case, for example, of the Trindade and Martin Vaz volcanic islands, which are the easternmost segment of the Vitória-Trindade Chain, which in turn are allocated on the eastern margin of the Abrolhos bank (Mohriak 2020). It is worth noting that despite Trindade and Martin Vaz having no occurrence of cacti in the present day, our ENM reveals a suitable occurrence for *C. fernambucensis* during the LGM (Fig. S8). This suggests that at least some elements of the restinga forest may have previously reached this archipelago under the possibility of a "bridge" scenario.

Speciation models

Theory suggests that founder effects may be a driver of insular evolution, highlighting an association between genetic drift and selection in diversification (e.g. Mayr 1954; Carson 1968; Templeton 1981). However, the relative importance of these microevolutionary processes remains a fundamental question in island biogeography (Patiño et al. 2017).

Here, we recovered a "budding" branching pattern supporting *C. insularis* as a subclade nested within *C. fernambucensis*, in agreement with the hypothesis of the progenitor-derivative speciation model (Table 1). This result is consistent with a prediction about the role of genetic drift in the speciation process,

i.e., that the descendant species is monophyletic within the paraphyletic progenitor (Bock et al. 2023). However, demographic modeling rejected a scenario of severe short-term colonization of FNI (as discussed above). Thus, although our results indicate a peripheral differentiation of *C. insularis* after colonization of FNI at approximately 80 Ka (Table 2), it does not appear to have occurred through a single founder event. Indeed, growing evidence suggests that the genetic connectivity between islands and mainland source areas will depend on several physical and biological factors (Stuessy et al. 2014; Alsos et al. 2015; García-Verdugo et al. 2017; 2019; Kimura et al. 2022), which may vary over time (García-Verdugo et al. 2015; Fernández-Palacios et al. 2016; Flantua et al. 2020), and may contribute to the maintenance of genetic diversity in recently colonized islands.

C. insularis seems to have diverged morphologically in the presence of recurrent gene flow with the *C. fernambucensis* mainland population (Fig. 2), as suggested previously (Franco et al. 2022). Models of speciation with gene flow (isolation-with-migration) have shown an important role in divergent selection (Nosil 2008; 2009). Indeed, selection may play a more significant role than drift in many cases of rapid evolutionary divergence in peripheral populations (García-Ramos and Kirkpatrick 1997; Grossenbacher et al. 2014). We have detected 59 candidate loci under selection (Fig. S8), of which 23 are annotated (Table S6). The enrichment analysis of gene functions and the analysis of metabolic biosynthetic pathways for these gene products revealed associations with photosynthetic electron transport, photorespiration, and the metabolism of carbohydrates, amino acids, vitamins, terpenoids, and polyketides, including phosphate and lipids. These associations may be linked to nutrient starvation and abiotic stress, common conditions expected in an insular environment due to limited soil resources, extreme temperatures, salinity, and exposure to winds. (Pichersky and Raguso 2018; Smirnoff 2018; Pan et al. 2019).

The genes under positive selection may have played a role in the speciation of *C. insularis*. Each of these genes has specific functions that have probably contributed to the adaptation and survival of this species in the challenging conditions of the FNI. For instance, cellulose-related genes like cellulose synthase A catalytic subunit 6 [UDP-forming]-like may impact cellular structure and dehydration resistance (Chen et al. 2005; Wang et al. 2016). Similarly, genes such as GTP cyclohydrolase 1 and adenylate isopentenyltransferase 5, are involved in some crucial metabolic processes, potentially aiding in the adaptation to limited soil resources (Hung et al. 2022; Zhang et al. 2021). Furthermore, the positive selection on stress-responsive genes, including alpha-amylase, dehydroascorbate reductase 2, and Zinc finger CCCH-type, also suggests a potential role in the adaptation to challenging island conditions, encompassing temperature fluctuations and oxidative stress (Lu et al. 2022; Zulfiqar and Ashra 2021).

In this context, the rapid anagenetic differentiation of the *C. insularis* lineage in FNI might be associated with divergent selection pressures experienced by mainland versus insular populations. Fine-scale climatic niche differentiation has been demonstrated between *C. insularis* and *C. fernambucensis* (Franco et al. 2022). Despite both species occurring in the same biome (restinga forest), *C. insularis* grows on volcanic substrates, while *C. fernambucensis* does not. Vegetative differences, such as the higher number of ribs and spines on *C. insularis* (Taylor and Zappi 2004), have been proposed to protect cactus tissue from excessive solar radiation (Menezes et al. 2015; Aquino et al. 2021), which is extreme in the equatorial region of FNI. Taylor and Zappi (2004) hypothesize that *C. insularis* may be a neotenic derivative form of *C. fernambucensis*, such that their morphological differences could be explained by a few regulatory changes to development.

However, it is essential to acknowledge that our FST outliers' results should be cautiously interpreted, considering limitations related to sampling design acquisition bias inherent in RAD data, which may vary depending on the species system, population structure, and

statistical methods applied (e.g., Lotterhos and Whitlock 2014; Lotterhos and Whitlock 2015; Cariou et al. 2016; Lowry et al. 2017). Thus, future studies using genomic and transcriptomic data are needed to properly address the role of selection during *C. insularis* divergence, and how ecological divergence contributes to reproductive isolation with *C. fernambucensis*.

In conclusion, our work corroborates the role of historical effects of eustatic Pleistocene sea-level in the transient-land bridges of the FNC to explain the origin of *C. insularis* (the "bridge" hypothesis; Franco et al. 2017). The progenitor-derivative pattern in the *C. fernambucensis* - *C. insularis* system is typical of a "budding speciation" process, which may be particularly common among plants (Crawford 2010; Anacker and Strauss 2014; Zhao et al. 2020), and provides a valuable system in which to study the early stages of the speciation process (Crawford 2010). It provides a natural experiment because the progenitor species tend to retain the ancestral type (Stuessy et al. 2006) and can be sampled as a gradation of paraphyletic lineages with increasing genetic distance from the derivative species (Rieseberg and Brouillet 1994; Gottlieb 2004). In addition, the paraphyletic pattern observed in many peripheral species illustrates how the acquisition of reciprocal monophyly, which is a fundamental criterion for systematics, can be understood as a property resulting from complementary stages in the process of continuous differentiation in incipient species, in accordance with the unified species concept (De Queiroz 2007). The species-rich cactus family is an interesting taxon for investigating such progenitor-derivative speciation model. Apart from some reported cases of peripheral speciation in cactus species (e.g., Majure et al. 2021), there are many species pairs with asymmetrical geographic ranges, involving both widespread and microendemic relatives, perhaps as a consequence of adaptation to challenging environments such as insular habitats.

Data archiving
The raw data is available in the SRA public database under the project number PRJNA785598.

REFERENCES

Alsos IG, Ehrich D, Eidesen PB, Solstad H, Westergaard KB, Schönswetter P et al. (2015) Long-distance plant dispersal to North Atlantic islands: colonization routes and founder effect. AoB Plants 15:plv036

Amaral DT, Minhós-Yano I, Oliveira JVM, Romeiro-Brito M, Bonatelli IAS, Taylor NP et al. (2021) Tracking the xeric biomes of South America: the spatiotemporal diversification of Mandacaru cactus. J Biogeogr 48:3085–3103

Anacker BL, Strauss SY (2014) The geography and ecology of plant speciation: range overlap and niche divergence in sister species. Proc R Soc Lond B Biol Sci 281:20132980

Aquino D, Moreno-Letelier A, González-Botello MA, Arias S (2021) The importance of environmental conditions in maintaining lineage identity in *Epithelantha* (Cactaceae). Ecol Evol 11:4520–4531

Barton NH, Charlesworth B (1984) Genetic revolutions, founder effects, and speciation. Annu Rev Ecol Evol Syst 15:133–164

Bielejec F, Rambaut A, Suchard MA, Lemey P (2011) SPREAD: spatial phylogenetic reconstruction of evolutionary dynamics. Bioinformatics 27:2910–2912

Bintanja R, Van de Wal RSW (2008) North American ice-sheet dynamics and the onset of 100,000-year glacial cycles. Nature 454:869–872

Bock DG, Cai Z, Elphinstone C, González-Segovia E, Hirabayashi K, Huang K et al. (2023) Genomics of plant speciation. Plant Commun 4:100599

Bombonato JR, Amaral DT, Silva GAR, Khan G, Moraes EM, da Silva Andrade SC et al. (2020) The potential of genome-wide RAD sequences for resolving rapid radiations: a case study in Cactaceae. Mol Phylogenet Evol 151:106894

Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu CH, Xie D et al. (2014) BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput Biol 10:e1003537

Branner JC (1888) Notes on the Fauna of the Islands of Fernando de Noronha. Am Nat 22:861–871

Brown RM, Siler CD, Oliveros CH, Esselstyn JA, Diesmos AC, Hosner PA et al. (2013) Evolutionary processes of diversification in a model island archipelago. Annu Rev Ecol Evol Syst 44:411–435

Cariou M, Duret L, Charlat S (2016) How and how much does RAD-seq bias genetic diversity estimates? BMC Evol Biol 16:1–8

Carson HL (1968) The population flush and its genetic consequences. In: Lewontin RC (Ed) Population Biology and Evolution. Syracuse University Press, Syracuse, NY, p 123–137

Chen Z, Hong X, Zhang H, Wang Y, Li X, Zhu JK et al. (2005) Disruption of the cellulose synthase gene, AtCesA8/IRX1, enhances drought and osmotic stress tolerance in Arabidopsis. Plant J 43:273–283

Chifman J, Kubatko L (2014) Quartet inference from SNP data under the coalescent model. Bioinformatics 30:3317–3324

Cooper JC, Soberón J (2018) Creating individual accessible area hypotheses improves stacked species distribution model performance. Glob Ecol Biogeogr 27:156–165

Crawford DJ (2010) Progenitor-derivative species pairs and plant speciation. Taxon 59:1413–1423

Csilléry K, François O, Blum MG (2012) abc: an R package for approximate Bayesian computation (ABC). Methods Ecol Evol 3:475–479

De Queiroz K (2007) Species concepts and species delimitation. Syst Biol 56:879–886

de la Torre AR, Li Z, de Peer YV, Ingvarsson PK (2017) Contrasting rates of molecular evolution and patterns of selection among gymnosperms and flowering plants. Mol Biol Evol 34:1363–1377

Diver KC (2008) Not as the crow flies: assessing effective isolation for island bio-geographical analysis. J Biogeogr 35:1040–1048

Eaton DA, Overcast I (2020) ipyrad: Interactive assembly and analysis of RADseq datasets. Bioinf 36:2592–2594

Eaton DA, Spriggs EL, Park B, Donoghue MJ (2017) Misconceptions on missing data in RAD-seq phylogenetics with a deep-scale example from flowering plants. Syst Biol 66:399–412

Emerson BC, Patiño J (2018) Anagenesis, cladogenesis, and speciation on islands. Trends Ecol Evol 33:488–491

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14:2611–2620

Fernández-Palacios JM, Rijsdijk KF, Norder SJ, Otto R, de Nascimento L, Fernández-Lugo S et al. (2016) Towards a glacial-sensitive model of island biogeography. Glob Ecol Biogeogr 25:817–830

Flantua SG, Payne D, Borregaard MK, Beierkuhnlein C, Steinbauer MJ, Dullinger S et al. (2020) Snapshot isolation and isolation history challenge the analogy between mountains and islands used to understand endemism. Glob Ecol Biogeogr 29:1651–1673

Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. Genetics 180:977–993

Fordham DA, Saltré F, Haythorne S, Wigley TM, Otto-Bliesner BL, Chan KC et al. (2017) PaleoView: a tool for generating continuous climate projections spanning the last 21000 years at regional and global scales. Ecography 40:1348–1358

Franco FF, Jojima CL, Perez MF, Zappi DC, Taylor N, Moraes EM (2017) The xeric side of the Brazilian Atlantic Forest: The forces shaping phylogeographic structure of cacti. Ecol Evol 7:9281–9293

Franco FF, Silva FS, Khan G, Bonatelli IAS, Amaral DT, Zappi DC et al. (2022) Generalizations of genetic conservation principles in islands are not always likely: a case study from a Neotropical insular cactus. Bot J Linn Soc 199:210–227

Frankham R, Lees, K, Montgomery ME, England PR, Lowe EH, Briscoe DA (1999) Do population size bottlenecks reduce evolutionary potential? In Animal Conservation Forum. Cambridge University Press, UK, pp. 255-260

Frey JK (1993) Modes of peripheral isolate formation and speciation. Syst Biol 42:373–381

Futuyma DJ (1989) Speciational trends and the role of species in macroevolution. Am Nat 134:318–321

García-Ramos G, Kirkpatrick M (1997) Genetic models of adaptation and gene flow in peripheral populations. Evolution 51:21–28

García-Verdugo C, Sajeva M, La Mantia T, Harrouni C, Msanda F, Caujapé-Castells J (2015) Do island plant populations really have lower genetic variation than mainland populations? Effects of selection and distribution range on genetic diversity estimates. Mol Ecol 24:726–41

García-Verdugo C, Mairal M, Monroy P, Sajeva M, Caujapé-Castells J (2017) The loss of dispersal on islands hypothesis revisited: implementing phylogeography to investigate evolution of dispersal traits in Periploca (Apocynaceae). J Biogeogr 44:2595–2606

García-Verdugo C, Caujapé-Castells J, Mairal M, Monroy P (2019) How repeatable is microevolution on islands? Patterns of dispersal and colonization-related plant traits in a phylogeographical context. Ann Bot 123:557–568

Gamisch A (2019) Oscillayers: a dataset for the study of climatic oscillations over Plio-Pleistocene time-scales at high spatial-temporal resolution. Glob Ecol Biogeogr 28:1552–1560

Gehara M, Garda AA, Werneck FP, Oliveira EF, da Fonseca EM, Camurugi F et al. (2017) Estimating synchronous demographic changes across populations using hABC

and its application for a herpetological community from northeastern Brazil. Mol Ecol 26:4756–4771

Gomes VGN, Quirino ZGM, Araujo HFP (2014) Frugivory and seed dispersal by birds in Cereus jamacaru DC. ssp. jamacaru (Cactaceae) in the Caatinga of Northeastern Brazil. Braz J Biol 74:32–40

Gottlieb LD (2004) Rethinking classic examples of recent speciation in plants. N. Phytol 161:71–82

Grant V (1971) Plant speciation. Columbia Univ. Press, NY and London, p 435

Grant V (1989) The theory of speciational trends. Am Nat 133:604–612

Grossenbacher DL, Veloz SD, Sexton JP (2014) Niche and range size patterns suggest that speciation begins in small, ecologically diverged populations in North American monkeyflowers (Mimulus spp.). Evolution 68:1270–1280

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. Int J Climatol 25:1965–1978

Hung CY, Zhu C, Kittur FS, He M, Arning E, Zhang J et al. (2022) A plant-based mutant huntingtin model-driven discovery of impaired expression of GTPCH and DHFR. Cell Mol Life Sci 79:553

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

Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24:1403–1405

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

Kimura T, Yamada T, Sakaguchi S, Ito M, Maki M (2022) Multiple colonizations and genetic differentiation in goldenrod populations on recently formed nearshore islands. J Biogeogr 49:836–852

Lambeck K, Rouby H, Purcell A, Sun Y, Sambridge M (2014) Sea level and global ice volumes from the Last Glacial Maximum to the Holocene. Proc Natl Acad Sci USA 111:15296–15303

Ledru M, Araújo F (2023) The Restinga and Cerrado pathways, two ancient biotic corridors in the Neotropics. Front Biogeogr 15:e59398

Leite YL, Costa LP, Loss AC, Rocha RG, Batalha-Filho H, Bastos AC et al. (2016) Neo-tropical forest expansion during the last glacial period challenges refuge hypothesis. Proc Natl Acad Sci USA 113:1008–1013

Leutner B, Horning N (2017) Rstoolbox: Tools for Remote Sensing Data Analysis. R Package Version 0.1.8. URL https://cran.r-project.org/web/packages/RStoolbox/index.html. Accessed 30 November 2020.

Losos JB, Glor RE (2003) Phylogenetic comparative methods and the geography of speciation. Trends Ecol Evol 18:220–227

Losos JB, Ricklefs RE (2009) Adaptation and diversification on islands. Nature 457:830–836

Lotterhos KE, Whitlock MC (2014) Evaluation of demographic history and neutral parameterization on the performance of FST outlier tests. Mol Ecol 23:2178–2192

Lotterhos KE, Whitlock MC (2015) The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. Mol Ecol 24:1031–46

Lowry DB, Hoban S, Kelley JL, Lotterhos KE, Reed LK, Antolin MF et al. (2017) Breaking RAD: an evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. Mol Ecol Resour 17:142–152

Lu F, Duan W, Cui Y, Zhang J, Zhu D, Zhang M et al. (2022) 2D-DIGE based proteome analysis of wheat-Thinopyrum intermedium 7XL/7DS translocation line under drought stress. BMC Genom 23:369

Luu K, Bazin E, Blum MG (2017) pcadapt: an R package to perform genome scans for selection based on principal component analysis. Mol Ecol Resour 17:67–77

MacArthur RH, Wilson EO (1967) The theory of island biogeography. Princeton University Press.

Majure LC, Barrios D, Díaz E, Zumwalde BA, Testo W, Negrón-Ortíz V (2021) Pleisto-cene aridification underlies the evolutionary history of the Caribbean endemic, insular, giant Consolea (Opuntioideae). Am J Bot 108:200–215

Marques MC, Silva SM, Liebsch D (2015) Coastal plain forests in southern and southeastern Brazil: ecological drivers, floristic patterns and conservation status. Rev Bras Bot 38:1–18

Mayr E (1954) Change of genetic environment and evolution. In: Huxley J, Hardy AC, Ford EB (Eds) Evolution as a process. Collier Books, New York, p 157–180

Maya-Lastra CA, Eaton DAR (2021) Genetic incompatibilities do not snowball in a demographic model of speciation. Biorxiv. https://doi.org/10.1101/2021.02.23.432472.

Méndez-Castro FE, Mendieta-Leiva G, Rao D, Bader MY (2020) Island-biogeographic patterns of spider communities on epiphytes depend on differential space use among functional groups. J Biogeogr 47:1322–1332

Menezes MO, Taylor NP, Zappi DC, Loiola MIB (2015) Spines and ribs of Pilosocereus arrabidae (Lem.) Byles and GD Rowley and allies (Cactaceae): ecologic or genetic traits? Flora 214:44–49

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A et al. (2020) IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol 37:1530–1534

Mohriak W (2020) Genesis and evolution of the South Atlantic volcanic islands off-shore Brazil. Geo-Mar Lett 40:1–33

Nazareno AG, Bemmels JB, Dick CW, Lohmann LG (2017) Minimum sample sizes for population genomics: an empirical study from an Amazonian plant species. Mol Ecol Resour 17:1136–1147

Nosil P (2008) Ernst Mayr and the integration of geographic and ecological factors in speciation. Biol J Linn Soc Lond 95:26–46

Nosil P (2009) Adaptive population divergence in cryptic color-pattern following a reduction in gene flow. Evolution 63:1902–1912

Ossowski S, Schneeberger K, Lucas-Lledó JI, Warthmann N, Clark RM, Shaw RG et al. (2010) The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. Science 327:92–94

Otto-Bliesner BL, Brady EC, Clauzet G, Tomas R, Levis S, Kothavala Z (2006) Last glacial maximum and Holocene climate in CCSM3. J Clim 19:2526–2544

Pan W, Wu Y, Xie Q (2019) Regulation of ubiquitination is central to the phosphate starvation response. Trends Plant Sci 24:755–769

Papadopoulou A, Knowles LL (2015) Genomic tests of the species-pump hypothesis: recent island connectivity cycles drive population divergence but not speciation in Caribbean crickets across the Virgin Islands. Evolution 69:1501–1517

Patiño J, Whittaker RJ, Borges PA, Fernández-Palacios JM, Ah-Peng C, Araújo MB et al. (2017) A roadmap for island biology: 50 fundamental questions after 50 years of The Theory of Island Biogeography. J Biogeogr 44:963–983

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O et al. (2011) Scikit-learn: machine learning in Python. J Mach Learn Res 12:2825–2830

Peres EA, Pinto-da-Rocha R, Lohmann LG, Michelangeli FA, Miyaki CY, Carnaval AC (2020) Patterns of species and lineage diversity in the Atlantic Rainforest of Brazil. In: Rull V, Carnaval AC (Eds) Neotropical diversification: patterns and processes. Springer, Cham, p 415–447

Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. Ecol Model 190:231–259

Pichersky E, Raguso RA (2018) Why do plants produce so many terpenoid compounds? N. Phytol 220:692–702

Pinheiro F, de Barros F, Palma-Silva C, Fay MF, LexerC, Cozzolino S (2011) Phylogeography and genetic differentiation along the distributional range of the orchid *Epidendrum fulgens*: A Neotropical coastal species not restricted to glacial refugia. J Biogeogr 38:1923–1935

Pritchard JK, Stephens M, Rosenberg NA, Donnelly P (2000) Association mapping in structured populations. Am J Hum Genet 67:170–181

Ramos-Fregonezi AM, Fregonezi JN, Cybis GB, Fagundes NJ, Bonatto SL, Freitas LB (2015) Were sea level changes during the Pleistocene in the South Atlantic Coastal Plain a driver of speciation in Petunia (Solanaceae)? BMC Evol Biol 15:1–11

Rieseberg LH, Brouillet L (1994) Are many plant species paraphyletic? Taxon 43:21–32

Rijsdijk KF, Hengl T, Norder SJ, Otto R, Emerson BC, Ávila SP et al. (2014) Quantifying surface-area changes of volcanic islands driven by Pleistocene sea-level cycles: Biogeographical implications for the Macaronesian archipelagos. J Biogeogr 41:1242–1254

Schluter D (2009) Evidence for ecological speciation and its alternative. Science 323:737–41

Sendell-Price AT, Ruegg KC, Robertson BC, Clegg SM (2021) An island-hopping bird reveals how founder events shape genome-wide divergence. Mol Ecol 30:2495–2510

Shcheglovitova M, Anderson RP (2013) Estimating optimal complexity for ecological niche models: a jackknife approach for species with small sample sizes. Ecol Model 269:9–17

Smirnoff N (2018) Ascorbic acid metabolism and functions: a comparison of plants and mammals. Free Radic Biol Med 122:116–129

Snir S, Rao S (2012) Quartet MaxCut: a fast algorithm for amalgamating quartet trees. Mol Phylog Evol 62:1–8

Soares MO, Lemos VB, Kikuchi RKP (2011) Aspectos biogeomorfológicos do Atol das Rocas, Atlântico Sul Equatorial. Rev Bras Biocienc 41:85–94

Solís-Lemus C, Ané C (2016) Inferringphylogenetic networks with maximum pseudolikelihood under incomplete lineage sorting. PLoS Genet 12:e1005896

Solís-Lemus C, Bastide P, Ané C (2017) PhyloNetworks: a package for phylogenetic networks. Mol Biol Evol 34:3292–3298

Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313

Stuessy TF, Crawford DJ, Marticorena C (1990) Patterns of phylogeny in the endemic vascular Flora of the Juan Fernandez Islands, Chile. Syst Bot 15:338–346

Stuessy TF, Jakubowsky G, Salguero-Gómez R, Pfosser M, Schluter PM, Fer T et al. (2006) Anagenetic evolution in island plants. J Biogeogr 33:1259–1265

Stuessy TF, Takayama K, López-Sepúlveda P, Crawford DJ (2014) Interpretation of patterns of genetic variation in endemic plant species of oceanic islands. Bot J Linn Soc 174:276–288

Takayama K, López-Sepúlveda P, Greimler J, Crawford DJ, Peñailillo P, Baeza M et al. (2015) Genetic consequences of cladogenetic vs. anagenetic speciation in endemic plants of oceanic islands. AoB Plants 7:plv102

Tan DJ, Gyllenhaal EF, Andersen MJ (2022) PleistoDist: a toolbox for visualising and quantifying the effects of Pleistocene sea-level change on island archipelagos. Methods Ecol Evol 14:496–504

Taylor NP (2021) Notes on cultivated Pereskia quisqueyana and related species (Cactaceae). Bradleya 2021:69–110

Taylor NP, Hoxey P, Gdaniec A (2021) Dendrocereus Britton and Rose–one species or two? Bradleya 2021:16–22

Taylor NP, Zappi DC (2004) Cacti of eastern Brazil. Royal Botanic Gardens, Kew

Templeton AR (1981) Mechanisms of speciation-a population genetic approach. Annu Rev Ecol Evol Syst 12:23–48

Templeton AR (2008) The reality and importance of founder speciation in evolution. Bioessays 30:470–479

Thuiller W, Georges D, Engler R, Breiner F, Georges MD, Thuiller CW (2016) Package 'biomod2'. Species distribution modeling within an ensemble forecasting framework. https://cran.r-project.org/web/packages/biomod2/index.html

Varela S, Anderson RP, García-Valdés R, Fernández-González F (2014) Environmental filters reduce the effects of sampling bias and improve predictions of ecological niche models. Ecography 37:1084–1091

Vital H (2014) The north and northeast Brazilian tropical shelves. In: Chiocci FL, Chivas AR (Eds.) Continental shelves of the world: Their evolution during the last glacio-eustatic cycle. Geological Society, Memoirs, London, p 35–46

Wang T, McFarlane HE, Persson S (2016) The impact of abiotic factors on cellulose synthesis. J Exp Bot 67:543–552

Weber MM, Stevens RD, Diniz-Filho JAF, Grelle CEV (2017) Is there a correlation between abundance and environmental suitability derived from ecological niche modelling? A meta-analysis. Ecography 40:817–828

Weigelt P, Steinbauer MJ, Cabral JS, Kreft H (2016) Late Quaternary climate change shapes island biodiversity. Nature 532:99–102

Whittaker RJ, Fernández-Palacios JM, Matthews TJ, Borregaard MK, Triantis KA (2017) Island biogeography: taking the long view of nature's laboratories. Science 357:6354

Zhang C, Rabiee M, Sayyari E, Mirarab S (2018) ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinforma 19:15–30

Zhang L, Li M, Yan P, Fu J, Zhang L, Li X et al. (2021) A novel adenylate isopentenyltransferase 5 regulates shoot branching via the ATTTA motif in Camellia sinensis. BMC Plant Biol 21:1–14

Zhao YJ, Yin GS, Pan YZ, Tian B, Gong X (2020) Climatic Refugia and Geographical Isolation Contribute to the Speciation and Genetic Divergence in Himalayan-Hengduan Tree Peonies (*Paeonia delavayi* and *Paeonia ludlowii*). Front Genet 11:595334

Zulfiqar F, Ashraf M (2021) Bioregulators: unlocking their potential role in regulation of the plant oxidative defense system. Plant Mol Biol 105:11–41

Zhbannikov IY, Hunter SS, Foster JA, Settles ML (2017) SeqyClean: a pipeline for high-throughput sequence data preprocessing. In: Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics (ACM-BCB '17). ACM, New York, NY, USA, p 407–416. https://doi.org/10.1145/3107411.3107446

AUTHOR CONTRIBUTIONS
FFF, DTA and DE conceived the idea. DE, DTA, FFF, JM and IASB performed data collection and analyses. FFF and DE led the writing. DZ, EMM, NT, DTA and IASB contributed to many conceptions and writing. All authors contributed to the intellectual development of the paper, made multiple revisions, and approved the final draft. Funding acquisition, FFF.

COMPETING INTERESTS
The authors declare no competing interests.

ETHICS APPROVAL
Sampling was carried out in accordance with Brazilian law through special permits provided to FFF (permit from Chico Mendes Biodiversity Conservation Institute, SISBIO n° 38141-1).

ADDITIONAL INFORMATION
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41437-024-00683-4.