Geography, climate and shifts in host plants distribution explain the genomic variation in the cactus moth

Daniel Poveda-Martínez 1,2*, Víctor Noguerales 3, Stephen D. Hight 4, Guillermo Logarzo 1, Brent C. Emerson 3, Laura Varone 1† and Esteban Hasson 2†

1Fundación para el Estudio de Especies Invasivas (FuEDEI), Hurlingham, Argentina, 2Instituto de Ecología Genética y Evolución de Buenos Aires (IEGEBA–CONICET), Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires, Argentina, 3Island Ecology and Evolution Research Group, Instituto de Productos Naturales y Agrobiología (IPNA–CSIC), San Cristóbal de La Laguna, Canary Islands, Spain, 4The U.S. Department of Agriculture – Agricultural Research Service (USDA-ARS), Insect Behavior and Biocontrol Research Unit (IBBRU), Tallahassee, FL, United States

Landscape heterogeneity and the host plant use are factors suggested to play determinant roles in shaping the evolutionary history of herbivorous insects. However, the role of the reconfiguration of host plants distributions linked to Quaternary climate oscillations as drivers of contemporary population genetic structure is still poorly understood. Here, we formally examine the relative contribution of such factors on intraspecific diversification using the South American cactus moth, Cactoblastis cactorum, an herbivore insect specialized in the use of cacti as host plants. We assessed genomic variation using genome-wide SNPs and mitochondrial data in populations sampled across a broad geographical gradient where moths feed on different cactus species. We integrated demographic simulations and ecological niche modeling into a landscape genomics framework, to test alternative hypotheses of past and current population connectivity for both C. cactorum and its host plants. Regions exhibiting higher genomic diversity were evaluated for congruence with areas where suitable climatic conditions remained stable through time. Our results revealed that past spatial configuration of suitable habitat conditions and shifts of host plants distributions are the factors that better explain the intraspecific diversification. Genomic data also supported the hypothesis that areas of long-term habitat stability served as refugia for C. cactorum, enabling the maintenance of high levels of genetic diversity over time. Overall, our study highlights the importance of integrating inter-specific interactions and their spatio-temporal dynamics to better understand the relative importance of abiotic and biotic factors driving the diversification processes in herbivorous insects with broad geographical and restricted host ranges.

KEYWORDS
cactus pest, cactoblastis cactorum, ddRAD, landscape genomics, paleoclimate, opuntia
1 Introduction

The spatial distribution of genomic variation in natural populations is influenced by a variety of processes, including the dispersal ability of individuals, gene flow, biotic interactions, and landscape features. Unveiling the relative roles of these processes is necessary to understand how organisms interact with the environment and how the latter shapes their genomic makeup. In this matter, herbivorous insects distributed along environmental gradients feeding on a set of host plants offer the opportunity of disentangling the ecological and geographical processes driving population differentiation and gene flow dynamics in heterogeneous landscapes (Borer et al., 2012; Laukkonen et al., 2014).

Among the ecological factors, the use of alternative host plants is thought to shape patterns of gene flow among populations in herbivorous insects (Funk et al., 2006; Forbes et al., 2017). Moreover, changes in host plant use together with adaptation to different environmental conditions can lead to ecological specialization and diversification (isolation by environment,IBE; Wang and Bradburd, 2014). Geography (isolation by distance, IBD; Slatkin, 1993) and the spatial distribution of suitable habitats (isolation by resistance, IRR; McRae, 2006; McRae and Beier, 2007) may also be determinants of population genetic differentiation across the landscape (Peterman et al., 2014; Vidal et al., 2019).

Insects with narrow diet breadths and environmental requirements are expected to exhibit marked genetic structure as a consequence of gene flow disruption between populations exploiting alternative hosts (Vidal and Murphy, 2018) and/or inhabiting a fragmented landscape of suitable habitats (Ortego et al., 2021). However, in widely distributed insects feeding on a narrow array of host plants across the landscape, contemporary patterns of genomic variation can also be influenced by a range of factors, including dispersal ability, patterns of host plant use and/or components of landscape heterogeneity. Moreover, less studied spatio-ecological factors, such as distributional shifts of host plants driven by climate oscillations throughout the Quaternary period, may have also promoted cycles of geographical isolation and secondary contact, shaping contemporary patterns of genetic variation in herbivorous insects (Noguerales et al., 2018). Yet, the relative contribution of these ecological factors and spatio-temporal landscape dynamics in driving populations differentiation in taxa with broad climatic niches and narrow feeding requirements are still open questions in evolutionary biology. To understand the role of such spatial-ecological processes in species diversification and shaping gene flow in heterogeneous landscapes, an integrative approach is necessary, taking into account historical and contemporary environmental data, combined with a suitable insect model with narrow trophic requirements and distributed along environmental gradients.

Here, we investigate patterns of genomic variation in the cactus moth, Cactoblastis cactorum (Berg) (Lepidoptera: Pyralidae), a southern South American herbivore insect, that inhabits heterogeneous landscapes, faces a variety of climate conditions, and feeds on several species of the genus Opuntia (Cactaceae, Opuntioideae). This moth can be found across wide latitudinal and longitudinal gradients, from the arid and semi-arid lands of northwestern and northeastern Chaco, to the grasslands of the Pampa, and north of the Patagonian steppe (McFadyen, 1985; Morrone, 2014; Varone et al., 2014). Along such broad distribution, C. cactorum feeds on native species, such as O. anacantha Speg., O. bonaerensis Speg., O. elata Link & Otto ex Salm-Dyck, O. megapotamica Arechav., O. penicilligera Speg., O. quimilo K. Schum., and O. rioplatense Font., and also the cultivated O. ficus-indica (L.) Miller, an exotic species in the region (Varone et al., 2012; Varone et al., 2014).

Soon after its introduction the prickly pear O. ficus-indica was quickly adopted as a crop in southern South America (Ervin, 2012). The availability of this new host plant represents a relatively recent expansion of host use for C. cactorum. Currently this moth is a serious threat to the regional industry of prickly pear fruit production (Ochoa et al., 2007). The introduction of the new host led to the suggestion that the spread of O. ficus-indica promoted the dispersion and geographical range expansion of C. cactorum (Varone et al., 2014).

Previous studies revealed that populations of C. cactorum are genetically differentiated in at least four haplogroups correlated with geography and variable levels of genetic diversity among native and non-native moth populations (Marsico et al., 2011), and that patterns of coloration of mature larvae and host plant use vary across regions (Brooks et al., 2012). However, only limited conclusions can be drawn from these studies since (i) only a single mitochondrial genetic marker was used; (ii) sampling did not cover the entire species range; and (iii) the relative abundance of host species was not considered (McFadyen, 1985; Marsico et al., 2011; Brooks et al., 2012). More recently, an extensive survey evaluating host plant preferences across the broad range of C. cactorum concluded that patterns of host use reflect variation in host availability rather than female preference (Varone et al., 2014).

In this study, we use high-throughput nuclear sequence data and Sanger sequence data of mitochondrial DNA to investigate the processes that shaped patterns of diversification in the cactus moth, C. cactorum. Our sampling strategy extended over a broad geographic range taking care of including the diversity of native and non-native hosts. We combine inferences from phylogenomics and demographic modeling to uncover the tempo and mode of diversification, patterns of gene flow and population genetic structure. By integrating ecological niche modeling into a landscape genomic framework, we test a comprehensive suite of competing scenarios of population isolation. Specifically we evaluate the hypotheses that contemporary spatial patterns of population genetic differentiation in C. cactorum are predominantly a consequence of the spatial reconfiguration of suitable areas and shifts in the distribution of host plants caused by Quaternary climatic oscillations whilst the current pattern of host plant use and topoclimatic factors are less influential in its diversification. Since we estimated historical and contemporary habitat suitability, and long-term suitable environments tend to support larger effective population sizes (Carnaival et al., 2009; Soley-Guardia et al., 2019), we also address whether regions where suitable habitat conditions remained stable since the last glacial maximum (LGM, ca. 21 ka) to the present exhibit greater genetic diversity than populations thriving in less stable areas. Finally, we evaluated whether the expansion of O. ficus-indica facilitated
contemporary gene flow among geographically distant populations. In this matter, our expectation is to find footprints of genetic admixture among distant regions, especially among areas highly impacted by prickly pear agroindustry.

2 Methods

2.1 Population sampling and DNA extraction

Between 2018 and 2019, we collected 140 individuals of *C. cactorum* from 28 sampling sites spanning a broad range of its native distribution in southern South America (Figure 1; Table S1). Individuals were collected directly from host plants, including seven native taxa, *O. anacantha*, *O. bonaerensis*, *O. elata*, *O. megapotamica*, *O. penicilligera*, *O. rioplatense*, and *O. quimilo*, and the exotic *O. ficus-indica*. To avoid the inclusion of siblings, only a single individual per sampled plant and eggstick was used for genetic analysis. Before DNA extraction, individual larvae were dissected to remove any food residue, and ~30 mg of clean tissue was used for DNA extraction with Qiagen DNeasy Blood & Tissue Kit following manufacturer’s instructions (Qiagen, Inc.).

2.2 Genomic libraries preparation, filtering and assembling

Individuals were processed into two genomic libraries following the double digest restriction-site Associated DNA (ddRAD) procedure described in Peterson et al. (2012). Genomic DNA was fragmented using *NspI* and *MboI* restriction enzymes, and then purified and ligated to barcoded adapters. Libraries were sequenced in paired-end 125 bp mode on a HiSeq2500 Illumina instrument.

Demultiplexed raw reads were checked, compiled and summarized on MultiQC v1.10.1 (Ewels et al., 2016). Reads were assembled with IPYRAD v.0.9.59 (Eaton and Overcast, 2020). A reference based assembly method was implemented for mapping the filtered reads against the *C. cactorum* reference genome. DNA extracted from an adult male (clean DNA, free of food contaminants can be obtained from males, as they do not feed) was used for library preparation and sequenced in an Illumina NovaSeq 6000 platform. The de novo assembly was accomplished with CLC Genomics Workbench v7.1 (see Poveda-Martínez et al., 2022 for details). Thus, the reference genome served also as a filter for exogenous reads derived from larval food.

In subsequent IPYRAD assembly steps, we allowed 20% of SNPs per RAD locus, and shared heterozygous sites occurring across a maximum of 25% of samples. A minimum of 90% was set for the minimum number of samples scored per locus. We used vcftools v1.15 (Danecek et al., 2011) to remove SNPs with minimum allele frequency lower than 3%, missing data per site across individuals exceeding 25%, and to keep SNPs with read depths between 6x and 100x. Individuals with more than 20% missing data were excluded from further analyses. To prune SNPs in linkage disequilibrium, PLINK v1.9 (Purcell et al., 2007) was used with a window size of 50 bp, window shift of 5 and VIF threshold of 2. SNPs under selection were identified using Bayescan v.2.1 (Foll and Gaggiotti, 2008) and excluded from the dataset. A fragment of the mitochondrial gene encoding the Cytochrome oxidase subunit I

![FIGURE 1](image-url)

**FIGURE 1**

Geographic location and host species of the *Cactoblastis cactorum* populations sampled in this study (A). Regions represent biogeographic provinces according to Morrone et al. (2014). Panels (B, C) represent the maps of habitat suitability for *C. cactorum* during the present (B) and during the Last Glacial Maximum (LGM, ca. 21ky) (C) as inferred by the most-supported ENM model in MAXENT.
(COI) (mtDNA) was amplified using primers C1-J-2183 and PatII. Amplifications, PCR-product purification and sequencing were performed as detailed in Poveda-Martinez et al. (2022).

2.3 Assessing genomic population structure

Population genetic structure was inferred using genome-wide SNP data (nDNA hereafter) and the sparse non-negative matrix factorization approach (sNMF) (Frichot et al., 2014), as implemented in R using the LEA package (Frichot and François, 2015). The number of genetic clusters that best described our data was assessed with 100 repetitions for each possible K value (from K = 1 to K = 10) and using the cross-entropy criterion. In addition to sNMF, genetic structure was approximated by using the major axes of genomic variation obtained from a Principal Component Analysis (PCA), as implemented in the hierfstat (Goudet, 2005) R package. We used mtDNA data to construct a haplotype network using the Median Joining algorithm (Randel et al., 1999) as implemented in PopArt v.1.7.1 (Leigh and Bryant, 2015).

2.4 Phylogenetic analyses

Phylogenetic relationships were reconstructed among the main genetic groups as inferred in sNMF using nDNA data and the coalescent-based method implemented in SNAPP (Bryant et al., 2012). Due to computational burden, SNAPP analyses were run including only five individuals per genetic cluster, selecting those with an ancestry coefficient higher than 0.85, according to sNMF, and individuals that had the lowest proportion of missing data. One SNP per locus was randomly selected resulting in a new matrix of 5,000,000 MCMC generations, sampling every 1000 generations in SNAPP for U and V equal to one, and the analysis was run with 50 replicates, considering 100,000–250,000 simulations for the calculation of the composite likelihood, 10–40 expectation–conditional maximization (ECM) cycles, and a stopping criterion of 0.001. Once maximum likelihood was estimated per run, the best demographic model was selected according to Akaikes information criterion (AIC). AIC values were rescaled in terms of AIC differences (Δi) according to the formula: Δi = AICi – AICmin. A model with a ΔAIC value of 0 and the highest AIC weight (ωi) served as the best model. A parametric bootstrapping approach was used to construct 95% confidence intervals of the estimated parameters running 100 bootstrap replicates using initialized values from the best model (Excoffier et al., 2013).

2.5 Demographic history

A coalescent-based simulation approach was implemented in Fastsimcoal2 (Excoffier et al., 2013) based on the site frequency spectrum (SFS) of nDNA to further investigate the demographic history of the three main lineages identified in clustering analysis and species tree inference (Center, East and South, see results section). Initially, we decided to investigate whether the model selection criteria resulted in the evaluation of different topologies in Fastsimcoal2 and SNAPP. For Fastsimcoal2, six models were considered (models 1–6), while SNAPP considered the same three scenarios of symmetric gene flow among lineages while models 7–9 considered asymmetric gene flow among lineages. For these analyses, 20 individuals were selected with the highest ancestry coefficient (>0.85) according to sNMF (e.g., Ortego et al., 2021). The folded joint SFS was calculated considering a single SNP per locus using the Python script written by Isaac Overcast and available at GitHub (https://github.com/isaacovercast/easySFS). To remove missing data and minimize errors with allele frequency estimates, each genetic cluster was downsampled to 16 individuals yielding a total of 2,565 variable sites. Assuming that invariable sites were not considered in the SFS calculation, we used the “removeZeroSFS” option in Fastsimcoal2 and fixed the effective population size of one of the demes (South lineage, NE_SOUTH = 824,626) to enable the estimation of other parameters with Fastsimcoal2. NE_SOUTH was calculated according to the equation NE = π/4u (Lynch and Conery, 2003). Nucleotide diversity (π = 0.004) was estimated using variant and invariant sites with DNAsp v6 (Rozas et al., 2017). The mutation rate (μ) was considered to be 2.9 × 10⁻⁹ substitutions per site per generation, as previously estimated for the butterfly Heliconius melpomene (Keightley et al., 2015) (Lepidoptera: Nymphalidae). Each model was run with 50 replicates, considering 100,000–250,000 simulations for the calculation of the composite likelihood, 10–40 expectation–conditional maximization (ECM) cycles, and a stopping criterion of 0.001. One maximum likelihood was estimated per run, the best demographic model was selected according to Akaikes information criterion (AIC). AIC values were rescaled in terms of AIC differences (Δi) according to the formula: Δi = AICi – AICmin. A model with a ΔAIC value of 0 and the highest AIC weight (ωi) served as the best model. A parametric bootstrapping approach was used to construct 95% confidence intervals of the estimated parameters running 100 bootstrap replicates using initialized values from the best model (Excoffier et al., 2013).

2.6 Environmental niche modeling

An environmental niche modeling approach was implemented to predict contemporary and historical geographic distributions of suitable areas for the cactus moth in its native range. Environmental niche models (ENMs) were used to determine the impact of historical and contemporary components of landscape heterogeneity on C. cactorum diversification as is detailed below in the landscape genomic analysis section. ENMs were built using MaxEnt v.3.4.1 (Phillips et al., 2006) and the bioclimatic data available in CHELSA v1.2 database (https://chelsa-climate.org/bioclim/). Model parameters in MaxEnt were selected and optimized using the kuenm (Cobos et al., 2019) R package. For both temporal bioclimatic conditions, present-day and Last Glacial Maximum (LGM, ca. 21 kya), we evaluated 15 environmental variables retrieved at 30 arc-sec (~1 km) of resolution. Four bioclimatic variables (Bio 8, 9, 18 and 19) were discarded for
having artificial breaks (Oliveira et al., 2020). Highly correlated variables (R > 0.9) according to variance inflation factor criterion were excluded for downstream analyses resulting in a final dataset of six bioclimatic variables (Bio 2, 3, 5, 13, 14 and 15). Suitability maps during the LGM were obtained by projecting the present-day ENM onto LGM bioclimatic conditions derived from the Community Climate System Model (CCSM4; Braconnot et al., 2007) resulting in two suitability maps based on bioclimatic variables data (Set 1, Climatic$_{CURR}$ and Climatic$_{LGM}$ from hereafter).

Additionally, we constructed ENMs for five of the seven native host species of C. cactorum considering the aforementioned approach for variable selection and model building. ENM for O. penicilligera and O. bonaerensis could not be implemented due to limited occurrence data. The resulting ENMs obtained for current and LGM conditions from each host species were used as variables to build an additional ENM for C. cactorum (Set 2, Host$_{CURR}$ and Host$_{LGM}$ from hereafter). A third ENM was constructed for C. cactorum considering the input of both the bioclimatic variables (Set 1) and the host plant ENMs (Set 2) (Set 3; Climatic-Host$_{CURR}$ and Climatic-Host$_{LGM}$ from hereafter).

Occurrence data for the host species and C. cactorum were obtained from field surveys made from 2007 to 2019 and complemented, after detailed curation, with distribution information available at GBIF (www.gbif.org). Redundant occurrences (e.g. points occurring within 1,500 km$^2$) were excluded using spThin (Aiello-Lammens et al., 2015) R package. After thinning occurrence data, 81 records for C. cactorum locations and their associated occurrence of 205 host species remained [O. quimilo (50), O. megapotamica (47), O. rioplatense (39), O. elata (38), and O. anacantha (31)] and were used to conduct ENMs. Model performance for each scenario was evaluated independently based on statistical significance (Partial ROC), omission rates (OR), and the Akaike information criterion corrected for small sample sizes (AICc) using the kuenm (Cobos et al., 2019) R package.

### 2.7 Landscape genomic analyses

A landscape genomic approach was implemented to study potential factors that could explain patterns of genetic differentiation within C. cactorum. As a measure of genetic differentiation, pairwise F$_{ST}$ estimates were derived from genome-wide SNP data using the Weir and Cockerham (1984) method with the STAMPP (Pemberton et al., 2013) R package and 9,999 bootstrap replicates. We evaluated several plausible scenarios of population connectivity based on historical and contemporary spatial and ecological data.

Isolation by resistance scenarios (IBR) were tested by calculating resistance surfaces based on the suitability maps obtained from ENMs for present-day and LGM conditions considering different subsets of factors: (i) only climatic variables (Climatic$_{CURR}$ and Climatic$_{LGM}$); (ii) climate-based habitat suitability maps considering the host species distribution (Host$_{CURR}$ and Host$_{LGM}$); and (iii) combining climatic variables and climate-based habitat suitability maps for host species distribution (Climatic-Host$_{CURR}$ and Climatic-Host$_{LGM}$). Resistance distances for all pairs of populations were calculated using an eight-neighbor cell connection scheme in Circuitscape v.5 (Hall et al., 2021) through Julia v.1.5.2 (https://julialang.org/). We also calculated resistance distances based on a “flat landscape” where all cells have an equal resistance value (=1) representing a null model of isolation by resistance (IBR$_{NULL}$).

An isolation by distance scenario was also evaluated using weighted topographic distances (IBD$_{WTLD}$), which incorporate an additional overland distance covered by an organism due to changes in elevation imposed by the topography. The IBD$_{WTLD}$ distance matrix was calculated on a digital elevation model (DEM) at ~1 km of resolution retrieved from WorldClim v.2.1 (Fick and Hijmans, 2017) dataset using the TopoWeightedDist function implemented in the topoDistance (Fick and Hijmans, 2017; Wang, 2020) R package. We assumed a linear function to weight aspect changes (hFunction parameter) and an exponential function to weight the slope (vFunction parameter) (e.g. Nogueras et al., 2021).

Climatic dissimilarities were estimated between populations to evaluate an isolation by environment scenario (IBE$_{CLI}$). Environmental data was extracted from the 15 CHELSA bioclimatic variables for each of the 28 sampling sites, as well as from 500 random points covering our study area to avoid potential biases resulting from only considering conditions at focal sites. Due to collinearity among the bioclimatic variables, we ran a PCA using the ade4 R package and summarized the environmental variation in the three first axes accounting for ca. 82% (PC1 = 43.79%; PC2 = 23.96%; PC3 = 14.31%) of total variation. The environmental dissimilarity matrix was obtained by calculating the Euclidean distances for PC scores between pairs of sampling sites (Ortego et al., 2021).

To test the relative role of host use in population genetic structure in C. cactorum, a host plant distance matrix (IBH) was built considering the host species where moths were collected. We constructed a binary matrix coded with each host species per sampling site, and ran a PCA using the ade4 R package to build a distance matrix with total PCs variation. IBH was obtained by calculating the Euclidean distances for PC scores between pairs of sampling sites.

Relationships between explanatory distance matrices based on landscape heterogeneity (IBRs, IBD$_{WTLD}$, IBE$_{CLI}$), and the contemporary patterns of host use (IBH) with the genetic differentiation between C. cactorum populations (F$_{ST}$) were evaluated using univariate and multiple matrix regressions with randomization using the function MMRR (Wang, 2013) as implemented in R. An initial full model was constructed considering all significant explanatory terms identified previously in univariate analysis, and a final best-fit model was selected using a backward-stepwise procedure by progressively removing non significant variables until all retained terms within the model were significant (Wang, 2013; Ortego et al., 2014). The result was the minimal most adequate model for explaining the variability in the response variable, where only the significant explanatory terms were retained.
2.8 Population genetic diversity and climate/habitat stability

Given that suitable environments tend to support larger effective population sizes (Carnaval et al., 2009), we tested whether the regions exhibiting high genetic diversity were those where suitable climatic conditions remained stable through time, from the LGM to the present. To test this hypothesis, environmental stability maps were constructed by averaging current and LGM suitability maps following Soley-Guardia et al. (2019). For each sampling site, values of climate/habitat stability were extracted for each of the three scenarios considering (i) only climatic variables (StabilityENV), (ii) climate-based habitat suitability maps for host species (StabilityHOST), and (iii) combining climatic variables and climate-based habitat suitability maps for host species (StabilityENV-HOST). Raster calculations were conducted using the raster R package.

Nuclear and mitochondrial diversity were characterized for each of the 28 sampling sites. Genetic diversity estimates were normalized to three individuals per sampling location to avoid potential bias resulting from uneven population sample size (range: 3–11 individuals). For nDNA, expected heterozygosities (H_e) and nucleotide diversity (π) were calculated using the diveRsity (Keenan et al., 2013) R package, and DNAsp v6 (Rozas et al., 2017), respectively. We also estimated π and haplotype diversity (H_D) for mtDNA data using DNAsp. Correlations between stability values and estimates of nuclear (H_e, π) and mtDNA (H_D, π) population genetic diversity were tested using linear regressions. Longitude and latitude were also included as explanatory factors to account for potential geographical clines of genetic diversity (Guo, 2012).

3 Results

3.1 Genomic data

We successfully genotyped 138 individuals of C. cactorum using ddRAD sequencing which are representative of populations feeding on both exotic O. ficus-indica (71 individuals), and native host plants O. megapotamica (34), O. rioplatense (11), O. elata (10), O. bonaerensis (4), O. penicilligera (3), O. quimilo (3), and O. anacantha (2) along three biogeographic regions. After filtering steps, the average number of paired-end reads retained per individual was 1,859,116, of which an average of 1,253,277 reads mapped to the C. cactorum reference genome. After discarding loci in linkage disequilibrium and under selection, we recovered a total of 3,506 biallelic SNPs with an average coverage of 30x (Table S2). A total of 143 individuals from the same 28 locations were successfully sequenced for a fragment of 790 bp of the COI gene. The analysis of mtDNA data revealed 66 haplotypes, 171 variable sites, 97 of which were parsimony-informative sample-wide.

3.2 Assessing population genomic structure

Clustering analysis with sNMF indicated K = 6 as the most likely number of genetic clusters. While all genetic groups were spatially structured (North, Central, Western, Southeast, Southwest and East), no genetic structure based on host plants was observed. Population genetic structure analyses also revealed a certain degree of admixture among geographic groups; some Northern and Central individuals exhibited a relatively high degree of admixture with both the Western and Eastern clusters (Figures 2A–C). Interestingly, Northern, Central and Western populations appeared to form a unique cluster (Central lineage from hereafter) when assuming K = 3, whereas both Southern (South lineage from hereafter) and Eastern (East lineage from hereafter) populations remained as separated panmictic groups. With K = 4, the Southern population appeared subdivided into two geographic units; Southeastern and Southwestern. Finally, with K = 5, Northern populations appeared separated from the Central ones (Figure S2).

In the PCA the Eastern lineage appears separated from the Central and South lineages along PC1, whereas the South lineage can be distinguished from the Central lineage along PC2 (Figure 2D).

The median-joining network obtained with mtDNA sequence data revealed strong spatial genetic structure consistent with the clusters identified with nDNA data (Figure 2E). The Central haplogroup identified with a star-like pattern had one of the most frequent haplotypes and was separated by seven mutational steps from the Eastern haplogroup, whereas at least four mutational steps separated the Central and Northern haplogroups. Western, southeastern and Southwestern haplogroups appeared close to the Central haplogroup separated by few mutational steps (Figure 2F).

3.3 Species tree reconstruction

Phylogenetic relationships inferred with SNAPP were consistent with the hierarchical spatial genetic structure observed with sNMF (Figures 2B, C; Figure S2). The most ancestral split corresponded to the separation of South and East lineages from the Central lineage. Southwestern and Southeastern clusters, as well as West, Central and North clusters diverged subsequently from their respective lineages, as observed in clustering analyses.

3.4 Demographic inference using coalescent-based simulations

The scenario considering full asymmetric interlineage gene flow (model 8) was the most supported among the nine demographic models tested with Fastsimcoal2 (Table S3). Under this model, divergence among the three main lineages of C. cactorum was estimated to have occurred during the Late Pleistocene (Figure 3; Table 1). Specifically, this model involves an initial split (T_DIV1) of the Central lineage ~75 kya (considering two generations per year in native host species, as reported by Varone et al., 2014), and a more recent split (T_DIV2) separating the South and East lineages occurring ~19 kya. Demographic simulations estimated an effective population size close to 250,000 for the Central lineage (N_E.CENTER), which was six times higher than the value obtained for the East lineage (N_E.EAST = 41,551). Estimates of gene flow rates varied among lineages with the highest rate of contemporary gene...
FIGURE 3
Visual representation of the most supported demographic model (Model 8) estimated using Fastsimcoal2. Parameter estimates included timing of population divergence (TDIV1 and TDIV2), historical and contemporary effective population sizes (NEANC1 and NEANC2), and asymmetric rates of gene flow (MSC, MCE, MES, MEC, MSE, and MEC).

FIGURE 2
Population genetic structure and phylogenetic relationships of the C. cactorum populations. (A) Pie charts represent the average ancestry coefficient of individuals belonging to each of 28 sampling sites distributed along three biogeographic regions. (B) Species tree reconstructed by SNAPP; numbers in nodes denote posterior probability. (C) Results of individual assignment in genetic clusters using sNMF. Vertical bars represent the ancestry coefficient of each individual to the corresponding genetic cluster. (D) PCA using the two major axes of genomic variation. (E) Median-joining network obtained with mtDNA data. Image shows an adult male of C. cactorum.
flow estimated to occur from the Central to the East lineage \( (M_{CE} = 1.50 \times 10^{-5} \text{ migration rate by generation}) \) and the lowest from the Central lineage to the South lineages \( (M_{CS} = 2.71 \times 10^{-8} \text{ migration rate by generation}) \) (Table 1).

### 3.5 Ecological niche modeling through time

Ecological niche models (ENMs) for the focal species (C. cactorum) and host species exhibited an overall good performance as suggested by the relatively high AUC scores (Tables S4, S5). Suitability maps for host species were concordant with their respective current distributions (Figure S3). Among host species, the areas of higher suitability for *Opuntia megapotamica* were predicted in central and southern Argentina; for *O. elata* in central-eastern, and for *O. rioplatense* in eastern Argentina (in the area of influence of the Paraná-La Plata river basin). *Opuntia anacantha* and *O. quimilo* have more restricted suitability areas in northwestern and central Argentina, respectively. Projections during the LGM indicated changes in host distribution (Figure S3) that suggest an increased habitat suitability for almost all host plants, except for *O. anacantha* whose suitability areas have been restricted to the central-eastern region during the LGM. Instead, habitat suitability for *O. megapotamica* and *O. elata* appears extended into the northern region, whereas the projections indicate an extended distribution in central Argentina for *O. quimilo*. Finally, suitable habitats for *O. rioplatense* extended to northern and central Argentina.

Models for *C. cactorum* (Set 1: ClimaticCURR, Set 2: HostCURR, Set 3: Climatic-HostCURR) indicated similar predicted distributions (Figure S4). However, the most supported model for *C. cactorum* according to AIC values was one that considered as input variables the predicted distribution of the host plants (HostCURR). Additional information on the results of ENMs for both *C. cactorum* and host species are detailed in Tables S4, S5. The present-day distribution inferred by the most supported scenario (HostCURR) was in line with the contemporary range of the moth. Under this model, higher suitability areas were predicted in both Chaco and north of Pampa biogeographic regions whereas areas of lower suitability were predicted in the southern portions of the Pampas and the Monte biogeographic regions (Figure 1B). Other biogeographic provinces (Yungas, Puna and Prepuna) had extremely low suitability values for *C. cactorum*. Projections of ENM to LGM predicted high environmental suitability along the Chaco biogeographic province to the most southern portion of Pampas, with areas of low suitability located in eastern Argentina, a confluence zone of La Plata river basin (Figure 1C).

### 3.6 Landscape genomic analyses

Estimates of genetic differentiation \( (F_{ST}) \) among sampling sites ranged from 0.023 to 0.448 for nDNA data (Table S6). Univariate matrix regressions indicated that nuclear genetic differentiation was significantly correlated with all distance matrices except with the current pattern of host plant use (IBH) consistent with clustering analysis, and both explanatory terms of landscape heterogeneity: ClimaticCURR and Climatic-HostCURR (Table S7). Yet, only Climatic-HostLGM was significantly retained in the best-fit model in the backward-stepwise procedure (Table 2).

### 3.7 Population genetic diversity and climate/habitat stability

We found higher nuclear and mitochondrial diversity in Northern, Central and Eastern populations, than in Southern

---

**TABLE 1 Parameters inferred from coalescent simulations with FastSimcoal2 under the best-supported demographic model (Model 8; Figure 5).**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Point estimate</th>
<th>95% Confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lower bound</td>
<td>Upper bound</td>
</tr>
<tr>
<td><strong>NEANCE</strong></td>
<td>3,504,318</td>
<td>2,576,620 3,858,949</td>
</tr>
<tr>
<td><strong>NEANC2</strong></td>
<td>11,746</td>
<td>3,773 21,102</td>
</tr>
<tr>
<td><strong>NECENTER</strong></td>
<td>245,957</td>
<td>208,570 380,730</td>
</tr>
<tr>
<td><strong>NEEAST</strong></td>
<td>41,551</td>
<td>35,596 54,151</td>
</tr>
<tr>
<td><strong>TDIV1</strong></td>
<td>75,513</td>
<td>61,524 120,375</td>
</tr>
<tr>
<td><strong>TDIV2</strong></td>
<td>19,207</td>
<td>17,202 23,310</td>
</tr>
<tr>
<td><strong>MCE</strong></td>
<td>1.18 ( \times 10^{-6} )</td>
<td>8.86 ( \times 10^{-7} ) 1.75 ( \times 10^{-6} )</td>
</tr>
<tr>
<td><strong>MSE</strong></td>
<td>4.68 ( \times 10^{-8} )</td>
<td>5.41 ( \times 10^{-11} ) 1.65 ( \times 10^{-8} )</td>
</tr>
<tr>
<td><strong>MCS</strong></td>
<td>2.71 ( \times 10^{-8} )</td>
<td>3.55 ( \times 10^{-11} ) 6.83 ( \times 10^{-8} )</td>
</tr>
<tr>
<td><strong>MSE</strong></td>
<td>6.37 ( \times 10^{-6} )</td>
<td>4.37 ( \times 10^{-6} ) 7.32 ( \times 10^{-8} )</td>
</tr>
<tr>
<td><strong>MCSE</strong></td>
<td>1.50 ( \times 10^{-5} )</td>
<td>1.23 ( \times 10^{-5} ) 1.67 ( \times 10^{-5} )</td>
</tr>
<tr>
<td><strong>MCE</strong></td>
<td>2.98 ( \times 10^{-5} )</td>
<td>3.78 ( \times 10^{-11} ) 4.12 ( \times 10^{-8} )</td>
</tr>
</tbody>
</table>

For each parameter, point estimate and lower and upper 95% confidence intervals are shown. \( T_{DIV} \) and \( T_{DIV2} \), population divergence; \( NEANCE \) and \( NEANC2 \), historical and contemporary effective population sizes; \( MSE, MCE, MSEE, MCE, \) and \( MSEC \), asymmetric rates of gene flow among lineages.
4 Discussion

Nuclear SNPs and mtDNA data revealed significant population structure in *C. cactorum*. Landscape genomic analyses provided support for the hypothesis that habitat connectivity based on a combination of climatic conditions and habitat suitability, influenced by shifts in host species distributions during LGM, was the main factor shaping population genetic structure. Although patterns of host species use appeared to be non-influential on *C. cactorum* diversification, shifts in the distribution of Opuntia hosts, mediated by climatic changes during the Quaternary, had a direct influence on the distribution and genomic variation of the focal species. Such shifts in host distributions may have generated fragmented ranges that led to reductions of gene flow among moth populations, promoting genetic differentiation. Our results also supported the habitat stability hypothesis, whereby populations inhabiting regions of greater habitat suitability during the last 21,000 years harbored more genetic diversity than those that persisted in areas of lower habitat stability. The hypothesis that the geographically widespread cultivation of *O. ficus-indica* facilitated contemporary gene flow among otherwise geographically distant populations can be rejected, on the basis of the results of population genomic analyses revealing limited genetic admixture and a hierarchical pattern of population differentiation mainly concordant with geography.

4.1 Effects of Quaternary landscape composition on populations differentiation

Genetic clustering analyses revealed three major lineages defined by geography: Central, East and South (Figures 2 and 3). Coalescent-based demographic modeling suggested that these lineages began to diverge during the Late Quaternary, approximately 75 kya. The earliest split involved populations of the Chacoan biogeographic domain (hereafter “Chaco”), including populations from North, West and Central Argentina, from the Pampean biogeographic region (hereafter “Pampa”) that includes the South and East lineages (Figures 1, 2). Prior to this split, the regional contraction of subtropical and tropical biomes made the Chaco and the Pampa biomes more alike (Ortiz-Jaureguizar and Cladera, 2006). Subsequently, during the Quaternary, major climatic shifts took place in the area, leading to the isolation of the Chacoan xeric woodlands from the typical grasslands of the Pampean region. Such climatic changes were promoted by the topographic reconfiguration of Andean and sub-Andean Piedmont and the uplift of the eastern orographic systems (known as the Sierras Pampeanas) associated with the Peripampasic orogenic arc (Ortiz-Jaureguizar and Cladera, 2006; Speranza et al., 2007; Calatayud-Mascarell et al., 2022). These events produced a rain-shadow effect that resulted in the extreme xeric conditions that prevailed in this area. Climate changes during the Quaternary, which included cold, dry glacial cycles alternating with warm, moist interglacial periods, affected these regions causing the expansion and contraction of open woodlands and grasslands (Ortiz-Jaureguizar and Cladera, 2006). This dynamic likely promoted changes in floristic composition of such regions, affecting the distributions of host species and consequent habitat suitability for *C. cactorum*, promoting the divergence between the Central lineage (Chaco) and East, South lineages (Pampean) (Figures 2, 3).

The split between East and South lineages was estimated to have occurred approximately 19 kya. Regional climate models suggest that this timing is coincident with an increase in precipitation along the eastern foothills of the Andes (Cook and Vizy, 2006; Ortiz-Jaureguizar and Cladera, 2006). These topographical features together with Quaternary climatic dynamics in the region maintained a fragmented distribution of *C. cactorum* host species from the Pampa (Moureille and Ezcurra, 1997). This likely limited dispersal of *C. cactorum*, restricting the East lineage to the more humid environments with deep fertile soils typical of the Pampa, and the South lineage to the drier grasslands of the northern border of the Patagonian steppe, both regions with diverse patterns of host species composition (Varone et al., 2014).

<table>
<thead>
<tr>
<th>Variable</th>
<th>R²</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>LGMENV-HOST</td>
<td>0.416</td>
<td>16.365</td>
<td>0.001</td>
</tr>
<tr>
<td>Rejected terms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Climatic dissimilarity (IBECLI)</td>
<td>−0.476</td>
<td>0.831</td>
<td></td>
</tr>
<tr>
<td>Topographic distance (IBDWTDD)</td>
<td>−2.058</td>
<td>0.301</td>
<td></td>
</tr>
<tr>
<td>IBR - HostCURR</td>
<td>3.251</td>
<td>0.278</td>
<td></td>
</tr>
<tr>
<td>IBR - ClimaticCURR</td>
<td>4.808</td>
<td>0.117</td>
<td></td>
</tr>
<tr>
<td>IBR - HostLGM</td>
<td>−5.834</td>
<td>0.087</td>
<td></td>
</tr>
<tr>
<td>Flat landscape – IBRNULL</td>
<td>−3.654</td>
<td>0.278</td>
<td></td>
</tr>
</tbody>
</table>

IBR scenarios consist of current and past habitat suitability based on (i) only climatic variables (*ClimaticCURR* and *ClimaticLGM*), (ii) climate-based habitat suitability maps for host species (*HostCURR* and *HostLGM*), and (iii) combining climatic variables climate-based habitat suitability maps for host species (*ClimaticCUMCURR* and *ClimaticCUMLGM*). IBRNULL represents an IBR scenario where all pixel values are equal to 1 (flat landscape). R², regression coefficient; t, t-statistic; p, significance level.

*Table 2. Results of multiple matrix regression with randomization (MMRR) testing for the relationships between population genetic differentiation (FST) and explanatory variables representing weighted topographic distance (isolation by distance, IBDWTDD), climatic dissimilarity (isolation by environment, IBECLI), host plant distance (*IBH*), and alternative isolation by resistance (IBR) scenarios.*
Climatic changes during the Quaternary also promoted complex phylogeographic patterns in other southern South American herbivorous insects, such as the grasshopper Dichroplus vittatus Bruner (Rosetti et al., 2022) and the beetle Naupactus cervinus Boheman (Rodriguero et al., 2016). Such phylogeographic patterns would also be a direct consequence of changes in habitat and host plant distributions during the Quaternary. Moreover, similar phylogeographic consequences of climate change over the last 21,000 years have been recently documented in Opuntia bonaerensis, one of the hosts of C. cactorum (Köhler et al., 2020). Such concordance in the timing of intraspecific diversification within O. bonaerensis and C. cactorum is strongly suggestive of a fundamental role of Quaternary climate on the evolutionary history of the cactus-moth system.

Intraspecific genetic differentiation in C. cactorum and its close relatives C. doddi and C. bucyrus also dates back to the late Quaternary (Poveda-Martínez et al., 2022). Diversification within C. bucyrus, a specialist on the columnar cactus Trichocereus atacamensis, is estimated to have initiated during the Marine Isotopic Stage 3 (~42 kya), an interstadial during the last glacial period. Divergence in C. doddi, which feeds upon Opuntia sulphurea, is estimated to have occurred very close to the end of the Pleistocene (~19 kya), coincident with estimates for the split between East and South lineages of C. cactorum.

### TABLE 3 Results of linear regressions testing for the relationships between population genetic diversity, and habitat stability models during the last 21,000 years, according to ENMs based on (i) only climatic variables (StabilityENV), (ii) climate-based habitat suitability maps for host species (StabilityHOST), and (iii) combining climatic variables climate-based habitat suitability maps for host species (StabilityENV-HOST).

<table>
<thead>
<tr>
<th></th>
<th>StabilityENV</th>
<th>StabilityHOST</th>
<th>StabilityENV-HOST</th>
<th>Latitude</th>
<th>Longitude</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R²</td>
<td>p</td>
<td>R²</td>
<td>p</td>
<td>R²</td>
</tr>
<tr>
<td>nDNA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expected heterozygosity (Hₑ)</td>
<td>0.466</td>
<td>0.000</td>
<td>0.292</td>
<td>0.002</td>
<td>0.245</td>
</tr>
<tr>
<td>Nucleotide diversity (π)</td>
<td>0.509</td>
<td>0.000</td>
<td>0.266</td>
<td>0.004</td>
<td>0.219</td>
</tr>
<tr>
<td>mtDNA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Haplotype diversity (Hₑ)</td>
<td>0.084</td>
<td>0.133</td>
<td>0.021</td>
<td>0.453</td>
<td>0.031</td>
</tr>
<tr>
<td>Nucleotide diversity (π)</td>
<td>0.076</td>
<td>0.154</td>
<td>0.005</td>
<td>0.722</td>
<td>0.008</td>
</tr>
</tbody>
</table>

Correlations with geographical variables (latitude and longitude) were also assessed. Genetic diversity was calculated for both nuclear (nDNA) and mitochondrial (mtDNA) data. R², regression coefficient; p, significance level.
Maps of environmental suitability estimated using ecological modeling suggested that the Chaco and northern Pampean regions harbored larger climatically suitable areas than other biogeographic regions such as the Yungas, Monte, Puna and Prepuna (Figure 1). Both *C. doddii* and *C. bucyrus* are found in the Prepuna and Monte regions where *C. cactorum* is absent, supporting the idea that disparate environmental/habitat conditions and disjunct host plant distributions played a role in divergence within *C. cactorum* and closely related species (Poveda-Martinez et al., 2022). Overall, these findings suggest that intraspecific diversification within this group of closely related herbivorous insects have been strongly influenced by shifts in the distribution of host species in response to Quaternary climate dynamics.

### 4.2 Landscape genomic analysis

In agreement with inferences of divergence times among the three major lineages inferred using coalescent-based demographic modeling, landscape genomic analyses revealed that the spatial pattern of population genetic differentiation can be mainly explained by a Quaternary landscape scenario that takes into account the distribution of climatically suitable habitats and predicted ranges of host species during the LGM (Figures 1–3). The resistance map defined by habitat connectivity based on historical host distribution provided the best fit to genetic data, with no significant additional effects of topography, current environmental suitability or environmental dissimilarity (Table 2). The influence of historical landscape composition on contemporary patterns of genomic variation is illustrated by the fact that weakly differentiated populations had high habitat connectivity in LGM projections. This is the case of JUI (northwestern) and FOG (northeastern) populations (*F_{ST} = 0.088*; Table S6), currently separated by 740 km but with high past habitat connectivity according to suitability maps. In contrast, a discontinuous habitat is expected to promote isolation and, thus, genetic differentiation even between geographically close populations. This would be particularly true for *C. cactorum*, given the limited dispersal ability in a species with a short-lived adult stage (Zimmerman et al., 2000), and illustrated by the southern populations LPS (southwestern) and BAP (southeastern) which exhibit one of the highest *F_{ST}* values (*F_{ST} = 0.389*), despite the relatively shorter geographic distance of approximately 290 km. Projections during LGM suggest that dispersal, and consequently gene flow, between LPS and BAP was likely limited by low environmental suitability during glacial periods. Thus, our results suggest that environmental tolerance along with limited dispersal interacting with landscape features are responsible for current patterns of population genetic structure in the cactus moth (e.g. Broquet and Petit, 2009; Sherpa et al., 2020).

Despite signatures for shifts in host species distributions, which may have influenced contemporary population genetic structure in *C. cactorum*, patterns of host use did not appear to be a determinant factor. This finding is consistent with the idea that the use of alternative hosts in this specialist herbivorous moth does not promote divergence, likely due to relatively low fitness variation across the narrow array of hosts (Vidal and Murphy, 2018; Vidal et al., 2019). This is in line with field and laboratory studies showing that host use was proportional to the abundance of hosts in locations where more than one potential host coexist (Varone et al., 2014). Moreover, multiple choice experiments revealed that *C. cactorum* females do not exhibit oviposition preference for any *Opuntia* species. Altogether these results indicate that host use is not an important selective agent in *C. cactorum*, consistent with evidence in other herbivorous insects (Vidal and Murphy, 2018; Vidal et al., 2019; Tischechkin, 2020).

### 4.3 Climatic stability and patterns of genomic diversity

Regression analyses revealed that population genetic diversity was significantly correlated with habitat stability. The Chaco region, representing the northernmost distribution limit of *C. cactorum*, is an open vegetation biome characterized by high endemity and diversity for both plants and animal species (Wernick, 2011; Brusquetti et al., 2019; Bonatelli et al., 2022). Analyses of environmental habitat suitability between the present and the LGM revealed relative habitat stability within the Chaco (Figure 4). Consistent with this, nuclear genetic diversity in *C. cactorum* was higher in the Chaco as compared to southern regions, supporting the hypothesis that areas characterized by a high climate stability over the last glacial and interglacial periods tend to accumulate and harbor more genetic diversity (Carnaual et al., 2009; Hewitt, 2011; Barros et al., 2015; Rocha et al., 2020). The greater effective population size estimated for the Central lineage, as estimated by coalescent demographic modeling, is in line with the existence of Pleistocene refugia in Chaco. This region also contains a larger number of *Opuntia* species than any other region in the sampled area (Varone et al., 2014). Such host plant availability together with long-term habitat stability has likely benefited the persistence of relatively large *C. cactorum* populations in this region.

### 4.4 Effect of the prickly pear on contemporary gene flow of the cactus pest

Overall, our results argue against the idea that the human-driven introduction and intensive cultivation of *O. ficus-indica* enabled the rapid expansion of *C. cactorum*, promoting contemporary gene flow among geographically distant populations. Population genetic analyses yielded evidence for gene flow between diverging Chacoan and North Pampean lineages but not for current gene flow among distant populations feeding on *O. ficus-indica* and/or alternative native hosts. In fact, we detected nuclear admixture and sharing of mtDNA haplotypes across populations located along the contact area between northern Pampa and Chaco (Figure 2). This area is coincident with a transition between Chaco and Pampa regions, referred to as the Espinal (Bucher, 1982). Historical secondary contact among previously isolated Central and East lineages, prompted by post glacial host range expansions, provides a
plausible explanation for the observed pattern of genetic admixture. Evidence for admixture was also detected between populations of East and South lineages, both distributed within the Pampean and Monte regions. This area has been reported as a landscape corridor facilitating connectivity between previously separated populations of the grasshopper *D. vittatus* inhabiting grassland and savanna biomes (Rosetti et al., 2022), and the southern leaf-cutting black ant (*Sánchez-Restrepo et al.*, 2023). In this respect, our results emphasize the extensive habitat connectivity between these regions during the Quaternary, but not as a product of growing prickly pear agroindustry in the study area.

**Data availability statement**

The original contributions presented in the study are publicly available. This data can be found here: Figshare, https://doi.org/10.6084/m9.figshare.18270866.v1.

**Ethics statement**

The manuscript presents research on animals that do not require ethical approval for their study.

**Author contributions**

DP-M: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Validation, Writing – original draft, Writing – review & editing. VN: Investigation, Methodology, Writing – original draft, Writing – review & editing, Conceptualization, Supervision. SH: Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing. GL: Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing. BE: Conceptualization, Supervision, Validation, Writing – review & editing. LV: Conceptualization, Funding acquisition, Investigation, Project administration, Resources, Supervision, Validation, Writing – review & editing, Data curation, Writing – original draft. EH: Conceptualization, Investigation, Methodology, Project administration, Supervision, Validation, Writing – original draft, Writing – review & editing.

**Funding**

Funding was obtained from FONCyT through grant PICT1447/2016 and USDA APHIS-PPQ, Farm Bill 10201.

**Acknowledgments**

We are grateful to the Associate Editor and two reviewers for suggestions that helped to improve previous versions of this paper. We would like to thank Mariel Guala and Malena Fuentes Corona for fieldwork support. We are also grateful to the Centro de Computo de Alto Rendimiento (CeCAR) and Biocóides S.A. for granting use of computational resources. DP-M is the recipient of a postdoctoral scholarship awarded by CONICET. VN was supported by a Juan de la Cierva-Formación postdoctoral fellowship (grant FJCI2018-035611-I) funded by MCIN/AEI/10.13039/501100011033. LV and EH are members of Carrera del Investigador CONICET.

**Conflict of interest**

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher’s note**

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

**Supplementary material**

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fevo.2023.1260857/full#supplementary-material

**References**


genetic variation across species ranges. Mol. Ecol. Notes


