Biodiversity and Conservation
https://doi.org/10.1007/s10531-023-02627-6

ORIGINAL RESEARCH

Using genome-wide diversity and population structure to define management units in the cirio (*Fouquieria columnaris*), an emblematic tree of the Sonoran Desert

Josué J. Martínez-Noguez[1] · **Francisco J. García-De León**[1]

Received: 15 June 2022 / Revised: 8 April 2023 / Accepted: 14 May 2023

Abstract

An important research goal for plant species of conservation concern is to identify and spatially delimit genetically distinct populations and understand their degree of connectivity. Here, we evaluated structure, diversity, and genetic connectivity using neutral genetic markers in cirio (*Fouquieria columnaris*), a unique succulent tree restricted to the northwestern Sonoran Desert in Mexico. Using genome-wide sequencing, 1,137 biallelic single nucleotide polymorphisms (SNPs) were identified in 55 individuals from 7 sites (6 on the Baja California Peninsula, 1 in coastal Sonora), belonging to populations that were previously shown to constitute distinct lineages according to chloroplast DNA. There was relatively minor cytonuclear discordance, which was attributed to the northernmost peninsular site; when that site was removed, there was concordance between chloroplast and nuclear DNA matrices. We recognize four genetic groups as demographically independent units. Pairwise comparisons of nuclear genetic differentiation between localities (Jost's D) ranged from 0.035 to 0.008. We detected positive inbreeding values and diversity patterns associated with historical bottlenecks and habitat fragmentation. There was asymmetric gene flow from the small Sonoran mainland population to the much larger northern peninsular population, an unexpected result that may be influenced by pollinators. The most strongly differentiated populations based on SNPs were Sierra Bacha on the Sonoran coast with the southernmost peninsular site, Tres Vírgenes. Within the Peninsula there was a statistically significant pattern of isolation by distance. Our findings provide a detailed view of the genetic structure and connectivity of this emblematic tree that may help guide conservation actions.

Keywords *Genetic population structure* · Management units · Single nucleotide polymorphism · Conservation genomics

---

Communicated by Daniel Sanchez Mata.

✉ Francisco J. García-De León
fgarciadl@cibnor.mx

[1] Laboratorio de Genética para la Conservación, Centro de Investigaciones Biológicas del Noroeste (CIBNOR), Calle IPN #195, La Paz BCS 23096, Mexico

Introduction

Cirio (*Fouquieria columnaris* (Kellogg) Kellogg ex Curran) is an endemic non-model species of the northwestern Sonoran Desert and is considered a charismatic regional symbol. The species is distributed in the central desert of the Baja California Peninsula, as well as isolated patches in Sierra de San Francisco, the Tres Vírgenes volcano complex in Baja California Sur, Isla Ángel de la Guarda in the Gulf of California, and Sierra Bacha in Sonora, Mexico (Humphrey 1974; Webb et al. 2014). Cirio seeds are dispersed by wind (Humphrey 1974). Cirios receive floral visitors including bees, arthropods, and hummingbirds (Humphrey and Werner 1964; Nabhan et al. 2000), although their ability to participate as pollinators has not been evaluated. A previous study using chloroplast DNA (cpDNA) sequences defined four different evolutionary lineages with a fragmented distribution, associated with a complex regional geological and climatic history (Martínez-Noguez et al. 2020). However, it is not clear whether this pattern continues to be maintained by contemporary ecological processes. Genetic structure does not depend exclusively on geological or climatic barriers; genetic connectivity and patterns of diversity are also defined by factors such as type of reproduction, longevity, gene flow mediated by the dispersal of pollen and seeds (Duminil et al. 2007; Wessinger 2021). Furthermore, results of these analyses can depend on the type of molecular marker used (Kellogg et al. 1996). For example, the patterns of genetic divergence revealed by nuclear and cytoplasmic markers (such as cpDNA sequences) may not match. This phenomenon, known as cytonuclear discordance, can be caused by processes such as introgression, incomplete lineage sorting or cytonuclear incompatibility (Drouin et al. 2008; Renoult et al. 2009; Rose et al. 2021).

Cirio is an important species from a conservation perspective. Almost the entire peninsular distribution is found within the Mexican natural protected areas Valle de los Cirios Flora and Fauna Protection Area, the El Vizcaíno Biosphere Reserve, the Protected Area of the Gulf of California Islands and the Marine Zone Biosphere Reserve, Bahía de Los Ángeles, Canales Ballenas, and Salsipuedes, all of which are protected by Mexican federal law (Supplementary Information 1). The only locality where cirios are found on the continental massif, Sierra Bacha to the south of Puerto Libertad, is not contained within any type of protected territory, although it has been proposed as a Priority Terrestrial Region for conservation (Arriaga et al. 2000). This species is listed in Appendix II of the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES), where it is recognized as threatened with extinction in the future. It is also listed as a vulnerable species on the International Union for Conservation of Nature (IUCN) Red List of Threatened Species. There is little information available to assess the state of its conservation, but the most recent data indicate that the species is experiencing a significant population decline (Bullock et al. 2005).

The interest in protecting species and improving management of protected areas makes it essential to understand the changes and challenges faced by natural populations in the face of global climate change. New methods of reduced-representation of genomes provide many biallelic molecular markers, known as single nucleotide polymorphisms (SNPs), which provide valuable information on genetic diversity that can help in the management and monitoring of endangered species and their habitats (Hohenlohe et al. 2021). It is important to highlight that genetic diversity is known as one of the fundamental components of biodiversity (United Nations 2011). Under these principles, concepts such as conservation units (CUs) have been developed to conserve biodiversity. CUs defined based on neutral molecular markers represent the first step toward the management and conservation of species (Fraser and Bernatchez 2001, Funk et al. 2012; Barbosa et al. 2018), such that they are also known as Management Units (MUs). The demographic history of MUs is shaped by microevolutionary forces such as mutation, genetic drift, recombination, and migration (Palsbøll et al. 2007; Schwartz et al. 2007). Recognizing genetically distinct plant populations may help preserve the genetic integrity and inform decision-making for conservation actions (Volkmann et al. 2014; Benestan et al. 2016).

In this study, we use reduced-representation genome sequencing to evaluate cytonuclear discordance and determine whether the four evolutionary lineages previously defined based on cpDNA in cirio (Martínez-Noguez et al. 2020) represent independent and contemporary MUs. We aimed to understand whether geological and climatic events that shaped the phylogeographic lineages detected using cpDNA sequences continue to be evident in the population dynamics of cirio today. We also sought to unravel the dynamics of structure, gene flow, inbreeding, and kinship in the species, and how the species' biological capabilities influence them. We anticipate that the information generated will allow the development of strategies for the management and conservation of contemporary genetic diversity, defining populations with greater relevance for diversity contribution to the species.

Methods

Sample collection

We used samples from 64 of the 108 individuals analyzed by Martínez-Noguez (et al., 2020). We included 34 individuals from four localities of the northern lineage, which has a continuous distribution on the Baja California Peninsula: 8 individuals from North of Cataviña, 11 from Bahía de los Ángeles, 7 from North of Rosarito, and 8 from Piedras Blancas. We also included 20 individuals from two southern lineages—10 from Tres Vírgenes and 10 from Sierra de San Francisco—which were isolated from the northern lineage by the activity of the Tres Vírgenes Volcano Complex (≈ 1.20 Mya) and the mid-peninsula seaway during the late Pleistocene (≈ 1.53 Mya). Finally, we included 10 individuals from the mainland locality of Sierra Bacha in Sonora, which was isolated from the Baja California lineages by the formation of the Gulf of California during the late Miocene (≈ 7.24 Mya, Martínez-Noguez et al., 2020; see Fig. 1). Samples were obtained by collecting approximately 5 g of leaf tissue from adult trees separated by a minimum distance of 20 meters to avoid sampling close relatives. DNA was extracted from dried leaf samples with the 2% cetyltrimethylammonium bromide method (Doyle and Doyle 1990) and the purification protocol used magnetic beads from the Thermo Scientific KingFisher Plant DNA kit to obtain 15 µL of DNA (50 ng/µL).

Libraries, single nucleotide polymorphism calling and filtering

We used the 3RADseq methodology to obtain SNPs throughout the genome. This method is based on double-digesting restriction site-associated DNA (ddRAD), followed by digestion with a third restriction enzyme for dimer removal (Bayona-Vásquez et al. 2019). The libraries were made using three restriction enzymes (*NheI*, E*coRI*, and X*baI*). Paired-end

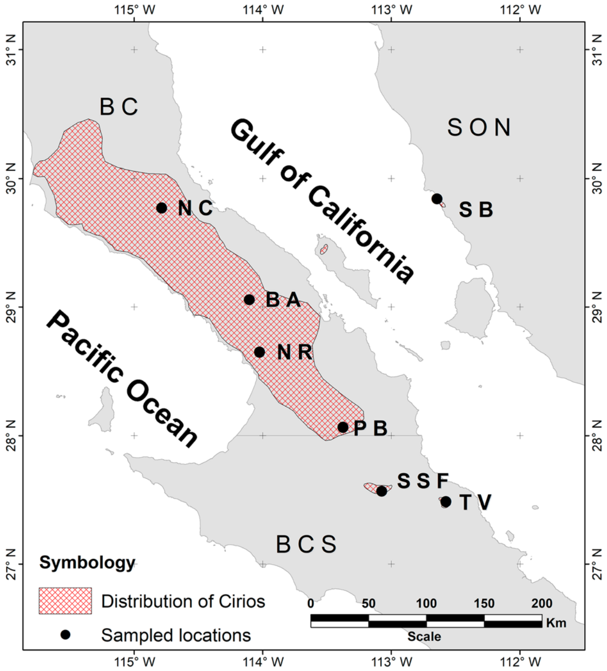

Fig. 1 Distribution of cirio (*Fouquieria columnaris*). Black dots represent the sampled locations. Legend for localities: Sierra Bacha (SB), North of Cataviña (NC), Bahía de los Ángeles (BA), North of Rosarito (NR), Piedras Blancas (PB), Sierra de San Francisco (SSF), Tres Vírgenes (TV), and the states of Baja California (BC), Baja California Sur (BCS), and Sonora (Son). The coordinates of the localities are given in Supplementary Information 1

sequencing was performed on the Illumina Hiseq 4000 platform in the Georgia Genomics and Bioinformatics Core at the University of Georgia. The raw Illumina data were quality filtered and demultiplexed using the *process_radtags* function in the STACKS v2.4 pipeline (Catchen et al. 2013). The barcodes and RAD-Tags (*-r*) were recovered and cleaned of any reads with an uncalled base (*-c*), and we discarded low-quality reads (*-q*) and trimmed to 140 pb (*-t*). Due to the lack of a reference genome from a closely related species, we made replicas to optimize de novo assembly. De novo assembly was performed with varying parameter values in the wrapper *denovo_map.pl* in STACKS. To minimize SNP genotyping errors, we determined the optimal *ustacks* parameter *-m* = 3–5 (minimum depth of coverage to create a stack), *-M* = 1–3 (maximum distance between stacks) and *cstacks* parameter *-n* = 1–4 (number of mismatches allowed between stacks when creating a catalogue) (Mastretta-Yanes et al. 2015; Paris et al. 2017). We chose as the optimal parameter set the set that maximized the number of variant sites, resulted in the best-structured data set (Gugger et al. 2018), reduced the percentage of missing data in individuals, and minimized differences inferred from the ordination of the resulting SNPs in a principal component analysis. The optimal parameters were -m = 3, -M = 2 and -n = 4 (Supplementary Information 2). For *ustacks*, we implemented a bounded-error model with the option bounded (*--model_type*) with a *-bound_high* of 0.05, and a maximum stack allowed per locus of 3 (*--max_locus_stacks*). We implemented a quality filter, retaining only one SNP per stack to prevent physically linked loci (*--write_single_snp*), and requiring that the locus be present in at least 70% of the total of individuals (-r) with a minor allele frequency > 0.05. We applied three different approaches to detect and delete atypical loci that could affect the neutrality of our final data set. A complete description of the methods used can be consulted in Supplementary Information 3. VCFtools v1.16 (Danecek et al. 2011) was used to filter sites that deviated from Hardy-Weinberg equilibrium (HWE, $P<0.001$). Linkage disequilibrium was identified by r2 values > 0.5 (--genor2, --min-r2 0.5). Only biallelic SNPs (--max-alleles 2) were kept, and individual samples with ≥30% missing data across loci were discarded. From the final filtered data set, we generated all of the file formats required for the subsequent analyses using PGDSPIDER v.2.1.1.5 (Lischer and Excoffier 2012).

Genetic diversity

To estimate neutral genetic diversity in the localities of North of Cataviña, Bahía de los Ángeles, Sierra Bacha (Sonora), Piedras Blancas, North of Rosarito, Sierra de San Francisco, and Tres Vírgenes (Fig. 1) and in the four populations detected by the clustering methods (see Results), we estimate rarefied allelic richness (Ar), observed heterozygosity (H_o), and the inbreeding coefficient (F_{IS}) using *basicStats* in the diveRsity R package (Keenan et al. 2013). To evaluate differences among localities and populations we used a Kruskal-Wallis test and posterior Gao test in the R package nparcomp v.3.0 (Konietschke et al. 2015). We also estimated inbreeding coefficients (*F statistics*) using the *--het* flag. Since some populations in our sample may be isolated, we also assessed the averaged pairwise relatedness index between each pair of individuals using the *--relatedness* method, with values around ~ 1.0 for self and ~ 0.00 for unrelated pairs, based on Yang et al. (2010), and visualized this using a heatmap in gplot in R. These two analyses were evaluated using VCFtools v1.16 (Danecek et al. 2011).

Population genetic analyses

We explored the population genetic structure using three different approaches. First, we implemented a discriminant analysis of principal components (DAPC) using the *dapc* function in the R package adegenet v.2.1.1 (Jombart and Ahmed 2011). To choose the optimal number of principal components to retain, we performed a cross-validation analysis with the *xvalDapc* function, and to select the most appropriate population structure (K) we used the Bayesian information criterion method (BIC) in the *find.cluster* function. As a second approach to infer the most likely value of K, we used STRUCTURE v2.3.4 (Pritchard et al. 2000). This Bayesian clustering method determines the number of ancestral populations based on allele frequencies under linkage disequilibrium and HWE assumptions. The STRUCTURE analysis was carried out in parallel by the StrAuto v1.0 package (Chhatre and Emerson 2017), which combines the analysis of STRUCTURE with STRUCTURE Harvester. We conducted an analysis using an admixture model and correlated allele frequencies with an alpha value of 1/k, as is recommended when sample sizes differ among localities (Wang 2017), and we used 100,000 burn-in and 1,000,000 MCMC steps with 15 replicates for each population value ($k=1$–7), inferred from the seven different localities. The value of the delta K method (Evanno et al. 2005) was used to determine the optimal value of K, and to visualize the results we merged the replicates run with CLUMPP v1.1.2 (Jakobsson and Rosenberg 2007).The third method we used to evaluate the genetic composition and divergence between paired sites was Jost's D estimates of pairwise genetic differentiation, which is recognized as a good predictor to assess these evolutionary processes (Jost et al. 2018). The genetic differentiation between sampling sites and their corresponding p-values were calculated with the R package strataG, based on 1000 bootstrap replicates

(Archer et al. 2017). The resulting p-values were corrected for false positives using the false discovery rate (FDR) correction (Benjamini and Yekutieli 2001). Heatmap and hierarchical clustering based on pairwise Jost's D were performed to determine the relationships among groups with the package pheatmap (v1.0.12) (https://www.rdocumentation.org/packages/pheatmap/versions/1.0.12/topics/pheatmap-package).

Spatial patterns of population structure and connectivity

Mantel's correlation tests were used to explore the influence of isolation by distance, i.e., to determine whether genetic similarity decreases with increasing geographic distance. We performed two variants of this analysis: the first using samples from the species' entire distribution area, and the second restricted to the peninsular area. A Euclidean geographic pairwise distance matrix was generated using the Geographic Distance Matrix Generator v.1.2.3 (Erst 2015). For the genetic distance, we used the previously described method of Jost's D. In addition, we performed a Mantel's test to evaluate isolation by environment, which analyzes the influence of the environment on the genetic diversity of all localities and peninsular localities. For the environmental distances between localities, we used the five bioclimatic variables previously chosen and refined as described in Supplementary Information 3. All Mantel's tests were performed in the R package adegenet v2.1.1 (Jombart and Ahmed 2011).

To assess the spatial patterns of genetic structure and the existence of barriers to gene flow, the Bayesian method of Estimating Effective Migration Surfaces (EEMS) was implemented. This method employs the concept of effective migration to graphically model the relationship between genetics and geography and identify sites where genetic similarity declines rapidly, which are then interpreted as barriers to gene flow between the populations (Petkova et al. 2016). We used the *make_eems_plots* function to visualize the results based on estimated effective migration rates on the log10-transformed scale after mean centering (m) (https://github.com/dipetkov/reemsplots2). Finally, to evaluate the relative migration between localities, we used divMigrate-online (Sundqvist et al. 2016), (https://popgen.shinyapps.io/divMigrate-online/). This method makes it possible to calculate the paired values of Jost's D as a measure of genetic distance; we employed a filter threshold that discarded values of relative migration rates (m_R) < 0.40 to determine which sites had the strongest migration signal. In addition, to test whether relative migration led to significant gene flow between localities, we executed 1,000 bootstrap replicates for all pairs of localities ($\alpha = 0.05$). Symmetric migration was assumed to occur in sites that had the same relative migration rate and significant gene flow; in cases where these conditions were not met, migration was considered asymmetric (Sundqvist et al. 2016).

Congruence among genetic distances and genetic distinctiveness

We used the "congruence among distance matrices" method (CADM; Campbell et al. 2011) to evaluate the concordance between the pairwise genetic distance matrix (Jost's D) composed of 1,137 neutral SNPs and the pairwise genetic distance matrix (F_{ST}) from cpDNA sequences reported in Martínez-Noguez et al. (2020). This analysis was conducted using the R package ape 3.0 (Paradis et al. 2004) and was performed with 100,000 permutations for the significance test. The concordance statistic (Kendall's W statistic) provides the degree of congruence of two distance matrices with a value on a scale from 0 (no congruence) to 1 (complete congruence; Kendall and Babington 1939).

We calculated a measure of the distinctive character of the population structure determined by the analyses described above. For this, we used the prioritization classification based on the Shapley metric (*SH*; Haake et al. 2008). This method generated a classification of distinctiveness that reflects the relative degree of isolation and ranks the populations in order of their relative contribution to future diversity (Volkmann et al. 2014). For this, the genetic distance Jost's *D* was calculated for each pair of populations using the method described above to generate a distance matrix, then NeighborNet network splits were computed using SplitsTree5 version 5.3.0 (Huson and Bryant 2006) and *SH* values were estimated using the method proposed by Volkmann et al. (2014).

Results

Libraries, single nucleotide polymorphism calling and filtering

3dRADseq generated a total of 132,855,470 paired-end reads in 64 individuals with an average of 1.02 million reads per individual. Three individuals were discarded because they exhibited coverage $<$ 0.1 million reads. The average depth for processed samples was 31.2x. The population-based secondary filtering step detected a total of 1,006,801 genotyped sites that after filtering yielded a data set composed of 1,405 variant sites in 61 individuals. In this data set, we detected and discarded 82 loci that were putatively under selection by the three different approaches. VCFtools filtering discarded six additional individuals that showed more than 30% missing data. Thus, after all filtering steps, the final neutral loci data contained 1,137 SNPs from 55 individuals (Supplementary Information 3).

Genetic diversity

Rarified allelic richness (*Ar*) ranged from 1.349 (1.328–1.370, CI 95%) for Piedras Blancas to 1.380 (1.356–1.403, CI 95%) for Bahía de los Ángeles (Table 1). Observed heterozygosity (H_o) ranged from 0.114 (0.104–0.123, CI 95%) for Piedras Blancas to 0.158 (0.145–0.170, CI 95%) for Bahía de los Ángeles. The inbreeding coefficient (F_{IS}) for Sierra Bacha presented the highest values with an average of 0.116 (0.082–0.150, CI 95%), and Bahía de los Ángeles (average of -0.064) and Tres Vírgenes (average of -0.029) presented the lowest values and had an excess of heterozygotes. The genetic diversity per locality and per population showed significant difference in the values by the Kruskal–Wallis test (Supplementary Information 4). The results of the paired Gao test indicated that Bahía de los Ángeles presented the most significant pairwise differences in *Ar* and H_o, and for F_{IS} the different localities and populations showed significant differences in most of the paired comparisons (Supplementary Information 4). These results are consistent with the values obtained by statistical inbreeding coefficients (*F statistics*), which mostly reflected historical inbreeding in all localities, especially in Sierra Bacha (average $F = 0.320$). The localities of Bahía de los Ángeles and Tres Vírgenes exhibited the lowest inbreeding values (Bahía de los Ángeles mean *F* statistic $= 0.058$ and Tres Vírgenes mean *F* statistic $= 0.110$, Table 1). The analysis of relatedness between individuals showed that there was marked relatedness

Table 1 Summary statistics for five genetic measures averaged across 1,137 neutral SNPs at 7 sampling locations of the *Fouquieria columnaris*. The confidence interval is given in parentheses (CI 95%). Bold titles in the left column refer to populations, the titles under the population label and in regular text are the constituent localities

Population	Localities	N	*Ar*	H_o	F_{IS}	*F*	*Relatedness*
NP		14	1.619 (1.595–1.642)	0.139 (0.130–0.148)	0.117 (0.090–0.143)	0.160 (0.093–0.227)	0.136 (0.115–0.158)
	NC	8	1.370 (1.348–1.393)	0.131 (0.120–0.142)	0.096 (0.057–0.134)	0.237 (0.165–0.309)	0.151 (0.125–0.177)
	BA	6	1.380 (1.356–1.403)	0.158 (0.145–0.170)	-0.064 (-0.096 - -0.031)	0.058 (0.007–0.109)	0.110 (0.069–0.151)
SB	SB	10	1.501 (1.473–1.529)	0.116 (0.107–0.126)	0.116 (0.082–0.150)	0.320 (0.252–0.389)	0.109 (0.088–0.129)
CP		21	1.602 (1.582–1.622)	0.122 (0.114–0.129)	0.168 (0.144–0.192)	0.261 (0.207–0.314)	0.053 (0.039–0.067)
	PB	8	1.349 (1.328–1.370)	0.114 (0.104–0.123)	0.110 (0.073–0.147)	0.307 (0.221–0.392)	0.058 (0.036–0.080)
	SSF	7	1.355 (1.334–1.377)	0.125 (0.114–0.135)	0.034 (0.000–0.070)	0.260 (0.147–0.372)	0.046 (0.019–0.073)
	NR	6	1.361 (1.338–1.384)	0.132 (0.121–0.144)	0.003 (-0.034–0.041)	0.200 (0.073–0.327)	0.053 (0.026–0.081)
TV	TV	10	1.517 (1.489–1.545)	0.154 (0.142–0.166)	-0.029 (-0.061–0.001)	0.110 (0.029–0.191)	0.161 (0.132–0.190)

N: Number of samples successfully genotyped and which passed filters, Ar: Rarefied allelic richness, *Ho*: Observed heterozygosity, *FIS*: Inbreeding coefficient, *F*: F *statistic* of inbreeding coefficient and *relatedness* indices. NP: North Peninsular population, CP: Central Peninsular population, SB: Sierra Bacha, TV: Tres Vírgenes, BA: Bahía de Los Ángeles, NC: North of Cataviña, NR: North of Rosarito, PB: Piedras Blancas, and SSF: Sierra de San Francisco

between pairs of individuals in the North peninsular, Sierra Bacha and Tres Vírgenes localities (Supplementary Information 5).

Population genetic analyses

The DAPC analysis retained a total of 17 principal components and six discriminant functions (Fig. 2). The first two axes explained 79.5% of the variation, and the first and third axes explained a total of 68.3% of the total variation in the data set. The likelihood BIC method determined that the model with K=2 was the best-fit model of the genetic cluster (Supplementary Information 6). Discriminant function 1 of the DAPC separated two regional genetic groups, each with a substructure. The northern region included North of Cataviña, Bahía de los Ángeles, and Sierra Bacha, which each constituted a separate group; the southern region included Piedras Blancas, North of Rosarito, Sierra de San Francisco,

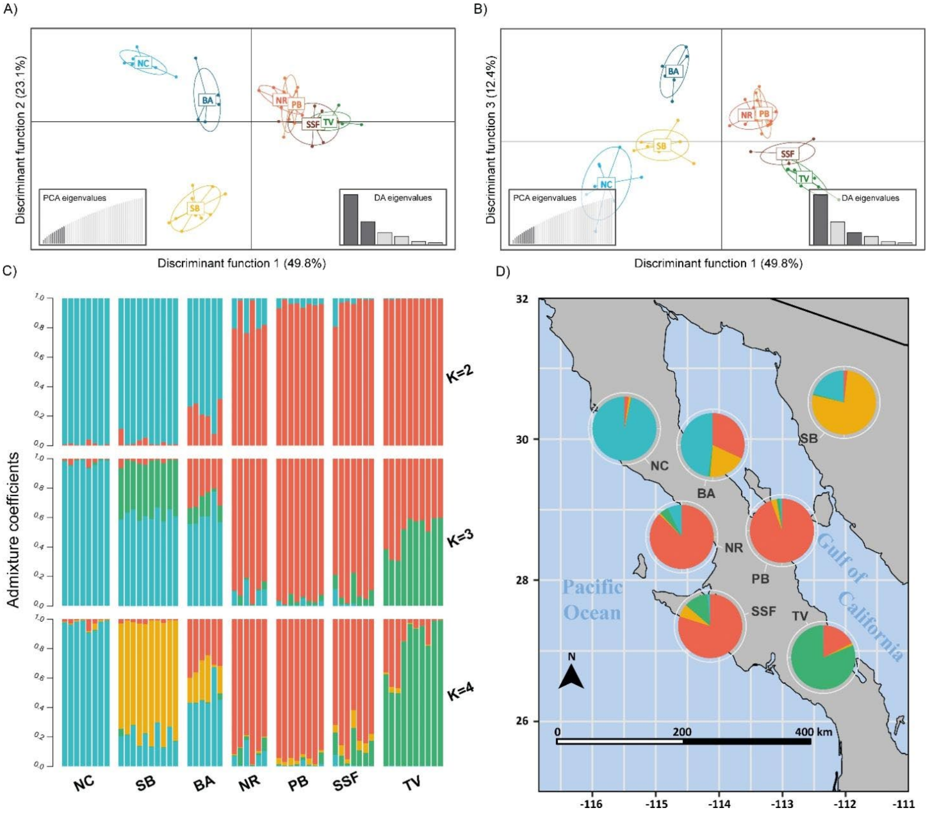

Fig. 2 Population structure of cirio (*Fouquieria columnaris*) based on 1,137 neutral SNPs at seven sampling sites. **(A)** DAPC scatterplot representing the first of two discriminant functions; **(B)** DAPC corresponding to the first and third discriminant functions; **(C)** STRUCTURE scatterplot for K = 2–4, each bar corresponds to an individual's membership in each cluster; **(D)** Pie chart composition indicating the distribution of geographic ancestry with K = 4. Legend for localities: Sierra Bacha (SB), North of Cataviña (NC), Bahía de los Ángeles (BA), North of Rosarito (NR), Piedras Blancas (PB), Sierra de San Francisco (SSF), and Tres Vírgenes (TV)

and Tres Vírgenes. Discriminant function 3 separated Piedras Blancas and North of Rosarito from Sierra de San Francisco and Tres Vírgenes (Fig. 2A & B).

Based on Evanno's delta K statistic, STRUCTURE found that the most probable number of K values was two (Supplementary Information 7), concordant with the clusters defined by the DAPC. A visual inspection of the additional K values of the STRUCTURE plot showed an additional substructure with K = 4; genetic admixing among localities was also evident. Unlike the DAPC, STRUCTURE detected that in the northern region, the species is conformed of two groups with distinct genetic structure—North of Cataviña and Sierra Bacha—while the Bahía de los Ángeles locality is a mixed population whose genetic signal is most similar to North of Cataviña, but which also presents genetic components of Sierra Bacha and of the central localities. The southern localities (North of Rosarito, Piedras Blancas, Sierra de San Francisco, and Tres Vírgenes) were differentiated into two main genetic groups: one composed of the central population (North of Rosarito, Piedras Blancas, and Sierra de San Francisco) and the other comprising only the Tres Vírgenes locality. The Tres

Vírgenes population showed some degree of genetic mixture with the Sierra de San Francisco locality (Fig. 2C & D).

All of the values of Jost's *D* pairwise genetic differentiation were statistically significant after FDR correction. The localities with the strongest pairwise genetic differentiation were North of Cataviña and Tres Vírgenes with a *D* value of 0.035, and Sierra Bacha and Tres Vírgenes, with a *D* of 0.028. The localities of Sierra de San Francisco, Piedras Blancas, and North of Rosarito presented the lowest Jost's *D* values in pairwise comparisons among them (maximum of 0.008 and minimum of 0.005, Fig. 3). Hierarchical clustering based on Jost's *D* confirmed the clustering detected by the previous analyses. Four different populations were recognized: a North peninsular population (composed of the localities of North of Cataviña and Bahía de los Ángeles), a Central peninsular population (composed of the localities of North of Rosarito, Piedras Blancas, and Sierra de San Francisco), and the localities of Tres Vírgenes (southern Peninsula) and Sierra Bacha (Sonora) each constituted independent populations (Figs. 2 and 3).

Spatial patterns of population structure and connectivity

There was a significant correlation between pairwise Jost's *D* and Euclidean geographic distance in both the analysis containing all localities (Mantel $r=0.650$, $p=0.014$) and the analysis restricted to the Peninsula (Mantel $r=0.816$ $p=0.003$). Mantel tests for isolation by environment for the complete data set (r = -0.386 $p=0.935$) and peninsular data set (r = -0.331 $p=0.774$) were not significant.

The EEMS analysis showed interesting migration estimates from Sierra Bacha on the mainland to a population in the North peninsular area. A barrier to gene flow was detected between the North and Central populations of the Baja California Peninsula. Gene flow was found between the localities of Sierra de San Francisco and Tres Vírgenes, suggesting that Sierra de San Francisco is the site of contact between Tres Vírgenes and Central populations (Fig. 4A). Based on DivMigrate, high gene flow values were detected from the mainland to the North and Central peninsular populations. Within the peninsula, there was symmetrical gene flow between the North and Central populations and between the Central and Tres Vírgenes populations (Fig. 4B). Finally, the significance test of asymmetrical gene flow

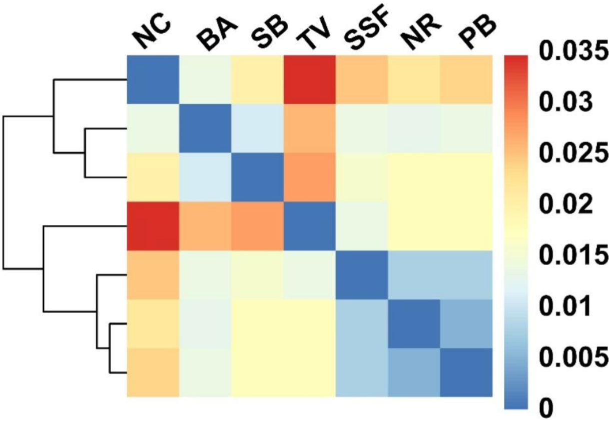

Fig. 3 Pairwise heatmap and hierarchical clustering based on Jost's *D*. All p-values were significant with a p-value < 0.05 after false discovery rate correction. The heatmap color code represents low differentiation with cold colors and higher differentiation with hot colors. Legend for localities: Sierra Bacha (SB), North of Cataviña (NC), Bahía de los Ángeles (BA), North of Rosarito (NR), Piedras Blancas (PB), Sierra de San Francisco (SSF), and Tres Vírgenes (TV)

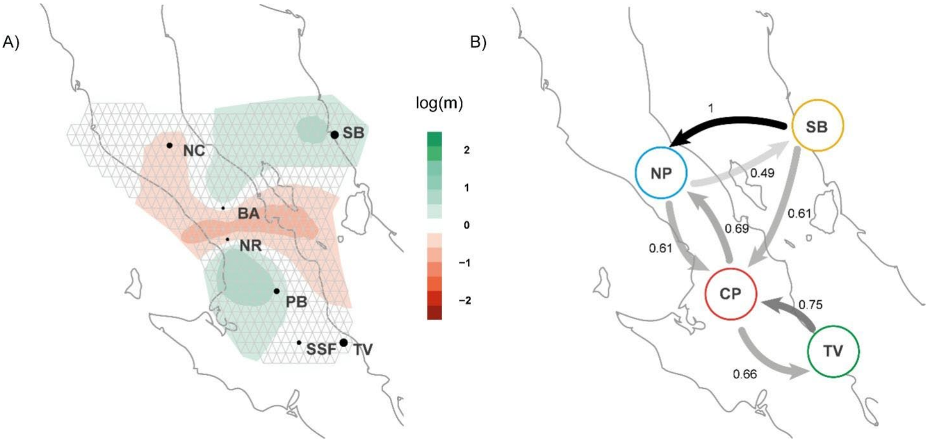

Fig. 4 **(A)** Estimation of effective migration surfaces (EEMS) of *Fouquieria columnaris*. The log(m) is the effective migration, where zero denotes the average effective migration, and red arrows indicate the area recognized as a possible barrier against gene flow. Dots correspond to the geographic sampling localities and are proportional in size to the number of individuals analyzed. SB: Sierra Bacha, NC: North of Cataviña, BA: Bahía de los Ángeles, NR: North of Rosarito, PB: Piedras Blancas, SSF: Sierra de San Francisco, and TV: Tres Vírgenes. **(B)** Relative migration network estimated by divMigrate. Only migration rates > 0.40 are represented between populations. NP/Blue: North Peninsular (composed of NC and BA), CP/Red: Central Peninsular (composed of NR, PB, and SSF), TV/Green: Tres Vírgenes, and SB/Yellow = Sierra Bacha

between localities showed statistically significant gene flow only from the locality of Sierra Bacha to the Northern peninsula population (Supplementary Information 8).

Congruence between genetic distances and genetic distinctiveness

Congruence in the genetic distance matrices between the SNP (Jost's D) and cpDNA (F_{ST}) data had a Kendall's W statistic = 0.616, that was not significant (permutation probability = 0.2). Seven elimination tests were performed to examine the influence of each of the sampled localities, excluding one of the seven localities in each test. The test in which North of Cataviña was excluded was the only one that rejected the null hypothesis of non-concordance; this test showed high and significant values between the distance matrices (Kendall's W statistic = 0.825, permutation probability = 0.01), which indicates more than 80% agreement between the two matrices when the locality of North of Cataviña is excluded.

The distinctiveness of the populations based on K = 4 and the NeighborNet tree showed that Tres Vírgenes (SH = 0.013) and Sierra Bacha (SH = 0.012) were the sites with the strongest genetic differences and may contribute the most strongly to the species' genetic diversity. Meanwhile, the North peninsular population (North of Cataviña and Bahía de Los Ángeles, SH = 0.009) and the Central peninsular population (North of Rosarito, Piedras Blancas, and Sierra de San Francisco, SH = 0.004) are the closest to the interior of the network and therefore make the smallest contribution to the species' genetic diversity (Supplementary Information 9).

Discussion

The results of the genetic variation of the nuclear genome (SNPs) revealed four genetic groups: a northern peninsular population, a central peninsular population, the populations of Tres Vírgenes (southern peninsula) and Sierra Bacha (mainland Sonora). Two of these groups—Tres Vírgenes and Sierra Bacha—coincided with evolutionary lineages determined using cpDNA sequences that were reported previously (Martínez Noguez et al. 2020). Low genetic differentiation, positive kinship and inbreeding patterns were evident, as was a strong pattern of isolation by distance on the Baja California Peninsula (but not when including the mainland). Cytonuclear mismatch was relatively minor and could be attributed mainly to the northernmost peninsular site, since removing that site resulted in concordance between the cpDNA and nuclear genomes (SNPs). Finally, there was more gene flow from the continental population to the northern peninsular population than in any other direction. This study highlights the importance of using reduced-representation genomic methods to provide information for reintroduction efforts for threatened species, resolving fine-scaled population structure and providing estimates of genetic connectivity based on neutral markers for conservation and management applications.

Genetic diversity

The mating system and life cycle shape genetic diversity in plants in such a way that a deficit of heterozygotes may be due to the loss of randomness in the crosses of a population (Glémin et al. 2006), and breeding among closely related individuals from the same population and self-fertilization can play a very important role in inbreeding (Abu Awad and Billiard 2017). As far as we know, there are no previous reports of genetic diversity of SNPs from species that are closely related to cirio, so the values reported here will be particularly useful as reference values for similar studies in the future ($H_o = 0.116–0.154$). These values are within the range of SNP H_o values ($H_o = 0.055–0.269$) reported in plant species with a variety of habits and life cycles and that have undergone historical bottlenecks, inbreeding, and habitat fragmentation (Borrell et al. 2018; Blyth et al. 2020; Jia et al. 2020; Stojanova et al. 2020; Brito et al. 2021).

Cirio has been impacted by various ancient geographical barriers that left signals of fragmentation in its cpDNA lineages (Martínez-Noguez et al. 2020). In addition, climatic changes during the transition from the Last Glacial Maximum to the Interglacial Period led to population contraction followed by a recent expansion of these lineages of the species (Martínez-Noguez et al. 2020). Biogeographical proposals for the Sonoran Desert suggest that the vegetation changed from forest to desert vegetation during the transition from the Last Glacial Maximum and the Holocene (Van Devender et al. 1994; Peñalba and Van Devender 1998; Wells 2000; Sankey et al. 2001). The four populations detected here have positive values in at least one of the inbreeding coefficients (F_{IS} and *F statistics*, Table 1), which suggests the populations have been under historical inbreeding processes. We also found positive and near-zero relatedness values (Table 1, Supplementary Information 5), confirming relatedness among the individuals sampled in the North peninsular, Sierra Bacha, and Tres Vírgenes populations, but the degree of kinship is lower than the relationship of half-siblings.

Mating systems can be a strong determinant of genetic diversity and structure in plants, making pollinators and seed dispersal, which determine gene flow, important mechanisms in determining the spatial genetic structure of tree species (Petit and Hampe 2006). Cirio produces sweet-smelling flowers with nectar around the base of the ovary (Henrickson 1972), which attracts large numbers of bees and hummingbirds, many of which are generalist pollinators (Nabhan et al. 2000). Bees have limited foraging movements, generating mating patterns between flowers of the same plant and neighboring plants (Wessinger 2021). Considering that the dispersal of cirio is limited (Humphrey 1974), nearby flowers likely belong to related individuals, producing patterns of genetic diversity consistent with our results. Although cirio can be a self-compatible species, it does not have mechanisms that facilitate self-pollination, since the branches of the style and the stigmas are located under the stamens and protected by the corolla (Henrickson 1972). However, it cannot be ruled out that pollinators may visit the flowers of the same panicle, which would generate assisted self-pollination, which could result in the signal of high inbreeding values that we have detected. The Sierra Bacha population had the highest inbreeding values, possibly due to its partial isolation on the mainland. This result is consistent with the high levels of genetic homogeneity that were observed in cirio from three different localities in Sierra Bacha based on an AFLP analysis, where the authors proposed isolation of a very low number of individuals in the mainland population (Gutiérrez-Ruacho et al. 2018). The Tres Vírgenes population had a negative inbreeding coefficient (F_{IS} = -0.061–0.001), indicating an excess of heterozygotes, and the lowest values of F_{IS} among all of the populations, but it also had positive indices of relatedness (Table 1, Supplementary Information 5). Although unpredictable and infrequent, outcrossing events can result in increased genetic diversity and decreased inbreeding signal, even in a single generation (Blyth et al. 2020). Therefore, it is probably a sign that the Tres Vírgenes population receives gene flow from the Central population, mainly from the locality of Sierra de San Francisco, which has decreased the inbreeding values and increased the values of heterozygosity of Tres Vírgenes (Table 1, Fig. 4). The same pattern was found with the cpDNA sequences, where Sierra de San Francisco appears to function as a contact area between the northern localities and Tres Vírgenes, where the most abundant haplotype in Sierra de San Francisco was also present in Tres Vírgenes (Martínez-Noguez et al. 2020).

Population structure and connectivity

Based on a genetic analysis of isoenzymes in cirio, Hall (1973) showed that Sierra Bacha and the northern peninsular zone are genetically similar, which led him to hypothesize that Sierra Bacha could have originated from the transport of seed from the northern peninsula. Sierra Bacha appears to be a relict lineage, with an isolation event associated with the formation of the Gulf of California, a signal that was evident in the results based on cpDNA with divergence times estimated during the Late Miocene, between 5.75–8.74 Mya (Martínez-Noguez et al. 2020).

Our current study suggests that the isolation caused by the formation of the Gulf of California was partial (rather than complete, as was previously thought; Martínez-Noguez et al. (2020). Our results show clear genetic connectivity between Sierra Bacha and the northern peninsular region, probably caused by physical, climatic, and biological processes that influence gene flow between the two populations (Fig. 4). One possibility is that the gene flow patterns in these regions could be mediated by pollinators that travel from the mainland to the peninsular region. Birds, for instance, are known to be able to transport pollen over long distances, especially species like hummingbirds that can spread pollen farther than bees (Wessinger 2021). For example, pollinators of ocotillos (*Fouquieria macdougalii* and *Fouquieria splendens*), such as the hummingbird *Selasphorus rufus*, travel from Sonora to Baja California between 29° and 30° N. Sightings of this hummingbird have been reported on islands in the Gulf of California, specifically Tiburón, San Esteban, San Lorenzo Sur, and Ángel de la Guarda Islands (Cody 1983). However, this is not a yearly migration route, but rather, it occurs when the previous season's rainfall was particularly high, leading to high floral abundance (Van Devender et al. 2004). Furthermore, cirio´s well-defined summer flowering season in Sierra Bacha (June-August, Humphrey 1975, Nabhan et al. 2000) does not overlap with the spring migration (February-April) of Rufous hummingbird. Other species of hummingbirds, such as *Calypte costae*, are resident on the peninsular and low Sonoran coasts (Grant and Grant 1967). This species was recognized as an abundant visitor during the flowering season of cirio (Nabhan et al. 2000) and could therefore participate in the transport of pollen from cirio from one coast to the other across the Gulf of California. Consequently, hummingbirds could be promoting pollination and gene flow from the mainland to the peninsular area, while arthropod floral visitors may be playing a stronger role in local breeding. Undoubtedly, the annual climatic dynamics and phenology of the vegetation regulate the abundance and type of pollinators; these phenomena could affect the routes and directionality of cirio pollinators. An alternative scenario is based on seed dispersion, where the strong winds during hurricane season could connect the peninsula with the Sonora through the transport of germplasm (Humphrey 1974). Plants of the Fouquieriaceae family are wind-dispersed, evidenced by the presence of a wing on their seeds (Henrickson 1972). However, in the case of cirio, the seed is heavy relative to its wing size, such that wind dispersal is driven mainly by the height of the tree from which the seed falls (Nabhan et al. 2000), which seems counter to the hypothesis that they could be moved long distances by hurricanes after falling to the ground. Also, the STRUCTURE results show that the individuals from the Baja California Peninsula are not purely individuals from the Sierra Bacha population (Fig. 2). It therefore seems that gene flow occurred mainly due to pollen dispersal, more than through seed dispersal. The cirio population distributed on the northern part of Ángel de la Guarda Island, which was not sampled in this study, could provide more information on the mechanisms by which this species was dispersed around the Gulf of California.

Historical barriers such as the hypothetical mid-peninsular maritime channel and the activity of the regional volcanic complex represented a true barrier against gene flow for Tres Vírgenes population, which diverged 1.23–1.86 Mya (Martínez-Noguez et al. 2020). The Tres Vírgenes volcano has been active for more than a million years and is currently active (Capra et al. 1998; Schmitt et al. 2006). The effects of volcanism have probably continued to impact the diversity and the genetic structure of cirio in this region. Furthermore, the distribution of cirio on the volcano is restricted to the eastern side of the volcano, and the volcano may act as natural resistance to northward air-assisted seed dispersal. Potential wide-range pollinators of cirio, such as hummingbirds, manage to access mountainous and complex habitats (Lagomarsino et al. 2017), so they have potential access to Tres Vírgenes population, though the volcanism historically affected floral visitors and pollinators through the tephra and ashes from the volcanic complex activity (Martínez-Noguez et al. 2020).

Isolation by distance was detected when considering only the peninsular localities (Mantel $r=0.816$ $p=0.003$), suggesting that the dispersal potential of cirio on the peninsula is caused by short-range wind dispersal and the effect of local pollination by bees, the most abundant floral visitors of cirio (Nabhan et al. 2000). Although no isolation by environment was found, cirio is distributed across a wide latitudinal gradient where the climatic conditions can be variable, and it is also one of the few species of the Fouquieriaceae family that is associated with a wide variety of soil types (Humphrey 1974). Future landscape genomics-oriented approaches would allow the detection of subtle processes of local adaptation along with geographic and environmental variables (Sork et al. 2013). These studies would allow the analysis of adaptive genetic variation to define Adaptive Units in the species, which would provide information that complements the MUs, such as relevant information on the species' ability to cope with future climatic conditions, and information to support and design an effective management strategy for threatened species (Hohenlohe et al. 2021).

Cytonuclear discordance

The congruence analysis revealed concordance when individuals from the North of Cataviña locality were removed from comparisons of cpDNA sequences and nuclear matrices. It is interesting to highlight that cpDNA data indicate that the cirios of Northern Cataviña are currently under the effect of isolation by distance established after the Last Glacial Maximum (Martínez-Noguez et al. 2020), when the species began to exhibit a distribution that is more similar to present (Sankey et al. 2001). Therefore, the cytonuclear discordance shown by the individuals from the North of Cataviña could be explained by the fact that the low mutation rates and the uniparental and non-recombinant inheritance of the chloroplast genome require more time than the nuclear genome to show genetic divergences (Drouin et al. 2008; Greiner et al. 2014). Another relevant mechanism that could explain the cytonuclear discordance is the different dispersal rates of seeds and pollen and the genetic signal of sex-linked bias, since the nuclear genome is inherited biparentally while the chloroplast genome is inherited matrilineally (Xu et al. 2021). The large number of private haplotypes in the cpDNA indicates a lack of gene flow across the Gulf of California, and thus an ancient isolation of the Sierra Bacha lineage (Martínez-Noguez et al. 2020). However, SNPs suggest asymmetric migration from the mainland to the Northern peninsula, probably due to pollen migration. The absence of gene flow between cpDNA lineages would demonstrate sex-biased migration since in angiosperms cpDNA is predominantly inherited maternally (Greiner et al., 2014).

Our analysis is pioneering in a non-model species and in a botanical family little studied at the molecular level. The lack of a closely related species reference genome increases uncertainty about the calling of SNPs loci. Future studies should aim to provide information on the complete genomes of cirio and/or closely related species, which would serve as templates for SNP screening. This will provide greater certainty in genotyping and therefore in estimates of intra- and interpopulation genetic diversity, especially in plants where genetic exchange between the cytoplasmic and nuclear genomes is common (Yoshida et al. 2014; Cui et al. 2021).

Conservation implications

One requirement for proper conservation management is the development of entities that classify diversity in a way that can be managed to ensure long-term persistence (Funk et al. 2012). Genomic data are an important tool for understanding biology and learning about population dynamics, which allows decision-makers to engage in specific actions for the conservation and management of flora, fauna, and their habitats in the short term (Hohenlohe et al. 2021). Our results define four genetic populations and the patterns of connectivity between them. These populations can be recognized as independent demographic units that can be managed differently and recognized as MUs. This information should be useful for conservation purposes since it can be used as a guide for reforestation or the improvement of assisted genetic diversity, as has occurred in endangered species, while minimizing the risks associated with outbreeding depression. It is important to stress that three of these MUs are already under the conservation efforts of the Mexican government. Part of the Central peninsular population and Tres Vírgenes population are located within the protected areas of El Vizcaíno, and the northern and part of the central peninsular populations are distributed within the Valle de los Cirios protected area (Supplementary Information 1). However, cirio faces several threats today. For example, it has been suggested that it has experienced historical bottlenecks caused by geological and climate change events (Martínez-Noguez et al. 2020), which our genetic diversity results seem to support. These factors could make the species more susceptible to future threats, since the constant loss of diversity affects the adaptive capacity and survival of the species. Contemporary climate change could be the most important factor facing the species, especially since it appears to be sensitive to specific climatic requirements (Humphrey 1974). Demographic censuses between latitudes 30.123° and 28.696° (the northern and part of the central peninsula populations) have shown that mortality exceeds recruitment (Bullock et al. 2005). These data suggest a population decline in a species that previously exhibited an abundant and geographically extensive distribution. In addition, the cirio isotype that is found in the Gray Herbarium (herbarium barcode: 00075341) was collected in 1860 on Cedros Island on the shores of Sebastián Vizcaíno Bay; there are currently no cirios present at that location, which could indicate a local extinction there over a relatively short time frame, probably due to recent climate changes.

Physiological requirements make cirio grow in areas with strong maritime influence from the Pacific Ocean, mainly on slopes and mountain ranges where it is subject to the influence of fog and environmental humidity (Webb et al. 2014; Webb and Turner 2020). Even in the Vizcaíno region, precipitation in the form of mist from the Pacific Ocean provides an additional contribution of water and cloudiness that reduces evaporation, despite being one of the driest regions of the Sonoran Desert (Peinado et al. 2005). The same occurs in Sierra Bacha, Sonora, where the influence of humidity from the Gulf of California restricts the distribution of cirio to mountain slopes to the north and towards the sea (Humphrey 1974). The influence of marine humidity is important for other groups of plants in the peninsular desert, as is the case of the genus *Cochemiea*/*Mammillaria* (Breslin et al. 2022), with which cirio cohabits in much of its distribution. Therefore, it would be interesting to carry out research on the change in the distribution of genetic populations under future climate scenarios to propose strategies for the protection of genetic diversity in these iconic species. Another threat is the anthropic impact of urbanization. Cirio has been subject to undocumented extraction for use as fuel and in rustic construction and decoration, and excessive logging for local use or export should be evaluated and limited (Bullock et al. 2005). An additional issue that needs to be evaluated is off-road races, which have taken place on the Baja California Peninsula since the 1960s (Fiolka, 2005). These same races were prohibited in the United States of America due to their serious environmental impact. However, on the Baja California Peninsula they remain a recurring practice, even within the protected natural areas of Valle de los Cirios and the Vizcaíno Reserve, with repercussions on the regional desert ecosystems (Vidarte 2012). Given that local extinctions have been documented in one of the areas where the species once showed its greatest abundance (Cedros Island), it is important to highlight the urgent need for demographic monitoring and management of populations with a restricted distribution such as Sierra Bacha and Tres Vírgenes. These same populations, according to our results with the Shapley metric (*SH*), contain an important pool of genetic diversity that may contribute to future diversity. Although Tres Vírgenes is in an area under national conservation programs, the population of Sierra Bacha is not, so we propose the designation of this region for conservation of a key part of the biodiversity of cirio. This locality has previously been proposed for conservation (“Sierra Seri” conservation area) due to its high endemism and high concentration of species at risk, and a high level of relictualism (Arriaga et al. 2000). We hope that our work contributes to the protection of cirio and the species with which it cohabits.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10531-023-02627-6.

Acknowledgements The authors thank Dr. José Luis León De La Luz for his helpful comments on the discussion. We acknowledge the support of the PhD scholarship provided by the Consejo Nacional de Ciencias y Tecnología (no. 549226) to JJMN.

Author contributions JJMN designed the study, conceived and designed the experiments, performed the experiments, prepared figures and tables, authored or reviewed drafts of the paper, and approved the final draft. FJGDL designed the study, conceived and designed the experiments, provided support for the data analysis, authored or reviewed drafts of the paper, and approved the final draft.

Funding This work was supported by the Rufford Foundation with the 1st Rufford Small Grant (Application ID: 32109-1). FJGDL provided funding for the development of the project.

Declarations

Competing interests The authors declare no conflict of interest. The authors have no relevant financial or non-financial interest to disclose.

References

Abu Awad D, Billiard S (2017) The double edged sword: the demographic consequences of the evolution of self-fertilization. Evolution 71:1178–1190. https://doi.org/10.1111/evo.13222

Archer FI, Adams PE, Schneiders BB (2017) Stratag: an r package for manipulating, summarizing and analysing population genetic data. Mol Ecol Resour 17:5–11. https://doi.org/10.1111/1755-0998.12559

Arriaga CL, Espinoza JM, Aguilar C et al (2000) Regiones Terrestres Prioritarias de México. Comisión Nacional para el Conocimiento y Uso de la Biodiversidad, México

Barbosa S, Mestre F, White TA et al (2018) Integrative approaches to guide conservation decisions: using genomics to define conservation units and functional corridors. Mol Ecol 27:3452–3465. https://doi.org/10.1111/mec.14806

Bayona-Vásquez NJ, Glenn TC, Kieran Troy J et al (2019) Adapterama III: Quadruple-indexed, triple-enzyme RADseq libraries for about $1USD per sample (3RAD). PeerJ 7:e7724. https://doi.org/10.7717/peerj.7724

Benestan LM, Ferchaud AL, Hohenlohe PA et al (2016) Conservation genomics of natural and managed populations: building a conceptual and practical framework. Mol Ecol 25:2967–2977. https://doi.org/10.1111/mec.13647

Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. Ann Stat 29:1165–1188

Blyth C, Christmas MJ, Bickerton DC et al (2020) Increased genetic diversity via gene flow provides hope for *Acacia whibleyana*, an endangered wattle facing extinction. Diversity 12:299. https://doi.org/10.3390/D12080299

Borrell JS, Wang N, Nichols RA, Buggs RJA (2018) Genetic diversity maintained among fragmented populations of a tree undergoing range contraction. Heredity (Edinb) 121:304–318. https://doi.org/10.1038/s41437-018-0132-8

Breslin PB, Wojciechowski MF, Majure LC (2022) Remarkably rapid, recent diversification of Cochemiea and Mammillaria in the Baja California, Mexico region. Am J Bot 109:1472–1487. https://doi.org/10.1002/ajb2.16048

Brito F, de Zucchi A, Pinheiro MI JB, et al (2021) Low diversity in the native populations of *Croton tetradenius* Baill. When using SNP markers: a future crop with an insecticidal activity. Genet Resour Crop Evol 68:3145–3159. https://doi.org/10.1007/s10722-021-01176-7

Bullock SH, Martijena NE, Webb RH, Turner RM (2005) Twentieth century demographic changes in cirio and cardón in Baja California, México. J Biogeogr 32:127–143. https://doi.org/10.1111/j.1365-2699.2004.01152.x

Campbell V, Legendre P, Lapointe FJ (2011) The performance of the congruence among Distance Matrices (CADM) test in phylogenetic analysis. BMC Evol Biol 11:64. https://doi.org/10.1186/1471-2148-11-64

Capra L, Macías J, Espíndola J, Siebe C (1998) Holocene plinian eruption of La Virgen volcano, Baja California, Mexico. J Volcanol Geotherm Res 80:239–266. https://doi.org/10.1016/S0377-0273(97)00049-8

Catchen J, Hohenlohe PA, Bassham S et al (2013) Stacks: an analysis tool set for population genomics. Mol Ecol 22:3124–3140. https://doi.org/10.1111/mec.12354

Chhatre VE, Emerson KJ (2017) StrAuto: automation and parallelization of STRUCTURE analysis. BMC Bioinformatics 18:1–5. https://doi.org/10.1186/s12859-017-1593-0

Cody ML (1983) The land birds. In: Case TJ, Cody ML (eds) Island biogeography in the Sea of Cortez. University of California Press, pp 210–264

Cui H, Ding Z, Zhu Q et al (2021) Comparative analysis of nuclear, chloroplast, and mitochondrial genomes of watermelon and melon provides evidence of gene transfer. Sci Rep 11:1–9. https://doi.org/10.1038/s41598-020-80149-9

Danecek P, Auton A, Abecasis G et al (2011) The variant call format and VCFtools. Bioinformatics 27:2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. Focus (Madison) 12:13–15. https://doi.org/10.3923/rjmp.2012.65.73

Drouin G, Daoud H, Xia J (2008) Relative rates of synonymous substitutions in the mitochondrial, chloroplast and nuclear genomes of seed plants. Mol Phylogenet Evol 49:827–831. https://doi.org/10.1016/j.ympev.2008.09.009

Duminil J, Fineschi S, Hampe A et al (2007) Can population genetic structure be predicted from life-history traits? Am Nat 169:662–672

Erst PJ (2015) Geographic Distance Matrix Generator (version 1.2.3). https://biodiversityinformatics.amnh.org/open_source/gdmg/. Accessed 15 June 2022

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. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Fiolka M (2005) 1000 Miles to Glory: the history of the Baja 1000, 3rd edn. David Bull Publishing, Hong Kong

Fraser DJ, Bernatchez L (2001) Adaptive evolutionary conservation: towards a unified concept for defining conservation units. Mol Ecol 10:2741–2752. https://doi.org/10.1046/j.1365-294X.2001.t01-1-01411.x

Funk WC, McKay JK, Hohenlohe PA, Allendorf FW (2012) Harnessing genomics for delineating conservation units. Trends Ecol Evol 27:489–496. https://doi.org/10.1016/j.tree.2012.05.012

Glémin S, Bazin E, Charlesworth D (2006) Impact of mating systems on patterns of sequence polymorphism in flowering plants. Proc R Soc B Biol Sci 273:3011–3019. https://doi.org/10.1098/rspb.2006.3657

Grant KA, Grant V (1967) Effects of hummingbird migration on plant speciation in the California flora. Evol (N Y) 21:457. https://doi.org/10.2307/2406607

Greiner S, Sobanski J, Bock R (2014) Why are most organelle genomes transmitted maternally? BioEssays 37:80–94. https://doi.org/10.1002/bies.201400110

Gugger PF, Liang CT, Sork VL et al (2018) Applying landscape genomic tools to forest management and restoration of hawaiian koa (*Acacia koa*) in a changing environment. Evol Appl 11:231–242. https://doi.org/10.1111/eva.12534

Gutiérrez-Ruacho O, Coronado ML, Sánchez-Teyer F et al (2018) Abundance of rhizospheric bacteria and fungi associated with *Fouquieria columnaris* at Punta Cirio, Sonora, Mexico. Rev Mex Biodivers 89:541–552. https://doi.org/10.22201/ib.20078706e.2018.2.1620

Haake CJ, Kashiwada A, Su FE (2008) The Shapley value of phylogenetic trees. J Math Biol 56:479–497. https://doi.org/10.1007/s00285-007-0126-2

Hall RL (1973) A physiological genetic study of the populations of Idria columnaris (Kellogg) of Baja California and Sonora, Mexico, Dissertation edn. The University of Arizona

Henrickson J (1972) A taxonomic revision of the Fouquieriaceae. Aliso 7:439–537. https://doi.org/10.5642/aliso.19720704.08

Hohenlohe PA, Funk WC, Rajora OP (2021) Population genomics for wildlife conservation and management. Mol Ecol 30:62–82. https://doi.org/10.1111/mec.15720

Humphrey RR (1974) The Boojum and its home. *Idria columnaris* Kellogg and its home. University of Arizona Press

Humphrey RR (1975) Phenology of selected Sonoran Desert plants at Punta Cirio, Sonora, Mexico. J Ariz Acad Sci 10:50–67

Humphrey RR, Werner GF (1964) Some records of bee visitations to the flowers of *Idria columnaris*. J Arizona Acad Sci 5:243–244

Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. Mol Biol Evol 23:254–267. https://doi.org/10.1093/molbev/msj030

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. https://doi.org/10.1093/bioinformatics/btm233

Jia KH, Zhao W, Maier PA et al (2020) Landscape genomics predicts climate change-related genetic offset for the widespread *Platycladus orientalis* (Cupressaceae). Evol Appl 13:665–676. https://doi.org/10.1111/eva.12891

Jombart T, Ahmed I (2011) Adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. Bioinformatics 27:3070–3071. https://doi.org/10.1093/bioinformatics/btr521

Jost L, Archer F, Flanagan S et al (2018) Differentiation measures for conservation genetics. Evol Appl 11:1139–1148. https://doi.org/10.1111/eva.12590

Keenan K, Mcginnity P, Cross TF et al (2013) DiveRsity: an R package for the estimation and exploration of population genetics parameters and their associated errors. Methods Ecol Evol 4:782–788. https://doi.org/10.1111/2041-210X.12067

Kellogg EA, Appels R, Mason-Gamer RJ (1996) When genes tell different stories: the diploid genera of Triticeae (Gramineae). Syst Bot 21:321–347. https://doi.org/10.2307/2419662

Kendall M, Babington BS (1939) The problem of m rankings. Ann Math Stat 10:275–287

Konietschke F, Placzek M, Schaarschmidt F, Hothorn LA (2015) Nparcomp: an R software package for nonparametric multiple comparisons and simultaneous confidence intervals. J Stat Softw 64:1–17. https://doi.org/10.18637/jss.v064.i09

Lagomarsino LP, Forrestel EJ, Muchhala N, Davis CC (2017) Repeated evolution of vertebrate pollination syndromes in a recently diverged Andean plant clade. Evolution 71:1970–1985. https://doi.org/10.1111/evo.13297

Lischer HEL, Excoffier L (2012) PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. Bioinformatics 28:298–299. https://doi.org/10.1093/bioinformatics/btr642

Martínez-Noguez JJ, De La León JL, Delgadillo Rodríguez J, Garciá-De León FJ (2020) Phylogeography and genetic structure of an iconic tree of the Sonoran Desert, the Cirio (*Fouquieria columnaris*), based on chloroplast DNA. Biol J Linn Soc 130:433–446. https://doi.org/10.1093/biolinnean/blaa065

Mastretta-Yanes A, Arrigo N, Alvarez N et al (2015) Restriction site-associated DNA sequencing, genotyping error estimation and de novo assembly optimization for population genetic inference. Mol Ecol Resour 15:28–41. https://doi.org/10.1111/1755-0998.12291

Nabhan GP, Donovan J, Buchmann S, Búrquez A (2000) Riqueza y estabilidad de los polinizadores de los cirios (*Fouquieria columnaris*) del desierto sonorense: Un análisis comparativo. Bot Sci 50:39. https://doi.org/10.17129/botsci.1593

Palsbøll PJ, Bérubé M, Allendorf FW (2007) Identification of management units using population genetic data. Trends Ecol Evol 22:11–16. https://doi.org/10.1016/j.tree.2006.09.003

Paradis E, Claude J, Strimmer K (2004) APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20:289–290. https://doi.org/10.1093/bioinformatics/btg412

Paris JR, Stevens JR, Catchen JM (2017) Lost in parameter space: a road map for stacks. Methods Ecol Evol 8:1360–1373. https://doi.org/10.1111/2041-210X.12775

Peinado M, Delgadillo J, Aguirre JL (2005) Plant associations of El Vizcaíno Biosphere Reserve, Baja California Sur, Mexico. Southwest Nat 50:129–149. https://doi.org/10.1894/0038-4909(2005)050[0129:PAOEVB]2.0.CO;2

Peñalba MC, Van Devender TR (1998) Cambios de vegetación y clima en Baja California, México, durante los últimos 20 000 años. Geol del Noroeste 2:21–23

Petit RJ, Hampe A (2006) Some evolutionary consequences of being a tree. Annu Rev Ecol Evol Syst 37:187–214. https://doi.org/10.1146/annurev.ecolsys.37.091305.110215

Petkova D, Novembre J, Stephens M (2016) Visualizing spatial population structure with estimated effective migration surfaces. Nat Genet 48:94–100. https://doi.org/10.1038/ng.3464

Pritchard JK, Stephens M, Peter D (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959. https://doi.org/10.1007/s10681-008-9788-0

Renoult JP, Kjellberg F, Grout C et al (2009) Cyto-nuclear discordance in the phylogeny of Ficus section Galoglychia and host shifts in plant-pollinator associations. BMC Evol Biol 9:1–18. https://doi.org/10.1186/1471-2148-9-248

Rose JP, Toledo CAP, Lemmon EM et al (2021) Out of sight, out of mind: widespread nuclear and plastid-nuclear discordance in the flowering plant genus Polemonium (Polemoniaceae) suggests widespread historical gene flow despite limited nuclear signal. Syst Biol 70:162–180. https://doi.org/10.1093/sysbio/syaa049

Sankey JT, Van Devender TR, Clark WH (2001) Late Holocene plants, Cataviña, Baja California. Southwest Nat 46:1–7

Schmitt AK, Stockli DF, Hausback BP (2006) Eruption and magma crystallization ages of Las Tres Vírgenes (Baja California) constrained by combined 230Th/238U and (U-Th)/He dating of zircon. J Volcanol Geotherm Res 158:281–295. https://doi.org/10.1016/j.jvolgeores.2006.07.005

Schwartz MK, Luikart G, Waples RS (2007) Genetic monitoring as a promising tool for conservation and management. Trends Ecol Evol 22:25–33. https://doi.org/10.1016/j.tree.2006.08.009

Sork VL, Aitken SN, Dyer RJ et al (2013) Putting the landscape into the genomics of trees: approaches for understanding local adaptation and population responses to changing climate. Tree Genet Genomes 9:901–911. https://doi.org/10.1007/s11295-013-0596-x

Stojanova B, Šurinová M, Zeisek V et al (2020) Low genetic differentiation despite high fragmentation in the endemic serpentinophyte Minuartia smejkalii (M. verna agg., Caryophyllaceae) revealed by RADSeq SNP markers. Conserv Genet. https://doi.org/10.1007/s10592-019-01239-4

Sundqvist L, Keenan K, Zackrisson M et al (2016) Directional genetic differentiation and relative migration. Ecol Evol 6:3461–3475. https://doi.org/10.1002/ece3.2096

United Nations (2011) Nagoya protocol on access to genetic resources and the fair and equitable sharing of benefits arising from their utilization to the convention on biological diversity. Secretariat of the convention on biological diversity. Montreal, Canada

Van Devender TR, Burgess TL, Piper JC, Turner RM (1994) Paleoclimatic implications of Holocene plant remains from the Sierra Bacha, Sonora. Mexico Quat Res 41:99–108

Van Devender TR, Calder WA, Krebbs K et al (2004) Hummingbird plants and potential nectar corridors of the Rufous hummingbird in Sonora, Mexico. In: Nabhan GP (ed) Conserving migratory pollinators and nectar corridors in western North America. The University of Arizona Press, Tucson, United States of America, pp 96–107

Vidarte AR (2012) Implicaciones sociales y ambientales de las carreras fuera de carretera en la Peninsula de Baja California. Dissertation, El Colegio de la Frontera Norte

Volkmann L, Martyn I, Moulton V et al (2014) Prioritizing populations for conservation using phylogenetic networks. PLoS ONE 9(2):e88945. https://doi.org/10.1371/journal.pone.0088945

Wang J (2017) The computer program structure for assigning individuals to populations: easy to use but easier to misuse. Mol Ecol Resour 17:981–990. https://doi.org/10.1111/1755-0998.12650

Webb RH, Turner RM (2020) Biodiversity of perennial vegetation in the desert regions of Baja California and Baja California Sur, Mexico. Encyclopedia of the World's Biomes, Volume 2. Elsevier Inc. https://doi.org/10.1016/B978-0-12-409548-9.11997-9

Webb RH, Salazar-Ceseña M, Turner M R (2014) The distribution of Cirio (*Fouquieria columnaris*) in Baja California, Mexico. Haseltonia 19:26–37. https://doi.org/10.2985/026.019.0105

Wells PV (2000) Pleistocene macrofossil records of four-needled pinyon or juniper encinal in the Northern Vizcaino Desert, Baja California del Norte. Madroño 47:189–194

Wessinger CA (2021) From pollen dispersal to plant diversification: genetic consequences of pollination mode. New Phytol 229:3125–3132. https://doi.org/10.1111/NPH.17073

Xu LL, Yu RM, Lin XR et al (2021) Different rates of pollen and seed gene flow cause branch-length and geographic cytonuclear discordance within asian butternuts. New Phytol 232:388–403. https://doi.org/10.1111/nph.17564

Yang J, Benyamin B, McEvoy BP et al (2010) Common SNPs explain a large proportion of the heritability for human height. Nat Genet 42:565–569. https://doi.org/10.1038/ng.608

Yoshida T, Furihata HY, Kawabe A (2014) Patterns of genomic integration of nuclear chloroplast DNA fragments in plant species. DNA Res 21:127–140. https://doi.org/10.1093/dnares/dst045