BAG | Journal of Basic & Applied Genetics

Vol XXXV (2); July-December 2024

Journal of the Argentine Society of Genetics

POPULATION EXPANSION OF *Prosopis alba* GRISEB. (LEGUMINOSAE) IN SOUTHERN SOUTH-AMERICA: PHYLOGEOGRAPHICAL AND ECOLOGICAL APPROACH BASED ON CPDNA SEQUENCES

EXPANSIÓN POBLACIONAL DE *Prosopis alba* GRISEB. (LEGUMINOSAE) EN EL SUR DE SUDAMÉRICA: APROXIMACIÓN FILOGEOGRÁFICA Y ECOLÓGICA BASADA EN SECUENCIAS DE ADNCP

Bessega C.[1,2*], Pometti C.[1,2], Saidman B.O.[1,2], Fortunato R.[3,4], Santoro C.M.[5,6], Mcrostie V.[6,7,8], Vilardi J.C.[1,2]

[1]Departamento Ecología, Genética y Evolución (EGE), Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires (UBA), Buenos Aires, Argentina.

[2]Instituto de Ecología, Genética y Evolución (IEGEBA), CONICET-UBA, Buenos Aires, Argentina.

[3]Instituto de Botánica Darwinion (CONICET/ANCEFYN) CONICET, Buenos Aires, Argentina.

[4]ESIlyCA, Universidad de Morón, Morón, Argentina

[5]Instituto de Alta Investigación, Universidad de Tarapacá, Arica, Chile

[6]Núcleo Milenio AFOREST, NCS2022_024, Chile

[7]Facultad de Ciencias Sociales, Escuela de Antropología, Pontificia Universidad Católica de Chile, Santiago, Chile

[8]Centro del Desierto de Atacama UC, Santiago, Chile

Corresponding author:
Cecilia Bessega
cecib@ege.fcen.uba.ar

ORCID 0000-0002-8575-1828

Cite this article as:
Bessega C., Pometti C., Saidman B.O., Fortunato R., Santoro C.M., Mcrostie V., Vilardi J.C. 2024. POPULATION EXPANSION OF *Prosopis alba Griseb.* (LEGUMINOSAE) IN SOUTHERN SOUTH-AMERICA: PHYLOGEOGRAPHICAL AND ECOLOGICAL APPROACH BASED ON cpDNA SEQUENCES. BAG. Journal of Basic and Applied Genetics XXXV (2).

Received: 08/23/2023
Revised version received: 07/08/2024
Accepted: 07/31/2024
General Editor: Elsa Camadro
DOI: 10.35407/bag.2024.35.02.01
ISSN online version: 1852-6233

Available online at
www.sag.org.ar/jbag

ABSTRACT

Genealogical relationships among DNA lineages considering their current geographic distributions are useful to infer historical events that have shaped the contemporary distributions of species and their genetic variation. In this study we analyzed the variation of the *nadhF-rpl32* intergenic spacer in *Prosopis alba* Griseb. samples (*algarrobo*) collected in Chile, Argentina and Bolivia in order to contribute to our understanding of the evolutionary history of this species in southern South-America. We assessed the influence of environmental conditions on the demographic history of them by using a Bayesian ecological clustering (BPEC) approach and simulations based on the theory of coalescence. The results obtained allowed us to identify nine haplotypes. The Tajima (*TD*= -1.35) and Fu (*Fs*= -2.36) tests were non-significant, suggesting absence of selection. On the other hand, the disparity between sequences or raggedness (*rg*=0.021) was also non-significant, compatible with the population expansion. The coalescent analysis using MCMC indicated that the best fit demographic model was the linear growth one, with a time to the most recent common ancestor, for the haplotypes sampled in the present analysis, TMRA=0.0072, that is, roughly 7,000 generations. The BPEC analysis identified two clusters whose distribution partially overlaps in the Atacama Desert (Chile) and allows us to postulate that the species would have expanded to the north and west from the Chaqueña Region in Argentina. The comparison of scenarios by means of ABC (Approximate Bayesian Computation) analyses is in accordance with this result as the cases where the East cluster or the Argentinean samples were postulated as ancestral, yielded the higher posterior probabilities. The analysis performed contributed to the *P. alba* historical reconstruction throwing light on the trans Andean movement considering direction, time and natural- and human-mediated dispersal agents.

Key words: *algarrobo*, Atacama Desert, Bayesian analysis, coalescent models, *nadhF-rpL32* intergenic spacer, population genetics

RESUMEN

Las relaciones genealógicas entre linajes de ADN considerando sus distribuciones geográficas actuales son útiles para inferir eventos históricos que han dado forma a las distribuciones contemporáneas de las especies y su variación genética. En el presente trabajo se analizó la variación del espaciador intergénico *nadhF-rpl32* en muestras de *Prosopis alba* Griseb. (*algarrobo*) coleccionadas en Chile, Argentina y Bolivia para contribuir a nuestra comprensión de la historia evolutiva de esta especie en el sur de Sudamérica. Hemos evaluado la influencia de las condiciones ambientales sobre la historia demográfica de esta especie mediante la utilización de un enfoque de agrupamiento ecológico Bayesiano (BPEC) y simulaciones basadas en la teoría de la coalescencia. Los resultados obtenidos permitieron identificar nueve haplotipos. Las pruebas de Tajima (*TD*= -1.35) y Fu (*Fs*= -2.36) no fueron significativas, sugiriendo ausencia de selección. Por otra parte, la disparidad entre secuencias (*raggedness*, *rg*=0.021) tampoco fue significativa, compatible con la expansión poblacional. El análisis coalescente utilizando MCMC indicó que el modelo demográfico de mejor ajuste fue el de crecimiento lineal, con un tiempo hasta el ancestro común más reciente, para los haplotipos muestreados en el presente análisis, TMRA=0,0072, es decir, aproximadamente 7.000 generaciones. El análisis BPEC permitió identificar dos grupos cuya distribución se superpone parcialmente en el Desierto de Atacama (Chile) y permite postular que la especie se habría expandido hacia el norte y el oeste desde la Región Chaqueña en Argentina. La comparación de escenarios mediante análisis ABC (Cómputos Bayesianos Aproximados) concuerda con este resultado ya que los casos en donde las muestras del grupo del Este o de Argentina fueron postuladas como ancestrales arrojaron las mayores probabilidades posteriores. El análisis realizado contribuyó en la reconstrucción histórica de *P. alba* y en el esclarecimiento del movimiento trasandino considerando la dirección, el tiempo y los agentes de dispersión naturales y humanos.

Palabras clave: *algarrobo*, análisis Bayesiano, Desierto de Atacama, espaciador intergénico *nadhF-rpL32*, modelos coalescentes

INTRODUCTION

Examining population genetic structure over historical, spatial, and temporal scales and its relationship with environmental changes is crucial for understanding species distributions and adaptations to the ongoing climatic changes (Rico *et al.,* 2021). Phylogeography (Avise *et al.,* 1987) is a widely applied discipline that seeks to integrate the genealogical relationships among DNA lineages (sequences) with their current geographic distributions to infer historical events that have shaped the contemporary distributions of species and their genetic variation. More recently, Manolopoulou *et al.* (2016) proposed the use of Bayesian Phylogeographic and Ecological Clustering (BPEC) to combine DNA sequence genealogies with geographical distribution, environmental data, and phenotypic measurements to reveal geographic structuring of genetic clusters consistent with migration events.

Prosopis (Leguminosae, Mimosoideae) is a genus composed of 44 species belonging to five sections (*Prosopis, Anonychium, Strombocarpa, Monilicarpa* and *Algarobia*) well represented in arid and semi-arid regions of the world (Burkart, 1976). Recently, based on phylogenetic results, a proposal to disaggregate the genus was performed and the names for the sections and species were proposed to be changed (Hughes *et al.,* 2022). Here we still use the classification of *Prosopis* according to Burkart (1976), to consider the nomenclatural stability of taxonomic groups: Arts. 14.1, 34 and 56 from the International Code of Nomenclature for algae, fungi, and plants (Shenzhen Code) (Turland *et al.*, 2018).

Several *Algarobia* species are both ecologically and economically multipurpose trees as they can grow on sandy soils and contribute to stabilize dunes, combat desertification, and reforest degraded areas. This section (*algarrobos*) includes most species of economic importance that have been widely introduced in arid and semiarid regions, either with negative or positive effects on local populations (Burkart, 1976; Roig, 1993; Barros, 2010). *Algarrobos* are very appreciated in hot and arid areas as they provide shade, wood and edible fruits. The wood, very hard and highly caloric, is useful as firewood, charcoal and for furniture. Legume (pod) properties are highly variable among species and some of them are used as human food and forage (Burkart, 1976; Roig, 1993; Cony, 1996; Capparelli, 2007; Pometti *et al.,* 2009).

Prosopis alba Griseb. is one of the most important species of *Algarobia* from an economic point of view. It has been claimed that it is currently distributed in plains of subtropical Argentina, Uruguay, Paraguay, Bolivia, Perú and Chile (Burkart, 1976). In Chile, isolated natural and planted *Prosopis* populations can be found up to 3,000 masl throughout the western Andean slope, being typically confined to discrete zones of groundwater discharge and/or on riverbanks of perennial/ephemeral watersheds that flows into the Pampa del Tamarugal basin and the oases in the Salar de Atacama basin. Several authors (McRostie *et al.,* 2017; Bessega *et al.,* 2021) also recorded its presence in the Atacama Desert. Although this desert is one of the most hyper-arid on earth (Dunai *et al.,* 2005) some species of *Prosopis* can grow in it.

The distribution and cultural significance of Chilean *algarrobos* support the assumption that they are native to the Atacama Desert. AMS dating of *algarrobo* endocarps throughout several archaeological and paleoecological sites of the Atacama Desert estimates the presence of these trees around 4,000 years before present, suggesting that humans could have acted as vectors, at least for some of these species (McRostie *et al.,* 2017). Additionally, the considerable exo-morphological similarities between the trees from Atacama and *P. alba* specimens found in Salta Province, in the north of Argentina, have led Palacios and Brizuela (2005) to suggest that these species were introduced from Argentina during pre-Hispanic times. These authors stated that the presence and similarities between *algarrobos* from different localities in the Americas are a consequence of their cultivation, and this appears to be a case for domestication that, as with other American crops, did not persist with the arrival of European settlers.

Non-coding fragments of the chloroplast genome are the most appropriate markers for phylogeographical studies due to their uniparental inheritance and their capacity to detect processes of neutral evolution (Avise, 2000). By comparing DNA sequences, evolutionary relationships, levels of variation and geographical sub-structuring within and between groups of populations may be derived (Avise *et al.,* 1987). In particular the variation of the *nadhF-rpl32* intergenic spacer has been shown to be very useful in elucidating the origin and diversification of populations of a related species to *P. alba: P. chilensis* (Aguilar *et al.,* 2020). Here we analyzed sequence variation of the same cpDNA region in *P. alba* samples collected in Chile, Argentina, and Bolivia in order to contribute to our understanding of the evolutionary history of this species in southern South-America. We assessed the influence of environmental conditions on the demographic history of these species by using a BPEC approach and simulations based on the theory of coalescence. Considering the extreme characteristics of the Atacama Desert and previous works that discuss the endemism of *P. alba* in Chile, our work is based on the following two hypotheses: (1) Trans Andean ("Chilean") *P. alba* populations have been colonized by gene flow from "Argentinean" populations, and (2) the distribution of *P. alba* populations is at least partially affected by environmental factors. The phylogeographic ecological analysis may throw light in reference to the origin of *P. alba* populations in the Atacama Desert in Chile.

MATERIALS AND METHODS

Plant material, DNA extraction, amplification and sequencing

The samples of *Prosopis alba* were collected in 12 sites: two in Argentina, three in Bolivia, and seven in Chile. Within each sampling site, 1–6 individuals were collected, and their geographical coordinates were recorded with a handheld GPS (Table 1, Fig. 1A). Taxonomic identification of the specimens was based on Burkart (1976) and carried out by Reneé Fortunato. Herbarium vouchers of the material analyzed here are listed in Bessega *et al.* (2016, 2018, 2021) and deposited at BAB herbarium, INTA, Hurlingham, Buenos Aires, Argentina.

Total genomic DNA was extracted from the leaves of each plant using the DNA easy Plant mini kit (QIAGEN Inc., Valencia, CA, USA) following the instructions of the manufacturer. The non-coding chloroplast region *ndhF-rpL32* was selected for the present study as this region showed the greatest variation among several surveyed loci (*trnQ-rps16, trnH-psbA, rpl32R-ndhF, rpl32F-trnL,*

Table 1. Sampling sites, geographic coordinates, number of sequences (n) and *ndhF-rpl32* haplotypes of *Prosopis alba* from Chile, Argentina and Bolivia.

Sampling site	Country	Code	n	Latitude (S)	Longitude (W)	Haplotypes (N° of individuals)
Zapiga	Chile	ZAPI	1	19° 38’ 3.64’’	69° 57’ 7.51’’	H1 (1)
Tiliviche	Chile	TILI	1	19° 33’ 14.3’’	69° 57’ 11.5’’	H1 (1)
Valle Viejo	Chile	VVJO	3	20° 46’ 8.4’’	69° 34’ 6.79’’	H1 (2) H2 (1)
Quillagua Sur	Chile	QUIS	1	21° 39’ 5.86’’	69° 32’ 2.54’’	H3 (1)
Chiu Chiu	Chile	CHIU	4	22° 19’ 3.81’’	68° 39’ 0.65’’	H1 (2) H5 (1) H6 (1)
Yaye	Chile	YAYE	6	22° 55’ 6.72’’	68° 12’ 7.3’’	H1 (6)
Camar	Chile	CAMA	5	23° 24’ 3.53’’	67° 57’ 4.27’’	H1 (2) H3 (1) H4 (2)
Huajchilla	Bolivia	HUAJ	2	16° 37’ 26.35’’	68° 3’ 17.46’’	H7 (1) H8 (1)
Mecapaca	Bolivia	MECA	2	16° 40’ 2.89’’	68° 0’ 15.42’’	H1 (1) H4 (1)
Tahuapalca	Bolivia	TAHU	1	16° 43’ 16.7’’	67° 52’ 36.57’’	H1 (1)
Campo Duran	Argentina	CDUR	2	22° 10’ 56.64’’	63° 42’ 49.68’’	H1 (1) H4 (1)
Fernandez-Forres	Argentina	FFOR	1	27° 54’ 24’’	63° 54’ 13’’	H9 (1)

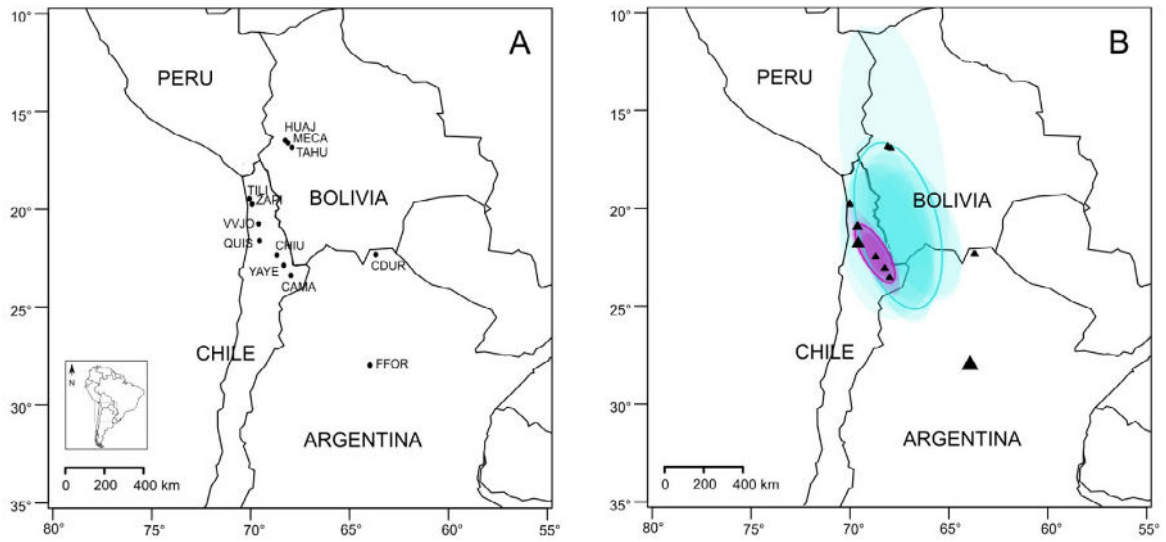

Figure 1. A: Location of the *Prosopis* alba sampling sites from Chile, Bolivia and Argentina in South America; **B**: Results from Bayesian phylogeographic and ecological clustering analysis (BPEC) based on *ndhF-rpl32* sequence data. The simulated contour areas are centered for each population cluster, and the shaded areas show the radius of 50% concentration contours around it. Note: light blue = cluster 1, violet = cluster 2. The triangles sizes represent the root probability of each sample. TILI: Tiliviche, ZAPI: Zapiga, VVJO: Valle Viejo, QUIS: Quillagua Sur, CHIU: Chiu-Chiu, YAYE: Yaye, CAMA: Camar, HUAJ: Huajchilla, MECA: Mecapaca, TAHU: Tahuapalca, CDUR: Campo Duran, FFOR: Fernandez-Forres.

trnD-trnT; Shaw *et al.,* 2007). The *ndhF-rpL32* intergenic region was amplified in 29 *P. alba* individuals, with the primers described by Shaw *et al.* (2007) and successfully used at intraspecific level before in *P. chilens*is (Aguilar *et al.,* 2020). The PCR amplifications were carried out in a 50 ul reaction volume containing 10–30 ng of DNA, 0.6 uM of each primer, 0.2 mM of dNTPs, 0.3 U of Taq DNA polymerase (Invitrogen, Foster City, CA, USA), and 1.5 mM of MgCl2. The amplifications were carried out in a T100 thermal cycler (Life Science Research, BioRad) with a cycling profile of initial denaturation at 94° C for 5 min followed by 35 cycles of 45 s at 94° (denaturation), 45 s at 55° (annealing) and 45 s at 72° C (extension), and a final extension step of 10 min at 72°.

Sequencing was performed by Macrogen Inc. (Seoul, South Korea, http://dna.macrogen.com). Electropherograms were visualized and edited using BioEdit software (Hall, 1999). Sequences were aligned with the multiple alignment option of BioEdit (Hall, 1999) and were adjusted manually. Gaps were not coded, and all the sequences are available upon request (cecib@ege.fcen.uba.ar), and accessible in the Genebank public database.

Data analysis

Genetic diversity, structure, and haplotype network

Considering that an isolation pattern due to physical barriers is expected mainly due to the Andean mountains, diversity among sequences was first quantified in each country by the mean number of Haplotypes (*H*), nucleotide diversity (π) (Nei and Li, 1979), haplotype diversity (*Hd*), the unique allele proportion and private alleles, using *strataG* package (Archer *et al.,* 2016) of R ver. 4.2.1 software (R Development Core Team, 2022).

Genetic structure was evaluated by analysis of molecular variance (AMOVA) using the function *poppr.amova* of the *poppr* package (Kamvar *et al.,* 2014, 2015). The genetic differentiation (Φ_{ST}) was estimated considering countries and BPEC clusters. The significance levels of Φ statistics were obtained with the function *randtest.amova* of the *ade4* package (Chessel *et al.,* 2004; Dray and Dufour, 2007; Dray *et al.,* 2007; Bougeard and Dray, 2018) of R software, with 2,000 permutations. To estimate the relationships between haplotypes, a minimum spanning network was constructed using the function *haploNet* from *pegas* package (Paradis, 2010) using R software.

Correlation between genetic and geographic distance

Pairwise Nei's (1978) genetic distances between populations (*D*) were estimated with the function *dist.genpop* of the package *adegenet* (Jombart, 2008; Jombart and Ahmed, 2011). The possible existence of isolation by distance (IBD) was analyzed by comparing pairwise population geographical distance with nucleotide divergence matrices by Mantel test (with 2,000 permutations) using the package *ade4* (Dray *et al.,* 2007).

Bayesian phylogeographical and ecological clustering (BPEC)

To verify the geographical structure of the populations and the most likely ancestral geographical locations, we analyzed the distribution of the *ndhF-rpL32* intergenic sequence variability by means of a BPEC approach, including geographical and environmental variables as potentially explanatory factors, using the package *BPEC* (Manolopoulou *et al.,* 2016) of R software. The analyses were conducted considering as covariates the geographical coordinates and the following seven environmental variables: (1) altitude, (2) mean temperature of the warmest quarter, (3) mean temperature of coldest quarter, (4) precipitation of driest quarter, (5) precipitation of the warmest quarter, (6) average spring normalized difference vegetation index (*NDVI*), and (7) average summer *NDVI*. *NDVI* data were obtained from NOP (NASA Earth Observations) from 200 m resolution images (https://neo.sci.gsfc.nasa.gov/view.php?datasetId=MOD_NDVI_M). Environmental climatic data were taken from WorldClim v.2 (https://www.worldclim.org/data/worldclim21.html).

The running conditions were 100,000 iterations, saving 10,000 posterior samples, with a parsimony relaxation parameter, *ds*= 0. To avoid bias owing to a scale effect, environmental variables were scaled to mean= 0 and variance= 1. The probable geographical distribution of the clusters identified was plotted with the function *bpec.contourPlot* in R software. For the resulting clusters, genetic diversity parameters, population structure and haplotype network were also estimated in the same way as described above for countries.

Population demographic history

In order to evaluate the population demographic history, we used several proxies: a) Tajima´s *D* (*TD*) (Tajima, 1989) and Fu's F_S (Fu, 1997) statistical tests; b) mismatch distribution plot and raggedness index (Harpending, 1994); c) coalescent Monte Carlo simulations of Markov chains; and d) Approximate Bayesian Computational (ABC) analysis (Beaumont *et al.*, 2002).

TD and F_S were obtained with the functions *tajimasD* and *fusFS* using the *strataG* package (Archer *et al.,* 2016) while the mismatch distribution plot was produced with the function *MMD* of the *pegas* package (Paradis, 2010) using R software. The statistic *rg* was calculated with the specific script described in Vilardi *et al.* (2021). The statistical significance of *rg*, *TD* and F_S was determined by comparison with the neutral expected distribution obtained from coalescent model simulations generated with the function *coal_model* of the *coala* package (Staab and Metzler, 2016) using R software.

The coalescent analysis of the *nadF-rpl32* fragment sequences to determine the most probable time-dependent demographic model was conducted by Monte Carlo simulations of Markov chains (MCMC)

with the *coalescentMCMC* package (Paradis, 2015). Four demographic models (constant, exponential, step, and linear growth) were compared simulating 15,000 phylogenetic trees, from which the last 10,000 were kept (burnin=5,000). Models were compared according to their deviance information criterion (*DIC*) (Spiegelhalter *et al.*, 2002). The time to the most recent common ancestor (TMRCA) was estimated using the function *branching.times* from *ape* package from R software (Paradis *et al.*, 2004; Paradis and Schliep, 2018) assuming ultrametricity.

The ABC analysis procedure (Beaumont *et al.*, 2002) was performed with DIYABC v2.1 (Cornuet *et al.*, 2014) and based on the cpDNA dataset. Five different scenarios were tested with DIYABC considering the samples from the different countries as ancestral possibilities (scenarios 1, 2, and 3; Fig. 2A) and the ones from the BPEC clusters retrieved (scenarios 1 and 2; Fig. 2B). Priors for the different parameters were adjusted after performing 10,000 short simulations (Table S1).

Table S1. Parameters used as prior settings for DIYABC analysis.

Parameter	Distribution	Scenario	
		Minimum	Maximum
By countries			
Effective population size			
N1	Uniform	100	30000
N2	Uniform	100	30000
N3	Uniform	100	30000
Time of event			
t1	Uniform	1	25000
By clusters			
Effective population size			
N1	Uniform	100	30000
N2	Uniform	100	30000
Time of event			
t1	Uniform	1	25000

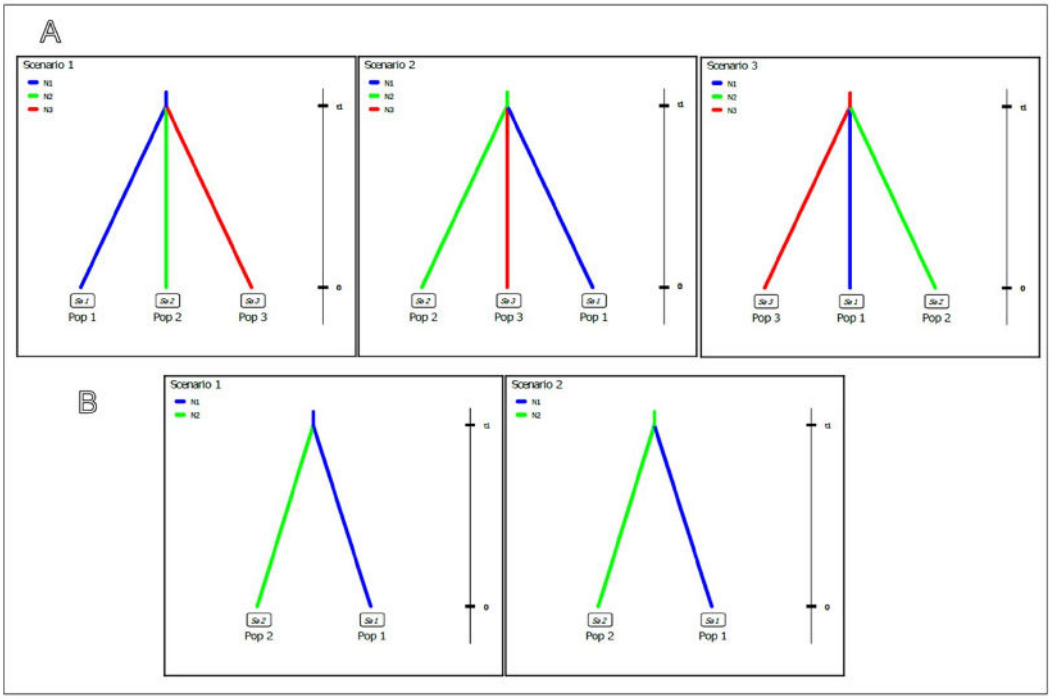

Figure 2. Five evolutionary scenarios (Sc) considered for comparison by ABC approach based on *ndhF-rpl32* sequence data. **A:** Sc1, Sc2 and Sc3 considering as ancestral the samples from different countries. Pop 1= Argentina, Pop 2= Bolivia and Pop 3= Chile. **B:** Sc1 and Sc2 considering as ancestral the samples from the different BPEC clusters. Pop 1= Cluster 1 (East) and Pop 2= Cluster 2 (West).

RESULTS

Genetic diversity, structure, and haplotype network by countries

The cpDNA *nadhF-rpl32* intergenic region was amplified and sequenced in all the *Prosopis alba* samples from Chile, Bolivia and Argentina allowing the recognition of 9 haplotypes (Table 1).Genetic variation was analyzed by country (Table 2). Haplotype diversity (*Hd*) was high in Bolivia and Argentina (0.9-1), and medium in Chile (0.55). The nucleotide diversity (p) was low in all the countries varying among 0.004 and 0.005. In reference to the private alleles, Chile was the country with the highest number of private alleles (4) followed by Bolivia (2) and Argentina (1). As the sample size was very different among countries, the unique allele proportion gives a better idea of the country diversity, being higher in Argentina (1) followed by Bolivia (0.6) and finally by Chile (0.14).

The differentiation among countries (Table 3) was estimated by the Φ_{ST} index that resulted low and not significant (Φ_{ST}=0.014, *P*=0.315). In agreement with this, the AMOVA analysis indicated that most of the variation occurs within countries (98.6%) and only 1.4% of the variation can be detected among countries.

The relationships among the cpDNA haplotypes were represented with a median-Joining network (Fig. 3A). The network shows that the two most frequent alleles (H4 and H1) differ in a single mutation, while the other haplotypes arose by the accumulation of several mutations. The network does not show a geographical association, there are haplotypes that are present in the three countries (H1 and H4) but there are also others that are only present in Chile (H3, H2, H5 and H6), in Bolivia (H7 and H8) and in Argentina (H9).

Correlation between genetic and geographic distance

The correlation between genetic and geographic distances was negative and non-significant (*r*=0.056, *P*=0.3), indicating that genetic differentiation between populations did not fit a spatial pattern as the expected by the isolation-by-distance model.

Bayesian phylogeographical and ecological clustering (BPEC)

The phylogeographical clustering determined that the most probable number of migration events was 1 (*P*=1) yielding two clusters (Fig. 1B) with high posterior probabilities for all the haplotypes (*PP*>0.999) to belong to cluster 1. Each individual was assigned unequivocally to its cluster (*P*=1). The network central haplotypes (H4, H1 and H9) exhibited the highest root node posterior probabilities (*PP*=0.173, 0.170 and 0.155, respectively) being H4 the most probable ancestral one.

The posterior distribution of the covariates showed significant differences between the two clusters for six out of seven environment variables (average summer *NDVI*, average spring *NDVI*, mean temperature of warmest quarter, mean temperature of coldest quarter, precipitation of driest quarter, precipitation of warmest quarter; $P<2x10^{-16}$). Only the altitude differentiation between the clusters was no significant (*t*=-3.43, P=0.00059). Considering the sample here analyzed, the most likely ancestral location was Fernández-Forres (FFOR) in Argentina (*P*=0.155, cluster 1, Fig. 1B).

Genetic diversity, structure and haplotype network by BPEC clusters

When genetic variability distribution was analyzed considering the clusters retrieved by BPEC (Table 2), the cluster 2 (from the west) exhibited less variation

Table 2. Molecular diversity indices considering the sampling sites (by country, A) and the genetic groups retrieved by (by cluster, B) of *Prosopis alba* from Chile, Argentina and Bolivia. H= number of haplotypes/alleles, π = nucleotide diversity, Hd = haplotype diversity, UA = unique alleles proportion, PA= private alleles, SD = standard deviation

	Grouping	Location	Number of Haplotypes (H)	Diversity: Nucleotide diversity (π) (SD)	Diversity: Haplotype diversity (Hd) (SD)	Diversity: Unique alleles proportion (UA)	Diversity: Private alleles (PA)
A	BY COUNTRY						
		CHILE	6	0.00406 (0.00124)	0.55714 (0.124)	0.14	4
		BOLIVIA	4	0.00580 (0.00206)	0.900 (0.161)	0.60	2
		ARGENTINA	3	0.00446 (0.00157)	1.000 (0.272)	1	1
B	BY CLUSTERS						
		CLUSTER 1 (EAST)	7	0.00553 (0.00165)	0.800 (0.083)	0.33	6
		CLUSTER 2 (WEST)	3	0.00275 (0.00103)	0.385 (0.149)	0.07	2
	GLOBAL		9	0.00439 (0.00109)	0.648 (0.096)	0.21	-

according to the haplotype diversity (*Hd*= 0.38 vs. 0.80), private alleles (2 vs. 6) and unique allele proportion (0.07 vs. 0.33).

The differentiation between the BPEC clusters (Table 3) was borderline significant (Φ_{ST} =0.08, P=0.053) and much higher than the differentiation among countries shown above. The partition of the variation was consistent in demonstrating high variation within clusters and low variation, about 8%, among clusters.

In the median-joining network analyzed according to the clusters recovered by *BPEC* (Fig. 3B), H2 and H3 haplotypes were only found in cluster 2, H1 was shared by both clusters and the rest of the haplotypes were located exclusively in cluster 1.

Population demographic history

The first proxy to infer the demographic history of *P. alba* and to detect past population growth was based on Tajima's *D* and Fu's F_S statistical test. Both indices were negative but not significant as they fall inside the $CI_{95\%}$ generated by the coalescent simulations (*TD*=-1.351 [-1.727–1.995], F_S=-2.326 [-3.831–4.622]).

A second approach to evaluate the demographic history was based on the mismatch distribution plot (Fig. 4) and the raggedness index. The mismatch distribution plot did not show a multimodal and ragged shape revealing that the population is not in demographic equilibrium. According to this, the raggedness index estimated under

Table 3. Analysis of molecular variance (AMOVA) and genetic differentiation (ΦST) based on *ndhF-rpl32* sequences of individuals of *Prosopis alba* from Chile, Argentina and Bolivia. Grouping by country (A) and by the retrieved clusters (B).

Grouping	df	Sum of squares	Estimated variance	%	Genetic differentiation
A. By Country					
Among countries	2	4.253	0.027	1.389	Φ_{ST} = 0.014
Within countries	26	50.781	1.953	98.61	(P= 0.315)
Total	28	55.035	1.980	100	
B. Considering BPEC clusters					
Among clusters	1	4.368	0.172	8.397	Φ_{ST} = 0.084
Within clusters	27	50.667	1.877	91.603	(P= 0.053)
Total	28	55.034	2.049	100	

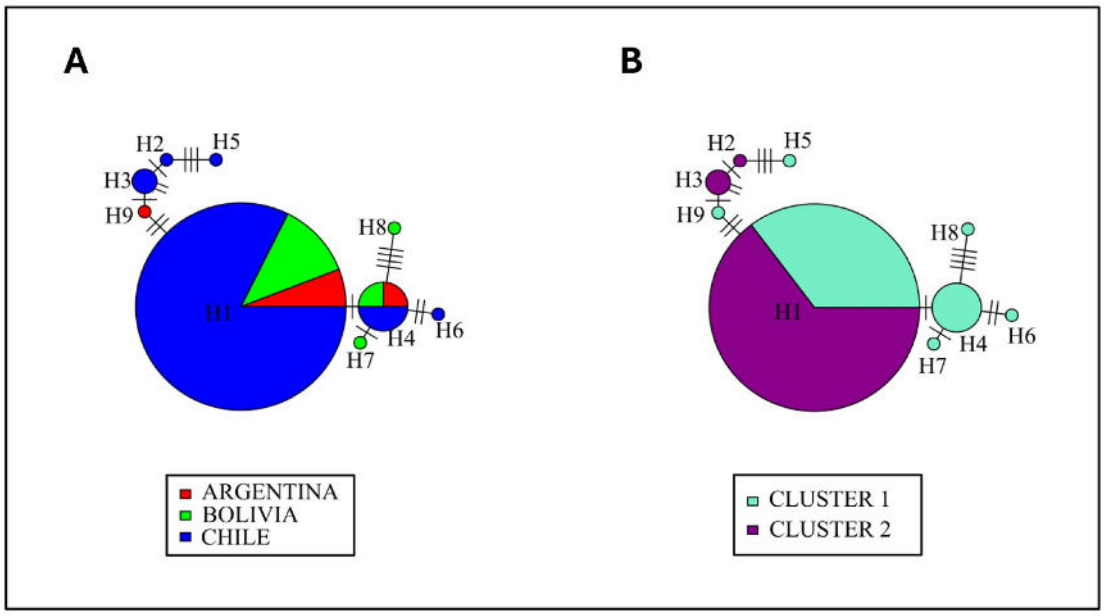

Figure 3. Haplotype network of the *Prosopis alba* based on *ndhF-rpl32* haplotypes. **A:** circles are colored according to the country of haplotype; **B:** circles are colored according to the genetic clusters obtained by BPEC. The size of the circles is proportional to the frequencies of each haplotype across all populations. Lines represent the mutational steps between haplotype sequences.

the demographic expansion model was non-significant (*rg*=0.021 [0.017–0.342]).

The third approach performed was based on coalescent trees simulated by MCMC procedures. The best fitting model was the linear one (*DIC*=-132.01). According to this, model theta (θ) grew from 1.15 x 10^{-16} to 1.98 x 10^{-2} and the time to the most recent ancestor (TMRA) was 0.0072 units of time before present (Fig. 5).

The same analysis was also applied considering only the samples included by BPEC in the cluster 2. The fittest model was the linear one (DIC=-445.09), with θ growing from 6.8 x 10^{-3} to 8.27 x 10^{-3} and a TMRA=0.0042.

Finally, ABC demographic inference analysis showed that among scenarios 1, 2 and 3, the scenario 1, that assumes the Argentinean populations as the ancestral ones, is the most likely scenario that would underline the observed genetic population divergence considering the direct and the logistic approach (Fig. 6A,B) with a posterior probability [$CI_{95\%}$] of 0.362 [0.0000-0.7832] and 0.3887 [0.3550-0.4223] respectively. When comparing scenarios 1 and 2 considering the BPEC retrieved clusters, the scenario 1 that considers the cluster 1 (east one, light blue) as ancestral, yielded a higher posterior probability [$CI_{95\%}$] for the direct (0.512 [0.0739-0.9501]) and the logistic regression (0.5655 [0.5111-0.6199]) approaches, suggesting that it is the best supported one (Fig. 6C,D).

DISCUSSION

In this study we analyzed from a phylogeographic standpoint samples from different *P. alba* populations from southern South-America in order to contribute to elucidate the evolutionary history of this important forest resource in arid and semiarid regions. We analyzed the genetic diversity, population structure, demographic history and Bayesian phylogeographical and ecological clustering (BPEC) based on the *nadhF-rpl32* intergenic spacer. We found evidence of two genetic lineages and signatures of demographic expansions.

Genetic diversity was analyzed considering the country distribution of the populations and the BPEC retrieved clusters. Genetic diversity in Chile and in cluster 2, the western one, was lower than in the remaining countries and cluster according to the haplotype diversity (*Hd*) and unique allele proportions estimators. Genetic diversity is expected to be lower in recent colonized areas in comparison with the original sites and the high presence of private haplotypes can be considered an indicator of long-term residency given the low mutation rate of organellar DNA (Wares *et al.,* 2002; Torre *et al.,* 2022). Additionally, the combination of high haplotype diversity and low nucleotide diversity, as observed in Argentina and Bolivia, can be signature of a rapid demographic expansion from a small effective population size (Avise, 2000; Joshi *et al.*, 2013). In accordance with this, we found signals of demographic expansion based on the different considered proxies. First, the negative Tajima's D (*TD*) and Fu's (*Fs*) estimates showed an excess of low frequencies polymorphisms relative to expectation, suggesting size expansion. Second, the mismatch distribution plot was unimodal and consequently the smoothness of the distribution (*rg* index) resulted no significant, and third, the coalescent analyses indicated that the linear model was the one that best fits the demographical populational expansion. The genetic diversity parameters and the signals of population expansion obtained may be interpreted in accordance with our first hypothesis that suggests that the Chilean populations have been colonized by populations from the east through trans-Andean migration. Finally, the comparison of several divergent scenarios by means of ABC analyses is in accordance with this result as the cases where the east cluster samples and the

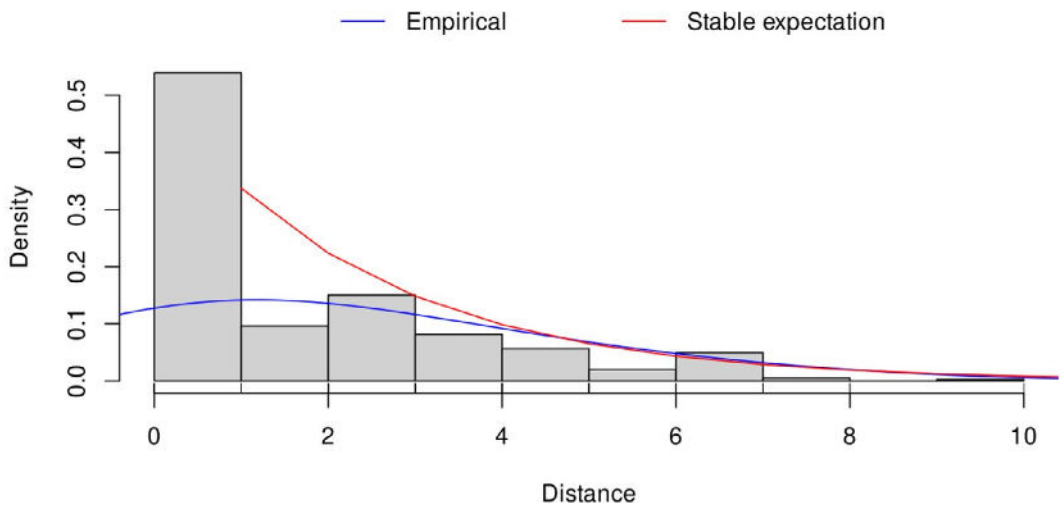

Figure 4. Pairwise mismatch distribution plot obtained based on the *ndhf-rpl32* sequences from *Prosopis alba.*

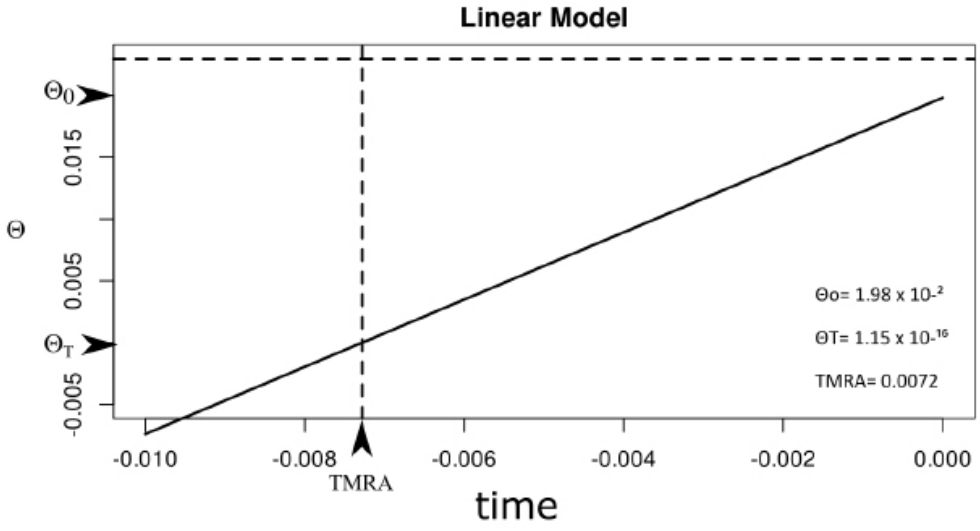

Figure 5. Expected variation of θ through time for the linear growing model. The predicted line is compared with the constant model (horizontal dashed line) and the vertical dashed line represents the time to the most recent common ancestor (*TMRCA*).

Argentinean ones were postulated as ancestral, yielded the higher posterior probabilities. This hypothesis is partially compatible with the proposal made by Palacios and Brizuela (2005). These authors suggested that morphological resemblance between some species of *algarrobo* trees found in Quilmes (Tucumán), Tolombón (Salta), Belén, Bolsón de Fiambalá and Chaschuil river (Catamarca) in Argentina and in Quillagua (Río Loa) and Canchones (Pampa del Tamarugal) in Chile, could be due to the introduction from east to west in the midst of a process of domestication in the Atacama Desert during pre-Hispanic times. This possibility was also suggested by McRostie *et al.* (2017) that provided evidence that the *algarrobo* species were not native to the Atacama Desert of Chile. It was suggested that in Atacama, the spread of species of section *Algarobia* was contemporaneous with the introduction of other crops by humans (McRostie, 2014; McRostie *et al.*, 2017). Based on AMS radiocarbon dates from archaeobotanical and palaeoecological records and settlement patterns in the Atacama Desert, it was suggested that the introduction of these trees occurred late in the Holocene (ca. 3,000 yr BP or later), and that the most likely vectors were humans. Moreover, Ugalde *et al.* (2021) analyzed the presence and abundance of *Prosopis* species in Atacama considering different cultural periods over the last 13,000 years and conclude that legume trees (i.e. chañar and algarrobo), probably introduced later in the Holocene, became a key resource from the Formative Period in the Pampa del Tamarugal, Loa and Salar de Atacama basins, linked with the Chaco region and the Northwest of Argentina. Seeds may have been circulated as part of intense mobility systems (caravans), which were rooted in traditional ways of life (McRostie *et al.*, 2022).

This aspect could be discussed considering the genetic differentiation estimations obtained in *P. alba* samples from different countries based on the *ndhF-rpL32* intergenic sequences. Our phylogeographic survey showed that only 1.4% of the genetic differentiation was found among countries. This percentage -which represents low estimates of genetic differentiation- could be indicative of a recent differentiation between the populations, compatible with the relative recent colonization suggested by Palacios and Brizuela (2005) and McRostie *et al.* (2017). The lack of detection of a clear spatial pattern goes in the same direction as no significant correlation was detected between the genetic and geographic distance matrices. However, the differentiation reaches up to 8% and the Φ_{ST} estimate results borderline significant when BPEC groups are considered. This result should be taken into account with greater certainty since climatic variables were considered for obtaining the clusters through BPEC. It is possible to consider that environmental variables such as temperature, precipitation and *NDVI* may partially explain the structure detected in *P. alba.* The Atacama area (that mainly overlaps with cluster 2, west) is one of the most arid deserts on Earth that has its own unique climatic characteristics. Two main causes may be contributing to explain the extreme aridity of the region. The cold water of the Humboldt Current, running parallel to the Chilean and southern Peruvian coasts and preventing precipitation in the coastal areas; and the occurrence of the Andes cord, that produces a rain-shadow effect, blocking moisture from the Amazon basin. As expected by the second hypothesis proposed, the environmental differences may be contributing to explain the geographical patterns of genetic variation retrieved.

The lack of association observed in the median joining network of haplotypes considering the countries (Fig. 3A) can also be interpreted considering the environmental variables and association expected in BPEC network under such scenario. However, in the network recovered by BPEC analysis (Fig. 3B) inconsistencies were also found. The eastern cluster (cluster 1) presents more exclusive haplotypes, suggesting indirectly that cluster 2 (from west) might be considered as derived. Indeed, our BPEC analysis lends some weight to this conclusion, since the most likely ancestral location detected, considering the sampling here conducted, was in the Chaqueña region in Argentina. Gene flow from Argentina to Chile has also been suggested for *P. chilensis*, a species related to *P. alba*, based on cpDNA sequence data, since Bolivian and Argentinean haplotypes were found in Chilean samples. A possible colonization (or introduction) of *P. chilensis* from the Bolivian Chaco and Argentinian Monte to the Chilean Matorral was proposed (Aguilar *et al.*, 2020). The complex pattern here found in *P. alba* is somehow inconsistent with a single migration event, as the haplotypes H2 and H3 are absent in the eastern populations, however this may be attributed to genetic drift or sampling error.

The low level of genetic differentiation among countries and clusters can be interpreted as signals of recent differentiation events although the low substitution rate of cpDNA and the long reproductive cycles of *Prosopis* species, should be considered. For woody plants with long reproductive cycle, their cpDNA mutation rate is much lower than in other species (Gaut *et al.*, 1996; Yan *et al.*, 2018). In oaks (*Quercus* spp.), for example, where seedlings need 3-4 years to produce their first acorns, the substitution rate resulted very low (0.19-0.96 x 10^{-9} s/s/y), even below the average values reported for non-coding regions in other angiosperm lineages (1.2-1.7 x 10^{-9} s/s/y). Additionally, in *Cercidium*, the average substitution rate reported is also much lower (0.318 x 10^{-9} s/s/y) than the average values generally reported for non-coding regions of the chloroplast genome, but also consistent with the notion that woody taxa should have slower rates of molecular evolution (Qi *et al.*, 2012).

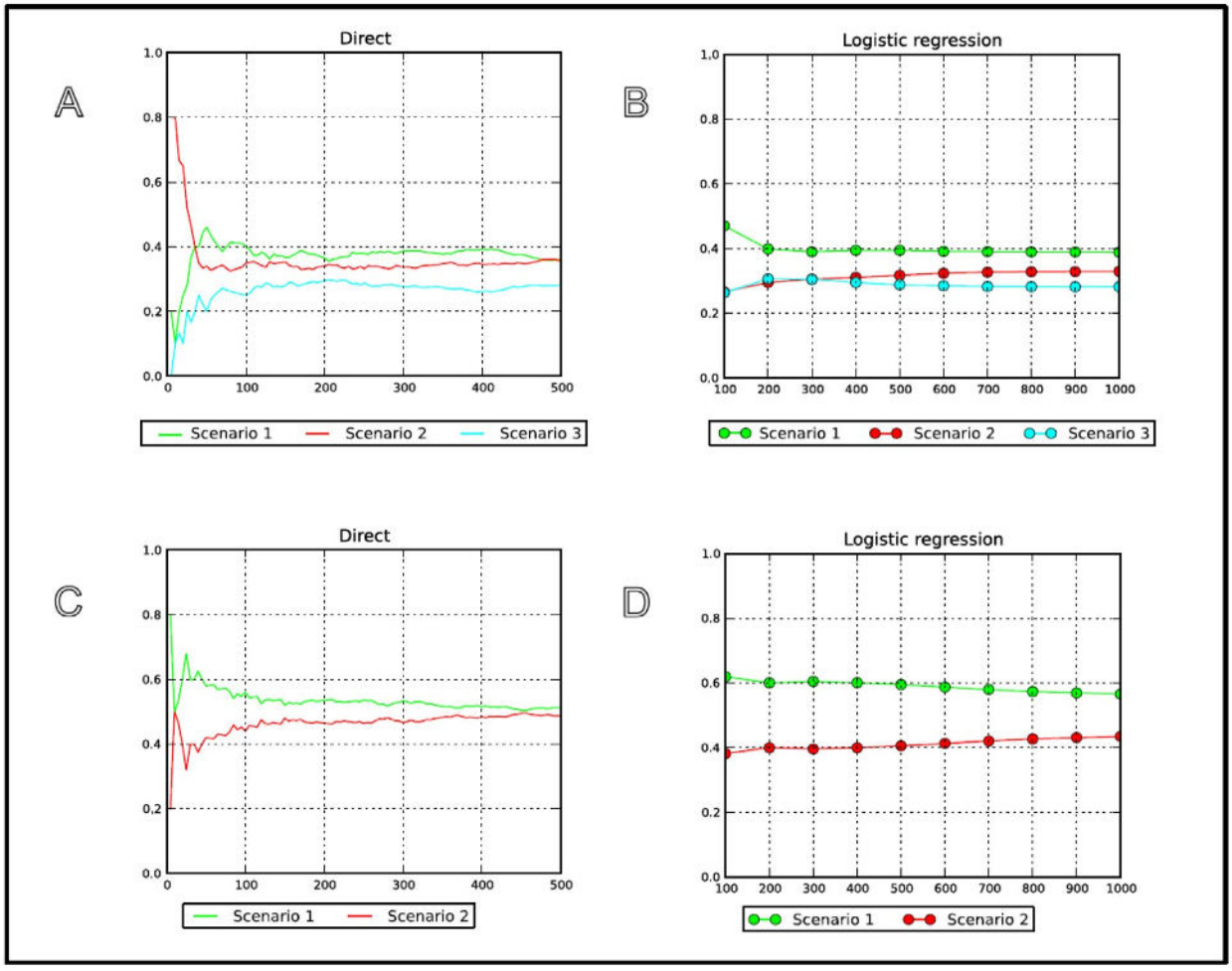

Figure 6. Plots showing fitness of competing scenarios based on direct estimates and logistic regression simulated in DIYABC based on *ndhF-rpl32* sequence data. **A and B:** comparison between scenarios 1, 2 and 3 by country. **C and D:** comparison among scenarios 1 and 2 by BPEC clusters.

The global nucleotide diversity estimated here for *ndhF-rpL32* was π =0.004, which may imply a divergence time among haplotypes of 4 to 6.2 Mya if we consider the mean substitution rate in oak and *Cercidium* one respectively. These time estimates are consistent with that obtained using the same cpDNA region in *P. chilensis* populations from Argentina and Chile, where the haplotype divergence in the phylogroups detected in Argentina was proposed in 5.22 Mya and was associated with the Paranaean Sea marine transgression (Aguilar *et al.,* 2020).

Luebert (2011) suggests that disjunct elements in arid and semi-arid areas on both sides of The Andes occurred through the mountain uprising, which created a barrier to the dispersal, promoting differentiation of species on both slopes. The timeline of the Andean uprising is currently a matter of controversy, but it seems to be agreed that the current elevation is early Pliocene (Luebert, 2011). The lineages distributed in the Desert of Atacama, the Chaco and the tropical and/or Mediterranean Andes could be explained considering that Andean areas were colonized after the vicariance from the basal areas of one or both sides of the mountain range. Long distance for the latter cases, via trans-Andean corridors, could not be ruled out *a priori*. Indeed, in *Prosopis* it was suggested that successive vicariance events split the ancestral wide area, and long-distance dispersal episodes led to recolonizations from North to South America, and vice versa (Bessega *et al.,* 2006). However, the estimated population expansion times estimated are much lower. The coalescent MCMC analysis yielded a TMRA value (Θ= 0.007) that suggest that 7,000 generations have occurred since the most recent common ancestor for the haplotypes sampled in the present analysis. As *Prosopis* trees may have long overlapping generations from near 50 to ~100 years, and it was seen and described that the first bloom and pods production occurs near the third and fourth year of growth (Castillo and Tarnowski, 2011), populations may have been expanding since about 28,000 years bp. When the analysis is restricted to the sequences belonging to cluster 2 (west), TMRCA estimate drops to near a half (Θ= 0.004), suggesting that the expansion time in Chilean territory would have started around 16,000 years bp.

Human-mediated migration may have played a substantial contribution to current distribution of *Prosopis* populations on both sides of The Andes given the traditional link between this resource and the original communities (Pasiecznik *et al.,* 2001; Giovannetti *et al.,* 2008; Rivera and Dodd, 2013; Uribe *et al.,* 2020). However, based on the present results, *P. alba* may have started to disperse to the north and west from the Chaqueña Region in Argentina much earlier, and the movement may have been mediated by natural trans-Andean dispersers such as guanacos, small mammals and/or birds.

Although the analysis here performed is only based on the cpDNA *nadhF-rpl32* intergenic region and can be consider weak, it contributes in the *P. alba* (*algarrobo)* historical reconstruction throwing light on the trans-Andean movement, considering direction and time. A more intensive sampling of the South-American populations in combination with more cpDNA and nuclear genotyping is needed to give a more complete picture of the phylogeography of *P. alba*. Up to our knowledge, this is the first phylogeographic study in *P. alba* which gives strong evidence of the introduction of *P. alba* in the Atacama Desert pointing the Chaqueña region in Argentina as the source area.

REFERENCES

Aguilar D.L., Acosta M.C., Baranzelli M.C., Sérsic A.N., Delatorre-Herrera J., Verga A., Cosacov A. (2020) Ecophylogeography of the disjunct South American xerophytic tree species *Prosopis chilensis* (Fabaceae). Biol. J. of Linn. Soc. 129(4): 793-809. doi: 10.1093/biolinnean/blaa006

Archer F.I., Adams P.E., Schneiders, B.B. (2016) *stratag*: An r package for manipulating, summarizing and analyzing population genetic data. Mol. Ecol. Resour. 17: 5–11.

Avise J.C. (2000) Phylogeography. The history and formation of species. Cambridge: Harvard University Press, Cambridge, Massachusetts, United States.

Avise J.C., Arnold J., Ball R.M., Bermingham E., Lamb T., Neigel J.E., Reeb C.A., Saunders N.C. (1987) Intraspecific Phylogeography: The Mitochondrial DNA Bridge Between Population Genetics and Systematics. Annu. Rev. Ecol. Syst. 18:489-522. doi: 10.1146/annurev.es.18.110187.002421

Barros S. (2010) El género *Prosopis*, valioso recurso forestal de las zonas áridas y semiáridas de América, Asia y Africa. Ciencia e Investigación Forestal 16(1): 91–128.

Beaumont M.A., Zhang W., Balding D.J. (2002) Approximate Bayesian computation in population genetics. Genetics 162: 2025–2035. doi: 10.1093/genetics/162.4.2025

Bessega C., Pometti C.L., Ewens M., Saidman B.O., Vilardi J.C. (2016) Fine-scale spatial genetic structure analysis in two Argentine populations of *Prosopis alba* (Mimosoideae) with different levels of ecological disturbance. Eur. J. For. Res. 135(3): 495–505.

Bessega C., Pometti C., Fortunato R., Greene F., Santoro C.M., McRostie V. (2021) Genetic studies of various *Prosopis* species (Leguminosae, Section *Algarobia*) co-occurring in oases of the Atacama Desert (northern Chile). Ecol. Evol. 11(5): 2375–2390. doi: 10.1002/ece3.7212

Bessega C., Pometti C., López R.P., Larrea-Alcázar D., Fortunato R.H., Saidman B.O., Vilardi J.C. (2018) Genetic diversity and differentiation among *Prosopis alba* (Leguminosae) populations from dry valleys of Bolivia with different levels of human disturbance and altitude. Ecol. Evol. 8(22): 11309–11321. doi: 10.1002/ece3.4610

Bessega C., Vilardi J.C., Saidman B.O. (2006) Genetic relationships among American species of the genus *Prosopis* (Mimosoideae, Leguminosae) inferred from ITS sequences: evidence for long-distance dispersal. J. Biogeogr. 33: 1905–1915.

Bougeard S., Dray S. (2018) Supervised multiblock analysis in R with the ade4 package. J. Stat. Softw. 86(1): 1–17. doi: 10.18637/ jss.v086.i01

Burkart A. (1976) A monograph of the genus *Prosopis* (Leguminosae subfam. Mimosoideae). J. Arnold Arboretum 57(3–4): 217–249; 450–525.

Capparelli A. (2007) Los productos alimenticios derivados de *Prosopis chilensis* (Mol.) Stuntz y *P. flexuosa* DC., Fabaceae, en la vida cotidiana de los habitantes del NOA y su paralelismo con el algarrobo europeo. Kurtz. 33(1): 1–19.

Castillo E.M., Tarnowski E. (2011) Cultivo experimental de *Prosopis chilensis* en el Pedemonte de las Yungas. https://inta.gob.ar/sites/default/files/script-tmp-prosopis_chilensis.pdf

Chessel D., Dufour A.B., Thioulouse J. (2004) The ade4 package – I: One-table methods. R News 4(1): 5–10.

Cony M.A. (1996) Genetic variability in *Prosopis flexuosa* D.C. a native tree of Monte phytogeography province, Argentina. For. Ecol. Manag. 87: 41–49.

Cornuet J.M., Pudlo P., Veyssier J., Dehne-Garcia A., Gautier M., Leblois R., Marin J.M., Estoup A. (2014) DIYABC v2.0: A software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. Bioinformatics 30: 1187–1189.

Dray S., Dufour A. (2007) The ade4 package: Implementing the duality diagram for ecologists. J. Stat. Softw. 22: 1–20. doi: 10.18637/ jss.v022.i04

Dray S., Dufour A., Chessel D. (2007) The *ade4* package – II: two-table and K-table methods. R News 7: 47–52.

Dunai T.J., Gonzalez Lopez G.A., Juez-Larre J. (2005) Oligocene–Miocene age of aridity in the Atacama Desert revealed by exposure dating of erosion-sensitive landforms. Geology 33(4): 321–324. doi: 10.1130/G21184.1

Fu Y.X. (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics 147: 915–925.

Gaut B.S., Morton B.R., McCaig B.C., Clegg M.T. (1996) Substitution rate comparisons between grasses and palms: synonymous rate differences at the nuclear gene *Adh* parallel rate differences at the plastid gene *rbcL*. P. Natl. Acad. Sci. (PNAS), USA. 93:10274–10279. doi: 10.1073/pnas.93.19.10274

Giovannetti M., Lema V., Bartoli C., Capparelli A. (2008) Starch grain characterization of *Prosopis chilensis* (Mol.) Stuntz and *P. flexuosa* DC, and the analysis of their archaeological remains in Andean South America. J. Archaeol. Sci. 35: 2973-2985. doi: 10.1016/j.jas.2008.06.009

Hall T.A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acid. S. 41: 95–98.

Harpending H.C. (1994) Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. Hum. Biol. 66: 591–600.

Hughes C.E., Ringelberg J.J., Lewis G.P., Catalano S.A. (2022) Disintegration of the genus *Prosopis* L. (Leguminosae, Caesalpinioideae, Mimosoid clade). PhytoKeys 205: 147-189. doi: 103897. phytokeys.205.75379

Jombart T. (2008) Adegenet: A R package for the multivariate analysis of genetic markers. Bioinformatics 24:1403–1405. doi:10.1093/bioinforma tics/btn129

Jombart T., Ahmed I. (2011) Adegenet 1.3-1: New tools for the analysis of genome wide SNP data. Bioinformatics 27(21): 3070–3071.

Joshi J., Salar R.K., Banerjee P., Upasna S., Tantia M.S., Vijh R.K. (2013) Genetic Variation and Phylogenetic Relationships of Indian Buffaloes of Uttar Pradesh. Asian-Australas. J. Anim. Sci. 26 (9):1229-1236. doi: 10.5713/ajas.2012.12669

Kamvar Z.N., Brooks J.C., Grünwald N.J. (2015) Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. Front. Genetics 6: 208.

Kamvar Z.N., Tabima J.F., Grünwald N.J. (2014) *Poppr*: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ 2: e281.

Luebert F. (2011) Hacia una fotogeografía histórica del Desierto de Atacama. Rev. Geogr. Norte Gd. 50: 105-133.

Manolopoulou I., Hille A., Emerson B. (2016) BPEC: an R package for Bayesian phylogeographic and ecological clustering. arXiv preprint arXiv:1604.01617.

McRostie V.B. (2014) Arboricultura y Silvopastoralismo en el Período Formativo (1.400 a.c.-500 D.C.) de la cuenca del Salar de Atacama. Chungara (Arica) 46(4): 543-557. doi: 10.4067/S0717-73562014000400002

McRostie V.B., Babot P., Calás E., Gayó E., Gallardo F., Godoy-Aguirre C., Labarca R., Latorre C., Núñez L., Ojeda K., Santoro C.M., Valenzuela D. (2022) Silvopastoralism and the shaping of forest patches in the Atacama Desert during the Formative Period (ca. 3000–1500 years BP). The Holocene 32(12): 1492–1502. doi: 10.1177/09596836221122636

McRostie V.B., Gayó E.M., Santoro C.M., De Pol-Holz R., Latorre C. (2017) The pre-Columbian introduction and dispersal of Algarrobo (*Prosopis*, Section Algarobia) in the Atacama Desert of northern Chile. PLoS ONE 12(7): e0181759. doi: 10.1371/journal.pone.0181759

Nei M. (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 23: 341–369.

Nei M., Li W.H. (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc. Natl. Acad. Sci. USA 76: 5269–5273.

Palacios R.A., Brizuela M.M. (2005) Fabaceae, parte 13. Subfam. II. Mimosoideae, parte 4. Tribu VI. Mimoseae, parte B. *Prosopis* L. Flora Fanerogámica Argentina, 92: 3–25.

Paradis E. (2010) *pegas*: an R package for population genetics with an integrated–modular approach. Bioinformatics 26: 419–420.

Paradis E. (2015) *coalescentMCMC*: MCMC algorithms for the coalescent. R package version 0.4-1. https://CRAN.R-project.org/package=coalescentMCMC (accessed March 2023).

Paradis E., Schliep K. (2018) *ape 5.0*: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 35: 526–528.

Paradis E., Claude J., Strimmer K. (2004) *APE*: analyses of phylogenetics and evolution in R language. Bioinformatics 20: 289-290.

Pasiecznik N., Felker P., Harris P.J.C., Harsh L.N., Cruz J.C., Tewari K., Cadoret K., Maldonado L.J. (2001) The *Prosopis juliflora-Prosopis pallida* complex: a monograph. HDRA, Coventry, UK.

Pometti C.L., Pizzo B., Brunetti M., Macchioni N., Ewens M., Saidman B.O. (2009) Argentinean native wood species: Physical and mechanical characterization of some *Prosopis* species and *Acacia aroma* (Leguminosae, Mimosoideae). Bioresource Technol. 100(6):1999-2004.

Qi X., Chen Ch., Comes H.P., Sakaguchi Sh., Liu Y., Tanaka N., Sakio H., Qiu Y. (2012) Molecular data and ecological niche modelling reveal a highly dynamic evolutionary history of the East Asian Tertiary relict *Cercidiphyllum* (Cercidiphyllaceae). New Phytol. doi: 10.1111/j.1469-8137.2012.04242.x

R Development Core Team (2022) R: A Language and Environment for Statistical Computing. R Foundation for Statistical computing. http://www.r-project.org/ (accessed January 2023)

Rico Y., León-Tapia M.Á., Zurita-Solís M., Rodríguez-Gómez F., Vásquez-Morales S.G. (2021) Influence of Pleistocene climatic oscillations on the phylogeography and demographic history of

endemic vulnerable trees (section Magnolia) of the Tropical Montane Cloud Forest in Mexico. PeerJ 9:e12181. doi: 10.7717/peerj.12181

Rivera M., Dodd J. (2013) Domesticando el Desierto: medio ambiente y ocupaciones humanas en Ramaditas, Desierto de Atacama. Diálogo Andino 41: 45-60. doi: 10.4067/S0719-26812013000100004

Roig F.A. (1993) Aportes a la Etnobotánica del Género *Prosopis*. In: Iadiza (Ed.), Contribuciones Mendocinas a la quinta Reunión Regional para América Latina y el Caribe de la Red de Forestación del CIID. Conservación y mejoramiento de especies del género *Prosopis*. Unidades de Botánica y Fisiología Vegetal (IADIZA) Editores, Mendoza, pp. 99-119.

Shaw J., Edgar B., Lickey E.B., Schilling E.E., Small R. (2007) Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in Angiosperms: the tortoise and the hare III. Am. J. Bot. 94: 275–288.

Spiegelhalter D.J., Best N.G., Carlin B.R., van der Linde A. (2002) Bayesian measures of model complexity and fit. J. R. Stat. Soc., Series B 64: 583–616.

Staab P.R., Metzler D. (2016) *coala*: an R framework for coalescent simulation. Bioinformatics 32: 1903–1904.

Tajima F. (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123: 585–595.

Torre S., Sebastiani F., Burbui G., Pecori F., Pepori A.L., Passeri I., Ghelardini L., Selvaggi A., Santini A. (2022) Novel Insights Into Refugia at the Southern Margin of the Distribution Range of the Endangered Species *Ulmus laevis*. Front. Plant Sci. 13: 826158. doi:10.3389/fpls.2022.826158

Turland N.J., Wiersema J.H., Barrie F.R., Greuter W., Hawksworth D.L., Herendeen P.S., Knapp S., Kusber W.-H., Li D.-Z., Marhold K., May T.W., McNeill J., Monro A.M., Prado J., Price M.J., Smith G.F. (2018) International Code of Nomenclature for algae, fungi, and plants (Shenzhen Code) adopted by the Nineteenth International Botanical Congress Shenzhen, China, July 2017. Koeltz Botanical Books, Germany. doi: 10.12705/Code.2018

Ugalde P.C., McRostie V.B., Gayo E.M., García M., Latorre C., Santoro C.M. (2021) 13,000 years of sociocultural plant use in the Atacama Desert of northern Chile. Veg. Hist. Archaeobot. 30: 213–230.

Uribe M., Angelo D., Capriles J., Castro V., De Porras M.E., García M., Gayo E., González J., Herrera M.J., Izaurieta R., Maldonado A., Mandakovic V., McRostie V., Razeto J., Santana F., Santoro C., Valenzuela J., Vidal A. (2020) El Formativo en Tarapacá (3000-1000 aP): Arqueología, naturaleza y cultura en la Pampa del Tamarugal, Desierto de Atacama, norte de Chile. Lat. Am. Antiq. 31(1): 81–102.

Vilardi J.C., Freilij D., Ferreyra L.I., Gómez-Cendra P. (2021) Ecological phylogeography and coalescent models suggest a linear population expansion of *Anastrepha fraterculus* (Diptera: Tephritidae) in southern South America, Biol. J. Linn. Soc. 133(3): 779–795. doi: 10.1093/biolinnean/blab029

Wares J.P., Goldwater D.S., Kong B.Y., Cunningham C.W. (2002) Refuting a controversial case of a human-mediated marine species introduction. Ecol. Lett. 5:577–584. doi: 10.1046/J.1461-0248.2002.00359.X

Yan M., Xiong Y., Liu R., Deng M., Songm J. (2018) The Application and Limitation of Universal Chloroplast Markers in Discriminating East Asian Evergreen Oaks. Front. Plant Sci. 9:569. doi:10.3389/fpls.2018.00569

FINANCIAL SUPPORT

This research was funded by Agencia Nacional de Investigación y Desarrollo, ANID, Chile (projetc FONDECYT 11180805 given to V McR), Núcleo Milenio, AFOREST (projetc NCS2022-24 given to V McR), Agencia Nacional de Promoción Científica y Tecnológica, ANPCYT, Argentina (project PICT 2020-1402 given to CB), and Universidad de Buenos Aires, Argentina, UBA (project UBACYT 20020190200106BA given to CB).

AUTHOR CONTRIBUTION

CB: data, conceptualization, formal analysis, funding acquisition, investigation, writing original draft; **CP:** formal analysis, methodology; **RF:** data curation; **BOS:** conceptualization, formal analysis, investigation, writing original draft; **CS:** funding acquisition, supervision; **V McR:** funding acquisition, investigation, supervision; **JCV:** conceptualization, formal analysis, investigation, writing original draft.