Received: 4 June 2019 | Revised: 5 December 2019 | Accepted: 13 December 2019

DOI: 10.1111/jbi.13808

RESEARCH PAPER

Journal of Biogeography | WILEY

Colonization history of the Canary Islands endemic *Lavatera acerifolia*, (Malvaceae) unveiled with genotyping-by-sequencing data and niche modelling

Irene Villa-Machío | **Alejandro G. Fernández de Castro** | **Javier Fuertes-Aguilar** | **Gonzalo Nieto Feliner**

Real Jardín Botánico, CSIC, Madrid, Spain

Correspondence
Irene Villa-Machío, Real Jardín Botánico, CSIC, Plaza de Murillo 2, 28014 Madrid, Spain.
Email: irvilma0@gmail.com

Funding information
Spanish ministry of Science, Technology and Innovation, Grant/Award Number: CGL2010-16138; Spanish ministry of Economy and Competitivity, Grant/Award Number: CGL2013-49097-C2-1-P and CGL2017-88500-P (AEI/FEDER, UE); FPI fellowship, Grant/Award Number: BES-2011-049852 "(co-financed by European Social Fund)".

Handling Editor: Hanno Schaefer

Abstract

Aim: Differentiation of island lineages from mainland relatives and radiation after initial colonization are two important processes generating island diversity. Both of these processes are influenced by colonization dynamics and contemporary connections between island species and their source populations. The classic island progression rule model, that is dispersal from mainland to older islands and subsequently to younger islands, can be considered a null hypothesis, particularly for islands that are geographically aligned perpendicular to the mainland coast with ages inversely proportional to distance from the mainland. Alterations to this pattern have been reported, particularly in archipelagos that are geographically closely adjacent to mainland source populations. Here we aim to integrate genomic and environmental niche data to infer the colonization history of a Canary endemic species and to understand its current diversity patterns.

Location: Canary Islands.

Taxon: *Lavatera acerifolia* (Malvaceae).

Methods: We used high-throughput genotyping-by-sequencing (GBS) combined with species distribution modelling (SDM) projected onto past conditions. Genetic structure (clustering methods), relatedness (coalescent and ML trees), nucleotide diversity and differentiation (population genetics) were assessed based on SNPs obtained from three alternative bioinformatics pipelines. The influence of environmental variables over time was assessed with a generalized linear model in which the response variable was amount of heterozygous sites per individual.

Results: Four genetic groups were identified arranged along a longitudinal gradient, and the earliest diverging coincides with the older, and easternmost, islands (Lanzarote and Fuerteventura). Genetic diversity is reduced in the westernmost islands, which are more distant from the mainland, host few populations and yet apparently offer more suitable habitats.

Main conclusions: The inferred colonization scenario generally fits the progression rule model, but suggests a more complex pattern for the central islands. For the westernmost islands, the contrast between high availability of suitable habitats and reduced genetic diversity and number of populations suggests a colonization front

moving at a slow pace, rather than local extinctions, as an explanation for the scarcity of populations in those islands. Historical projections of SDM lend support to this interpretation.

KEYWORDS

Canary Islands, colonization, genomics, genotyping-by-sequencing, polyploid, SDM

1 | INTRODUCTION

Impressive amounts of empirical data have been gathered on island biogeography during the last few decades (Patiño et al., 2017; Whittaker, Fernández-Palacios, Matthews, Borregaard, & Triantis, 2017). Past and contemporary relationships between mainland and island lineages and colonization history play important roles in determining current diversity patterns in islands, since they influence differentiation of island lineages from mainland relatives and radiation within the islands. The colonization process—where, when and how many times islands become colonized from mainland source populations—sets the foundation for the evolution of diversity patterns in island lineages (Díaz-Pérez, Sequeira, Santos-Guerra, & Catalán, 2008; García-Olivares et al., 2017; Givnish et al., 2008; Parent, Caccone, & Petren, 2008; Rumeu et al., 2011). A step-wise colonization pattern from the oldest islands to the youngest ones, following an 'island progression rule' (Juan, Emerson, Oromí, & Hewitt, 2000; Wagner & Funk, 1995), is the most classic model. This includes dispersal from the mainland to older islands and subsequently to younger islands. For islands with emerging ages that are inversely proportional to the distance from the mainland, chronology adds to relative proximity to the source areas. Accordingly, in archipelagos with these characteristics, such as the Canary Islands (Canaries), the island progression model may be considered a null hypothesis to test with empirical studies. This model implies a decrease in genetic diversity within populations as the distance to mainland increases, due to successive founder events and dispersal limitations from the source areas (García-Verdugo et al., 2015), although large effective population sizes and deep coalescence can counteract such patterns (Pillon et al., 2013). However, a number of studies have highlighted the possibility of more complex scenarios than a linear progression, involving recurrent island colonization or back-colonization and local extinction. Understanding these possibilities is important for accurate interpretations of evolutionary history as well as for designing conservation policies (Caujapé-Castells, 2011; García-Verdugo, Caujapé-Castells, Illera, et al., 2019). Consequently, a careful examination of the colonization scenario is needed even in cases where the chronological and spatial considerations favour a progression rule model.

The Canaries, composed of seven main islands that emerged sequentially from east to west within the last 20 Ma (Van Den Bogaard, 2013), is, together with Hawaii (Wagner & Funk, 1995), the most intensively studied ocean archipelago from an evolutionary perspective (e.g. Husemann, Deppermann, & Hochkirch, 2014; Valtueña, López, Álvarez, Rodríguez-Riaño, & Ortega-Olivencia, 2016). This study is focused on a Canary endemic, *Lavatera acerifolia* Cav. (Malvaceae), a shrub up to 5 m tall occurring in disturbed halo-nitrophilous soils within the thermophilous forest vegetation, frequently at the base of cliffs of all main islands except El Hierro (Bramwell, Bramwell, Sánchez-Pinto, & Croker, 1990). It is a hexaploid species ($2n$ = 44, 6×, Escobar, Schönswetter, Fuertes Aguilar, Nieto Feliner, & Schneeweiss, 2009), as are its closest relatives in the *Malva* alliance including its sister species, *Lavatera maritima* (Devesa Alcaraz & Luque, 1986). *Lavatera acerifolia* is a suitable system for investigating the colonization history of island lineages since the framing elements are relatively simple for an archipelago endemic and are specifically favourable for testing the linear progression rule. Suitable geological elements include the chronology of the islands and their location perpendicular to the mainland, with the distance of each island inversely proportional to its age of emergence as well as proximity to the mainland—for example as compared to Hawaii that is more than 3,600 km away from the mainland. Other characteristics of note include the fact that the sister species, *L. maritima*, exhibits its highest genetic diversity in the region closest to the archipelago (south-west Morocco), and the fact that its mericarps lack evident dispersal mechanisms, thus reducing the likelihood of multiple colonization events (Villa-Machío, Fernández de Castro, Fuertes-Aguilar, & Nieto Feliner, 2018).

Genotyping-by-sequencing (GBS, Elshire et al., 2011) has become a widely employed genomic technique in ecology and evolution and has been useful in the study of island endemics (Curto, Schachtler, Puppo, & Meimberg, 2018), as well as for evolutionary studies in other polyploids (McAllister & Miller, 2016; Qi et al., 2015; Tyler et al., 2016). Here we used GBS to study the population structure and diversity of *L. acerfolia*, and integrate environmental factors in reconstructing the colonization history of the species. In addition, we constructed a species distribution model (SDM) to aid in assessing dispersal histories. In previous SDM studies of the closest two relatives to the study species, *L. maritima* and *Navaea phoenicea*, the main determinant of the SDM model was slope, so that these plants are mainly found in steep habitats (González Fernández de Castro, 2016; Villa-Machío et al., 2018). Since *L. acerifolia* grows in similar habitats, we hypothesized

that slope also is a prominent determinant of its potential distribution, notwithstanding the fact that islands represent highly dynamic areas (Whittaker, Triantis, & Ladle, 2008) wherein stochastic events such as landslides and eruptions can substantially affect island topography and thus availability of suitable sites. These considerations were also included in our SDM.

The overarching purpose of this study is to contribute to our understanding of the past and contemporary connections between island and mainland lineages in a model island archipelago. Our hypothesis is that in a system for which both spatial and lineage-specific elements seem to be favourable for the classic progression rule colonization model, we should be able to document it along with potential processes—such as recurrent recolonization, back-colonization, Pleistocene extinction—that may have altered it, provided that niche characteristics are accounted for. To achieve these goals we: (a) estimated genetic structure and diversity patterns within *L. acerifolia* across the archipelago, (b) inferred its island colonization pathways and (c) explored the feasibility of specific scenarios using SDM for the present, mid-Holocene (6 ka) and Last Glacial Maximum (22 ka).

2 | MATERIALS AND METHODS

2.1 | Plant material and sampling

A total of 15 populations spanning the range of *L. acerifolia* in the Canary Islands were sampled. All populations were georeferenced during fieldwork (Table 1), and leaves were collected and preserved in silica gel. Fuerteventura and Lanzarote only hold one population each, confined to abrupt slopes and consisting of 4 and *c.* 30 individuals respectively (Gangoso, Donázar, Scholz, Palacios, & Hiraldo, 2006; Gil González & Peña Hernández, 2018).

2.2 | Genotyping-by-sequencing

After a pilot study in which we surveyed several plastid regions (*mat*K, *ndh*F, *psb*A-*trn*H, *trn*D-*trn*T, *trn*G, *trn*L-*trn*F, *trn*S-*trn*G) as well as the nuclear DNA ITS for Sanger sequences, without finding informative variation, we decided to follow a genomic approach. GBS data were generated for 28 individuals of *L. acerifolia* and one individual of its sister species, *L. maritima*. Two individuals per population were sampled except for Bajamar and Famara, each sampled with one. Total DNA was extracted from leaves using DNeasy Plant Minikit (QIAGEN Inc.) and concentrated using a precipitation protocol described in Sambrook, Fritsch, and Maniatis (1989). DNA samples were processed to obtain GBS libraries according to Elshire et al. (2011). Paired-end (2 × 101 bp) Illumina sequencing was performed on a HiSeq2000 platform at Centro Nacional de Análisis Genómico (CRG-CNAG, Barcelona). Demultiplexed reads were visualized with FastQC 0.11.5 (Andrews, 2010) for quality control, and edited and filtered with Trimmomatic 0.36 (Bolger, Lohse, & Usadel, 2014).

2.3 | Bioinformatic analysis

To achieve reliable SNP calling despite difficulties related to paralogy, three different approaches were used: (a) de novo assembly (Escudero, Eaton, Hahn, & Hipp, 2014; Mastretta-Yanes et al., 2015), (b) assembly with a reduced reference sequence generated from our own data, named as a mock reference (Melo, Bartaula, & Hale, 2016) and (c) assembly using the closest available reference genome, *Gossypium arboreum* (GenBank accession number GCA_000612285.2) (Li et al., 2014). All analyses were performed using the cluster Trueno (SGAI-CSIC).

The de novo assembly was implemented in PyRAD (Eaton, 2014), whereas the mock reference and the assembly with *G. arboreum* were implemented in GBS-SNP-CROP (Melo et al., 2016), one of the few pipelines designed to discover SNPs that consider polyploid species both with and without a close-relative reference genome. The presence of a hexaploid genome requires considering two important parameters to discriminate and filter paralogs: the depth of sequence coverage used in SNP calling (Andrews, Good, Miller, Luikart, & Hohenlohe, 2016) and the degree of variation detected in the reads from the same loci, which could indicate a possible mixture of paralogs (Eaton, 2014). Accordingly, in our filtering the search was restricted to potential biallelic SNPs as detailed in Appendix S1: Figure S1, Table S1 where the three workflows are described.

2.4 | Genetic structure analysis and relatedness

Here, we only show analyses of the SNP matrices obtained from the mock reference filtering approach. Those based on the other filtering pipelines are in the Supporting Information (Appendix S2: Figures S2–S4). Population structure of *L. acerifolia* was inferred using two Bayesian model-based genetic clustering approaches: Bayesian Analysis of Population Structure (BAPS 6.0, Corander, Cheng, Marttinen, Sirén, & Tang, 2013) and STRUCTURE 2.3.4 (Pritchard, Stephens, & Donnelly, 2000). For the first approach, three matrices were analysed (mock3, mock20x, mock50x, see Appendix S1). In BAPS, a mixture analysis was conducted to estimate the number of genetically diverse populations, the maximum being 15. Then, an admixture analysis was carried out with a minimum population size of 2, and the number of iterations for estimating the admixture coefficient for the individuals was set to 100. The number of reference individuals from each population was 200, and 20 iterations were used for estimating the admixture coefficient for the reference individuals. The three analysed matrices showed similar results regarding population structure, but differed in the number of SNPs recovered. We estimated that an average read depth of 20 is sufficient coverage for working with a polyploid genome (Melo et al., 2016). Consequently, the mock20x matrix, constructed with an average read depth of 20 and a maximum of three SNPs per cluster, was selected for subsequent analyses.

The number of genetic groups was estimated in STRUCTURE under an admixture model with allele frequencies correlated among

TABLE 1 Origin of the populations of *Lavatera acerifolia* sampled

Island	Locality	Code	Colector	Latitude	Longitude
Fuerteventura	Morro del Halconcillo	HAL	S. Scholz, I. Villa	28.35689	−13.926534
Gran Canaria	Vecindad de Enfrente (Agaete)	AGA	G. Nieto, I. Villa	28.082278	−15.671056
Gran Canaria	Barranco Guiniguada	GUIN	G. Nieto, I. Villa	28.066028	−15.462778
Gran Canaria	Barranco de Guayadeque 1	GUA	G. Nieto, I. Villa	27.93675	−15.511417
Gran Canaria	Barranco de Guayadeque 2	GUA	G. Nieto, I. Villa	27.935972	−15.499028
Gran Canaria	Barranco de Guayadeque 3	GUA	G. Nieto, I. Villa	27.933528	−15.476083
Gran Canaria	Hoya de Pineda (Barranco Anzo)	HP	G. Nieto, I. Villa	28.113139	−15.639472
La Gomera	Agulo	AGU	J. Fuertes, A. González	28.184778	−17.190167
La Palma	Barranco Jorado (Tijarafe)	TIJ	J. Fuertes, A. González	28.703444	−17.960486
Lanzarote	Barranco Famara	FAM	Jose D. Naranjo	29.21831	−13.47834
Tenerife	Barranco de los Infiernos	INF	J. Fuertes, A. González	28.13375	−16.711583
Tenerife	Masca	MAS	J. Fuertes, A. González	28.298556	−16.841278
Tenerife	Barranco Guaria (Acojeja)	ACO	J. Fuertes, A. González	28.194889	−16.755806
Tenerife	Güimar	GI	J. Fuertes, A. González	28.293833	−16.405819
Tenerife	Bajamar	BAJ	J. Fuertes, A. González	28.552086	−16.340794
Tenerife	Chamorga	CHA	J. Fuertes, A. González	28.578278	−16.140806
Tenerife	Punta de Teno (Buenavista)	TENO	J. Fuertes, A. González	28.349081	−16.89486

populations. Each run consisted of 10^6 iterations with a burn-in period of 10^5. Ten replicates were carried out for each *K* value (from 1 to 16, the number of populations plus 1). The optimal partition of the genetic dataset was estimated applying the Evanno criterion (Evanno, Regnaut, & Goudet, 2005) implemented in Structure Harvester (Earl, 2012) and the multiple replicates of each *K* were combined in the online application CLUMPAK (http://clumpak.tau.ac.il/). The overall distribution of genetic variation was also analysed without population genetics assumptions by a principal component analysis (PCA) using the package 'adegenet' (Jombart & Ahmed, 2011) in the R environment (R Core Team, 2015). In addition, to examine the effect of non-natural admixture, we replicated the STRUCTURE analysis without one individual from the Guiniguada population, which had likely undergone introgression from cultivated accessions (Table 1; Figure 1). This population is located on the edge of the Jardín Botánico Canario Viera y Clavijo, close to living specimens from different origins which could have been the source of introgression (J. Naranjo, JBCVC, comm. pers.).

To infer genetic relationships among populations, a species tree was estimated using the SVDquartets method (Chifman & Kubatko, 2014) implemented in PAUP 4 (Swofford, 2002), including *L. maritima* as the outgroup. This method infers the topology among randomly sampled quartets of taxa under the coalescent model using SNPs. All possible random quartets were sampled with 1,000 replicates of nonparametric bootstrapping to measure uncertainty in the relationships.

2.5 | Island genetic diversity estimates

Due to the low number of populations that inhabit some islands, we grouped them from west to east into four groups: A (populations from La Palma and La Gomera), B (Tenerife), C (Gran Canaria) and D (Fuerteventura and Lanzarote). Nucleotide diversity (π) within groups was estimated with the 'PopGenome' package (Pfeifer, Wittelsbürger, Onsins, & Lercher, 2014) in the R environment.

Differentiation between pairs of groups was assessed using the *rho* statistic, an F_{st} analogue that is independent of the ploidy level used for polyploids (Meirmans, Liu, & Van Tienderen, 2018). It was calculated using GENODive 2.0b23 (Meirmans & Van Tienderen, 2004).

To assess the influence of environmental and geographical factors on genetic diversity across the range of the species, a generalized linear model (GLM) was performed in the R environment, in which the response variable was proportion of heterozygous sites per individual and the predictor variables were distance to the mainland, relative topoclimatic suitability and geological substrate age. The proportion of heterozygous sites in each individual was calculated using 'vcflib' (Garrison, 2012). Relative topoclimatic suitability was calculated as follows: the total area of the archipelago was divided applying Voronoi tessellation using population coordinates to define the limits of each population. Relative suitability for each population was calculated as the ratio of suitable area with respect to the area of the Voronoi polygon delimiting that population. Several models of GLM were constructed with different combinations of predictor

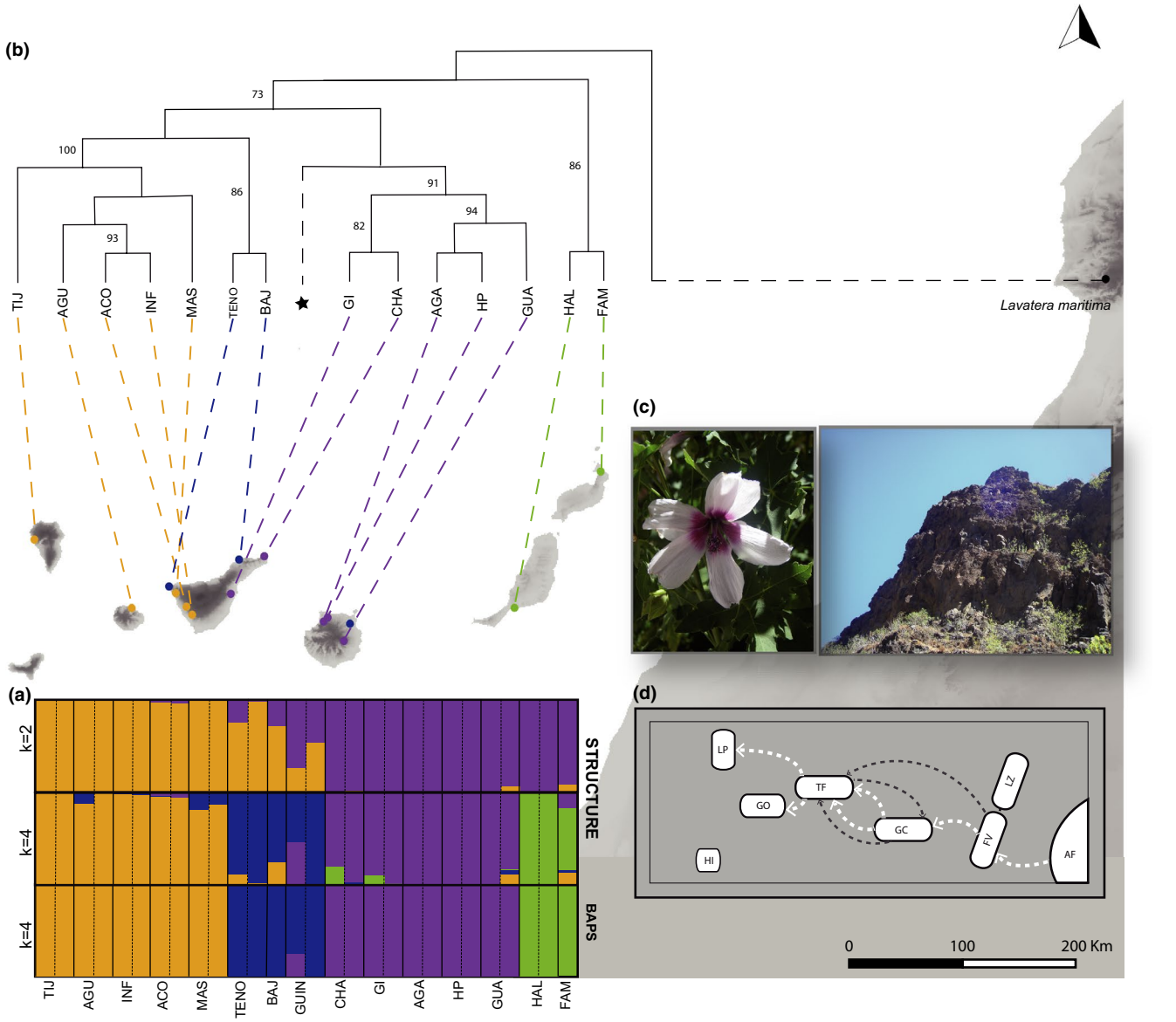


FIGURE 1 Analysis of 15 populations of *Lavatera acerifolia* based on 830 unlinked SNPs identified from a mock reference assembly. (a) Bayesian clustering of populations performed using two algorithms, STRUCTURE and BAPS. Samples are represented by rectangles where the colour indicates the probability of each sample belonging to each of the genetic groups. Two groups (K = 2) is the optimal partition according to STRUCTURE. The second most likely partition is K = 4, coinciding with the partition identified to BAPS. (b) SVDquartets tree constructed using *Lavatera maritima* as an outgroup with terminals projected to their geographical location by dash lines colour-coded according to the genetic groups identified by BAPS and STRUCTURE; bootstrap support values ≥75% are shown on the branches. The star marks a sample from the Guiniguada population that likely was introgressed in cultivation. (c) Flower and habitat of *Lavatera acerifolia* from Guayadeque (Gran Canaria). (d) Most likely colonization patterns for *L. acerifolia* (white arrows) based on the topology of a SVDquartets tree, Canary island ontogeny and species distribution modelling. Alternative pathways supported by a ML tree based on concatenated SNPs are marked in grey. AF, Africa; FV, Fuerteventura; GC, Gran Canaria; GO, La Gomera; HI, Hierro; LP, La Palma; LZ, Lanzarote; TF, Tenerife

variables and the model with the lowest Akaike information criterion (AIC) score was selected ('MuMIn' package in R; Barton, 2015).

2.6 | Species distribution model

2.6.1 | Development of bioclimatic variables

We developed a set of spatial climate layers at a 50-m resolution, based on the network of meteorological stations of the archipelago (data provided by AEMET, www.aemet.es) with 10 or more years of climate records. For the monthly variables of minimum and maximum temperature and precipitation, we developed a stepwise generalized additive model (GAM) with altitude, northness, latitude and longitude as predictor variables. Models were selected based on AIC scores and then projected onto the whole archipelago, including El Hierro. Residuals of each model in each meteorological station were used to interpolate by kriging the map of residuals for each variable. This interpolated layer was added to the predicted value of the GAM model to obtain the final layers of each monthly variable. The final dataset of monthly variables was used to develop the bioclimatic variables described by Hijmans, Cameron, Parra, Jones, and Jarvis (2005), using the 'dismo' package in R (Hijmans, Phillips, Leathwick, & Elith, 2015). We also assessed two topographical predictors: slope and topographical index (TPI), derived from the digital elevation model. We also derived these climate variables for past conditions by using the climate anomalies developed for the MIROC model (Hasumi & Emori, 2004) for the mid-Holocene (6 ka) and Last Glacial Maximum (22 ka). We downscaled the climate anomaly to a 50-m resolution for the study area using the Delta method (Field, 2014).

2.6.2 | Model calibration

At 50-m working resolution, 34 presence cells were recorded in total. We used biomod2 (Thuiller, Lafourcade, Engler, & Araújo, 2009) for which we developed five datasets containing presence points and 200 randomly generated pseudoabsence cells, weighted so as to contribute equally as presences.

To select topoclimatic predictor variables we conducted a correlation analysis with the R package 'ecospat' (Di Cola et al., 2017), which returned a value of eight predictor variables. To select those eight variables, we first performed a PCA in which 11 predictors obtained the highest scores along the first three axes. These were nine bioclimatic variables plus TPI and slope. Finally, we conducted a hierarchical partitioning approach (Chevan & Sutherland, 1991), as implemented in the R package 'hier.part' (Walsh & Mac Nally, 2013), to select the eight showing the highest independent contributions (TPI, slope, bio1: mean annual temperature, bio6: minimum temperature of the coldest month, bio7: temperature annual range, bio8: mean temperature of the wettest quarter, bio11: mean temperature of the coldest quarter, and bio15: precipitation seasonality).

We used five algorithms for niche modelling: GLM and gradient boosting machines (GBM), multivariate adaptive regression splines, artificial neural networks and random forest. For each of the algorithms chosen, we conducted 10 runs of each presence–pseudoabsence dataset. In each run, 85% of data was randomly selected for calibration and the rest for model evaluation. To evaluate the models, we used TSS and ROC scores. Models with scores over 0.8 for the two criteria were retained to build an ensemble model based on the contribution of each individual model, which was weighted according to the TSS score. Finally, this ensemble model was projected to the past climate conditions developed for the archipelago.

3 | RESULTS

3.1 | GBS output

Sequencing of 28 individuals of *L. acerifolia* from 15 populations generated 25,201,642 identified reads. The results of each pipeline using different parameters are shown in Appendix S1: Table S1.

The selected matrix from the de novo assembly (novo-c90m21) produced 1,279 loci for the whole dataset (1,236–1,268 per sample) including a total of 1,584 SNPs (453 unlinked), whereas the mock reference pipeline (mock20x) generated 1,101 (880–1,067) biallelic SNPs from 830 clusters. After applying the filters recommended for hexaploid genomes, the assembly using the *G. arboreum* reference genome resulted in only 164 SNPs recovered and this workflow was abandoned. Searching for robustness, downstream analyses were conducted on both novo-c90m21 and mock20x datasets (Appendix S1: Table S1). We only present here those from the mock reference pipeline (see Appendix S2).

3.2 | Genetic structure, diversity and relatedness

The PCA based on 1,101 SNPs revealed four distinct groups, eastern (E), central-eastern (CE), central-western (CW) and western populations (W), distributed along a longitudinal geographical gradient (Figure 2). The first two principal components explained 24.89% of the total genetic variance. The 830 unlinked SNPs in the BAPS analysis identified the same four genetic clusters (Figure 1a) with one individual (Guiniguada, CW) showing admixture from the CE group. STRUCTURE recognized two groups (E + CE vs. CW + W), with populations from Acojeja, Teno, Bajamar, Guiniguada, Guayadeque and Famara exhibiting admixture (Figure 1a). The second most likely partition for STRUCTURE, $K = 4$, identified the same groups as BAPS and the PCA, but with 11 of the 15 populations displaying admixture (Figure 1a).

The topology of the coalescent-based SVDquartets tree is consistent with BAPS and STRUCTURE (Figure 1b). The eastern group (86% BS) is sister to the remaining populations, which are split into two clades of similar size. The first one includes the CE group (91% BS)

sister to Guiniguada (without support). The second clade includes the W group (100% BS) sister to two populations from the CW group (Teno and Bajamar, 86% BS).

The current distribution of nucleotide diversity (π) for island groups increased from La Palma and La Gomera (A; 0.067), to Tenerife (B; 0.079) and Gran Canaria (C; 0.079), and then slightly decreased in Fuerteventura and Lanzarote (D; 0.071); when using the de novo assembly the east–west decreasing pattern for π is clearer (Appendix S2: Table S2). The *rho* statistic was maximum for the easternmost—westernmost island group pair and its value ranged from 0.084 (A–B pair; Table 2) to 0.415 (A–D pair; Table 2; see also Appendix S2: Table S3).

The proportion of heterozygous sites ranged from 0.20 to 0.36 (Figure 3). For this analysis, we excluded the individuals from Guiniguada. The GLM model showed that the proportion of heterozygous sites was negatively correlated with distance to the mainland (estimated slope of the predictor = −5.10e-04 ± 1.14e-04 *SD*, $p < .001$) and was not related to the relative topographical suitability of the niche. The proportion of heterozygous sites was significantly related to substrate age. However, the estimated slope of the predictor (−1.74e-8 ± 4.25e-09 *SD*, $p < .001$) was much smaller than the one for distance to mainland, and was thus considered negligible.

3.3 | Species distribution modelling

More than 90% of the model runs (229 over 250) were retained for the ensemble model (Appendix S3: Figure S5). TSS scores ranged from 0.383 to 1 and ROC from 0.25 to 1. There were no significant differences in any of the two scores, either between presence–absence datasets or runs ($F_{4,48} = 1.445$, $p = .233$).

TABLE 2 *Rho* pairwise values estimated for *Lavatera acerifolia* employing SNPs extracted from GBS data using a mock reference genome. *Rho* pairwise statistics were calculated for different geographical group pairs including A (populations from La Palma and La Gomera islands), B (populations from Tenerife island), C (populations from Gran Canaria island) and D (Fuerteventura and Lanzarote islands)

Geographical groups	A	B	C	D
A	0.000	0.084	0.339	0.415
B		0.000	0.144	0.247
C			0.000	0.296
D				0.000

The ensemble model was explained mainly by slope with an independent contribution of 0.77 compared to TPI (0.076). Climatic variables contributed less, ranging from 0.05 (mean temperature of the coldest quarter; bio11) to 0.01 (annual temperature range; bio7). The suitability threshold was 0.79 for ROC and 0.81 for the TSS score. In the projection of the model assembled with the TSS score, suitable areas in the eastern islands (Lanzarote and Fuerteventura) are restricted to a few steep areas (Figure 4a). In contrast, central and western islands (Gran Canaria, Tenerife, La Palma, La Gomera and El Hierro) host a higher proportion of suitable areas, mainly corresponding to ravines running across wide altitudinal ranges. In Gran Canaria, suitability is higher in the western part of the island, whereas in Tenerife it is concentrated in the north-eastern and north-western parts, which are the oldest mountain massifs on the island, rather than in younger central areas (Figure 4a). Suitability per island significantly increased with distance to the continent (GAM, $F_{1,440,685} = 19,660$,

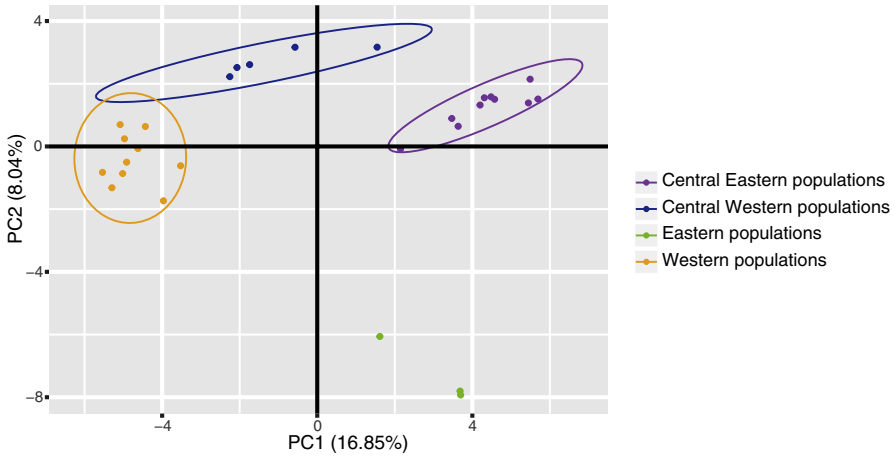


FIGURE 2 Principal component analysis (PCA) of 28 samples of *Lavatera acerifolia* based on 1,101 SNPs generated through genotyping-by-sequencing from a mock reference assembly. Scatter diagrams of the samples against the first two axes explaining 24.89% of variance. Ellipses enclosing sample points constructed with 95% confidence, matching genetic groups discussed in the main text. Eastern group includes populations from Famara and Halconcillo, central-eastern consists of Chamorga, Güimar, Agaete, Hoya Pineda and Guayadeque populations, central-western contains Teno, Bajamar, Guiniguada populations and western group includes Tijarafe, Agulo, Infierno, Acojeja and Masca populations

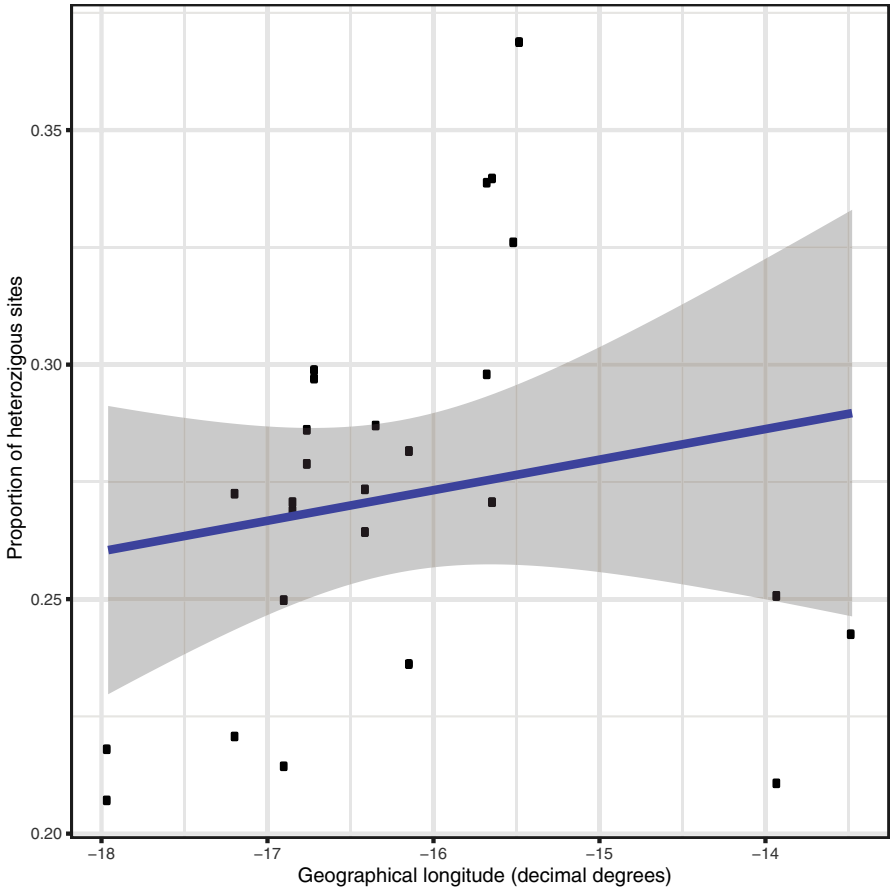


FIGURE 3 Proportion of heterozygous sites found in the genome of 26 individuals from 14 native populations of *Lavatera acerifolia* based on SNPs data identified through genotyping-by-sequence (GBS) using a mock reference genome (mock20x matrix; see text). The solid line indicates values of heterozygous sites predicted by a generalized linear model (GLM). The *x*-axis indicates the minimum distance from each island to the mainland, measured in decimal degrees

$p < .001$). The relative suitable surface per island plotted against distance to the continent (Appendix S3: Figure S6a) showed that the potential area for *L. acerifolia* reached its maximum at the central part of the archipelago, coinciding with the location of Tenerife. Similarly, topoclimatic suitability showed a significant relationship with age, peaking in geological units dating from the Miocene period (GAM, $F_{1,326} = 41.24$, $p = 4.8 \times e^{-10}$, Appendix S3: Figures S6b and S7).

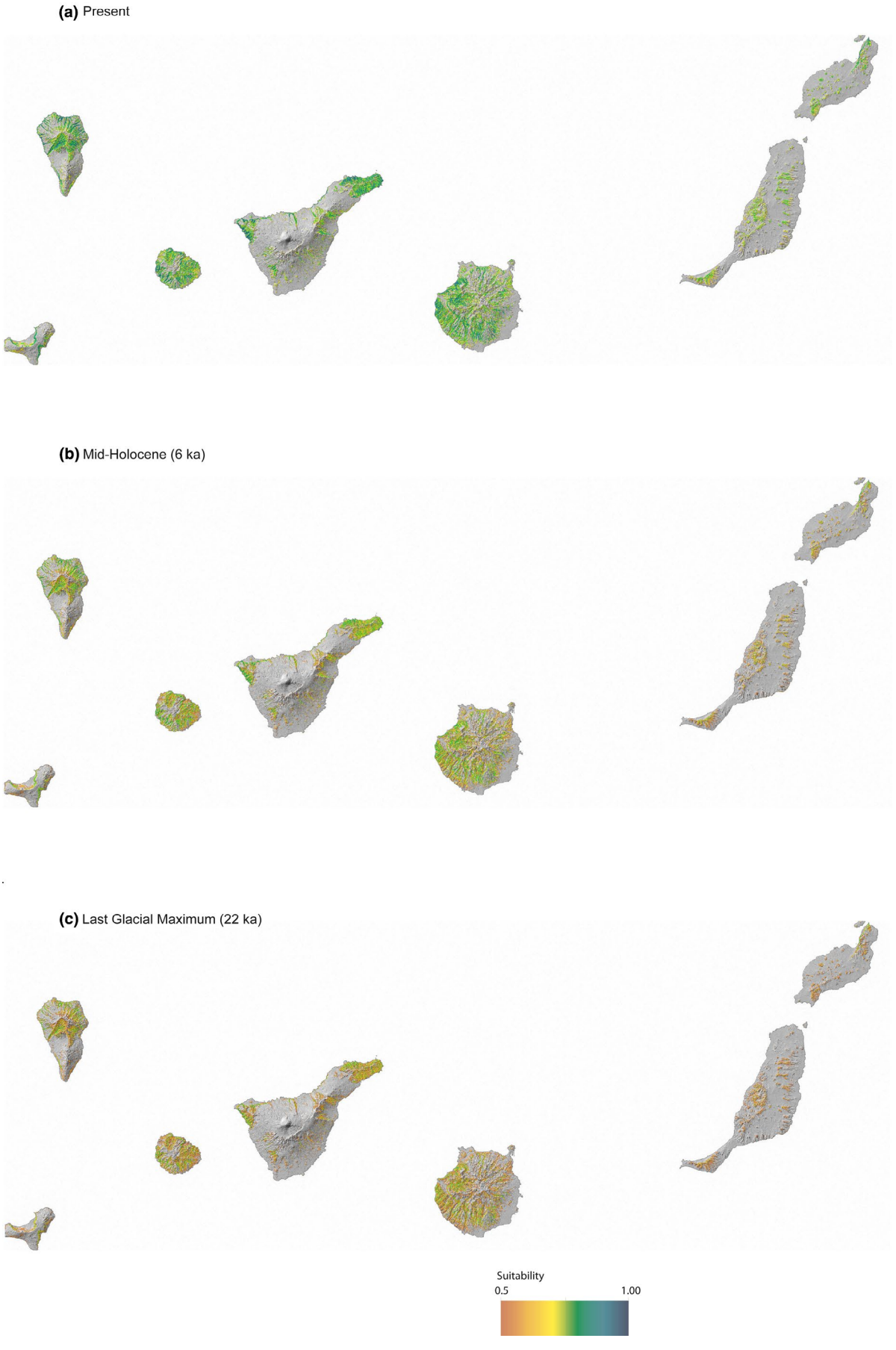
The projected suitability of the MIROC climate model to the mid-Holocene and Last Glacial Maximum followed a similar spatial pattern to the present (Figure 4b,c). Potential suitable areas in both periods were found in the same regions and the altitudinal range of these potential areas was also conserved over time. However, the topoclimatic suitability decreased with age in both temporal scenarios.

4 | DISCUSSION

The genetic structure and diversity patterns inferred for *L. acerifolia* across the Canary archipelago based on GBS data, together with phylogeographical reconstruction and the SDM projected to past scenarios provide a window into the colonization history of the species. Overall this history seems to fit the classic island progression rule scheme for oceanic islands (Whittaker et al., 2017). In the Canaries this pattern involves initial dispersal to the eastern islands—the closest to the mainland, Lanzarote and Fuerteventura—followed by subsequent westward dispersal to the remaining islands (Juan et al., 2000).

In *L. acerifolia*, differentiation between groups of islands is low, and the highest values are found between the most distant groups, eastern versus western (Table 2). An overall decline in nucleotide diversity (π, Appendix S2: Table S2) and proportion of heterozygous sites (Figure 3) is detected as the distance from the mainland increases. However, this trend for π is clearer when analysing the SNPs from the de novo assembly than from the mock reference (Appendix S2: Table S2). This pattern is consistent with island colonization following a stepping-stone model, as reported for other Canary plant species such as *Phoenix canariensis* (Saro, González-Pérez, García-Verdugo, & Sosa, 2015) and *Rumex bucephalophorus* (Talavera, Navarro-Sampedro, Ortiz, & Arista, 2013), as well as for other archipelagos, particularly Hawaii (Croucher, Oxford, Lam, Mody, & Gillespie, 2012; Fleischer, Mcintosh, & Tarr, 1998; Pillon et al., 2013).

FIGURE 4 Niche suitability for an ensemble species distribution model of *Lavatera acerifolia* constructed using five algorithms and seven variables, based on the TSS evaluation score: (a) current climate conditions; (b) projection of the ensemble model for the mid-Holocene (6 ka) climate conditions; (c) projection for the Last Glacial Maximum (22 ka). Values of niche suitability <0.5 are not coloured. Presence–absence threshold is set up by the ensemble model in 0.65


(a) Present
(b) Mid-Holocene (6 ka)
(c) Last Glacial Maximum (22 ka)
Suitability
0.5
1.00

As an alternative to the island progression rule model for *L. acerifolia*, García-Verdugo, Caujapé-Castells, Illera, et al. (2019) proposed extinctions in the eastern islands during the Pleistocene followed by recolonization events. Such extinctions and recolonizations, in principle, may have been facilitated by the low topographical complexity of Lanzarote and Fuerteventura, which offer few micro-refuges except for the Famara and Morro del Halconcillo massifs (Reyes-Betancort, Guerra, Guma, Humphries, & Carine, 2008). These micro-refuges coincide with the environmentally suitable areas for *L. acerifolia* and with their presence in those islands (Figure 4; Appendix S3: Figure S8). These two populations seem to have withstood demographic decline. However, with only 4 (Morro del Halconcillo, Fuerteventura) and *c*. 30 (Famara, Lanzarote) known individuals, the two islands have greater nucleotide diversity than the western islands, also represented by a small number of populations, or even than the rest of the island groups, when the de novo assembly is considered (Appendix S2: Table S2). Based on our data, the fact that the earliest diverging populations form a genetic group (Figure 1b) suggests that neither recolonizations from western islands nor extirpation of original populations has occurred. The highest nucleotide diversity in Lanzarote-Fuerteventura (D, Appendix S2: Table S2) is also consistent with non-extirpation. According to García-Verdugo, Caujapé-Castells, Illera, et al. (2019), persistence of eastern populations is more frequent in recently diverged groups (<0.8 Ma) like *Phoenix* (Saro et al., 2015) and *Rumex* (Talavera et al., 2013). Since Villa-Machío et al. (2018) inferred an early Pleistocene split of *L. acerifolia* from its sister species, *L. maritima*, the persistence of old populations in the eastern islands could be noteworthy. One caveat, however, is the possibility that time of differentiation from the sister species is not a good proxy for the time of colonization, as recently pointed out in García-Verdugo, Caujapé-Castells, and Sanmartín (2019), although such a mismatch should be more pronounced in other archipelagos that are farther from the mainland.

From a phylogenetic perspective, the SVDquartets tree, placing the easternmost populations as the earliest diverging (Figure 1b, Appendix S2: Figure S4), is consistent with this being the location of original colonization from the mainland. The overall topology also supports an east–west colonization pattern, albeit with some uncertainties regarding a strict linear progression rule. The central islands—the ones with the highest carrying capacity at present—seem to provide signatures of a non-strictly linear scheme. Specifically, two independent colonization events from Gran Canaria to Tenerife and independent colonization of La Palma and Gomera from Tenerife seem to have occurred. Furthermore, uncertainties about the role played by the two large islands in the colonization of the archipelago are also suggested by the topology of a ML tree based on concatenated SNPs, which is more asymmetric than that of the SVDquartets tree (Figure S4). According to this topology, an early colonization of Tenerife directly from the eastern islands could have occurred, followed by dispersal from Tenerife to Gran Canaria and subsequent dispersal in the opposite direction (Figure 1d). Also La Palma could have been colonized from Gomera, not Tenerife. Still one additional potentially intervening factor in these two islands is summarized by the hypothesis of the Syngameon (Caujapé-Castells et al., 2017), which posits that gene flow between populations (or in other groups between congeneric species) has contributed to high levels of genetic diversity and potential for colonization. This is compatible with Tenerife and Gran Canaria housing a larger number of populations and exhibiting high nucleotide diversity (Appendix S2: Table S2).

SDM projected onto past scenarios has caveats. In addition to the implicit assumption that niche has not changed over time (Peterson, Soberón, & Sánchez-Cordero, 1999), the recent time periods onto which the model can be projected preclude covering a significant part of the history of the species. In this study we have not projected our model to the last interglacial because the fine-scale data used for the present time cannot be meaningfully projected to *c*. 130,000 Ma. Despite these limitations, SDM can still complement the phylogeographical reconstruction inferred from the GBS data and provide insights into the specifics of the colonization pattern. First, projections of the model to the latest glacial period show stability in areas with a suitable niche, even if they are smaller, suggesting that *L. acerifolia* could have persisted within the same areas instead of becoming extinct followed by recolonization from other islands. A second finding based on the SDM is the importance of steeply sloped habitats, at the base of which organic matter accumulates, found in the SDM of *L. acerifolia* and previously in those of its closest relatives *N. phoenicea* (González Fernández de Castro, 2016) and *L. maritima* (Villa-Machío et al., 2018). A third finding concerning SDM is that topoclimatic suitability is significantly related to geological age, with peaks in Miocene areas, possibly because younger areas corresponding to recent eruption events created topographically less complex habitats that are unsuitable (Figure 4, Appendix S3: Figure S6). This is evident in southern La Palma, eastern Gran Canaria and especially in Tenerife. Flat, senescent islands also reduce opportunities for the occurrence of *L. acerifolia*, which thus is constrained to remnant mountain enclaves in Lanzarote and Fuerteventura. In contrast, topoclimatic suitability tends to increase towards the western and younger islands where, despite the availability of extensive areas, only one population per island is known or, in the case of El Hierro, none. Such a pattern is puzzling and may be indicative of the existence of a colonization front taking place at a slow pace. Lack of efficient dispersal mechanisms in *L. acerifolia* could require longer time periods to achieve successful colonization of these western islands (Arjona, Nogales, Heleno, & Vargas, 2018). However, the topographical dynamism due to volcanic activity in these younger islands may have also contributed to the observed pattern.

Available genomic resources for already a considerable number of organisms show that heterogeneous processes and signals can be picked up when we sample genomes, thus compromising the reconstruction of unified evolutionary histories (Bravo et al., 2018). Such realization not only stresses the need for careful filtering of genomic data but also for integrating independent evidence that can help sort out those processes.

As shown here, environmental factors can help to filter and modulate hypotheses in island biogeography studies, for example in reconciling dispersal capacity and probability of successful colonization.

In the case of *L. acerifolia*, the combined use of GBS and SDM has allowed us to unravel the relationships between mainland and island lineages, which are central to understanding island biogeography (Patiño et al., 2017). Our work also has posed new questions about the apparent discrepancies between suitable areas in the westernmost islands and their current populations. Thus, our study illustrates the diversity of factors that influence colonization history, even in cases for which geological and biological settings seem nearly optimal for matching an island progression rule model. Finally, compared to studies in Hawaii (Croucher et al., 2012; Fleischer et al., 1998), our results suggest that association between island age and chronology of colonization is a consistent pattern across oceanic archipelagos differing in fundamental spatial elements such as the closeness to mainland founding populations. The veracity of this proposal will become more clear as additional taxa become studied in this context.

ACKNOWLEDGEMENTS

The authors are grateful to Juli Caujapé Castells, Mario Mairal, Águedo Marrero, José Naranjo, Alfredo Reyes Betancort, Arnoldo Santos Guerra, Ángel B. Fernández López and Stephan Scholz for providing samples, information or assistance with fieldwork; José Naranjo for information on possible sources of introgression for the population of Guiniguada; Yurena Arjona, Mario Fernández-Mazuecos, Marcial Escudero and Isabel Sanmartín for help in some analyses; Joaquina M. García-Martín for helpful advice. The helpful suggestions on the manuscript by Jonathan F. Wendel and Rosemary Gillespie are also much appreciated. The Spanish State Meteorological Agency (AEMET) provided climate station data within the frame of the Dissertation project to I. V-M. Funding was provided by the Spanish ministries of Science, Technology and Innovation (grant CGL2010-16138) and Economy and Competitivity (grant CGL2013-49097-C2-1-P, CGL2017-88500-P (AEI/FEDER, UE)] and through the FPI fellowship BES-2011-049852 (co-financed by European Social Fund) to I.V-M.

DATA AVAILABILITY STATEMENT

Genomic data generated for this study are available in vcf format on Dryad: https://doi.org/10.5061/dryad.2fqz612kc

ORCID

Irene Villa-Machío https://orcid.org/0000-0002-4281-5975
Alejandro G. Fernández de Castro https://orcid.org/0000-0001-5239-7490
Javier Fuertes-Aguilar https://orcid.org/0000-0001-5649-4845
Gonzalo Nieto Feliner https://orcid.org/0000-0002-7469-4733

REFERENCES

Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., & Hohenlohe, P. A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics*, *17*, 81–92. https://doi.org/10.1038/nrg.2015.28

Andrews, S. (2010). FastQC: A quality control tool for high throughput sequence data. Retrieved from http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Arjona, Y., Nogales, M., Heleno, R., & Vargas, P. (2018). Long-distance dispersal syndromes matter: Diaspore–trait effect on shaping plant distribution across the Canary Islands. *Ecography*, *41*, 805–814. https://doi.org/10.1111/ecog.02624

Barton, K. (2015). MuMIn: multi-model inference: R package ver. 1.15.1. https://cran.r-project.org/web/packages/MuMIn.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bramwell, D., Bramwell, Z. I., Sánchez-Pinto, L., & Croker, E. O. (1990). *Flores silvestres de las Islas Canarias* (vol. 1). Madrid, Spain: Ed. Rueda.

Bravo, G. A., Antonelli, A., Bacon, C. D., Bartoszek, K., Blom, M., Huynh, S., … Edwards, S. V. (2018). Embracing heterogeneity: Coalescing the tree of life and the future of phylogenomics. *PeerJ*, *7*, e6399. https://doi.org/10.7717/peerj.6399

Caujapé-Castells, J. (2011). Jesters, red queens, boomerangs and surfers: A molecular outlook on the diversity of the Canarian endemic flora. In D. Bramwell & J. Caujapé-Castells (Eds.), *The biology of island floras* (pp. 284–324). Cambridge, UK: Cambridge University Press.

Caujapé-Castells, J., García-Verdugo, C., Marrero-Rodríguez, Á., Fernández-Palacios, J. M., Crawford, D. J., & Mort, M. E. (2017). Island ontogenies, syngameons, and the origins and evolution of genetic diversity in the Canarian endemic flora. *Perspectives in Plant Ecology, Evolution and Systematics*, *27*, 9–22. https://doi.org/10.1016/j.ppees.2017.03.003

Chevan, A., & Sutherland, M. (1991). Hierarchical partitioning. *The American Statistician*, *45*, 90–96.

Chifman, J., & Kubatko, L. (2014). Quartet inference from SNP data under the coalescent model. *Bioinformatics*, *30*, 3317–3324. https://doi.org/10.1093/bioinformatics/btu530

Corander, J., Cheng, L., Marttinen, P., Sirén, J., & Tang, J. (2013). *BAPS: Bayesian analysis of population structure V. 6.0*. Helsinki, Finland: Department of Mathematics and Statistics. University of Helsinki, Finland.

Croucher, P. J. P., Oxford, G. S., Lam, A., Mody, N., & Gillespie, R. G. (2012). Colonization history and population genetics of the color-polymorphic Hawaiian happy-face spider *Theridion grallator* (Araneae, Theridiidae). *Evolution*, *66*(9), 2815–2833. https://doi.org/10.1111/j.1558-5646.2012.01653.x

Curto, M., Schachtler, C., Puppo, P., & Meimberg, H. (2018). Using a new RAD-sequencing approach to study the evolution of Micromeria in the Canary islands. *Molecular Phylogenetics and Evolution*, *119*, 160–169. https://doi.org/10.1016/j.ympev.2017.11.005

Devesa Alcaraz, J. A., & Luque, T. (1986). Contribución al estudio citotaxonómico del género Lavatera (Malvaceae) en España. *Lagascalia*, *14*(2), 227–240.

Di Cola, V., Broennimann, O., Petitpierre, B., Breiner, F. T., D'Amen, M., Randin, C., … Guisan, A. (2017). ecospat: An R package to support spatial analyses and modeling of species niches and distributions. *Ecography*, *40*, 774–787. https://doi.org/10.1111/ecog.02671

Diaz-Pérez, A., Sequeira, M., Santos-Guerra, A., & Catalán, P. (2008). Multiple colonizations, in situ speciation, and volcanism-associated stepping-stone dispersals shaped the phylogeography of the Macaronesian red fescues (*Festuca* L., Gramineae). *Systematic Biology*, *57*, 732–749. https://doi.org/10.1080/10635150802302450

Earl, D. A., & von Holdt, B. M. (2012). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, *4*, 359–361. https://doi.org/10.1007/s12686-011-9548-7

Eaton, D. A. (2014). PyRAD: Assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics*, *30*, 1844–1849. https://doi.org/10.1093/bioinformatics/btu121

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE*, *6*, e19379. https://doi.org/10.1371/journal.pone.0019379

Escobar, P., Schönswetter, P., Fuertes Aguilar, J., Nieto Feliner, G., & Schneeweiss, G. M. (2009). Five molecular markers reveal extensive morphological homoplasy and reticulate evolution in the *Malva* alliance (Malvaceae). *Molecular Phylogenetics and Evolution*, *50*, 226–239. https://doi.org/10.1016/j.ympev.2008.10.015

Escudero, M., Eaton, D. A. R., Hahn, M., & Hipp, A. L. (2014). Genotyping-by-sequencing as a tool to infer phylogeny and ancestral hybridization: A case study in *Carex* (Cyperaceae). *Molecular Phylogenetics and Evolution*, *79*, 359–367. https://doi.org/10.1016/j.ympev.2014.06.026

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology*, *14*, 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Field, C. B. (Ed.) (2014). *Climate change 2014 – Impacts, adaptation and vulnerability: Regional aspects*. Cambridge, UK: Cambridge University Press.

Fleischer, R. C., Mcintosh, C. E., & Tarr, C. L. (1998). Evolution on a volcanic conveyor belt: Using phylogeographic reconstructions and K-Ar-based ages of the Hawaiian Islands to estimate molecular evolutionary rates. *Molecular Ecology*, *7*(4), 533–545. https://doi.org/10.1046/j.1365-294x.1998.00364.x

Gangoso, L., Donázar, J. A., Scholz, S., Palacios, C. J., & Hiraldo, F. (2006). Contradiction in conservation of island ecosystems: Plants, introduced herbivores and avian scavengers in the Canary Islands. *Biodiversity & Conservation*, *15*, 2231. https://doi.org/10.1007/s10531-004-7181-4

García-Olivares, V., López, H., Patiño, J., Alvarez, N., Machado, A., Carracedo, J. C., … Emerson, B. C. (2017). Evidence for mega-landslides as drivers of island colonization. *Journal of Biogeography*, *44*(5), 1053–1064. https://doi.org/10.1111/jbi.12961

García-Verdugo, C., Caujapé-Castells, J., Illera, J. C., Mairal, M., Patiño, J., Reyes-Betancort, A., & Scholz, S. (2019). Pleistocene extinctions as drivers of biogeographical patterns on the easternmost Canary Islands. *Journal of Biogeography*, *46*, 845–859. https://doi.org/10.1111/jbi.13563

García-Verdugo, C., Caujapé-Castells, J., & Sanmartín, I. (2019). Colonization time on island settings: Lessons from the Hawaiian and Canary Island floras. *Botanical Journal of the Linnean Society*, *191*(2), 155–163. https://doi.org/10.1093/botlinnean/boz044

García-Verdugo, C., Sajeva, M., La Mantia, T., Harrouni, C., Msanda, F., & Caujapé-Castells, J. (2015). Do island plant populations really have lower genetic variation than mainland populations? Effects of selection and distribution range on genetic diversity estimates. *Molecular Ecology*, *24*, 726–741. https://doi.org/10.1111/mec.13060

Garrison, E. (2012). Vcflib: A C++ library for parsing and manipulating VCF files. GitHub. https://github.com/ekg/vcflib

Gil González, J., & Peña Hernández, M. (2018). *Guía visual de la flora vascular de la reserva de la biosfera de Lanzarote*. Arrecife, Spain: Cabildo de Lanzarote.

Givnish, T. J., Millam, K. C., Mast, A. R., Paterson, T. B., Theim, T. J., Hipp, A. L., … Sytsma, K. J. (2008). Origin, adaptive radiation and diversification of the Hawaiian lobeliads (Asterales: Campanulaceae). *Proceedings of the Royal Society B: Biological Sciences*, *276*, 407–416.

González Fernández de Castro, A. (2016). Islas dentro de islas: biología y conservación del paleoendemismo macaronésico *Navaea phoenicea* (Vent) Webb & Berthel (PhD Thesis). Universidad Autónoma de Madrid, Madrid. Retrieved from https://repositorio.uam.es/handle/10486/671804

Hasumi, H., & Emori, S. (2004). K-1 coupled model (MIROC) description (K-1 Technical Report 1). Center for Climate System Research, University of Tokyo.

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, *25*(15), 1965–1978. https://doi.org/10.1002/joc.1276

Hijmans, R. J., Phillips, S., Leathwick, J., & Elith, J. (2015). *dismo: Species distribution modeling. R package version 1.0-12*. Vienna, Austria: The R Foundation for Statistical Computing. http://cran.R-project.org

Husemann, M., Deppermann, J., & Hochkirch, A. (2014). Multiple independent colonization of the Canary Islands by the winged grasshopper genus *Sphingonotus* Fieber, 1852. *Molecular Phylogenetics and Evolution*, *81*, 174–181. https://doi.org/10.1016/j.ympev.2014.09.017

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

Juan, C., Emerson, B. C., Oromí, P., & Hewitt, G. M. (2000). Colonization and diversification: Towards a phylogeographic synthesis for the Canary Islands. *Trends in Ecology & Evolution*, *15*, 104–109. https://doi.org/10.1016/S0169-5347(99)01776-0

Li, F., Fan, G., Wang, K., Sun, F., Yuan, Y., Song, G., … Yu, S. (2014). Genome sequence of the cultivated cotton *Gossypium arboreum*. *Nature Genetics*, *46*(6), 567–572. https://doi.org/10.1038/ng.2987

Mastretta-Yanes, A., Arrigo, N., Alvarez, N., Jorgensen, T. H., Piñero, D., & Emerson, B. C. (2015). Restriction site-associated DNA sequencing, genotyping error estimation and de novo assembly optimization for population genetic inference. *Molecular Ecology Resources*, *15*, 28–41.

McAllister, C. A., & Miller, A. J. (2016). Single nucleotide polymorphism discovery via genotyping by sequencing to assess population genetic structure and recurrent polyploidization in *Andropogon gerardii*. *American Journal of Botany*, *103*, 1314–1325.

Meirmans, P. G., Liu, S., & Van Tienderen, P. H. (2018). The analysis of polyploid genetic data. *Journal of Heredity*, *109*, 283–296. https://doi.org/10.1093/jhered/esy006

Meirmans, P. G., & Van Tienderen, P. H. (2004). GENOTYPE and GENODIVE: Two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes*, *4*, 792–794. https://doi.org/10.1111/j.1471-8286.2004.00770.x

Melo, A. T., Bartaula, R., & Hale, I. (2016). GBS-SNP-CROP: A reference-optional pipeline for SNP discovery and plant germplasm characterization using variable length, paired-end genotyping-by-sequencing data. *BMC Bioinformatics*, *17*, 29. https://doi.org/10.1186/s12859-016-0879-y

Parent, C. E., Caccone, A., & Petren, K. (2008). Colonization and diversification of Galápagos terrestrial fauna: A phylogenetic and biogeographical synthesis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*, 3347–3361. https://doi.org/10.1098/rstb.2008.0118

Patiño, J., Whittaker, R. J., Borges, P. A. V., Fernández-Palacios, J. M., Ah-Peng, C., Araújo, M. B., … Emerson, B. C. (2017). A roadmap for island biology: 50 fundamental questions after 50 years of The Theory of Island Biogeography. *Journal of Biogeography*, *44*, 963–983.

Peterson, A., Soberón, J., & Sánchez-Cordero, V. (1999). Conservatism of ecological niches in evolutionary time. *Science*, *285*(5431), 1265–1267. https://doi.org/10.1126/science.285.5431.1265

Pfeifer, B., Wittelsbürger, U., Onsins, S. E. R., & Lercher, M. J. (2014). PopGenome: An efficient Swiss army knife for population genomic analyses in R. *Molecular Biology and Evolution*, *31*, 1929–1936. https://doi.org/10.1093/molbev/msu136

Pillon, Y., Johansen, J., Sakishima, T., Chamala, S., Barbazuk, W. B., Roalson, E. H., … Stacy, E. A. (2013). Potential use of low-copy nuclear genes in DNA barcoding: A comparison with plastid genes in two Hawaiian plant radiations. *BMC Evolutionary Biology*, *13*(1), 35. https://doi.org/10.1186/1471-2148-13-35

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*, 945–959.

Qi, Z. C., Yu, Y., Liu, X., Pais, A., Ranney, T., Whetten, R., & Xiang, Q. Y. J. (2015). Phylogenomics of polyploid *Fothergilla* (Hamamelidaceae) by RADtag based GBS—Insights into species origin and effects of software pipelines. *Journal of Systematics and Evolution*, *53*, 432–447.

R Core Team. (2015). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Reyes-Betancort, J. A., Guerra, A. S., Guma, I. R., Humphries, C. J., & Carine, M. A. (2008). Diversity, rarity and the evolution and conservation of the Canary Islands endemic flora. *Anales del Jardín Botánico de Madrid*, *65*, 25–45. https://doi.org/10.3989/ajbm.2008.v65.i1.244

Rumeu, B., Caujapé-Castells, J., Blanco-Pastor, J. L., Jaén-Molina, R., Nogales, M., Elias, R. B., & Vargas, P. (2011). The colonization history of *Juniperus brevifolia* (Cupressaceae) in the Azores Islands. *PLoS ONE*, *6*, e27697. https://doi.org/10.1371/journal.pone.0027697

Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989). *Molecular cloning: A laboratory manual* (2nd ed., vol. *3*). New York, NY: Cold Spring Harbor Laboratory Press.

Saro, I., González-Pérez, M. A., García-Verdugo, C., & Sosa, P. A. (2015). Patterns of genetic diversity in *Phoenix canariensis*, a widespread oceanic palm (species) endemic from the Canarian archipelago. *Tree Genetics & Genomes*, *11*, 1–13. https://doi.org/10.1007/s11295-014-0815-0

Swofford, D. L. (2002). *PAUP*. Phylogenetic analysis using parsimony (* and other methods)*. Version 4. Sunderland, MA: Sinauer Associates.

Talavera, M., Navarro-Sampedro, L., Ortiz, P. L., & Arista, M. (2013). Phylogeography and seed dispersal in islands: The case of *Rumex bucephalophorus* subsp. *canariensis* (Polygonaceae). *Annals of Botany*, *111*, 249–260. https://doi.org/10.1093/aob/mcs284

Thuiller, W., Lafourcade, B., Engler, R., & Araújo, M. B. (2009). BIOMOD – A platform for ensemble forecasting of species distributions. *Ecography*, *32*, 369–373. https://doi.org/10.1111/j.1600-0587.2008.05742.x

Tyler, L., Lee, S. J., Young, N. D., Delulio, G. A., Benavente, E., Reagon, M., … Caycedo, A. L. (2016). Population structure in the model grass *Brachypodium distachyon* is highly correlated with flowering differences across broad geographic areas. *The Plant Genome*, *9*, https://doi.org/10.3835/plantgenome2015.08.0074

Valtueña, F. J., López, J., Álvarez, J., Rodríguez-Riaño, T., & Ortega-Olivencia, A. (2016). *Scrophularia arguta*, a widespread annual plant in the Canary Islands: A single recent colonization event or a more complex phylogeographic pattern? *Ecology and Evolution*, *6*, 4258–4273.

Van Den Bogaard, P. (2013). The origin of the Canary Island Seamount Province – New ages of old seamounts. *Scientific Reports*, *3*, 2107. https://doi.org/10.1038/srep02107

Villa-Machío, I., Fernández de Castro, A. G., Fuertes-Aguilar, J., & Nieto Feliner, G. (2018). Out of North Africa by different routes: Phylogeography and species distribution model of the western Mediterranean *Lavatera maritima* (Malvaceae). *Botanical Journal of the Linnean Society*, *187*, 441–455. https://doi.org/10.1093/botlinnean/boy025

Wagner, W. L., & Funk, V. A. (1995). *Hawaiian biogeography: Evolution on a hotspot archipelago*. Washington, VA: Smithsonian Institute Press.

Walsh, C., & MacNally, R. (2013). *hier. part: Hierarchical partitioning. R package version 1.0-4. Part of the R software*. Vienna, Austria: R Foundation for Statistical Computing.

Whittaker, R. J., Fernández-Palacios, J. M., Matthews, T. J., Borregaard, M. K., & Triantis, K. A. (2017). Island biogeography: Taking the long view of nature's laboratories. *Science*, *357*, eaam8326. https://doi.org/10.1126/science.aam8326

Whittaker, R. J., Triantis, K. A., & Ladle, R. J. (2008). A general dynamic theory of oceanic island biogeography. *Journal of Biogeography*, *35*, 977–994.

BIOSKETCH

All authors contributed to the design of the study, conducted the fieldwork and wrote the manuscript. In addition, G.N.F. helped with interpretation of results, A.G.F. carried out the SDM and I.V.M. performed the laboratory work, filtered raw data using different bioinformatic pipelines and analysed all genetic data.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Villa-Machío I, Fernández de Castro AG, Fuertes-Aguilar J, Nieto Feliner G. Colonization history of the Canary Islands endemic *Lavatera acerifolia*, (Malvaceae) unveiled with genotyping-by-sequencing data and niche modelling. *J Biogeogr*. 2020;00:1–13. https://doi.org/10.1111/jbi.13808