Received: 11 November 2022 | Accepted: 12 September 2024

DOI: 10.1111/mec.17537

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

Phylogenomics and phylogeographic model testing using convolutional neural networks reveal a history of recent admixture in the Canarian *Kleinia neriifolia*

Mario Rincón Barrado[1] | **Manolo Perez**[2] | **Tamara Villaverde**[1,3,4] | **Carlos García-Verdugo**[5] | **Juli Caujapé-Castells**[6] | **Ricarda Riina**[1] | **Isabel Sanmartín**[1]

[1]Department of Biodiversity and Conservation, Real Jardín Botánico (RJB), CSIC, Madrid, Spain

[2]Department of Life Sciences, Imperial College London, Silwood Park, Ascot, UK

[3]Instituto de Investigación en Cambio Global (IICG-URJC), Universidad Rey Juan Carlos, Móstoles, Spain

[4]Departamento de Biología y Geología, Física y Química Inorgánica, Universidad Rey Juan Carlos, Móstoles, Spain

[5]Departamento de Botánica, Universidad de Granada, Granada, Andalucía, Spain

[6]Dept. of Molecular Biodiversity & DNA Bank, Jardín Botánico Canario Viera y Clavijo-UA de I+D+i al CSIC, Las Palmas de Gran Canaria, Spain

Correspondence
Mario Rincón Barrado, Isabel Sanmartín and Ricarda Riina, Real Jardín Botánico, Plaza de Murillo 2, 28014 Madrid, Spain.
Email: mario.rincon@cnb.csic.es; rriina@rjb.csic.es and isanmartin@rjb.csic.es

Funding information
Consejo Superior de Investigaciones Científicas, Grant/Award Number: BES-2013-065389 and CGL2015-67849-P; Ministerio de Ciencia e Innovación, Grant/Award Number: PID2019-108109GB-I00

Handling Editor: Loren Rieseberg

Abstract

Multiple-island endemics (MIE) are considered ideal natural subjects to study patterns of island colonization that involve recent population-level genetic processes. *Kleinia neriifolia* is a Canarian MIE widespread across the archipelago, which exhibits a close phylogenetic relationship with species in northwest Africa and at the other side of the Sahara Desert. Here, we used target sequencing with plastid skimming (Hyb-Seq), a dense population-level sampling of *K. neriifolia*, and representatives of its African–southern Arabian relatives to infer phylogenetic relationships and divergence times at the species and population levels. Using population genetic techniques and machine learning (convolutional neural networks [CNNs]), we reconstructed phylogeographic relationships and patterns of genetic admixture based on a multilocus SNP nuclear dataset. Phylogenomic analysis based on the nuclear dataset identifies the northwestern African *Kleinia anteuphorbium* as the sister species of *K. neriifolia*, with divergence starting in the early Pliocene. Divergence from its sister clade, comprising species from the Horn of Africa and southern Arabia, is dated to the arid Messinian period, lending support to the climatic vicariance origin of the Rand Flora. Phylogeographic model testing with CNNs supports an initial colonization of the central island of Tenerife followed by eastward and westward migration across the archipelago, which resulted in the observed east/west phylogeographic split. Subsequent population extinctions linked to aridification events, and recolonization from Tenerife, are proposed to explain the patterns of genetic admixture in the eastern Canary Islands. We demonstrate that CNNs based on SNPs can be used to discriminate among complex scenarios of island migration and colonization.

KEYWORDS
Canary Islands, convolutional neural network, Hyb-Seq, island biogeography, Rand Flora, Syngameon hypothesis

Isabel Sanmartín and Ricarda Riina contributed equally to this study and are also joint senior authors.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2024 The Author(s). *Molecular Ecology* published by John Wiley & Sons Ltd.

1 | INTRODUCTION

Oceanic islands, despite representing a small percentage of the emergent land, are important biodiversity hotspots that concentrate a disproportionate amount of terrestrial endemism (Whittaker et al., 2023). They are geographically limited by the ocean and were never connected to continental landmasses as they originated via volcanism, and were never in contact with the continents. These attributes confer them important advantages for the study of biogeography (MacArthur & Wilson, 1963) and also help to expand our knowledge in fields as diverse as community ecology, macroevolution and population genetics (Overcast et al., 2023; Warren et al., 2015).

The Canarian flora is probably one of the best studied island floras in the world (Cardoso et al., 2010; Florencio et al., 2021; Reyes-Betancort et al., 2008) and has provided key insights into patterns of colonization, speciation and species extinction in insular systems (Caujapé-Castells et al., 2017; Fernández-Palacios et al., 2016; García-Verdugo, Caujapé-Castells, & Sanmartín, 2019; Hooft van Huysduynen et al., 2021; Sanmartín et al., 2008; Valente et al., 2017; Whittaker et al., 2008). However, studies have mostly focused on radiations of single-island endemic (SIEs) because they represent the majority of endemic species in the archipelago (c. 70%; Carine & Schaefer, 2010). On the contrary, multiple-island endemics (MIE), that is species that are present in two or more islands within the archipelago, have gained attention recently (Barres et al., 2017; Caujapé-Castells et al., 2022). These species allow exploring younger and ongoing population-level processes, including genetic admixture, geographic extirpation (population extinction), and events of demographic expansion or population bottlenecks (Caujapé-Castells et al., 2017, 2022; García-Verdugo, Caujapé-Castells, Illera, et al., 2019; Mairal, Pokorny, et al., 2015).

In highly isolated oceanic archipelagos such as Hawaii, the pattern of island colonization typically follows the 'stepping-stone' colonization (SSC) model: dispersal takes place between adjacent islands, following their order of geological appearance (Shaw & Gillespie, 2016; Wagner & Funk, 1995). However, in the Canary Islands, phylogeographical studies have revealed a more complex history of island migration and diversification. An east/west phylogeographical split has been described in numerous species radiations composed of multiple SIEs (Allan et al., 2004; Sanmartín et al., 2008), as well as, at the population level, within several Canarian MIEs (García-Verdugo, Caujapé-Castells, Mairal, & Monroy, 2019; Mairal, Sanmartín, et al., 2015; Saro et al., 2015; Villaverde et al., 2018). This pattern, referred to here as the 'Central Island Hub' (CIH) model, posits that the central islands in the archipelago, especially Tenerife, which is the largest and highest, acted as the source of colonization events towards the other islands, eastwards and westwards (Sanmartín et al., 2008). Often, a within-island geographic split is observed between the populations in central and eastern Tenerife and those in Teno, the westernmost part of the island (García-Verdugo, Caujapé-Castells, Mairal, & Monroy, 2019; Mairal, Sanmartín, et al., 2015; Sanmartín et al., 2008; Saro et al., 2015). These 'western Tenerife' populations are usually more closely related to populations to the west of the archipelago (La Gomera, La Palma and El Hierro), whereas populations from 'eastern/central Tenerife' are frequently grouped with those from other islands to the east (Gran Canaria, Fuerteventura, Lanzarote). The east/west CIH model is, however, at odds with the finding of higher values of population genetic diversity (i.e. expected heterozygosity and number of alleles per locus) in the eastern islands of Lanzarote and Fuerteventura in several plant species (Caujapé-Castells, 2011; González-Pérez et al., 2004; Kim et al., 1998; Oliva-Tejera et al., 2005). The eastern islands are the lowest in elevation and the driest, and for many plant and animal groups, they also harbour the lowest number of endemic species (García-Verdugo, Caujapé-Castells, Mairal, & Monroy, 2019; Sanmartín et al., 2008). To explain this discrepancy between patterns of genetic diversity and phylogeographic patterns, Caujapé-Castells (2011) proposed an alternative model: the 'Surfing Syngameon Hypothesis' (SSH). According to this model, initial colonization likely followed an east-to-west stepping-stone model, congruent with the temporal and geographic position of the islands within the archipelago. Later, selection and stochasticity associated with the islands' geological ontogeny determined population bottlenecks and even the extinction of some populations. In the easternmost islands of Fuerteventura and Lanzarote, where the climatic oscillations of the Pleistocene led to a major increase in aridity, these two processes (bottlenecks and population extinction) probably facilitated the recolonization by populations from Tenerife and Gran Canaria, and from the mainland (García-Verdugo, Caujapé-Castells, Mairal, & Monroy, 2019). As a result, the initial pattern of east-to-west genetic differentiation across the archipelago got erased. Concurrently, all islands but especially Lanzarote and Fuerteventura, accumulated greater levels of genetic variation through frequent inter-island dispersal and ongoing colonization events from nearby northwestern Africa. Within-island gene flow and secondary contact among initially isolated genotypes also contributed to patterns of genetic admixture (Caujapé-Castells, 2011; Caujapé-Castells et al., 2017).

Unlike the Hawaiian archipelago, the proximity of the Canarian Archipelago to a major continental landmass has resulted in the Canarian endemics having their closest phylogenetic relatives in nearby northwestern Africa, other Macaronesian archipelagos, or in the Mediterranean region (Caujapé-Castells et al., 2022). One exception is the 'Rand Flora' species (Fernández-Palacios et al., 2024; Sanmartín et al., 2010). These are Canarian endemics that have their closest relatives in geographically distant regions on the eastern and southern margins of the African continent: for example the Horn of Africa/southern Arabia, eastern Africa and southern Africa (Andrus et al., 2004; Pokorny et al., 2015; Thiv et al., 2010). Recent studies based on molecular phylogenies and species distribution models suggest that the Rand Flora pattern was formed by fragmentation of a former widespread African distribution by climatic extinction linked to aridification events (Culshaw, Mairal, & Sanmartín, 2021; Culshaw, Villaverde, et al., 2021; Mairal et al., 2017; Mairal, Pokorny, et al., 2015; Pokorny et al., 2015; Riina et al., 2021; Rincón-Barrado et al., 2021; Villaverde et al., 2018). Global climatic changes starting in the mid-Miocene led

to the aridification of northern, southern and eastern Africa, which climaxed with the formation of the Sahara Desert c. 8–7 Ma. The result was the replacement of the original subtropical African ('Gondwanan') vegetation by more xeric elements (Milne & Abbott, 2002; Senut et al., 2009; Sepulchre et al., 2006) and the current ring-like distribution of the Rand Flora clades, with species isolated in the continental African margins (Pokorny et al., 2015).

Many Canarian plant endemics are only found in the more humid and higher western-central islands (Caujapé-Castells et al., 2022). *Kleinia neriifolia* Haw. (Asteraceae) and *Euphorbia balsamifera* Ait. (Euphorbiaceae) are exceptions to this trend because they are present in all seven major islands, including the dry eastern Canary Islands, thus representing rare 'ubiquitous MIEs' (Caujapé-Castells et al., 2022). Both species belong to clades exhibiting the Rand Flora pattern: close phylogenetic relationships between species occurring on opposite sides of the Sahara Desert (Pokorny et al., 2015). *Kleinia neriifolia* is a succulent shrub that grows in lowland areas of the archipelago (Halliday, 1986; Figure 1). *Kleinia* Mill. is a genus within the Senecioneae tribe (Cicuzza et al., 2018; Compositae Working Group, 2024; Pelser et al., 2007; Timonin et al., 2015), which comprises c. 50 succulent species distributed in the arid and semiarid regions of southern Africa and Madagascar, in the Horn of Africa, Arabian Peninsula and Socotra Island in the east, as well as in western Sahel, the coasts of Morocco and Western Sahara, and the Canary Islands in the west (Figure 1); a few species are restricted to India and southeastern Asia (Compositae Working Group, 2024). Phylogenetic analyses based on Sanger sequencing of a limited number of hypervariable plastid markers and the ITS multicopy nuclear region, recovered *K. neriifolia* as the sister species of the northwestern African endemic *Kleinia anteuphorbium* (L.) DC. and, in turn, these two species form the sister clade to a group of species distributed in the Horn of Africa and southern Arabia (Cicuzza et al., 2018; Pelser et al., 2007; Pokorny et al., 2015; Timonin et al., 2015). At the population level, two phylogeographic studies, based on microsatellites (Sun & Vargas-Mendoza, 2017) or a few plastid loci (García-Verdugo, Caujapé-Castells, Mairal, & Monroy, 2019), recovered an east–west phylogeographic split within *K. neriifolia*, similar to the one observed in the Rand Flora endemic *Canarina canariensis* (Campanulaceae, Mairal, Pokorny et al., 2015; Mairal, Sanmartín et al., 2015).

Statistical comparison of phylogeographic scenarios have typically employed Approximate Bayesian Computation (ABC; Beaumont, 2010; Overcast et al., 2017; Patiño et al., 2015; Perez et al., 2016). These are likelihood-free, coalescent-based approaches in which genetic data simulated under alternative phylogeographic scenarios are compared with the empirical data to approximate the posterior distribution of the model parameters and identify the most likely scenario. In recent

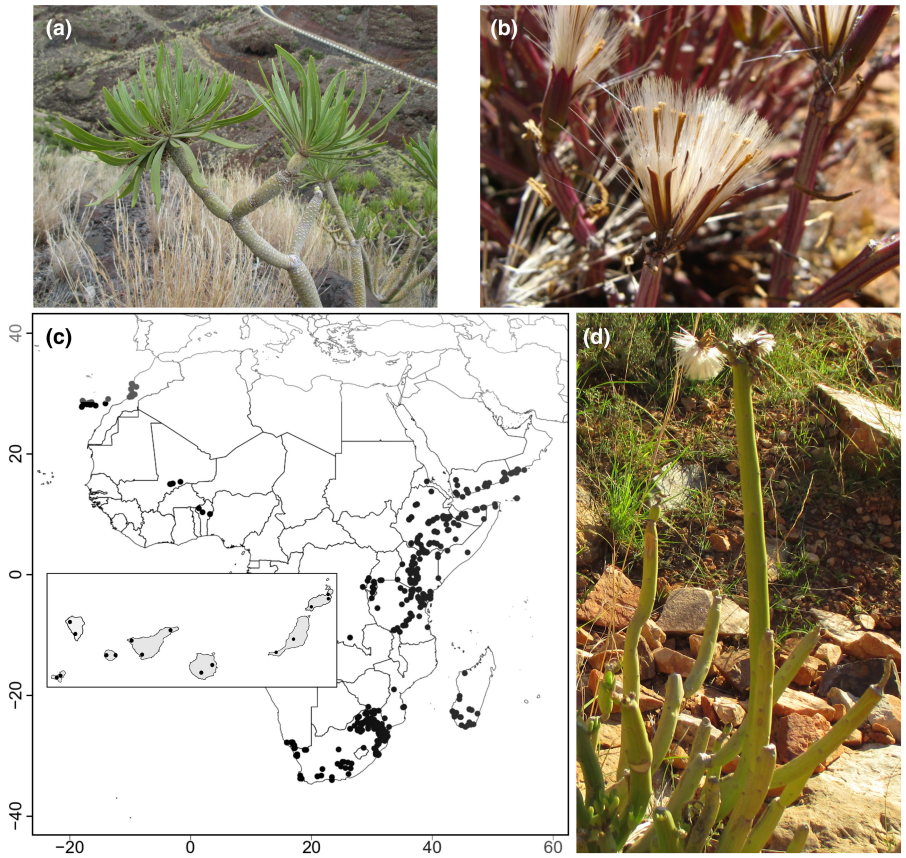


FIGURE 1 (a) Leafy shrub of *Kleinia neriifolia*. (b) Detail of the pappus of *K. longiflora*. (c) Distribution of genus *Kleinia* in Africa and nearby archipelagos. (d) Succulent branches and pappus of *K. anteuphorbium*. Distribution data were obtained from GBIF.org (7 June 2021; Download https://doi.org/10.15468/dl.2ww8dk), and from herbarium specimens studied. The inset shows the Canary Islands with dots marking the sampling localities. Photographs by C. García-Verdugo (*K. neriifolia* and *K. anteuphorbium*) and A. Dreyer (*K. longiflora*).

years, (deep) machine learning algorithms such as convolutional neural networks (CNN) (LeCun et al., 1998) have been proposed as an alternative to ABC approaches for phylogeographic inference. This machine learning approach relies on image pattern recognition and has proven useful to capture the patterns of variance/covariance among sites within a genetic dataset (Flagel et al., 2019; Perez et al., 2022). In contrast, ABC approaches rely on one or several summary statistics (e.g. heterozygosity level, allele richness) and suffer from the 'curse of dimensionality', that is the difficulty of reducing a complex genetic dataset to a few axes of genetic variation (Flagel et al., 2019). So far, CNNs have been successfully used to infer the history of population splits, the timing of demographic expansion/contraction events, and the magnitude of subsequent gene flow events in continental and island settings (Kirschner et al., 2022; Perez et al., 2022; Rincón-Barrado et al., 2024).

Here, we used high throughput genomic techniques, CNN, and a representative sample of *K. neriifolia* and its African–southern Arabian relatives to: (a) test the sister-group relationship between *K. neriifolia* and northwestern African *K. anteuphorbium*, (b) infer phylogenetic relationships and divergence times among *K. neriifolia* and its relatives, and (c) compare the models of island colonization and diversification postulated in previous studies to infer the evolutionary history of *K. neriifolia* in the Canarian archipelago.

2 | MATERIALS AND METHODS

2.1 | Taxon sampling

Through several field campaigns (2016, 2017 and 2018), we obtained a total of 103 specimens of *K. neriifolia* and *K. anteuphorbium*, and we added 11 samples from cultivated plants to represent the species *K. butleri* Lavranos & T.A. McCoy, *K. lauschsii* Lavranos & T.A. McCoy, *K. odora* DC., *K. saginata* P. Halliday, *K. scottii* (Balf.fil.) P. Halliday, *K. squarrosa* Cufod. and *K. semperviva* (Forssk.) DC. from eastern Africa, southern Arabia and Socotra, and three herbarium samples to include closely related genera within Senecioneae (*Solanecio* Sch. Bip. ex Walp., *Gynura* Cass., *Senecio* L.; Pelser et al., 2007, 2010; Timonin et al., 2015), totalling 121 samples for the study. Of these samples, 84 were from *K. neriifolia*, obtained from 16 localities across all seven main Canary islands (Figure 1c). We sequenced individuals from two localities per island, except in the case of Lanzarote and Tenerife, for which we obtained sequences of plants from three localities on each island. Nineteen samples of *K. anteuphorbium* were collected from three localities in Morocco and Western Sahara. Details on herbarium vouchers and geographic origin (provenance) for each sample are provided in Table S1.

2.2 | Probes and sequence capture

We used the bait design of Mandel et al. (2014) to capture the coding sequences of hundreds of orthologous genes across the genome of *Kleinia*. These baits were designed to perform Hyb-Seq in Asteraceae, using conserved orthologous genomic sequences from three species: *Lactuca sativa* L., *Helianthus annuus* L. and *Carthamus tinctorius* L., allowing hybrid capture of 1061 low-copy nuclear loci (Mandel et al., 2014). Genomic DNA extraction was performed on freshly collected, silica-dried samples or from dry herbarium material. To eliminate polysaccharides and secondary metabolites from the fresh, silica-dried *Kleinia* samples, we pre-treated them as indicated in Sun et al. (2009). Then, the genomic DNA extraction was performed as described in Dellaporta et al. (1983).

The construction of Illumina libraries, hybridization and next-generation sequencing steps were performed by Daicel Arbor Biosciences (Ann Arbor, Michigan, USA). DNA was sonicated using a QSonica Q800R instrument and size-selected to the desired target fragment of 550 base pairs (bp) using SPRI beads. Sonicated samples were then converted to an Illumina TruSeq-style sequencing library (Illumina Inc., San Diego, CA, USA). For the hybridization, the indexed libraries were pooled in equimolar amounts and accounting for proximity (i.e. same or nearby population/s) within *K. anteuphorbium* and *K. neriifolia* to maximize capture success; an exception was made with the eastern African samples, which were pooled together because we only sampled one or two individuals per species (macroevolutionary level). In total, we used 10 pooled libraries: eight for *K. neriifolia*, with 10 samples per pool; one for *K. anteuphorbium*, with 25 samples; and one pool for all the eastern African sampled species. Pooled libraries were enriched with the Asteraceae baits following the myBaits manual version 4. Enriched pools were then sequenced in the Illumina NovaSeq S4 platform.

2.3 | Data processing and phylogenetic analyses based on DNA sequences

Sequences were filtered according to their quality, and adapter sequences were removed using TRIMMOMATIC (Bolger et al., 2014). To assemble reads into contigs, we used the HybPiper pipeline 1.3 (Johnson et al., 2016), using default setting –bwa. This pipeline assembles the sequences with SPAdes 3.13.1 (Bankevich et al., 2012) and then uses exonerate v2.2 (Slater & Birney, 2005) to eliminate introns sequences from the contigs. We obtained DNA sequences from 952 of the 1061 targeted loci (Mandel et al., 2014). Percentages of captures can be consulted in Table S2 and Figure S1. Capture success per sample ranged from 66.22% to 86.34% and per locus from 100% to 0.82%. We used the paralog investigator tool included in the same pipeline to determine the number of possible paralog sequences for each locus. Two loci with paralogs were detected and removed from subsequent analyses.

Orthologous exon sequences were aligned into individual matrices using MAFFT 7.310 (Katoh & Standley, 2013). We visually evaluated the alignments in Geneious 10.2 (Kearse et al., 2012) and discarded those with poor quality, those pairwise alignments with less than 30% of identical sites, and if the aligned pairwise identity was less than 65%. As a result, 180 loci were removed, leaving a

dataset of 679 loci with a concatenated total length of 427,992 bp. The length of the aligned sequences of these loci ranged from 93 to 3098 bp. We constructed two different datasets and analysed them using different methods for reconstructing phylogenetic relationships using large-scale genomic data. For each of the 679 matrices, an individual gene tree was generated using maximum likelihood (ML) for statistical inference, as implemented in RAxML 8.2.11 (Stamatakis, 2014), with the GTRCAT model, 200 bootstraps and slow ML optimization (-fa option). In a second step, these individual ML gene trees were input into ASTRAL 2.5.7.3 (Mirarab & Warnow, 2015) to estimate a phylogenetic species tree using the default options. Clade support was assessed by estimating local posterior probability (LPP) values per node. In addition, the 679 loci individual matrices were concatenated into a single supermatrix in Geneious, and analysed under ML inference to build a phylogenetic tree, as implemented in IQ-TREE 1.6.1 (Nguyen et al., 2015) with options: 1000 ultrafast bootstrap replications (Hoang et al., 2018; -bb option), standard model selection using ModelFinder (Kalyaanamoorthy et al., 2017; -m TEST option), and an SH-like approximate likelihood ratio test (-alrt 1000 option). In the IQ-TREE and ASTRAL analyses, the phylogenetic tree was rooted at the most distant clade, represented by the genus *Senecio*.

One advantage of Hyb-Seq is the possibility to obtain plastid data as a by-product (Weitemier et al., 2014), even if the final libraries are not re-enriched with unfiltered libraries (see Villaverde et al., 2018). Chloroplast sequences were recovered using closely related plastomes as target regions in HybPiper with default parameters. We constructed our target file using the coding sequences of plastid genes from the three Asteraceae species (*H. annuus* [NC_007977], *L. sativa* [NC_007578] and *C. tinctorius* [KM207677]) for a total of 216 target sequences, 72 sequences per species. We obtained 72 exon matrices from 121 individuals belonging to 12 *Kleinia* species and 4 outgroups. The capture percentages for each locus and each individual can be found in Figure S2 and Table S3. The largest sequence length was 53,625 bp, while the one with the highest percentage of missing data included 3789 bp. The 72 exon matrices were concatenated into a single supermatrix of 54,276 bp and analysed in IQ-TREE, using the same settings as above. The raw reads are available in GenBank under BioProject PRJNA1167908.

2.4 | SNP calling

A reference genome for *Kleinia* is currently unavailable, and therefore, SNP calling was performed using a modified protocol from SECAPR (Andermann et al., 2018), which employs a pseudo-reference genome built from the target raw data. The SECAPR tool is based on PHYLUCE (Faircloth, 2016), and it executes SNP calling analysis in three different steps. First, the SECAPR pipeline assembles the DNA reads into de novo contigs instead of using reference sequences (as in HybPiper). Next, for each locus, the pipeline creates an aligned matrix with the sequences of all individuals and generates a consensus sequence per locus. This consensus sequence is used as a pseudo-reference to reassemble the reads again, increasing the coverage of the sequences. With the sequences of each individual assembled, SECAPR phases the reads into two different sequences, one for each allele, and then extracts the SNPs from these phased sequences. We modified the original pipeline, skipping the first step, the de novo DNA read assembly, and using instead the nuclear sequences previously obtained from HybPiper. The HybPiper nuclear sequences were then aligned in SECAPR to obtain a consensus sequence per locus. This way, we could ensure that the DNA-aligned sequences used for SNP calling do not contain assembling errors or paralogs, while skipping a computing-intensive step in SECAPR. Reads used for the reassembly were trimmed and quality filtered as mentioned before for the phylogenetic analyses. We discarded sequences with a coverage lower than five reads.

All retrieved SNPs were filtered by discarding all sites with more than 35% of missing data, that is, accepting sites represented in more than 65% of samples analysed, which have also a minimum coverage of 5 and a quality cut-off of 20. In our experience with target sequencing, within a recovered gene sequence there are regions containing a large number of SNPs, which might be due to either 'true' nucleotide changes or artefacts introduced by sequencing, and which are separated by regions with none or fewer SNPs. To minimize the possibility of linkage among our SNPs, a requirement for many of the genetic algorithms used below, we carried out a strategy based on 'windows' of genetic variance. See Supporting Information for the complete SNP filtering methodology.

2.5 | Divergence time estimation

Lineage divergence times among species of *Kleinia* were estimated using the nuclear exon matrix and a reduced dataset of 18 individuals, after removing most of the samples of *K. neriifolia*, *K. anteuphorbium* and *K. butleri*; we left two samples to test the species' monophyly. We only had one sample for the remaining species in *Kleinia*: *K. abyssinica* (A. Rich.) A. Berger, *K. semperviva*, *K. pendula* (Forssk.) DC., *K. lauschsii*, *K. odora*, *K. saginata*, *K. squarrosa* and *K. scottii*, and for the outgroup species, *Senecio adonidifolius* Loisel., *S. legionensis* Lange, *Solanecio mannii* (Hook.f.) C. Jeffrey and *Gynura valeriana* Oliv. After discarding alignments of poor quality, that is the aligned pairwise identity was less than 65% and the percentage of identical sites was less than 30%, we ended up with a matrix of 809 nuclear exon loci. This matrix was analysed in BEAST 1.8 (Drummond et al., 2012) using a birth–death process with incomplete taxon sampling (Stadler, 2009), a GTR substitution model with across-site rate heterogeneity, and an uncorrelated lognormal molecular clock as Bayesian priors, based on previous studies of *Kleinia* and other Rand Flora lineages (Pokorny et al., 2015; Rincón-Barrado et al., 2021). Because there are no known fossils from *Kleinia*, secondary age estimates were used as calibration points. The most complete and recent dating of the genus can be found in Pokorny et al. (2015). As the root node calibration, we used the crown age of subtribe Senecioninae (Pelser et al., 2007), which has an age of 27.20 Ma and a 95% high

probability density (HPD) = 14.67–48.01 Ma in Pokorny et al. (2015); this node encompasses all sequenced taxa in our dataset. We used a normal distribution prior with a mean age of 27.2 Ma and a standard deviation of 10.00, spanning the HPD in Pokorny et al.'s dating. Monophyly was not enforced for any clade. A Bayesian Markov Chain Monte Carlo (MCMC) analysis was run for 100 million generations, sampling every 1000th generation. Tracer 1.7.2 was used to check for adequate mixing and that all parameters reached at least a value of 200 for the effective sample size (Suchard et al., 2018). A maximum clade credibility (MMC) tree with clade posterior support was obtained with TREEANNOTATOR 1.8.0, after discarding 10% of samples as burn-in.

2.6 | SNP-based phylogenetics and population genetic analysis

Phylogenetic relationships based on the nuclear SNPs matrix were reconstructed using the SVDquartets method (Chifman & Kubatko, 2014), implemented in PAUP 4.0 (Swofford, 2018). Each site in the alignment is assumed to evolve under its own underlying gene tree, generated under the coalescent model from the species tree, so it can accommodate incomplete lineage sorting and different gene coalescence times. SVDQ can also accommodate other sources of gene-tree/species-tree incongruence (Peng et al., 2022) such as gene flow (e.g. introgression between sister taxa; Long & Kubatko, 2018, 2019). We used the option 'evaluating 20,000 quartets', or 'all possible quartets if lower than 20,000', and evaluated clade support by running 100 bootstrap replicates, using the multispecies coalescent (MSC) option and without assigning samples to distinct lineages.

To evaluate population structure within *K. neriifolia*, we used the population genetics method sNMF (Frichot et al., 2014) included in R Package *Lea* (Frichot & François, 2015). This method has the advantage over other methods that estimate shared ancestry (Alexander et al., 2009; Pritchard et al., 2000) that it makes no assumption about the absence of genetic drift or Hardy–Weinberg (HW) equilibrium among the ancestral populations (Frichot et al., 2014). We run sNMF on our SNPs dataset with 65% representation to estimate the percentage of homozygous and heterozygous loci, and determine the number of ancestral populations (K) using the cross-entropy criterion. We did this 100 times for different levels of K (from 1 to 15) using the recommended value of α (non-negative regularization parameter) = 100 for matrices of moderate size (Frichot et al., 2014). We selected the run with the least entropy, which had a K level of 3. This run separated individuals into an 'eastern group' formed by seven populations from Fuerteventura (2), Lanzarote (3) and Gran Canaria (2); a 'western group' comprising six populations (two per island) from La Palma, El Hierro and La Gomera; and a third group composed of three populations from Tenerife. To capture fine-scale population structure, we performed two subsequent analyses within the eastern and western groups, using the same parameters as above. Samples from Tenerife were not included in these subsampling analyses because their inclusion consistently resulted in a split ($K = 2$) between Tenerife and the other islands, hindering the detection of fine population structure, and contributing no additional information to the previous analysis.

To detect genetic barriers between populations and represent them graphically, we used Monmonier's maximum difference algorithm (Barbujani et al., 1996), implemented in the R Package *adegenet* 2.1.3 (Manni & Barrai, 2000): This method identifies the edges or breaks between populations where genetic differences are maximal. Genetic distance was calculated using the SNPs matrix and geographic coordinates included in Table S1. We used the Delauney triangulation (Brassel & Reif, 1979) algorithm for the geometric connection network between the different populations. Delaunay triangulation only works when all geographic coordinates are different, and so, it is not suitable for datasets with multiple individuals from the same population. To avoid this limitation, we used the jitter function (jitter(factor = 1.5)), included in R, which adds some noise to the coordinates.

2.7 | Genetic diversity analysis

We used the R Package *DartR* (Gruber et al., 2018) to calculate genetic diversity statistics using the SNP dataset, such as allelic richness (A_r), observed heterozygosity (H_o) and expected heterozygosity (H_e). Theoretical and empirical studies show that island populations displaying higher migration rates and introgression have higher values of allelic richness and heterozygosity (Caujapé-Castells et al., 2017, 2022). Therefore, assuming no extinction or population decline, islands that were colonized first should have higher values of genetic diversity if they have kept receiving further colonizers and hosting more introgression than more recently colonized islands, where we may generally expect smaller values of genetic diversity. We also estimated Tajima's D parameter using the software DNAsp 6.0 (Rozas et al., 2017) and the aligned nuclear exon matrices. This parameter is often used to obtain information about potential underlying processes explaining population dynamics, such as population bottleneck, demographic expansion or allele-level selection.

2.8 | Phylogeographic model testing using convolutional neural networks

We tested four phylogeographic scenarios that represent the three main hypotheses about the history of colonization and diversification in the Canary Islands described in the introduction (Figure 2). The first scenario depicts the SSC model, in which colonization starts on the easternmost islands of Lanzarote and Fuerteventura and proceeds along the island chain, from east to west. The second scenario simulates the 'SSH' model (Figure 2), with initial colonization following a stepping-stone model as in SSC. However, subsequent bottlenecks or even extinction of some populations in the easternmost

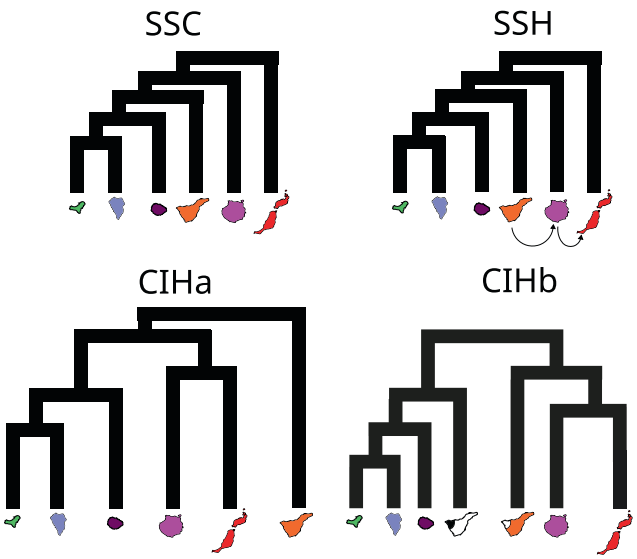


FIGURE 2 Schematic representation of the phylogeographic models of island colonization and diversification compared in our study using Convoluted Neural Networks. SSC, stepping-stone colonization model, with dispersal proceeding from east to west across the archipelago. SSH, Surfing Syngameon Hypothesis, with initial east-to-west colonization followed by recolonization of the eastern islands of Lanzarote and Fuerteventura from Tenerife and Gran Canaria. The arrows depict the direction of secondary migrations as outlined by this model. CIHa, Central Island Hub colonization model from Tenerife, with all populations from Tenerife grouped into a single island unit. CIHb, Central Island Hub colonization model, with the populations from Tenerife separated into an eastern-central and a western half. Under each model, a schematic representation of the islands and island units is depicted.

islands results in reduced competition, and facilitates recolonization by propagules from the central (particularly, Gran Canaria) to the easternmost islands. The third scenario is the CIH model (Figure 2), in which Tenerife is the first island to be colonized. We tested two different versions of this model: CIHa: all populations from Tenerife are grouped into a single unit; CIHb: populations from Tenerife are separated into an eastern-central and a western half (Figure 2). After the colonization of Tenerife, island dispersal proceeded in a stepwise sequence, to the west, from La Gomera to La Palma and El Hierro, and to the east, towards Gran Canaria, Fuerteventura and Lanzarote. Because we found that populations within each island formed a monophyletic group (see Section 3), each island was considered a 'single population' in our CNN models. The only exception was Lanzarote and Fuerteventura: phylogenetic analysis using either DNA sequences or SNPs showed individuals from these islands intermingled and embedded in a single clade, without any geographic structure. This fact, and the geological history of these islands (i.e. they were united into a single island in some periods, Fernández-Palacios et al., 2011), led to their inclusion in our analyses as a single population. We performed 10,000 coalescent simulations for each model in Figure 2 using the software *ms* (Hudson, 2002), with sample sizes according to the empirical dataset and 1000 SNPs simulated independently, as in Perez et al. (2022). The overall mutation rate was set to 0.001 substitution events per Ma, based on results from the BEAST analysis. We sampled the root age for our island diversification scenarios from a uniform distribution containing the estimated crown age and 95% HPD interval for the start of island diversification in *K. neriifolia* from the BEAST analysis. Generation time was sampled from a uniform distribution between 5 and 6 years, according to average values suggested by Pelser et al. (2010) for genus *Kleinia*. The migration events in the SSH model were sampled from a uniform distribution ranging from 0 to 10 migrants per generation. We followed Kirschner et al. (2022) to randomly add missing SNPs to our simulations, mimicking the empirical dataset. For other island scenario-specific priors, see Appendix S1.

We used a similar network architecture to Perez et al. (2022), including three 1D-convolutional layers with a kernel size of 3; the first layer had 125 neurons, the second and the remaining layers, 250 neurons each. CNN performance was assessed with a categorical cross-entropy loss function. We randomly divided the simulated datasets into two sets: 75% of the simulations were used as the training set (i.e. to train the CNN) and the remaining 25% were used as the validation set (i.e. to assess the accuracy of the CNN algorithm in predicting the simulated scenario). To avoid overfitting, we used two approaches based on the validation step: model checkpoint, which saves only the best model, and early stopping, which tolerates a maximum of 150 epochs without any improvement in the accuracy before stopping. Appendix S1 include further methodological details on the neural network architecture. In the last step, the trained network was used to predict the most probable model of island colonization for *K. neriifolia*, using 100 randomly sampled datasets of 1000 SNPs from the empirical data, and a new set of 10,000 independent simulations per scenario (the test set) not evaluated by the network during the training. Cross-validation runs were performed with 10,000 simulations per scenario. All scripts used to perform our CNN approaches are available at: https://github.com/manolofperez/CNN_Kleinia.

3 | RESULTS

3.1 | Phylogenetic analyses based on DNA sequences

Phylogenetic analyses based on the nuclear exon matrices yielded highly resolved phylogenetic trees, irrespective of the method of phylogenetic inference employed (Figures 3 and 4). Many backbone nodes received high support values, and relationships at the species level were generally congruent between methods approximating the MSC model (ASTRAL; Figure 3 and Figure S3) and those based on ML using concatenated sequences (IQ-TREE; Figure 4). Our analysis places the representative species of the genus *Solanecio* as the sister-genus of *Kleinia* and the genus *Gynura* as sister to them (Figures 3 and 4); however, the relationship between *Solanecio* and *Kleinia* is only well supported in IQ-TREE (Bootstrap Support [BS] = 98; Figure 4). *Kleinia* was recovered as a monophyletic group with high support

in all phylogenies (LPP/BS = 1.0/100). Within *Kleinia*, we recovered three main clades with the highest support (LPP/BS = 1.0/100), highlighted as A, B and C in Figures 3 and 4. Clade A comprises two individuals of *K. abyssinica* and is sister to the remaining *Kleinia* species in clades B and C (LPP/BS = 1.0/100). Clade B groups the eastern African species *K. pendula* with the southern Arabian species *K. butleri*, *K. lauschsii* and *K. saginata* (LPP/BS = 1.0/100), although relationships among the three last species were different between the ASTRAL and IQ-TREE phylogenies (Figures 3 and 4). Within clade C, the northwestern African species *K. anteuphorbium* forms a monophyletic group with the Canarian endemic *K. neriifolia* (LPP/BS = 1.0/100). In ASTRAL (Figure 3), this clade is sister to *K. squarrosa* IS1779 (LPP = 0.84). In IQ-TREE (Figure 4), the sister group to *K. neriifolia*-*K. anteuphorbium* is a clade (BS = 100), grouping the specimens of the southern Arabian species *K. odora* (Oman), the Socotran endemic *K. scottii*, and the two specimens of *K. squarrosa* (BS = 97). Sister to these species in clade C is *K. semperviva* from Yemen (LPP/BS = 1.0/100; Figures 3 and 4). The two samples of *K. squarrosa* included in our analyses (IS1779 and IS1776) did not form a monophyletic group, with the position of accession IS1779 disagreeing between ASTRAL (Figure 3) and IQ-TREE (Figure 4). Unlike IS1779, the position of *K. squarrosa* IS1776 was strongly supported (LPP/BS = 0.99/100) in both the ASTRAL and IQ-TREE analyses (Figures 3 and 4).

Within *K. neriifolia*, population-level relationships were generally recovered with high clade support, especially at the island level. Both ASTRAL and IQ-TREE recover three major 'island' clades (Figures 3 and 4): an 'Eastern clade' (EC) grouping populations from the islands east of Tenerife, that is, Gran Canaria, Fuerteventura and Lanzarote (LPP/BS = 1.0/100), and a 'Western clade' (WC), grouping populations from the islands west of Tenerife, that is La Gomera, La Palma and El Hierro (LPP/BS = 0.98/100). Populations from the island of Tenerife formed their own clade (LPP/BS = 0.92/100; Figure 4 and Figure S3).

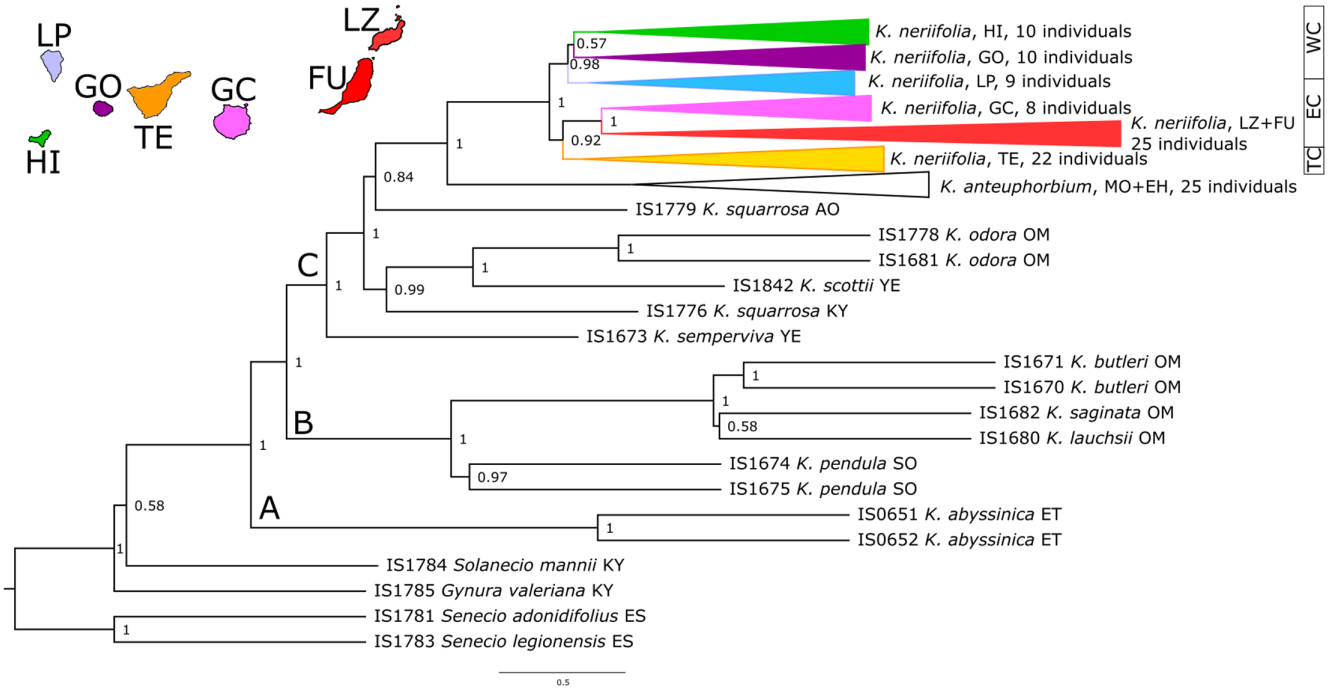
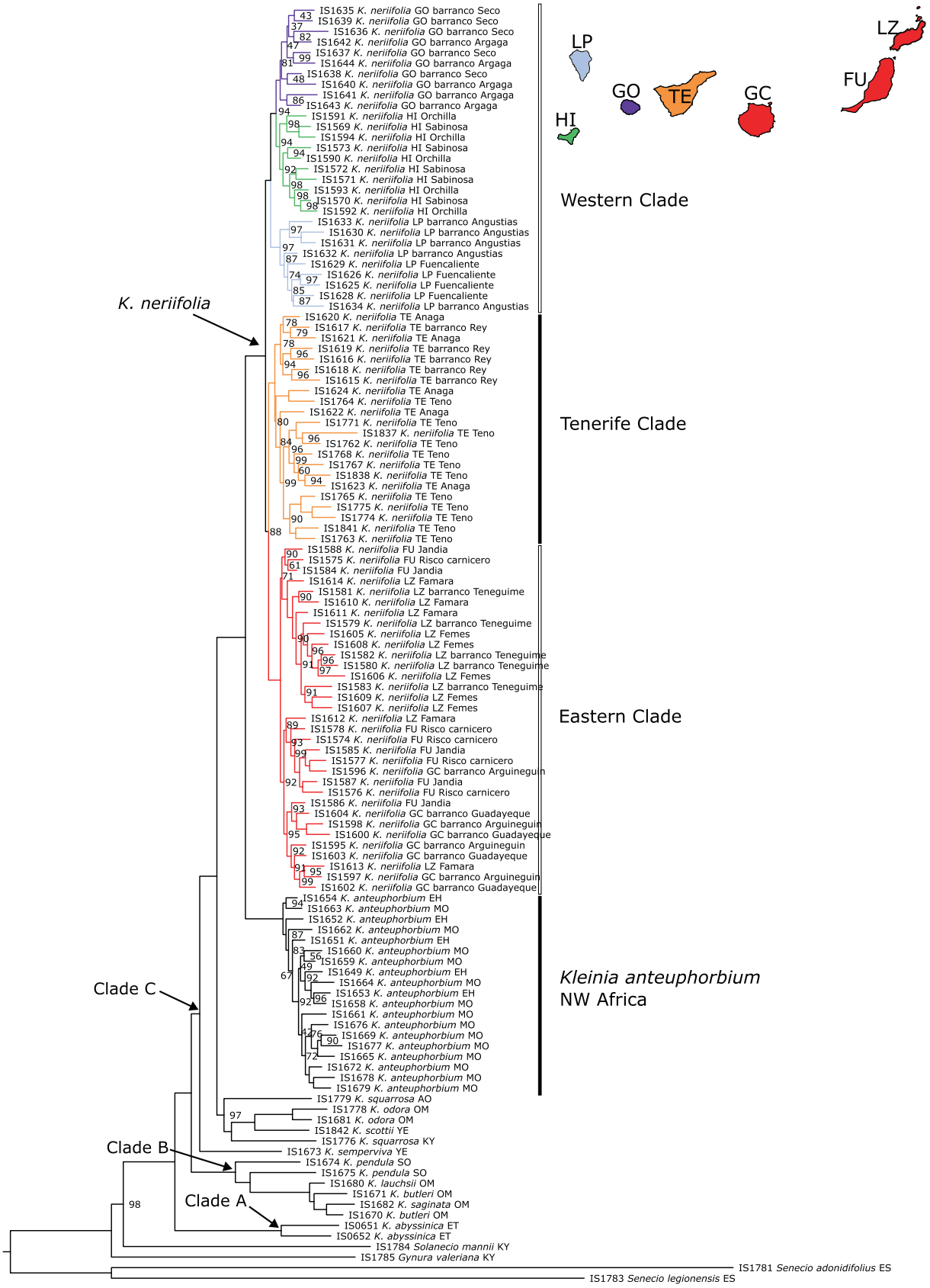


FIGURE 3 Multispecies coalescent model analysis in ASTRAL based on 679 nuclear exon gene trees generated with RAxML from 121 samples. Numbers at nodes indicate the Local Posterior Support (LPP) value; the support values for the collapsed clades are given in Figure S3. Major clades within *Kleinia neriifolia* are colour-coded according to the island of origin, as indicated in the inset map; these clades received the maximum support value (LPP = 1.0). FU, Fuerteventura; GC, Gran Canaria; GO, La Gomera; HI, El Hierro; LP, La Palma; LZ, Lanzarote; TE, Tenerife. Other regions: AO, Angola; EH, Western Sahara; ES, Spain; ET, Ethiopia; KY, Kenya; MO, Morocco; OM, Oman; SO, Somalia; YE, Yemen. The island clades recovered in our analysis are also indicated: Western Clade (WC), Eastern Clade (EC) and Tenerife Clade (TC). Detailed information on the origin of these samples can be found in Table S1.

FIGURE 4 Maximum likelihood tree obtained with IQ-TREE based on a concatenated nuclear DNA matrix of 427,992 bp with 679 loci from 121 samples of *Kleinia*. Node numbers indicate the Bootstrap Support (BS) value. Nodes without values received the maximum clade support (BS = 100). Branches of *K. neriifolia* are colour-coded according to the island of origin, as indicated in the inset map. FU, Fuerteventura; GC, Gran Canaria; GO, La Gomera; HI, El Hierro; LP, La Palma; LZ, Lanzarote; TE, Tenerife. Other regions: AO, Angola; EH, Western Sahara; ES, Spain; ET, Ethiopia; KY, Kenya; MO, Morocco; OM, Oman; SO, Somalia; YE, Yemen. Island clades recovered in our analysis (western, eastern and Tenerife clades) are also indicated. Detailed information on the origin of these samples can be found in Table S1.



The 'Tenerife clade' (TC) is sister to the EC clade in ASTRAL (LPP = 092, Figure 3 and Figure S3) and IQ-TREE (BS = 88; Figure 4). These two clades, in turn, are recovered as sister to the WC clade in both analyses (LPP/BS = 1.0/100). Populations within the WC clade were further structured into three individual subclades, corresponding to the three islands (i.e. individuals

from the same island were grouped together). The 'La Palma' subclade was recovered as the sister group to two sister subclades formed by individuals from La Gomera and El Hierro, respectively (Figures 3 and 4); however, this relationship is only supported in the IQ-TREE phylogeny (BS = 94, Figure 4). In the case of the EC clade, individuals from Lanzarote and Fuerteventura appear intermingled, regardless of their island of origin, in both ASTRAL and IQ-TREE (Figures 3 and 4 and Figure S3). Also, individuals from Gran Canaria form a subclade embedded within the EC clade in ASTRAL (Figure 3), albeit with low support (LPP = 0.51; Figure S3). Within each island, individuals belonging to different populations were clustered together, with very little geographic structure (Figure 4 and Figure S3).

Phylogenetic trees from the recovered exon plastid sequences (Figure 5) showed high phylogenetic support for many of the backbone nodes. However, some species-level relationships were different to those obtained from the nuclear exons (Figures 3 and 4). For example, *K. anteuphorbium* was recovered as the sister species of *K. semperviva* in the plastid tree (LPP = 0.99, Figure 5), instead of forming a clade with *K. neriifolia* (Figures 3 and 4). *Kleinia abyssinica* does not appear as sister to the remaining *Kleinia* species, as in the nuclear tree (clade A; Figures 3 and 4), but instead it is sister to *K. squarrosa* IS1776 (LPP = 0.99; Figure 5). Unlike in the IQ-TREE nuclear phylogeny (Figure 4), individuals of *K. neriifolia* in the IQ-TREE plastid tree did not appear grouped according to their island of origin (Figure S4).

3.2 | Divergence time estimation

The topology obtained from the BEAST analysis (Figure 6) was similar to the one recovered by ASTRAL and IQ-TREE (Figures 3 and 4). Clade support (Bayesian Posterior Probabilities) was highest (1.0), except in a few nodes. Confidence intervals for lineage divergence times were quite broad, especially for the basal nodes. The most recent common ancestor (MRCA) of all *Kleinia* samples was dated in the Tortonian period, during the Late Miocene (mean age = 11.63 Ma; 95% HPD 18.73–5.43 Ma). The age of the Rand Flora disjunction, represented by the split between *K. neriifolia* and *K. anteuphorbium* in Macaronesia-northwest Africa, and the clade formed by *K. scottii*, *K. odora* and *K. squarrosa* on the eastern side of the Sahara, was estimated at 6.36 Ma (95% HPD 10.52–1.86 Ma), in the Late Miocene Messinian period. The stem node of *K. neriifolia* (i.e. the divergence from *K. anteuphorbium*) was estimated at 4.23 Ma (95% HPD 7.01–1.30 Ma), in the Early Pliocene. The crown age (i.e. the MRCA of all sampled populations of *K. neriifolia*) was dated in the Early Pleistocene (2.21 Ma, 95% HPD 3.94–0.70 Ma).

3.3 | SNP-based phylogenetics and population genetic analysis

With our approximation to SNP calling, in which SNPs are separated from one another by at least 25 bp, we retrieved 989 SNPs



FIGURE 5 Phylogenetic relationships among the sampled species of genus *Kleinia* and selected outgroups (121 individuals) using the plastid supermatrix (72 loci) and analysed with IQ-TREE. Branches within *K. neriifolia* and *K. anteuphorbium* appear collapsed, since there was very little resolution among populations. Bootstrap support values are indicated at nodes. *Kleinia* species occurring in eastern Africa and southern Arabia are marked in blue; those inhabiting northwest Africa and the Canary Islands are marked in red. AO, Angola; ET, Ethiopia; KY, Kenya; OM, Oman; SO, Somalia; YE, Yemen. Detailed information on the origin of the samples can be found in Table S1.

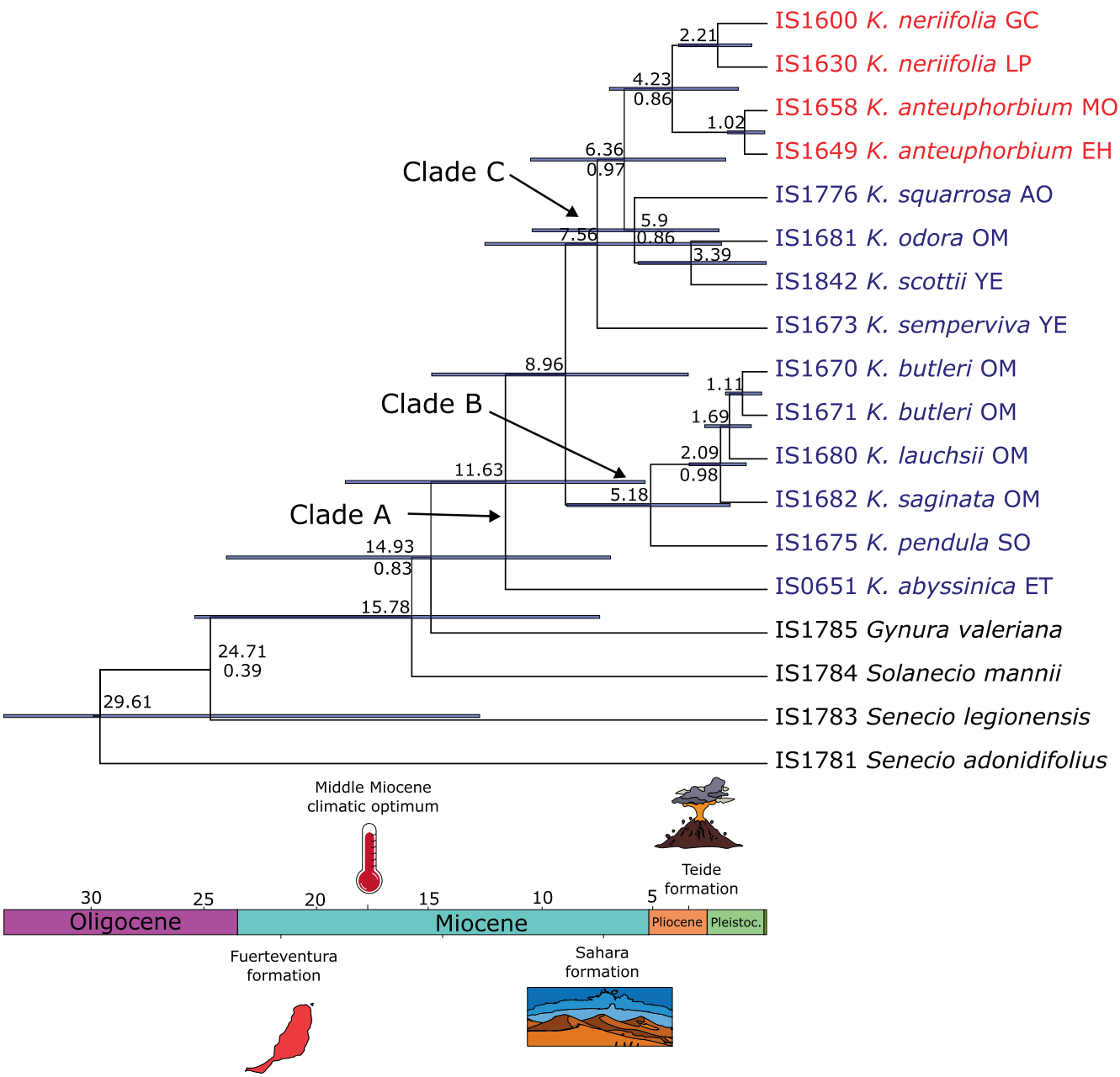
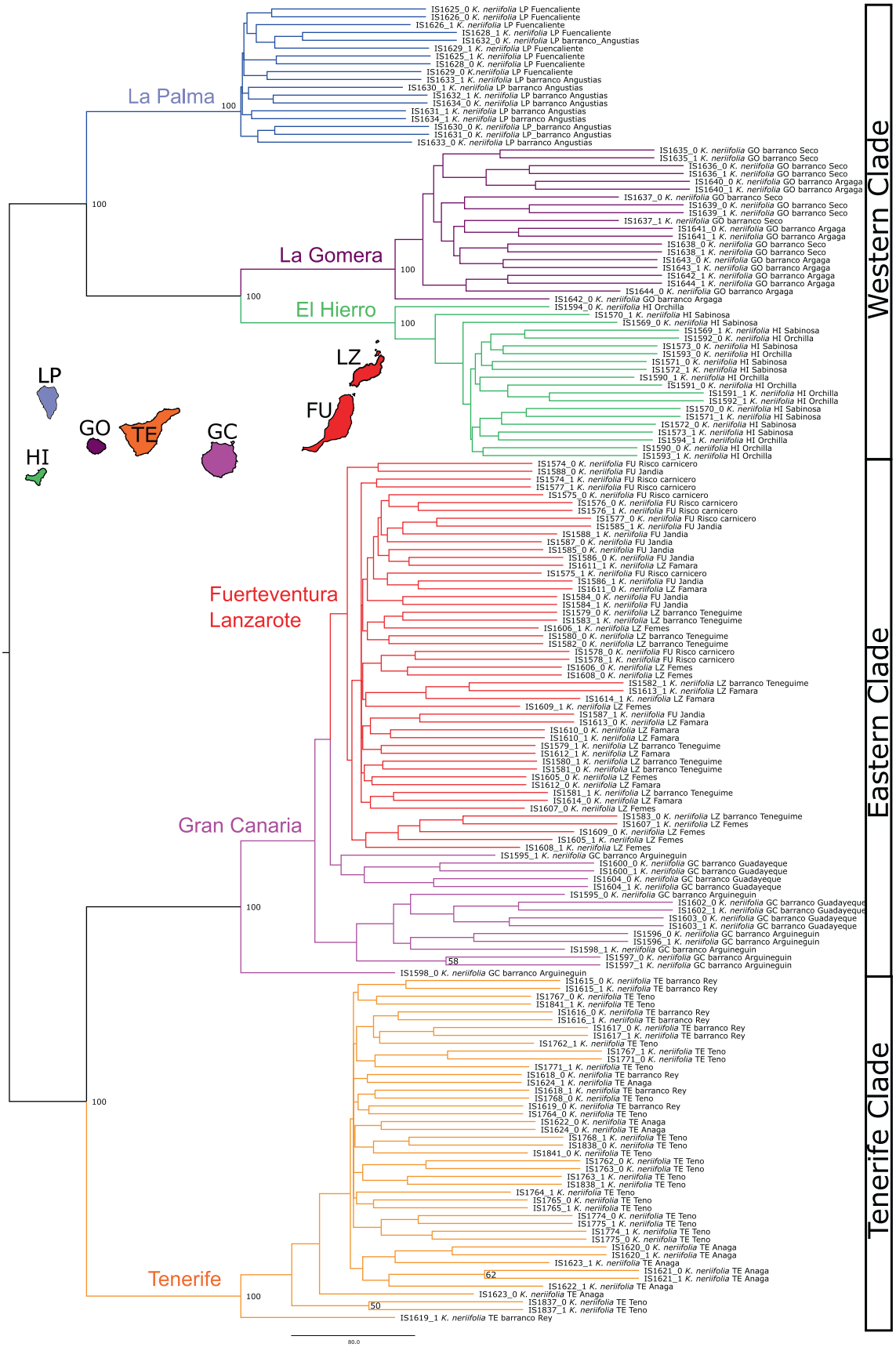


FIGURE 6 Maximum clade credibility (MCC) tree of *Kleinia* with lineage divergence times estimated in BEAST using 809 nuclear loci and a concatenated matrix of 459,659 bp. Numbers above nodes show the mean age estimates in million years; those below nodes show Bayesian posterior probabilities (PP) but only for values below 1.0. Node bars represent the 95% HPD credibility intervals around the mean age. *Kleinia* species from eastern Africa and southern Arabia are marked in blue, and those from northwest Africa and the Canary Islands in red. AO, Angola; EH, Western Sahara; ET, Ethiopia; GC, Gran Canaria; LP, La Palma; MO, Morocco; OM, Oman; SO, Somalia; YE, Yemen. Detailed information on the origin of these samples can be found in Table S1.

from the nuclear target loci that were present in all 84 samples of *K. neriifolia* (100% representation). When including SNPs represented in at least 65% of samples (i.e. a maximum of 35% missing data), we recovered a larger dataset of 1495 SNPs. The analysis of this latter dataset with SVDquartets inferred a phylogenetic tree (Figure 7) that was very similar in topology to those obtained with ASTRAL and IQ-TREE from the nuclear DNA sequences (Figures 3 and 4). In particular, we recovered the three major clades (WC, EC and TC), with TC being the sister to the EC clade, grouping Gran Canaria with the eastern islands of Fuerteventura and Lanzarote (Figure 7). Within EC, samples from Gran Canaria were recovered as paraphyletic relative to the samples from Fuerteventura and Lanzarote, which in turn formed one clade without any clear geographic structure, in line with the alignment-based nuclear analyses (Figures 3 and 4). Within WC, we recovered a similar phylogeographic structure to that obtained with ASTRAL and IQ-TREE for the full nuclear sequence datasets (Figures 3 and 4). Populations were grouped by island into three separate subclades:



the subclade formed by La Palma populations was recovered as sister to a clade comprising the two subclades of La Gomera and El Hierro (Figure 7).

This geographic structure was also captured by the sNMF analysis, where the most probable *K* value was 3 groups, corresponding to the three island clades: EC, WC and TC. Within the EC clade, there

FIGURE 7 SVDquartets phylogenetic tree based on 1495 single-nucleotide proteins (SNPs), showing relationships among the populations of *Kleinia neriifolia*. Numbers at nodes indicate the bootstrap support values when equal or higher than 50%. Branches are colour-coded according to the island of origin, as indicated in the inset map. In the absence of an outgroup, the tree was rooted using the results from the ASTRAL and IQ-TREE analyses, showing the Tenerife clade more closely related to the Eastern Clade than to the Western Clade. FU, Fuerteventura; GC, Gran Canaria; GO, La Gomera; HI, El Hierro; LP, La Palma; LZ, Lanzarote; TE, Tenerife. Island clades recovered in our analysis (Western, Eastern and Tenerife) are also indicated. Detailed information on the origin of these samples can be found in Table S1.

is a signal of admixture from the TC clade (Figure 8a). Within the EC and WC clades, there is also geographic structure: $K=3$ for the WC clade, corresponding to the islands La Palma, La Gomera and El Hierro (Figure 8b). A $K=2$ was recovered for the EC clade, supporting admixture from the Gran Canarian genotype into the admixed Lanzarote-Fuerteventura populations (Figure 8c). Monmonier algorithm (Figure 8d) detected a significant east/west genetic barrier separating the populations of EC and TC, to the east, from those of WC, to the west, in agreement with the SVDquartets analysis (Figure 7).

Analyses of genetic diversity within islands (Table 1) revealed high levels of genetic variation in the central island of Tenerife, both in terms of allelic richness (A_r) and expected heterozygosity levels (H_e). La Palma and Lanzarote, at opposite sides of the archipelago, also exhibited large diversity values for both H_e and A_r. The Tajima *D* value was found to be negative in all islands, indicating demographic expansion, but it was particularly large in Tenerife (−0.719) and Fuerteventura (−0.392), although none of these values were significant.

3.4 | Phylogeographic model testing using convolutional neural networks

The cross-validation procedure, using a test set of simulations not evaluated during the training step, showed a percentage of correct assignment to the simulated scenario larger than 80% for all phylogeographic models (Table 2), indicating high overall model accuracy for our CNN. The confusion matrix indicated the highest confusion levels between the SSC and the SSH model, whereas the two CIH models (CIHa and CIHb; Figure 2) showed a percentage of correct assignments higher than 98% (Table 2). Given the high level of accuracy in model prediction, we used the trained network in our empirical genetic dataset to identify the best-fit phylogeographic model for *K. neriifolia*. The CIHa scenario, with all Tenerife populations grouped into a single island unit, was selected as the phylogeographic scenario with the highest support (softmax = 1), whereas none of the other models received any support (softmax = 0).

4 | DISCUSSION

4.1 | Hyb-Seq phylogenomics support the Rand Flora origin of the Canarian endemic *Kleinia neriifolia*

Disentangling the origins of island endemic species, whether the product of ancient or recent colonization events from nearby or distant regions, can offer important insights into processes acting on community assembly in insular systems (García-Verdugo, Caujapé-Castells, Mairal, & Monroy, 2019). Our phylogenomic analysis, based on 679 low-copy nuclear loci and 427,992 nucleotides, supports Pokorny et al.'s (2015) assertion that the succulent genus *Kleinia* is one of several angiosperm lineages exhibiting the Rand Flora Afro-Macaronesian pattern. As in previous analyses based on Sanger sequencing (Ozerova et al., 2017; Pelser et al., 2007, 2010; Pokorny et al., 2015), the northwestern African endemic *K. anteuphorbium* is recovered as the sister species of the Canarian endemic *K. neriifolia*. There is not yet a taxonomically complete phylogeny of genus *Kleinia*. The Pelser et al. (2007) molecular phylogeny of Senecioneae, based on nuclear ITS, included 14 out of 50 species. *Kleinia neriifolia* was inferred as a sister to either *K. anteuphorbium* or a clade formed by this species and Madagascar *K. descoingsii* C. Jeffrey, but these relationships received low clade support (BS < 50%; PP < 0.70). Subsequent studies (Ozerova et al., 2017; Pelser et al., 2010) did not improve sampling or clade support. Pokorny et al. (2015) recovered a well-supported *K. neriifolia*–*K. anteuphorbium* clade in their ITS-trnLF time-tree, including 17 species. In their phylogeny, the two southern African species were not closely related to the 'Rand Flora clade' formed by *K. neriifolia*–*K. anteuphorbium* and a group of eastern African species (*K. descoingsii*, *K. semperviva*, *K. longiflora* DC. and *K. squarrosa*).

Although our taxon sampling of *Kleinia* is incomplete (we did not include any southern African species), we sequenced four of the six eastern African species represented within the Rand Flora clade of Pokorny et al.'s (2015) study. We additionally sequenced five species that have not been included in previous phylogenetic analyses: the eastern African/southern Arabian endemics *K. saginata*, *K. lauschsii* and *K. butleri*, as well as the Socotran endemic *K. scottii*; the latter is recovered within the Rand Flora clade in our analysis. Our study is also the first phylogenomic analysis including species and populations within genus *Kleinia*, based on the genome-wide sequencing of hundreds of protein-coding genes, which resulted in high levels of clade support for species relationships.

With the exception of the accession *K. squarrosa* IS1779, whose position varies depending on the inference method, all our results using nuclear information recovered a sister-group relationship between the Macaronesian-northwestern African clade formed by *K. neriifolia* and *K. anteuphorbium* and a clade of three species distributed on both sides of the Gulf of Aden and in the Eastern Arc Mountains, from Ethiopia to Tanzania (Figures 3, 4, 6 and Figure S3). The placement of *K. squarrosa* within the eastern Rand Flora clade (Pokorny et al., 2015) is more ambiguous in our analysis, though we noticed that the accession receiving the highest clade support,

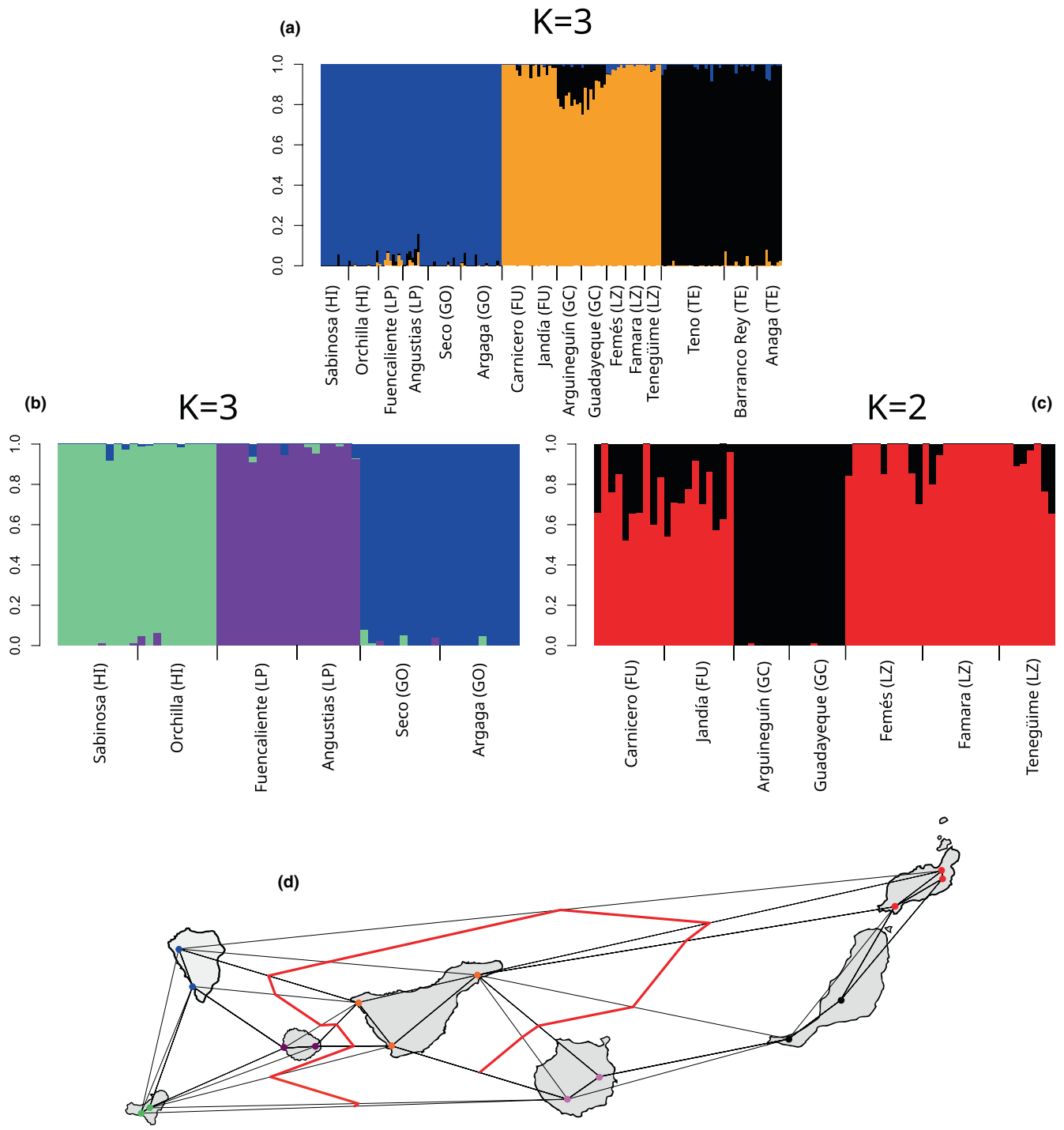


FIGURE 8 Clustering of individuals of *Kleinia neriifolia* based on 1495 SNPs using sNMF. Bar colours represent their estimated membership fractions. The *K* value was estimated using the entropy criterion. (a) SNMF analysis including populations from all islands ($K = 3$). (b) Analysis including only individuals from Western clade (WC) ($K = 3$). (c) Analysis including only populations the Eastern Clade (EC) ($K = 2$). In (b, c), Tenerife was excluded because the inclusion of this island resulted in a lack of geographically structured pattern. (d) Boundary detection using Monmonier's algorithm and 1495 SNPs of *K. neriifolia*. Black lines represent the connections between islands based on the Delaunay triangulation. The red lines represent the boundary found by the algorithm, separating the populations from the EC and WC groups; notice that Tenerife is not included in any of EC and WC analysis. FU, Fuerteventura; GC, Gran Canaria; GO, La Gomera; HI, El Hierro; LP, La Palma; LZ, Lanzarote; TE, Tenerife.

IS1776, is grouped with *K. odora* and *K. scottii* in the ASTRAL, IQ-TREE and BEAST phylogenies (Figures 3, 4 and 6). It is possible that the second accession of *K. squarrosa* (IS1779) is incorrectly identified in the private living collection that provided the samples. Both plants look very similar vegetatively (we were unable to observe these specimens in flower) and a misidentification cannot be ruled out.

TABLE 1 Population genetic diversity indices for populations of *Kleinia neriifolia* grouped by island; also indicated are the island clades recovered in our analysis: EC, Eastern Clade; TC, Tenerife Clade; WC, Western Clade.

	N° ind	H_o	H_e	A_r	N° sites	Tajima *D*
Fuerteventura (EC)	10	0.082	0.083	1.280	567.254	−0.392
Lanzarote (EC)	15	0.115	0.100	1.410	566.600	−0.144
Gran Canaria (EC)	8	0.061	0.078	1.272	573.035	−0.261
Tenerife (TC)	22	0.100	0.104	1.571	566.600	−0.719
La Gomera (WC)	10	0.070	0.081	1.279	566.600	−0.028
La Palma (WC)	9	0.124	0.108	1.390	569.850	−0.224
El Hierro (WC)	10	0.100	0.084	1.274	567.254	−0.090

Note: Allelic richness (A_r), observed heterozygosity (H_o), expected heterozygosity (H_e) and Tajima *D*; N° ind: number of individuals used to obtain the information in each island; N° sites: average number of sites per matrix used to calculate Tajima *D*.

TABLE 2 Cross-validation results for the four phylogeographic models depicted in Figure 2, using a convolutional neural network (CNN) approach.

		Predicted model			
		SSC	SSH	CIHa	CIHb
Simulated model	SSC	0.811	0.173	0.017	0
	SSH	0.140	0.842	0.018	0
	CIHa	0.000	0.013	0.987	0
	CIHb	0.0002	0.0026	0.0124	0.9848

Note: The green diagonal values in the confusion matrix indicate the percentage of datasets that are simulated and correctly predicted under the same model; the off-diagonal values indicate the confusion levels: The percentage of datasets that are simulated under one model and predicted as generated by a different model in the validation step.

Abbreviations: CIHa, Central Island Hub model with Tenerife as a single population; CIHb, Central Island Hub model with Tenerife divided into an eastern-central and a western half; SSC, stepping-stone colonization model; SSH, Surfing Syngameon Hypothesis model.

The crown age of the trans-Saharan disjunction in our analysis (6.36 Ma, Figure 6) is congruent with Pokorny et al.'s estimate (6.8 Ma) and falls around a period (Messinian) characterized by major aridification in Africa and worldwide (Bradshaw et al., 2012; Hansen et al., 2013; Jiménez-Moreno et al., 2019). Our results, therefore, support the climatic extinction hypothesis (Pokorny et al., 2015) as the most likely explanation for the divergence between the northwestern African/Canarian endemics and their sister clades in the distant regions of east Africa and southern Arabia/Socotra.

One advantage of Hyb-Seq is the possibility of obtaining plastid genes as a by-product (Villaverde et al., 2018). Similar to other phylogenomic studies on plants (Herrando-Moraira et al., 2019; Morales-Briones et al., 2018; Pessoa et al., 2021; Villaverde et al., 2023), we found a significant phylogenetic conflict between the nuclear and the plastid genomes. For example, none of the three clades recovered by the nuclear dataset (Figures 3 and 4) were fully supported by the plastid sequences (Figure 5). The most striking result is the position of *K. anteuphorbium*, which appears as a sister to *K. semperviva* (Figure 5), instead of forming a clade with Canarian *K. neriifolia* (Figures 3 and 4). This suggests the intriguing possibility of chloroplast capture events after the initial Rand Flora divergence. In the genus *Senecio*, which is closely related to *Kleinia*, Pelser et al. (2012) identified two separate hybridization events of the Canarian and Moroccan species *Senecio massaicus*, with two geographically distant lineages from the Mediterranean region and from southern Africa; these events took place more than 6 Ma, supporting ancient hybridization as a source of discordance among genomic signals (Pelser et al., 2012). The Sahara Desert was not an insurmountable barrier: it experienced several episodes of 'greening' during the Pliocene and the humid Pleistocene epochs, when migration and secondary contact became possible (e.g. Hopcroft et al., 2017).

4.2 | Syngameons, Central Island Hubs and geographic extirpation in a widespread Canarian species

Target sequencing with genome skimming, Hyb-Seq (Weitemier et al., 2014), has been revealed as a useful technique to solve phylogenetic relationships at both the interspecies (Mandel et al., 2014; Villaverde et al., 2023) and the intra-species level (Slimp et al., 2021; Villaverde et al., 2018) using a single source of genomic DNA. Recently, it has been employed to generate single-nucleotide polymorphism (SNPs) and document population patterns of genetic variation (Helmstetter et al., 2020; Karimi et al., 2020; Peláez et al., 2020). This versatility makes Hyb-Seq a useful technique to study patterns of island colonization at different evolutionary scales, from inferring the divergence of a MIE species from its closest mainland relative to reconstructing its phylogeographic history within the archipelago. Here, the inclusion for the first time of more than one individual of *K. anteuphorbium* allowed us to establish the reciprocal genetic isolation of *K. neriifolia* and *K. anteuphorbium* since their time of divergence, which is dated in the Early Pliocene (Figure 6). The stem age of *K. neriifolia* is thus older than that of other Canarian endemics (e.g. Albaladejo et al., 2021), but agrees with estimates for other Rand Flora species (Culshaw, Villaverde, et al., 2021; Rincón-Barrado et al., 2021; Villaverde et al., 2018).

Our dense intraspecific sampling of *K. neriifolia* also facilitated the analysis of population-level relationships and inter-island allelic admixture using genome-wide data. This revealed a strong

geographic structure in patterns of genetic variance: an east/west split among island populations that is observed in other Rand Flora MIEs: *C. canariensis* (Mairal, Pokorny, et al., 2015) or *E. balsamifera* (Rincón-Barrado et al., 2024). However, unlike *E. balsamifera* or *C. canariensis*, we did not find a clear geographic division within Tenerife between populations located in the western half of the island and those present in the central/eastern half. Instead, Tenerife populations were clustered into a single unit, separated from two island clades formed by populations to the east of Tenerife (Gran Canaria, Lanzarote and Fuerteventura) and to the west (La Gomera, La Palma and El Hierro). Alignment-based and SNP-based phylogenies (Figures 3, 4 and 7) support a closer relationship of Tenerife populations with the eastern islands clade. However, in the admixture and genetic-barrier analyses (Figure 8), Tenerife is recovered as an independent genetic unit placed between the eastern and western clades EC and WC.

This prominent position of Tenerife as a 'hinge' or bridge between the two opposite ends of the archipelago agrees well with the CIH hypothesis, postulating an initial colonization of the central islands (Tenerife, Gran Canaria) by migrant individuals, followed by eastward and westward dispersal (Sanmartín et al., 2008). This is indeed the phylogeographic model selected as the one with the highest explanatory power by our CNN (CIHa; Table 2). It also agrees with the population genetic diversity indexes (Table 1), which show the highest levels of allele richness and heterozygosity on the island of Tenerife. Previous population-level genetic studies on *K. neriifolia* based on plastid DNA or microsatellite markers (García-Verdugo, Caujapé-Castells, Mairal, & Monroy, 2019; Sun & Vargas-Mendoza, 2017) also found higher levels of genetic diversity on this island. *Kleinia neriifolia* is dispersed by wind by means of a wind-borne pappus. It has been observed that this kind of dispersal is not a particularly effective form of inter-island migration in the Canary Islands (Arjona et al., 2018), and even the species with the highest wind-dispersal capacity present a marked inter-island population structure (García-Verdugo, Caujapé-Castells, Mairal, & Monroy, 2019). At the time when *K. neriifolia* is estimated to have colonized Tenerife, the island was fully formed, having reached the largest area and highest altitude in the archipelago (Carracedo & Perez-Torrado, 2013). This would have facilitated the initial colonization event and later dispersal to each side of the archipelago, even for species with a low wind-dispersal capacity. For example, the trade winds blowing from the north and the northwest Atlantic would have favoured the arrival of propagules from northwest Africa to the central islands (Rognon & Coudé-Gaussen, 1996). On the contrary, the fact that the CIHa model was selected over the CIHb model in our CNN analysis suggests that the dispersal capacities of *Kleinia* have been enough to overcome intra-island dispersal barriers: the typical split between western and eastern/central Tenerife is not observed in *K. neriifolia*. In fact, this species forms an almost uninterrupted open-lowland belt around each of the Canary Islands, implying that populations on islands are well connected. Lack of phylogeographic structure in Tenerife was also found in Sanger-sequencing studies of this species (García-Verdugo, Caujapé-Castells, Illera, et al., 2019; Sun & Vargas-Mendoza, 2017).

Even if the CIH is the best-fit model to explain island colonization in *K. neriifolia*, evidence of genetic admixture within the eastern island clade (EC; Figure 8) supports the Surfing Syngameon Hypothesis of multiple events colonizations and secondary contact enhancing genetic diversity in these islands (Caujapé-Castells et al., 2017). The SSH model predicts that the easternmost islands of Lanzarote and Fuerteventura were recolonized from the west by populations from Tenerife and Gran Canaria, often after the extinction or geographic extirpation of the ancestral stock. During the Pleistocene, Lanzarote and Fuerteventura suffered several geological upheavals, such as eruptions (Carracedo et al., 2002), landslides (Carracedo et al., 2009), tsunamis (León et al., 2017) and rising sea levels during the Pleistocene interglacial periods (Caujapé-Castells, 2011). However, the climatic fluctuations introduced by the glacial–interglacial cycles in the Pleistocene had the greatest effect on these islands (Fernández-Palacios et al., 2016; Norder et al., 2019) and were probably the major cause of the local extinction of species and populations (García-Verdugo, Caujapé-Castells, Illera, et al., 2019).

It should be noted that the areas of Lanzarote and Fuerteventura that provided refugia during the Pleistocene, Famara and Jandía, and that today harbour the highest species diversity (Reyes-Betancort et al., 2008) were sampled in our study. Therefore, the possibility of non-sampled populations is remote. The fact that Tenerife sequences form a monophyletic clade in the phylogenetic analyses (TC; Figures 3, 4 and 7) and exhibit their own genetic cluster in the population genetic analysis (Figure 8), with no evidence of admixture with Lanzarote and Fuerteventura (or with Gran Canaria), does not support a SSH-like recolonization event from the west. It is more likely that the initial colonization followed the CIH model, from Tenerife to Gran Canaria, and that Lanzarote and Fuerteventura were colonized from this latter island, as suggested by the CNN model.

On the contrary, in the case of Gran Canaria, sequences from this island do not form a monophyletic group but are paraphyletic relative to the samples from Lanzarote and Fuerteventura in the SVDquartets analysis (Figure 7), and there is evidence of admixture with the eastern islands in the sNMF analysis (Figure 8). This suggests subsequent secondary contact and frequent migration events, from Gran Canaria to the eastern islands, and also between Lanzarote and Fuerteventura, resulting in the current patterns of genetic admixture. Again, this agrees with García-Verdugo, Caujapé-Castells, Mairal, and Monroy (2019), who inferred at least two colonization events of Lanzarote and Fuerteventura from Gran Canaria in very recent times based on plastid data.

Our study represents the first attempt of reconstructing phylogenomics relationships within the genus *Kleinia* based on a large dataset of nuclear genes, and a nearly complete representation of plastid coding genes. Though our species sampling is incomplete, our study provides a backbone and valuable dataset for future analyses encompassing a more comprehensive taxon sampling. We also demonstrate that machine learning based on convolutional neural networks can be used to discriminate among complex scenarios of island migration and colonization with high accuracy.

AUTHOR CONTRIBUTIONS

M.R.B., R.R. and I.S. designed the research, performed research, analysed the data and wrote the paper (original and revision). M.P. analysed the data and edited the paper. T.V. performed research and edited the paper. C.G.-V. contributed new reagents or analytical tools and edited the paper. J.C.-C. contributed new reagents or analytical tools and edited the paper.

ACKNOWLEDGEMENTS

Funding was provided by the Spanish Government, through grants CGL2015-67849-P (MINECO/FEDER), and project PID2019-108109GB-I00 funded by MCIN/AEI/10.13039/501100011033/ and FEDER 'A way to make Europe', to IS and RR. MR was supported by a MINECO FPI Fellowship (BES-2013-065389), jointly supervised by IS and RR. We are very grateful to people that provided tissue samples: J. Foster, J. Lodé, P. Pavelka, L. Pokorny, B. Rodríguez and A. Rzepecky. Thanks also to Herbarium MA for allowing DNA extraction from two specimens belonging to the outgroup taxa. Finally, we would like to thank M. García-Gallo for performing the DNA extractions.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Genetic data: Raw sequence reads are deposited in the SRA (BioProject PRJNA1167908). Sample metadata: Metadata are also stored in the SRA (BioProject PRJNA1167908) and in Table S1. Scripts: All scripts used to perform our CNN approaches are available at: https://github.com/manolofperez/CNN_Kleinia.

BENEFIT-SHARING STATEMENT

Benefits from this research accrue from the sharing of our data and results on public databases as described above.

ORCID

Mario Rincón Barrado https://orcid.org/0000-0001-5571-1473
Manolo Perez https://orcid.org/0000-0002-4642-7793
Tamara Villaverde https://orcid.org/0000-0002-9236-8616
Carlos García-Verdugo https://orcid.org/0000-0003-0332-5583
Juli Caujapé-Castells https://orcid.org/0000-0003-0600-1496
Ricarda Riina https://orcid.org/0000-0002-7423-899X
Isabel Sanmartín https://orcid.org/0000-0001-6104-9658

REFERENCES

Albaladejo, R. G., Martín-Hernanz, S., Reyes-Betancort, J. A., Santos-Guerra, A., Olangua-Corral, M., & Aparicio, A. (2021). Reconstruction of the spatio-temporal diversification and ecological niche evolution of *Helianthemum* (Cistaceae) in the Canary Islands using genotyping-by-sequencing data. *Annals of Botany*, *127*(5), 597–611.

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*(9), 1655–1664.

Allan, G. J., Francisco-Ortega, J., Santos-Guerra, A., Boerner, E., & Zimmer, E. A. (2004). Molecular phylogenetic evidence for the geographic origin and classification of Canary Island Lotus (Fabaceae: Loteae). *Molecular Phylogenetics and Evolution*, *32*(1), 123–138.

Andermann, T., Cano, Á., Zizka, A., Bacon, C., & Antonelli, A. (2018). SECAPR—A bioinformatics pipeline for the rapid and user-friendly processing of targeted enriched Illumina sequences, from raw reads to alignments. *PeerJ*, *6*, e5175.

Andrus, N., Trusty, J., Santos-Guerra, A., Jansen, R. K., & Francisco-Ortega, J. (2004). Using molecular phylogenies to test phytogeographical links between East/South Africa-Southern Arabia and the Macaronesian Islands – A review, and the case of *Vierea* and *Pulicaria* section *Vieraeopsis* (Asteraceae). *Taxon*, *53*(2), 333–346.

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*(5), 805–814.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., & Pevzner, P. A. (2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*, *19*(5), 455–477.

Barbujani, G., Stenico, M., Excoffier, L., & Nigro, L. (1996). Mitochondrial DNA sequence variation across linguistic and geographic boundaries in Italy. *Human Biology*, *68*, 201–215.

Barres, L., Galbany-Casals, M., Hipp, A. L., Molero, J., & Vilatersana, R. (2017). Phylogeography and character evolution of *Euphorbia* sect. *Aphyllis* subsect. *Macaronesicae* (Euphorbiaceae). *Taxon*, *66*(2), 324–342.

Beaumont, M. A. (2010). Approximate Bayesian computation in evolution and ecology. *Annual Review of Ecology, Evolution, and Systematics*, *41*(1), 379–406.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120.

Bradshaw, C. D., Lunt, D. J., Flecker, R., Salzmann, U., Pound, M. J., Haywood, A. M., & Eronen, J. T. (2012). The relative roles of CO_2 and palaeogeography in determining late Miocene climate: Results from a terrestrial model–data comparison. *Climate of the Past*, *8*(4), 1257–1285.

Brassel, K. E., & Reif, D. (1979). A procedure to generate Thiessen polygons. *Geographical Analysis*, *11*(3), 289–303.

Cardoso, P., Arnedo, M. A., Triantis, K. A., & Borges, P. A. (2010). Drivers of diversity in Macaronesian spiders and the role of species extinctions. *Journal of Biogeography*, *37*(6), 1034–1046.

Carine, M. A., & Schaefer, H. (2010). The Azores diversity enigma: Why are there so few Azorean endemic flowering plants and why are they so widespread? *Journal of Biogeography*, *37*(1), 77–89.

Carracedo, J. C., Ancochea Soto, E., Pérez Torrado, F. J., Meco, J., Hernán, F., Cubas, C. R., Casillas, R., Rodríguez Badiola, E., & Ahijado, A. (2002). Cenozoic volcanisme II: The Canary Islands. Geological Society of London, London.

Carracedo, J. C., Pérez Torrado, F. J., Paris, R., & Rodríguez Badiola, E. (2009). Megadeslizazmientos en las Islas Canarias, Enseñanza de las Ciencias de la Tierra, vol. 17.

Carracedo, J. C., & Perez-Torrado, F. J. (2013). Geological and geodynamic context of the Teide volcanic complex. In J. C. Carracedo & V. R. Troll (Eds.), *Teide Volcano* (pp. 23–36). Springer Berlin Heidelberg.

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 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.

Caujapé-Castells, J., García-Verdugo, C., Sanmartín, I., Fuertes-Aguilar, J., Romeiras, M. M., Zurita-Pérez, N., & Nebot, R. (2022). The late Pleistocene endemicity increase hypothesis and the origins of diversity in the Canary Islands Flora. *Journal of Biogeography*, *49*(8), 1469–1480.

Chifman, J., & Kubatko, L. (2014). Quartet inference from SNP data under the coalescent model. *Bioinformatics*, *30*(23), 3317–3324.

Cicuzza, D., Sombra Stäheli, D., Nyffeler, R., & Eggli, U. (2018). Morphology and anatomy support a reclassification of the African succulent taxa of *Senecio* s.l. (Asteraceae: Senecioneae). *Haseltonia*, *2017*(23), 11–26.

Compositae Working Group (CWG). (2024). Global compositae database. http://www.compositae.org/gcd

Culshaw, V., Mairal, M., & Sanmartín, I. (2021). Biogeography meets niche modeling: Inferring the role of deep time climate change when data is limited. *Frontiers in Ecology and Evolution*, *9*, 662092.

Culshaw, V., Villaverde, T., Mairal, M., Olsson, S., & Sanmartín, I. (2021). Rare and widespread: Integrating Bayesian MCMC approaches, sanger sequencing and Hyb-Seq phylogenomics to reconstruct the origin of the enigmatic Rand Flora genus *Camptoloma*. *American Journal of Botany*, *108*(9), 1673–1691.

Dellaporta, S., Wood, J., & Hicks, J. (1983). A rapid method for DNA extraction from plant tissue. *Plant Molecular Biology Reporter*, *1*, 19–21.

Drummond, A. J., Suchard, M. A., Xie, D., & Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, *29*(8), 1969–1973.

Faircloth, B. C. (2016). PHYLUCE is a software package for the analysis of conserved genomic loci. *Bioinformatics*, *32*(5), 786–788.

Fernández-Palacios, J. M., De Nascimento, L., Otto, R., Delgado, J. D., García-del-Rey, E., Arévalo, J. R., & Whittaker, R. J. (2011). A reconstruction of Palaeo-Macaronesia, with particular reference to the long-term biogeography of the Atlantic island laurel forests. *Journal of Biogeography*, *38*(2), 226–246.

Fernández-Palacios, J. M., Otto, R., Capelo, J., Caujapé-Castells, J., de Nascimento, L., Duarte, M. C., Elias, R. B., García-Verdugo, C., Menezes de Sequeira, M., Médail, F., Naranjo-Cigala, A., Patiño, J., Price, J., Romeiras, M. M., Sánchez-Pinto, L., & Whittaker, R. J. (2024). In defence of the entity of Macaronesia as a biogeographical region. *Biological Reviews*. https://doi.org/10.1111/brv.13112

Fernández-Palacios, J. M., Rijsdijk, K. F., Norder, S. J., Otto, R., de Nascimento, L., Fernández-Lugo, S., Tjørve, E., & Whittaker, R. J. (2016). Towards a glacial-sensitive model of island biogeography. *Global Ecology and Biogeography*, *25*(7), 817–830.

Flagel, L., Brandvain, Y., & Schrider, D. R. (2019). The unreasonable effectiveness of convolutional neural networks in population genetic inference. *Molecular Biology and Evolution*, *36*(2), 220–238.

Florencio, M., Patiño, J., Nogué, S., Traveset, A., Borges, P. A. V., Schaefer, H., Amorim, I. R., Arnedo, M., Ávila, S. P., Cardoso, P., de Nascimento, L., Fernández-Palacios, J. M., Gabriel, S. I., Gil, A., Gonçalves, V., Haroun, R., Illera, J. C., López-Darias, M., Martínez, A., … Santos, A. M. C. (2021). Macaronesia as a fruitful arena for ecology, evolution, and conservation biology. *Frontiers in Ecology and Evolution*, *9*, 718169.

Frichot, E., & François, O. (2015). LEA: An R package for landscape and ecological association studies. *Methods in Ecology and Evolution*, *6*(8), 925–929.

Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G., & François, O. (2014). Fast and efficient estimation of individual ancestry coefficients. *Genetics*, *196*(4), 973–983.

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*(5), 845–859.

García-Verdugo, C., Caujapé-Castells, J., Mairal, M., & Monroy, P. (2019). How repeatable is microevolution on islands? Patterns of dispersal and colonization-related plant traits in a phylogeographical context. *Annals of Botany*, *123*(3), 557–568.

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.

González-Pérez, M. A., Caujapé-Castells, J., & Sosa, P. A. (2004). Allozyme variation and structure of the Canarian endemic palm tree *Phoenix canariensis* (Arecaceae): Implications for conservation. *Heredity*, *93*(3), 307–315.

Gruber, B., Unmack, P. J., Berry, O. F., & Georges, A. (2018). Dartr: An r package to facilitate analysis of SNP data generated from reduced representation genome sequencing. *Molecular Ecology Resources*, *18*, 691–699.

Halliday, P. (1986). The genus *Kleinia* (Compositae) in North Africa and the Canary Islands. *Kew Bulletin*, *41*(2), 279–285.

Hansen, J., Sato, M., Russell, G., & Kharecha, P. (2013). Climate sensitivity, sea level and atmospheric carbon dioxide. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *371*(2001), 20120294.

Helmstetter, A. J., Béthune, K., Kamdem, N. G., Sonké, B., & Couvreur, T. L. P. (2020). Individualistic evolutionary responses of Central African rain forest plants to Pleistocene climatic fluctuations. *Proceedings of the National Academy of Sciences*, *117*(51), 32509–32518.

Herrando-Moraira, S., Calleja, J. A., Galbany-Casals, M., Garcia-Jacas, N., Liu, J.-Q., López-Alvarado, J., López-Pujol, J., Mandel, J. R., Massó, S., Montes-Moreno, N., Roquet, C., Sáez, L., Sennikov, A., Susanna, A., & Vilatersana, R. (2019). Nuclear and plastid DNA phylogeny of tribe Cardueae (Compositae) with Hyb-Seq data: A new subtribal classification and a temporal diversification framework. *Molecular Phylogenetics and Evolution*, *137*, 313–332.

Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: Improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution*, *35*(2), 518–522.

Hooft van Huysduynen, A., Janssens, S., Merckx, V., Vos, R., Valente, L., Zizka, A., Larter, M., Karabayir, B., Maaskant, D., Witmer, Y., Fernández-Palacios, J. M., de Nascimento, L., Jaén-Molina, R., Caujapé Castells, J., Marrero-Rodríguez, Á., del Arco, M., & Lens, F. (2021). Temporal and palaeoclimatic context of the evolution of insular woodiness in the Canary Islands. *Ecology and Evolution*, *11*(17), 12220–12231.

Hopcroft, P. O., Valdes, P. J., Harper, A. B., & Beerling, D. J. (2017). Multi vegetation model evaluation of the green Sahara climate regime. *Geophysical Research Letters*, *44*, 6804–6813.

Hudson, R. R. (2002). Generating samples under a wright-fisher neutral model of genetic variation. *Bioinformatics*, *18*(2), 337–338.

Jiménez-Moreno, G., Pérez-Asensio, J. N., Larrasoaña, J. C., Sierro, F. J., Garcia-Castellanos, D., Salazar, Á., Salvany, J. M., Ledesma, S., Mata, M. P., & Mediavilla, C. (2019). Early Pliocene climatic optimum, cooling and early glaciation deduced by terrestrial and marine environmental changes in SW Spain. *Global and Planetary Change*, *180*, 89–99.

Johnson, M. G., Gardner, E. M., Liu, Y., Medina, R., Goffinet, B., Shaw, A. J., Zerega, N. J., & Wickett, N. J. (2016). HybPiper: Extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. *Applications in Plant Sciences*, *4*(7), 1600016.

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K., Von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods*, *14*(6), 587–589.

Karimi, N., Grover, C. E., Gallagher, J. P., Wendel, J. F., Ané, C., & Baum, D. A. (2020). Reticulate evolution helps explain apparent homoplasy in floral biology and pollination in baobabs (*Adansonia*; Bombacoideae; Malvaceae). *Systematic Biology*, *69*(3), 462–478.

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, *30*(4), 772–780.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., & Drummond, A. (2012). Geneious basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, *28*(12), 1647–1649.

Kim, H.-G., Keeley, S. C., Vroom, P. S., & Jansen, R. K. (1998). Molecular evidence for an African origin of the Hawaiian endemic *Hesperomannia* (*Asteraceae*). *Proceedings of the National Academy of Sciences of the United States of America*, *95*(26), 15440–15445.

Kirschner, P., Perez, M. F., Záveská, E., Sanmartín, I., Marquer, L., Schlick-Steiner, B. C., Alvarez, N., STEPPE Consortium, Arthofer, W., Frajman, B., Gamisch, A., Hilpold, A., Paun, O., Trucchi, E., Záveská, E., Steiner, F. M., & Schönswetter, P. (2022). Congruent evolutionary responses of European steppe biota to late quaternary climate change. *Nature Communications*, *13*(1), 1921.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278–2324.

León, R., Somoza, L., Urgeles, R., Medialdea, T., Ferrer, M., Biain, A., García-Crespo, J., Mediato, J. F., Galindo, I., Yepes, J., González, F. J., & Gimenez-Moreno, J. (2017). Multi-event oceanic Island landslides: New onshore-offshore insights from El Hierro Island, Canary Archipelago. *Marine Geology*, *393*, 156–175.

Long, C., & Kubatko, L. (2018). The effect of gene flow on coalescent-based species-tree inference. *Systematic Biology*, *67*(5), 770–785.

Long, C., & Kubatko, L. (2019). Identifiability and reconstructibility of species phylogenies under a modified coalescent. *Bulletin of Mathematical Biology*, *81*(2), 408–430.

MacArthur, R. H., & Wilson, E. O. (1963). An equilibrium theory of insular zoogeography. *Evolution*, *17*(4), 373.

Mairal, M., Pokorny, L., Aldasoro, J. J., Alarcón, M., & Sanmartín, I. (2015). Ancient vicariance and climate-driven extinction explain continental-wide disjunctions in Africa: The case of the Rand Flora genus *Canarina* (Campanulaceae). *Molecular Ecology*, *24*(6), 1335–1354.

Mairal, M., Sanmartín, I., Aldasoro, J. J., Culshaw, V., Manolopoulou, I., & Alarcón, M. (2015). Palaeo-islands as refugia and sources of genetic diversity within volcanic archipelagos: The case of the widespread endemic *Canarina canariensis* (Campanulaceae). *Molecular Ecology*, *24*(15), 3944–3963.

Mairal, M., Sanmartín, I., Herrero, A., Pokorny, L., Vargas, P., Aldasoro, J. J., & Alarcón, M. (2017). Geographic barriers and Pleistocene climate change shaped patterns of genetic variation in the Eastern Afromontane biodiversity hotspot. *Scientific Reports*, *7*(1), 1–13.

Mandel, J. R., Dikow, R. B., Funk, V. A., Masalia, R. R., Staton, S. E., Kozik, A., Michelmore, R. W., Rieseberg, L. H., & Burke, J. M. (2014). A target enrichment method for gathering phylogenetic information from hundreds of loci: An example from the Compositae. *Applications in Plant Sciences*, *2*(2), 1300085.

Manni, F., & Barrai, I. (2000). Patterns of genetic and linguistic variation in Italy: A case study. In C. Renfrew & K. Boyle (Eds.), *Archaeogenetics: DNA and the population prehistory of Europe* (pp. 333–338). McDonald Institute for Archeological Research Monographs.

Milne, R. I., & Abbott, R. J. (2002). The origin and evolution of tertiary relict floras. *Advances in Botanical Research*, *38*, 281–314.

Mirarab, S., & Warnow, T. (2015). ASTRAL-II: Coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. *Bioinformatics*, *31*(12), i44–i52.

Morales-Briones, D. F., Liston, A., & Tank, D. C. (2018). Phylogenomic analyses reveal a deep history of hybridization and polyploidy in the Neotropical genus *Lachemilla* (Rosaceae). *New Phytologist*, *218*(4), 1668–1684.

Nguyen, L.-T., Schmidt, H. A., Von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, *32*(1), 268–274.

Norder, S. J., Proios, K., Whittaker, R. J., Alonso, M. R., Borges, P. A. V., Borregaard, M. K., Cowie, R. H., Florens, F. B. V., de Frias Martins, A. M., Ibáñez, M., Kissling, W. D., deNascimento, L., Otto, R., Parent, C. E., Rigal, F., Warren, B. H., Fernández-Palacios, J. M., vanLoon, E. E., Triantis, K. A., & Rijsdijk, K. F. (2019). Beyond the last glacial maximum: Island endemism is best explained by long-lasting archipelago configurations. *Global Ecology and Biogeography*, *28*(2), 184–197.

Oliva-Tejera, F., Caujapé-Castells, J., Naranjo-Suárez, J., Navarro-Déniz, J., Acebes-Ginovés, J. R., & Bramwell, D. (2005). Population genetic differentiation in taxa of *Lotus* (Fabaceae: Loteae) endemic to the Gran Canarian pine forest. *Heredity*, *94*(2), 199–206.

Overcast, I., Achaz, G., Aguilée, R., Andújar, C., Arribas, P., Creedy, T. J., Economo, E. P., Etienne, R. S., Gillespie, R., Jacquet, C., Jay, F., Kennedy, S., Krehenwinkel, H., Lambert, A., Meramveliotakis, E., Noguerales, V., Perez-Lamarque, B., Roderick, G., Rogers, H., … Morlon, H. (2023). Towards a genetic theory of Island biogeography: Inferring processes from multidimensional community-scale data. *Global Ecology and Biogeography*, *32*(1), 4–23.

Overcast, I., Bagley, J. C., & Hickerson, M. J. (2017). Strategies for improving approximate Bayesian computation tests for synchronous diversification. *BMC Evolutionary Biology*, *17*, 203.

Ozerova, L. V., Schanzer, I. A., & Timonin, A. C. (2017). *Curio* alliance (Asteraceae: Senecioneae) revisited. *Wulfenia*, *24*, 29–52.

Patiño, J., Carine, M., Mardulyn, P., Devos, N., Mateo, R. G., González-Mancebo, J. M., Shaw, A. J., & Vanderpoorten, A. (2015). Approximate Bayesian computation reveals the crucial role of oceanic islands for the assembly of continental biodiversity. *Systematic Biology*, *64*(4), 579–589.

Peláez, P., Ortiz-Martínez, A., Figueroa-Corona, L., Montes, J. R., & Gernandt, D. S. (2020). Population structure, diversifying selection, and local adaptation in *Pinus patula*. *American Journal of Botany*, *107*(11), 1555–1566.

Pelser, P. B., Abbott, R. J., Comes, H. P., Milton, J. J., Möller, M., Looseley, M. E., Cron, G. V., Barcelona, J. F., Kennedy, A. H., Watson, L. E., Barone, R., Hernández, F., & Kadereit, J. W. (2012). The genetic ghost of an invasion past: Colonization and extinction revealed by historical hybridization in *Senecio*: Historical hybridization in *Senecio massaicus*. *Molecular Ecology*, *21*(2), 369–387.

Pelser, P. B., Kennedy, A. H., Tepe, E. J., Shidler, J. B., Nordenstam, B., Kadereit, J. W., & Watson, L. E. (2010). Patterns and causes of incongruence between plastid and nuclear Senecioneae (Asteraceae) phylogenies. *American Journal of Botany*, *97*(5), 856–873.

Pelser, P. B., Nordenstam, B., Kadereit, J. W., & Watson, L. E. (2007). An ITS phylogeny of tribe Senecioneae (Asteraceae) and a new delimitation of *Senecio* L. *Taxon*, *56*(4), 1077–1104.

Peng, J., Swofford, D. L., & Kubatko, L. (2022). Estimation of speciation times under the multispecies coalescent. *Bioinformatics*, *38*(23), 5182–5190.

Perez, M. F., Bonatelli, I. A., Romeiro-Brito, M., Franco, F. F., Taylor, N. P., Zappi, D. C., & Moraes, E. M. (2022). Coalescent-based species delimitation meets deep learning: Insights from a highly fragmented cactus system. *Molecular Ecology Resources*, *22*(3), 1016–1028.

Perez, M. F., Bonatelli, I. A. S., Moraes, E. M., & Carstens, B. C. (2016). Model-based analysis supports interglacial refugia over long-dispersal events in the diversification of two South American cactus species. *Heredity*, *116*(6), 550–557.

Pessoa, E. M., Nollet, F., Magalhães, R. F., Viruel, J., Pinheiro, F., & Chase, M. W. (2021). Nuclear–plastid discordance indicates past introgression in *Epidendrum* species (Laeliinae: Orchidaceae) with highly variable chromosome numbers. *Botanical Journal of the Linnean Society*, *199*, 357–371. boab080.

Pokorny, L., Riina, R., Mairal, M., Meseguer, A. S., Culshaw, V., Cendoya, J., Serrano, M., Carbajal, R., Ortiz, S., Heuertz, M., & Sanmartín, I. (2015). Living on the edge: Timing of Rand Flora disjunctions congruent with ongoing aridification in Africa. *Frontiers in Genetics*, *6*, 154.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945–959.

Reyes-Betancort, J. A., Santos Guerra, A., 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 Jardin Botánico de Madrid*, *65*(1), 25–45.

Riina, R., Villaverde, T., Rincón-Barrado, M., Molero, J., & Sanmartín, I. (2021). More than one sweet tabaiba: Disentangling the systematics of the succulent dendroid shrub *Euphorbia balsamifera*. *Journal of Systematics and Evolution*, *59*(3), 490–503.

Rincón-Barrado, M., Olsson, S., Villaverde, T., Moncalvillo, B., Pokorny, L., Forrest, A., Riina, R., & Sanmartín, I. (2021). Ecological and geological processes impacting speciation modes drive the formation of wide-range disjunctions within tribe Putorieae (Rubiaceae). *Journal of Systematics and Evolution*, *59*(5), 915–934.

Rincón-Barrado, M., Villaverde, T., Perez, M. F., Sanmartín, I., & Riina, R. (2024). The sweet tabaiba or there and back again: phylogeographical history of the Macaronesian *Euphorbia balsamifera*. *Annals of Botany*, *133*(5–6), 883–904.

Rognon, P., & Coudé-Gaussen, G. (1996). Paleoclimates off northwest Africa (28°–35° N) about 18,000 yr B.P. based on continental eolian deposits. *Quaternary Research*, *46*(2), 118–126.

Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., & Sánchez-Gracia, A. (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Molecular Biology and Evolution*, *34*(12), 3299–3302.

Sanmartín, I., Anderson, C. L., Alarcon, M., Ronquist, F., & Aldasoro, J. J. (2010). Bayesian Island biogeography in a continental setting: The Rand Flora case. *Biology Letters*, *6*(5), 703–707.

Sanmartín, I., van der Mark, P., & Ronquist, F. (2008). Inferring dispersal: A Bayesian approach to phylogeny-based Island biogeography, with special reference to the Canary Islands. *Journal of Biogeography*, *35*(3), 428–449.

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), 815.

Senut, B., Pickford, M., & Ségalen, L. (2009). Neogene desertification of Africa. *Comptes Rendus Geoscience*, *341*(8–9), 591–602.

Sepulchre, P., Ramstein, G., Fluteau, F., Schuster, M., Tiercelin, J.-J., & Brunet, M. (2006). Tectonic uplift and Eastern Africa aridification. *Science*, *313*(5792), 1419–1423.

Shaw, K. L., & Gillespie, R. G. (2016). Comparative phylogeography of oceanic archipelagos: Hotspots for inferences of evolutionary process. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(29), 7986–7993.

Slater, G., & Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics*, *6*(1), 31.

Slimp, M., Williams, L. D., Hale, H., & Johnson (2021). On the potential of Angiosperms353 for population genomic studies.. *Applications in Plant Sciences*, *9*(7), 11419.

Stadler, T. (2009). On incomplete sampling under birth–death models and connections to the sampling-based coalescent. *Journal of Theoretical Biology*, *261*(1), 58–66.

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312–1313.

Suchard, M. A., Lemey, P., Baele, G., Ayres, D. L., Drummond, A. J., & Rambaut, A. (2018) Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evolution*, *4*(1), vey016.

Sun, Y., Liu, Y.-F., & Huang, H.-W. (2009). Isolation and characterization of polymorphic microsatellite markers in *Schisandra chinensis* (Turcz.) Baill. (Schisandraceae). *Conservation Genetics Resources*, *1*(1), 119–121.

Sun, Y., & Vargas-Mendoza, C. F. (2017). Population structure, genetic diversity, and evolutionary history of *Kleinia neriifolia* (Asteraceae) on the Canary Islands. *Frontiers in Plant Science*, *8*, 1180.

Swofford, D. (2018). *PAUP* version 4.0 a163*. Sinauer.

Thiv, M., Thulin, M., Hjertson, M., Kropf, M., & Linder, H. P. (2010). Evidence for a vicariant origin of Macaronesian–Eritreo/Arabian disjunctions in *Campylanthus* Roth (Plantaginaceae). *Molecular Phylogenetics and Evolution*, *54*(2), 607–616.

Timonin, A. C., Ozerova, L. V., & Schanzer, I. A. (2015). Evolution of the succulent organization of the Southern African Senecioneae (Asteraceae). *Biology Bulletin Reviews*, *5*(1), 17–27.

Valente, L., Illera, J. C., Havenstein, K., Pallien, T., Etienne, R. S., & Tiedemann, R. (2017). Equilibrium bird species diversity in Atlantic islands. *Current Biology*, *27*(11), 1660–1666.

Villaverde, T., Larridon, I., Shah, T., Fowler, R. M., Chau, J. H., Olmstead, R. G., & Sanmartín, I. (2023). Phylogenomics sheds new light on the drivers behind a long-lasting systematic riddle: The figwort family Scrophulariaceae. *New Phytologist*, *240*(4), 1601–1615.

Villaverde, T., Pokorny, L., Olsson, S., Rincón-Barrado, M., Johnson, M. G., Gardner, E. M., Wickett, N. J., Molero, J., Riina, R., & Sanmartín, I. (2018). Bridging the micro- and macroevolutionary levels in phylogenomics: Hyb-Seq solves relationships from populations to species and above. *New Phytologist*, *220*(2), 636–650.

Wagner, W. L., & Funk, V. A. (1995). *Hawaiian biogeography*. Smithsonian Institution Press.

Warren, B. H., Simberloff, D., Ricklefs, R. E., Aguilée, R., Condamine, F. L., Gravel, D., Morlon, H., Mouquet, N., Rosindell, J., Casquet, J., Conti, E., Cornuault, J., Fernández-Palacios, J. M., Hengl, T., Norder, S. J., Rijsdijk, K. F., Sanmartín, I., Strasberg, D., Triantis, K. A., … Thébaud, C. (2015). Islands as model systems in ecology and evolution: Prospects fifty years after MacArthur-Wilson. *Ecology Letters*, *18*(2), 200–217.

Weitemier, K., Straub, S. C., Cronn, R. C., Fishbein, M., Schmickl, R., McDonnell, A., & Liston, A. (2014). Hyb-Seq: Combining target enrichment and genome skimming for plant phylogenomics. *Applications in Plant Sciences*, *2*(9), 1400042.

Whittaker, R. J., Fernández-Palacios, J. M., & Matthews, T. J. (2023). *Island biogeography: Geo-environmental dynamics, ecology, evolution, human impact, and conservation*. Oxford University Press.

Whittaker, R. J., Triantis, K. A., & Ladle, R. J. (2008). A general dynamic theory of oceanic island biogeography: A general dynamic theory of oceanic island biogeography. *Journal of Biogeography*, *35*(6), 977–994.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Rincón Barrado, M., Perez, M., Villaverde, T., García-Verdugo, C., Caujapé-Castells, J., Riina, R., & Sanmartín, I. (2024). Phylogenomics and phylogeographic model testing using convolutional neural networks reveal a history of recent admixture in the Canarian *Kleinia neriifolia*. *Molecular Ecology*, *33*, e17537. https://doi.org/10.1111/mec.17537