Botanical Journal of the Linnean Society, 2021, **XX**, 1–18. With 5 figures.

Genetic and morphological differentiation in *Washingtonia* (Arecaceae): solving a century-old palm mystery

LORENA VILLANUEVA-ALMANZA[1,*], JACOB B. LANDIS[1], DANIEL KOENIG[1,2,3] and EXEQUIEL EZCURRA[1]

[1]*Botany and Plant Sciences, University of California-Riverside, Riverside, CA 92521, USA*
[2]*Institute for Integrative Genome Biology, University of California-Riverside, Riverside, CA 92521, USA*
[3]*Center for Plant Cell Biology, University of California-Riverside, Riverside, CA 92521, USA*

Received 28 June 2020; revised 17 November 2020; accepted for publication 8 January 2021

Washingtonia is a genus of palms currently composed of two species, *W. filifera* and *W. robusta*, distributed in the States of Baja California, Baja California Sur, and Sonora in Mexico, and Southern California and Arizona in the United States. The group has been a taxonomic challenge due to a lack of type specimens, incomplete protologues, highly variable vegetative morphology, human dispersal of seeds, limited fieldwork in native populations and poor representation in herbaria. Here, we analyse the population structure and phylogenetic relationships and test whether morphological traits correlate with genetic variation throughout its distributional range. We used genotyping-by-sequencing (GBS) data to identify population structure and delimit species. We further used these data to determine whether morphological traits varied among genetic regions. We analysed 188 individuals from 21 populations of *Washingtonia* across its distribution range using multivariate and Bayesian methods. Our results showed great consistency in the discovery of four genetic groups: (1) Southern Peninsula, (2) Mid-Peninsula, (3) Northern Peninsula and (4) Sonoran mainland. The geographical limits to these clusters coincide very well with the large natural regions of the Sonoran Desert. Our analyses indicate that *Washingtonia* populations are highly structured within four major geographical regions. Even when no single morphological trait can be used to determine the genetic identity of *Washingtonia* palms, leaf greenness, a novel morphological trait, can be useful. Our results provide a robust phylogenetic analysis of *Washingtonia* settling a taxonomic debate that has lasted over a century.

ADDITIONAL KEYWORDS: Arecaceae – Baja California Peninsula – genotyping-by-sequencing – population structure – Sonoran Desert – taxonomy.

INTRODUCTION

Washingtonia H.Wendl. is a North American genus of widely cultivated diploid ($n = 18$) palms. It is naturally distributed in desert oases in Southern California and Arizona in the US, and in the States of Baja California, Sonora, and Baja California Sur, in Mexico (throughout this paper we will refer to the Baja California Peninsula in its strict geological definition, incorporating all terrestrial parts of North America that lie west of the San Andreas Fault, i.e. Southern California in the US and the States of Baja California and Baja California Sur in Mexico, frequently referred to by geologists as Peninsular California). High morphological variability initially led to the description of numerous species in the description of numerous species in the genus, many of which were later synonymized with one of the two currently recognized species: *W. filifera* (Linden ex André) H.Wendl. and *W. robusta* H.Wendl., both species were described on the basis of cultivated specimens. Closer inspection of the genus reveals an old, unsolved mystery: what is a species in *Washingtonia* and how is speciation occurring across its range? This widely distributed group of palms has been a taxonomic challenge since the species were described due to lack of type specimens, incomplete protologues, highly variable vegetative morphology, human dispersal,

---

*Corresponding author. E-mail: lvill017@ucr.edu

limited fieldwork in native populations and poor representation in herbaria. The genus is highly variable along its native range, presenting a taxonomic challenge (Henderson, Galeano-Garces & Bernal, 1997) that has only been partially clarified. The most recent study of the Mexican populations of *Washingtonia* (Klimova *et al*., 2018) using single nucleotide polymorphisms (SNPs) obtained through GBS (Elshire *et al*., 2011) found that the populations in the northern half of the State of Baja California to be genetically distinct from those in the center and south of the peninsula. However, this study did not sample the palm populations found in the U.S., the northernmost part of the distribution of the genus, and did not include morphological data. A morphometric study of *Washingtonia* across its whole distribution range done by our research team (Villanueva-Almanza *et al*., 2018) found the genus to vary gradually from South to North along Baja California Peninsula (23–34° latitude), suggesting that the genus is composed of one highly variable taxon and providing support to a previous idea that *W. robusta* could be a variant of *W. filifera* (Nabhan, 1986).

Because species delimitation in *Washingtonia* has been problematic, accurate geographical distribution data for each putative species is missing and, at times, contradictory. There seems to be some agreement that *W. filifera* occurs naturally in Southern California and Arizona in the US, and in the northern part of the State of Baja California in Mexico (Turner, Bowers & Burgess, 1995), whereas *W. robusta* is present in the southern part of the Baja California Peninsula and in Sonora, mainland Mexico, where it has a narrow distribution in the Sierra del Aguaje near Guaymas (28° latitude; Felger & Joyal, 1999). However, there seems to be no agreement in the published literature on the exact transition limit between *W. robusta* and *W. filifera* along the peninsular gradient.

The challenge in delimiting the geographical distribution of *W. robusta* is not new. André (1900) mentioned: 'the home of *W. robusta* is Baja California' apparently referring to the northern part of the State of Baja California in Mexico, but the exact place he meant is unclear. The problem of the geographical range of *W. robusta* also comes from a misinterpretation in modern literature. In *Los oasis de la península de Baja California* (Arriaga *et al*., 1997), the groundbreaking work for the study of desert oases in the Peninsula, the distribution range of *W. robusta* is reported as the State of Baja California Sur, south of parallel 28°. Although the authors reported that they followed the *Flora of Baja California* (Wiggins, 1980), that work does not mention *W. robusta* as only occurring in Baja California Sur. Rather, Wiggins described the range of *W. robusta* extending as far north as Cataviña, at latitude 30° in the State of Baja California, as others had previously reported (Bailey, 1936; Shreve & Wiggins, 1964; Moran, 1978) and other authors have reported more recently (Cornett, 1987; Minnich, Franco-Vizcaíno & Salazar-Ceseña, 2011). That *W. robusta* is restricted to Baja California Sur, following Arriaga *et al*. (1997), was accepted in some later works (Felger & Joyal, 1999; Zona, 2000; Felger, Johnson & Wilson, 2001; León de la Luz *et al*., 2014), whereas others have left the issue open to interpretation: the *Jepson Manual* and the *Palms of the Americas* ambiguously mention *W. robusta* as being native to Baja California (Henderson *et al*., 1997; Simono, 2012). Thus, there is an expanse of *c*. 230 km along the peninsula, from 28° to 30° latitude, where the taxonomic identity of the palms is uncertain (Fig. 1). Having a better understanding of the taxonomy and geographical distribution of *Washingtonia* will ultimately have consequences for its conservation. The IUCN Red List conservation status version 2.3 of *W. filifera* is lower risk/near threatened, whereas *W. robusta* has not been assessed (Johnson, 1996).

Regardless of where the northern limit of *W. robusta* is placed, the populations of *Washingtonia* are distributed more or less discontinuously in oases, 'islands' within a 'sea' of desert scrub along a 1300 km environmental gradient from latitude 23°N to 34°N. Southern populations experience dry subtropical climate with summer rains while northern populations are subject to arid temperate climate dominated by winter rains (Vogl & McHargue, 1966; Garcillán & Ezcurra, 2003). Despite the species not occurring in the same area, high connectivity among populations of *Washingtonia*, apparent overlapping flowering in May and June (Wiggins, 1980; Felger *et al*., 2001; Rebman & Roberts, 2012) and genetic compatibility between the species under cultivation (Ishihata & Murata, 1971; Hodel, 2014) would suggest there is no inherent barrier to gene flow. *Washingtonia* spp. are pollinated by insects (McClenaghan & Beauchamp, 1986), and the fleshy, edible fruits are dispersed mostly by coyotes (Vogl & McHargue, 1966; Felger *et al*., 2001) and foxes (Bullock, 1980). Human dispersal could explain the geographical distribution of *W. robusta* since it is a widely cultivated palm in the whole Baja California Peninsula and around the world (Johnson, 1996; Felger & Joyal, 1999).

Given the possible existence of ecotypes in *Washingtonia* (Villanueva-Almanza *et al*., 2018), the group represents an ideal system in which to analyse their population structure, providing insights into the phytogeography of the Sonoran Desert. Two hypotheses to explain the observed morphological variation in *Washingtonia* were previously put forth: (1) *Washingtonia* is composed of two clearly distinguishable species, one in the north and one in the south, that have differentiated in disjunction and have expanded into secondary contact resulting in a possible hybrid zone (secondary intergradation), which we called the 'hybrid

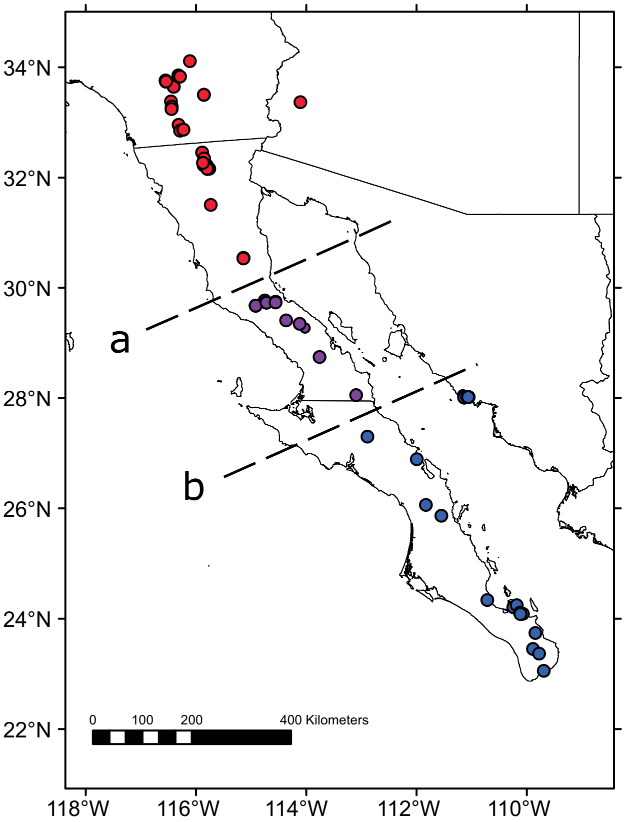


Figure 1. Major *Washingtonia* palm oases in the Baja California Peninsula, Arizona, and Sonora according to herbarium records and own collections. Dashed lines mark distribution limits according to: A, Bailey (1936); Shreve & Wiggins (1964); Moran (1978); Wiggins (1980); Cornett (1987); Minnich *et al*. (2011) and Klimova *et al*. (2018) and B, Arriaga *et al*. (1997); Felger & Joyal (1999); Zona (2000) and Felger *et al*. (2001). Red: *Washingtonia filifera*; blue: *Washingtonia robusta*. Purple circles show oases where authors have divergently identified *Washingtonia* individuals either as *W. filifera* or *W. robusta*.

zone model'; and (2) *Washingtonia* is composed of one highly variable species clustered in local and patchy populations in desert oases that change gradually along a clinal continuum (primary intergradation), which we referred to as the 'clinal model' (Villanueva-Almanza *et al*., 2018). Morphological results have shown that vegetative traits vary continuously along the latitudinal cline in the Baja California Peninsula giving support to the clinal model hypothesis.

In this study, we test the clinal model hypothesis using molecular data. In particular, the goals of this study are to: (1) determine whether morphological and genetic markers reveal a similar pattern of clinal variation connecting palm populations along the long peninsular gradient and (2) analyse whether genetic variation corresponds to the geographical distribution of the currently recognized species. The concept for delimiting provisional species that we apply in our work follows the genotypic cluster definition of Mallet (1995), which states that species are distinguishable groups of organisms that, when growing in sympatry, have a unimodal cluster of similar genotypes.

MATERIAL AND METHODS

SAMPLE COLLECTION

We collected 174 leaf samples of *Washingtonia* from 18 wild populations in Mexico and the USA between March 2017 and October 2017. We also included ten samples from the Bailey Hortorium Herbarium at Cornell University (BH) knowing that DNA quality from herbarium specimens is not as good as that from more recent collections. Using herbarium material was the only way to include historical populations that were not visited during fieldwork due to inaccessibility (Guadalupe and Tajo Canyons). We also included four samples of plants grown in the greenhouse from seeds collected in San Javier, Baja California Sur (Mexico) in 2015. Collection information and equivalent Sequence Read Archive accessions for specimens collected in Mexico and the USA and all other specimens included in the phylogenetic and morphometric analyses are provided in the Supporting Information. Together, these collections cover the entire range of *Washingtonia*. Fresh material from the field and cultivated specimens was stored in silica gel until DNA extraction. We did not assign taxonomic determination since we were interested in elucidating phylogenetic relationships among populations. Four samples of *Brahea armata* S.Watson cultivated at the University of California Riverside campus were used as the outgroup.

DNA EXTRACTION AND LIBRARY PREPARATION

Individual samples were randomized into two plates, each corresponding to the pools of final libraries sequenced. Total genomic DNA was extracted using a modified CTAB method (Doyle & Doyle, 1987) from the 192 samples. Total genomic DNA was checked for degradation using a 1.5% agarose electrophoresis gel. Samples were standardized following quantification using the Quant-It PicoGreen assay (Thermofisher, catalogue number P11496). Libraries were prepared following Rowan *et al*. (2017). Each DNA sample was digested using the restriction enzyme *KpnI* to which sequence adapters and sample-specific barcodes were ligated for sample identification. Ninety-six unique barcodes were used. All samples were sequenced at the University of California Davis using 1 × 150 Illumina HiSeq4000, one library per lane.

BIOINFORMATICS ANALYSIS

After sequencing, raw reads in FASTQ format were processed using Stacks software version 2.3 (Catchen *et al.*, 2013). Barcodes were removed, raw reads were quality filtered using the -q flag, adapters were removed, and trimmed to a uniform read length of 125 base pairs (bp) due to the need for uniform sequence length for *de novo* SNP calling using the program *process_radtags*. Demultiplexed data were assembled using *de novo* and reference-based methods using the date palm (*Phoenix dactylifera* L.) genome as a reference since no genome of *Washingtonia* is available, and date palm is the closest relative for which a genome has been sequenced. All analyses were performed using a computer cluster from the UCR High-Performance Computing Center. Untrimmed raw reads have been deposited in the Short Read Archive with individual accession numbers provided in Supporting Information, S1. Except where otherwise indicated, all statistical analyses were performed in R statistical software v.3.3.3 (R Development Core Team, 2018) using RStudio environment v.1.0.136 (RStudio Team, 2018).

For the *de novo* assembly, the *denovo_map.pl* wrapper was used to identify SNPs. Optimal parameters were identified by running the *denovo_map.pl* wrapper several times following Paris, Stevens & Catchen (2017) and Rochette & Catchen (2017). For all runs, the minimum number of raw reads required to create a stack in *ustacks* remained fixed ($m = 3$), but the number of nucleotides between alleles (M) varied from 2 to 3, and the number of mismatches allowed between loci for building the catalogue in *cstacks* (n) varied from 1 to 4. These various settings resulted in different numbers of loci and SNPs, but our final data set was obtained by setting $M = 2$ and $n = 2$. Our decision followed Paris *et al.* (2017) in using low values for M and n parameters if the genome shows low levels of polymorphism, as reported for *Washingtonia* by McClenaghan & Beauchamp (1986) and Klimova *et al.* (2018). Data was filtered using VCFtools v.0.1.15 (Danecek *et al.*, 2011) to require a minimum allele frequency of 0.05 and to remove loci with > 30% missing data, monomorphic loci, samples with at least 50% missing data and two samples with negative inbreeding coefficients and unusually high read depths suggestive of contaminated sequencing. Pairwise F_{ST} values were obtained re-reading the filtered *de novo* data set in Stacks using the *populations* program.

The SNP set was also pruned for linkage disequilibrium using Plink v.1.90b3.38 (Purcell *et al.*, 2007) with a 5000 bp sliding window and an r^2 value of 0.5. The final SNP data set was then reread into Stacks using the populations command to generate the necessary input files for downstream analyses.

Before calling SNPs in Stacks for the reference-guided approach, reads for each individual were mapped to the date palm reference using BWA-MEM v.0.7.12 (Li, 2013) with associated read groups. Created Sequence Alignment Map files were converted to BAM (Binary Alignment Map) and then sorted using Samtools v.1.3 (Li *et al.*, 2009). Sorted BAM files were then read into the *ref_map.pl* wrapper of Stacks using default parameters and exported with the populations command. Filtering was done in VCFtools using the same parameters as for the *de novo* SNP calling and LD pruned with Plink. The final VCF file was then reread to produce the appropriate files for downstream analyses.

POPULATION STRUCTURE

The *de novo* data set with 21 746 SNPs was used to calculate Nei's genetic distance among individuals through the R package *poppr* (Kamvar, Tabima & Grünwald, 2014). The genetic distance matrix was then used to construct a UPGMA (unweighted pair group method with arithmetic means) dendrogram.

Multivariate analysis

To summarize the genetic variation within *Washingtonia*, we performed a centred principal component analysis (PCA) of the individuals × SNPs matrix using the glPCA algorithm from the *adegenet* v.2.1.1 R package (Jombart & Ahmed, 2011). We used the broken-stick distribution to test for non-randomness of the axes (Jackson, 1993).

We used stepwise regression to estimate the best, non-redundant set of explanatory variables (latitude, region, sites, sites-nested-within-regions) to test their association with the first three PC axes. At each step, the significance of the variables included in the model was evaluated using F values and r^2. We assessed the effect of region and site within region using a nested ANOVA.

When viewed as a histogram, the PC scores of the first axis revealed a clear multi-modal pattern formed by a mixture of distributions. To find the number of distributions and the parameters that best described the first axis, we used a Hooke & Jeeves direct search method (Himmelblau, 1972) implemented in compilable QBX-Basic. We then fitted five different distribution models with different hypotheses: (1) sample frequency is normally distributed along the first PC axis (null model); (2) sample frequency is bimodal, formed by a mixture of two normal distributions; (3) sample frequency is tri-modal and (4) sample frequency is formed by a mixture of four normal distributions, and so forth. To compare distributions, we used the observed and expected frequencies to calculate the log-likelihood deviance and significance of the residuals of each model, as well as the Akaike information criterion.

Bayesian coalescent analysis

Population structure was implemented in the R package LEA v.2.4.0 (Frichot & François, 2015). Determination of the best *K* was determined using a cross-entropy criterion from 1–25 using both the *de novo* and reference-based data sets. Barplots from *K* = 1–12 were generated.

PHYLOGENETIC RELATIONSHIPS

Species tree inference was done using SNAPP (Bouckaert *et al*., 2014) implemented in BEAST v.2.4.5 (Bouckaert *et al*., 2014). Due to the computationally intensive nature of the program, three individuals from each population were included when available, except for ALA, which all samples were included due to the polyphyletic nature of the population (based on distance trees). The SNP data set was also further pruned using *vcftools* max-missing = 0.97. The VCF file was then converted to a nexus format using the *vcf2phylip* program (Ortiz, 2019). This file consisted of 57–63 individuals and 1094 bp for the *de novo* data set and 62–68 individuals and 3514 SNPs for the reference-based data set. Five different data sets were analysed to obtain a robust topology with and without the hybrid population, San Ignacio (ALA), based on distance trees clustering: (1) without San Ignacio; (2) San Ignacio as a monophyletic cluster; (3) one cluster of a single individual of San Ignacio (ALA13); (4) two ALA clusters (ALA13(ALA12 + ALA18)); and (5) three clusters of ALA with the two previously described clusters plus a third (ALA10, ALA15, ALA14). SNAPP was run using default parameters for 5 million generations to achieve large enough ESS values (> 200). Trees were visualized with *DensiTree* (part of the *BEAST* package). To test the number of species/subspecies, species delimitation was tested on the *de novo* data set using Bayes factor delimitation (BFD*) as implemented in SNAPP (Leaché *et al*., 2014). Five models were tested: a two-species model grouping northern and Mid-Peninsular populations together; a two-species model grouping Southern and Mid-Peninsular populations; a two-species model grouping Northern and Southern Peninsular populations; a three-species model and a four-species model. The three-species model grouped NACA with the Southern Peninsular populations, whereas the four-species model kept NACA as a distinct species. Each model consisted of 48 steps with 200 000 generations and a 20% burn-in following a 10 000 generation pre-burn-in. Model comparisons used Bayes factors and the reported marginal likelihood estimate (MLE).

HISTORICAL RELATIONSHIPS AMONG POPULATIONS

Using the filtered SNP set via the *de novo* approach and the reference-guided approach, historical migration events were inferred using *TreeMix* v.1.13 (Pickrell & Pritchard, 2012). For each data set, zero to ten migration events were tested, and the optimal value was determined using a likelihood ratio test as implemented in BioGeoBEARS v.0.2.1 (Matzke, 2013) in conjunction with the *D* statistic, a tool for detecting ancient admixture events (Durand *et al*., 2011). The model with the highest likelihood value, together with a positive admixture coefficient (*D* statistic), was selected as the most likely, or optimal, solution. Results for the optimal number of migration events was plotted in R using the *plotting_funcs.R* script packaged with the *TreeMix* software. For the reference-guided approach, the *Brahea* individuals were set as the outgroup. For the *de novo* analyses, NACA was set as the outgroup based on the positioning of this population in the distance dendrogram and phylogenetic tree (see previously).

ISOLATION BY DISTANCE

We tested for patterns of isolation by distance (IBD) using a simple linear model. The predictor matrix was the geographical distance between populations and the response matrix was made up of populations pairwise Nei's genetic distances. We performed a Mantel test with 10 000 permutations to correlate Nei's genetic distances matrix with the geographical distance matrix using the ade4 package v.1.7–13 (Jombart, 2008) in R. Populations were then assigned a proximity factor: 0 if the pair of populations belonged to the same region; 1 if they were in adjacent regions; 2 if they were in regions separated by a region and 3 if they were found in extreme regions. A generalized linear model was used to test for associations between genetic distance and geographical distance and proximity, as the independent predictors. This allowed us to see how genetic distance varied with genetic distance and what proportion of the residual variance was explained by the effect of class belonging.

ECOLOGICAL NICHE MODELLING AND ENVIRONMENTAL EFFECTS

To explore environmental niche differentiation (ENM) and understand how ecological diversification is driving speciation in *Washingtonia* we performed an ENM analysis. To highlight the genetic heterogeneity and population structure of *Washingtonia* and to avoid introducing any taxonomic bias, we used genetic regions instead of species names (Ikeda *et al*., 2017). We used occurrence data from 162 individuals belonging to the four genetic regions collected during fieldwork and kept after sequencing and filtering (54, 45, 10 and 52 from the central, northern, Sonoran and southern regions, respectively). We used the 19 bioclimatic variables from WorldClim v.2 with a resolution of 30 arc-seconds (*c.* 1 km) (Fick & Hijmans, 2017) to project current niches for each genetic region. Due to high correlation among bioclimatic variables, we discarded those with Pearson correlation coefficients > 0.80. Out of 19 variables, only seven were selected as evaluator variables. We used MaxEnt v.3.4.1 (Phillips, Anderson & Schapire, 2006) to determine niche variation among the four genetic regions of *Washingtonia* using a subsample procedure with a 25% random test percentage, 5000 maximum iterations and specifying the output to a logistic format. The potential error in species distribution estimations induced by spatially biased sampling was minimized by ensuring that individuals from all major oases within each genetic region were included in the analysis (Araújo & Guisan, 2006; Phillips *et al.*, 2009; Inman *et al.*, 2018).To quantify the relative importance of the bioclimatic variables to the suitability of each genetic region, we ran a jackknife test for the training and testing data sets.

To determine whether bioclimatic differed among genetic regions, we compared the effect of the seven uncorrelated bioclimatic variables used in the ENM on each of the genetic regions. Statistical analyses were performed with the R package v.3.5.1 (R Development Core Team, 2018) using RStudio environment v.1.1.463 (RStudio Team, 2018). When significant differences were found among genetic regions, we performed Tukey's honestly significant difference (HSD) post hoc tests to determine statistically similar groups.

ASSOCIATION BETWEEN GENETIC REGIONS AND PHENOTYPIC TRAITS

We had previously collected measurements of 13 morphological traits commonly used in taxonomic literature to distinguish between *Washingtonia* spp. (Bailey, 1936; Felger & Joyal, 1999; Simono, 2012; Villanueva-Almanza *et al.*, 2018). They represent vegetative characters only, since these are the ones used in the taxonomic keys of most treatments (Bailey, 1936; Wiggins, 1980; Simono, 2012). Specific leaf area (SLA) was added to test whether there were any physiological differences among populations. We assigned genetic region only to those individuals for which we had collected morphological data and SLA whenever both genetic and morphological data came from the same individual. Some populations and individuals were excluded from the analysis if this was not the case, e.g. when individuals had morphological data but lacked genetic information or vice versa. This was the case of Cataviña where individuals from which genetic data was extracted were different from those from which morphological traits were measured on a later fieldtrip. This was also the case for Berrendo and San Juan de la Costa, where leaves used for genetic analyses did not always have data on DBH.

To determine whether traits varied among genetic regions, morphological traits were compared using a one-way ANOVA. As above, when significant differences were found among genetic regions, we performed Tukey's HSD *post hoc* tests to determine statistically similar groups.

RESULTS

To investigate whether genetic variation in *Washingtonia* followed a clinal pattern that would refute the existence of two species, we performed GBS on 188 *Washingtonia* samples and four samples of *Brahea* resulting in 649 071 324 high-quality reads. From these data, 3685,07 variant sites were called using the *de novo* pipeline and 177 567 variant sites were called with the reference-based approach. After filtering, our final data set for the *de novo* approach included 173 *Washingtonia* individuals. We removed 17 individuals having at least 50% missing data (13 of *Washingtonia* and four of *Brahea*) and two *Washingtonia* samples with unusually high read depths (PALM19 and GER17; see Table 1). For the reference-based approach, four additional samples (GAS1, GAS2, JAV3 and COMO16) that were removed in the *de novo* approach were retained in the reference-guided approach, albeit with < 10 000 SNPs each. Fraction missing genotypes and average sequencing depth for each individual pre- and post-filtering is presented in the Supporting Information, Table S1.

POPULATION STRUCTURE

Our results from multivariate analysis and Bayesian coalescent methods showed great consistency in the discovery of four genetic groups: (1) Southern Peninsular: San José del Cabo, Sierra Cacachilas, San José de la Costa, San Javier, Comondú and Mulegé; (2) Mid-Peninsular: San Ignacio, Santa Gertrudis, San Francisco Borja, Bocana, Cataviña, Santa María Mission and Berrendo; (3) Northern Peninsular: Palomar, Guadalupe, Tajo, Gaskills, Anza, Palm Canyon, Joshua Tree and Kofa and (4) Sonoran mainland: Nacapule Canyon, as the sister group of the Southern Peninsular palms. Our results showed that the genetic variation in *Washingtonia* does not correspond to the geographical distribution of the two currently recognized species. The geographical limits to these clusters are shown in Figure 2 and coincide well with the large natural regions of the Sonoran Desert around the Gulf of California: (1) Southern Baja

Table 1. Characterization of GBS of *Washingtonia* and *Brahea* individuals from the Baja California Peninsula, Arizona (USA) and Sonora (Mexico)

	De novo approach	Reference-based approach
Reads	649 071 324	649 071 324
Variant sites	368 507	177 567
Post-filtering		
Individuals	173 *Washingtonia* spp.	177 *Washingtonia* spp., 3 *Brahea armata*
SNPs	21746	13459
Mean read depth	8.08 × –34.75 × (16.62×)	5.1 × –38.8 × (17×)

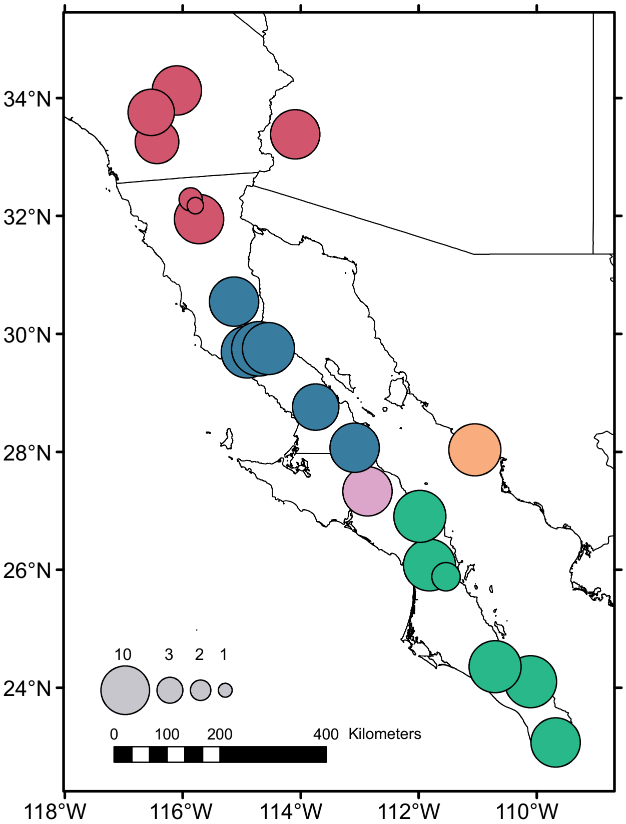


Figure 2. Map. Geographical distribution of sampled populations and their haplotypes colour coded by their genotype derived from Nei's genetic distance calculated using 21 746 SNPs via GBS using a *de novo* approach. The area of each pie slice occupied by that colour represents the proportion of individuals sharing the same haplotype. Colour codes: green: Southern Peninsula; pink: San Ignacio; blue: Mid-Peninsula; red: Northern Peninsula; yellow: Sonora.

California, including the Cape Region and La Giganta ranges; (2) the Central Desert of the peninsula; (3) the Lower Colorado Valley, including the Mojave and the Kofa Mountains in Arizona and (4) the Gulf Coast in Sonora, in the Mexican mainland.

Multivariate analysis

We used PCA to determine whether genetic markers of *Washingtonia* reveal a pattern of clinal variation connecting palm populations along the long peninsular gradient. The broken-stick test on the PCA of SNP data found three axes that jointly account for 62.52% of the total variance using the *de novo* data set. The first axis of the PCA (PC1, 38.42% of total genetic variance) was strongly correlated with latitude ($r^2 = 0.87$). The regression analysis (Table 2) showed that fit of the simple linear model provided an *F*-value than that of the logistic model although the latter showed a smaller value for the Akaike information criterion (AIC). Linear and non-linear models gave closely similar r^2 values when regressed against latitude (0.87 vs. 0.89; Table 2). However, the frequency histogram of the PC1 scores was best fit using a four-mode mixture distribution, a fact that suggests that genetic variation in *Washingtonia* populations from south to north is the result of four distinct haplotypes occurring along the latitudinal cline (Fig. 3B). The four-mode population distribution along PC1, together with the good fit of the linear (clinal) model, questions the two-species hypothesis in support of the idea that there are more than two distinct populations along the latitudinal cline.

The second and third axes reinforced the existence of the four genetic groups discussed in the previous section. The second axis (17.47% of variance) separated the populations in the Central Desert from the general N–S cline described by axis 1, and the third axis (6.63% of total variance) separated the Sonoran mainland (Nacapule) from the general trend (Fig. 3A), providing evidence for the uniqueness of the Sonoran population.

The linear model regression showed that, for all three axes, the region in which each population was found (a four-level factor dividing the whole set of sample sites into Southern, Sonoran, Central and Northern regions) was a better predictor than latitude or the individual oasis. That is, the sampled populations are genetically differentiated into four regions with a small, but significant ($P < 0.0001$) amount of the variation explained by the fixed

Table 2. Comparison of the clinal model (linear model) ANOVA and the hybrid zone model (non-linear model) ANOVA of PC1

	Predictor	SS	d.f.	MS	*F*	*P*	r^2	AIC
Linear model	latitude	910.11	1	910.11	1101.17	< 0.0001	0.86	459.98
	residuals	141.33	171	0.83				
	total	1051.44	172					
Non-linear model	latitude	936.75	3	312.25	460.11	< 0.0001	0.89	427.89
	residuals	114.69	169	0.68				
	total	1051.44	172					

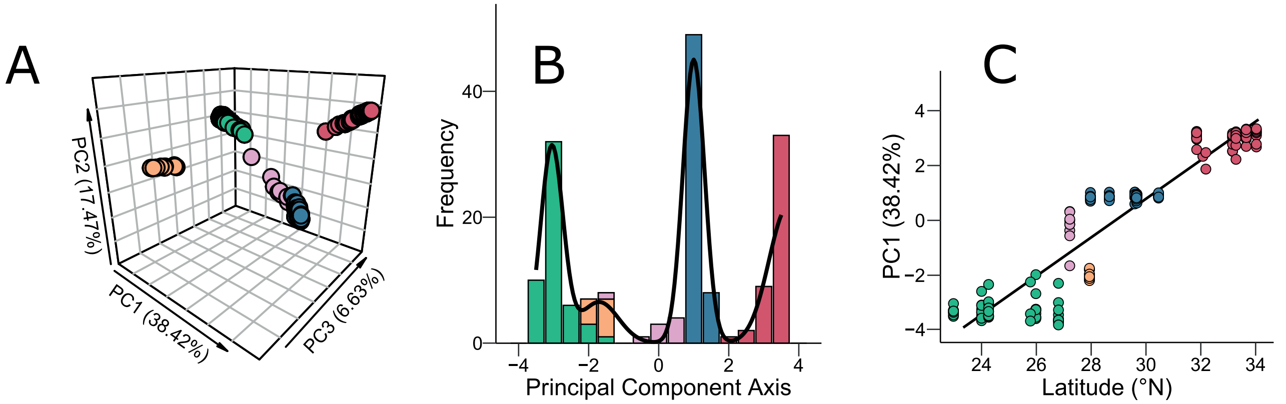


Figure 3. A, PCA plot obtained from a centred PCA of 21 746 SNPs and 173 individuals of 21 populations of *Washingtonia* from the Baja California Peninsula, Southern California, Arizona, and Sonora. B, Histogram of PC1 scores. Individuals present in each haplotype class calculated as the area under the curve of the four-modal density curve adjusted to the distribution frequency. C, Linear model of PC1 against latitude. $r^2 = 0.87$. Populations colour coded by inferred phylogeographic regions. Colour codes: green: Southern Peninsula; pink: San Ignacio; blue: Mid-Peninsula; red: Northern Peninsula; yellow: Sonora.

effect of sites-nested-within-regions (Table 3). In the case of axis 3, which separates Nacapule from the remaining sites, the effect of the site was aliased with that of the region, as the Sonoran region has only one site with palms. Similarly, the dendrogram based on Nei's genetic distance gave a similar, four-group pattern except for San Ignacio (ALA), which formed a loose group, mostly nested in the Mid-Peninsular region, but with one individual genetically different and closer to the Southern Peninsular cluster (Fig. 4A).

Bayesian coalescent analysis

For both the *de novo* and reference-based data sets, the change in cross-entropy started to level off at 4–5 clusters ($K = 4–5$) and remained low until $K = 11$ clusters in the *de novo* data set and $K = 12$ for the reference-based data set, suggesting that the genetic contribution from four ancestral populations of *Washingtonia* (plus the *Brahea* outgroup) can predict the currently observed population structure (Supporting Information S1, S2). The barplots depicting the population structure at $K = 4$ coincided with the UPGMA cladogram based on Nei's genetic distances (Fig. 4).

Phylogenetic relationships

To test whether the genus *Washingtonia* is composed of two species using molecular markers, we conducted phylogenetic analyses of the populations of *Washingtonia* using the *de novo* and the reference data sets producing similar results. For the reference data set, the *Brahea* population was, as expected, identified as an outgroup by the algorithm. For the *de novo* data set, Nacapule (NACA) in the Sonoran mainland was chosen as the root for the tree. With previous evidence that palms in San Ignacio (ALA) are of hybrid origin, the clearest picture of phylogenetic relationships is the analysis excluding San Ignacio. This tree topology suggests that Nacapule and the Southern Peninsular populations are sister to the *Brahea* outgroup. The Mid-Peninsular populations are sister to the Northern Peninsular populations, with that larger clade being

Table 3. Results of nested ANOVA on the three first PCA axes of 21 populations of *Washingtonia* sampled across the Baja California Peninsula and Arizona (USA) and Sonora (Mexico) ($N = 173$ individuals)

	Predictor	d.f.	F	P	r^2	AIC
PC1	region	3	3847.3	***	0.98	168.3
	+site(region)	17	8.0	***	0.01	
	full model	20	583.9	***	0.99	93.2
PC2	region	3	4397.5	***	0.98	13.4
	+site(region)	17	8.4	***	0.01	
	full model	20	666.8	***	0.99	–65.8
PC3	region	3	8995.8	***	0.99	–390.5
	+site(region)	17	7.6	***	0.00	
	full model	20	1355.8	***	0.99	–355.1

Note: *** $P < 0.0001$

sister to the Southern Peninsular + outgroup samples (Fig. 5). Incorporating three individuals from San Ignacio (ALA) into the analysis underscored the non-monophyletic nature of the palms in this oasis, as two individuals grouped with the Mid-Peninsular populations and the third grouped with the Southern Peninsular clade. Alternative trees incorporating all individuals from San Ignacio using both approaches are included in the Supporting Information (S3, S4).

The BFD framework selected three regional populations as the best model describing our dataset, with a four-regional populations model as a close runner-up, but rejected any two-species model (Table 4).

HISTORICAL RELATIONSHIPS AMONG POPULATIONS

For the *de novo* data set, there was strong statistical support for one migration event from the Comondú, San José del Cabo and San Javier clades to the formation of the San Ignacio population (Supporting Information S5). A second, statistically weaker, event is supported by the log-likelihood test from the Joshua Tree, Palm Canyon and Anza Borrego clades to Bocana. The reference-based approach shows strong support for the same migration event inferred in the *de novo* data set, with a second, statistically weaker event from Guadalupe Canyon and Gaskill Tanks clade back to the outgroup samples of *Brahea*, and a third event from Santa Gertrudis to Nacapule (Supporting Information S5).

Pairwise F_{ST} values among populations ranged from 0.001 between Joshua Tree and Anza Borrego to 0.96 between Anza Borrego and Nacapule (Supporting Information Table). Nacapule has the highest F_{ST} values ranging from 0.68 to 0.96 with an average of 0.88. F_{ST} values between Northern and Southern Peninsular populations were > 0.81. Populations within the same region had F_{ST} values ranging from 0.001 to 0.700 (Table 5).

ISOLATION BY DISTANCE

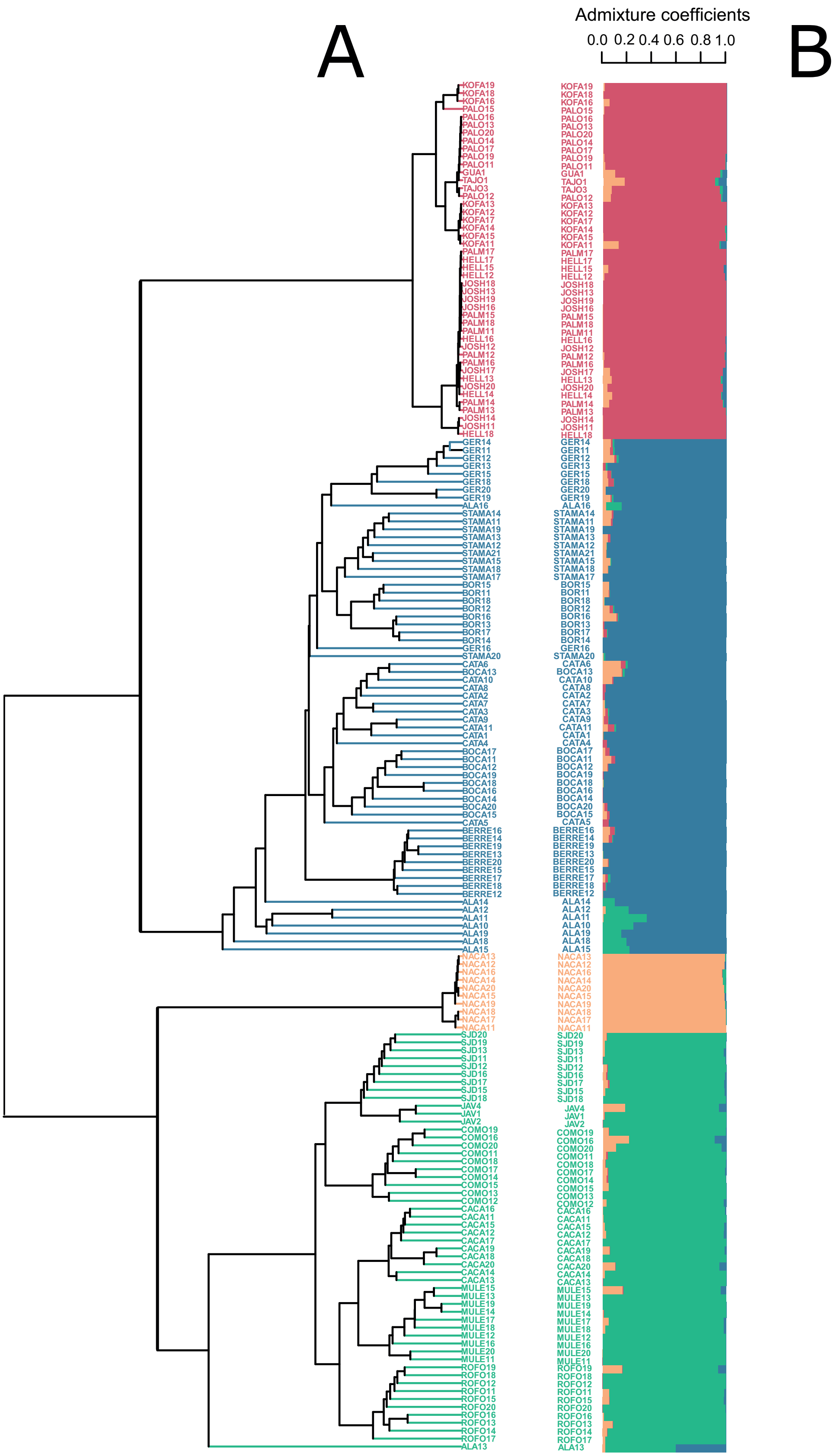
There was a strong association between genetic and geographical distances (Mantel's $r = 0.6$, $P < 0.0001$) considering all pairwise comparisons between populations (Supporting Information, Table S2) and therefore evidence for IBD (Supporting Information S6). The overall pattern also shows four distinct groups corresponding to the proximity class in which each pairwise comparison belongs.

ECOLOGICAL NICHE MODELLING AND ENVIRONMENTAL EFFECTS

For the Northern region, precipitation during the warmest quarter and isothermality had the largest contribution to the MaxEnt model fit (92.3% together). Precipitation during the driest quarter and mean temperature during the warmest quarter were the most critical factors on the current distribution of the Mid-Peninsular populations (79.6%). Finally, the distribution of the Southern Peninsular populations is shaped by precipitation seasonality and precipitation of the driest quarter (87.8%) (Supporting Information, Tables S3, S8). The regression analysis showed that precipitation seasonality, precipitation during the driest quarter, isothermality and mean temperature during the coldest quarter were highly associated with genetic region implying that there is an environmental effect on genetic identity in *Washingtonia* (Supporting Information, Table S4)

ASSOCIATION BETWEEN GENETIC REGIONS AND PHENOTYPIC TRAITS

To determine whether morphological traits varied among genetic regions, we performed an analysis of variance. Our results show that leaf greenness, SLA, DBH, hastula length, petiole width and basal diameter


A
B
Admixture coefficients
0.0 0.2 0.4 0.6 0.8 1.0
KOFA19
KOFA18
KOFA16
PALO15
PALO16
PALO13
PALO20
PALO14
PALO17
PALO19
PALO11
GUA1
TAJO1
TAJO3
PALO12
KOFA13
KOFA12
KOFA17
KOFA14
KOFA15
KOFA11
PALM17
HELL17
HELL15
HELL12
JOSH18
JOSH13
JOSH19
JOSH16
PALM15
PALM18
PALM11
HELL16
JOSH12
PALM12
PALM16
JOSH17
HELL13
JOSH20
HELL14
PALM14
PALM13
JOSH14
JOSH11
HELL18
GER14
GER11
GER12
GER13
GER15
GER18
GER20
GER19
ALA16
STAMA14
STAMA11
STAMA19
STAMA13
STAMA12
STAMA21
STAMA15
STAMA18
STAMA17
BOR15
BOR11
BOR18
BOR12
BOR16
BOR13
BOR17
BOR14
GER16
STAMA20
CATA6
BOCA13
CATA10
CATA8
CATA2
CATA7
CATA3
CATA9
CATA11
CATA1
CATA4
BOCA17
BOCA11
BOCA12
BOCA19
BOCA18
BOCA16
BOCA14
BOCA20
BOCA15
CATA5
BERRE16
BERRE14
BERRE19
BERRE13
BERRE20
BERRE15
BERRE17
BERRE18
BERRE12
ALA14
ALA12
ALA11
ALA10
ALA19
ALA18
ALA15
NACA13
NACA12
NACA16
NACA14
NACA20
NACA15
NACA19
NACA18
NACA17
NACA11
SJD20
SJD19
SJD13
SJD11
SJD12
SJD16
SJD17
SJD15
SJD18
JAV4
JAV1
JAV2
COMO19
COMO16
COMO20
COMO11
COMO18
COMO17
COMO14
COMO15
COMO13
COMO12
CACA16
CACA11
CACA15
CACA12
CACA17
CACA19
CACA18
CACA20
CACA14
CACA13
MULE15
MULE13
MULE19
MULE14
MULE17
MULE18
MULE12
MULE16
MULE20
MULE11
ROFO19
ROFO18
ROFO12
ROFO11
ROFO15
ROFO20
ROFO16
ROFO13
ROFO14
ROFO17
ALA13

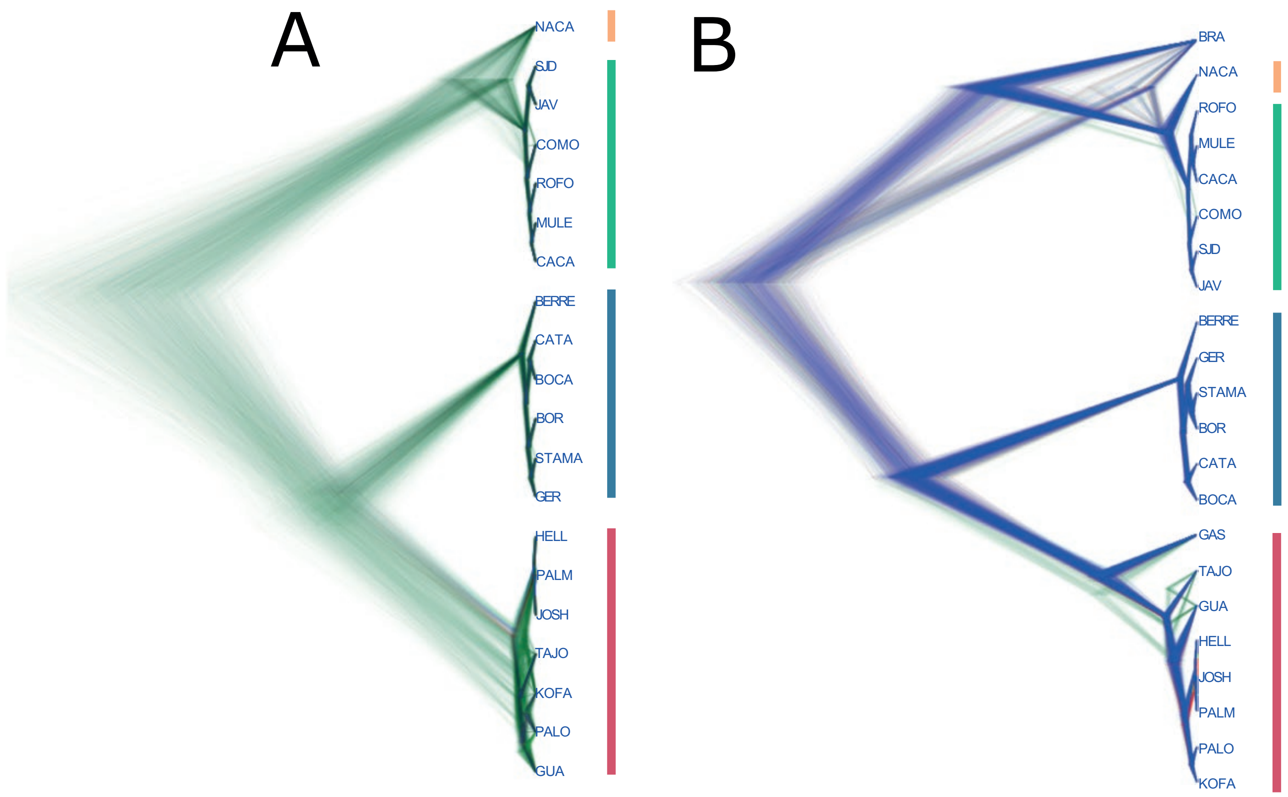


Figure 5. Species-level phylogenetic tree of *Washingtonia* inferred by SNAPP excluding San Ignacio (ALA) using A, 21 746 SNPs obtained through a *de novo* approach and B, 13 459 SNPs obtained through a reference-based approach using *Brahea armata* (BRA) as the outgroup. Colour codes: green: Southern Peninsula; pink: San Ignacio; blue: Mid-Peninsula; red: Northern Peninsula; yellow: Sonora.

were significantly different among regions (Supporting Information S9, Table S5) indicating genetically based differences. Leaf greenness, SLA and DBH were traits on which genetic region had the strongest effect having partial r^2 values ranging from 0.41 to 0.50. Although no morphological trait distinguishes the same four regions, we found using genetic markers that leaf greenness revealed three groups. Unlike genetic markers, leaf greenness was not statistically different between the Southern Peninsular and the Sonoran mainland groups.

SLA was different between the Southern Peninsular region and the Sonoran, Mid-Peninsular and Northern Peninsular regions. Finally, DBH revealed differences between the Southern Peninsular and the Sonoran regions and the Mid-Peninsular and Northern regions (Supporting Information S9, Table S5).

DISCUSSION

Our results of multivariate and Bayesian coalescent methods (Figs 3–5) and those of genetic distance (Supporting Information S6) consistently show four well-differentiated genetic regions that do not follow currently recognized species boundaries in *Washingtonia*, but coincide with the biogeographic subdivisions of the Sonoran Desert first recognized by Nelson and Shreve (Garcillán, Abraham & Ezcurra, 2010). The breaks between the Northern and the Mid-Peninsular regions lie approximately at 31°N, at 27°N between the Mid- and Southern Peninsular regions, and along the Gulf of California between the Nacapule and the Peninsular populations (Fig. 2). Of the four peninsular haplotypes identified in this study, two

---

Figure 4. A, Genetic distance UPGMA tree and B, estimated admixture proportions of 173 *Washingtonia* individuals using STRUCTURE (21746 SNP dataset for $K = 4$ obtained through a *de novo* approach). Each vertical bar represents an individual and each colour represents a genetic cluster. Colour codes: green: Southern Peninsula; blue: Mid-Peninsula; pink: Northern Peninsula; yellow: Sonora.

are consistent with vicariant events proposed for the Peninsula.

Palms in Nacapule comprise a genetically unique group as had been previously shown (Klimova *et al.*, 2018) that is consistent with the separation of the Peninsula from mainland Mexico starting *c.* 5.5 Mya (Riddle *et al.*, 2000). The genetic uniqueness of Nacapule is also evidenced by the high F_{ST} values of pairwise comparisons with peninsular populations and the number of private alleles (Supporting Information Table). Our Bayesian results show that Nacapule is most closely related to the Southern Peninsular region, as has been reported for other plants of the Cape Region with tropical affinity to those of Shreve's Sonora's Central Gulf Coast (Shreve & Wiggins, 1964).

Our Mid-Peninsular region falls in Shreve's Vizcaíno region (Shreve & Wiggins, 1964) and Aschmann's Central Desert (Aschmann, 1967), separated from the Southern Peninsular region at latitude 27°N by a putative seaway near the Vizcaíno Desert that existed *c.* 1 Mya isolating the northern part of the Baja California Peninsula from its southern section (Upton & Murphy, 1997). This seaway has been used to explain genetic differences in other groups of plants such as some Euphorbiaceae and Santalaceae (Garrick *et al.*, 2009; Lira-Noriega *et al.*, 2015).

Table 4. Results of the five BFD* models tested with reported marginal likelihood estimates (MLE) and Bayes factor values comparing the best model to all other models obtained using the *de novo* data set

BFD* Model	MLE	Bayes factor
Three species	–16 314.58	-
Four species	–17 069.81	1510.46
Two species, north and mid-grouped	–20 204.27	7779
Two species, south and mid-grouped	–21 550.6	10 472.04
Two species, north and south grouped	–28 234	23 838.84

Although we suspected the existence of a hybrid zone between latitude 28°N and 29.7°N, encompassing the area between Santa Gertrudis and Cataviña, respectively (Fig. 1), we did not find evidence of admixture except in San Ignacio (ALA), where all individuals show genetic influence of Southern Peninsular and Mid-Peninsular regions (Fig. 4). Although we do not know what the genetic structure of *Washingtonia* was before the arrival of the European missionaries, San Ignacio has been a mission since 1728 and our sampling site is close to the town making introduction of plant material from the south highly likely as our TreeMix results suggest (Supporting Information S5). Furthermore, San Ignacio is found in the transition of Shreve's Vizcaíno and La Giganta regions (Shreve & Wiggins, 1964) or Aschmann's Central Desert, where rainfall is bi-seasonal (Turner *et al.*, 1995; Búrquez *et al.*, 1999). Our Northern Peninsular region is equivalent to Shreve's Lower Colorado Valley (Shreve & Wiggins, 1964). Its separation from the Mid-Peninsular region at latitude 31°N does not coincide with any vicariant event. However, high pairwise F_{ST} values between populations at the edges of these two inter-regional breaks would suggest low genetic flow between them that cannot be explained by linear distance alone.

Genetic structure at the subclade level also shows more complex patterns that cannot be explained by vicariant events alone. Contrary to what has been found in other plant groups (Nason, Hamrick & Fleming, 2002; Garrick *et al.*, 2009; Lira-Noriega *et al.*, 2015) we did not find any evidence to support population structure by a break at the Isthmus of La Paz. In the Southern Peninsular region two subclades can be distinguished from our Bayesian results: one made up of Cacachilas, Mulegé and San Juan de la Costa and one made up of San José del Cabo, San Javier and Comondú. This is not in accordance with what would be expected under the phylogeographic break due to the inundation of the Isthmus of La Paz that would result in isolation of the Cape from other peninsular populations. This pattern could be evidence of human and animal long-distance dispersal. The Cape has been an area of human occupation since before the

Table 5. Mean values of pairwise fixation indices (F_{ST} ± SE) among populations both within a single region (table diagonal) or between regions (lower triangle). Note that because the Sonoran region contains only one site (Nacapule), no pairwise comparisons among sites could be done for this region

	South	Sonoran	Centre	North
South	0.235 (±0.013)			
Sonoran	0.612 (±0.019)	–		
Centre	0.497 (±0.013)	0.589 (±0.038)	0.175 (±0.011)	
North	0.720 (±0.007)	0.878 (±0.001)	0.486 (±0.014)	0.284 (±0.038)

arrival of Europeans (Barco, 1780) and transportation of plant material is likely to have occurred some time ago allowing genetic variation to become uniform across the region.

Although Hodel (2014) suggested that there is no genetic impediment for gene flow between *Washingtonia* spp. in cultivated landscapes, our results show that hybrids are not frequent in natural populations (with the exception of San Ignacio) and that regions maintain genetic cohesiveness (Table 3, Supporting Information, S8), adding evidence to the existence of ecotypes or syngameons (Lotsy, 1931). Increasing geographical distance does not necessarily lead to higher genetic distance if the pair of populations are within the same region (proximity factor = 0). For example, genetic distance between Berrendo and Palomar (found in adjacent regions) is 0.33, despite being 163 km apart, whereas genetic distance between San Juan de la Costa and San José del Cabo (both within the same region), 176 km apart, is 0.084.

The high degree of population structure in *Washingtonia* suggests little gene flow between regions, but high pollen flow within regions among sympatric populations. *Washingtonia* is monoecious with hermaphroditic flowers and, although self-compatibility in *Washingtonia* has not been studied, other species of Coryphoideae, e.g. *Thrinax parviflora* Sw., are self-compatible (Henderson, 1986). Pollination is probably carried out by insects as suggested by its loose inflorescences with fragrant flowers similar to pollination in several genera of Trachycarpeae (Henderson *et al.*, 1997) including *Brahea* (Barrett *et al.*, 2019). Additional fieldwork is needed to identify pollinator specificity and test hypotheses regarding gene flow between regions. Reduced gene flow between regions may also be the result of differences in flowering times (Ellstrand, 2014) driven by ecological conditions as in other palm genera including *Howea* Becc. (Savolainen *et al.*, 2006). It is unknown whether climate change has shifted the phenology of *Washingtonia* increasing differences in flowering times in each region. Genetic structure could also be the result of limited fruit dispersal. Birds are responsible for short distance dispersal, whereas mammals are probably the main colonization agents of *Washingtonia* in Southern California (Bullock, 1980). Although breaks among regions for *Washingtonia* occur at well-known biogeographic limits, reproductive isolation cannot be discarded until a formal phenological investigation in natural populations is conducted.

Herbarium specimen labels and bibliographic sources reveal *W. filifera* generally flowers from February to June (Simono, 2012) or in May and June (Wiggins, 1980; Rebman & Roberts, 2012). In some cases, herbarium specimens show flowering of *W. filifera* could be as early as April in Tajo Canyon (*C.F. Harbison Apr. 1, 1953*, SD) or as late as August in Murray Canyon in the Agua Caliente Indian Reservation (*O. F. Cook Aug. 6, 1906*, BH) or even October in Valle de las Palmas, State of Baja California (*R. Moran 25147*, SD). Park rangers mention flowering in the Agua Caliente palm canyons occurs in June (pers. comm. J. Conley, 2019). Flowering times for *W. robusta* are also inconsistent among sources. Some report that *W. robusta* flowers from April to June (Simono, 2012), but others report flowering from May through June (Wiggins, 1980; Felger *et al.*, 2001) and herbarium specimens show that flowering in *W. robusta* may occur as early as March in Sonora (*L.H. Bailey 16*, BH). Changes in flowering times derived from differences in environmental conditions (an interplay of genotype and environment) is at play as evidenced by a fertile (flowers and developing fruits) cultivated specimen from Corrientes, Argentina, collected in November (*Maruñak and Arbo 167, UC*). Effect of climate change on *Washingtonia* phenology remain to be studied. A flowering specimen of *W. robusta* was collected in September in San Bernardino (*S.B. Parish 5537, UC*). Flowering of cultivated specimens of *W. robusta* in Riverside was not seen in September 2019.

MaxEnt accurately predicted the current distribution of *Washingtonia* palms using seven bioclimatic variables (Supporting Information S7). Out of the seven bioclimatic variables adopted in the model, precipitation during the warmest quarter and isothermality had the largest contribution to the distribution of the Northern Peninsular populations. The large isothermality allows northern palms to photosynthesize during the day even at high temperatures, while maintaining low metabolic rates at night when the temperature is lower.

ECOLOGICAL NICHE MODELLING

For the Northern region, precipitation during the warmest quarter and isothermality made the the largest contribution to the MaxEnt model fit (92.3% together). Precipitation during the driest quarter and mean temperature during the warmest quarter were the most critical factors on the current distribution of the Mid-Peninsular populations (79.6%). Finally, the distribution of the Southern Peninsular populations is shaped by precipitation seasonality and precipitation of the driest quarter (87.8%) (Supporting Information, Table S3).

ASSOCIATION BETWEEN GENETIC REGIONS AND PHENOTYPIC TRAITS

Even when no morphological trait can be used to determine the genetic identity of *Washingtonia* individuals, leaf greenness was the best trait to separate between genetic regions for the exception of the Southern and Sonoran regions. These two groups are distinct using molecular markers, but indistinguishable with morphological traits such as leaf greenness or DBH. Even when morphologically indistinguishable, our work shows individuals in these two regions are using different ecological strategies as evidenced by the difference in SLA mean values. Our results should be interpreted cautiously given the small sample size of the Sonoran population, but they offer insight to the ecological differentiation of palms in the two regions. We found no significant difference in annual precipitation between the Southern Peninsular Cape Region (211.0 mm, SE ± 16.8 mm, $n = 79$) and the Sonoran region at Guaymas (191.3 mm, SE ± 29.8 mm, $n = 29$; $t = 0.58$, $P = 0.28$). The Sonoran region, however, is much hotter in summer. Daily maximum temperatures in July are significantly higher (37.8 °C, SE ± 1.4 °C) in Guaymas than in the Southern Peninsular region (33.8 °C, SE ± 1.7 °C; $t = 12.5$, $P < 0.0001$) and extreme heatwaves of > 45 °C have occurred in Guaymas in 40% of the recorded years, whereas no records > 44 °C have ever been made in the Cape Region. This difference in extreme high temperatures could be driving the difference in SLA between the two regions and be indicative of genetic divergence associated with adaptation to different temperature regimes. Finally, the Southern Peninsular and Sonoran regions had different mean DBH than that of the Mid-Peninsular and Northern regions, showing that DBH is not a useful taxonomic character to distinguish between genetic regions.

TAXONOMIC IMPLICATIONS FOR *WASHINGTONIA*

As currently defined, *Washingtonia* is composed of two species recognized by their geographical distribution and vegetative characters, stem diameter being the most important trait. As others have mentioned (Hausdorf & Hennig, 2010, 2020; Bacon *et al.*, 2012), and to avoid incorrect conclusions regarding species delimitation, we used an eclectic approach including geographical, ecological, genetic and morphological lines of evidence and classical statistical and Bayesian methods to answer the question of what is a species in *Washingtonia*. Also, as recommended for cases of species delimitation where data on fitness is lacking, as in ours, we identified phenotypic and genotypic clusters (Hausdorf & Hennig, 2010) following Mallet's genotypic cluster criteria. In this framework, a species is a 'distinguishable groups of individuals that have few or no intermediates when in contact' (Mallet, 1995: 296). We used SNP data together with morphological traits and found support for the recognition of new and undescribed taxonomic groups of *Washingtonia*. Additionally, we found no evidence of genotypic intermediates in any of the regions examined (with the exception of San Ignacio, probably resulting from a deliberate introduction). Our results also show that the genotypic clusters we found are monophyletic. Following the phylogenetic species concept stating that species share a common ancestor, our results are also congruent using this definition (de Queiroz & Donoghue, 1988). This, added to the gradual morphological variation (morphological clusters), results in our decision to recognize three subspecies in *Washingtonia* for the Baja California Peninsula, plus an additional population in Sonora that should be recognized because of its unique genetic diversity in addition to its geographical isolation.

The cohesion of each group is probably maintained by gene flow within regions and an absence of gene flow between regions, as evidenced by F_{ST} values, and by ecological adaptations that allow them to persist in the diverse niches they currently occupy. However, information on gene flow will greatly increase our understanding or the forces driving speciation in *Washingtonia*.

We have shown that *Washingtonia* was only partially understood since no connecting populations belonging to the Mid-Peninsular region had previously been properly characterized. Considering how little is known about flowering in *Washingtonia* in general and that palms in this region have been considered either as *W. filifera* or *W. robusta*, learning more about flowering times in the Mid-Peninsular region is necessary to understand isolation barriers in the genus.

CONCLUSIONS

Using a widely distributed palm genus of the Sonoran Desert, we have shown that *Washingtonia* populations are highly structured in four phylogeographic regions that do not correspond to current species. These regions show high levels of genetic differentiation that might be a consequence of differences in environmental conditions. Our results did not reveal the existence of a hybrid zone, giving further support to our morphological results in that *Washingtonia* is composed of a single species along the latitudinal transect we examined. Even when no morphological trait can be used to determine the genetic identity of individuals of *Washingtonia*, leaf greenness, a novel morphological trait, can be useful. Our work lays the groundwork for speciation with ongoing genetic exchange in *Washingtonia* and offers possibilities to investigate gene flow and species cohesion in the group.

ACKNOWLEDGEMENTS

The authors thank the rancheros in Baja California and Sonora, Mexico, for providing access to study sites and aid in the field. Special thanks to Ejido Tepentú in Punta Coyote, Aarón Esliman from Rancho Cacachilas, Sergio Aguilar in San Ignacio, Miguel Dávila Pacheco in Cañón del Nacapule, José Ángel Gerardo Monteón and Genaro Gerardo Gaxiola in San Borja and the Velazco Espinosa family in Cataviña. We thank the Agua Caliente Band of Cahuilla Indians, the National Park Service, the California State Parks System and the US Fish and Wildlife Service for providing permits to access and collect in California and Arizona. We also thank Craig Barrett and an anonymous reviewer for extremely helpful criticism to a previous version of this paper. Special thanks to Silvia Bottger and Elisabet V. Wehncke for the translations from German and Peter Fraissinet and staff at BH for help consulting the collection. We also thank the Koenig Laboratory at UCR for helpful discussions, Sue Sherman-Broyles for helpful analysis suggestions and Aleyda Acosta-Rangel, Eleinis Ávila-Lovera, Jean-Christophe Arbonne, Stephen H. Bullock, Jesús Pablo Carrillo León, Sara Ceccarelli, Eric Focht, Pedro P. Garcillán, Peter Ibsen, Israel Jiménez-Luna, Anastasia Klimova, Eulogio López Reyes, Richard Minnich, Vanessa Pérez, Fidel Rojo de Santiago, Marta Ruiz-Valdés, Mario Salazar Ceseña, José Jesús Sánchez-Escalante, Andrew Sanders, Andrew Semotiuk, Christian Silva Bejarano, Alan Vidali, Tatiane Vieira and Antonio J. Zerpa for field assistance. We are grateful to the High-Performance Computing Center (HPCC) at UC Riverside for providing computing resources. Funding was provided by the University of California Institute for Mexico and the United States (UC MEXUS) and the Consejo Nacional de Ciencia y Tecnología (Conacyt) under fellowship number 382840 (LV), the National Science Foundation Plant Genome Fellowship 1711807 (JBL) and UC Riverside startup funds (DK).

AUTHOR CONTRIBUTIONS

LV-A designed research, conducted fieldwork, performed wet laboratory work, analysed and interpreted data and wrote the manuscript. JL designed research, performed wet laboratory work, analysed and interpreted data and wrote the manuscript. EE designed research, analysed and interpreted data and wrote the manuscript. DK designed research, provided funding, interpreted data and wrote the manuscript.

DATA AVAILABILITY STATEMENT

The sequences in this study are available in NCBI.

REFERENCES

André ÉF. 1900. *Washingtonia robusta*. *Revue Horticole* **72:** 61–65. Available at: https://www.biodiversitylibrary.org/item/197178#page/71/mode/1up

Araújo MB, **Guisan A. 2006.** Five (or so) challenges for species distribution modelling. *Journal of Biogeography* **33:** 1677–1688.

Arriaga L, **Díaz S**, **Domínguez Cadena R**, **León de la Luz JL. 1997.** Composición florística y vegetación. In: Arriaga L, Rodríguez Estrella R, eds. *Los oasis de la península de Baja California.* La Paz: Centro de Investigaciones Biológicas del Noroeste.

Aschmann H. 1967. *The Central Desert of Baja California: demography and ecology.* Riverside: Manessier Publishing Company.

Bacon CD, **McKenna MJ**, **Simmons MP**, **Wagner WL. 2012.** Evaluating multiple criteria for species delimitation: an empirical example using Hawaiian palms (Arecaceae: *Pritchardia*). *BMC Evolutionary Biology* **12:** 23.

Bailey LH. 1936. *Washingtonia*. *Gentes Herbarum* **4:** 53–82.

del Barco M. 1780. *Correcciones y adiciones a la historia o noticia de la California en su primera edición de Madrid, año de 1757.* Mexico City: Universidad Nacional Autónoma de México, Instituto de Investigaciones Históricas, 1988.

Barrett CF, **Sinn BT**, **King LT**, **Medina JC**, **Bacon CD**, **Lahmeyer SC**, **Hodel DR. 2019.** Phylogenomics, biogeography and evolution in the American genus *Brahea* (Arecaceae). *Botanical Journal of the Linnean Society* **190:** 242–259.

Bouckaert R, **Heled J**, **Kühnert D**, **Vaughan T**, **Wu CH**, **Xie D**, **Suchard MA**, **Rambaut A**, **Drummond AJ. 2014.** BEAST **2:** a software platform for Bayesian evolutionary analysis. *PLoS Computational Biology* **10:** e1003537.

Bullock SH. 1980. Dispersal of a desert palm by opportunistic frugivores. *Principes* **24:** 29–32.

Búrquez A, **Martinez-Yrizar A**, **Felger RS**, **Yetman D. 1999.** Vegetation and habitat diversity at the southern edge of the Sonoran Desert. In: Robichaux RH, ed. *Ecology of Sonoran Desert plants and plant communities.* Tucson: The University of Arizona Press, 36–67.

Catchen J, **Hohenlohe PA**, **Bassham S**, **Amores A**, **Cresko WA. 2013.** Stacks: an analysis tool set for population genomics. *Molecular Ecology* **22:** 3124–3140.

Cornett JW. 1987. Three palm species at Cataviña. *Principes* **31:** 12–13.

Danecek P, **Auton A**, **Abecasis G**, **Albers CA**, **Banks E**, **DePristo MA**, **Handsaker RE**, **Lunter G**, **Marth GT**, **Sherry ST**, **McVean G**, **Durbin R**; **1000 Genomes Project Analysis Group**. **2011.** The variant call format and VCFtools. *Bioinformatics (Oxford, England)* **27:** 2156–2158.

Doyle J, **Doyle J. 1987.** A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin, Botanical Society of America* **19:** 11–15.

Durand EY, **Patterson N**, **Reich D**, **Slatkin M. 2011.** Testing for ancient admixture between closely related populations. *Molecular Biology and Evolution* **28:** 2239–2252.

Ellstrand NC. 2014. Is gene flow the most important evolutionary force in plants? *American Journal of Botany* **101:** 737–753.

Elshire RJ, **Glaubitz JC**, **Sun Q**, **Poland JA**, **Kawamoto K**, **Buckler ES**, **Mitchell SE. 2011.** A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* **6:** e19379.

Felger RS, **Johnson MB**, **Wilson MF. 2001.** *The trees of Sonora, Mexico.* New York: Oxford University Press.

Felger RS, **Joyal E. 1999.** The palms (Arecaceae) of Sonora, Mexico. *Aliso: a Journal of Systematic and Evolutionary Botany* **18:** 1–18.

Fick SE, **Hijmans RJ. 2017. WorldClim 2: new 1**-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology* **37: 4302**–4315.

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

Garcillán PP, **Abraham CEG**, **Ezcurra E. 2010.** The cartographers of life: two centuries of mapping the natural history of Baja California. *Journal of the Southwest* **52:** 1–40.

Garcillán PP, **Ezcurra E. 2003.** Biogeographic regions and β-diversity of woody dryland legumes in the Baja California peninsula. *Journal of Vegetation Science* **14:** 859–868.

Garrick RC, **Nason JD**, **Meadows CA**, **Dyer RJ. 2009.** Not just vicariance: phylogeography of a Sonoran Desert euphorb indicates a major role of range expansion along the Baja peninsula. *Molecular Ecology* **18:** 1916–1931.

Hausdorf B, **Hennig C. 2010.** Species delimitation using dominant and codominant multilocus markers. *Systematic Biology* **59:** 491–503.

Hausdorf B, **Hennig C. 2020.** Species delimitation and geography. *Molecular Ecology Resources* **20:** 950–960.

Henderson AJ. 1986. A review of pollination studies in the Palmae. *The Botanical Review* **52:** 221–259.

Henderson AJ, **Galeano-Garces G**, **Bernal R. 1997.** *Field guide to the palms of the Americas.* Princeton: Princeton University Press.

Himmelblau DM. 1972. *Applied nonlinear programming.* New York: McGraw-Hill.

Hodel DR. 2014. *Washingtonia ×filibusta* (Arecaceae: Coryphoideae), a new hybrid from cultivation. *Phytoneuron* **68:** 1–7.

Ikeda DH, **Max TL**, **Allan GJ**, **Lau MK**, **Shuster SM**, **Whitham TG. 2017.** Genetically informed ecological niche models improve climate change predictions. *Global Change Biology* **23:** 164–176.

Inman R, **Franklin J**, **Esque T**, **Nussear K. 2018.** Spatial sampling bias in the *Neotoma* paleoecological archives affects species paleo-distribution models. *Quaternary Science Reviews* **198:** 115–125.

Ishihata K, **Murata H. 1971.** Morphological studies in the genus *Washingtonia*: on the intermediate form between *Washingtonia filifera* (L. Linden) H. Wendland and *Washingtonia robusta* H. Wendland. *Memoirs of the Faculty of Agriculture, Kagoshima University* **8:** 331–354.

Jackson DA. 1993. Stopping rules in principal components analysis: a comparison of heuristical and statistical approaches. *Ecology* **74:** 2204–2214.

Johnson D. 1996. *Palms: their conservation and sustained utilization. Status survey and conservation action plan.* Gland, Cambridge: IUCN.

Jombart T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics (Oxford, England)* **24:** 1403–1405.

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

Kamvar ZN, **Tabima JF**, **Grünwald NJ. 2014.** Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* **2:** e281.

Klimova A, **Ortega-Rubio A**, **Vendrami DLJ**, **Hoffman JI. 2018.** Genotyping by sequencing reveals contrasting patterns of population structure, ecologically mediated divergence, and long-distance dispersal in North American palms. *Ecology and Evolution* **8:** 5873–5890.

Leaché AD, **Fujita MK**, **Minin VN**, **Bouckaert RR. 2014.** Species delimitation using genome-wide SNP data. *Systematic Biology* **63:** 534–542.

León de la Luz JL, **Domínguez Cadena R**, **Domínguez León M**, **Coria Benet RC. 2014.** *Flora iconográfica de Baja California Sur.* La Paz: Centro de Investigaciones Biológicas del Noroeste.

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint arXiv:1303.3997v2 [q-bio.GN]. Available at: https://arxiv.org/abs/1303.3997.

Li H, **Handsaker B**, **Wysoker A**, **Fennell T**, **Ruan J**, **Homer N**, **Marth G**, **Abecasis G**, **Durbin R**; **1000 Genome Project Data Processing Subgroup**. **2009.** The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)* **25:** 2078–2079.

Lira-Noriega A, **Toro-Núñez O**, **Oaks JR**, **Mort ME. 2015.** The roles of history and ecology in chloroplast phylogeographic patterns of the bird-dispersed plant parasite *Phoradendron californicum* (Viscaceae) in the Sonoran Desert. *American Journal of Botany* **102:** 149–164.

Lotsy JP. 1931. On the species of the taxonomist in its relation to evolution. *Genetica* **13:** 1–16.

Mallet J. 1995. A species definition for the modern synthesis. *Trends in Ecology & Evolution* **10:** 294–299.

Matzke NJ. 2013. Probabilistic historical biogeography: new models for founder-event speciation, imperfect detection, and fossils allow improved accuracy and model-testing. *Frontiers of Biogeography* **5:** 242–248.

McClenaghan LR, **Beauchamp AC. 1986.** Low genic differentiation among isolated populations of the California fan palm (*Washingtonia filifera*). *Evolution* **40:** 315–322.

Minnich RA, **Franco-Vizcaíno E**, **Salazar-Ceseña M. 2011.** Distribution and regional ecology of Californian palm oases interpreted from Google Earth images. *Aliso: a Journal of Systematic and Evolutionary Botany* **29:** 1–12.

Moran R. 1978. Palms in Baja California. *Principes* **22:** 47–55.

Nabhan GP. 1986. The palms in our hands. In: Nabhan GP. *Gathering the desert.* Tucson: University of Arizona Press, 21–35.

Nason JD, **Hamrick JL**, **Fleming TH. 2002.** Historical vicariance and postglacial colonization effects on the evolution of genetic structure in *Lophocereus*, a Sonoran Desert columnar cactus. *Evolution; International Journal of Organic Evolution* **56:** 2214–2226.

Ortiz ME. 2019. *vcf2phylip v2.0: convert a VCF matrix into several matrix formats for phylogenetic analysis.* Available at: https://zenodo.org/record/2540861#.X_icTtj7Q2w.

Paris JR, **Stevens JR**, **Catchen JM. 2017.** Lost in parameter space: a road map for stacks. *Methods in Ecology and Evolution* **8:** 1360–1373.

Phillips SJ, **Anderson RP**, **Schapire RE. 2006.** Maximum entropy modeling of species geographic distributions. *Ecological Modelling* **190:** 231–259.

Phillips SJ, **Dudík M**, **Elith J**, **Graham CH**, **Lehmann A**, **Leathwick J**, **Ferrier S. 2009.** Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological Applications: a Publication of the Ecological Society of America* **19:** 181–197.

Pickrell JK, **Pritchard JK. 2012.** Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genetics* **8:** e1002967.

Purcell S, **Neale B**, **Todd-Brown K**, **Thomas L**, **Ferreira MA**, **Bender D**, **Maller J**, **Sklar P**, **de Bakker PI**, **Daly MJ**, **Sham PC. 2007**. PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics* **81:** 559–575.

de Queiroz K, **Donoghue MJ. 1988.** Phylogenetic systematics and the species problem. *Cladistics* **4:** 317–338.

R Development Core Team. 2018. *R: a language and environment for statistical computing.* Vienna: The R Foundation for Statistical Computing.

Rebman JP, **Roberts NC. 2012.** *Baja California plant field guide.* San Diego: San Diego Natural History Museum and Sunbelt Publications.

Riddle BR, **Hafner DJ**, **Alexander LF**, **Jaeger JR. 2000.** Cryptic vicariance in the historical assembly of a Baja California peninsular desert biota. *Proceedings of the National Academy of Sciences of the United States of America* **97:** 14438–14443.

Rochette NC, **Catchen JM. 2017.** Deriving genotypes from RAD-seq short-read data using Stacks. *Nature Protocols* **12:** 2640–2659.

Rowan BA, **Seymour DK**, **Chae E**, **Lundberg DS**, **Weigel D. 2017.** Methods for genotyping-by-sequencing. *Methods in Molecular Biology* **1492:** 221–242.

RStudio Team. 2018. *RStudio: integrated development environment for R.* Boston: RStudio, Inc.

Savolainen V, **Anstett MC**, **Lexer C**, **Hutton I**, **Clarkson JJ**, **Norup MV**, **Powell MP**, **Springate D**, **Salamin N**, **Baker WJ. 2006.** Sympatric speciation in palms on an oceanic island. *Nature* **441:** 210–213.

Shreve F, **Wiggins IL. 1964.** *Vegetation and flora of the Sonoran Desert.* Stanford: Stanford University Press.

Simono S. 2012. *Washingtonia*. In: Project JF, ed. *The Jepson manual: vascular plants of California.* Berkeley: University of California Press.

Turner RM, **Bowers JE**, **Burgess TL. 1995.** *Sonoran Desert plants: an ecological atlas.* Tucson: University of Arizona Press.

Upton DE, **Murphy RW. 1997.** Phylogeny of the side-blotched lizards (Phrynosomatidae:*Uta*) based on mtDNA sequences: support for midpeninsular seaway in Baja California. *Molecular Phylogenetics and Evolution* **8:** 104–113.

Villanueva-Almanza L, **Garcillán PP**, **Ávila-Lovera E**, **Pérez V**, **Silva Bejarano C**, **Focht E**, **Ezcurra E. 2018.** A Hollywood palm icon unmasked: clinal variation in *Washingtonia* (Arecaceae) of Peninsular California. *Botanical Journal of the Linnean Society* **188:** 406–425.

Vogl RJ, **McHargue LT. 1966.** Vegetation of California fan palm oases on the San Andreas Fault. *Ecology* **47:** 532–540.

Wiggins IL. 1980. *Flora of Baja California.* Stanford: Stanford University Press.

Zona S. 2000. Arecaceae. In: Flora of North America Editorial Committee, ed. 1993+. *Flora of North America North of Mexico [Online]. Vol. 21+, Vol. 22.* New York and Oxford. Available at: http://www.efloras.org/florataxon.aspx?flora_id=1&taxon_id=10061. Accessed 26 January 2021.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Figure S1. A, Cross-entropy plot of the dataset obtained through a *de novo* approach. B–K, Barplots from $K = 2–11$ of estimated admixture proportions of 173 *Washingtonia* palms using a dataset of 21 746 SNP obtained through a *de novo* approach. Each vertical bar represents an individual and each colour represents a genetic cluster.

Figure S2. A, Cross-entropy plot of the dataset obtained through a reference-based approach. B–L, Barplots from $K = 2–12$ of estimated admixture proportions of 173 *Washingtonia* palms using a dataset of 21 746 SNP obtained through a reference-based approach. Each vertical bar represents an individual and each colour represents a genetic cluster.

Figure S3. Species-level phylogeny of *Washingtonia* inferred by SNAPP using 21746 SNPs obtained through a de novo approach: A, three San Ignacio clusters (ALA10, ALA15, ALA14); B, two San Ignacio clusters (ALA13(ALA12 + ALA18)); C, one cluster of a single individual of San Ignacio (ALA13) and D, San Ignacio as a monophyletic cluster.

Figure S4. Species-level phylogeny of *Washingtonia* inferred by SNAPP using 23514 SNPs obtained through a reference-based approach: A, two San Ignacio clusters, B, three San Ignacio clusters; C, San Ignacio as a monophyletic cluster and D, one cluster of a single individual of San Ignacio.

Figure S5. MLE trees showing patterns of genetic divergence among *Washingtonia* populations using a: A, *de novo* approach with NACA set as the outgroup based on the positioning of this population in the distance dendrogram and phylogenetic tree; and B, reference-based approach with *Brahea* individuals set as the outgroup. Inferred historical migration events shown with arrows with gene flow direction reflected by the position of the arrow. Higher colour intensity shows the likelihood of the migration event.

Figure S6. Isolation by distance (IBD) of *Washingtonia* populations. Analysis was based on the correlation between Nei's genetic distance and geographical distance of 21 populations from the Baja California peninsula, Southern California, Arizona and Sonora. $N = 210$ unique sample pairs. $r^2 = 0.68$. Red: genetic distance among pairs of populations within the same region; dark orange: genetic distance among pairs of populations from adjacent regions; light orange: genetic distance among pairs of populations separated by one region; yellow: genetic distance among pairs of populations at extreme regions.

Figure S7. Predicted potential distribution of *Washingtonia* in the Baja California peninsula, Southern California, Arizona, and Sonora under current niches using MaxEnt applying genetic regions instead of species names: A, Northern Peninsular region, B, Mid-Peninsular region and C, Southern Peninsular region. The Sonoran region is not included because there were not enough samples for testing. Red colours represent areas with high probability of suitable conditions, green show areas with conditions typical of those where *Washingtonia* is found and blue are areas with low probability of suitable conditions. Occurrence data comes from 162 individuals belonging to the four genetic regions: 54 from the central region, 45 from the northern, ten from the Sonoran and 52 from the Southern Peninsular region.

Figure S8. Bioclimatic variables of the four genetic regions of *Washingtonia* palms in the Baja California peninsula, Southern California, Arizona and Sonora. Tukey multiple comparison of means: means with the same letter are not statistically significant, $P < 0.05$. Bars show mean ± standard error of the mean (SE).

Figure S9. Morphological traits of the four genetic regions of *Washingtonia* palms in the Baja California peninsula, Southern California, Arizona and Sonora. Tukey multiple comparison of means: means with the same letter are not statistically significant, $P < 0.05$. Bars show mean ± standard error of the mean (SE). Note: a, b, c, or d is group divided by Tukey's HSD test.

Table S1. D statistic (admixture coefficient values) and significance (i.e. probabilities under a null hypothesis of random variation with no admixture) from TreeMix (v.1.13; Pickrell & Pritchard, 2012) models for zero to ten migration events, as well the log-likelihood values for each model from package BioGeoBEARS v.0.2.1 (Matzke, 2013). From all models with a significantly positive admixture coefficient (D statistic), the one with the highest log-likelihood value was selected as the most likely number of past migration events.

Table S2. Results of a GLM for genetic distance of 21 populations of *Washingtonia* sampled across Peninsular California, Arizona and Sonora, Mexico ($N = 210$ unique pairwise comparisons).

Table S3. Percentage contributions and permutation importance of the bioclimatic variables included in the MaxEnt models for the four genetic regions of *Washingtonia*. Note that because the Sonoran region contains only ten occurrences, no model could be generated for this region.

Table S4. One-way ANOVA of bioclimatic variables with genetic region as predictor. Traits ordered by ascending significance level.

Table S5. One-way ANOVA of morphological traits with genetic region as predictor. Traits ordered by ascending significance level.