Conservation Genetics (2024) 25:335–355
https://doi.org/10.1007/s10592-023-01569-4

RESEARCH ARTICLE

Conservation genomics of *Dioon holmgrenii* (Zamiaceae) reveals a history of range expansion, fragmentation, and isolation of populations

Brian L. Dorsey[1] · Silva H. Salas-Morales[2] · Timothy J. Gregory[3]

Received: 21 February 2023 / Accepted: 5 September 2023 / Published online: 5 October 2023

Abstract
Many Cycad species may not survive the current extinction crisis, despite belonging to the oldest living seed plant lineage. Conservation of endangered and threatened species will require a combination of in situ and ex situ programs, both of which will benefit from better knowledge of species' population genetic structure, as will assessments of threatened status. Here we develop a cost-effective method of obtaining population-level genomic data from across the range of the Mexican cycad, *Dioon holmgreni*, and use these data to characterize the genetic structure and diversity of the species. We also reconstruct aspects of the demographic history of the species and evaluate the taxonomic cohesion of populations across the range using genomic and morphological data. We find that *D. holmgrenii* harbors moderate genetic diversity across genetically and geographically isolated populations that each possess a substantial percentage of private alleles. We further find that the history of this species likely includes a widespread range expansion followed by fragmentation due to population contraction. These results argue for conservation of all populations and their unique alleles but also suggest an unexpected ability of this species to maintain genetic diversity despite dramatic reductions in population size.

Keywords Bottleneck · Cycads · *Dioon* · Range expansion · RADseq

Introduction

Cycads (Cycadales) belong to one of the oldest living seed plant lineages with a fossil record dating back at least to the Permian (Norstog and Nicholls 1997; Taylor et al. 2009). They are also the most threatened lineage of plants on the planet with 65% of the species considered at least Vulnerable to Extinction by the IUCN (Calonje et al. 2013). So, it is possible that this lineage, which has survived multiple mass extinction events, will not survive the current extinction crisis. The major threats to cycad species' survival are illegal collection for the horticultural trade and individual collectors, and land conversion/loss of habitat (IUCN 2021). Efforts to prevent further loss of cycad species include establishing reserves to protect habitat and, when this is not yet possible, growing plants ex situ as assurance colonies that will provide a source for reestablishment when reserves are possible (Donaldson 2003). Both approaches will benefit from a better understanding of the genetic diversity and structure within and among populations of target species.

Dioon (Zamiaceae) is a cycad genus mostly endemic to Mexico (one species occurs in Honduras) with 17 currently recognized species (Calonje et al. 2013). Eleven of these species are categorized as Vulnerable or Endangered in the IUCN Red List (IUCN 2021), although both the taxonomy and assessments of *Dioon* in the Red List need updating. A good example of the outdated status of our understanding of threats to these species is *Dioon holmgrenii*. The Red List information for this species includes only two populations and a very small extent of occurrence. However, since the last assessment, at least ten populations ranging over approximately 4600 km^2 in the Sierra Madre del Sur and Pacific slope of Oaxaca are now known (Velasco-García

✉ Brian L. Dorsey
bdorsey@huntington.org

[1] The Huntington Library, Art Museum, and Botanical Gardens, 1151 Oxford Rd, San Marino, CA 91108, USA

[2] Sociedad Para El Estudio de los Recursos Bioticos de Oaxaca (SERBO), Camino Nacional No. 80-B, AP 533, San Sebastián Tutla, 71246 Oaxaca, Mexico

[3] Wild Cycad Conservancy, c/o 280 Woodside Dr., Woodside, CA 94062, USA

et al. 2016; T.G. and S.S. pers. obs.) In addition, the species is quite fragmented—a feature not reported in the last Red List assessment.

Genetic diversity is a major target for conservation in global strategies aimed at preserving biodiversity (e.g. Convention on Biological Diversity, Aichi Biodiversity Targets, 2010; https://www.cbd.int/sp/targets/; https://www.cbd.int/article/cop15-cbd-press-release-final-19dec 2022). Assessments of the threat to survival can benefit from population genetic information (Garner et al. 2020) and the genetic diversity and structure within a species can be used to delimit entities worthy of conservation efforts (Moritz 1994; Coates et al. 2018; García-Dorado and Caballero 2021). An increase in the known number of populations and individuals, as has been documented in *D. holmgrenii*, may justify a change to a less threatened Red List status. However, given the importance of the current genetic diversity within a species for its future adaptation, the distribution of genetic diversity should be assessed and factored into any listing or management decisions. If the genetic diversity of the species is low (e.g., due to bottleneck effects) or is highly structured because of limited dispersal, then the increase in known individuals may not translate into a lower threat to long-term survival and adaptability. In fact, a species that is highly structured into more or less discrete genetic clusters (populations) will require a more extensive reserve system and any ex situ conservation collections will need to represent the diversity and maintain the structure into the future.

Given the fragmented distribution of *D. holmgrenii* we expect some genetic structure across the range unless this distribution is extremely recent (Bohonak 1999; Wright 1949). An important question regarding genetic diversity and structure is how any such structure came to be. Historical demographic processes such as gene flow, dispersal, and demographic change can all affect geographic distribution and genetic structure. A better understanding of these processes can help us infer the history of the species, which will in turn improve our understanding of the possibilities for such processes in the future as we enter another era of rapid climate change.

Taxonomic status is also an important factor in species conservation (Mace 2004). Several new species of *Dioon* have been recognized recently, based on morphology and genetic divergence (Gutiérrez-Ortega, et al. 2018a, b; Gutiérrez-Ortega, et al. 2020). These studies have increased our understanding of *Dioon* evolution and helped delineate potential targets for conservation efforts. They also reinforce the need for critical evaluation of species definitions in this genus. Distribution patterns of current *Dioon* taxa vary widely and include both widespread and highly restricted species. For example, *D. argenteum* is known only from three populations restricted to a single canyon, whereas *D. edule* comprises dozens of populations from coastal Veracruz through the Sierra Madre Oriental to Nuevo Leon. Moreover, no less than six species of *Dioon* occur throughout the Tehuacán-Cuicatlán Valley, an area comparable in size to the area of occupation of *D. holmgrenii*. Given these disparate distributions and the evidence that most species are of recent and contemporaneous origin, it is possible that more segregate species exist among the more widespread taxa (Dorsey et al. 2018). In fact, our observations across the known range of *D. holmgrenii* suggested that there may be divergent morphologies among localities that could indicate within species divergence or the presence of multiple species. Specifically, it appears that plants in the Ixtayutla area may be distinct from others we've observed. This locality is at the extreme western edge of the known distribution of the species and so may be an isolated and diverging population and perhaps a new taxon.

Characterizing the genetic diversity of a species and detecting how that diversity is spatially structured requires sufficient genetic data across the species' range. To date, no study has included the full range of *D. holmgrenii* when assessing genetic diversity and most studies in *Dioon* species have relied on allozyme data or microsatellites (González-Astorga et al. 2003, 2005; González-Astorga et al. 2008a, b; Cabrera-Toledo et al. 2008, 2010; González-Astorga et al. 2008a, b; Velasco-García et al. 2021). However, several groups have utilized single nucleotide polymorphism (SNP) data in *Dioon* and other cycad species (Clugston et al. 2019, 2022; Gutiérrez-Ortega et al. 2020). A number of recent studies have attempted to evaluate the utility of microsatellites vs. SNPs for population genetics and related fields and a few have looked at haplotypes from short read data (e.g. Fischer et al. 2017; Baetscher et al. 2018; Zimmerman et al. 2020). While these studies generally find correlations between summary statistics derived from microsatellites and SNPs, the latter do show some advantages such as smaller confidence intervals and greater ability to detect finer patterns of population structure. More promising still are loci that contain several SNPs and multiple alleles. These 'microhaplotypes' (Baetscher et al. 2018) may offer the advantages of both microsatellites and SNPs in that they are more variable and contain more information than single SNPs but are more easily obtained in larger number and cover a greater proportion of the genome than is possible for microsatellites. Here we develop a cost-effective method to obtain thousands of microhaplotypes across hundreds of individuals from across the entire range of *D. holmgrenii*.

In this study we use haplotype and SNP data to characterize the genetic diversity and structure of *D. holmgrenii* in order to evaluate its taxonomic status and provide a genomic basis for any future conservation programs. Specifically, we aim to test whether known stands of *D. holmgrenii* represent distinct genomic clusters and if there exists higher order structure among them, assess the level of gene flow between

localities, compare any genetic structure with patterns of morphological diversity among populations, and test for patterns of past demographic and/or distributional change that may help explain the current phylogeography.

Materials and methods

Sampling for genomic analysis

We collected leaf tissue from six localities across the known range of *D. holmgrenii* (Fig. 1) and one outgroup population of *D. planifolium*. Sample sizes ranged from 27 to 43 individuals per population. DNA was extracted according to Dorsey et al. (2018).

RADseq library preparation

Cycad genomes are very large with the average size in *Dioon* estimated at approximately 49 Gb (Zonneveld 2012). In order to cost effectively sequence a subset of the genome for a population level sampling of this species we developed a RADseq method that maximizes high coverage loci and number of individuals for a given sequencing effort by combining ddRAD and RESTseq techniques (Peterson et al. 2012; Stolle and Moritz 2013). Briefly, we used the R package simRAD (Lepais and Weir 2014; R Core Team 2019) to simulate and digest in silico a genome with the same GC content measured across loci sequenced in Dorsey et al. (2018). Tests of combinations of restriction enzymes and size selection ranges identified MspI and MboI (Thermo Scientific, Waltham, MA USA) as the combination that would produce a majority of fragments between 300 and 500 bp long. Using these enzymes we followed the ddRAD protocol of Peterson et al. (2012) through the P1/P2 adapter ligation and pooling step, modified for the requirements of our enzymes. We then used three additional enzymes (AluI, SaqAI, and Csp6I; Thermo Scientific, Waltham, MA USA) in a second round of digestion following the RESTseq method of Stolle and Mortiz (2013). This creates smaller fragments with only one P1/P2 adapter to which the Illumina flow cell adapter sequence cannot be incorporated in

Fig. 1 Map of southwestern Oaxaca, MX showing sampling localities of *D. holmgrenii*. These localities cover the known geographic extent of occurrence for the species

subsequent PCR because this process requires both adapters. The result is a smaller library enriched for a random fraction of loci from the original digest, but that are within our target size range. This smaller number of loci allowed us to include more individuals per lane while still ensuring sufficient coverage per locus for genotyping, resulting in cost-effective sequencing of population level sampling. Seven libraries were pooled and sequenced on one S1 lane of a Novaseq producing 150 bp paired end reads. Quality control and sequencing were done at the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley.

Filtering sequences

We used the process_radtags program from the Stacks (Catchen et al. 2011, 2013) pipeline for demultiplexing and initial filtering of raw reads (--barcode_dist_1 0 --retain_header --adapter_1 AGATCGGAAGAGCACACGTCTGAACTCCAGTCA --adapter_2 AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT --renz_1 dpnII --renz_2 mspI -D --adapter_mm 1 --inline_index -q -c -r). We then used bbduk (k = 15, hdist = 1) from the BBTools suite v. 38.42 (Bushnell 2019) to filter out any adapter sequences not removed initially. Reads whose mates were removed by bbduk were added to the single read files from process_radtags.

Assembling and filtering RADseq loci

We used the Stacks pipeline to assemble reads into loci and call genotypes. We included one technical replicate in our sampling and compared the genotypes of the shared loci between these replicates to estimate error rates in genotyping using a custom R script. We then used this rate, the number of loci assembled, and mean coverage as the criteria for optimizing the -m, -M, and -n parameters of the Stacks pipeline. Final assembly parameters chosen were -m 8, -M 2, and -n 2. We made a preliminary run of the populations program with parameters --min_mac 3, -p 1, -r 0.5 and then filtered the resulting loci for paralogs and organellar loci. We identified putative paralogs among the assembled loci using ParalogFinder (--maxH 0.5 --minD -4 --maxD 4; Ortiz 2018) and organelle sequences using a BLAST search (Altschul et al. 1990) against a combined data base of the mitochondrial genome of *Cycas taitungensis* (NCBI Accession NC_010303; Chaw et al. 2008) and the plastome genome of *Dioon spinulosum* (NCBI Accession NC_027512; Wu and Chaw 2015). The populations program in Stacks was used to filter the assembled loci using a blacklist of paralogs and organellar sequences and output haplotypes with the parameters --min_mac 3, -p 6, -r 0.5 and all others at defaults. We also removed sequences containing SNPs potentially under selection using Bayescan (FDR = 0.1, pos = 0.1; Foll and Gaggiotti 2008). We checked for post burn-in stationarity of the Bayescan run using Tracer v. 1.7.1 (Rambaut et al. 2018). Finally, we removed loci present in < 70% of individuals and individuals with < 70% of loci sequenced as well as 2 loci that were monomorphic after other filters were applied. We used the populations program in Stacks to construct two data sets: The first included only populations of *D. holmgrenii* and consisted of haplotypes from each locus. The second data set was made of single SNPs from each locus and included the *D. planifolium* outgroup population. This second set was necessary for a population expansion test that requires polarized SNP data (see below).

Genetic structure

We performed two analyses of our haplotype data to identify genetic clusters and compare these to sampled localities. First, we performed discriminant analysis using principal components (DAPC) to infer clusters of genetically similar individuals in the R package adegenet (Jombart 2008; Jombart et al. 2010; Jombart and Ahmed 2011). Allele counts were transformed to frequencies and missing data were replaced with the mean allele frequency. We inferred the number of genetic clusters using the successive k-means method implemented in the find.clusters function in adegenet, with 100 random starting centroids. The resulting groups were used in the dapc function. We used 90 principal components in this analysis as this was the optimal number given by the cross-validation procedure suggested by the authors and we retained four discriminant functions. Results were plotted using ggplot2 (Wickham 2016). To identify alleles that are more strongly correlated to each discriminant function we took advantage of the variable contributions to each discriminant function reported by dapc. Because these individual contributions sum to one, we could calculate the proportional contribution of each allele to each function and compared that to a threshold value assuming equal contributions from all alleles. Those alleles that contribute more than this threshold are considered important variables associated with a particular discriminant function.

We also used the fineRADstructure package (Malinsky et al. 2018) to estimate coancestry among haplotypes and to cluster individuals based on this estimate. The sampleLD.R script (-n 500) provided with the package was used as recommended to reorder loci based on estimated LD. We then ran the RADpainter program (ver. 0.3.2 r109) to estimate coancestry and the finestructure program on the output, with parameters -x 200,000, -y 10,000,000, and -z 4000, to cluster individuals based on this coancestry estimate. The MCMC output from four independent runs was compared to evaluate convergence using the program Tracer (Rambaut et al. 2018). We visualized the coancestry among individuals and the resulting genetic clusters using modified versions of R scripts supplied with fineRADstructure. Following the

authors' suggestions, we compared the proportion of missing data between clusters using a paired Wilcoxon rank sum test and found no significant differences.

Genetic diversity and differentiation

The following analyses used the populations/clusters inferred in the clustering analyses. We used the hierfsat R package to calculate the median observed heterozygosity (Ho), gene (haplotype) diversity (Hs), and inbreeding coefficient (F) within each population (De Meeûs and Goudet 2008; Goudet and Jombart 2018). We calculated mean allelic richness using rarefaction (A) and allelic diversity ($A_S = A-1$), using the Metapop2 program (Pérez-Figueroa et al. 2009; López-Cortegano et al. 2019). We used the poppr R package (Kamvar et al. 2014, 2015) to count private alleles in each population and a custom R script to calculate the proportion of distinct alleles in each population that are private to that population.

The diveRsity R package (Keenan et al. 2013) was used to calculate global estimates of Weir and Cockerham's (1984) F-statistics as well as pairwise estimates of Fst. We estimated two parameters of allelic differentiation among populations, Jost's D (D_{Jost}) (Jost 2008; Jost et al. 2018) using diveRsity, and the allelic differentiation coefficient (A_{ST}) from total allelic diversity (A_T), within allelic diversity (A_S), and average allelic distance between populations (D_A) (Caballero and Rodríguez-Ramilo 2010) using Metapop2. Jost's D and A_{ST} are based on allele frequency and (rarefacted) allelic richness, respectively, and are meant to capture the actual differentiation between populations in terms of differences in allelic composition, as opposed to measures of fixation such as F_{ST}. The former two are also complementary because, in Jost's D, alleles are weighted by their frequency whereas for A_{ST} they are not (Jost et al. 2018). Comparing these two statistics provides a measure of the importance of rare vs. common alleles for the differentiation between populations. The contribution of each population to overall allelic diversity and differentiation was also estimated in Metapop2 by removing each population and recalculating the statistics. Bootstrap confidence intervals of the median or mean across loci were also calculated for all statistics.

We performed an analysis of molecular variance (AMOVA) in poppr using populations as the grouping variable and removing loci with > 5% missing data. To test for significance of the variance components we used the randtest function with 999 permutations (Excoffier et al. 1992).

We used a Mantel test as implemented in ade4 v. 1.7–13 (Dray and Dufour 2007) to test for isolation by distance between populations. Great circle geographic distances between localities were based on sampling site coordinates and calculated using the distm function in the R package geosphere, while genetic distances were calculated as Cavalli-Sforza and Edwards Chord distance (Takezaki and Nei 1996) in hierfstat. We used 9999 permutations to assess significance.

Population demographic history

We tested for deviations from Hardy–Weinberg equilibrium (HWE) at each locus in the species as a whole and within populations using Fisher's exact test implemented in pegas (Paradis 2010) and tested for significance following Waples (2015).

Estimates of recent migration rates and population origins of individuals were calculated in BayesAss v. 3.0.4 (Wilson and Rannala 2003) using modifications by Musselman et al. (2019) for use with RADseq data, including their method for choosing optimal mixing parameters. We performed four independent runs of 3E8 iterations each, sampling every 10,000 steps and discarding a burn-in period of 1E7 steps. Output of parameters from all four runs were compared using Tracer (ver. 1.7.1; Rambaut et al. 2018) to confirm convergence and stationarity.

We tested for recent reductions in effective population sizes with the program Bottleneck ver. 1.2.02. We ran 1000 replicates assuming the infinite alleles model and used both the sign test and standardized difference test to assess significance. Population bottlenecks can be the result of a reduction in population size in situ or of founder events by dispersal and range expansion. We tested for a history of range expansion via founder events using our SNP data set in the R package rangeExpansion (Peter and Slatkin 2013; 2015). This method estimates a directionality index, Ψ, which is the difference in average frequency of derived alleles between two populations and is expected to be positive in the direction of an expansion. To assess the significance of the directionality index for each pairwise comparison we used a permutation test in which we randomly assigned empirical allele frequencies to populations and then recalculated Ψ. Ten thousand permutations were run, and the empirical cumulative distribution of the results was used to calculate a p-value for the original estimates.

Morphological diversity and divergence

In order to assess the level of morphological divergence between populations we measured twelve morphological characters in 20 individuals at each locality (Table S1). We first removed outliers by comparing individual Mahalanobis distances to the critical value of a chi-squared distribution with alpha = 0.05. We again used DAPC to cluster individuals, this time based on the multivariate morphological data. We clustered individuals into groups inferred from the data using the k-means method find.clusters implemented in adegenet and, alternatively, into populations defined by our

genetic clustering results. Eleven principal components were used following the results of the cross-validation procedure and we retained two discriminant functions to cluster individuals. We used the same 'equal contribution' procedure described above to determine 'important' variables for each discriminant function.

The DAPC procedure also calculates the posterior probability of assignment of individuals to groups based on the discriminant functions. We used the proportion of individuals assigned (based on the discriminant functions of the morphological data) to their predefined populations (based on genetic data and locality) as a measure of population distinction. That is, if populations defined by genetic clustering and locality are morphologically distinct then we would expect individuals to be placed in these same clusters with high posterior probability based on morphological data as well. If not, then there is low support for a correspondence between genetic and morphological divergence or structure.

We also tested for significant differences in morphological traits between populations with a permutational MANOVA using the adonis2 function in the R package vegan (ver. 2.5–6; Anderson 2001; Oksanen et al. 2019). We used the function vegdist to create a Euclidean distance matrix among samples. Populations from the genetic clustering were used as groups and 9999 permutations were run to assess significance. We also ran pairwise permutational MANOVAs between all populations using the R package pairwiseAdonis, a wrapper for the adonis function that includes a correction for multiple tests (Arbizu 2017).

Results

Sequencing and locus assembly

After removing 1994 putative paralogs, 596 organellar sequences, and six loci potentially under selection, our sequencing and filtering pipeline resulted in a final haplotype data set of 2314 loci across 151 individuals (18–36 individuals per population) and a SNP data set with 2056 loci and 218 individuals (27–43 per population). Detailed sequencing and filtering information is given in Tables S2–S3. Sequences will be deposited in a public archive upon manuscript acceptance.

Population genetic structure, diversity and differentiation

Plotting the BIC for each chosen number of clusters (k) indicates that there are six genetic clusters among our samples (Fig. 2). We used the first four discriminant functions to cluster individuals using dapc. The proportion of alleles with outsized contributions to these functions was 0.25, 0.24, 0.24, and 0.23, respectively, which we interpret as a relatively large signal across the genome as opposed to very few loci contributing to the analysis. The first four discriminant functions of the DAPC analysis tightly cluster individuals from the same locality with no overlap among clusters (Fig. 3). The two eastern populations in our sample, Rancho

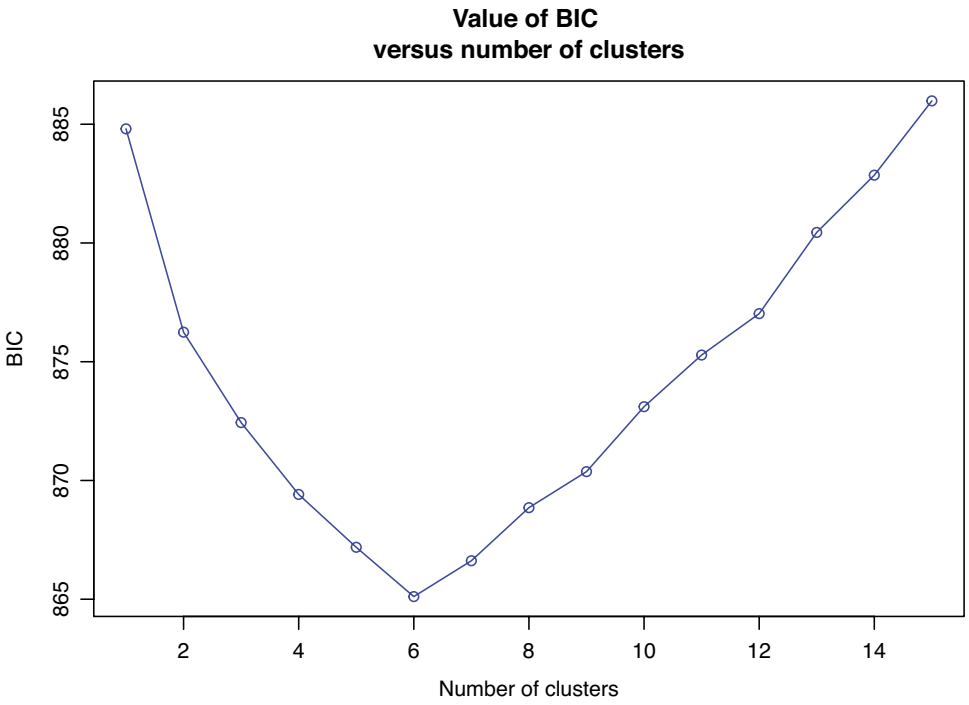

Fig. 2 Bayesian information criterion for varying numbers of clusters of individuals from find.clusters function based on genomic data

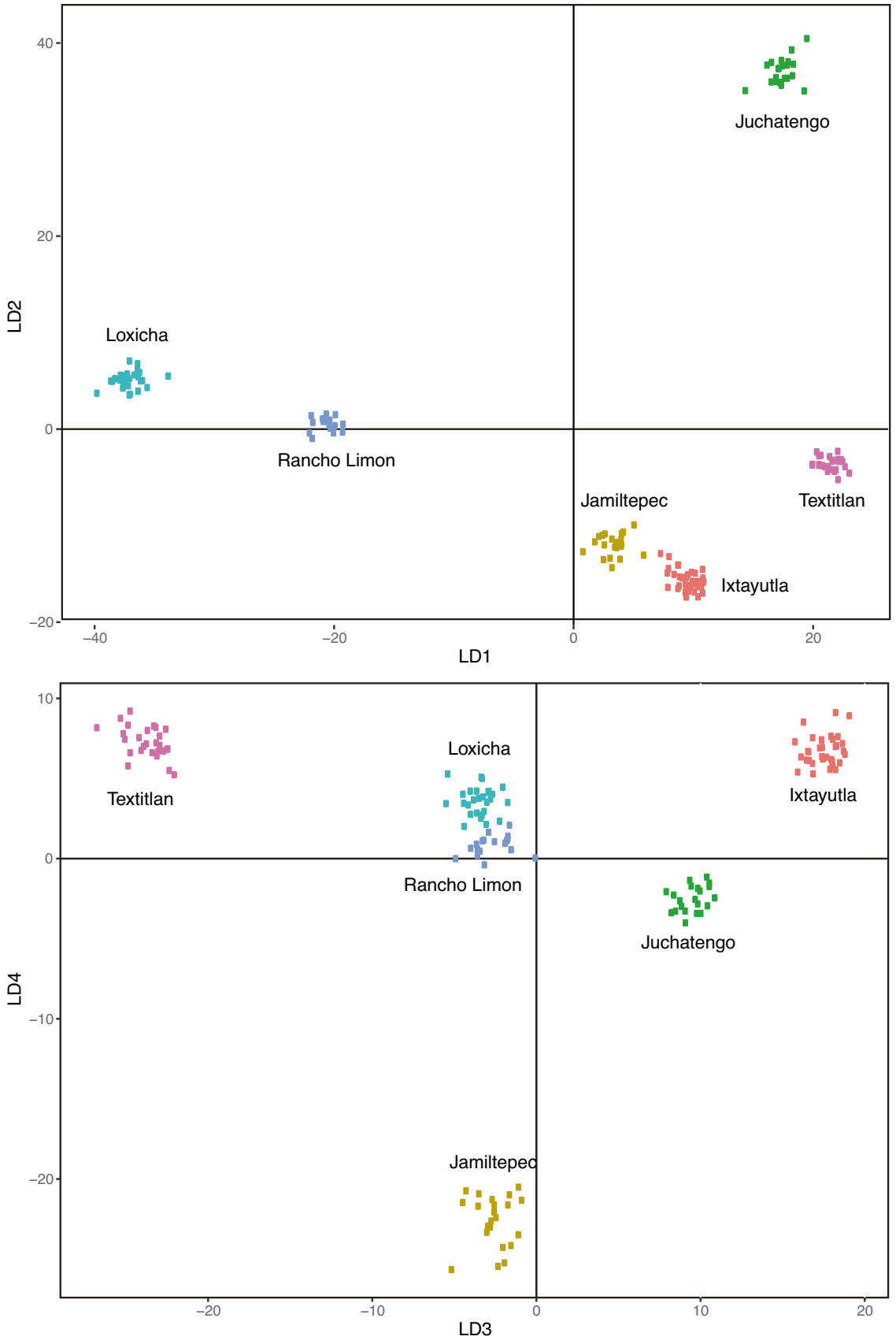

Fig. 3 Scatterplots of four discriminant functions showing genomic clusters of individuals corresponding to sampling localities

Limón and Loxicha, are relatively close along axes 3 and 4 but do not overlap.

All four independent runs of the fineRADstructure analysis converged on the same posterior probability distribution. The clustering of individuals based on inferred coancestry recovers the same sampled populations as the DAPC analysis indicating higher levels of coancestry within than between populations (Fig. 4). Other patterns also emerge from Fig. 4. Two main groups can be delineated with Rancho Limón and Loxicha populations in one and the rest of the populations in the other. A shared history between these populations is not surprising given their proximity but we also see that Jamiltepec, a population that is geographically quite removed from the former two, has an intermediate amount of coancestry with them. This is also reflected in the positions of these three populations along the third discriminant function in Fig. 3. Levels of coancestry vary within the populations as well. In two of the populations, Juchatengo and Textitlán, we see evidence of substructure indicated by higher shared coancestry among clusters of individuals within each population. The remaining four populations have a more even distribution of coancestry among all individuals, for the most part.

The mantel test for isolation by distance was not significant ($P = 0.08$). Indeed, the two greatest genetic distances were between Juchatengo and each of Rancho Limón and Loxicha, three species that are relatively close in geographic distance. We note, however, that in this case the barrier to gene flow may, in fact, be the Sierra del Sur mountain range rather than linear distance.

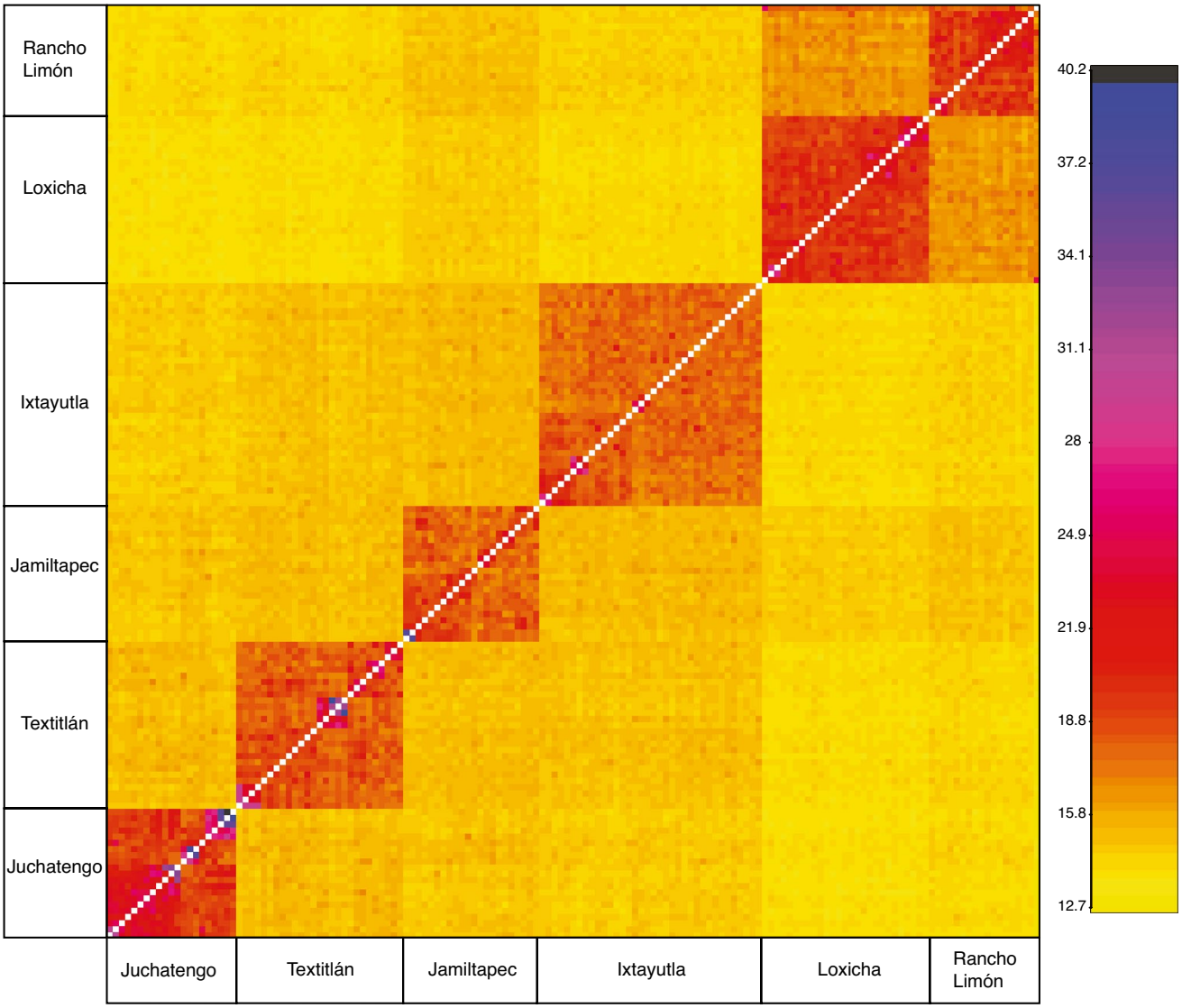

Fig. 4 Heatmap showing the coancestry of individuals across six populations of *D. holmgrenii* calculated from 2314 loci using fineRADstructure. Colors indicate the sum of nearest-neighbor alleles shared between individuals

Taking the populations inferred from our two clustering procedures, we calculated several summary statistics of genetic diversity and differentiation. Within population diversity is generally moderate (Tables 1, 2). Allelic richness is similar across populations (mean = 2.11–2.39) but the percent private alleles is more dissimilar, although substantial in all, ranging from 3 to 7%. Median observed heterozygosity (H_O; range 0.23–0.27) and median gene diversity (H_S; range 0.25–0.31) are similar to each other and fairly similar across populations. The median inbreeding coefficient (F) across loci is close to zero (− 0.03–0.0) in all populations, providing little evidence of inbreeding. When possible, we present the distribution of these statistics across loci (Figs. S1–S4).

Components of allelic frequency variance and allelic composition are given in Table 2. Estimates of global F-statistics suggest moderate inbreeding in the species overall ($F_{it} = 0.17$) but this is not concentrated within populations ($F_{IS} = 0.036$). The global fixation index (F_{ST}) indicates moderate fixation among populations ($F_{ST} = 0.139$) and pairwise F_{ST} values range from 0.10 to 0.20 (Table 3). Population differentiation based on allelic composition is moderate according to our estimates of A_{ST} (0.239) but much less when measure with Jost's D (0.039; Table 2) and pairwise estimates range from 0.19–0.32 to 0.015–0.039, respectively (Table 4). Individual population contributions to overall allelic diversity vary widely (Table 5). Ixtayutla, Textitlan and, to a lesser degree, Juchatengo, each increase overall allelic diversity whereas the other three each reduce it when included in the analysis. The AMOVA apportioned 86.27% (P = 0.001) of the genetic variation within individuals and 13.69% (P = 0.001) among populations while the within population variation was not

Table 1 Population specific estimates of genetic diversity

Population	N	% Private Alleles	Median [95% BS CI]			
			A	F	H_O	H_S
Ixtayutla	33.3	5.4	2.32 [2.29,2.36]	0.00 [− 0.001,0.001]	0.25 [0.23,0.26]	0.28 [0.26,0.31]
Jamiltapec	20.4	2.7	2.29 [2.25,2.32]	− 0.024 [− 0.034,− 0.019]	0.27 [0.24,0.29]	0.28 [0.26,0.30]
Juchatengo	19.1	6.9	2.37 [2.33,2.41]	− 0.026 [− 0.028,-0.023]	0.29 [0.27,0.31]	0.31 [0.29,0.34]
Loxicha	25.6	3.2	2.11 [2.08,2.15]	0.00 [− 0.004,0.01]	0.23 [0.22,0.25]	0.25 [0.23,0.28]
Rancho Limón	16.7	2.1	2.25 [2.22,2.28]	− 0.030 [− 0.043,− 0.023]	0.28 [0.26,0.30]	0.29 [0.27,0.30]
Textitlán	24.9	6.0	2.39 [2.34,2.42]	− 0.006 [− 0.017,0.006]	0.27 [0.24,0.29]	0.31 [0.29,0.32]
Mean	23.3	4	2.29	− 0.01	0.27	0.29

N mean number of individuals/locus, *A* allelic richness using rarefaction, *F* inbreeding coefficient, H_O Observed heterozygosity, H_S Gene diversity, *BS CI* bootstrap confidence interval of the estimates

Table 2 Global estimates and 95% bootstrap confidence intervals (CI) of Wier and Cockerham's F-statistics (F_{IS}, F_{IT}, F_{ST}) and allelic diversity and differentiation statistics for *Dioon holmgrenii*

	F_{IS}	F_{IT}	F_{ST}	A_S	D_A	A_T	A_{ST}	D_{Jost}
Estimate	0.036	0.170	0.139	1.290	0.407	1.697	0.239	0.039
95% CI	[0.031,0.042]	[0.164,0.178]	[0.135,0.142]	[1.257,1.318]	[0.396,0.417]	[1.655,1.735]	[0.237,0.242]	[0.036,0.042]

A_S allelic diversity within populations, D_A allelic distance between populations, A_T total allelic diversity, A_{ST} allelic differentiation, D_{Jost} Jost's D, *CI* 95% bootstrap confidence intervals

Table 3 Pairwise estimates of F_{ST} among six populations of *Dioon holmgrenii* (below the diagonal)

	Ixtayutla	Jamiltapec	Juchatengo	Loxicha	Rancho limón	Textitlán
Ixtayutla		[0.09,0.11]	[0.13,0.15]	[0.17,0.18]	[0.13,0.15]	[0.10,0.11]
Jamiltapec	0.10		[0.13,0.14]	[0.15,0.16]	[0.11,0.12]	[0.09,0.11]
Juchatengo	0.14	0.13		[0.19,0.20]	[0.16,0.17]	[0.11,0.12]
Loxicha	0.17	0.16	0.20		[0.11,0.12]	[0.16,0.18]
Rancho Limón	0.14	0.12	0.17	0.12		[0.13,0.15]
Textitlán	0.11	0.10	0.12	0.17	0.14	

Bootstrap CI are above the diagonal

Table 4 Pairwise estimates of allelic differentiation between six populations of *Dioon holmgrenii*

	Ixtayutla	Jamiltapec	Juchatengo	Loxicha	Rancho limón	Textitlán
Ixtayutla		0.22 [0.21,0.22]	0.26 [0.26,0.27]	0.31 [0.30,0.31]	0.28 [0.27,0.29]	0.24 [0.24,0.25]
Jamiltapec	0.015 [0.013,0.017]		0.25 [0.24,0.25]	0.29 [0.29,0.30]	0.27 [0.26,0.27]	0.24 [0.23,0.24]
Juchatengo	0.026 [0.023,0.028]	0.025 [0.023,0.028]		0.32 [0.31,0.33]	0.30 [0.30,0.31]	0.25 [0.24,0.26]
Loxicha	0.033 [0.030,0.036]	0.026 [0.023,0.029]	0.039 [0.036,0.043]		0.19 [0.18,0.19]	0.32 [0.31,0.32]
Rancho Limón	0.025 [0.022,0.027]	0.018 [0.016,0.021]	0.032 [0.029,0.036]	0.016 [0.014,0.018]		0.29 [0.29,0.30]
Textitlán	0.017 [0.016,0.019]	0.016 [0.014,0.017]	0.022 [0.020,0.024]	0.033 [0.030,0.037]	0.026 [0.023,0.028]	

A_{ST} above the diagonal and D_{Jost} below. Brackets contain 95% bootstrap confidence intervals of the estimates

Table 5 Contribution of each population to allelic diversity within populations (A_S), between populations (D_A) and total (A_T) estimated by removing one population and recalculating the statistics

Population	Percent contribution		
	A_S	D_A	A_T
Ixtayutla	1.3814	0.641	2.0223
Jamiltepec	− 0.1908	− 0.0331	− 0.2239
Juchatengo	0.5363	1.5115	2.0479
Loxicha	− 1.9202	1.4349	− 0.4853
Rancho Limón	− 1.1636	0.8041	− 0.3595
Textitlan	1.3568	1.0831	2.4399

significantly different from random (P = 0.5). The Phi-stats were comparable to their F-stat analogs (Table 6).

Population Demographic History

Tests of HW proportions showed that the species taken as a whole is not in HW equilibrium (HWE) with 992 loci (42%) significantly out of HWE, well above the critical value of 133 loci calculated from the binomial distribution. Two individual populations also had more than the critical number of loci significantly out of HWE, Ixtayutla (163 loci; 7%), and Textitlán (162 loci, 7%).

The median estimated proportion of migrants in each population was 0.01 for most population pairs and the 95% HDP of all pairs included zero, indicating very little to no migration in the last few generations (Table S4). One possible exception is between Rancho Limón and Loxicha where the median proportion is 0.03 (95% HPD = [0.00, 0.06]). Indeed, the only individual inferred as a migrant occurred in Rancho Limón but shared most of its coancestry with individuals of Loxicha (data not shown).

The bottleneck test was highly significant ($P < 10^{-5}$) for all populations indicating a reduction in effective population size in the relatively recent past (i.e., 2–3Ne generations; Table S5). The rangeExpansion analysis resulted in six significant Ψ values (P < 0.05; Fig. 5) and the origin of expansion was inferred to be in the southeast of the range, near Loxicha (data not shown). Most significant signals of

Table 6 Results of the analysis of molecular variance among populations of *Dioon holmgrenii*

AMOVA results	Df	Sum Sq	Mean Sq
Between population	5	12333.34	2466.67
Between samples within population	145	40196.06	277.21
Within samples	151	41822.62	276.97
Total	301	94352.02	313.46
Components of covariance	Sigma	%	P-value
Variation between populations	43.97	13.69	0.001
Variation between samples within populations	0.12	0.04	0.489
Variation within samples	276.97	86.27	0.001
Total variation	321.06	100	
Phi-stats			Phi
Phi-samples-total (Phi_{IT})			0.1373
Phi-samples-population (Phi_{IS})			0.0004
Phi-population-total (Phi_{ST})			0.1369

Loci with > 5% missing data were removed. Significance was assessed using a permutation test with 999 replicates

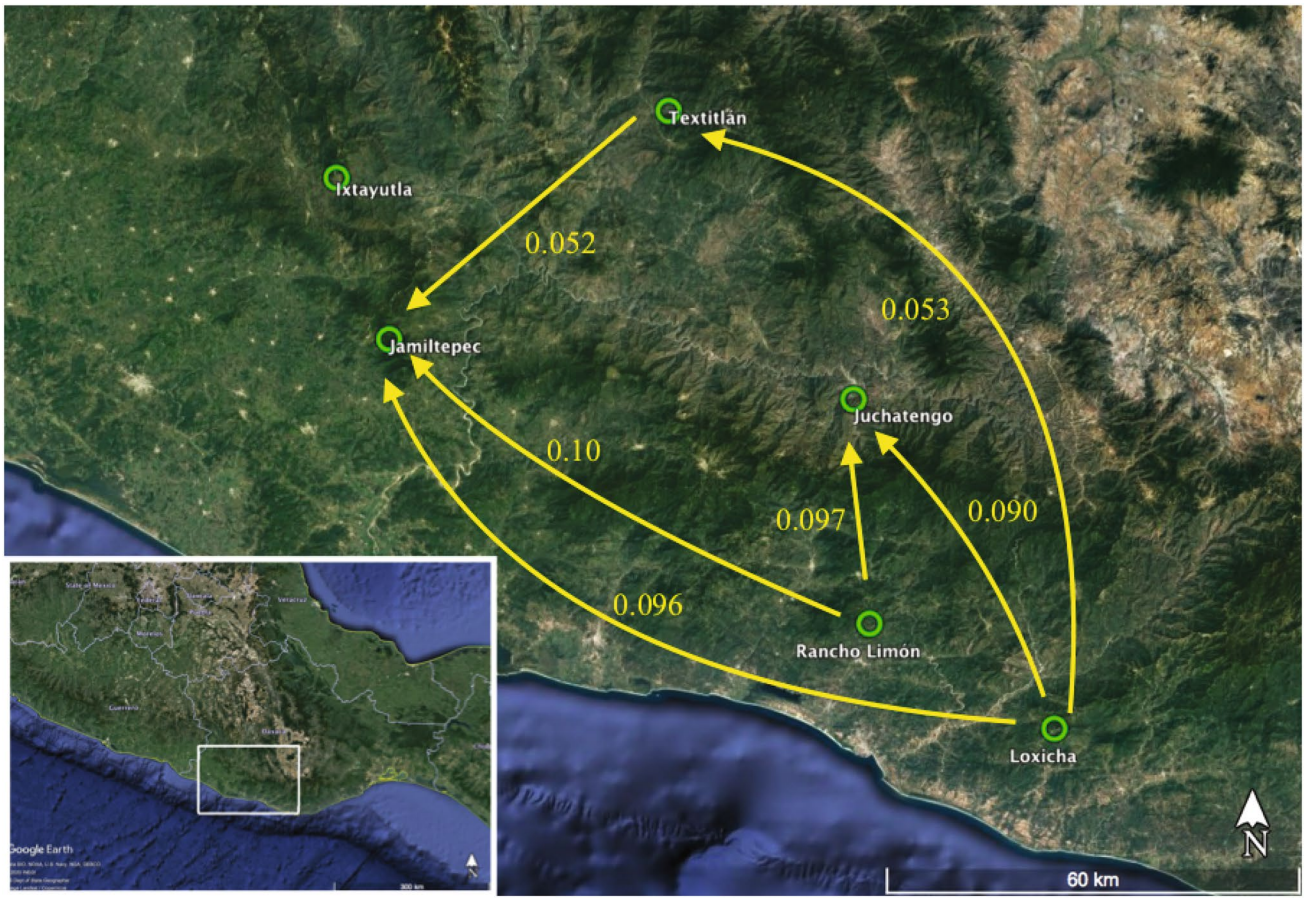

Fig. 5 Map of *D. holmgrenii* distribution showing significant directionality indices, Ψ, between populations. Arrows point to the population within a pair furthest from the inferred ancestral geographic source, as indicated by a positive Ψ. Overall, the population has expanded from southeast to northwest. Significance was determined with 10,000 permutations (see text). Note that the arrows do not necessarily represent direct dispersal events but rather the relative distance from the ancestral source

expansion are between either Rancho Limón or Loxicha and populations to the north and west, with one exception going from Textitlan to Jamiltepec. Interestingly, there is no significant evidence for expansion to or from Ixtayutla.

Morphology

The optimal number of clusters inferred using the k-means method was not clear from the BIC scores (Fig. 6). We chose eight clusters to use in the DAPC procedure because the BIC score more or less plateaued after this number. We also ran the DAPC procedure multiple times to gauge the consistency of the grouping and these runs produced multiple distinct scatterplots (data not shown). One solution is shown in Fig. 7a. Most of the eight groups had clearly separated centroids but nearly all overlap another cluster. Figure 7b shows the prior group assignments from the k-means method and the posterior probabilities of those assignments from the DAPC. Each of these prior groups consists of individuals from multiple populations and every population is split among multiple groups.

In the DAPC analysis using the populations defined by our haplotype data (and locality) as a priori groups, the first two discriminant functions do not clearly separate these populations, in contrast to the haplotype data (Fig. 8). With the exception of Ixtayutla, there is substantial overlap among all populations. The contributions of the original variables to the discriminant functions are given in Table S6 and indicate that the first two axes correlate to overall leaf and plant size. The overall proportion of individuals assigned a posteriori to their original localities was 0.68 (range across localities = 0.50–0.88; Table S7) and many individuals could not be assigned to their original population, nor any other, with high posterior probability. This pattern is clear in the assignment plot (Fig. 7b) where Rancho Limón, Loxicha, and Textitlán all have individuals assigned to multiple populations with relatively equal posterior probability.

Despite the overall ambiguous assignment of individuals to populations, the permutational MANOVA results indicate

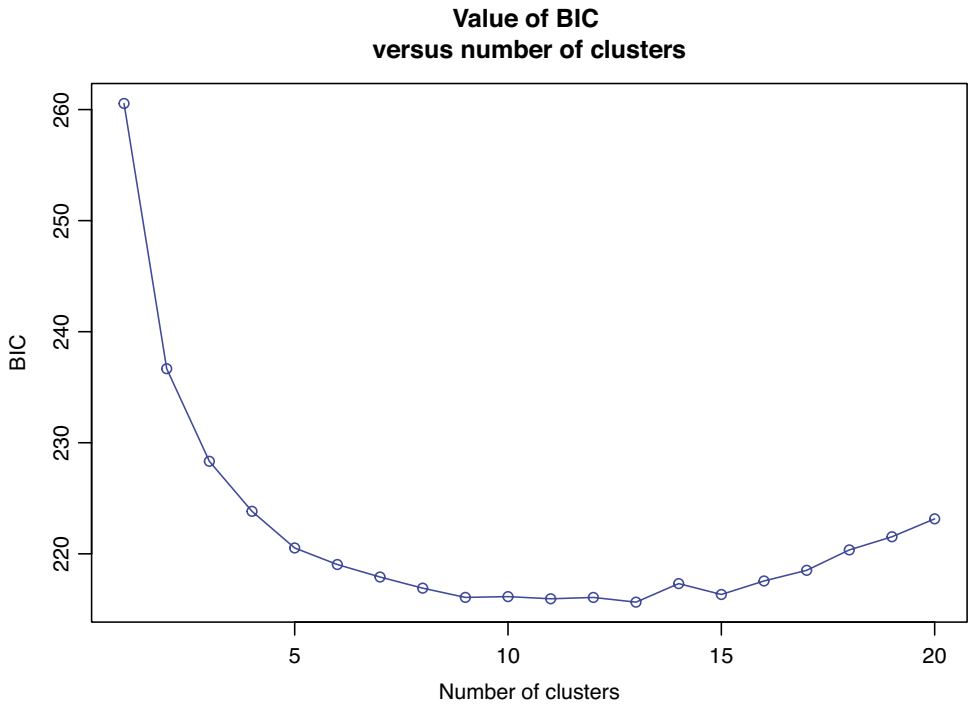

Fig. 6 Bayesian information criterion (BIC) for varying numbers of clusters of individuals based on morphological traits

significant differences among populations in morphological space (Table S8). However, the pairwise MANOVA was significant in only 8 of 15 comparisons and 5 of those included Ixtayutla (Table S9).

Discussion

Cycads face multiple threats to their survival, including habitat loss and illegal collecting, that have led to a majority of species facing extinction (IUCN 2021). Specifically, 68% of 337 species are listed as Vulnerable to Extinction or worse, with 56 of these Critically Endangered and 5 Extinct in the Wild. To prevent further population declines and ultimately reverse this trend many species will require human interventions, such as reserve creation, altered land use practices, and/or the development of ex situ assurance colonies (Donaldson 2003). But with limited resources for conservation, accurate and up to date assessments of the threats to species survival are critical for prioritizing the allocation of these resources and informing strategic planning. These priorities and potential interventions require an understanding of the distribution of standing genetic diversity to be fully effective (Hoban et al. 2022). Stands of *D. holmgrenii* occur in a patchy distribution of at least 10 known localities in the Sierra del Sur and the Pacific Coastal Plain of southwestern Oaxaca (Velasco-García et al. 2016). This is a much greater number of localities and extent of occurrence than is reported in the last Red List assessment and previous population genetic work has focused on only sections of the distribution. We sampled from localities across the entire known range to characterize the genetic structure and diversity of *D. holmgrenii*, and test for differences in morphology among populations. Major findings of our work are that the species comprises at least six distinct, isolated populations; that these populations are the result of an historic range expansion from an ancestral population, most likely in the southeast of the current range; and that moderate allelic diversity, driven primarily by rare and private allele frequencies, exists within and among these populations. These results are here discussed in the context of previous work in *Dioon* to evaluate the taxonomic status and genomic cohesion of the species in order to inform any future conservation efforts. We also present a potential phylogeographic history for the species that is consistent with the current study and previous work on divergence in the genus.

Fragmented populations with distinct alleles

Our results suggest that the sampled localities of *D. holmgrenii* represent cohesive and relatively distinct genomic clusters. Two separate genomic clustering methods using loci from across the genome clearly separate each locality into a distinct and separate genomic cluster. In terms of hierarchical structure, it appears that Rancho Limón and Loxicha share more coancestry than any other two populations and accordingly occupy the closest space in the DAPC biplots (Figs. 3, 4). This is not surprising given their close proximity and the fact that there are potentially connecting populations between them (Velasco-García et al. 2016).

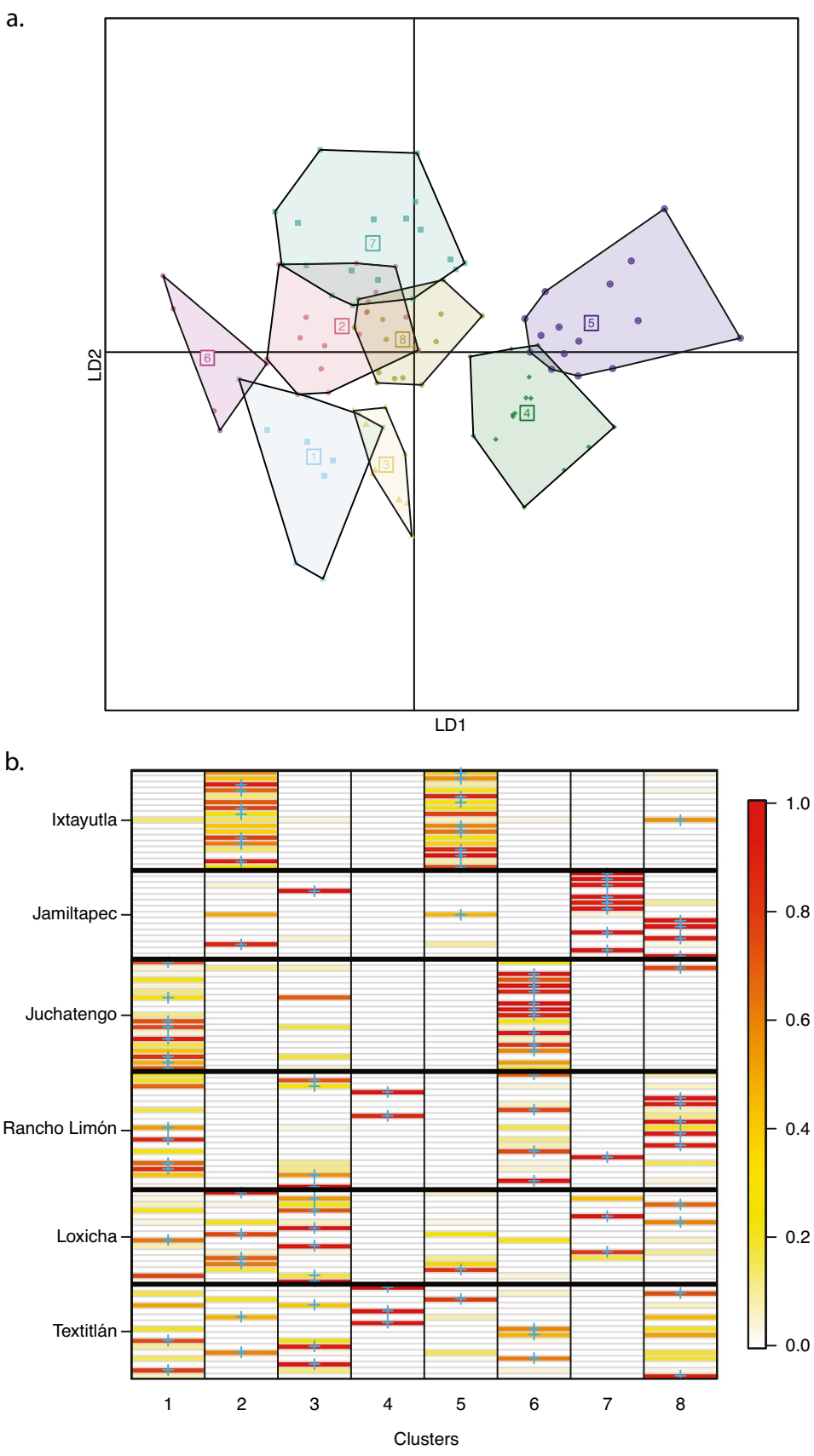

Fig. 7 One example result from the DAPC analysis of morphological traits using inferred groups. **a** Scatterplot of individuals clustered de novo based on groups inferred using the k-means method. Polygons are convex hulls of each cluster and boxed numbers are cluster centroids. **b** Assignment plot showing posterior probabilities (PP) for each possible individual assignment. Samples (rows) are arranged along the y-axis according to their collection localities and clusters are given along the x-axis. Blue crosses indicate the inferred group assignment (highest PP) and cell colors indicate the posterior probability of assignment to a given cluster

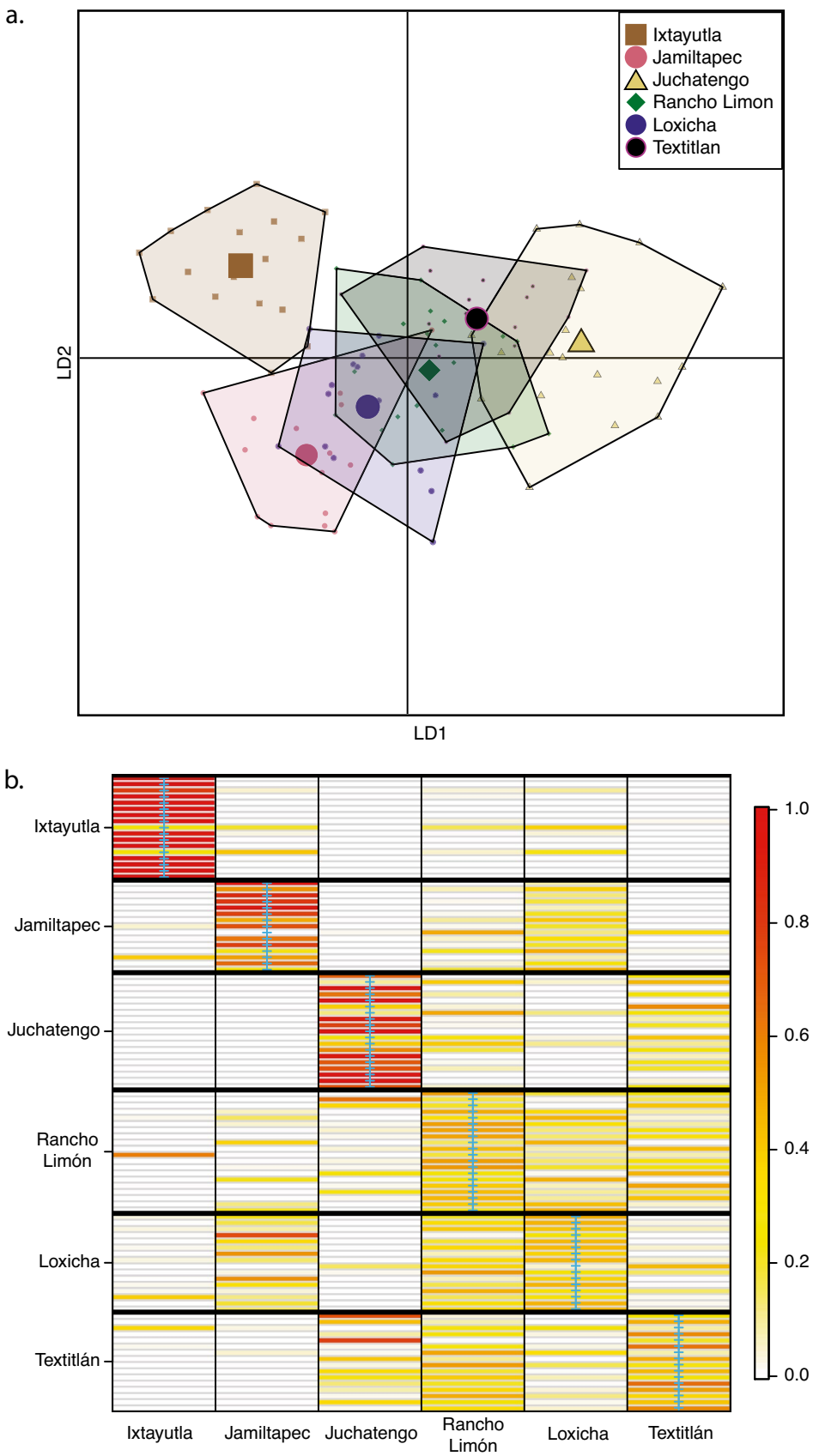

Fig. 8 Results from the DAPC analysis of morphological traits using genetic clusters/localities as a priori groups. **a** Scatter plot showing clusters with convex hulls and their centroids (large symbols). **b** Assignment plot showing posterior probabilities for each possible individual assignment. Symbols as in Fig. 7

Velasco-Garcia et al. (2021) identified what appear to be distinct but geographically close populations between our sampled localities of Rancho Limón and Loxicha. These could potentially form a conduit of gene flow between Rancho Limón and Loxicha but our migration results do not bear that out. Our bootstrapped BayesAss analysis of migrant proportions could not reject zero migrants between these two localities. Further, Velasco-Garcia et al. (2021), using six microsatellites, suggested that this set of small populations is divided between two distinct genetic clusters, one of which was more allied with a population tens of kilometers to the west rather than with its closest neighbors. We suggest that the greater genomic coverage of our data set has provided a clearer signal of coancestry and genetic structure and that the higher level of coancestry between these two populations is potentially a remnant of previous connectivity that has been reduced due to population contractions.

Results from the AMOVA reveal that most of the variance in allele frequencies is within individuals and among populations, rather than between individuals within populations (Table 6). This lack of divergence among individuals is consistent with our estimates of F (Table 1) that essentially show random mating within populations, which follows from cycads being dioecious and entomophilous. This contrasts with an earlier study by González-Astorga et al. (2008a, b) that included only the Rancho Limón and Loxicha populations. Using allozymes they found an excess of heterozygotes within each of these two populations and in total ($F_{IS} = -0.201$, $F_{IT} = -0.116$). (Heterozygote excess has also been found in *D. caputoi*, again using allozyme data (Cabrera-Toledo et al. 2008).) Our genome-wide RADseq haplotype data indicate just the opposite pattern in *D. holmgrenii*. Median population-specific estimates of inbreeding for all populations are essentially zero while species-wide estimates also indicate little local inbreeding but substantial global inbreeding ($F_{IS} = 0.036$ and $F_{IT} = 0.17$), likely a result of lack of gene flow among populations. The global estimate of F_{ST} is 0.139 indicating moderate levels of allelic frequency variation between populations relative to the total, consistent with the AMOVA results.

We also find comparable differentiation between populations based on allelic composition (i.e., average number of different alleles per locus within and between populations). The total allelic diversity (i.e., average pairwise allelic diversity,) is moderate ($A_T = 1.697$) and much of it is found within populations ($A_S = 1.29$). Population differentiation based on allelic distance (average number of pairwise private alleles) is somewhat low ($D_A = 0.4$), but this between population diversity represents a fair proportion of the total ($A_{ST} = 0.239$; Table 2). Our population differentiation estimate that considers the frequency as well as composition of alleles is much lower ($D_{Jost} = 0.039$). Because this statistic primarily measures the differentiation among the most common alleles (Jost et al. 2018), we can conclude that the differentiation we do see among populations (measured by D_A and A_{ST}) is due to the presence/absence of rare alleles. This is also supported by our estimates of the percentage of alleles that are private to each population (ca. 2–7%).

Dioon holmgrenii populations harbor a substantial, but not excessive, level of standing genetic variation, as measured by observed heterozygosity and gene diversity, and this is fairly even across populations (Table 1). Several other studies using allozyme data have estimated genetic diversity in *D. holmgrenii* and other *Dioon* species (González-Astorga et al. 2003, 2005; González-Astorga et al. 2008a, b; Cabrera-Toledo et al. 2008, 2010; González-Astorga et al. 2008a, b). In contrast to studies in *D. merolae* and *D. caputoi* (Cabrera-Toledo et al. 2008, 2010), we find no evidence of heterozygote excess in *D. holmgrenii*. Gonzalez-Astorga et al. (2008a, b) reported somewhat lower gene diversity in the Loxicha and Rancho Limón populations ($H_S = 0.14$ and 0.19, respectively), relative to our estimates. But we find that the mean gene diversity among *D. holmgrenii* populations using RAD loci (mean $H_S = 0.26$) is somewhat lower than the mean of previously reported values across all *Dioon* species using allozymes (mean $H_S = 0.32$). Gutiérrez et al. (2018a, b) also estimated much lower gene diversity in *D. sonorense* using SNPs vs. previous estimates using allozymes (mean He = 0.082 vs. 0.341, respectively; Jorge González-Astorga et al. 2008a, b).

We note that central tendency summary statistics do not inform us about the variation across the genome, which can help explain the discrepancy between previous studies and the present one. Among our RAD loci the F statistic has a strong central tendency around the median value, but H_O and H_S have much broader distributions with a large number of loci having estimates of 0 to 0.05 (Figs. S2–S4). Importantly, these results show how easy it would be to estimate quite different statistics from a small set of loci, depending on where that set fell in the distribution. It is a clear advantage of reduced representation genomic techniques that they allow us to view the range of values for summary statistics across loci. But, this also highlights the limits of these single summary statistics to capture this variation.

Summarizing these results, we find that much of the standing genetic diversity within the species resides within individuals and that despite the fragmented nature of the populations they are only moderately diverged. However, an important component of genetic diversity is held between populations. This component consists of the substantial proportion of uncommon and private alleles. To the extent that alleles that are rare across the species or unique to populations are the target of conservation (i.e., may be important for future adaptation), this particular partitioning of allelic diversity justifies providing appropriate protection to each population of *D. holmgrenii*.

Taxonomic status of D. holmgrenii

While divergence among populations of *D. holmgrenii* is moderate, the isolation illustrated by our clustering, migration rate, and bottleneck analyses begs the question of whether all the populations in southwestern Oaxaca belong to the same taxon, which is important for conservation of the species. Along the first two axes of the DAPC plot (Fig. 3), one could potentially interpret three population groupings based on geography: northeastern, southwestern and Juchatengo alone. As mentioned above, the Rancho Limón and Loxicha populations are relatively closely positioned along all four axes of the DAPC plot and clearly have a higher level of coancestry (Figs. 3 and 4). However, the two other potential groups do not hold up along the third and fourth DAPC suggesting that the overall genetic variation is not hierarchically structured. Patterns of genetic structure in recently described species of *Dioon* have shown clear breaks between groups of populations and/or more substantial differentiation than we observe in *D. holmgrenii* (Gutiérrez-Ortega et al. 2018b, 2020).

Leaf traits are commonly used to delimit species in *Dioon* (de Luca et al. 1981; Gregory et al. 2003; Salas-Morales et al. 2016; Gutiérrez-Ortega et al. 2018a). The morphological traits that we compared among individuals and populations do not provide evidence of strong population differentiation, with one possible exception. This is clear from the DAPC analysis of morphological data (Figs.7, 8). Using the k-means method to define groups does not clearly result in an optimal number of groups nor do the groups chosen correspond to populations (Fig. 7a). Further, these groups are not clearly separated by the DAPC (Fig. 7b). Using populations (genomic clusters) as a priori groups we find general overlap among populations and relatively low posterior probabilities of assignment to population based on morphology indicating that most populations are not distinct nor homogeneous morphologically (Fig. 8). Ixtayutla is the only population that is potentially differentiated from the rest. Indeed, the permutational MANOVA found significant pairwise differences between Ixtayutla and each of the other populations. This is consistent with our initial field observations that suggested to us this population might represent a distinct taxon. However, because none of our genetic analyses indicate greater genetic divergence between this population and the others, we must conclude that the Ixtayutla population is part of the *D. holmgrenii* species. *Dioon holmgrenii* has been described anecdotally as a variable species and this is what we think our morphological data show. This variability may be intrinsic to the species but may also be the result of human activity and/or habitat quality (Lopez-Gallego 2008; Velasco-García et al. 2016). Garcia et al. (2016) found significant differences in leaf and stem traits of *D. holmgrenii* individuals between sites with different levels of human disturbance. Plants at Ixtayutla tend to have smaller leaves, according to our DAPC results but larger trunks than other populations, such that the leaf size is not due to smaller plant size. The Ixtayutla population occurs in an area of disturbance including conversion to pastures and burning for planting (G. Juárez pers. obs.). García et al. (2016) note that in areas of cattle grazing leaves and even stem apices are removed to prevent injury to the cattle. This, along with the open and potentially less humid environment of this population, may contribute to the differential leaf morphology. Taken together, we see no patterns in our haplotype data nor in our morphological that would suggest splitting *D. holmgrenii* into multiple species.

Expansion and fragmentation led to current phylogeography

Given that the known stands of *D. holmgrenii* most likely comprise a single species, currently distributed in multiple genetically and geographically distinct populations, the question becomes: how did the current distribution arise? Combining the results discussed in the previous sections with those of our historical and current migration analyses suggests to us a history of expansion coupled with, or followed by, fragmentation. Our estimates of zero current or recent migration among all populations reinforce the genetic clustering results indicating all populations are currently genetically isolated. However, given the moderate Fst values and the results from the AMOVA, as well as the lack of evidence for inbreeding within populations, it is likely that this is a relatively recent situation. The lack of significant evidence of isolation by distance among populations also suggests that there has been at least some connectivity across the species' range. The entomophilous pollination system of cycads (Terry et al. 2012; Toon et al. 2020) likely allows for some distance between pollen and seed cones but what this distance is in *Dioon* species, we do not yet know. However, Valencia-Montoya et al. (2017) measured in situ flight distances of *Pharaxonotha* beetles that pollinate *Zamia incognita* and found a maximum flight distance of 22 m. This does not necessarily represent the actual maximum possible flight distance but does give a first approximation of within population pollen movement. Importantly though, the very low migration rates among populations of *D. holmgrenii* more recently, suggests that their weevil pollinators are no longer able to maintain the historic genetic connection between the current populations.

Comparable gene diversity among populations and only low to moderate genetic differentiation suggests that much of the present genetic diversity was present in a single ancestral (meta)population. A scenario of greater historical panmixia is also consistent with our Bottleneck analysis showing a recent reduction in populations sizes and with the signal of

range expansion we found. The method implemented by the Bottleneck program is meant to detect recent reductions in population sizes and not a strict bottleneck scenario where a secondary recovery of population size is assumed. As such, the significant results of this analysis suggests that every population of *D. holmgrenii* is the result of a dramatic loss of individuals (and genetic diversity) in the relatively recent past (i.e. 2Ne–4Ne generations; Piry et al. 1999), relative to an ancestral population.

This ancestral population was likely centered in the southeast of the current species' range, according to our demographic expansion analysis. The distribution of derived alleles suggests a general expansion of the species from southeast to northwest, from the Pacific Coastal plain upward into the Sierra del Sur and the valley of the Rio Verde. There exist at least two possibilities for how this scenario played out and both rely on one or more seed dispersal agents that are as yet unknown. In the first scenario, serial founder events from an ancestral population created satellite populations further from the source in a net northwesterly direction. These populations would be small, and each would necessarily experience a bottleneck. It may be that the populations we find today are the result of (some of) these dispersal events. The rate of dispersal would influence the likelihood of population establishment and consequently the density of individuals. Also, the nature of the disperser would influence the pattern of populations across the landscape. For example, roosting birds would likely disperse more seeds to fewer sites. This would fit with the serial, isolated founder events and a lack of gene flow being the norm from the beginning.

One problem with this history is that it is not likely to produce the levels of genetic diversity we observe. The derived allele proportions observed may be consistent with this scenario, but we would not expect the current level of observed heterozygosity or gene diversity in recently established founder populations, especially in the case of serial founder events. This scenario would also require the simultaneous establishment of multiple founder populations in the relatively recent past to account for the concurrent bottlenecks we detected in all populations. Finally, this scenario does not account for the fate of the ancestral population in the southeast of the distribution. Where has it gone, or if it is represented by one of the current populations, what caused its bottleneck if not a founder event?

The second possibility is almost the converse of the first. In this scenario there is a wave front of cycads spreading west and north from an ancestral site in the southeast, in a more even and connected pattern across the landscape, resulting in a widespread panmictic (meta) population. The evenness of this distribution would, of course, depend on habitat availability but gene flow would mostly be restricted only by distance and/or environmental limitations for pollinators and dispersers. A necessary second step to arrive at the present situation would be a contraction of the species and survival in isolated mini-refugia that would become the current populations, or their progenitors.

This history fits our results better regarding the points raised above. It is consistent with our range expansion results because derived alleles are expected to occur at the edge of an expanding population, whether the mechanism is long distance dispersal or simply population growth. Further, populations that result from fragmentation of larger population would more likely retain ancestral rates of heterozygosity but would still potentially harbor private or rare alleles that arose since or just before isolation. The predominance of rare alleles contributing to the population differentiation we observe supports this view as well. Lastly, this history is also consistent with our estimates of strong, simultaneous bottlenecks in all current populations as they are the result of the same process during the same time period.

Dioon holmgrenii belongs to one of four main clades in the genus, whose members occur from central and southern Oaxaca to western Chiapas. Four species in this clade occur in the Tehuacán Valley region and together occupy and area comparable to that of *D. holmgrenii*. Previous work has provided support for a history of diversification in *Dioon* due to shifting habitats, whereby potentially extensive populations were fragmented, eventually leading to speciation (Dorsey et al. 2018; but see also Gutiérrez-Ortega et al. 2018a, b). This process coincided with, and was potentially the result of, glacial dynamics of the Pleistocene causing upward habitat shifts and elevational tracking of species, which led to fragmentation due to topology. Depending on the circumstances, this process could have formed new species if fragmentation resulted in isolation. So, why do we find multiple species in central Oaxaca while *D. holmgrenii* seems to be a well-defined, if more widespread, species? One possibility is that there is less opportunity for isolation in southern Oaxaca. In areas of less topological variation, it is possible that these habitat shifts led to the current distribution and genetic variation within a species but that the populations may have been less isolated than those in the canyons of central Oaxaca. Also, timing of the expansion and subsequent contraction of populations could limit the isolation among them. If *D. holmgrenii* has more recently undergone these phylogeographic shifts, then perhaps there simply hasn't been time for speciation to occur. Finally, the species boundaries and phylogeography have not yet been investigated in the majority of species in *Dioon*. It may well be that the populations in central Oaxaca that we place in separate species are not so differentiated but rather are somewhere in the middle of the speciation process. Further work aims to test these possibilities.

Conservation Implications

Based on our field observations and discussions with local communities, the six populations of *D. holmgrenii* sampled for this work are under minimal threat at the moment. However, land use practices can affect the demography and distribution of these populations (Velasco-García et al. 2016) and we do not know the influence the current climate crisis will have. The fact that *D. holmgrenii* has experienced range expansion and contraction in the past and still has genetically diverse populations that are mostly intact is a positive sign for *Dioon* conservation as we move through the current period of climate change. However, a key component of this history that remains unknown is the actual rate of population movement. We need finer scale, and absolute, estimates of the timing of any range expansion before we know how the species might respond to future changes in habitat suitability. We also lack any knowledge of the dispersal agents that were responsible for past migrations and whether any exist today. In future considerations of land use practices or conservation efforts for this species we would emphasize the rare and private alleles present in all populations and the clear separation of all populations into distinct genetic clusters as indicative of unique genetic contributions that each population may make to the future of the species.

Conclusions

In this study we have characterized the genomic structure and diversity, as well as the morphological variation of *D. holmgrenii*. We find significant structure among sampled populations and no evidence for recent gene flow. We also find that all sampled populations have undergone a significant reduction in size in the relatively recent past. Despite this, there remains moderate genetic diversity within populations and no evidence of inbreeding. Further, while *D. holmgrenii* shows a range of leaf morphology (mostly size), this does not clearly differentiate populations with the possible exception of plants at Ixtayutla. We suggest that the history of *D. holmgrenii* is one of population expansion across its current range, followed by contraction into smaller populations. Because of the contribution of rare and unique alleles present in all populations to the overall genetic diversity among populations, we think it is important to maintain populations throughout the species' range and monitor the effects of land use and climate change.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10592-023-01569-4.

Author contributions BD, TG and SS conceived the study. BD and SS collected the data. BD analyzed the data and wrote the manuscript. All authors reviewed the manuscript.

Funding This work was funded by The Huntington and Sociedad para el Estudio de los Recursos Bioticos de Oaxaca.

Data availability Illumina reads used in this study are available under the NIH Bioproject PRJNA1023921. Rscripts for filtering data and running analyses are available at https://github.com/dorseyb/D_holmgrenii_congen.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215(3):403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Anderson MJ (2001) A new method for non-parametric multivariate analysis of variance. Austral Ecol 26(1):32–46. https://doi.org/10.1111/j.1442-9993.2001.01070.pp.x

Arbizu PM (2017) "PairwiseAdonis: Pairwise Multilevel Comparison Using Adonis." R

Baetscher DS, Clemento AJ, Ng TC, Anderson EC, Garza JC (2018) Microhaplotypes provide increased power from short-read DNA sequences for relationship inference. Mol Ecol Resour 18(2):296–305. https://doi.org/10.1111/1755-0998.12737

Bohonak AJ (1999) Dispersal, gene flow, and population structure. Q Rev Biol 74(1):21–45

Bushnell, Brian. 2019. "BBMap." SourceForge. April 7, 2019. https://sourceforge.net/projects/bbmap/

Caballero A, Rodríguez-Ramilo ST (2010) A new method for the partition of allelic diversity within and between subpopulations. Conserv Genet 11(6):2219–2229. https://doi.org/10.1007/s10592-010-0107-7

Cabrera-Toledo D, González-Astorga J, Vovides AP (2008) Heterozygote excess in ancient populations of the critically endangered Dioon Caputoi (Zamiaceae, Cycadales) from Central Mexico. Bot J Linn Soc 158(3):436–447. https://doi.org/10.1111/j.1095-8339.2008.00868.x

Cabrera-Toledo D, González-Astorga J, Nicolalde-Morejon F, Vergara-Silva F, Vovides AP (2010) Allozyme diversity levels in two Congeneric Dioon Spp. (Zamiaceae, Cycadales) with contrasting rarities. Plant Syst Evol 290(1–4):115–125. https://doi.org/10.1007/s00606-010-0354-6

Calonje M, Stevenson DW, Osborne R (2013) The World List of Cycads, Online Edition. 2021 2013. https://cycadlist.org/index.php

Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH (2011) Stacks: building and genotyping loci de novo from short-read sequences. G3: Genes Genomes Genetics 1(3):171–182. https://doi.org/10.1534/g3.111.000240

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: an analysis tool set for population genomics. Mol Ecol 22(11):3124–3140. https://doi.org/10.1111/mec.12354

Chaw S-M, Shih A-C, Wang D, Yu-Wei Wu, Liu S-M, Chou T-Y (2008) The mitochondrial genome of the gymnosperm *Cycas taitungensis* contains a novel family of short interspersed elements, bpu sequences, and abundant RNA editing sites. Mol Biol Evol 25(3):603–615. https://doi.org/10.1093/molbev/msn009

Clugston JAR, Kenicer GJ, Milne R, Overcast I, Wilson TC, Nagalingum NS (2019) RADseq as a valuable tool for plants with large genomes—a case study in cycads. Mol Ecol Resour 19(6):1610–1622. https://doi.org/10.1111/1755-0998.13085

Clugston JAR, Ruhsam M, Kenicer GJ, Henwood M, Milne R, Nagalingum NS (2022) Conservation genomics of an Australian Cycad *Cycas calcicola*, and the absence of key genotypes in botanic gardens. Conserv Genet 23(3):449–465. https://doi.org/10.1007/s10592-022-01428-8

Coates DJ, Byrne M, Moritz C (2018) Genetic diversity and conservation units: dealing with the species-population continuum in the age of genomics. Front Ecol Evol. https://doi.org/10.3389/fevo.2018.00165

de Luca P, Sabato S, Torres MV (1981) Dioon Merolae (Zamiaceae), a New Species from Mexico. Brittonia 33(2):179–185. https://doi.org/10.2307/2806317

De Meeûs T, Goudet J (2008) A step-by-step tutorial to use HierFstat to analyse populations hierarchically structured at multiple levels. Infect Genetics Evol 7(January):731–735. https://doi.org/10.1016/j.meegid.2007.07.005

Donaldson J (2003) Cycads : status survey and conservation action plan. IUCN. https://portals.iucn.org/library/node/8203

Dorsey BL, Gregory TJ, Sass C, Specht CD (2018) Pleistocene diversification in an ancient lineage: a role for glacial cycles in the evolutionary history of Dioon Lindl. (Zamiaceae). Am J Bot 105(9):1512–1530. https://doi.org/10.1002/ajb2.1149

Dray S, Dufour A-B (2007) "The Ade4 package: implementing the duality diagram for ecologists. J Stat Softw 22(4):1–20. https://doi.org/10.18637/jss.v022.i04

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131(2):479–491

Fischer MC, Rellstab C, Leuzinger M, Roumet M, Gugerli F, Shimizu KK, Holderegger R, Widmer A (2017) Estimating genomic diversity and population differentiation – an empirical comparison of microsatellite and SNP variation in arabidopsis halleri. BMC Genomics 18(1):69. https://doi.org/10.1186/s12864-016-3459-7

Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. Genetics 180(2):977–993. https://doi.org/10.1534/genetics.108.092221

García-Dorado A, Caballero A (2021) Neutral genetic diversity as a useful tool for conservation biology. Conserv Genet 22(4):541–545. https://doi.org/10.1007/s10592-021-01384-9

Garner BA, Hoban S, Luikart G (2020) IUCN red list and the value of integrating genetics. Conserv Genet 21(5):795–801. https://doi.org/10.1007/s10592-020-01301-6

González-Astorga J, Vovides AP, Ferrer MM, Iglesias C (2003) Population genetics of *Dioon edule* Lindl. (Zamiaceae, Cycadales): biogeographical and evolutionary implications. Biol J Lin Soc 80(3):457–467. https://doi.org/10.1046/j.1095-8312.2003.00257.x

González-Astorga J, Vovides AP, Cruz-Angon A, Octavio-Aguilar P, Iglesias C (2005) Allozyme variation in the three extant populations of the narrowly endemic cycad *Dioon angustifolium* Miq. (Zamiaceae) from North-Eastern Mexico. Ann Bot 95(6):999–1007. https://doi.org/10.1093/aob/mci106

González-Astorga J, Vergara-Silva F, Vovides AR, Nicolalde-Morejon F, Cabrera-Toledo D, Perez-Farrera MA (2008a) Diversity and genetic structure of three species of Dioon Lindl. (Zamiaceae, Cycadales) from the Pacific Seaboard of Mexico. Biol J Lin Soc 94(4):765–776. https://doi.org/10.1111/j.1095-8312.2008.01034.x

González-Astorga J, Vovides AP, Cabrera-Toledo D, Nicolalde-Morejón F (2008b) Diversity and genetic structure of the endangered Cycad *Dioon sonorense* (Zamiaceae) from Sonora, Mexico: evolutionary and conservation implications. Biochem Syst Ecol 36(12):891–899. https://doi.org/10.1016/j.bse.2008.11.006

Goudet J, Jombart T (2018) Hierfstat: Estimation and Tests of Hierarchical F-Statistics. http://www.r-project.org, http://github.com/jgx65/hierfstat

Gregory TJ, Chemnick J, Salas-Morales S, Vovides AP (2003) A new species in the genus Dioon (Zamiaceae) from North-Central Oaxaca, Mexico. Bot J Linn Soc 141(4):471–476. https://doi.org/10.1046/j.1095-8339.2003.t01-1-00151.x

Gutiérrez-Ortega JS, Jiménez-Cedillo K, Pérez-Farrera MA, Martínez JF, Molina-Freaner F, Watano Y, Kajita T (2018a) Species definition of Dioon Sonorense (Zamiaceae, Cycadales), and description of D. Vovidesii, a new cycad species from Northwestern Mexico. Phytotaxa 369(2):107–114. https://doi.org/10.11646/phytotaxa.369.2.4

Gutiérrez-Ortega JS, Jiménez-Cedillo K, Pérez-Farrera MA, Vovides AP, Martínez JF, Molina-Freaner F, Imai R et al (2018b) Considering evolutionary processes in cycad conservation: identification of evolutionarily significant units within *Dioon Sonorense* (Zamiaceae) in Northwestern Mexico. Conserv Genetics. https://doi.org/10.1007/s10592-018-1079-2

Gutiérrez-Ortega JS, Salinas-Rodríguez MM, Martínez JF, Molina-Freaner F, Pérez-Farrera MA, Vovides AP, Matsuki Yu et al (2018c) The phylogeography of the cycad genus Dioon (Zamiaceae) clarifies its cenozoic expansion and diversification in the Mexican transition zone. Ann Bot 121(3):535–548. https://doi.org/10.1093/aob/mcx165

Gutiérrez-Ortega JS, Salinas-R M, Ito T, Farrera MA, Vovides A, Martínez J, Molina-Freaner F et al (2020) "Niche conservatism promotes speciation in cycads: the case of Dioon Merolae (Zamiaceae) in Mexico. New Phytol. https://doi.org/10.1111/nph.16647

Hoban S, Archer FI, Bertola LD, Bragg JG, Breed MF, Bruford MW, Coleman MA et al (2022) Global genetic diversity status and trends: towards a suite of essential biodiversity variables (EBVs) for genetic composition. Biol Rev 97(4):1511–1538. https://doi.org/10.1111/brv.12852

IUCN (2021) The IUCN Red List of Threatened Species. IUCN Red List of Threatened Species. 2021. https://www.iucnredlist.org/en

Jombart T (2008) Adegenet: A R package for the multivariate analysis of genetic markers. Bioinformatics 24(11):1403–1405. https://doi.org/10.1093/bioinformatics/btn129

Jombart T, Ahmed I (2011) Adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics 27(21):3070–3071. https://doi.org/10.1093/bioinformatics/btr521

Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet 11(1):94. https://doi.org/10.1186/1471-2156-11-94

Jost L (2008) GST and its relatives do not measure differentiation. Mol Ecol 17(18):4015–4026. https://doi.org/10.1111/j.1365-294X.2008.03887.x

Jost L, Archer F, Flanagan S, Gaggiotti O, Hoban S, Latch E (2018) Differentiation measures for conservation genetics. Evol Appl 11(7):1139–1148. https://doi.org/10.1111/eva.12590

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. https://doi.org/10.7717/peerj.281

Kamvar ZN, Brooks JC, Grünwald NJ (2015) Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. Front Genet 6:208. https://doi.org/10.3389/fgene.2015.00208

Keenan K, McGinnity P, Cross TF, Crozier WW, Prodöhl PA (2013) DiveRsity: an R package for the estimation and exploration of population genetics parameters and their associated errors. Methods Ecol Evol 4(8):782–788. https://doi.org/10.1111/2041-210X.12067

Lepais O, Weir JT (2014) SimRAD: an r package for simulation-based prediction of the number of loci expected in RADseq and similar genotyping by sequencing approaches. Mol Ecol Resour 14(6):1314–1321. https://doi.org/10.1111/1755-0998.12273

López-Cortegano E, Pérez-Figueroa A, Caballero A (2019) Metapop2: re-implementation of software for the analysis and management of subdivided populations using gene and allelic diversity. Mol Ecol Resour 19(4):1095–1100. https://doi.org/10.1111/1755-0998.13015

Lopez-Gallego C (2008) Demographic variation in cycad populations inhabiting contrasting forest fragments. Biodivers Conserv 17:1213–1225. https://doi.org/10.1007/s10531-007-9263-6

Mace GM (2004) The role of taxonomy in species conservation. Philos Trans Royal Soc Lond Ser B Biol Sci 359(1444):711–719. https://doi.org/10.1098/rstb.2003.1454

Malinsky M, Trucchi E, Lawson DJ, Falush D, Takezaki N (2018) RADpainter and FineRADstructure: population inference from RADseq data. Mol Biol Evol 35(5):1284–1290. https://doi.org/10.1093/molbev/msy023

Moritz C (1994) Defining 'evolutionarily significant units' for conservation. Trends Ecol Evol 9(10):373–375. https://doi.org/10.1016/0169-5347(94)90057-4

Mussmann SM, Douglas MR, Chafin TK, Douglas ME (2019) BA3-SNPs: contemporary migration reconfigured in BayesAss for next-generation sequence data. Methods Ecol Evol 10(10):1808–1813. https://doi.org/10.1111/2041-210X.13252

Norstog K, Nicholls TJ (1997) The biology of the cycads. Comstock Pub Associates, Ithaca, N.Y

Oksanen J, Blanchet FG, Friendly M, Kindt M, Legendre P, McGlinn D, Minchin PR et al (2019) Vegan: Community Ecology Package. R. https://CRAN.R-project.org/package=vegan

Ortiz, Edgardo M (2018) "Paralog-Finder V1.0: Detect And Blacklist Paralog Rad Loci." Zenodo. https://doi.org/10.5281/ZENODO.1209328

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

Pérez-Figueroa A, Saura M, Fernández J, Toro MA, Caballero A (2009) METAPOP—a software for the management and analysis of subdivided populations in conservation programs. Conserv Genet 10(4):1097–1099. https://doi.org/10.1007/s10592-008-9718-7

Peter BM, Slatkin M (2013) Detecting range expansions from genetic data. Evolution 67(11):3274–3289. https://doi.org/10.1111/evo.12202

Peter BM, Slatkin M (2015) the effective founder effect in a spatially expanding population. Evolution 69(3):721–734. https://doi.org/10.1111/evo.12609

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double Digest RADseq: an inexpensive method for de novo snp discovery and genotyping in model and non-model species. PLoS ONE 7(5):e37135. https://doi.org/10.1371/journal.pone.0037135

Piry S, Luikart G, Cornuet J-M (1999) Computer note. BOTTLENECK: a computer program for detecting recent reductions in the effective size using allele frequency data. J Hered 90(4):502–503. https://doi.org/10.1093/jhered/90.4.502

R Core Team. 2019. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/

Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA (2018) Posterior summarization in Bayesian phylogenetics using tracer 1.7. Syst Biol 67(5):901–904. https://doi.org/10.1093/sysbio/syy032

Salas-Morales SH, Chemnick J, Gregory TJ (2016) A new cycad species in the genus Dioon (Zamiaceae) from the Mixteca Region of Oaxaca, Mexico. Cactus and Succulent Journal 88(1):35–42. https://doi.org/10.2985/015.088.0104

Stolle E, Moritz RFA (2013) RESTseq – Efficient Benchtop population genomics with RESTriction Fragment SEQuencing. PLoS ONE 8(5):e63960. https://doi.org/10.1371/journal.pone.0063960

Takezaki N, Nei M (1996) Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. Genetics 144(1):389–399

Taylor TN, Taylor EL, Krings M (2009) Paleobotany: the biology and evolution of fossil plants, 2nd edn. Academic Press, Amsterdam Boston

Terry I, William Tang A, Taylor JD, Singh R, Vovides A, Cibrian-Jaramillo A (2012) An overview of cycad pollination studies. Mem N Y Bot Gard 106(January):352–394

Toon A, Irene Terry L, Tang W, Walter GH, Cook LG (2020) Insect pollination of cycads. Austral Ecol 45(8):1033–1058. https://doi.org/10.1111/aec.12925

Valencia-Montoya WA, Tuberquia D, Guzmán PA, Cardona-Duque J (2017) Pollination of the Cycad Zamia Incognita A. Lindstr. & Idárraga by Pharaxonotha BEetles in the Magdalena Medio valley, Colombia: a mutualism dependent on a specific pollinator and its significance for conservation. Arthropod-Plant Interactions 11(5):717–729. https://doi.org/10.1007/s11829-017-9511-y

Velasco-García MV, Valdez-Hernández JI, Ramírez-Herrera C, Hernández-Hernández ML, López-Upton J, López-Mata L, López-Sánchez H (2016) Estructura, heterogeneidad de estadios y patrón de dispersión espacial de Dioon holmgrenii (Zamiaceae). Bot Sci 94(1):75–87. https://doi.org/10.17129/botsci.258

Velasco-García MV, Ramírez-Herrera C, López-Upton J, Valdez-Hernández JI, López-Sánchez H, López-Mata L (2021) Diversity and genetic structure of Dioon Holmgrenii (Cycadales: Zamiaceae) in the Mexican pacific coast biogeographic province: implications for conservation. Plants 10(11):2250. https://doi.org/10.3390/plants10112250

Waples RS (2015) Testing for hardy-weinberg proportions: have we lost the plot? J Hered 106(1):1–19. https://doi.org/10.1093/jhered/esu062

Weir BS, Clark Cockerham C (1984) Estimating F-statistics for the analysis of population structure. Evolution 38(6):1358–1370. https://doi.org/10.1111/j.1558-5646.1984.tb05657.x

Wickham H (2016) Ggplot2: elegant graphics for data analysis. Springer-Verlag, New York

Wilson GA, Rannala B (2003) Bayesian inference of recent migration rates using multilocus genotypes. Genetics 163(3):1177–1191

Wright S (1949) The genetical structure of populations. Ann Eugen 15(1):323–354. https://doi.org/10.1111/j.1469-1809.1949.tb02451.x

Wu C-S, Chaw S-M (2015) Evolutionary stasis in cycad plastomes and the first case of plastome GC-biased gene conversion. Genome Biol Evol 7(7):2000–2009. https://doi.org/10.1093/gbe/evv125

Zimmerman SJ, Aldridge CL, Oyler-McCance SJ (2020) An Empirical comparison of population genetic analyses using microsatellite

and SNP data for a species of conservation concern. BMC Genomics 21(1):382. https://doi.org/10.1186/s12864-020-06783-9
Zonneveld BJM (2012) Genome sizes for all genera of cycadales. Plant Biol 14(1):253–256. https://doi.org/10.1111/j.1438-8677.2011.00522.x