



Assessing evolutionarily significant units in a widespread species: morphology, demography, and adaptive divergence support multiple lineages in the southern range of *Dioon tomasellii* (Zamiaceae)

Brian L. Dorsey¹ · José Said Gutiérrez-Ortega² · Miguel Angel Pérez-Farrera³ · Timothy J. Gregory⁴

Received: 4 June 2026 / Accepted: 5 October 2026 / Published online: 13 October 2026
© The Author(s) 2026

Abstract

A majority of cycad species face multiple challenges to their long-term survival, including habitat loss and illegal collecting. Confronting these threats is made more difficult by a lack of understanding of species genetic diversity and an unstable taxonomy. A comprehensive and uniform assessment of species boundaries and population genetic structure is necessary to ensure that both *ex-situ* and in-situ conservation programs successfully maintain genetic diversity in threatened species. The genus *Dioon* (Zamiaceae) is nearly endemic to Mexico (one Honduran species), and most species are designated Vulnerable to Critically Endangered. As part of our program to understand the evolution and conservation status of *Dioon* species, we measured levels of genetic diversity, estimated population genetic structure, and demographic history of *D. tomasellii* populations in western Mexico. We also compare morphological traits among populations in both univariate and multivariate analyses and compare these to patterns of genetic differentiation. Our results indicate that these populations are isolated, with relatively high genetic diversity, and show no evidence of inbreeding. Surprisingly high levels of private alleles along with significant allelic and morphological differentiation and local adaptation suggest that *D. tomasellii*, as currently circumscribed, may comprise multiple species. We compare our results to previous studies in *Dioon* and suggest future work aimed at revising the taxonomy of the species and informing conservation efforts.

Keywords Conservation · Cycads · Divergence · Genomics · Morphology

Introduction

A well-established taxonomy and clear species boundaries can help define evolutionary significant units of conservation and determine their correct level of threat. Characterizing

the genetic diversity and structure within a species can provide estimates of past demography, migration, inbreeding, and adaptive potential, all of which are helpful to improve assessments of extinction risk (Garner et al. 2020). Recently adopted strategies and recommendations for best practices in plant conservation have recognized the importance of genetic variation as an indicator of population longevity and adaptability, and a target for conservation efforts (e.g., Frankham et al. 2014; Garner et al. 2020; Hoban et al. 2024; Kardos et al. 2021). Taxonomy and population genetics are interdependent as taxonomy is often informed by measures of genetic divergence among populations, especially in cryptic species or species complexes, while inaccurate taxonomy can affect measures of genetic structure within species. Addressing these challenges is both highly necessary and most challenging when morphological distinctions among taxa are quite subtle, while at the same time, the morphological diversity of particular traits within species are substantially variable. In these systems, identifying key

✉ Brian L. Dorsey
bdorsey@huntington.org

¹ The Huntington, 1151 Oxford Rd, CA 91016 San Marino, USA

² RIKEN Center for Interdisciplinary Theoretical and Mathematical Sciences, Hirosawa 2-1, Wako 351-0198, Saitama, Japan

³ Laboratorio de Ecología Evolutiva, Herbario Eizi Matuda, Instituto de Ciencias Biológicas, Universidad de Ciencias y Artes de Chiapas, Tuxtla Gutiérrez, CS, Mexico

⁴ Wild Cycad Conservancy, c/o 280 Woodside Dr., Woodside, CA 94062, USA

morphological traits that align with genetic and ecological differentiation is critical to determining practical species boundaries that can be used to inform conservation assessments and recovery plans.

These are common challenges facing the conservation of many species of Cycads (Cycadales), arguably the most threatened lineage of plants on the planet, with approximately 70% of all species assessed as vulnerable to extinction or worse (IUCN 2025). The taxonomy of several cycad genera has been in flux for some time. Historically, cycad species labels have often been assigned to herbaria specimens without regard for population-level variability. This practice has created vague taxonomic boundaries, causing geographically adjacent but distinct populations to be arbitrarily grouped under the same name, when perhaps they belong to distinct species. As a consequence, over the past decade, numerous new taxa have been described, often as a result of splitting previously recognized species, based on analyses of genetic structure, morphological variation, and ecological niche differentiation (e.g., Gutiérrez-Ortega et al. 2018a, b, 2020a, b, 2021a, b; Pérez-Farrera et al. 2024). While recent studies have assessed genetic diversity across genera (e.g., Cabrera-Toledo et al. 2012; Clugston et al. 2022; Da Silva et al. 2012; Dorsey et al. 2023; González-Astorga et al. 2008a, b; Wang et al. 2019) the conservation genetic status of most cycad species is unknown and we cannot rely on closely related species to infer genetic diversity or to set conservation goals (Hoban et al. 2020). Further, phylogenetic work has revealed high levels of morphological homoplasy within cycad genera (Lindstrom et al. 2024), which highlights the difficulty faced by botanists, regulators, and conservationists when trying to identify and assess these species.

There is a clear need for critical studies of species boundaries and genetic diversity in cycads using multiple complementary data sets - a program we have undertaken in the genus *Dioon* (Zamiaceae) (Gutiérrez-Ortega et al. 2018a, b; 2020a, b; Gutiérrez-Ortega et al. 2021a, b; Dorsey et al. 2023).

Dioon includes 18 species, all but one endemic to Mexico (*D. mejiae* occurs in Honduras). Of the 15 taxa that are assessed for the IUCN Red List, eight are considered endangered, and six are vulnerable to extinction (IUCN 2025). Unassessed taxa are recently described species that have been split from other taxa, which suggests that these smaller species will be similarly ranked, or worse. In this study, we focus on *Dioon tomasellii*, a species from western Mexico that was once defined quite broadly and included populations from Guerrero to Sonora (Fig. 1) (de Luca et al. 1984). It is now clear that the previous circumscription included at least four species: *D. tomasellii* (de Luca et al. 1984), *D. sonorensis* (Chemnick et al. 1997), *D. vovidesii* (Gutiérrez-Ortega

et al. 2018a, b), and *D. stevensonii* (Nicolalde-Morejón et al. 2009). As currently circumscribed, documented populations of *D. tomasellii* are restricted to Jalisco, Nayarit, Sinaloa, and western Durango. Our observations and collection locality data, along with publicly available occurrence data, suggest a patchy distribution of apparently isolated populations across this range (GBIF.org 2025). These occur mostly on the western slopes of the Sierra Madre Occidental in the northern end of the range, but also into the Trans-Mexican Volcanic Belt in Jalisco, generally in pine-oak forest and tropical deciduous forest.

Given the widespread and discontinuous distribution of this species, we hypothesize that the connectivity among populations is low and that this may have led to inbreeding and population divergence. We also suspect that there may be multiple evolutionary lineages among southern populations of *D. tomasellii*, based on our field observations of morphological variation. To test these hypotheses and characterize the conservation genetic status of the species, we produced population-level RADseq data and measured putatively diagnostic leaf traits. Genomic data were used to assess the genetic structure among populations, as well as their genetic diversity and demographic history. Leaf trait data were used to test for morphological divergence and to confirm the taxonomic utility of these characters. This work represents an initial reassessment of the integrity of the species as currently circumscribed and adds valuable genetic information for the continued conservation of this species.

Methods

Sampling and sequencing protocols

We collected leaf tissue from six localities (18–48 individuals per locality) in Jalisco and Nayarit, including the type locality (Fig. 1, Appendix 1). Tissue was preserved in silica gel in the field and stored at -80° C upon returning to our lab. DNA was extracted and RADseq libraries were prepared according to Dorsey et al. (2023). Paired-end 150 bp reads were sequenced on one S1 lane of a NovaSeq 6000 sequencer (Illumina, San Diego, CA) by the Genomics and Cell Characterization Core Facility at the University of Oregon. Reads were filtered for quality and adapter sequences using the `process_radtags` program from the Stacks pipeline (Davey et al. 2011; Catchen et al. 2013; Rochette et al. 2019) and again using `bbduk` ($k=15$, $hdist=1$) from the BBTools suite v. 38.42 (Bushnell 2019). Raw read and `process_rad_tag` demultiplexing statistics are reported in Tables S1–S2. We used the Stacks pipeline to assemble loci after optimizing the main assembly parameters – *m*, *M*, and *N*. Parameter optimization for Stacks locus assembly



Fig. 1 Map of Jalisco state and southern Nayarit state showing collection localities of *D. tomasellii*. Carrizal locality is represented only by morphological data. The black rectangle in the inset indicates the region displayed in the map. The cyan lines delimit the distribution

range of *D. tomasellii* as originally described (De Luca et al. 1984), and the yellow lines delimit the distribution of *D. tomasellii* as currently circumscribed

was performed as in Dorsey et al. (2023). Briefly, loci were assembled from a subset of 61 individuals from all populations, including two technical replicates, at varying values for each parameter. Error rates defined in Mastretta-Yanes et al. (2015) were calculated between technical replicates at each parameter combination using a custom R script. Parameters that minimized error were chosen as optimal for this data set: $m=7$, $M=N=1$. Other parameters were left at defaults except the minimum minor allele count (`--min-mac`), which was set to 3 to ensure any SNP would be present in at least two individuals. We removed organellar loci that were identified using a `blastn` search of a combined data base of the *Cycas taitungensis* mitochondrion genome (GenBank NC_010303.1) and *D. spinulosum* chloroplast genome (GenBank NC_027512.1). Potential paralogs were identified using the R package `rCNV` (Karunaratne et al. 2023) and removed. All loci present in less than 70% of individuals were also removed. The R package `pcadapt` was used to identify potentially non-neutral SNPs. Because we found that minor allele frequency can affect the number of non-neutral loci identified, we took a conservative approach to assemble neutral and non-neutral sets of loci. For the neutral data set we included all SNPs, regardless

of minor allele frequency (MAF), in the `pcadapt` analysis. Any locus that contained a non-neutral SNP in this analysis was removed from the neutral set. We then ran `pcadapt` ignoring SNPs with $MAF < 0.05$ to reduce the number of loci in our non-neutral data set that may be flagged as such due to MAF. This strategy was applied to both the ingroup only and the entire set of samples to produced three sets of loci: (1) neutral (2507 loci) and (2) non-neutral (85 loci) from *D. tomasellii* samples only and (3) neutral loci (2422) including six *D. holmgrenii* samples as an outgroup. We also constructed a reduced data set from the neutral ingroup set with 16 individuals per population and a minimum allele frequency of 0.05 to reduce computational burden in two analyses (reduced set).

Genetic structure among populations

We used several methods to estimate genetic structure among sampled localities. First, `fineRADstructure` (Malinsky et al. 2018) was used to estimate a coancestry matrix (`-x 500000 -y 20000000 -z 5000`) using the neutral ingroup data set, and associated R scripts were used to cluster individuals by coancestry and visualize these results in a heat map.

Four independent analyses were run, and the Markov chain results were compared using Tracer (Rambaut et al. 2018).

Second, for both the neutral and non-neutral ingroup data sets, we used the R package adegenet (Jombart and Ahmed 2011; R Core Team 2024) to cluster individuals using discriminant analysis of principal components components (DAPC; Jombart et al. 2010). We used the find.clusters function, which implements K-means clustering, to test if individuals clustered by allele frequency data correspond to sampling localities (populations). We then used the dapc function to test the assignment of individuals to these clusters and to separate individuals based on discriminant functions of principal components. These results were plotted with ggplot2 (Wickham 2016).

Third, we used Structure v. 2.3.4 (Pritchard et al. 2000; Falush et al. 2003) to estimate genomic components and admixture within individuals. For this analysis we randomly chose one SNP per locus in the ingroup data set. We ran 10 independent analyses for each number of genomic clusters (K)=1–10, setting burnin to 5000 and number of MCMC reps to 500,000. We used the online Clumpak suite (Kopelman et al. 2015) to identify and plot modes of each K and to estimate the best K overall using the methods from Pritchard et al. (2000) and Evanno et al. (2005).

Finally, we estimated the phylogenetic relationships among individuals and populations using the neutral outgroup data set. We first created consensus sequences of the alleles at each locus within each individual and then aligned each locus using mafft (v7.526; Katoh et al. 2002; v7.526; Katoh and Standley 2013). For each locus we identified the best model of sequence evolution and estimated a gene tree with branch using the ultrafast bootstrap in iqtree2 (-B 2000 -T 1 --quiet -pers 0.2 -nstop 500; Hoang et al. 2018; Kalyanamoorthy et al. 2017; Minh et al. 2020). We then used these gene trees as input for a weighted ASTRAL (v1.19.3.5; Zhang et al. 2017; Zhang and Mirarab 2022) analysis to estimate a species tree.

Genetic diversity and differentiation

We estimated parameters of genetic diversity within each identified genomic cluster and parameters of differentiation and fixation between them for both the neutral and non-neutral data sets. We used the R package poppr (Kamvar et al. 2014, 2015) to count private alleles within each population and a custom R script to calculate the percentage of alleles found in each population that were private. We used hierfstat (Goudet and Jombart 2018) to estimate F -statistics, observed heterozygosity and gene diversity; mmod (Winter 2012) to estimate G'_{ST} ; and Jost's D , and metapop2 (López-Cortegano et al. 2019) to estimate rarefied allelic richness and allelic divergence. Among-population parameters

were estimated globally and pairwise. We also estimated proportions of genomic variation within and among populations and Phi-statistics by running an AMOVA as implemented in the R package poppr. Loci with >5% missing data were excluded from the AMOVA, resulting in a data set of 943 neutral loci and 34 non-neutral loci. The significance of covariance components was evaluated using 9999 permutations.

Demographic history

We tested for recent population size reductions using the Bottleneck (v. 1.2.02; Piry et al. 1999) under both the infinite alleles and the step-wise mutation models using the sign test and the standardized differences test, as recommended in Piry et al. (1999) for the large number of loci in our data set.

Recent historical migration was estimated using Bayes-Ass3 (Wilson and Rannala 2003) and the coalescent-based method implemented in MIGRATE-N v.3.6 (Beerli and Felsenstein 1999, 2001). We ran the BayesAss3 analysis on the ingroup data set with either the sampling localities or the inferred genomic clusters as populations. For each of these analyses, we filtered out loci with >4 alleles then ran four independent runs of 3E7 iterations, with a burn in of 1E7, sampling interval of 10,000, and mixing parameters: $dM=0.08$, $dA=0.3$, $dF=0.04$. We verified convergence and stationarity using Tracer (Rambaut et al. 2018).

The reduced dataset was used as input for the MIGRATE-N analysis of historic gene flow among the four genomic clusters. A stepwise mutation model with constant mutation rates across loci was assumed. Observed F_{ST} values were used as starting parameters for the estimation of mutation-scaled migration rates over $4N_e$ generations (M). The prior distribution to estimate M was uniform between 0 and 2000 with $\delta=200$. The prior to estimate theta (Θ) consisted of a uniform distribution from 0 to 0.25 with $\delta=0.025$. Metropolis-coupled Markov chain Monte Carlo consisted of one long chain of 25,000 steps, sampling every 20 steps, which gave a total of 500,000 sampled genealogies. The first 50,000 genealogies were discarded as burnin. The number of migrants per generation (N_m) was estimated via $\Theta_a M_{a \rightarrow b} = 4N_m$.

Morphological variation

We measured 23 leaf traits in 148 individuals across seven localities of *D. tomasellii*. For one population (Carrizal), we were not able to produce DNA data, so it is not included in the above genomic analyses. A second population (Rancho Eureka) was not included in the morphological data set.

Most traits are self-explanatory, but several calculated indices require explanation (Fig. 2). We define the falcate distance as the distance from the proximal margin of a leaflet, at its midpoint, to the chord of the arc defined by that margin; it gives a measure of how falcate the leaflet is. We also calculated an index of leaf recurve by measuring the chord of the arc defined by the abaxial side of a leaflet and the length of that leaflet when pressed flat and taking the ratio of these two measurements (recurved length/flat length). A recurved leaflet will have a shorter chord distance than flat length and thus a ratio < 1 , while in a flat or straight leaflet, those measurements are the same and the ratio = 1. We also counted the maximum number of spines found on the distal and proximal leaflet margins of a single leaflet. The following traits were recorded as discrete states and so were ordinal: leaflet imbricate index where $< 25\%$ leaflet overlap = 0, $25\text{--}75\%$ = 1, and $> 75\%$ = 2; leaf aspect where straight = 0, arching = 1, bent laterally = 2; tomentum (rachis and leaflet surfaces) where glabrous = 0, pubescent = 1, tomentose = 2.

Analyses of the morphological data were carried out in R (R Core Team 2024). Data for each continuous trait were centered by median and scaled by median absolute deviation and then winsorized to reduce the effect of outliers. Most traits failed the Shapiro test for normality, so we test for differences among populations for each trait using the Kruskal-Wallis test and the Dunn post-hoc test to identify significant pairwise differences after Holm's (1979) correction for multiple tests. The non-normality of most trait distributions also precluded the use of typical linear discriminant analysis (LDA) to separate populations based on

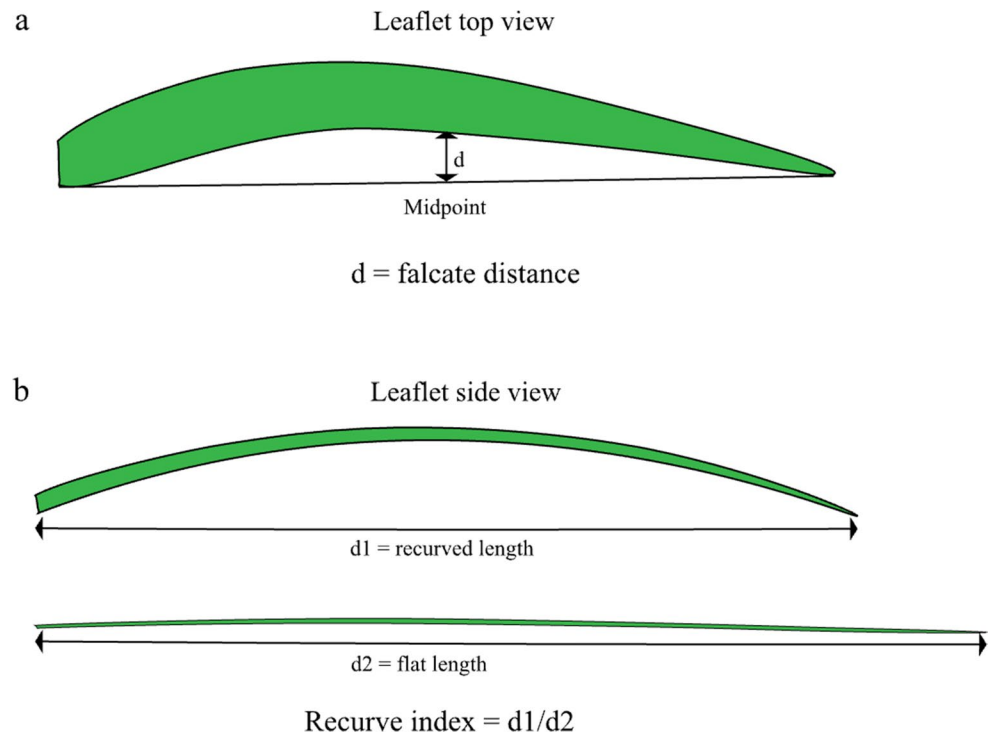
our multivariate data. So, we again used DAPC because the principal components are normally distributed and can be used as input to LDA. For this and further tests we combined Pedro Moreno and Santuario into one group based on individual trait comparisons, preliminary multivariate analyses, and genomic clustering (see Results). Only continuous traits were used for the DAPC. We then used the scores of the first two discriminant functions to create a Euclidean distance matrix among individuals and used this as input for a permanova analysis (Anderson 2001; McArdle and Anderson 2001) using the function `adonis2` in the R package `vegan` v. 2.6-8 (Oksanen et al. 2024) with 9999 permutations. Because differences in dispersion among groups can affect the results of the permanova method (Anderson 2001), we tested for significantly different dispersion with an ANOVA on the output of the `betadisper` function in `vegan`, although permanova is generally less sensitive to the dispersion effect (Anderson and Walsh 2013). We used pairwise permanovas as post hoc tests for significant differences between populations (Marinez Arbizu 2020).

Results

Genetic structure across localities

The K-mean clustering method found four genomic clusters for both the neutral and non-neutral data sets. The DAPC clearly separated the four neutral clusters with no overlap among them (Fig. 3a), while the smaller, non-neutral set

Fig. 2 (a) Falcate distance: between the leaflet margin and the line connecting the leaflet base to the apex, at midpoint of leaf. (b) Recurve index: the ratio between the length of the leaf in its natural decurved state when viewed from the side (d_1) and the length when pressed flat (d_2)



produced more dispersed but clearly identifiable clusters, with some slight overlap among some (Fig. 3b). Individuals from Sierra de Quila and Soledad all clustered separately, while individuals from Pedro Moreno and Santuario form one cluster, as did those from Rancho Eureka and Miravalles (Fig. 3).

Examination of the Markov chain output from the fineRADstructure replicate runs indicates that three of the four found a similar globally optimal posterior probability. A single run with the highest ESS values after burnin (K and posterior > 300) was chosen to represent the optimal solution. These results indicate that the same groups found in the dapc share greater coancestry within them than between (Fig. 4). There is also evidence of substructure within these groups where several individuals share a high degree of coancestry, perhaps indicating parent-offspring groups or siblings.

Comparisons among the chosen number of genomic clusters (K) in the Structure analysis using the methods of Pritchard et al. (2000) and Evanno et al. (2005) suggest 4 groups as the best K . Figure 5 shows that these correspond to the groups identified in the DAPC and fineRADstructure analyses. There are only a few individuals in the Miravalles and Rancho Eureka populations that show evidence of admixture and this is slight.

The ASTRAL species tree estimated from the neutral gene trees clearly separates the same four groups found in our clustering analyses and the outgroup individuals into clades with local posterior probabilities (LPP) for these main quartets = 1 (Fig. S1). Consistent with our clustering analyses, individuals from Miravalles are interspersed among those of Rancho Eureka, and those of Pedro Moreno are similarly distributed among individuals of Santuario, although internal branches of the main clades have low LPP, generally. Rooting the tree with the *D. holmgrenii* clade indicates a

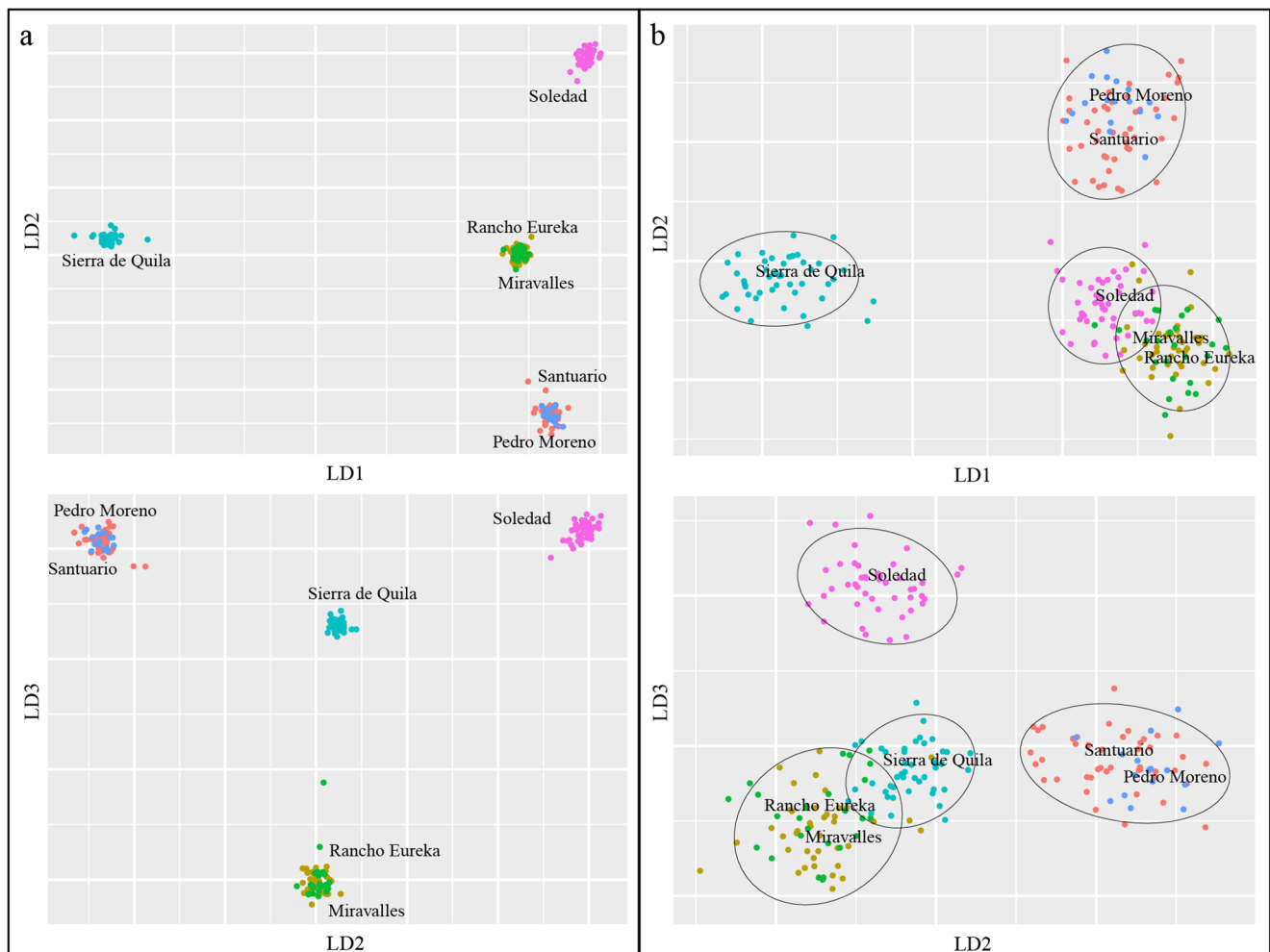


Fig. 3 Scatter plots showing the position of *D. tomasellii* individuals along the three linear discriminant axes from a discriminant analysis of principal components of neutral (a) and non-neutral (b) loci. Colors indicate sample localities, and the locality names are given at the

centroid of their respective individuals. Bayesian information criterion from the find_clusters method indicated four as the optimal number of clusters for both data sets

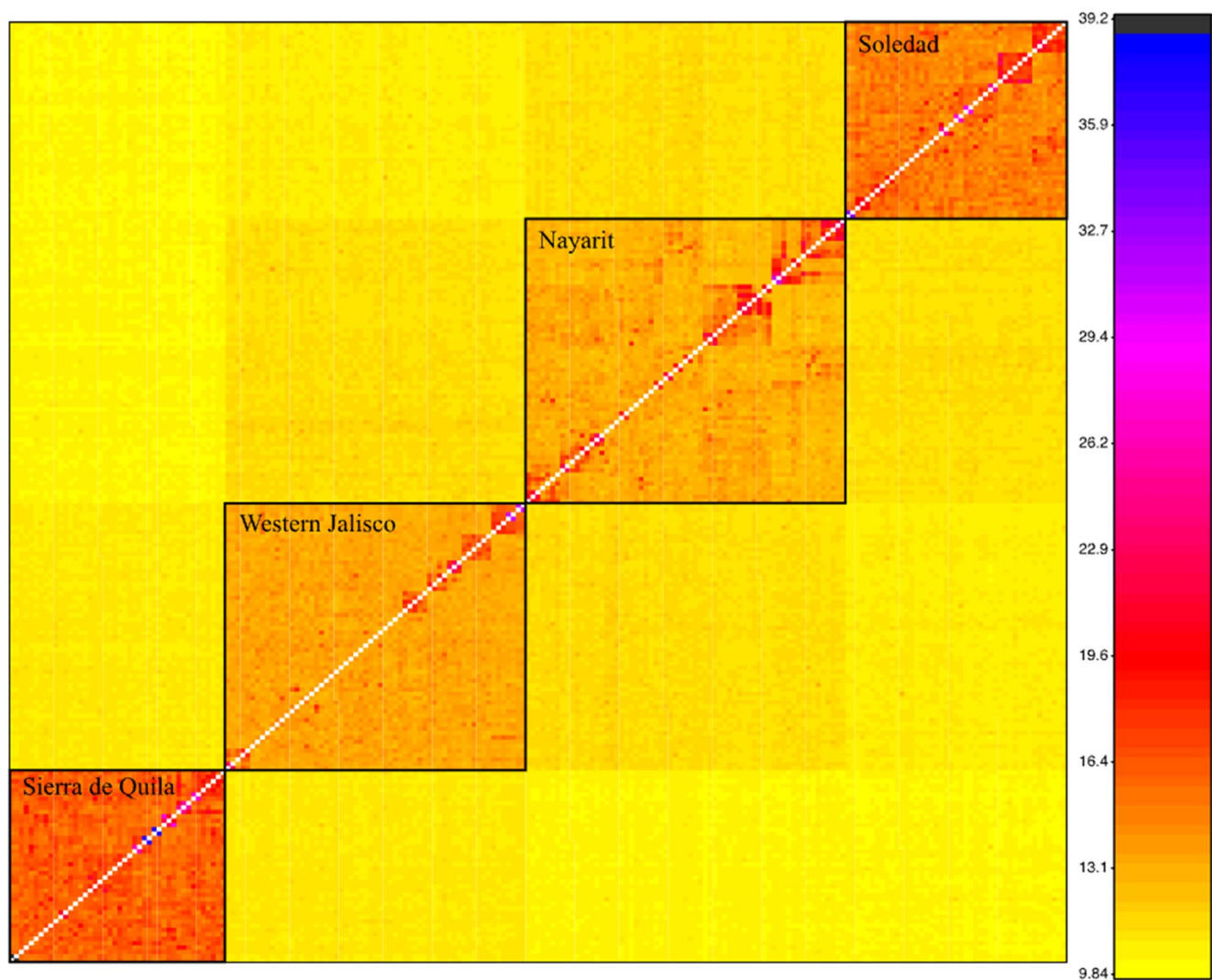


Fig. 4 Heat map showing the pairwise coancestry between individuals of *D. tomasellii* based on nearest-neighbor alleles as calculated in fineRADstructure. Cells represent pairwise estimates. Scale is in summed

proportions of nearest-neighbor alleles across all loci. Individuals are arranged according to a MCMC clustering algorithm based on coancestry and major clusters are outlined in black

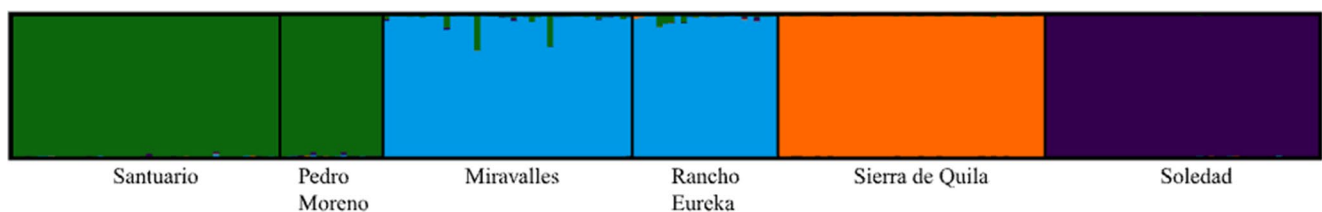


Fig. 5 Results of the structure analysis for populations of *D. tomasellii* showing the proportion of individuals' genomes (colors in columns) assigned to one of the optimal $K=4$ genomic clusters. Individuals are arranged by locality

branching order where Sierra de Quila was first to diverge, followed by Santuario/Pedro Moreno, and finally, Soledad and Miravalles/Rancho Eureka split.

Given the consistency of grouping Miravalles with Rancho Eureka and Santuario with Pedro Moreno, we refer to them as the Nayarit and Western Jalisco populations, respectively, when considering genomic clusters.

Genetic diversity and differentiation

We used several measures to examine genetic diversity within and among genomic clusters identified here. The per population percentage of private alleles, allelic richness, inbreeding coefficient, observed heterozygosity, and gene diversity (expected heterozygosity) are reported in Table 1

Table 1 Estimates of genetic diversity within genetic clusters/populations of *Dioon tomasellii*

Population	Neutral Loci						Non-Neutral Loci			
	<i>N</i>	% Private Alleles	<i>A</i>	<i>F</i>	<i>H_O</i>	<i>H_S</i>	% Private Alleles	<i>A</i>	<i>H_O</i>	<i>H_S</i>
Nayarit	65	15.22	2.49 [2.45,2.53]	0 [-0.002,0.002]	0.16 [0.14,0.17]	0.18 [0.16,0.2]	21.68	3.49 [3.15,3.85]	0.30 [0.16,0.37]	0.36 [0.23,0.48]
Western Jalisco	61	14.66	2.34 [2.30,2.39]	-0.01 [-0.02,0.00]	0.13 [0.12,0.15]	0.14 [0.13,0.15]	20.96	3.30 [2.97,3.58]	0.30 [0.21,0.38]	0.32 [0.24,0.42]
Soledad	45	17.96	2.22 [2.18,2.25]	-0.01 [-0.02,0.00]	0.16 [0.14,0.17]	0.17 [0.15,0.19]	20.42	2.78 [2.56,3.03]	0.25 [0.18,0.31]	0.25 [0.15,0.32]
Sierra de Quila	44	24.60	2.16 [2.12,2.19]	0 [0.00,0.01]	0.14 [0.12,0.15]	0.16 [0.15,0.18]	35.93	3.13 [2.80,3.46]	0.35 [0.27,0.44]	0.39 [0.29,0.5]
Mean	54	18.11	2.32	0.00	0.14	0.16	24.75	3.21	0.30	0.33

N number of individuals, *A* mean allelic richness using rarefaction, *F* median inbreeding coefficient, *H_O* median observed heterozygosity, *H_S* median gene diversity, *BS CI* bootstrap confidence interval of the estimates.

Table 2 Global differentiation statistics for *Dioon tomasellii*

Φ_{IS}	Φ_{ST}	Φ_{IT}	G'_{ST}	D_{Jost}	A_{ST}
0.01	0.15	0.16	0.19	0.05	0.30

D_{Jost} Jost's *D*, A_{ST} Allelic differentiation

for both neutral and non-neutral loci. Neutral loci show moderate levels of heterozygosity (range *H_O*: 0.13–0.16 and *H_S*: 0.14–0.18) and no evidence of inbreeding. Notably, we see very high proportions of private alleles in each population (neutral loci range=16–25%). Estimates of these parameters are much higher for non-neutral loci. Observed heterozygosity and gene diversity are roughly twice the neutral estimates and the percentage of non-neutral private alleles within populations is 20–35%.

Estimates of fixation and allelic diversity are presented in Tables 2, 3 and 4. From the neutral data, global

measures of fixation are moderate to high among populations (Φ_{ST} =0.15, G'_{ST} = 0.19, Table 2) and pairwise measures of G'_{ST} range from 0.13 to 0.25 (Table 3, lower), with the greatest value between Sierra de Quila and Soledad. We also calculated two estimates of allelic differentiation among populations: the allelic differentiation, A_{ST} , and Jost's *D* (D_{Jost}). The global A_{ST} value from neutral loci is 0.30 (Table 2) and pairwise A_{ST} values range from 0.29 to 0.40 (Table 3A). Jost's *D* estimates are much lower, with a global estimate of 0.05 and pairwise estimates from 0.03 to 0.07. These two statistics are complementary in that D_{Jost} is a weighted measure that de-emphasizes rare alleles, whereas A_{ST} is not weighted and measures the effect of all alleles equally. Global estimates from the non-neutral data are comparable to the neutral ones (Table 2) as are the pairwise estimates of

Table 3 Pairwise estimates of Hedrick's G'_{ST} between genetic clusters of *Dioon tomasellii*

	Nayarit	Western Jalisco	Soledad	Sierra de Quila
Nayarit		0.12	0.15	0.25
Western Jalisco	0.13		0.21	0.24
Soledad	0.14	0.20		0.26
Sierra de Quila	0.21	0.20	0.25	

Neutral loci below, non-neutral loci above

Table 4 Pairwise estimates of allelic differentiation between genetic clusters of *Dioon tomasellii* from neutral (A) and non-neutral (B) loci

A	Nayarit	Western Jalisco	Soledad	Sierra de Quila
Nayarit		0.29	0.33	0.39
Western Jalisco	0.03		0.37	0.38
Soledad	0.04	0.05		0.40
Sierra de Quila	0.06	0.05	0.07	
B	Nayarit	Western Jalisco	Soledad	Sierra de Quila
Nayarit		0.31	0.35	0.41
Western Jalisco	0.04		0.38	0.39
Soledad	0.06	0.08		0.41
Sierra de Quila	0.11	0.10	0.11	

Jost's *D* below diagonal and A_{ST} above

G'_{ST} except those involving Sierra de Quila, which are somewhat higher (Table 3). Measures of allelic differentiation for non-neutral loci are slightly elevated when measured by A_{ST} (range = 0.31–0.41) and up to twice the neutral estimate for D_{Jost} (range = 0.04–0.11, Table 4).

Demography

The results of the Bottleneck analysis were mixed among models and statistical tests. Under a stepwise mutation model, there is significant evidence for recent population reduction in all four populations according to both the sign test and the standardized differences test. Under the infinite alleles model, however, on we find a significant result for the Nayarit, Sierra de Quila, and Soledad populations from the sing test but only Sierra de Quila and Soledad from the standardized differences test. Since our data are microhaplotypes that more closely align with the IAM, we have high confidence in a past size reduction in Sierra de Quila and Soledad and consider it likely in Western Jalisco but not so in Nayarit.

In both BayesAss analyses, all runs within an analysis found the same distributions among all parameters and ESS values were > 1000 for all except the posterior probability. When assuming localities as populations, the overall proportions of migrants per generation were near zero for all populations except those from Santuario into Pedro Moreno (0.261) and from Miravalles into Rancho Eureka (0.268) (Table 5A). Seventeen individuals in Pedro Moreno were identified as last-generation migrants from Santuario (posterior probability = 1.0), and 24 individuals in Rancho Eureka were identified as migrants from Miravalles, all in the previous generation except for one individual from two generations back (posterior probability = 1.0). Surprisingly, one individual in Soledad and one from Santuario are identified as migrants from Rancho Eureka, both two generations back (posterior probability = 1.0). When designating

the genomic clusters identified above as populations, which combines the localities with migrants, we found essentially zero migration among all clusters (Table 5B). The Migrate-N analysis for gene flow among genomic clusters estimated < 1 migrants over the past 4 N generations, again indicating essentially no migration (Fig. S2).

Morphological diversity

The DAPC separated populations of *D. tomasellii* into four clusters with limited overlap among them, indicating that the measured traits can potentially be used to distinguish populations in this species (Fig. 6). The first cluster includes the Western Jalisco population (Santuario and Pedro Moreno combined) along with Miravalles from Nayarit. The other three are the Carrizal, Soledad, and Sierra de Quila populations. The Carrizal population occupies a position intermediate between the western cluster and Soledad, and while there is much less overlap among the distributions of Sierra de Quila and Soledad with most other localities, none of the populations occupy a wholly distinct space. Quantitative assessments of population assignment based on the morphology are also provided by the dapc function. The overall proportion of individuals assigned to their populations based on the discriminant functions was 0.67. Proportions of individuals correctly assigned to their original populations are given in Table S6, and the posterior probabilities (PP) of these assignments are summarized in an assignment plot (Fig. S3), which shows the original source population for each individual and the PP of every possible assignment. All individuals from Sierra de Quila were correctly assigned with PP = 1.0, whereas the correct assignment proportions ranged from 0.44 to 0.74 for the other populations.

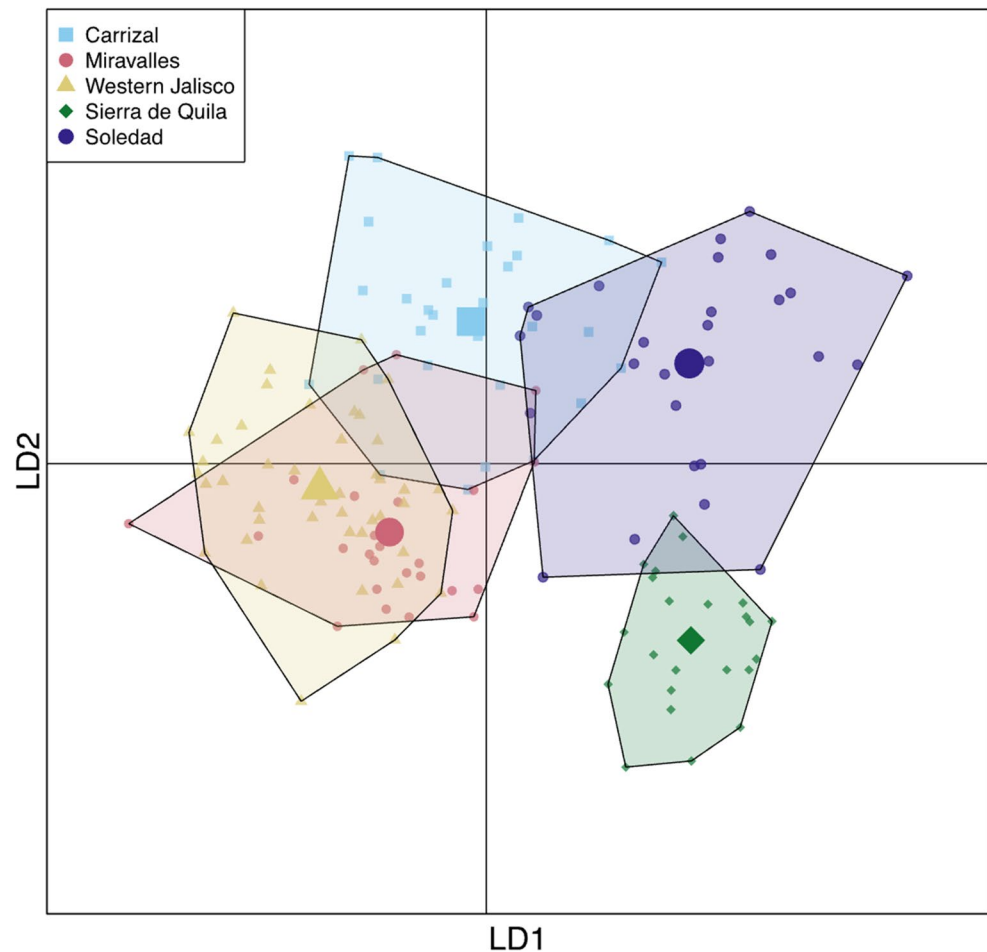
The PERMANOVA on positions in the linear discriminant space shows that significant differences exist among the population centroids (Table S7). The ANOVA

Table 5 Estimates of migration among (A) localities and (B) genomic clusters of *Dioon tomasellii*

A	Soledad	Miravalles	Santuario	Pedro Moreno	Rancho Eureka	Sierra de Quila
Soledad	0.960 (0.015)	0.007 (0.007)	0.007 (0.006)	0.007 (0.007)	0.013 (0.009)	0.007 (0.006)
Miravalles	0.007 (0.007)	0.964 (0.015)	0.008 (0.008)	0.007 (0.007)	0.007 (0.007)	0.007 (0.007)
Santuario	0.007 (0.007)	0.007 (0.007)	0.960 (0.015)	0.007 (0.007)	0.013 (0.009)	0.007 (0.007)
Pedro Moreno	0.015 (0.014)	0.014 (0.014)	0.261 (0.028)	0.681 (0.014)	0.015 (0.014)	0.015 (0.014)
Rancho Eureka	0.011 (0.010)	0.268 (0.024)	0.011 (0.010)	0.011 (0.011)	0.688 (0.014)	0.011 (0.011)
Sierra de Quila	0.007 (0.007)	0.007 (0.007)	0.007 (0.007)	0.007 (0.006)	0.007 (0.007)	0.967 (0.014)
B	Nayarit	Western Jalisco	Soledad	Sierra de Quila		
Nayarit	0.9851 (0.008)	0.005 (0.005)	0.005 (0.005)	0.005 (0.005)		
Western Jalisco	0.005 (0.005)	0.985 (0.009)	0.005 (0.005)	0.005 (0.005)		
Soledad	0.007 (0.007)	0.007 (0.007)	0.980 (0.011)	0.007 (0.007)		
Sierra de Quila	0.007 (0.007)	0.007 (0.007)	0.007 (0.007)	0.979 (0.012)		

Proportions of migrants per generation in row, from column estimated using BayesAss. Mean (SD) across loci. Bold cells are those with non-trivial estimates of migration

Fig. 6 Scatter plot showing the position of *D. tomasel-lii* individuals along the two linear discriminant axes from a discriminant analysis of principal components of 24 morphological traits from six localities in Jalisco and Nayarit. Locality of individuals is indicated by symbols (see legend) and polygons enclose individuals from each locality



for differences in dispersion among populations was also significant; however, comparing the difference in centroid position among the three pairs with significantly different dispersions (t-test, $p < 0.05$; Table S8), it seems unlikely that the difference in centroids is a false positive. Post-hoc pairwise PERMANOVA tests found significant differences between each pair of centroids (Table S9; Fig. 6). The contributions of the original variables to the linear functions are given in Table S10. We compared these values to the expected value, assuming equal contribution among traits, and several traits had outsized contributions. Along the first linear function, leaf length, leaflet width, and leaflet length/width were above this threshold. These traits most strongly separate Soledad and Sierra de Quila from the Western Jalisco and Miravalles populations. In the second linear function, petiole width, petiole length, leaf width, leaflet curve ratio, falcate distance, and leaf length (in descending order) were above the threshold and separated Carrizal and Soledad from the other populations according to their relative positions.

To more directly test for differences in specific traits among populations, we used the Kruskal-Wallis test among

all populations and the Dunn test between population pairs. Significant differences among populations were found for all traits measured according to the Kruskal-Wallis test (Table S11). We compiled the results from the Dunn pairwise tests in three ways. First, violin plots of each trait showing the significant differences between populations are found in Figures S4-S27. Second, we tallied the number of significant comparisons for each trait (Table S12), and third, we tallied the number of traits that were significantly different between population pairs (Table S13). Raw P-values for pairwise comparisons of every trait are in Table S14. Ranking traits by the number of significant pairwise comparisons, we find that most of the traits that had an outsized contribution to the linear discriminant functions also vary among most populations.

In general, Sierra de Quila differs from most populations in the highest number of characters (11–14 significant differences), except Soledad. Sierra de Quila and Soledad differ from each other in the fewest traits (4 significant differences) and share a number of traits that distinguish them from the other populations. They tend to have straighter, shorter leaves with glabrous, narrower leaflets, more leaflets

per meter, and the highest leaflet length to width ratio. They differ only in four traits, with Soledad having shorter, wider petioles; more decurved leaflets; and a more tomentose rachis.

Among the other three populations, Western Jalisco and Carrizal have the most deflexed leaflets and Miravalles is intermediate among all populations. Western Jalisco and Miravalles both tend to have somewhat pubescent and moderately falcate leaflets and long petioles, and Western Jalisco also has the greatest forward leaflet angle. Miravalles tends to have more widely spaced leaflets than Western Jalisco, less recurved leaflets than Western Jalisco and Carrizal, and fewer leaflet spines overall. Western Jalisco has the lowest leaflet length to width ratio, while Carrizal and Miravalles are intermediate among all populations. Carrizal is characterized by shorter, wider petioles similar in length to Soledad, and a consistently tomentose rachis.

Discussion

In this study we examined the genomic and morphological variation within the southern populations of *Dioon tomasellii* to identify independent evolutionary lineages and to characterize the conservation genomic status of populations. Our data focus on the states of Nayarit and Jalisco, the region surrounding the type locality (De Luca et al. 1984), and so represent an initial study until we are able to include the northernmost populations in Sinaloa and Durango. We found clear population genetic structure and evidence of local adaptation. We also find complementary patterns in morphological differentiation among populations, with several characters distinguishing each. Genetic diversity within populations is moderate, with no evidence of inbreeding but with a very high percentage of private alleles in each. Taken together, these results suggest that there are multiple evolutionary lineages represented by *D. tomasellii* in its southern range.

Drivers of divergence among populations of *D. tomasellii*

Our results provide two possible complementary mechanisms that have produced these genetically distinct, isolated, and relatively small lineages in Nayarit and Jalisco. First, as suggested by the very high percentage of private neutral alleles in each population and confirmed by our estimates of migration rates, there is apparently no gene flow among any of the four recognized populations. We did detect migration between sampling localities within the Nayarit and Western Jalisco populations (Appendix Table 6), but these localities are geographically quite close (1.5–5.5 km apart), whereas

the distances between the four populations range from approximately 100–200 km. This distribution pattern, the high levels of allelic differentiation, and the geographically limited migration rates among localities likely result from a fundamental limit to gene flow in cycads that is the result of pollination and seed dispersal mechanisms. Pollination in *Dioon* is by weevils (Tang et al. 2018) whose maximum flight distance is not known, but Terry et al. (2001) found that over half the mean pollen load is lost from the thrip pollinators between male and female cones of *Macrozamia communis*, with a mean distance of 5 m. If pollen retention is similar between these pollinators, this suggests that effective pollination is highly restricted among populations of *D. tomasellii*. Further, the dispersal mechanism for the relatively large seeds of cycads is not known for most species, but several studies have suggested that birds and/or mammals may provide some limited dispersal (Snow and Walter 2007; Hall and Walter 2013; Yáñez-Espinosa et al. 2021; Gómez-Domínguez et al. 2022). However, even when animal dispersal has been observed, the vast majority of seeds remain close to the mother plant. This suggests that long-distance seed dispersal is uncommon, but it cannot be discarded as impossible. However, the lack of evidence for recent gene flow among geographically distant localities and only partial connectivity between close ones, in this and previous studies, is consistent with limited seed dispersal and pollination distance in cycads (Gutiérrez-Ortega et al. 2020a, b; Gutiérrez-Ortega et al. 2021a, b; Clugston et al. 2022; Dorsey et al. 2023). Despite the evidence for recent population reductions, the levels of differentiation between populations seem unlikely to have arisen since the bottleneck event, which is inferred to be 2Ne–4Ne generations in the past (Piry et al. 1999). It is therefore likely that these populations have been isolated for a long time or that there was a significant level of isolation by distance within the ancestral metapopulation, or both. The fundamental limit on metapopulation connectivity due to pollination and dispersal biology may be one of the main drivers of high spatial turnover of *Dioon* species across Mexico (Rossetto et al. 2008; Wessinger 2021), a pattern that is becoming more evident within the genus as more taxa are re-examined using genomic tools.

We also find evidence of adaptive divergence among populations from several analyses. The non-neutral data set clustered individuals in the same groups as the neutral set and we find a slight increase in measures of population divergence (G_{ST} , A_{ST} , D_{Jost}) for the non-neutral data set. Also, the percentage of private, non-neutral alleles is even higher than that of the neutral set for all populations, ranging from 21 to 36%. These results, along with the morphological variation discussed below, suggest that the population divergence is driven not only by

demographic processes but also local adaptation. If true, this would be in contrast to previous studies in *Dioon* that found niche conservatism to be a driving factor in speciation among closely distributed taxa (Gutiérrez-Ortega et al. 2020a, b; Gutiérrez-Ortega et al. 2021a, b). Further tests of the extent and drivers of local adaptation should focus on environmental and morphological correlates of genetic diversification across the species' entire range.

Genetic diversity within populations

The genomic clusters/populations identified here have moderate and comparable levels of neutral genetic diversity as measured by allelic richness (A , mean 2.32, range=2.16–2.49), observed heterozygosity (H_O , mean=0.14, range=0.13–0.16) and gene diversity (H_S , mean=0.16 range=0.14–0.18). Previous studies have used various types of markers to estimate genetic diversity in *Dioon* and this makes comparison difficult. Dorsey et al. (2023) used the same microhaplotypes derived from RAD-seq data as in this study for *D. holmgrenii* and found similar estimates of allelic richness but twice the level of H_O and H_S and substantial yet smaller levels of private alleles (range=0.015–0.039). *D. holmgrenii* also comprises multiple isolated populations but this distribution apparently has not resulted in the same level of divergence as we find in *D. tomasellii*.

Estimates of inbreeding (F) are near zero for all populations of *D. tomasellii* in the present study, which is somewhat surprising given the likely bottleneck experienced in at least half of the populations and the lack of gene flow among all. One explanation is that cycads are dioecious and less susceptible to inbreeding due to selfing. However, we would expect the low dispersibility and limited pollination distance discussed above to produce clusters of closely related individuals. If related individuals are spatially clustered within populations, one could imagine that, rather than increasing the level of inbreeding, a reduction in population size that is evenly distributed in effect could potentially decrease inbreeding by thinning the population such that a less dense cohort of individuals that are more distantly related remains. Also, given the slow growth and presumed long generation times in cycads, the effect of the bottleneck may be delayed, and our estimates of inbreeding may actually be that of the populations before reductions in size. One could test this by comparing coancestry among individuals and inbreeding in populations that have and haven't experienced a recent bottleneck. Unfortunately, both species for which we have genomic data have experienced a recent reduction in population sizes.

Morphological characterization and differentiation

Having established that the populations sampled are genetically differentiated to various degrees, we would like to know if they are also distinct morphologically. This will facilitate further work addressing local adaptation and help us to evaluate our suspicions of multiple potential species within *D. tomasellii* as currently defined. To get at this question, we measured or calculated 24 leaf traits, many of which are typically used to identify species of *Dioon*. This data set included an additional population, Carrizal, from southwest Jalisco, and excluded Rancho Eureka. The multivariate PERMANOVA analysis confirms that the localities can be distinguished by a combination of the traits we measured. Pairwise PERMANOVAs suggest that all populations are morphologically distinct; however, the statistical difference between Miravalles and Western Jalisco populations is surprising given the overlap among individuals in the linear discriminant space.

Multivariate tests are helpful to determine if the clustering we infer from LDA plots is non-random and can help develop hypotheses of evolutionary trends. But for investigating the evolution of specific characters and the identification of distinct morphologies in the field, we would like to know the particular characters, if any, that distinguish populations. To this end, it is helpful to consider the variation of traits within populations as well as the differences in central tendencies. Reviewing the Dunn test violin plots (Figs. S4–S27), we find that several traits tend to be more consistent within and different between some populations. Traits that distinguish Sierra de Quila, and slightly less so Soledad, and which are less variable within these populations are leaflet width, leaflet length/width, abaxial leaflet surfaces, and leaflet curve ratio, and to a lesser degree, leaf length. Sierra de Quila has consistently thin, glabrous, straight leaflets compared to the western populations. Soledad also has thinner, glabrous leaflets, but they can be deflexed, more similar to the western populations. Leaf length is generally shorter in these eastern populations, but they do overlap with the lower end of the distribution in western populations. Given the history of population divergence, characters that separate Sierra de Quila and Soledad from the western populations may represent retention of ancestral traits. Among the three western populations, while Carrizal does occupy a fairly distinct position in the dapc plot, there are few characters that do not overlap, despite any significant differences between medians. One trait that may distinguish Carrizal from Miravalles and Western Jalisco plants is rachis tomentum. Carrizal plants consistently had tomentose rachis, whereas vestiture could range from glabrous to tomentose in the other two western populations.

Taxonomic and conservation implications

The taxonomy of the genus *Dioon* continues to be refined, mostly due to reevaluation of taxa using new genomic, morphological, and environmental data (Gutiérrez-Ortega et al. 2018a, b; Gutiérrez-Ortega et al. 2020a, b; Gutiérrez-Ortega et al. 2021a, b; Calonje et al. 2026). An emerging pattern from these studies is that the distributions of *Dioon* species are quite limited. This seems to be the case for the populations in this study as well, despite being included in one of the most widespread taxa as originally circumscribed (de Luca et al. 1984). We have demonstrated that there are indeed multiple evolutionary lineages among the southern populations of *D. tomasellii* and that at least two of these, Soledad and Sierra de Quila, can be distinguished morphologically. Accordingly, we are in the process of confirming that these two populations should be recognized as new species. However, splitting taxa can have important effects on their conservation status (Morrison et al. 2009). *Dioon tomasellii* is currently assessed as Vulnerable on the IUCN Red List under criterion A2 because of population declines and loss of habitat (IUCN 2025). Any segregate taxa will have smaller population sizes, smaller extent of occurrence, and fewer populations, all potentially increasing their threat level. To the extent that conservation efforts can address the needs of these local endemic lineages, this recognition of increased threat represents progress. However, because the populations of *D. tomasellii* are isolated and genetically distinct but retain moderate levels of genetic diversity, we feel they should each be considered irreplaceable and high priority for conservation, regardless of taxonomy. This is further supported by the potentially adaptive genetic divergence among populations, which means that the loss of any one

would mean a reduction, not only in the number of living individuals, but also of ecological and phylogenetic diversity within the genus. Fortunately, Sierra de Quila population is within the Sierra de Quila Flora and Fauna Protection Area and so has a level of protection and the Santuario locality is on a private reserve. Ther remaining populations are on private land and so potentially subject to land clearing or other habitat changes.

Conclusions

We have shown here that the populations of *Dioon tomasellii* in the southern portion of its range are isolated and genetically distinct. These populations harbor high proportions of unique alleles, the maintenance of which is likely due to the absence of pollen flow and seed dispersal and local adaptation. Further, we find that two of the populations, Sierra de Quila and Soledad, diverge morphologically, particularly from Western Jalisco and Nayarit populations, with Carrizal plants being intermediate. Notably, Sierra de Quila occupies the most distinct and narrow section of morphospace and is the first diverging lineage in the phylogeny. This strongly suggests to us that it may be a distinct taxon. Future work will incorporate populations from the northern half of the range (i.e., northern Nayarit and Sinaloa) to determine the extent of divergence across the distribution and to further evaluate the possibility that this taxon consists of multiple species. Given the moderate levels of heterozygosity and the very high proportion of private alleles in each population, conservation of each population at current or higher numbers of individuals will be necessary to prevent the loss of genetic, ecological an evolutionary diversity represented by these populations.

Appendix 1

Table 6 Voucher information for tissue sampled for DNA extraction from populations of *Dioon tomasellii*

Locality	N DNA	N Morph.	Voucher	Deposited
Santuario	48	28	BLD 308	HNT
			BLD 309	SERBO
Pedro Moreno	18	14	-	-
Miravalles	40	25	BLD 314	HNT
			BLD 315	HNT
Rancho Eureka	25	-	BLD 316	HNT
Sierra de Quila	44	23	BLD 310	HNT
			BLD 311	HNT
			BLD 312	HNT
Soledad	48	25	BLD 313	HNT
			-	-
Carrizal	-	30	-	-

N DNA number of samples for DNA extraction, N Morph. number of individuals measured for morphological traits, - no data or voucher

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10592-026-01845-z>.

Acknowledgements We thank Alejandra Flores-Argüelles, Heather Huggins, and Neil Gerlowski for assistance in the field. We also thank Silvia Salas-Morales and Vallarta Botanic Garden for logistical support.

Author contributions B.D. designed the study; B.D., T.G., and M.P. collected data, B.D. and J.G. analyzed the data, B.D. wrote the initial manuscript, all authors edited and approved the final manuscript.

Funding The authors have no relevant financial or non-financial interests to disclose.

Data availability Raw sequence data are available at the NIH SRA under BioProject PRJNA1543029. R code for data filtering and analyses is available at <https://github.com/dorseyb/D.-tomasellii-Conservation-Genomics>.

Declarations

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

- Anderson MJ (2001) A new method for non-parametric multivariate analysis of variance. *Austral Ecol* 26:32–46. <https://doi.org/10.1111/j.1442-9993.2001.01070.pp.x>
- Anderson MJ, Walsh DCI (2013) PERMANOVA, ANOSIM, and the Mantel test in the face of heterogeneous dispersions: What null hypothesis are you testing? *Ecol Monogr* 83:557–574. <https://doi.org/10.1890/12-2010.1>
- Berli P, Felsenstein J (1999) Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. *Genetics* 152:763–773. <https://doi.org/10.1093/genetics/152.2.763>
- Berli P, Felsenstein J (2001) Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. *Proc Natl Acad Sci* 98:4563–4568. <https://doi.org/10.1073/pnas.081068098>
- Bushnell B (2019) BBMap. <https://sourceforge.net/projects/bbmap/>. 7 Apr. 2019
- Cabrera-Toledo D, González-Astorga J, Carlos Flores-Vazquez J (2012) Fine-scale spatial genetic structure in two Mexican cycad species *Dioon caputoi* and *Dioon merolae* (Zamiaceae, Cycadales): Implications for conservation. *Biochem Syst Ecol* 40:43–48. <https://doi.org/10.1016/j.bse.2011.09.004>
- Calonje M, Stevenson DW, Osborne R (2026) The world list of cycads, online edition. <https://cycadlist.org/index.php>. 10 July 2025
- Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: an analysis tool set for population genomics. *Mol Ecol* 22:3124–3140. <https://doi.org/10.1111/mec.12354>
- Chemnick J, Gregory TJ, Salas-Morales SH (1997) A revision of *Dioon tomaselli* (Zamiaceae) from western Mexico, a range extension of *D. morolae*, and clarification of *D. purpusii*. *Phytologia* 83:1–6
- Clugston JAR, Ruhsam M, Kenicer GJ, Henwood M, Milne R, Nagalingum NS (2022) Conservation genomics of an Australian cycad *Cycas calicicola*, and the absence of key genotypes in botanic gardens. *Conserv Genet* 23:449–465. <https://doi.org/10.1007/s10592-022-01428-8>
- da Silva JM, Donaldson JS, Reeves G, Hedderson TA (2012) Population genetics and conservation of critically small cycad populations: a case study of the Albany Cycad, *Encephalartos latifrons* (Lehmann). *Biol J Linn Soc* 105:293–308. <https://doi.org/10.1111/j.1095-8312.2011.01801.x>
- Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat Rev Genet* 12:499–510. <https://doi.org/10.1038/nrg3012>
- de Luca P, Sabato S, Torres MV (1984) *Dion tomasellii* (Zamiaceae), a new species with two varieties from western Mexico. *Brittonia* 36:223–227. <https://doi.org/10.2307/2806511>
- Dorsey BL, Salas-Morales SH, Gregory TJ (2023) Conservation genomics of *Dioon holmgrenii* (Zamiaceae) reveals a history of range expansion, fragmentation, and isolation of populations. *Conserv Genet*. <https://doi.org/10.1007/s10592-023-01569-4>
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. *Mol Ecol* 14:2611–2620. <https://doi.org/10.1111/j.1365-294X.2005.02553.x>
- Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164:1567–1587. <https://doi.org/10.1093/genetics/164.4.1567>
- Frankham R, Bradshaw CJA, Brook BW (2014) Genetics in conservation management: revised recommendations for the 50/500 rules, red list criteria and population viability analyses. *Biol Conserv* 170:56–63. <https://doi.org/10.1016/j.biocon.2013.12.036>
- Garner BA, Hoban S, Luikart G (2020) IUCN Red List and the value of integrating genetics. *Conserv Genet* 21:795–801. <https://doi.org/10.1007/s10592-020-01301-6>
- GBIF.org (2025) *Dioon* occurrence download GBIF. <https://doi.org/10.15468/DL.KJY8XB>
- Gómez-Domínguez H, Hernández-Tapia JE, Ortiz-Rodríguez AE (2022) Seed predation and potential seed dispersers of the narrow endemic *Ceratozamia norstogii* (Zamiaceae). *Biodivers Data J* 10:e86007. <https://doi.org/10.3897/BDJ.10.e86007>
- González-Astorga J, Vergara-Silva F, Vovides AR, Nicolalde-Morejón F, Cabrera-Toledo D, Pérez-Farrera MA (2008a) Diversity and genetic structure of three species of *Dioon* Lindl. (Zamiaceae, Cycadales) from the Pacific seaboard of Mexico. *Biol J Linn Soc* 94:765–776. <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 sonorensis* (Zamiaceae) from Sonora, Mexico: Evolutionary and conservation implications. *Biochem Syst Ecol* 36:891–899. <https://doi.org/10.1016/j.bse.2008.11.006>
- Goudet J, Jombart T (2018) hierfstat: estimation and tests of hierarchical F-statistics
- Gutiérrez-Ortega JS, Jiménez-Cedillo K, Pérez-Farrera MA, Vovides AP et al (2018a) Considering evolutionary processes in cycad conservation: identification of evolutionarily significant units within *Dioon sonorensis* (Zamiaceae) in northwestern Mexico. *Conserv Genet* 1–13. <https://doi.org/10.1007/s10592-018-1079-2>

- Gutiérrez-Ortega JS, Jiménez-Cedillo K, Pérez-Farrera MA, Martínez JF et al (2018b) Species definition of *Dioon sonorense* (Zamiaceae, Cycadales), and description of *D. vovidesii*, a new cycad species from northwestern Mexico. *Phytotaxa* 369:107–114. <https://doi.org/10.11646/phytotaxa.369.2.4>
- Gutiérrez-Ortega JS, Farrera MA, Vovides A, Salas S, Chemnick J (2020a) *Dioon oaxacensis* (Zamiaceae): a new cycad species from the arid central valleys of Oaxaca (Mexico). *Phytotaxa* 474:51–61. <https://doi.org/10.11646/phytotaxa.474.1.5>
- Gutiérrez-Ortega JS, Salinas-R M, Ito T et al (2020b) Niche conservatism promotes speciation in cycads: the case of *Dioon merolae* (Zamiaceae) in Mexico. *New Phytol* 227. <https://doi.org/10.1111/nph.16647>
- Gutiérrez-Ortega JS, Molina-Freaner F, Martínez JF et al (2021a) Speciation along a latitudinal gradient: The origin of the Neotropical cycad sister pair *Dioon sonorense*–*D. vovidesii* (Zamiaceae). *Ecol Evol* 11:6962–6976. <https://doi.org/10.1002/ece3.7545>
- Gutiérrez-Ortega JS, Pérez-Farrera MA, Chemnick J, Gregory TJ (2021b) A reassessment of *Dioon merolae* (Zamiaceae) leads to the description of *Dioon salas-moralesae*, a new cycad species from Southeastern Oaxaca, Mexico. *Phytotaxa* 528:93–110. <https://doi.org/10.11646/phytotaxa.528.2.3>
- Hall JA, Walter GH (2013) Seed dispersal of the Australian cycad *Macrozamia miquelii* (Zamiaceae): Are cycads megafauna-dispersed grove forming plants? *Am J Bot* 100:1127–1136. <https://doi.org/10.3732/ajb.1200115>
- Hoang DT, Chernomor O, Von Haeseler A, Minh BQ, Vinh LS (2018) UFBBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol* 35:518–522. <https://doi.org/10.1093/molbev/msx281>
- Hoban S, Callicrate T, Clark J et al (2020) Taxonomic similarity does not predict necessary sample size for ex situ conservation: a comparison among five genera. *Proc Royal Soc B: Biol Sci* 287:20200102. <https://doi.org/10.1098/rspb.2020.0102>
- Hoban S, da Silva JM, Hughes A et al (2024) Too simple, too complex, or just right? Advantages, challenges, and guidance for indicators of genetic diversity. *Bioscience* 74:269–280. <https://doi.org/10.1093/biosci/biae006>
- Holm S (1979) A Simple Sequentially Rejective Multiple Test Procedure. *Scand J Stat* 6:65–70
- IUCN (2025) The IUCN red list of threatened species. <https://www.iucnredlist.org/en>. 10 July 2025
- Jombart T, Ahmed I (2011) adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* 27:3070–3071. <https://doi.org/10.1093/bioinformatics/btr521>
- Jombart T, Devillard S, Ballou F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet* 11:94. <https://doi.org/10.1186/1471-2156-11-94>
- Kalyanamoorthy S, Minh BQ, Wong TKF, Von Haeseler A, Jermini LS (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 14:587–589. <https://doi.org/10.1038/nmeth.4285>
- 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. <https://doi.org/10.3389/fgene.2015.00208>
- Kardos M, Armstrong EE, Fitzpatrick SW et al (2021) The crucial role of genome-wide genetic variation in conservation. *Proc Natl Acad Sci* 118:e2104642118. <https://doi.org/10.1073/pnas.2104642118>
- Karunarathne P, Zhou Q, Schliep K, Milesi P (2023) A comprehensive framework for detecting copy number variants from single nucleotide polymorphism data: ‘rCNV’, a versatile R package for paralogue and CNV detection. *Mol Ecol Resour* 23:1772–1789. <https://doi.org/10.1111/1755-0998.13843>
- Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30:772–780. <https://doi.org/10.1093/molbev/mst010>
- Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 30:3059–3066. <https://doi.org/10.1093/nar/gkf436>
- Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I (2015) Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Mol Ecol Resour* 15:1179–1191. <https://doi.org/10.1111/1755-0998.12387>
- Lindstrom A, Habib S, Dong S et al (2024) Transcriptome sequencing data provide a solid base to understand the phylogenetic relationships, biogeography and reticulated evolution of the genus *Zamia* L. (Cycadales: Zamiaceae). *Ann Botany* 134:747–768. <https://doi.org/10.1093/aob/mcae065>
- 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* 19:1095–1100. <https://doi.org/10.1111/1755-0998.13015>
- Malinsky M, Trucchi E, Lawson DJ, Falush D, Takezaki N (2018) RADpainter and fineRADstructure: Population Inference from RADseq Data. *Mol Biol Evol* 35:1284–1290. <https://doi.org/10.1093/molbev/msy023>
- Marinez Arbizu P (2020) pairwiseAdonis: pairwise multilevel comparison using adonis
- Mastretta-Yanes A, Arrigo N, Alvarez N, Jorgensen TH, Piñero D, Emerson BC (2015) Restriction site-associated DNA sequencing, genotyping error estimation and de novo assembly optimization for population genetic inference. *Mol Ecol Resour* 15:28–41. <https://doi.org/10.1111/1755-0998.12291>
- McArdle BH, Anderson MJ (2001) Fitting Multivariate Models to Community Data: A Comment on Distance-Based Redundancy Analysis. *Ecology* 82:290–297. [https://doi.org/10.1890/0012-9658\(2001\)082\[0290:FMMTCJ\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2001)082[0290:FMMTCJ]2.0.CO;2)
- Minh BQ, Schmidt HA, Chernomor O et al (2020) IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol* 37:1530–1534. <https://doi.org/10.1093/molbev/msaa015>
- Morrison WR, Lohr JL, Duchon P et al (2009) The impact of taxonomic change on conservation: Does it kill, can it save, or is it just irrelevant? *Biol Conserv* 142:3201–3206. <https://doi.org/10.1016/j.biocon.2009.07.019>
- Nicolalde-Morejón F, Vergara-Silva F, González-Astorga J, Vovides AP, Monteros AE (2009) Reciprocal illumination of morphological characters upon a molecular hypothesis supports the proposal of a new species of cycad from Mexico. *Syst Biodivers* 7:73–79. <https://doi.org/10.1017/S1472200008002879>
- Oksanen J, Simpson GL, Blanchet FG et al (2024) vegan: community ecology package
- Pérez-Farrera MA, Ramirez-Oviedo SM, Martínez-Martínez MG, Moreno Mendez G, Rocha Loredó AG, Gutiérrez-Ortega JS (2024) *Ceratozamia chinantlensis* (Zamiaceae): A New Cycad Species from La Chinantla, Oaxaca, Mexico. *Taxonomy* 4:733–747. <https://doi.org/10.3390/taxonomy4040039>
- 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* 90:502–503. <https://doi.org/10.1093/jhered/90.4.502>
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959

- R Core Team (2024) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria
- Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA (2018) Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst Biol* 67:901–904. <https://doi.org/10.1093/sysbio/syy032>
- Rochette NC, Rivera-Colón AG, Catchen JM (2019) Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. *Mol Ecol* 28:4737–4754. <https://doi.org/10.1111/mec.15253>
- Rossetto M, Kooyman R, Sherwin W, Jones R (2008) Dispersal limitations, rather than bottlenecks or habitat specificity, can restrict the distribution of rare and endemic rainforest trees. *Am J Bot* 95:321–329. <https://doi.org/10.3732/ajb.95.3.321>
- Snow E, Walter G (2007) Large seeds, extinct vectors and contemporary ecology: testing dispersal in a locally distributed cycad, *Macrozamia lucida* (Cycadales). *Aust J Bot* 55:592–600. <https://doi.org/10.1071/BT07009>
- Tang W, Xu G, O'Brien CW et al (2018) Molecular and Morphological Phylogenetic Analyses of New World Cycad Beetles: What They Reveal about Cycad Evolution in the New World. *Diversity* 10:38. <https://doi.org/10.3390/d10020038>
- Terry I (2001) Thrips and Weevils as Dual, Specialist Pollinators of the Australian Cycad *Macrozamia communis* (Zamiaceae). *Int J Plant Sci* 162:1293–1305. <https://doi.org/10.1086/321929>
- Wang X-H, Li J, Zhang L-M et al (2019) Population differentiation and demographic history of the *Cycas taiwaniana* complex (Cycadaceae) Endemic to South China as Indicated by DNA sequences and microsatellite markers. *Front Genet* 10. <https://doi.org/10.3389/fgene.2019.01238>
- Wessinger CA (2021) From pollen dispersal to plant diversification: genetic consequences of pollination mode. *New Phytol* 229:3125–3132. <https://doi.org/10.1111/nph.17073>
- Wickham H (2016) ggplot2: Elegant Graphics for Data Analysis. Springer International Publishing: Imprint: Springer, Cham
- Wilson GA, Rannala B (2003) Bayesian inference of recent migration rates using multilocus genotypes. *Genetics* 163:1177–1191
- Winter DJ (2012) mmod: an R library for the calculation of population differentiation statistics. *Mol Ecol Resour* 12:1158–1160. <https://doi.org/10.1111/j.1755-0998.2012.03174.x>
- Yáñez-Espinosa L, Barragán-Torres F, Hernández ABI, Morales JI (2021) Dispersal of *Dioon edule* cycad seeds by rodents in a tropical oak forest in Mexico. *Trop Zool* 34. <https://doi.org/10.4081/tz.2021.105>
- Zhang C, Mirarab S (2022) Weighting by gene tree uncertainty improves accuracy of quartet-based species trees. *Mol Biol Evol* 39:msac215. <https://doi.org/10.1093/molbev/msac215>
- Zhang C, Sayyari E, Mirarab S (2017) ASTRAL-III: increased scalability and impacts of contracting low support branches. *Comparative Genomics. Lecture Notes in Computer Science*. Springer, Cham, pp 53–75

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.