2022

Forest cover and geographic distance influence fine-scale genetic structure of leaf-toed geckos in the tropical dry forests of western Mexico

Connor M. French
CUNY Graduate Center

Casey-Tyler Berezin
CUNY City College

Isaac Overcast
CUNY Graduate Center

Fausto R. Méndez De La Cruz
Universidad Nacional Autonoma de Mexico

Saptarsi Basu
CUNY New York City College of Technology

See next page for additional authors

Forest cover and geographic distance influence fine-scale genetic structure of leaf-toed geckos in the tropical dry forests of western Mexico

Running title: Spatial genomics of Mexican geckos

Connor M. French[1*], Casey-Tyler Berezin[2*], Isaac Overcast[1,3,4], Fausto R. Méndez de la Cruz[5], Saptarsi Basu[6], Roberto Lhemish Martínez Bernal[5], Robert W. Murphy[7], Michael J. Hickerson[1,2], Christopher Blair[1,6]

[1]Biology PhD Program, CUNY Graduate Center, 365 5th Ave., New York, NY 10016

[2]Department of Biology, City College of New York, 160, Convent Avenue, New York, NY, 10031 USA

[3]Institut de Biologie de l'Ecole Normale Superieure, 46 Rue d'Ulm, 75005 Paris

[4]Division of Vertebrate Zoology, American Museum of Natural History, 200 Central Park West, New York, NY 10024

[5]Instituto de Biología, UNAM. Ciudad Universitaria, CdMx. C.P. 04510, A.P. 70-153. México.

[6]Department of Biological Sciences, New York City College of Technology, The City University of New York, 285 Jay Street, Brooklyn, NY 11201

[7]Centre for Biodiversity, Royal Ontario Museum, 100 Queen's Park, Toronto, ON M5S 2C6

**Joint first author*

Corresponding author: Christopher Blair: cblair@citytech.cuny.edu

ABSTRACT

The biodiversity within tropical dry forests (TDF) is astounding and yet poorly cataloged due to inadequate sampling and the presence of cryptic species. In the Mexican TDF, endemic species are common, and the landscape has been continually altered by geologic and anthropogenic changes. To understand how landscape and environmental variables have shaped the population structure of endemic species, we study the recently described species of leaf-toed gecko, *Phyllodactylus benedettii*, in coastal western Mexico. Using ddRADseq data, we first explore population structure and estimate the number of ancestral populations. Results indicate a high degree of genetic structure with little admixture, and patterns corresponding to both latitudinal and altitudinal gradients. We find that genetic structure cannot be explained purely by geographic distance, and that ecological corridors may facilitate dispersal and gene flow. We then model the spatial distribution of *P. benedettii* in the TDF through time and find that the coastline has been climatically suitable for the species since the last glacial maximum (LGM). Landscape genetic analyses suggest that the combined influence of isolation by distance (IBD) and isolation by resistance (IBR; forest cover) influence the spatial genetic structure of the species. Overall, our genomic data demonstrate fine-scale population structure in TDF habitat, a complex colonization history, and spatial patterns consistent with both IBD and other ecological factors. These results further highlight the Mexican TDF as a diversity hotspot and suggest that continued anthropogenic changes are likely to impact native fauna.

KEYWORDS

climate, landscape genetics, *Phyllodactylus*, population structure, tropical dry forest

INTRODUCTION

A fundamental goal for evolutionary biology and landscape genetics involves documenting fine-scale patterns of population genetic structure and elucidating the geographic and ecological causes of these patterns (Manel et al., 2003; Storfer et al., 2007; Holderegger & Wagner, 2008; Sork & Waits, 2010; Petren, 2013; Storfer et al., 2018). This not only informs us about how diversity is generated and maintained in nature, but also is fundamental to conservation efforts (Keller et al., 2015; Bowman et al., 2016). Geographic (Euclidean) distance between samples or populations (IBD; Wright, 1943) and both landscape and environmental differences can influence patterns of gene flow in diverse taxa (Storfer et al., 2010; Wang, 2012; Wang et al., 2013; Sexton et al., 2014; Wang & Bradburd, 2014). Furthermore, the incorporation of ecological niche models (ENMs) into landscape genetic studies can provide researchers with the tools to assess the relative importance of IBD versus the contemporary or historical climate in shaping patterns of genetic structure (Ortego et al., 2012; Oliveira et al., 2018). Since landscape genetics emerged in 2003 (Manel et al., 2003), numerous studies have tested hypotheses regarding landscape and environmental effects on gene flow (see Storfer et al., 2010; Manel & Holderegger, 2013; Storfer et al., 2018 for review). However, the vast majority of studies have focused on taxa inhabiting temperate environments, with few targeting species living in diverse yet threatened tropical and subtropical habitats (Storfer et al., 2010; Rico, 2019).

The preservation of global biodiversity is essential (Pimm, 1995). Empirical studies and conservation efforts in tropical latitudes have typically focused on rainforests, but tropical dry forests (TDFs) are similarly threatened and more poorly understood (Mooney, 1995). TDFs are distributed in tropical regions throughout the world and are recognized as hyper-diverse hubs for endemic plants, mammals, insects, and reptiles (Janzen, 1988). The expansive TDF in Mexico,

which formed roughly 20 to 30 million years ago (Becerra, 2005), has been understudied both geographically and taxonomically, precluding a thorough understanding of evolutionary patterns and processes throughout the biome. The current genomics revolution has substantial potential to increase our power to document fine-scale genetic structure and test alternative historical and contemporary evolutionary hypotheses impacting species inhabiting this system.

Mexico is home to 8.7% of the world's reptiles (Flores-Villela & Garcia-Vazquez, 2014), but Neotropical lowland taxa have received relatively little attention as compared to montane biota, and most studies have tested hypotheses at the phylogenetic or phylogeographic level (e.g. Devitt, 2006; Zarza et al. 2008; Bryson et al., 2011a, b; Bryson & Riddle, 2012; Blair et al., 2015, Blair et al., 2022). However, evidence is accumulating that suggests the presence of cryptic, ancient lineages within widespread TDF taxa (Devitt, 2006; Zarza et al. 2008; Blair et al., 2013, 2015). Anthropogenic alteration of TDF drives an urgent need to characterize the region's biodiversity and test hypotheses regarding fine-scale patterns of population structure, which is further exemplified by the introduction of non-native species (Trejo & Dirzo, 2000).

Few landscape genetic studies have focused on taxa inhabiting Mexican TDF (Rico, 2019). Using microsatellites, Blair et al. (2013) found that multiple landscape and climatic variables played critical roles in shaping patterns of gene flow in a leaf-toed gecko species (*Phyllodactylus*) in northwestern Mexico. To get a better understanding of landscape genetic relationships throughout the TDF, we utilize genomic data from the recently described *P. benedettii* (Ramirez-Reyes et al., 2018), which is endemic to Jalisco, Mexico. Most diversification within the *P. lanei* complex, of which *P. benedettii* is a member, dates to the Miocene and Pliocene epochs when both the Sierra Madre Occidental (SMO) and the Trans-Mexican Volcanic Belt (TMVB) were forming (Blair et al., 2014 & 2015; Ramirez-Reyes et al.,

2020), although some uncertainties remain (Ramirez-Reyes et al. 2017). Beyond introducing elevational and climatic gradients, uplifting of the SMO increased the extent of the Mexican TDF overall (Becerra, 2005), and the formation of the TMVB caused higher diversification rates in many species (Blair et al., 2015; Bryson & Riddle, 2012; Ruiz-Sanchez & Specht, 2013; Zarza et al., 2018). Southwestern Mexico, in particular, is a diversity hotspot with many taxa exhibiting relatively high diversification rates (Becerra & Venable, 2008). Although an increasing number of empirical studies are beginning to shed light into historical evolutionary patterns and processes in the region, few have focused on testing hypotheses on more contemporary ecological timescales.

We use double-digest restriction site-associated DNA sequencing data (ddRADseq; Peterson et al., 2012) from *P. benedettii* to further characterize spatial patterns of molecular diversity throughout the region and determine if statistical models favor the inclusion of landscape and environmental variables over pure IBD (our null model). We specifically use our ddRADseq data set to test three primary hypotheses: (1) population structure occurs over a relatively fine spatial scale in southwestern Mexico; (2) there is limited gene flow between contemporary populations; and (3) population structure and gene flow are best explained by a combination of IBD, forest cover, and climatic variables (temperature and precipitation). An alternative hypothesis is that forest cover and environmental variables have a limited influence on gene flow, and IBD or other variables are the driver of patterns. To this end, we use ENMs to hypothesize the likely habitat suitability of *P. benedettii* in the Jalisco-Colima region, and assess how the species' range has changed since the Last Glacial Maximum (LGM). We parameterize a resistance model to determine the relative importance of geographic and environmental/landscape predictors (current ENM, ENM projected to the LGM, forest cover) on

genetic differentiation. The results indicate that historical and contemporary range shifts, Euclidean distance, and ecological corridors have influenced the population structure of these geckos. Further, average annual temperature is a major variable influencing the distribution of the species, which has direct implications in light of future climate change.

MATERIALS AND METHODS

Sample Collection and ddRAD Assembly

Genomic DNA was extracted from 161 individuals collected from nine sampling locations throughout Jalisco, Mexico (Fig. 1, Supplementary Table 1). Animal care protocols were approved by the Animal Care Committee at the Royal Ontario Museum (Number 2010-07; Toronto, Ontario, Canada). The ddRADseq libraries were prepared following the protocol of Peterson et al. (2012) and then submitted for paired-end 100 bp sequencing on two lanes of an Illumina Hi-Seq 2500 platform. Genomic DNA was digested using two restriction enzymes, *SphI* and *MluCI*, which recognize GCATGC and AATT sequences respectively. The DNA was then separated into seven pools of 43-48 samples, each with a unique index sequence which was ligated to one end of the DNA. At the other end of the indexed DNA, an inline-barcode was ligated. Size selection was performed for each pool using a Blue Pippin prep 2% dye-free gel cassette (V1; BDF2010, Sage Science) with the size set to 'narrow' at 400 bp. The DNA was multiplexed and amplified after direct size selection. All library preparation and sequencing was performed by the University of Arizona Genetics Core (UAGC).

Raw ddRADseq reads were subsequently assembled *de novo* using ipyrad v.0.6.11 (Eaton & Overcast, 2020). Raw reads were demultiplexed to individuals based on unique barcode sequences, allowing no mismatches in the barcodes. Next, sequences were filtered using the

following parameters: Phred score of 33, no more than five low quality bases per read, a strict filter for adapter sequences, and a clustering threshold of 0.85. We retained a minimum of 20 samples per **dd**RAD locus and used default settings for the remaining parameters.

After assembling and quality filtering the reads, we performed further single nucleotide polymorphism (SNP) filtering using vcftools (Danecek et al., 2011). Only biallelic SNPs were retained. To assess the effect of missing data, we used two missingness thresholds: a 30% complete data matrix and a 50% complete data matrix. More conservative missingness thresholds resulted in matrices with very few loci. For both data sets, we thinned for a single SNP per locus to reduce the impact of linkage disequilibrium (LD) among SNPs.

Population structure

We used principal component analysis (PCA) to visualize genetic structure among individuals and assess the impact of missing data on downstream inference. We treated missing data in two ways. The first approach was to replace the missing genotype with the ancestral allele and the second was to impute missing genotypes by sampling from the distribution of allele frequencies per locality for each SNP. The use of both approaches allowed us to assess the bias introduced by the treatment of missing data.

We additionally used ADMIXTURE v1.3.0 to explore population structure (Alexander et al., 2009). The software implements model-based estimation of ancestry proportions akin to STRUCTURE (Pritchard et al. 2000), with a considerably faster algorithm. In addition, ADMIXTURE is more suitable for moderately sized data sets with high missing data compared to model-free methods like sNMF (Frichot et al., 2014; Frichot & François 2015). In short, ADMIXTURE estimates ancestry coefficients for each individual and the proportion of the

individual's genome that originated from each of *K* possible ancestral pools. The *K*-value with the lowest cross-validation error is inferred to be the most likely number of ancestral pools. We used *K*-values from 2 to 9 to test the hypothesis that the nine sampling locations represented isolated populations. Additionally, we included an sNMF (LEA R package, v3.4.0) analysis for comparison with ADMIXTURE results, acknowledging its limitations with high missing data (Frichot et al., 2014; Frichot & François 2015). Ancestry proportions (*Q*-matrices) were visualized as bar plots using the R package ggplot2 (Wickham 2016) in R v4.1.0 (R Core Team 2021).

Spatially explicit genetic structure

To evaluate the contribution of geographic distance and potential barriers to dispersal among sampling sites, we estimated effective migration surfaces using the EEMS pipeline (Petkova et al., 2015). This method jointly evaluates genetic and geographic distance under a null hypothesis of isolation-by-distance (IBD) to help identify putative corridors and barriers to gene flow. Effective migration was modeled through the comparison of expected and observed genetic dissimilarities between demes, which are regularly spaced and densely packed across the landscape. The model parameterizes a resistance distance matrix by integrating over all possible migration routes between a pair of demes and adjusting potential values to closely match empirical data (McRae 2006). We constructed a pairwise dissimilarity matrix and used the program 'runeems_snps' to implement EEMS with both 200 and 500 demes. We ran 2,000,000 Markov chain Monte Carlo (MCMC) iterations for each, with 1,000,000 burn-in iterations and 9999 thinning iterations between writing steps. We used all default values for hyperparameters, and over multiple replicates, tuned the proposal variances related to diversity parameters,

qEffctProposalS2 = 0.05 (cell effects) and qSeedsProposalS2 = 0.2 (cell locations), and those related to migration parameters, mEffctProposalS2 = 0.25 (cell effects) and mSeedsProposalS2 = 0.05 (cell locations), to get results within the recommended acceptance proportions.

Ecological niche modelling

We used the R application Wallace (Kass et al., 2018) to estimate the extent of habitat suitability for *P. benedettii* across the landscape. Wallace represents a highly reproducible and flexible workflow for species distribution modeling. ENMs require geo-referenced occurrence data of sampled individuals and environmental data to predict areas of ecological suitability for a species (Elith & Leathwick, 2009). Models were trained using the coordinate locations of occurrence data collected for this study. We reduced the effects of spatial autocorrelation by spatially thinning the localities with a 5 km buffer (Aiello-Lammens et al., 2015). We obtained 19 annual temperature and precipitation raster layers at 30 arc-second resolution from the CHELSA database (v1.2; Karger et al., 2017). These rasters were downscaled from a global circulation model and summarized from monthly temperature and precipitation climatology for the years 1979-2013 (Karger et al., 2017). To reduce model overfitting and increase interpretability of the final model, we reduced the climate data set by selecting variables that likely limit the species' range and removed variables highly correlated ($r > 0.7$) with these variables. The study area was considered as a minimum convex polygon around the sampling localities with a 0.50 degree buffer. We sampled pseudo-absence environmental data from 10,000 randomly sampled background points contained within the study area. Given the low number of localities after spatial thinning ($N = 8$), we used a non-spatial jackknife approach to assess model fit (Shcheglovitova & Anderson, 2013). We used the Maxent algorithm for all modelling (Phillips,

Anderson, & Shapire, 2006), and assessed L (linear), LQ (linear-quadratic), H (hinge), and LQH (linear-quadratic-hinge) models with regularization multipliers from 0.5 to 5 in 0.5 intervals to test models of varying complexity. Clamping was used to prevent extrapolation of our models to environmental conditions outside of the training set (Phillips et al., 2005). We discretized model predictions into a binary presence-absence raster with a 10^{th} percentile training threshold for visualizing changes in total occupied area.

The top model was chosen based on the lowest Akaike Information Criterion, corrected for finite sample sizes (AICc). AICc penalizes model complexity and corrects for small sample sizes, giving it an advantage over other model selection approaches (Warren & Seifert, 2011). Because it was unknown which species of the *P. lanei* complex exists in Colima (south of Jalisco), we hypothesized the presence of *P. benedettii* there and extended our model projection to reflect that. We did not include northern regions where other species of the *P. lanei* complex, namely *P. lupitae* and another unnamed species, were proposed to exist (Ramirez-Reyes et al., 2017). In addition, we conducted a Multivariate Environmental Similarity Surface (MESS) analysis to determine regions in the projected extent outside of the study area's range of environmental variation (Elith, Kearney, & Phillips, 2010). Values below zero indicated dissimilar environments and, therefore, more uncertain predictions.

Finally, we used ENMs to visualize how the distribution of *P. benedettii* may have changed since the LGM. Analyses used climatic data at 30 arc-second resolution from the CHELSA database, which applied an algorithm on Paleoclimate Modelling Intercomparison Project Phase III (PMIP3) global circulation model data (Karger et al., 2017). We conducted an additional MESS analysis following the steps listed above in order to hindcast to this time period.

Landscape genetics

Given that leaf-toed geckos are generally found in warm and seasonally wet habitats in forested areas, we hypothesized that past and/or current climatic conditions in addition to forest cover likely influenced gene flow. To complement our EEMS analysis, we estimated the relative contributions of IBD, forest cover, current climate, and LGM climate (jointly, isolation-by-resistance; IBR) on the pairwise genetic distances of *P. benedettii*. We quantified spatial genetic diversity by estimating genetic differentiation among localities with Weir and Cockerham's (1984) F_{ST} in the R package *diveRsity* v1.9.9 (Keenan et al., 2013). We linearized the data ($F_{ST}/1\text{-}F_{ST}$) according to Rousset (1997) to avoid the possibility of non-linear relationships between the genetic distance and the predictors.

Geographic distance was calculated as the Euclidean distance between localities on the WGS ellipsoid, using the *raster* v3.4 R package's function pointDistance (Hijmans, 2019). We derived a binary forest cover predictor from a land cover classification raster produced by the North American Land Change Monitoring System (NALCMS), using data from 2010 to 2015. We classified forested areas as having low resistance to dispersal and non-forested areas as having high resistance to dispersal. Several cost ratios were considered (1:10, 1:100, 1:1000, 1:10000) to understand the impact of parameterization. We inverted the logistic output of the current and LGM ENM projections so low ENM values indicate high resistance to movement and high ENM values indicate low resistance to movement (Spear et al., 2010). For the forest cover, current climate and historical climate variables, we created resistance matrices in the R package *gdistance* v1.3 (van Etten, 2018) that quantify the cost of dispersal between localities given the environmental conditions (Alvarado-Serrano & Knowles, 2014). Individuals were

assumed to disperse in a stochastic manner ("random walk") between two localities and the costs incurred for the journey were averaged to estimate the resistance distance (McRae, 2006).

We used a multiple matrix regression with randomization (MMRR) approach to assess the relationship of IBD and IBR with genetic distance (Wang, 2013). This approach reduced the effect of autocorrelation among pairwise comparisons and helped to partition the relative contribution of each predictor on spatial genetic diversity. Prior to assessing the predictors' fits with spatial genetic diversity, we assessed the Pearson correlation among predictors, removing those with an $r > 0.70$. We calculated AIC values (AICc) for candidate models. Candidate models were selected based on all combinations of IBR hypotheses with IBD and IBD alone. In combination with the multiple parameterizations of forest cover, 10 candidate models were considered.

RESULTS

ddRADseq assembly

From a total of 161 samples, the mean number of raw reads per individual was 883,774, with 880,287 retained after the initial quality control filtering. We obtained an average of 511,185 total clusters per sample after clustering reads within samples, and an average of 20,453 high depth clusters per sample with average read coverage of 10. The final assembly was composed of 153,014 pre-filtered loci, with an average of 21,013 loci per sample. After filtering for duplicates, maximum number of indels and SNPs, and minimum number of samples per locus, we retained 63,721 loci with 348,662 SNPs. We then removed five outlier individuals prior to filtering for SNP missingness. Our 30% complete matrix thinned to one SNP per locus resulted

in 34,447 SNPs, while the 50% complete matrix thinned to one SNP per locus resulted in 9,967 SNPs.

Population structure

The PCAs run on the 30% complete and 50% complete genotype matrices showed highly similar patterns of genetic structure (Supplementary Fig. S1). Replacing the missing data with imputed allele frequencies resulted in higher clustering around localities, but overall patterns were similar to replacing the missing data with the ancestral allele. We used the 30% complete data matrix for all subsequent analyses, since there was little to no difference in genetic structure according to the PCA.

The ADMIXTURE results indicated that the genetic structure of *P. benedettii* was best represented by five ancestral populations (K=5; cross-validation error = 0.184; Fig. 2). With slightly higher cross-validation scores, K=6 and K=7 were included for comparison (Fig. S2). Most individuals had low levels of mixed ancestry. Population structure appeared to follow latitudinal gradient (Fig. 2). In addition, the inland localities P-35, PZ, and P-30 shared the same ancestral population. All individuals from the LG locality shared some ancestry with this inland ancestral population, indicating an elevational coastal-inland gradient in ancestry (Fig. 2; Supplementary Table S1). sNMF results broadly agreed with those of the ADMIXTURE analysis (Supplementary Fig. S3).

Spatial patterns of genetic diversity

The EEMS method allowed us to spatially visualize the patterns and magnitudes of genetic connectivity and isolation across the landscape (Petkova et al., 2016). The estimated effective

298 migration rate across the landscape showed a small deviation from a pure IBD model, with weak
299 barriers and some evidence for dispersal corridors (Fig. 3A). We additionally visualized pairwise
300 FST and within-locality average pairwise nucleotide differences (Tajima's π) to contextualize the
301 EEMS results (Fig. 3B-C). Pairwise FST broadly reflects the EEMS model, with lower FST
302 across the inferred dispersal corridor (Fig. 3B). Tajima's π was lowest in the PC, ST, P-35, and
303 P-30 localities (Fig. 3C), which corresponds with areas of low inferred migration from EEMS
304 analysis and consistently high FST.
305
306 *Ecological niche models*
307 The top performing ENM predicted *P. benedettii* to occur along coastal Jalisco and Colima, with
308 some occurrences inland (Fig. 4A). The chosen model was simple, with a single linear feature
309 and a 2.5 regularization parameter. All models within two AICc of the top model (N=9)
310 contained the same feature and were qualitatively similar to the chosen model. *Phyllodactylus*
311 *benedettii*'s occurrence correlated positively with average annual temperature (Supplementary
312 Fig. S4).
313 Model projection to the LGM suggested overall range stability, with some evidence of
314 recent expansion in Colima (Fig. 4B). Our MESS analysis indicated predominantly similar
315 climate across projections, with dissimilar climatic conditions in the southern portion of the
316 species' predicted range during the LGM, predominantly in present-day Colima (Supplementary
317 Fig. S5). Therefore, predictions made in Colima were interpreted with caution.
318
319 *Landscape genetics*

Resistance matrices generated for current climate and LGM climate were highly correlated (R^2 = 0.962; $P < 0.001$). Because the independent effect of either was difficult to discern in the present analytical framework, we only retained the current climate, forest cover, and geographic distance matrices for the MMRR analyses. Spatial genetic diversity among the sampled populations of *P. benedettii* was jointly explained by IBD and forest cover (IBR) (AICc weight = 0.443, R^2 = 0.675; $\beta_{IBD} = 0.82$; $\beta_{IBR} = -0.322$; $P = 0.002$; Fig. 5; Table 1). IBD showed a statistically significant effect ($P_{IBD} = 0.001$), while IBR was statistically insignificant ($P_{IBR} = 0.099$) at $\alpha = 0.05$. The second most supported model, which was within two AICc of the top model, differed only in parameterization of the forest cover variable (ΔAICc = 0.365, AICc weight = 0.369).

DISCUSSION

The Tropical Dry Forest (TDF) of western Mexico falls within the Mesoamerican biodiversity hotspot (Myers et al., 2000), and previous studies have documented cryptic diversity in several taxa occupying the region (e.g. Devitt, 2006; Zarza et al., 2008; Blair et al., 2013, 2015; Suárez-Atilano et al., 2014; Card et al., 2016). Here, we use ddRADseq data to assemble a *de novo* reduced representation genomic dataset for a recently identified species of gecko, *P. benedettii*, endemic to the Mexican TDF (Ramirez-Reyes, 2018). We uncover fine-scale population structure for *P. benedettii*, and combine molecular and ecological data to test alternative hypotheses for the causes of differentiation. Results suggest substantial ancestral population structure, little admixture, and the presence of isolation by resistance (IBR) in addition to a strong signature isolation by distance (IBD) in influencing patterns of gene flow. Inference of the historical and contemporary processes underlying the patterns of population structuring is

essential for obtaining probabilistic estimates of dispersal patterns, colonization events, and periods of geographic isolation to inform conservation efforts (Lande, 1988).

Spatial population structure

Spatial population structure patterns suggest a latitudinal gradient in population turnover. In addition, three higher elevation inland localities (PZ, P-35, P-30) share ancestry with a fourth inland locality (LG)**.** This suggests an elevational gradient in population structure from lowland coastal to higher elevation inland populations, mediated by a central dispersal corridor found in the EEMS analysis.

Consistent with our population structure results, we find a strong pattern of IBD with a weak signal of IBR. Within the EEMS analysis, most of the sampled areas exhibit levels of effective migration slightly beyond IBD expectations. These altered patterns of migration and genetic diversity across the landscape are the drivers of structure within *P. benedettii*, and this presses further investigation into the specific environmental factors that influence gene flow. The presence of forest cover may facilitate gene flow in the presence of other strong environmental gradients from lowland to montane habitats (Blair et al., 2015). We considered climate and forest cover in a landscape genetic model of dispersal, as climate varies strongly along elevational gradients as the TDF near the coast gives way to pine-oak forest at higher elevations. Our best landscape genetic model includes both IBD + IBR (forest cover) with an AICc weight of 0.443, supporting the hypothesis that distance and landscape characteristics influence gene flow (Table 1). These results are corroborated by the second top model, which differs only in the way that forest resistance is parameterized (1:10 vs 1:10,000). Together, these two models have a cumulative AICc weight of 0.812. In contrast, the best model that includes all three predictors

(i.e. IBD, forest cover, current climate) has substantially lower support (AICc weight = 0.065) and a marginally better R^2 value (0.713 vs 0.675). Thus, forest cover and geographic distance influence gene flow to a greater extent than current climate variables. Notwithstanding, the regression coefficient for forest cover is non-significant in our best model, indicating forest cover predicts dispersal in *P. benedettii* only when considered in tandem with geographic distance. These results are similar to another landscape genetics study on a related species inhabiting TDF that also shows that patterns of gene flow are a function of geographic distance and landscape characteristics (Blair et al., 2013).

Our ENMs show that the lowland regions of Jalisco (i.e. the coastline) are particularly suitable for *P. benedettii* (Fig. 4). Interestingly, the species' predicted range extends beyond the species' northernmost coastal population (PC, Fig. 4), yet this locality has lower nucleotide diversity and lower effective migration than other localities (Fig. 3). Intervening historical environmental fluctuations or biotic factors not captured by our models may influence the genetic diversity of P. *benedettii,* and deserve further investigation (e.g. Kass et al. 2020). Similar situations may influence other species in the *P. lanei* complex, or other lowland taxa throughout the Mexican TDF (Devitt, 2006; Zarza et al., 2008; Blair et al., 2013; Suárez-Atilano et al., 2014; Ramirez-Reyes et al., 2017). This highlights the importance of both geographic distance and landscape features, such as forest configuration, in establishing spatial genetic structure and patterns of gene flow.

Environmental correlates of persistence and dispersal

ENMs uncover how a species' range relates to environmental variables to find new or unknown populations of species, identify barriers to dispersal, and to inform conservation efforts

(Peterson, 2006). ENMs are often explicitly incorporated into landscape genetic studies to test hypotheses regarding spatial patterns of genetic diversity and gene flow (e.g. Ortego et al., 2012; Velo-Antón et al., 2013). Landscape genetic hypotheses are usually tested by examining the effects of multiple explanatory variables in isolation, such as land cover and vegetation density, stream connectivity, and elevation (Spear et al., 2005; Vignieri, 2005; Blair et al., 2013; Trumbo et al., 2021). Although this approach can be useful to determine the best single variable influencing gene flow, the inclusion of many variables makes model testing more challenging while also introducing issues of multicollinearity among predictors (Trumbo et al., 2021). Further, it is more likely that dispersal and gene flow are a result of several landscape and climatic variables acting together. By combining multivariate ENMs with calculations of landscape resistance, our approach minimizes the number of predictor matrices while testing for the effects of climate on gene flow.

Although we explicitly utilize multivariate environmental data derived from current climatic conditions for our ENMs, only a single bioclimatic variable (average annual temperature) contributes to the distribution of *P. benedettii* (Supplementary Fig. S4). More specifically, warmer temperatures correlate with a higher probability of presence. The western coast of Jalisco has consistently higher temperatures than inland regions (Pongpattananurak et al., 2012), and thus temperature variables are also likely important for shaping patterns of dispersal and gene flow. Beyond the responses of the data to specific climate variables, our model hints at other ecological variables that may make the coast suitable for *P. benedettii*. The soil at the coastline of Jalisco is extremely sandy, with some of the lowest levels of clay and silt found in the entire state (Pongpattananurak et al., 2012). Because the niche of *P. benedettii* was

found to be heavily related to soil (Ramirez-Reyes et al., 2018), this characteristic of the coastline may be another factor making this region particularly suitable for *P. benedettii*.

Our finding that the coastline of Jalisco and Colima is climatically most suitable for *P. benedettii* raises two important points. First, the coastline is at a lower elevation than the inland, montane region of western Mexico, which bolsters our argument that vagility, and subsequent gene flow, is limited between populations separated by an elevation gradient. Second, because Colima has not yet been sampled, we hypothesize that *P. benedettii* is present there, or may extend to there soon. It follows that individuals may continue to expand southward along the Jalisco coast into Colima and form populations in this region, if not there already. Alternatively, it is possible that competition with other species of leaf-toed geckos and introduced *Hemidactylus* will limit range expansions into suitable habitats (Ramirez-Reyes et al., 2018).

Although our ENMs suggest that temperature is important in shaping the distribution of these geckos, our landscape genetic analyses indicate that patterns of genetic structure are driven primarily by a combination of IBD + IBR (forest cover). Given the presumed low vagility of these geckos, we expect a signal of IBD to build up quickly. However, a model containing geographic distance alone ranks poorly in our analysis (AICc weight = 0.047), indicating that additional processes are also contributing to patterns of gene flow, corroborating our EEMS analyses. IBR has been shown to predict patterns of gene flow in other species with low vagility (Wang, 2012; Sexton et al., 2014; Wang & Gideon, 2014), and our results lend additional support to the hypothesis that environmental and landscape heterogeneity can limit dispersal. Currently, there are several ways to parameterize models to determine the influence of environmental variables on genetic patterns. While some approaches include sophisticated methods of parameterizing landscape resistance surfaces (e.g., resistanceGA, Peterman, 2018) or

use mixed-effects models to control for non-independence of pairwise distance calculations (van Strien et al., 2012), they require large numbers of localities to effectively make inferences and are designed for individual-based analyses, two requirements that our sampling design does not satisfy. Given currently available analytical methods in landscape genetics, our approach provides the most power to detect strong correlates of dispersal and gene flow based on characteristics of the data (Wang, 2013).

Future directions

Moving forward, use of a latent factor mixed model (LFMM), such as that used by the package LEA (Frichot & François, 2015), could identify loci that correlate with particular SNPs (after controlling for population structure), and may indicate adaptation. The identification of SNPs associated with a particular population would likely provide information on how adaptation to the coastal lowlands has led to the genetic diversity and structure within this species complex. Additionally, because morphological variation is common within and among species of the *P. lanei* complex (Ramirez-Reyes et al., 2018), analysis of the SNPs associated with phenotypic variables could uncover the relationship between genetic and morphological variation (Raposo do Amaral et al., 2018). The high genetic diversity and population structure in such a small region of the landscape indicates that *P. benedettii* is filling a niche space (Ricklefs, 2010). A competitive, invasive gecko species could, therefore, present a threat to *Phyllodactylus* if they occupy the same niche and drive competitive exclusion dynamics (Ricklefs, 2010). One such invasive species may be *Hemidactylus frenatus* that have been seen "stalking, lunging towards and biting other geckos" (Ramirez-Reyes et al., 2018). Lastly, although not explored here, Ramirez-Reyes et al. (2018) suggest that *P. benedettii* has a distinct karyotype (2n=38) compared

to other species in the *P. lanei* complex (2n=33-34). This may present an additional variable one can use, in addition to morphological measurements, exploratory genetic analyses, and coalescent analyses, to accurately delimit species of *Phyllodactylus*. In sum, our genome-wide data provide additional evidence that the TDF of western Mexico harbors unrecognized diversity and a deep genetic history over small spatial scales. Such studies will continue to be vital, as deforestation and habitat loss threaten biodiversity throughout the region (Ceballos & Garcia, 1995; Trejo & Dirzo, 2000). Our results also indicate that ddRADseq (and related methods) continues to serve as an invaluable tool for identifying spatial patterns of genetic diversity and gaining a clearer picture of how gene flow, adaptation, and population structure form in non-model organisms in response to landscape and environmental characteristics.

ACKNOWLEDGEMENTS

Funding was provided to C. Blair as part of his start-up package at NYC College of Technology (CUNY). We thank C. Ané for discussions regarding multiple regression analysis. We thank the University of Arizona Genetics Core (UAGC) for help with library preparation and sequencing. Finally, we thank the anonymous reviewers for their helpful comments on the manuscript.

CONFLICT OF INTEREST

The authors have no conflict of interest to declare.

DATA AVAILABILITY STATEMENT

The raw sequence data are available on the Sequence Read Archive.

479

480 REFERENCES

481 Aiello-Lammens, M.E., Boria, R.A., Radosavljevic, A., Vilela, B., & Anderson, R.P. (2015).
482 spThin: an R package for spatial thinning of species occurrence records for use in
483 ecological niche models. *Ecography*, 38, 541–545.
484 Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in
485 unrelated individuals. *Genome Research*, 19, 1655–64.
486 Alvarado-Serrano, D.F., & Knowles, L.L. (2014). Ecological niche models in phylogeographic
487 studies: applications, advances and precautions. *Molecular Ecology Resources*, 14, 233–
488 248.
489 Andrews, S. (2014). FastQC: a quality control tool for high throughput sequence data. url:
490 https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
491 Becerra, J. X. (2005). Timing the origin and expansion of the Mexican tropical dry forest.
492 *Proceedings of the National Academy of Sciences USA*, 102, 10919–10923.
493 Becerra, J. X., & Venable, D. L. (2008). Sources and sinks of diversification and
494 conservation priorities for the Mexican tropical dry forest. *PLoS One.* 3, e3436.
495 Blair, C., Jiménez Arcos, V.H., Mendez de la Cruz, F. R., & Murphy, R.W. (2013). Landscape
496 genetics of leaf-toed geckos in the tropical dry forest of northern Mexico. *PLoS One*, 8,
497 e57433.
498 Blair, C., Jimenez Arcos, V. H., Mendez de la Cruz, F. R., & Murphy, R. W. (2014). Historical
499 and contemporary demography of leaf-toed geckos (Phyllodactylidae: *Phyllodactylus*
500 *tuberculosus saxatilis*) in the Mexican dry forest. *Conservation Genetics*, 19, 419–429.

501 Blair, C., Mendez de la Cruz, F. R., Law, C., & Murphy, R. W. (2015). Molecular phylogenetics
502 and species delimitation of leaf-toed geckos (Phyllodactylidae*: Phyllodactylus*)
503 throughout the Mexican tropical dry forest. *Molecular Phylogenetics and Evolution*, 84,
504 254–265.
505 Blair C, Bryson Jr RW, García-Vázquez UO, Nieto-Montes de Oca, A, Lazcano D, McCormack
506 JE & Klicka J. (2022). Phylogenomics of alligator lizards elucidate diversification
507 patterns across the Mexican Transition Zone and support the recognition of a new genus.
508 *Biological Journal of the Linnean Society*, 135, 25–39.
509 Bowman, J., Greenhorn, J.E., Marrotte, R.R., McKay, M.M., Morris, K.Y., Prentice, M.B., &
510 Wehtje M. (2016). On applications of landscape genetics. *Conservation Genetics*, 17,
511 753–760.
512 Bryson, R. W., Murphy, R. W., Graham, M. R., Lathrop, A., & Lazcano, D. (2011a). Ephemeral
513 Pleistocene woodlands connect the dots for highland rattlesnakes of the *Crotalus*
514 *intermedius* group. *Journal of Biogeography*, 38, 2299–2310.
515 Bryson, R. W., Murphy, R. W., Lathrop, A., & Lazcano-Villareal, D. (2011b). Evolutionary
516 drivers of phylogeographic diversity in the highlands of Mexico: a case study of the
517 *Crotalus triseriatus* species group of montane rattlesnakes. *Journal of Biogeography*, 38,
518 697–710.
519 Bryson, R. W. & Riddle, B. R. (2012). Tracing the origins of widespread highland species: a
520 case of Neogene diversification across the Mexican sierras in an endemic lizard.
521 *Biological Journal of the Linnean Society*, 105, 382–394.

522 Card, D. C., Schield, D.R., Adams, R. H., et al. (2016). Phylogeographic and population genetic
523 analyses reveal multiple species of *Boa* and independent origins of insular dwarfism.
524 *Molecular Phylogenetics and Evolution,* 102, 104–116.
525 Ceballos, G., & García, A. (1995). Conserving neotropical biodiversity: the role of dry forests in
526 western Mexico. *Conservation Biology,* 9, 1349–1356.
527 Cronin, M. A. (1993). Mitochondrial DNA in wildlife taxonomy and conservation biology:
528 cautionary notes. *Wildlife Society Bulletin,* 21, 339–348.
529 Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E.
530 et al. (2011). The variant call format and VCFtools. *Bioinformatics,* 27, 2156–2158
531 Devitt, T.J. (2006). Phylogeography of the Western Lyresnake (*Trimorphodon biscutatus*: testing
532 aridland biogeographical hypotheses across the Nearctic-Neotropical transition.
533 *Molecular Ecology,* 15, 4387–4407.
534 Dixon, J.R. (1964) The systematics and distribution of lizards of the genus *Phyllodactylus* in
535 North and Central America. *New Mexico State University Research Center, University*
536 *Park, New Mexico, Scientific Bulletin,* 64-1, 1–139.
537 Dubey, S., Croak, B., Pike, D., Webb., J., & Shine, R. (2012). Phylogeography and dispersal in
538 the velvet gecko (*Oedura lesueurii*), and potential implications for conservation of an
539 endangered snake (*Hoplocephalus bungaroides*). *BMC Evolutionary Biology,* 12, 67.
540 Duellman, W. E. (1958). A preliminary analysis of the herpetofauna of Colima, Mexico.
541 *Occasional Papers of the Museum of Zoology, University of Michigan*, No. 589.
542 Eaton, D. A., & Overcast, I. (2020). ipyrad: Interactive assembly and analysis of RAD-seq
543 datasets. *Bioinformatics,* btz966.

544 Elith, J., Kearney M., & Phillips, S. (2010). The art of modelling range-shifting species. *Methods*
545 *in Ecology and Evolution,* 1, 330–342.
546 Elith, J. & Leathwick, J. R. (2009). Species distribution models: ecological explanation and
547 prediction across space and time. *Annual Review of Ecology, Evolution, and*
548 *Systematics,* 40, 677–697.
549 Flores-Villela, O., & Garcia-Vazquez, U. O. (2014). Biodiversidad de reptiles en México.
550 *Revista Mexicana de Biodiversidad,* 85, S467–S475.
551 Frichot, E. & François, O. (2015). LEA: an R package for landscape and ecological
552 association studies. *Methods in Ecology and Evolution,* 6, 925–929.
553 Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G., & François, O. (2014). Fast and efficient
554 estimation of individual ancestry coefficients. *Genetics,* 196, 973–983.
555 Hijmans, R.J. (2019). raster: Geographic Data Analysis and Modeling. R package version 2.9-5.
556 https://CRAN.R-project.org/package=raster
557 Holderegger, R., & Wagner, H.H. (2008). Landscape genetics. *Bioscience,* 58,199–207.
558 Janzen, D. H. (1988). Tropical dry forests: the most endangered major tropical ecosystem. In E.
559 O. Wilson (Ed.), Biodiversity. National Academies Press, 130–137.
560 Karger, D.N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R.W., Zimmermann,
561 N.E., et al. (2017). Climatologies at high resolution for the earth's land surface areas.
562 *Scientific Data,* 4,170122.
563 Kass, J.M., Meenan, S.I., Tinoco, N., Burneo, S.F., & Anderson, R.P. (2021). Improving area of
564 occupancy estimates for parapatric species using distribution models and support vector
565 machines. Ecological Applications 31, e02228.
566 Kass, J.M., Vilela, B., Aiello-Lammens, M.E., Muscarella, R., Merow, C., & Anderson, R.P.

(2018). Wallace: A flexible platform for reproducible modeling of species niches and distributions built for community expansion. *Methods in Ecology and Evolution*, 9,1151–1156.

Keenan, K., McGinnity, P., Cross, T.F., Crozier, W.W., & Prodöhl, P.A. (2013), diveRsity: An R package for the estimation of population genetics parameters and their associated errors. *Methods in Ecology and Evolution*, 4, 782-788. doi: 10.1111/2041-210X.12067

Keller, D., Holeregger, R., van Strien M.J., & Bolliger, J.(2015). How to make landscape genetics beneficial for conservation management? *Conservation Genetics*, 16, 503–512.

Lande, R. (1988). Genetics and demography in biological conservation. *Science.* 241, 1455–1460.

Manel, S., Schwartz, M.K., Luikart, G., & Taberlet, P. (2003). Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology and Evolution*, 18, 189–197.

Manel, S., & Holderegger, R. (2013). Ten years of landscape genetics. *Trends in Ecology and Evolution*, 28, 614–621.

McRae, B.H. (2006). Isolation by resistance. *Evolution*, 60, 1551–1561.

Mooney, H. A., Bullock, S. H., & Medina, E. (1995) Seasonally Dry Tropical Forests. Cambridge University Press.

Myers, N., Mittermeier, R.A., Mittermeier, C.G., da Fonseca G.A.B., & Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature*, 409, 853–858.

Oliveira, E.F., Martinez, P.A., São-Pedro, V.A., Gehara, M., Burbrink, F.T., Mesquita, D.O., Garda, A.A., et al. (2018). Climatic suitability, isolation by distance and river resistance explain genetic variation in a Brazilian whiptail lizard. *Heredity*, 120, 251–265.

590 Ortego, J., Riordan, E.C., Gugger, P.F., & Sork, V.L. (2012). Influence of environmental
591 heterogeneity on genetic diversity and structure in an endemic southern Californian oak.
592 *Molecular Ecology,* 21, 3210–3223.
593 Peterman, W.E. (2018). ResistanceGA: An R package for the optimization of resistance surfaces
594 using genetic algorithms. *Methods in Ecology & Evolution,* 9, 1638–1647.
595 Peterson, A. T. (2006). Uses and requirements of ecological niche models and related
596 distributional models. *Biodiversity Informatics,* 3, 59–72.
597 Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest
598 RADseq: an inexpensive method for de novo SNP discovery and genotyping in model
599 and non-model species. *PloS One,* 7, e37135.
600 Petkova, D., Novembre, J., & Stephens, M. (2015). Visualizing spatial population structure
601 with estimated effective migration surfaces. *Nature Genetics,* 48, 94–100.
602 Petren, K. (2013). The evolution of landscape genetics. *Evolution,* 67, 3383–3385.
603 Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2005). Maximum entropy modeling of
604 species geographic distributions. *Ecological Modelling,* 190, 231–259.
605 Pimm, S. L., Russell, G. J., Gittleman, J. L. & Brooks, T. M. (1995). The future of biodiversity.
606 *Science,* 269, 347–350.
607 Pongpattananurak, N., Reich, R. M., Khosla, R., & Aguirre-Bravo, C. (2012). Modeling the
608 spatial distribution of soil texture in the state of Jalisco, Mexico. *Soil Science Society*
609 *of America Journal,* 76, 199–209.
610 R Core Team (2021). R: A language and environment for statistical computing. R Foundation for
611 Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

612 Ramirez-Reyes, T., Pinero, D., Flores-Villela, O., & Vazquez-Dominguez, E. (2017). Molecular
613 systematics, species delimitation and diversification patterns of the *Phyllodactylus lanei*
614 complex (Gekkota: *Phyllodactylidae*) in Mexico. *Molecular Phylogenetics and*
615 *Evolution,* 115, 82–94.
616 Ramirez-Reyes, T. & Flores-Villela, O. A. (2018). Taxonomic changes and description of two
617 new species for the *Phyllodactylus lanei* complex (Gekkota: *Phyllodactylidae*) in
618 Mexico. *Zootaxa,* 4407, 151–190.
619 Ramirez-Reyes, T., Blair, C., Flores-Villela, O., Piñero, D., Lathrop, A., & Murphy, R. (2020).
620 Phylogenomics and molecular species delimitation reveals great cryptic diversity of leaf-
621 toed geckos (Phyllodactylidae: *Phyllodactylus*), ancient origins, and diversification in
622 Mexico. *Molecular Phylogenetics and Evolution,* 150, 106880.
623 Raposo do Amaral, F., Maldonado-Coelho, M., Aleixo, A., Luna, L. W., Sena do Rêgo, P.,
624 Araripe, J., Souze, T. O., Silva, W. A. G., & Thom, G. (2018). Recent chapters of
625 Neotropical history overlooked in phylogeography: shallow divergence explains
626 phenotype and genotype uncoupling in *Antilophia* manakins. *Molecular Ecology,*
627 27, 4108–4120.
628 Ricklefs, R. E. (2010). Evolutionary diversification, coevolution between populations and
629 their antagonists, and the filling of niche space. *Proceedings of the National Academy of*
630 *Sciences USA,* 107, 1265–1272.
631 Rico, Y. (2019). Landscape genetics of Mexican biodiversity: A review. *Acta Universitaria,* 29,
632 e1894.
633 Ruiz-Sanchez, E. & Specht, C. D. (2013). Influence of the geological history of the
634 Trans-Mexican Volcanic Belt on the diversification of *Nolina parviflora*

(Asparagaceae: Nolinoideae). *Journal of Biogeography,* 40, 1336–1347.

Rousset, F. (1997). Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics,* 145, 1219–1228.

Sexton, J.P., Hangartner, S.B., & Hoffman, A. A. (2014). Genetic isolation by environment or distance: which pattern of gene flow is most common? *Evolution,* 68, 1–15.

Shcheglovitova, M., & Anderson, R.P. (2013). Estimating optimal complexity for ecological niche models: A jackknife approach for species with small sample sizes. *Ecological Modelling,* 269, 9–17.

Sork, V.L., & Waits, L.P. (2010). Contributions of landscape genetics—approaches, insights, and future potential. *Molecular Ecology,* 19, 3489–3495.

Spear, S.F., Peterson, C.R., Matocq, M., & Storfer, A. (2005). Landscape genetics of the blotched tiger salamander (*Ambystoma tigrinum melanostictum*). *Molecular Ecology,* 14, 2553–2564.

Spear, S.F., Balkenhol, N., Fortin, M.-J., McRae, B.H., & Scribner, K. (2010). Use of resistance surfaces for landscape genetic studies: considerations for parameterization and analysis. *Molecular Ecology,* 19, 3576–3591.

Storfer, A., Murphy, M.A., Evans, J.S., Goldberg, C.S., Robinson, S., Spear, S.F., Dezzani, R., et al. (2007). Putting the ‘landscape’ in landscape genetics. *Heredity,* 98, 128–142.

Storfer, A., Murphy, M.A., Spear, S.P., Holderegger, R., & Waits, L.P. (2010). Landscape genetics: where are we now? *Molecular Ecology,* 19, 3496–3514.

Storfer, A., Patton, A., & Fraik, A.K. (2018). Navigating the interface between landscape genetics and landscape genomics. *Frontiers in Genetics,* 9, 68.

Suárez-Atilano, M., Burbrink, F., & Vázquez-Domínguez, E. (2014). Phylogeographical structure within *Boa constrictor imperator* across the lowlands and mountains of Central America and Mexico. *Journal of Biogeography*, 41, 2371–2384.

Trejo, I. & Dirzo, R. (2000). Deforestation of seasonally dry tropical forest: a national and local analysis in Mexico. *Biological Conservation*, 94, 133–142.

Trumbo, D.R., Funk, W.C., Pauly, G.B., & Robertson, J.M. (2021). Conservation genetics of an island-endemic lizard: low Ne and the critical role of intermediate temperatures for genetic connectivity. *Conservation Genetics*, 22, 783–797.

van Etten, J. (2018). gdistance: distances and routes on geographical grids. R package version 1.2-2. https://CRAN.R-project.org/package=gdistance

Van Strien, M.J., Keller, D., & Holderegger, R. (2012). A new analytical approach to landscape genetic modelling: least-cost transect analysis and linear mixed models. *Molecular Ecology*, 21, 4010–4023.

Velo-Antón, G., Parra, J.L., Parra-Olea, G., & Zamudio, K. (2013). Tracking climate change in a dispersal limited species: reduced spatial and genetic connectivity in a montane salamander. *Molecular Ecology*, 22, 3261–3278.

Vignieri, S.N. (2005). Streams over mountains: influence of riparian connectivity on gene flow in the Pacific jumping mouse (*Zapus trinotatus*). *Molecular Ecology*, 14, 1925–1937.

Wang, I.J. (2012). Environmental and topographical variables shape genetic structure and effective population sizes in the endangered Yosemite toad. *Diversity and Distributions*, 18, 1033–1041.

678 Wang, I.J. (2013). Examining the full effects of landscape heterogeneity on spatial genetic
679 variation: a multiple matrix regression approach for quantifying geographic and
680 ecological isolation. *Evolution,* 67, 3403–3411.
681 Wang, I.J., Glor, R.E., & Losos, J.B. (2013). Quantifying the roles of ecology and geography in
682 spatial genetic divergence. *Ecology Letters,* 16, 175–182.
683 Wang, I.J., & Bradburd, G.S. (2014). Isolation by environment. *Molecular Ecology,* 23, 5649–
684 5662.
685 Warren, D.L., & Seifert, S.N. (2011). Ecological niche modeling in Maxent: the importance of
686 model complexity and the performance of model selection criteria. *Ecological*
687 *Applications,* 21, 335–342.
688 Weir, B.S., & Cockerham, C.C. (1984). Estimating F-Statistics for the analysis of population
689 structure. *Evolution,* 38, 1358–1370.
690 Werneck, F. P., Costa, G. C., Colli, G. R., Prado, D. E., & Sites, Jr., J. W. (2010). Revisiting the
691 historical distribution of Seasonally Dry Tropical Forests: new insights based on
692 palaeodistribution modelling and palynological evidence. *Global Ecology and*
693 *Biogeography,* 20, 272–288.
694 Wickham H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.
695 ISBN 978-3-319-24277-4, https://ggplot2.tidyverse.org.
696 Wright, S. (1943). Isolation by distance. *Genetics,* 28, 139–156.
697 Zarza, E., Reynoso, V.H, & Emerson, B.C. (2008). Diversification in the northern Neotropics:
698 mitochondrial and nuclear DNA phylogeography of the iguana *Ctenosaura pectinata* and
699 related species. *Molecular Ecology,* 17, 3259–3275.

Zarza, E., Connors, E. M., Maley, J. M., Tsai, W. L. E., Heimes, P., Kaplan, M., & McCormack, J. E. (2018). Combining ultraconserved elements and mtDNA data to uncover lineage diversity in a Mexican highland frog (*Sarcohyla;* Hylidae). *PeerJ*, 6, e6045.

BIOSKETCH

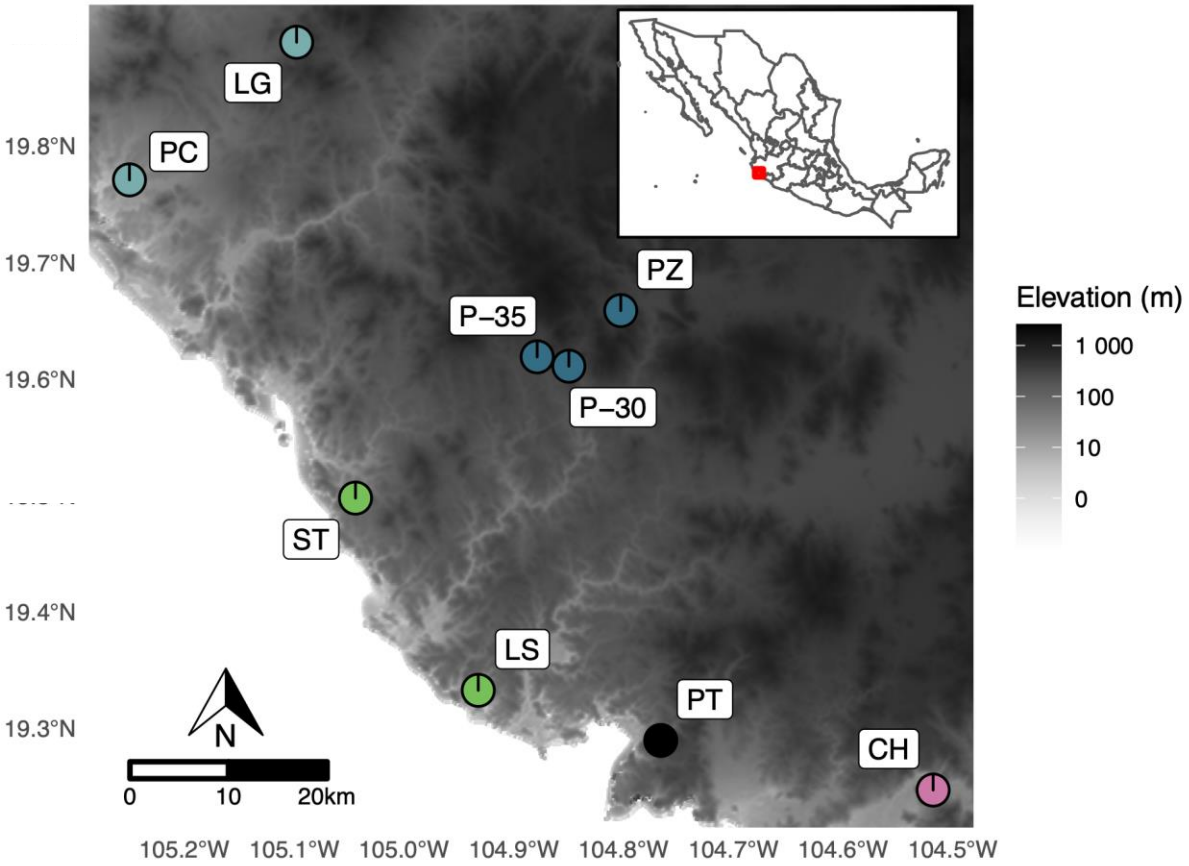
Connor M. French is a PhD candidate interested in understanding the geographic distribution of genetic diversity, from patterns to processes. This work represents a collaboration among researchers at CCNY and abroad with a shared interest in the influence of a changing environment on species genetic structure. In addition, this study is part of a series of work focusing on uncovering cryptic patterns of Mexican lizard diversity.

AUTHOR CONTRIBUTIONS

CB, RLMB and FRM collected the genetic samples. CB, RLMB, FRM, RWM and MJH conceived of the study and suggested experiments. CMF, CS, SB, and IO performed analyses. CS wrote the initial draft of the manuscript. CMF, IO, MJH, and CB were involved in the revision of the initial manuscript.

719 TABLES AND FIGURES

720
721



722
723

724 **Fig. 1**. Sampling localities of *P. benedettii* from nine lowland locations within Jalisco, Mexico.
725 Pie charts at each locality represent the proportion of individuals assigned to an ancestral genetic
726 population, further detailed in Fig. 2. Of note, LG individuals contain substantial admixture from
727 the [P-35, PZ, and P-30] ancestral population, and P-35 contain substantial admixture from the
728 [ST and LS] ancestral population. Marker labels are as follows: ST = Station; PC = Puente Cuate
729 I; LG = Road to Llano Grande; P-35 = Road to Purificación km 35; PZ = Puente El Zarco; CH =
730 El Charco House; PT = Puente Tigra; LS = Llano Seco; P-30 = Road to Purificación House km
731 30. Further detail on sampling is provided in Supplementary Table S1.

732
733
734
735
736
737
738
739
740

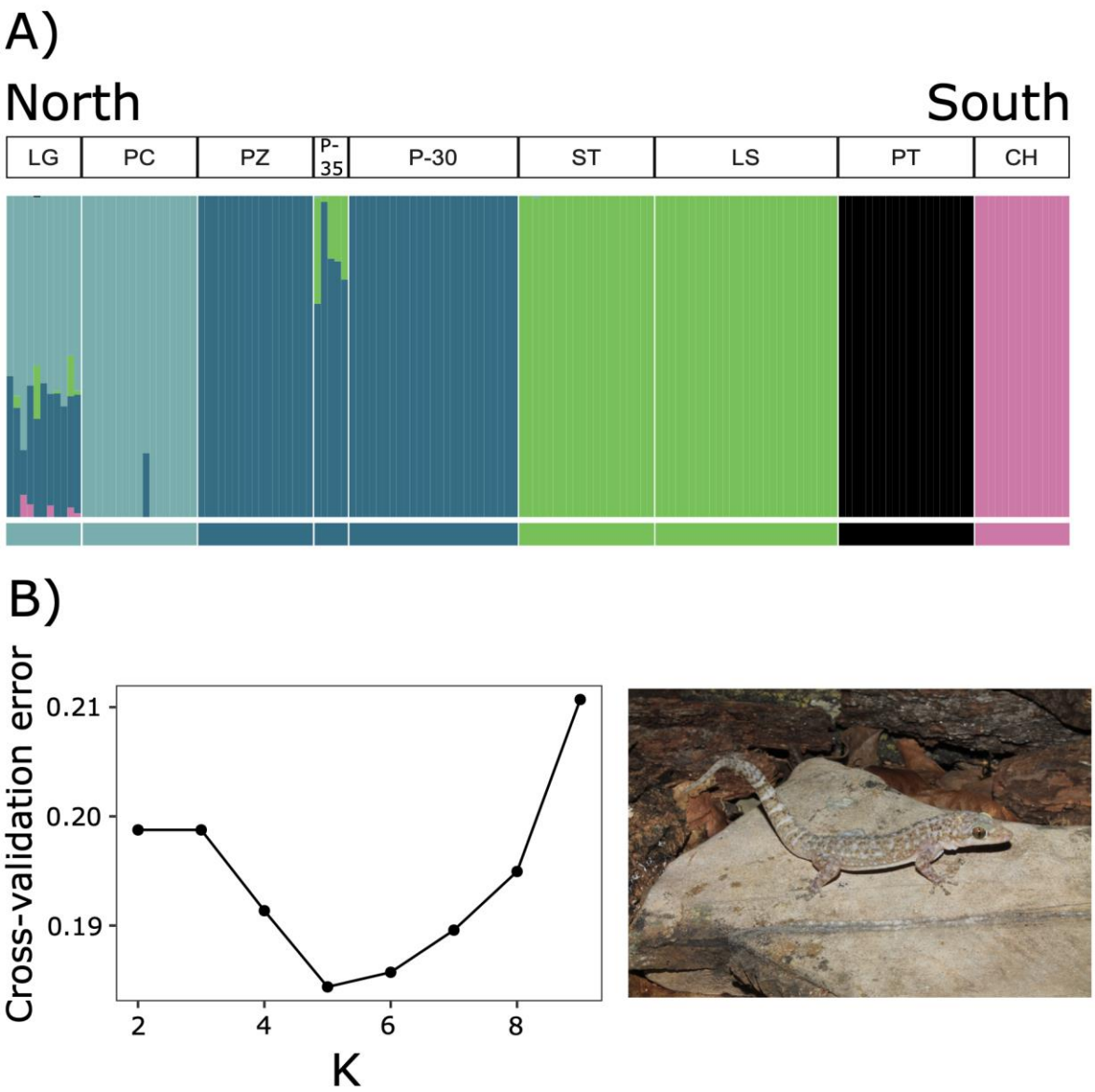


Fig. 2. ADMIXTURE analysis results for *P. benedettii*. A) Barplot of admixture proportions for the top performing run (*K*=5), where spatial structure is evident. Each vertical bar represents one individual, and individuals are grouped by sampling locality along the *x*-axis, arranged from the northernmost locality to the southernmost. Each individual was assigned to a most-likely ancestral population according to the population with the highest admixture proportion for that individual. The bottom row represents the proportion of individuals assigned to each ancestral population for each locality. The cross-entropy plot B) shows the support for different *K* values considered in the analysis, where lower cross-validation error indicates more support. The bottom-right inset is an *in-situ* photograph of the study species (Photo credit: Tonatiuh Ramirez).

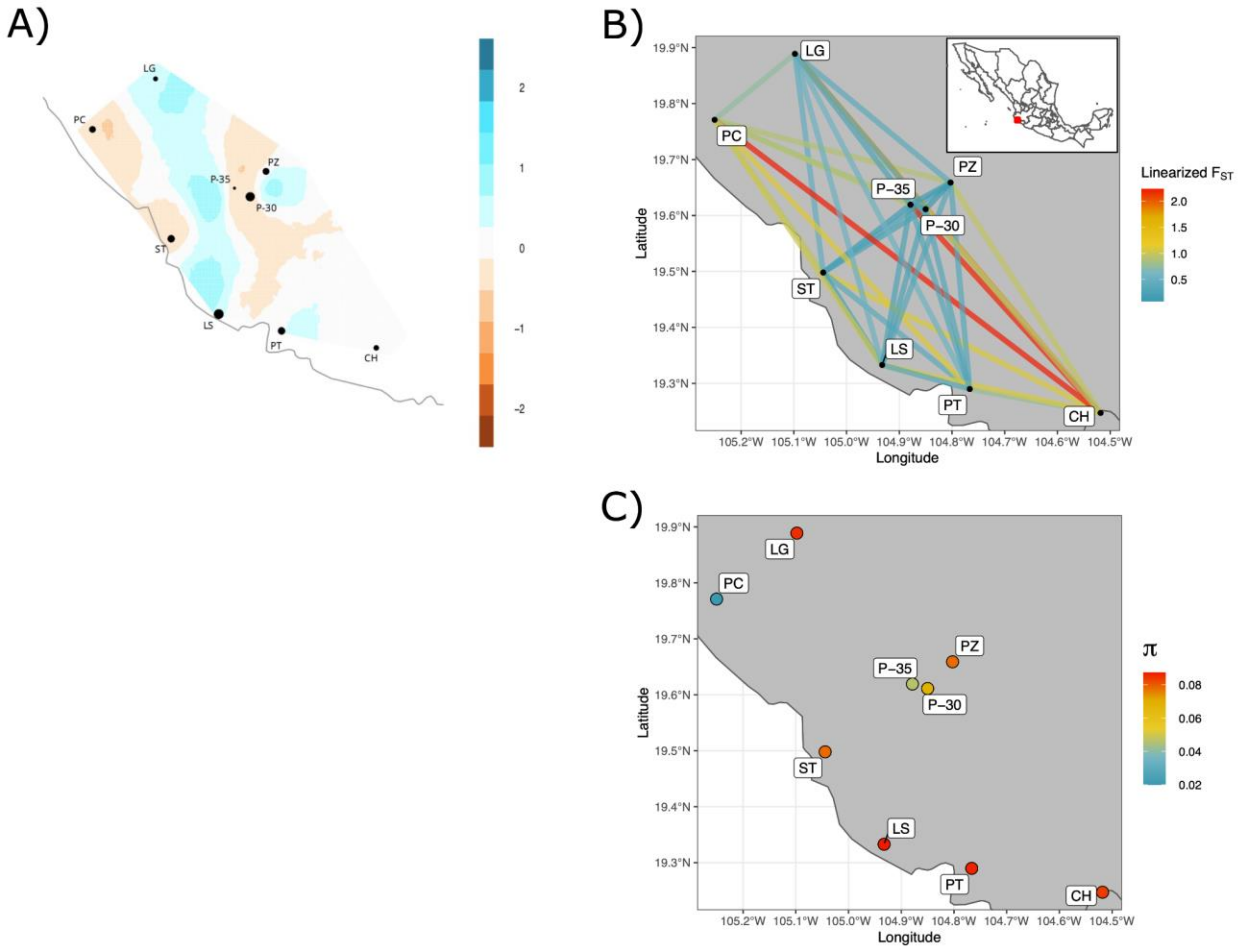


Fig. 3. EEMS migration surface and raw genetic diversity estimates, showing that effective migration is qualitatively similar to linearized F_{ST} ($F_{ST}/(1 - F_{ST})$) and localities with high connectivity generally have higher nucleotide diversity (Tajima's π). Black dots in A) indicate sampling locations and their sizes reflect the number of samples collected from that site (see Supplementary Table S1). Mean migration rates (*m*) A) are highest along a central corridor, bridging the LG and LS localities, while *m* is lower in parallel transects along the coast and further inland. This pattern is broadly reflected in B) pairwise linearized F_{ST}. The localities with the lowest pairwise divergence in B) also generally have the highest nucleotide diversity in C), with the southernmost locality, CH, being the exception.

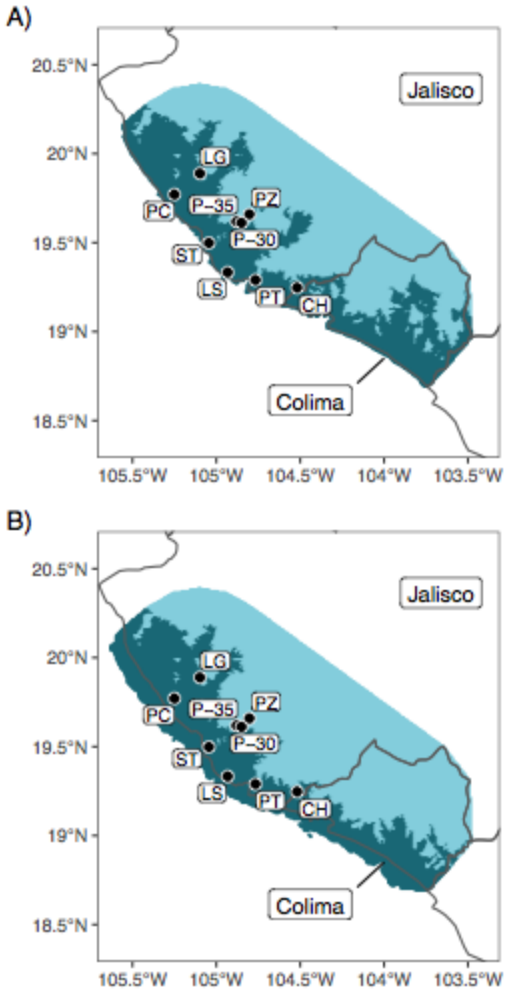


Fig. 4. ENM results indicating suitability of the southwestern Mexican coastline for *P. benedettii*. A) projection to current climatic conditions; B) projection to the last glacial maximum (LGM). The black dots represent the nine sampling sites. Dark blue indicates predicted presence based on a 10th percentile training threshold, while light blue indicates predicted absence. Sampling locality codes correspond to those in Fig. 1 and Supplementary Table S1. The species' range is predicted to have remained stable since the LGM, with some evidence for an inland expansion in Colima. Note that coastal land area during the LGM extends beyond current coastal boundaries.

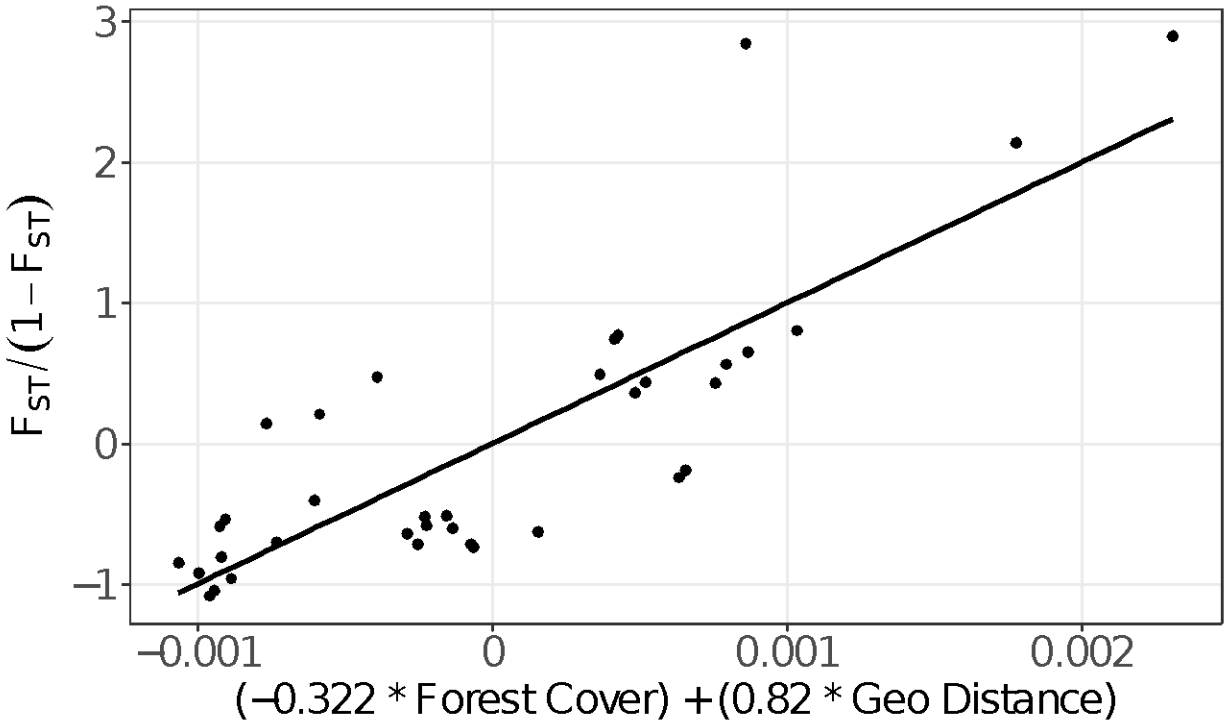


Fig. 5. Linearized F_{ST} among sites plotted as a function of the MMRR model. There is a stronger signature of isolation by distance (IBD) compared to isolation by resistance (IBR, Forest Cover) among populations of *P. benedettii* ($R^2 = 0.675$; $\beta_{IBD} = 0.82$; $\beta_{IBR} = -0.322$; $P = 0.002$). IBD shows a statistically significant effect ($P_{IBD} = 0.001$), while IBR is statistically insignificant ($P_{IBR} = 0.099$) at $\alpha = 0.05$.

787 **Table 1**. Multiple matrix regression with randomization model comparison results. These models
788 represent the relationship between genetic distance (linearized F_{ST}) and the predictors geographic
789 distance, current climate, and forest cover. We tested four cost ratios of forest cover (1:10, 1:100,
790 1:1000, 1:10000). The top two models with nearly identical support represent two forest cover
791 parameterizations and are qualitatively similar. Models are ranked by their AIC_C value. K =
792 number of model parameters, FC = Forest cover, Geo = geographic distance, Current = current
793 climate.
794

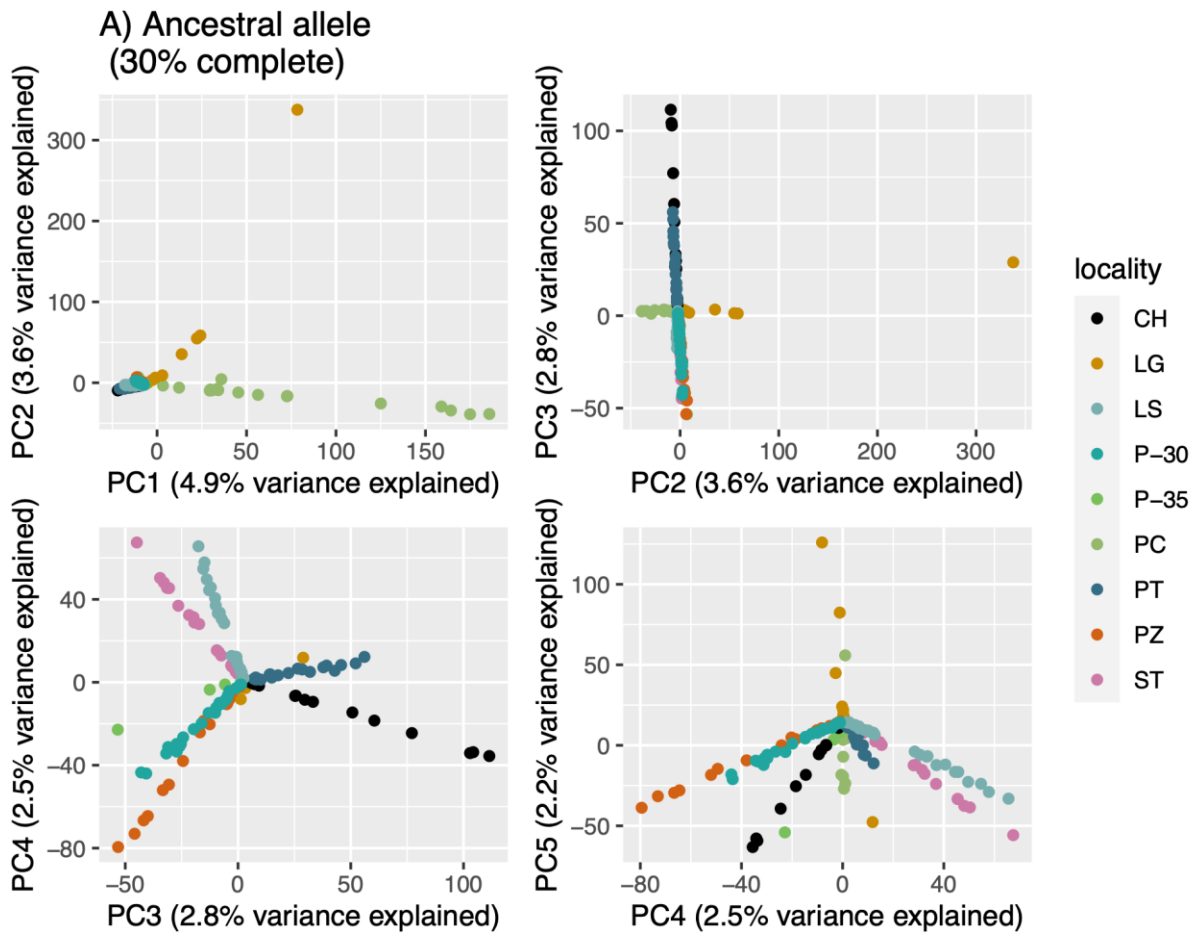
Model	**R^2**	**K**	**AICc**	**ΔAICc**	**AICc weight**	**Cumulative weight**	**Log Likelihood**	***P*-value**
FC 1:10000 + Geo	0.675	4	99.950	0.000	0.443	0.443	-45.330	0.002
FC 1:10 + Geo	0.725	4	100.314	0.365	0.369	0.812	-45.512	0.001
FC 1:100 + Current + Geo	0.713	5	103.786	3.836	0.065	0.877	-45.893	0.001
Geo	0.575	3	104.443	4.493	0.047	0.924	-48.846	0.001
FC 1:100 + Geo	0.694	4	105.373	5.423	0.029	0.954	-48.041	0.001
Current + Geo	0.587	4	105.845	5.895	0.023	0.977	-48.277	0.002
FC 1:1000 + Current + Geo	0.693	5	108.087	8.137	0.008	0.985	-48.043	0.003
FC 1:1000 + Geo	0.677	4	108.379	8.429	0.007	0.991	-49.544	0.002
FC 1:10000 + Current + Geo	0.690	5	108.649	8.700	0.006	0.997	-48.325	0.002
FC 1:10 + Current + Geo	0.752	5	109.853	9.904	0.003	1.000	-48.927	0.001

795
796

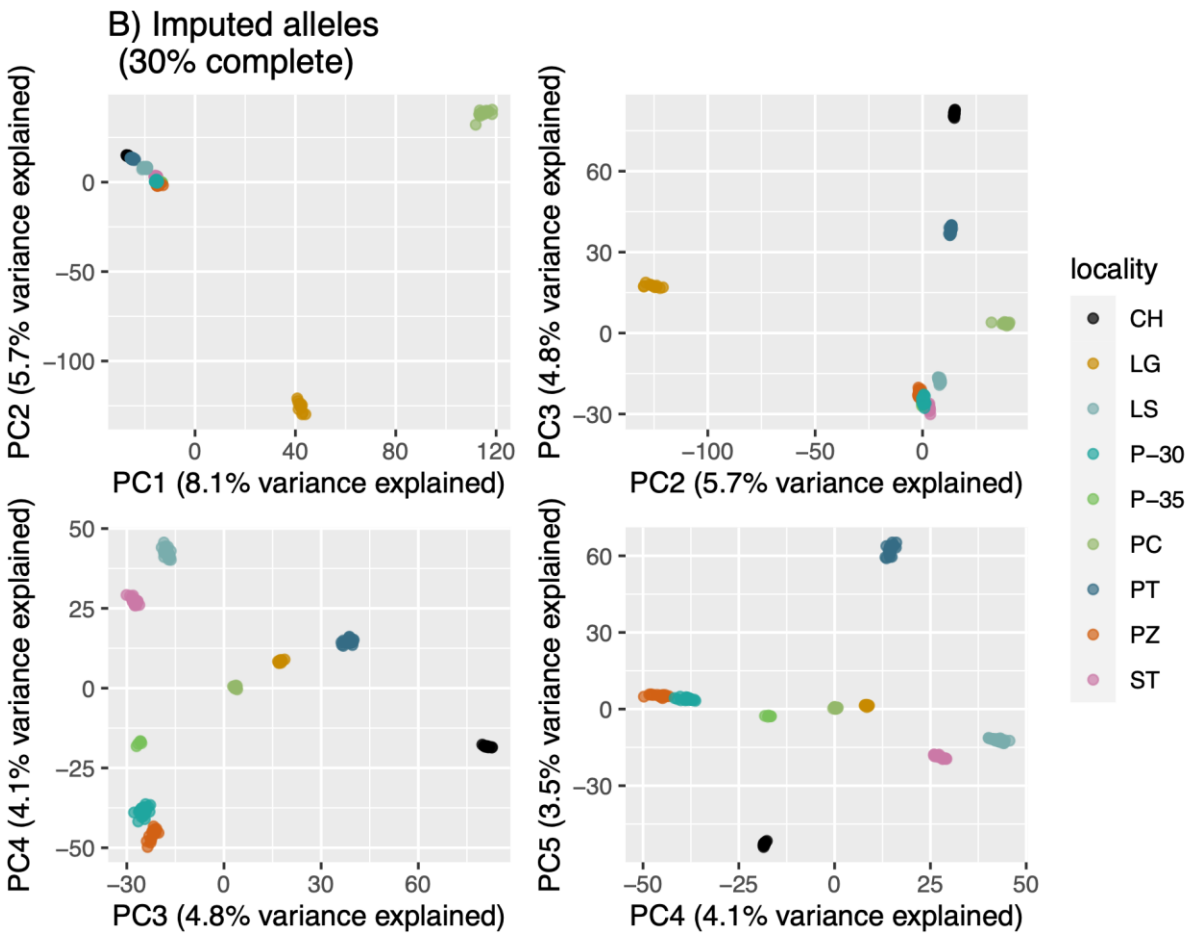
SUPPLEMENTARY TABLES & FIGURES

Supplementary Table S1. Sample information for each of the nine sampling localities in Jalisco, Mexico. In parentheses next to each abbreviated site name is a letter indicating whether the locality is indicated as coastal (C) or inland (I).

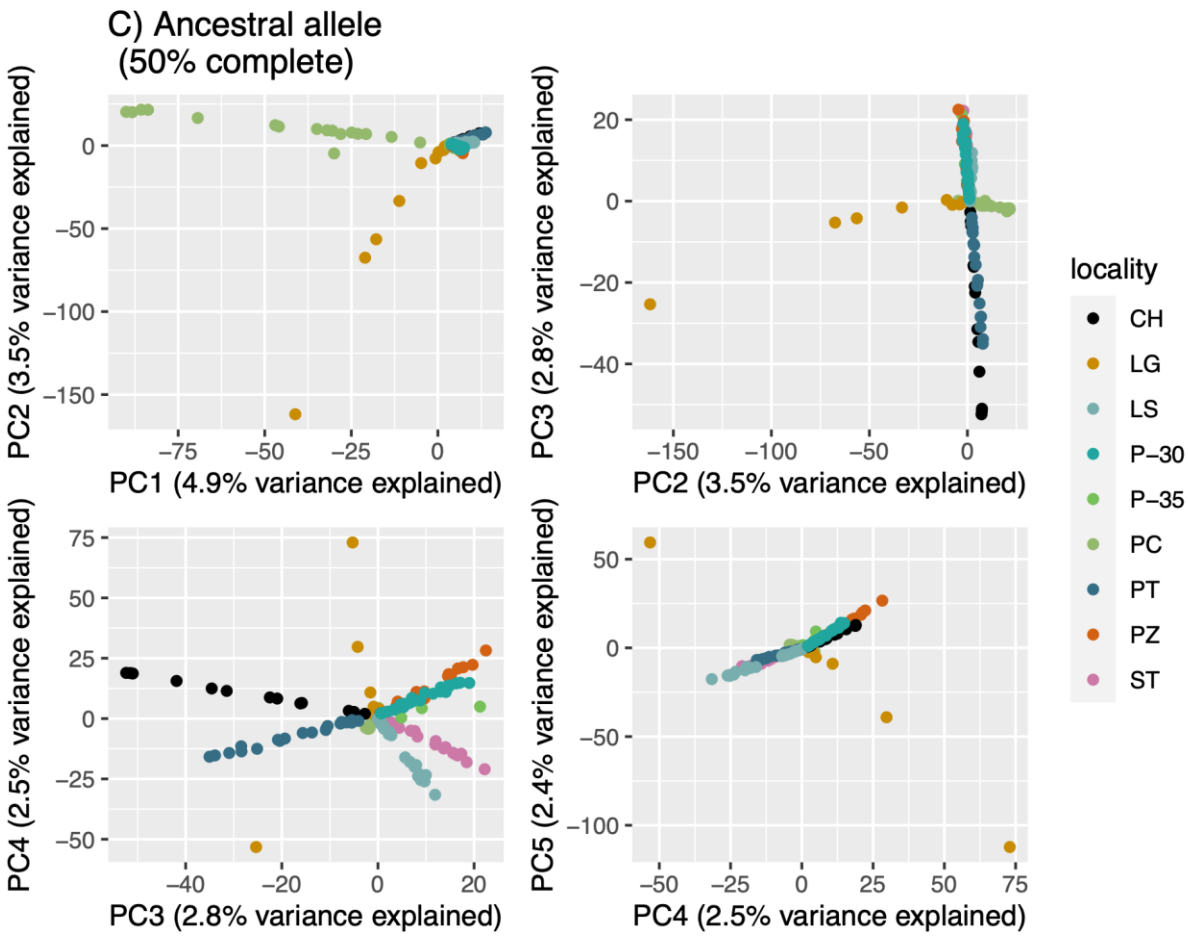
Site #	Full Site Name	Abbreviated Site Name	Coordinates	Elevation (m)	# Samples
1	Station	ST (C)	19.498028 N 105.044361 W	587	20
2	Puente Cuate I	PC (C)	19.771 N 105.250278 W	17	17
3	Road to Llano Grande	LG (I)	19.888917 N 105.098 W	327	11
4	Road to Purificación, km 35	P-35 (I)	19.61925 N 104.878861 W	377	5
5	Puente El Zarco	PZ (I)	19.658917 N 104.802722 W	401	17
6	El Charco House	CH (C)	19.247167 N 104.518139 W	154	14
7	Puente Tigra	PT (C)	19.289806 N 104.766306 W	40	20
8	Llano Seco	LS (C)	19.332889 N 104.932528 W	50	27
9	Road to Purificación House, km 30	P-30 (I)	19.611222 N 104.849972 W	296	25


A) Ancestral allele
(30% complete)
PC2 (3.6% variance explained)
PC1 (4.9% variance explained)
PC3 (2.8% variance explained)
PC2 (3.6% variance explained)
PC4 (2.5% variance explained)
PC3 (2.8% variance explained)
PC5 (2.2% variance explained)
PC4 (2.5% variance explained)
locality
CH
LG
LS
P–30
P–35
PC
PT
PZ
ST

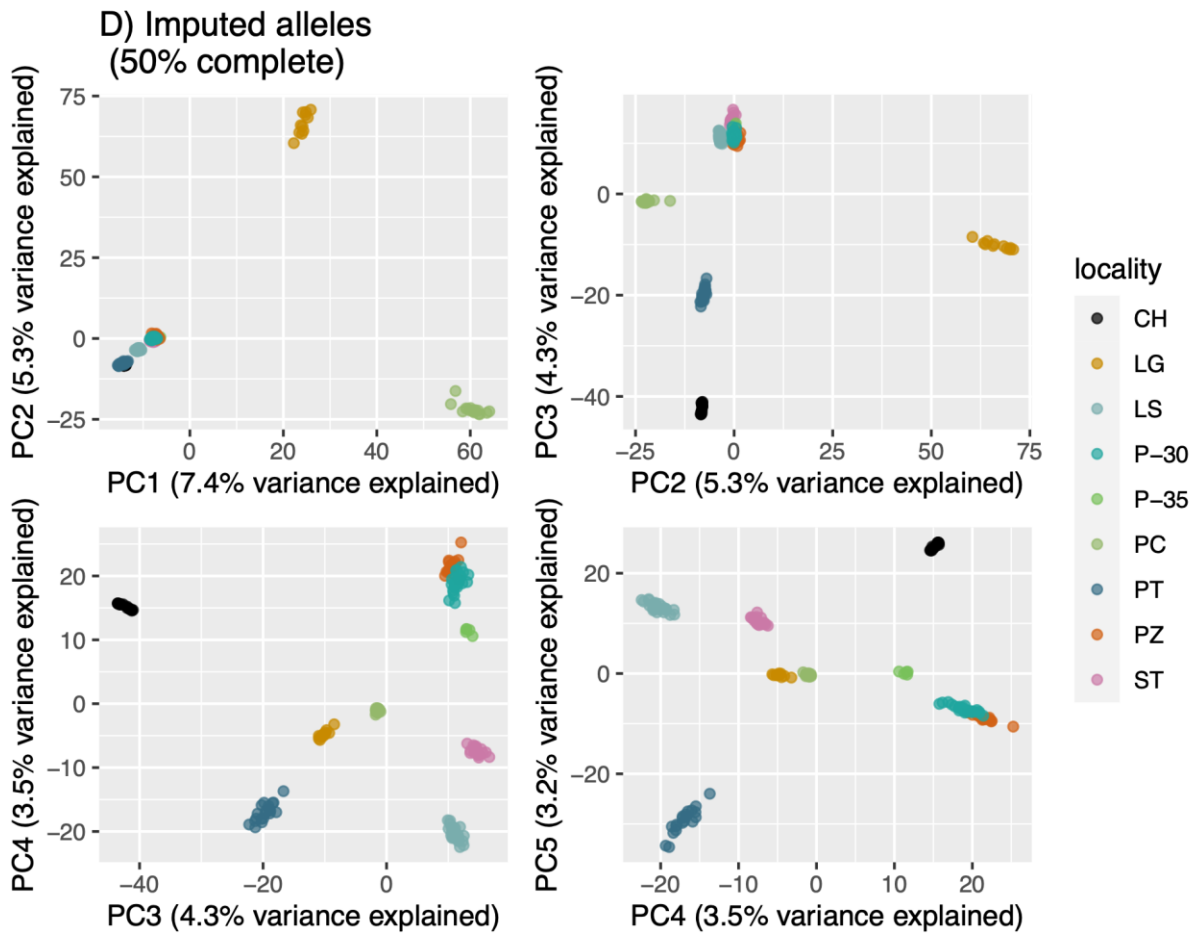
816


B) Imputed alleles
(30% complete)
PC2 (5.7% variance explained)
PC1 (8.1% variance explained)
PC3 (4.8% variance explained)
PC2 (5.7% variance explained)
PC4 (4.1% variance explained)
PC3 (4.8% variance explained)
PC5 (3.5% variance explained)
PC4 (4.1% variance explained)
locality
CH
LG
LS
P–30
P–35
PC
PT
PZ
ST

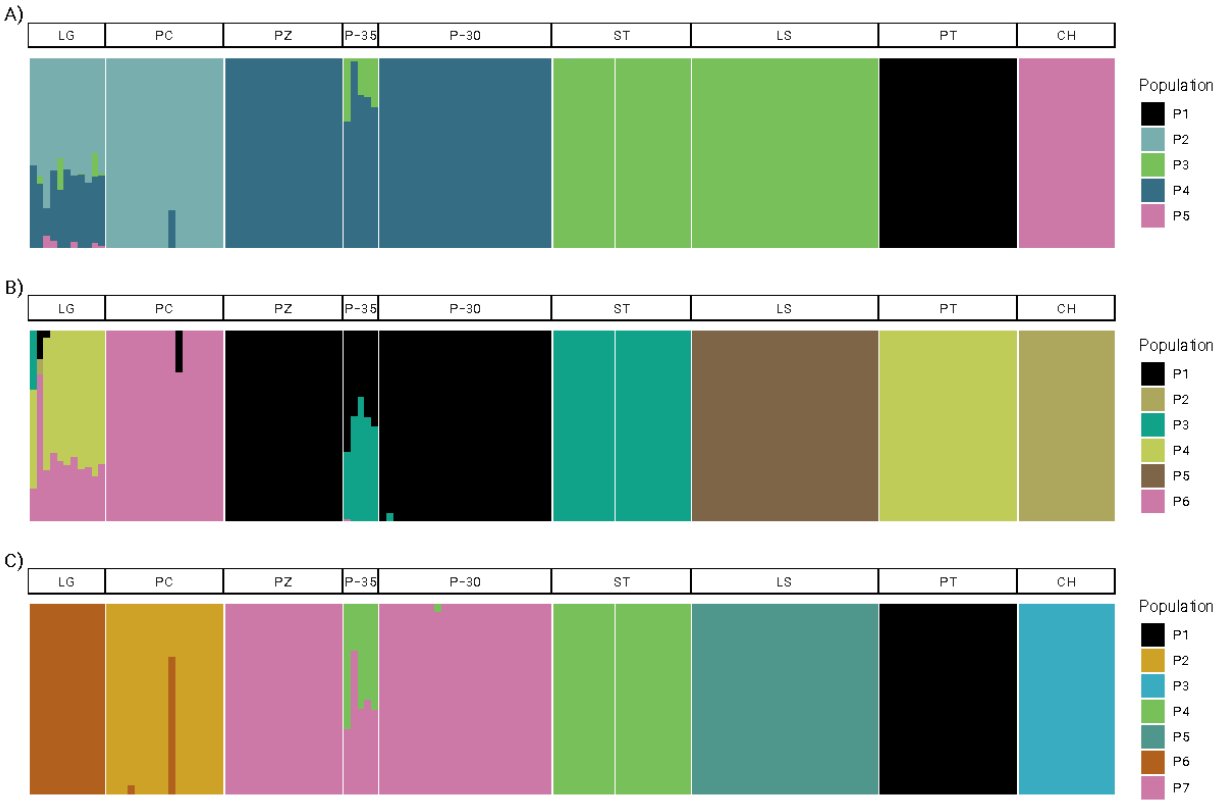
817


C) Ancestral allele
(50% complete)
PC1 (4.9% variance explained)
PC2 (3.5% variance explained)
PC3 (2.8% variance explained)
PC4 (2.5% variance explained)
PC5 (2.4% variance explained)
locality
CH
LG
LS
P–30
P–35
PC
PT
PZ
ST

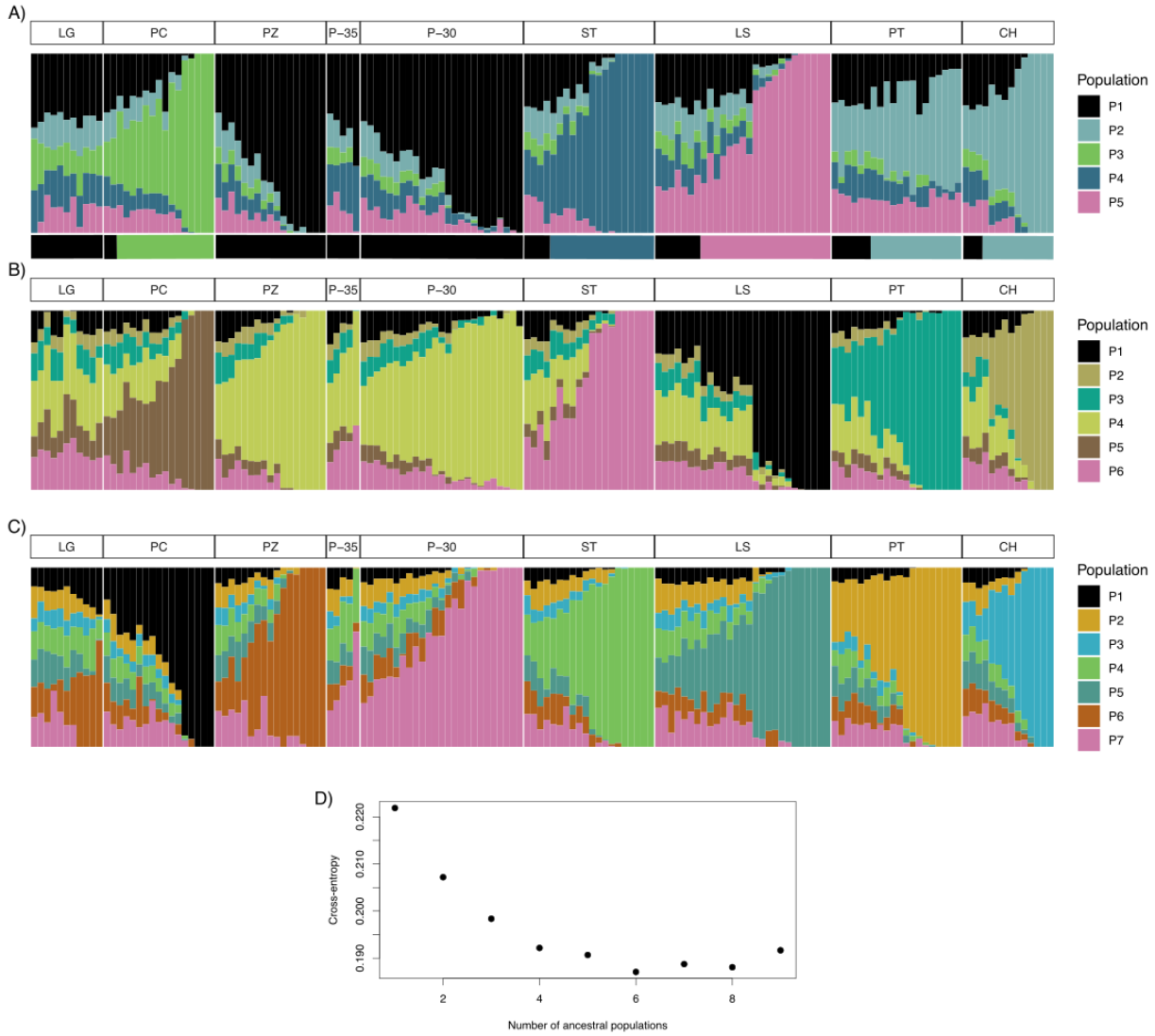
818



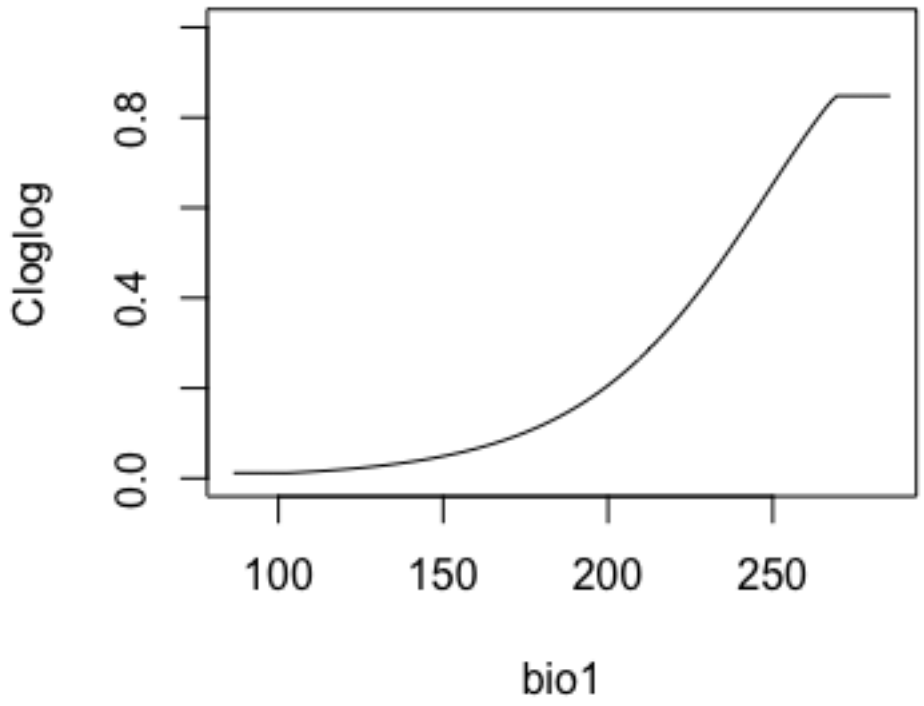
Supplementary Fig. S1. The first five principal components of the A-B) 30% complete and C-D) 50% complete genotype matrices, where missing data was either A-C) replaced with the ancestral allele, or B-D) imputed from allele frequencies sampled from the locality the individual was from. The overall patterns are highly similar between the two datasets, with tighter locality-specific clustering from the imputed matrices.



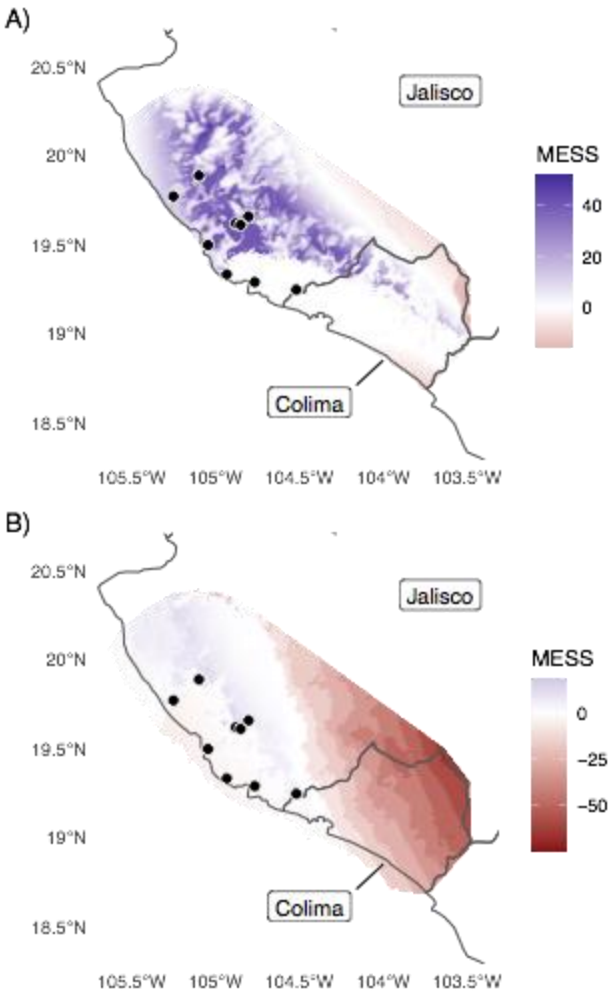
Supplementary Fig. S2. Barplots of admixture proportions for the three most supported *K* values, A) K=5 (see Fig. 2), B) K=6, and C) K=7. Evident from all is the high levels of genetic structure among sampling localities. Localities are oriented along a north to south axis, where the northernmost locality, LG, is on the left and the southernmost locality, CH, is on the right.



Supplementary Figure S3. sNMF barplots of admixture proportions for the three K values that correspond with ADMIXTURE results presented in the main text, A) K=5, B) K=6, and C) K=7. To further compare with the main results, we included ancestry proportion bars under panel A). While sNMF indicates higher proportions of individual admixture, we acknowledge that sNMF tends to perform poorly with the high levels of missing data we used in our analyses. Overall, the sNMF population assignments correspond with those in the ADMIXTURE analysis, although the corresponding populations sometimes occur for differing K-values. This is expected, given the stochastic nature of both methods. Localities are oriented along a north to south axis, where the northernmost locality, LG, is on the left and the southernmost locality, CH, is on the right.



Supplementary Fig. S4. Response curve for average annual temperature (bio1; °C * 10), the single feature used in our top ecological niche model (ENM). The *y*-axis represents the probability of occurrence, a cloglog transformation of Maxent's raw output.



Supplementary Fig. S5. Multivariate Environmental Similarity Surface (MESS) maps for A) current and B) last glacial maximum (LGM) climate. More negative values (red) indicate environments dissimilar from the environment used in training the model. The high level of dissimilarity in the southern region of the LGM projection indicates that predictions made in this area should be interpreted with caution.