Title: The effects of climate and demographic history in shaping genomic variation across populations of the Desert Horned Lizard (*Phrynosoma platyrhinos*)

Running title: Demography and climate in a desert lizard

Keaka Farleigh[1], Sarah A. Vladimirova[1], Christopher Blair[2,3], Jason T. Bracken[1], Nazila Koochekian[1], Drew R. Schield[4,5], Daren C. Card[4,6,7], Nicholas Finger[2], Jonathan Henault[1], Adam D. Leaché[8], Todd A. Castoe[4], and Tereza Jezkova[1]

[1]Department of Biology, Miami University, Oxford, OH, 45056, USA

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

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

[4]Department of Biology, University of Texas at Arlington, Arlington, TX, 76019, USA

[5]Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, CO, 80309, USA

[6]Department of Organismic & Evolutionary Biology, Harvard University, Cambridge, MA, 02138, USA

[7]Museum of Comparative Zoology, Harvard University, Cambridge, MA, 02138, USA

[8]Department of Biology & Burke Museum of Natural History and Culture, University of Washington, Seattle, WA 98195

*Corresponding author: jezkovt@miamioh.edu

Word count: 8,595 words, excluding references

Abstract (193 words)

Species often experience spatial environmental heterogeneity across their range, and populations may exhibit signatures of adaptation to local environmental characteristics. Other population genetic processes, such as migration and genetic drift, can impede the effects of local adaptation. Genetic drift in particular can have a pronounced effect on population genetic structure during large-scale geographic expansions, where a series of founder effects leads to decreases in genetic variation in the direction of the expansion. Here we explore the genetic diversity of a desert lizard that occupies a wide range of environmental conditions and that has experienced post-glacial expansion northwards along two colonization routes. Based on our analyses of a large SNP dataset, we find evidence that both climate and demographic history have shaped the genetic structure of populations. Pronounced genetic differentiation was evident between populations occupying cold versus hot deserts, and we detected numerous loci with significant associations with climate. The genetic signal of founder effects, however, is still present in the genomes of the recently expanded populations, which comprise subsets of genetic variation found in the southern populations.

Keywords

Bayesian Phylogenetics and Phylogeography, founder effects, genotype-environment association analysis, climate adaptation, population expansion, RADseq

1. Introduction

Genotypic and phenotypic variation among populations can be shaped by many processes, including natural selection, genetic drift, and gene flow. Phenotypic variation is often attributed to local adaptation to different environmental conditions occupied by distinct populations and species. Among these environmental conditions, temperature can be a major force that drives intraspecific phenotypic variation. For example, ectothermic species display increased thermal tolerance at higher latitudes and elevation, likely as an adaptation to more extreme climatic conditions (Campbell-Staton, Bare, Losos, Edwards, & Cheviron, 2018; Kealoha Freidenburg & Skelly, 2004; Munoz et al., 2014). Intraspecific differences in phenotypic traits can be

partly plastic (Ayrinhac et al., 2004; Ragland & Kingsolver, 2008) and partly adaptive (Campbell-Staton et al., 2017). Here, the identification of loci with signatures of selection and that have associations with environmental conditions may enable the identification of genes underlying adaptive phenotypic variation (Campbell-Staton et al., 2017, 2020; Frichot, Mathieu, Trouillon, Bouchard, & François, 2014; Hoban et al., 2016). Numerous genes showing signals of thermal adaptation have been identified (Campbell-Staton et al., 2017; Frydenberg, Hoffmann, & Loeschcke, 2003; Riehle, Bennett, & Long, 2001) and linked to important molecular functions and biological processes, such as lipoprotein metabolism, vasoconstriction and vasodilation, and response to oxidative stress (Akashi, Cádiz Díaz, Shigenobu, Makino, & Kawata, 2016; Valero et al., 2014).

Other evolutionary processes, such as genetic drift and gene flow, can also affect intraspecific genetic structure. The extent to which these processes shape the genetic diversity of populations will depend on population-specific characteristics, such as population size and connectivity. For example, small, isolated populations may experience substantial divergence from other populations through genetic drift alone (Knowles & Richards, 2005). Demographic changes triggered by fluctuating environmental conditions can also significantly impact genetic variation (Hewitt, 1996). During environmental change, many species experience large distributional shifts, such as range expansions into previously unsuitable geographic regions. Range expansion can be viewed as a series of founder events where subsets of individuals colonize new areas, leading to thinning of genotypes in the direction of the expansion (Excoffier, Foll, & Petit, 2009a; Hewitt, 2000). The signal of serial founder events can be especially apparent in mitochondrial DNA (mtDNA) population structure due to combined effects of small effective population size, maternal inheritance, and fast mutation rates (Cox, Stringer, Moseley, Chippindale, & Streicher, 2018; Excoffier et al., 2009a; Jezkova et al., 2016, 2015; Streicher et al., 2016). Additionally, a phenomenon known as allele surfing occurs when populations on the expansion front are small and results in individuals of those populations disproportionately contributing their alleles to the expansion (Hoban et al., 2016) which can enable mutations to establish at expanding range margins and reach high frequencies. This can result in genetic patterns that mimic selective sweeps and confound analyses of genetic differentiation (Excoffier et al., 2009a; Lotterhos & Whitlock, 2014). The influence of range expansion on nuclear genome-wide diversity has been less explored, though it is likely that range expansion reduces standing variation of expanding populations. Reductions in variation may in turn impact the adaptive potential of founder populations, and therefore fitness in novel environments (Fedorka, Winterhalter, Shaw, Brogan, & Mousseau, 2012).

The Desert Horned Lizard (*Phrynosoma platyrhinos*) currently inhabits areas of the warm Sonoran and Mojave Deserts as well as the cold Great Basin Desert (Figure 1). During the Last Glacial Maximum (LGM;

approximately 20,000 years ago), when the extent of the global ice sheets were at their largest (Clark et al., 2009), *P. platyrhinos* only inhabited the Mojave and Sonoran deserts (Jezkova et al., 2016) which were approximately 9°C cooler than they are today (Betancourt, Van Devender, & Martin, 1990). After post-LGM warming, this species persisted *in situ* despite a large increase in temperature over several thousand years. Simultaneously, this species expanded northward along two independent colonization routes into the western and eastern portions of the Great Basin (Jezkova et al., 2016). As a result, this species currently has a large distribution encompassing highly variable conditions, ranging from some of the hottest and driest places on Earth (e.g., Death Valley) to conditions where freezing temperatures may persist for six months of the year and precipitation increases during summer months (Figure 1; Fick & Hijmans, 2017). Populations of *P. platyrhinos* exhibit spatial and temporal differences in the number and size of clutches across their range (Pianka & Parker, 1975). Northern populations often lay a single, larger clutch averaging 9.9 eggs/clutch while southern populations may lay two or more smaller clutches averaging 6.8 eggs/clutch (Pianka & Parker, 1975). Body size is also different between populations, with southern males being larger (Pianka & Parker, 1975). Though these differences may be plastic and not adaptive, reciprocal experiments on several other species of lizards showed evidence that some natural history traits exhibit signals of genetic differentiation linked to local climatic conditions (Ma, Liu, Su, Luo, Zhao, & Ji, 2019, Iraeta, Monasterio, Salvador, & Diaz, 2006, Nievarowski & Roosenburg, 1993).

The mtDNA structure of *P. platyrhinos* is consistent with range expansions from the Mojave Desert into the low-elevation western and eastern portions of the Great Basin, as indicated by a decrease in mtDNA variation in the northward direction (Jezkova et al., 2016). Thus, *P. platyrhinos* currently comprises populations that inhabit very different climatic conditions which likely impose differential selective pressures across the species range, and which may have manifested in observed differences in morphology and other natural history traits. At the same time, different population histories have likely influenced levels of genetic diversity due to serial founder effects, which may have in turn placed constraints on adaptive potential during northward range expansions.

Here, we explore how divergent climatic conditions and population histories influence range-wide genetic diversity within *P. platyrhinos*. Our objective was to assess to what extent genome-wide variation among populations of *P. platyrhinos* is driven by different population histories (e.g., stable versus expanding populations) and differential climatic conditions (e.g., warm Mojave and cold Great Basin deserts; Figure 1). We predicted that the geographic signal of a postglacial range expansion should result in western and eastern Great Basin populations harboring random subsets of alleles from the southern populations and exhibiting lower levels of within-population genome-wide diversity. Loci underlying phenotypes under natural selection,

however, may show shifts in allele frequencies or allelic divergence corresponding to the environmental divergence between Mojave and Great Basin deserts.

We assess genetic variation of lizards throughout the entire range of the species using a SNP dataset derived from double digest restriction-site associated DNA sequencing (ddRADseq; Peterson, Weber, Kay, Fisher, & Hoekstra, 2012). We first assess the overall population genetic structure within *P. platyrhinos* and conduct tests of demographic history. To identify regions of the genome putatively underlying patterns of local adaptation, we scanned for loci with signals of high genetic variation (i.e., outlier loci) and loci significantly associated with climatic conditions via genotype-environment association analyses (Forester, Jones, Joost, Landguth, & Lasky, 2016; Frichot, Schoville, Bouchard, & François, 2013). We leverage a new, chromosome-level genome assembly for *P. platyrhinos* for mapping RAD loci and to explore the genomic distribution of loci affected by selection along individual chromosomes.

2. Materials and Methods

2.1 Reference genome

A chromosome-level reference genome for *P. platyrhinos* was recently generated as part of a separate study (Koochekian et al., *in review*), and is available at DDBJ/ENA/Genbank (BioProject PRJNA685451). This genome was sequenced using Illumina short paired-end reads and assembled and scaffolded using Dovetail Genomics Hi-C and Chicago long-range contact data (Table S1). The resulting genome assembly had a total length of 1,901.85 Mb, scaffold N50 length = 273.213 Mb, and included 17 large scaffolds (containing 99.56% of the genome assembly) that were assigned to chromosomes. Here, we use this high-quality genome assembly to map and interpret RADseq data.

2.2 Laboratory methods and SNP identification

We acquired tissue samples (toe or tail tissues) from 72 individuals of *P. platyrhinos* from 32 localities (1-5 samples / locality) across the species range spanning the Mojave, Sonoran, and Great Basin deserts (Figure 1; Jezkova et al., 2016). One locality (locality 7, Figure 1) represented a lineage *P. goodei*, which was split from *P. platyrhinos* and elevated to a species status based on mitochondrial (mtDNA) sequence data and morphology (Mulcahy, Spaulding, Mendelson, & Brodie, 2006). We extracted genomic DNA from these samples using the paramagnetic bead protocol (Faircloth, Glenn, & White, 2014). We generated double-digest RADseq libraries for all samples using the protocol of Peterson et al. (2012) with minor adjustments detailed below. We used 0.5 µg genomic DNA per sample that was digested overnight at 37°C using a rare-cutting restriction enzyme *Sbf*I (5'-CCTGCA*GG -3'; New England Biolabs Inc.) and a common-cutting restriction

enzyme *Msp*I (5'-C*CGG-3'; New England Biolabs Inc.). Digested samples were purified with SeraMag beads, eluted in 40 µL of TE, and quantified using a Qubit Fluorometer. We standardized samples to a common quantity and ligated double stranded indexed DNA adapters with a barcode sequence on the 5' end of the ligated adapter and an 8 bp unique molecular identifier region (UMI) that allows filtering of PCR clones. Following adapter ligation, we pooled samples into groups of nine samples, with each sample having a specific barcoded adapter. Pooled samples were purified and size-selected using a 1.5% agarose cassette on Pippin Prep (Sage Science) for fragments within a range of 435 – 535 bp. Size selected samples were amplified using primers with an Illumina index on the 5' end of the PCR primer sequence. Indexed pools were mixed in equimolar ratios and were sequenced together using 100 bp pair-end reads on an Illumina HiSeq 2500 at the Brigham Young University DNA Sequencing Center.

We removed PCR clones, quality-filtered, and demultiplexed data using the functions clone_filter and process_radtags in Stacks v1.42 (Catchen, Hohenlohe, Bassham, & Amores, 2013). We used the dDocent pipeline for quality trimming and to map the reads to the assembled genome of *P. platyrhinos* using BWA (Li, 2013; see Table S1 for information on genome assembly). The variants produced by dDocent using FreeBayes (Garrison & Marth, 2012) were filtered following the general steps recommended in the dDocent tutorial (Puritz, Hollenbeck, & Gold, 2014; Puritz et al., 2014) using vcftools and vcflib (Danecek et al., 2011). We filtered the variants (e.g., single nucleotide polymorphisms; SNPs) to include those called in 50% of individuals (--max-missing 0.50) and with a minimum mean read depth among samples of 10 (--min-meanDP 10). Individuals with 60% or greater missing data were excluded. We also removed minor alleles with a frequency lower than 0.05 (--maf 0.05) and SNPs that were found to be in linkage disequilibrium (LD; Rho ≥ 0.96). The resulting SNP matrix was used for all analyses apart from IQ-TREE and Bayesian Phylogenetics and Phylogeography (BPP), which require full locus genotypes (e.g., variant and invariant sites). These alignments were generated by assembling demultiplexed data with ipyrad v. 0.9.10 (Eaton, 2014) using a *de novo* assembly with a clustering threshold of 0.9, minimum Phred quality score of 33, and default parameters for all other options.

2.3 Phylogeographic structure

We analyzed the intraspecific genetic structure of *P. platyrhinos* using clustering and tree-based methods. We inferred the number of genetic groups (e.g., clusters) present in the dataset using Bayesian information criterion (BIC), sparse non-negative matrix factorization (sNMF), and principal component analysis (PCA). We also estimated the contribution of each cluster to the genetic composition of each individual using sNMF implemented in the R package *LEA* (Frichot & François, 2015). We first conducted the

PCA using the function *pca* to reduce the SNP dataset to principal components (PCs). Visual inspection of the PCA was used to inform the sNMF analysis about the number of genetic clusters to consider in subsequent analysis. We then used the least-squares estimates of ancestry proportions and the entropy criterion as implemented in the function *snmf* to calculate genotype frequencies and used those to impute missing data. We then ran the final sNMF analyses for one to ten genetic clusters (referred to as latent factors), to estimate the most likely number of genetic clusters and the assignment of each individual to respective genetic clusters (Frichot et al., 2014). We also used the find.clusters function in the R package *adegenet* to evaluate the number of genetic clusters using BIC (Jombart & Ahmed, 2011). Finally, we evaluated if sample size may be driving observed ancestry proportions by testing for a correlation between ancestry proportions and sample size at each locality. For example, the test allowed us to determine if there was a pattern of greater admixture at localities with a higher number of samples and less admixture among localities with fewer samples. The correlation test was executed using the *cor* function in R (R Core Team, 2020).

We inferred a maximum likelihood (ML) genealogy from the concatenated data using IQ-TREE v. 1.6.12 (Nguyen, Schmidt, Haeseler, & Minh, 2015). We used ModelFinder (Kalyaanamoorthy, Minh, Wong, Haeseler, & Jermiin, 2017) to select the best fitting substitution model (via BIC). Support for nodes was assessed through both the ultrafast bootstrap (Hoang, Chernomor, Haeseler, Minh, & Vinh, 2018) with 10,000 replicates and SH-aLRTs (Guindon et al., 2010) with 1,000 replicates. Bootstrap values >95% and SH-aLRT values >80% were indicative of strong support. The tree was rooted using ddRADseq data from *Phrynosoma modestum*, which is the sister taxon to *P. platyrhinos* (Leaché & McGuire, 2006), downloaded from the NCBI sequence read archive (SRR2240547; Leaché, Banbury, Felsenstein, De Oca, & Stamatakis, 2015).

2.4 Niche assessments

We assessed whether the climatic conditions occupied by individual genetic clusters are significantly different from each other and whether the boundary between genetic clusters corresponds to a transition between climatically distinct regions. We evaluated the climatic niche differences between populations of *P. platyrhinos* using multivariate methods (Kozak & Wiens, 2006; Mccormack, Zellmer, & Knowles, 2010). We first assigned each sampling locality to a genetic cluster based on the genetic cluster analyses (see Results). To expand our sample size, we accessed occurrence records from Jezkova et al., (2016) and assigned them to a particular genetic cluster if they fell within the convex hull of that genetic cluster. For each locality, we extracted values from 19 bioclimatic variables (Table S2; Kozak & Wiens, 2006; Waltari et al., 2007) derived from the WorldClim dataset v.1.4 with resolution of 30 seconds (Hijmans, Cameron, Parra, Jones, & Jarvis, 2005). We conducted a PCA using the *prcomp* function in R (R Core Team, 2020) to reduce the bioclimatic

variables to PCs and visualized the first two PCs. For statistical comparison of climatic niches occupied by the main genetic clusters, we calculated niche overlap using the identity test and conducted a range-breaking test using the R package ENMTools (Glor & Warren, 2011; Warren, Glor, & Turelli, 2010). The identity tests and range break tests were calculated using Schoener's *D* (Schoener, 1968) and similarity measure *I* derived from Hellinger's distance (Warren, Glor, & Turelli, 2008), which both range from 0 (no overlap) to 1 (complete overlap). The observed niche overlap was compared with a null distribution generated from 100 pseudoreplicates where occurrence records were randomly assigned to each group (identity test) or where the geographic area was divided randomly (linear range-break test; Glor & Warren, 2011).

2.5 Tests of demographic history

We compared the demographic history of the southern and northern populations by calculating four population genetic indices representative of genetic diversity and differentiation, by assessing the effective population sizes of southern and northern populations, and by testing support for 20 competing demographic models of diversification and gene flow. We tested for a signal of a range expansion from the southern deserts into northern deserts by calculating private allele frequencies, individual heterozygosity, genetic differentiation (F_{ST}) among populations, and residual genetic distances (e.g., genetic distances corrected for geographic distance) between neighboring sampling sites. Expanded populations should exhibit a genetic signal of a founder effect (Hewitt, 2000), such as a decrease in heterozygosity and private allele frequency in the direction of the expansion (Hewitt, 2000), and lower levels of population differentiation among recently expanded populations (e.g., lower F_{ST} and genetic distance; Excoffier et al., 2009a; Jezkova et al., 2015). It should be noted that in some cases, F_{ST} will not exhibit lower levels of population differentiation among expanded populations as it is dependent on variation within the expanding populations (Burri, 2017).

Private alleles were identified for each sampling locality and for each genetic cluster (see Results) using the *poppr* R package (Kamvar, Tabima, & Grunwald, 2014). We used Pearson's chi-squared test to assess significant differences in private allele frequency among the genetic clusters, and we used linear regression to test for an association between private allele frequency at each locality and latitude. We calculated average multilocus heterozygosity (MLH) for each individual using biallelic loci with 0% missing data among individuals (n = 1,950 SNPs) using the R package *inbreedR* (Stoffel et al., 2016). This method divides the number of heterozygous loci in an individual by the individual's total number of loci. For example, an MLH of 0.2 means that 20% of an individual's loci are heterozygous (Stoffel et al., 2016). We also investigated if genetic cluster assignment may influence heterozygosity results by rerunning this analysis using only individuals that exhibited no admixture among different genetic clusters (see Results).

We calculated F_{ST} (Weir & Cockerham, 1984) values for each pair of localities using the function *genet.dist* in the R package *hierfstat* (Goudet, 2005), excluding localities represented by only one individual. We tested whether levels of MLH and F_{ST} differ among the genetic clusters using Kruskal-Wallis and Dunn's *post-hoc* tests. Significance was evaluated based on p-values adjusted for multiple tests using Benjamini-Hochberg method (Benjamini & Hochberg, 1995). To measure genetic differentiation while accounting for geographical distance between populations, we calculated residual genetic distance using ALLELES IN SPACE v. 1.0 (Miller, 2005). Residual genetic distances were assigned to midpoints between adjacent localities using the Delaunay triangulation-based connectivity network, imported into ArcGIS, and interpolated across the geographic range of *P. platyrhinos*. We tested whether residual genetic distances significantly differ among the genetic clusters using analysis of variance (ANOVA).

We estimated current and ancestral population sizes and divergence times of the clades identified with IQ-TREE using the program Bayesian Phylogenetics and Phylogeography (BPP; Flouri, Jiao, Rannala, & Yang, 2018). BPP is a Bayesian Markov chain Monte Carlo (MCMC) program for analyzing DNA sequences under the multispecies coalescent model (MSC; Yang, 2015). The number of genetic clades was set to either three (S, E, W; based on sNMF) or four (S1, S2, E, W; based on IQ-TREE). For each analysis, we generated a dataset of 4 individuals per clade with evidence of little to no admixture. Two independent fixed tree parameter inferences (A00) were conducted to increase reliability of estimates of ancestral population sizes (Yang, 2015). All BPP analyses used inverse gamma priors for θ (IG[3,0.004]) and τ (IG[3,0.05]) parameters. Runs were based on an initial burn-in of 50,000 generations, a sampling frequency of twenty generations, and 50,000 samples from the posterior. Convergence was assessed by examining parameter estimates among independent runs and ESS values in Tracer v.1.6.0 (Rambaut, Drummond, Xie, Baele, & Suchard, 2018) for A00 analyses. Effective samples size (ESS) values were above 200 for most parameters in combined runs and represent the number of independent samples in the MCMC (Table 1).

We tested competing hypotheses for patterns of diversification and gene flow between southern and northern populations by simulating three dimensional (3D) and four dimensional (4D) joint site frequency spectra (JSFS) of genetic variation using the program Moments (Jouganous, Long, Ragsdale, & Gravel, 2017). Moments uses differential equations to simulate the evolution of allele frequency distributions over time and is closely related to the diffusion approximation method used in the program ∂a∂i (Gutenkunst, Hernandez, Williamson, & Bustamante, 2009). We tested 20 demographic models that differ in the number of lineages, patterns of divergence, and gene flow (Figure S1). For this analysis, all SNPs that were identified as candidates by outlier tests or genome environment association analyses (see below) were removed, as linked selection can lead to demographic mis-inference (Johri, Riall, Becher, Charlesworth, & Jensen, 2020). The number of

lineages was set either to three (S, E, W, based on population cluster analyses) or to four (S2, S1, E, W; based on the phylogenetic tree generated by IQ-TREE; see Results). The pattern of divergence was set to a simultaneous split among the three or four lineages, or to bifurcating divergence as suggested by the IQ-TREE. For each scenario, we assumed no gene flow, symmetrical gene flow, or asymmetrical gene flow, either among all or adjacent populations. Models were optimized and run following the methods of Leaché et al. (2019), and Python scripts for performing model fitting were downloaded from github.com/dportik/moments_pipeline (Portik et al., 2017). We performed demographic modeling with folded JSFS, meaning that we did not include an outgroup sequence. Four rounds of optimizations were performed for each model; parameters from the best replicate were used as starting values for the next round of optimization. Replicate analyses were conducted to ensure that the optimization routine was stable. Before final analysis, the replicate with the highest likelihood for each model was used to calculate Akaike information criterion (AIC) scores, ΔAIC scores, and Akaike weights (wAIC; Burnham & Anderson, 2004), and the model with the lowest AIC score was considered the best model. To maximize the number of segregating sites for analysis, populations were projected down to smaller sample sizes of 30 for 3D models and 8 for 4D models. These methods address missing data by averaging over all possible re-samplings of the larger samples size data and allow one to set a lower bound on the number of SNPs necessary to include a SNP in the analysis. This projection results in some SNPs becoming uninformative, and the authors suggest choosing a projection that maximizes the number of segregating sites for analysis.

2.6 Outlier analyses and genotype-environment association

We used two approaches to search for potential signals of selection in our dataset. First, we identified outlier loci that exhibit greater differentiation among populations than what is observed genome-wide (Luu, Bazin, & Blum, 2017) using a standard "outlier scan" approach. Outlier loci are expected to be enriched for loci under strong differential selection between populations, as alleles that are involved in local adaptation should occur at higher frequencies where they increase fitness (Hoban et al., 2016). Outlier scans provide flexibility by not requiring detailed climatic or natural history data and identify outlier loci by screening for alleles that exhibit greater than average genetic differentiation among populations (Hoban et al., 2016). Second, we identified SNPs significantly associated with varying environmental conditions across the species range, which are expected to enrich for regions of the genome that are important for adaptation to local climates. We scanned for outlier SNPs using the R package *pcadapt* (Luu et al., 2017). We first performed a PCA of the SNP dataset (PCA; Luu et al., 2017) and selected PCs to be retained by inspecting scree plots (Luu et al., 2017). PCA creates new independent axes (PCs) representing gradients of redundant genetic variation within the

dataset. Thus, one axis is a composite of similar variation from many loci. We then regressed individual SNPs against the retained PCs to determine the position of a particular locus (represented by a z-score) along each composite variable. The z-scores are used to calculate the Mahalanobis distance between a particular locus coordinate in PCA space and the barycenter (center of mass) of all loci, meaning that the greater the distance, the more likely it is that a locus is exhibiting a significant amount of genetic divergence, which can be indicative of selection. To determine which loci represent significant outliers, we transformed p-values associated with Mahanalobis distances into q-values using the R package *qvalue* with an expected false discovery rate of less than 1% (Storey, Bass, Dabney, Robinson, & Warnes, 2019).

Redundancy analysis (RDA) is a powerful method to detect candidate adaptive loci exhibiting strong associations with the environmental variables hypothesized to influence selection while exhibiting low false positive and high true positive rates (Forester, Lasky, Wagner, & Urban, 2018). The approach performs a PCA on the response variables (SNP matrix) while constraining the PCA axes as linear combinations of the predictor (environmental) variables. We implemented RDA to assess correlation between SNPs and environmental variables using the R package *vegan* (Oksanen et al., 2016). RDA is a constrained ordination method that is a multivariate analog of linear regression and examines the amount of variation in one set of variables that explains variation in another set. In our case, we aimed to measure how much genomic variation is explained by environmental predictor variables. Environmental variables were represented by four bioclimatic variables from WorldClim v2 (Fick & Hijmans, 2017): annual mean temperature (AMT), temperature seasonality (TS), precipitation of the warmest quarter (PWQ), and precipitation of the coldest quarter (PCQ). These variables were selected to account for major aspects of climate while avoiding autocorrelation among variables (see Appendix S1 for methods on selecting variables; Dormann et al., 2013). We conducted an additional analysis with latitude as a response variable. The significance of the entire model and each axis was evaluated using an ANOVA with 999 permutations. Effects of collinearity between environmental predictors were assessed using the function *vif.cca* to evaluate variance inflation factors. We then identified candidate SNPs based on locus scores that were ± 2.5 *SD* from the mean loading on all four constrained axes. We identified the environmental variables with the strongest associations with each candidate SNP using a Pearson's correlation coefficient (*r)*. We also identified candidate loci that were shared among RDA and the outlier analysis. To assess whether the number of shared loci is greater than what is expected due to stochasticity, we used a permutation approach to resample loci at random based on the number of empirical candidate loci identified in each of the datasets. The permutation test was executed using the sample and replicate functions in R (R Core Team, 2020). For each permutation, we randomly resampled the appropriate number of loci to reflect each dataset and calculated the number of shared loci. We generated 100,000 permutations of the data to generate a null distribution of the

number of shared loci and compared this distribution to our empirical results. If our empirical results are due to stochasticity, we would expect our observation to fall within the 95% confidence interval of our generated null distribution. Therefore, if our empirical results fall outside this area, we can conclude that there is an excess of shared loci between datasets. We also visualized the candidate SNPs shared between the RDA and outlier analysis on the genome of *P. platyrhinos* to assess the distribution of candidate SNPs across individual chromosomes. When searching for signals of selection using ddRADseq it is important to acknowledge that many of the ddRAD loci in a dataset will not sample functional regions of the genome, however, and ddRAD loci will be linked to adjacent blocks of the genome that may contain functional elements (e.g., genes; Brodie, Azaria, & Ofran, 2016). Therefore, it is important to consider both linkage disequilibrium and genome size when designing a study that seeks to identify signal of selection (McKinney, Larson, Seeb, & Seeb, 2017).

Finally, we performed functional annotation of RDA candidate SNPs to determine if SNPs exhibiting strong associations with the environmental variables may be associated with relevant biological processes. Functional annotation allows us to identify gene ontology (GO) terms associated with candidates which can link them to molecular functions, biological processes, and cellular components. Since the genome of *P. platyrhinos* has not been fully annotated, we performed functional annotation with Blast2GO which allows for the functional annotation of novel sequences (Conesa et al., 2005). We used the program BLASTX to search against the NCBI 'nr' protein database, requiring E-values below 10^{-5} for candidates to be considered during mapping and annotation. We then mapped and annotated candidates below the E-value threshold to assign GO terms.

3. RESULTS

3.1 ddRADseq and filtering

After alignment of RAD sequencing reads to the *P. platyrhinos* genome, a total of 614,332 SNPs were identified by dDocent; *de novo* assembly with ipyrad identified 32,037 SNPs. Eight samples were excluded from the dDocent dataset to maintain a high level of data completeness resulting in 64 remaining samples from 30 localities with less than 60% missing data.; all samples passed filtering in ipyrad. Sequencing efforts resulted in 4,270 contigs and 1,044,224 callable sites identified by dDocent. After filtering, ipyrad retained 23,120 loci representing 923,866 callable sites. The BPP analysis was conducted using 250 loci and the Maximum Likelihood (ML) analysis using IQ-TREE was based on a concatenated matrix of 923,866 bp. All other analyses were conducted with the dDocent dataset, after filtering for linkage disequilibrium, minor allele frequency, missing data, and read depth. The final dataset was represented by 8,094 SNPs, with an average read depth across loci per sample of 186 and an average of 7,661 assembled loci (Table S3). Therefore, if we

use the lowest linkage scenario from McKinney et al. (2017; 100 kb blocks), we expect that a majority (80-100%) of the 1,901.85 Mb genome of *P. platyrhinos* (Koochekian et al., *in review*) is sampled by the final dataset. The demultiplexed ddRADseq data are deposited at NCBI SRA (BioProject PRJNA743315).

3.2 Phylogeographic structure

The admixture, BIC, and PCA analyses suggest 3-5 genetic clusters (Figure 2a, 2b, S2). After further investigation, we chose to use a K of 3 to represent the major genetic clusters within *P. platyrhinos,* as higher values of K introduce additional substructure rather than geographically meaningful clusters of populations (Figure S3). The three clusters correspond to the geographic extents of the Mojave Desert, and the western and eastern arms of the Great Basin Desert (herein, we refer to these as Southern, Western, and Eastern genetic clusters). Evidence of genetic admixture among the three clusters is greatest in areas where populations belonging to the different clusters come into close proximity. Most populations belonging to the Southern genetic cluster possess traces of the other two clusters. Levels of admixture within the Western and Eastern cluster decreases with latitude, with the most northern populations showing no signal of admixture (Figure 1,2a). Samples that have been previously assigned to *P. goodei* based on mtDNA data (Jezkova et al., 2016) are nested within the southern genetic cluster. Tests for correlation between sample size and ancestry proportions did not reveal any significant relationships (all correlation coefficients were less than 0.7). However, it should be noted that some localities (e.g., locality 44 and 58) are only represented by a single individual and may not fully represent the ancestry proportion of that population. For comparison, we also analyzed the data using 4 and 5 genetic clusters. The traces of these additional genetic clusters can be found in individuals throughout the species range and do not separate samples that had been previously assigned to *P. goodei* (Figure S3).

The ML analysis (Figure 2c) was based on 923,866 bp, from which 905,985 sites were constant and 6,412 were parsimony informative, in addition to 145,912 singleton sites. ModelFinder chose the K3Pu+F+R6 model as the best substitution model according to BIC. The ML tree suggested that the Western and Eastern clusters each represents a well-supported clade with a sister relationship between them and nested within the Southern cluster. The Southern cluster is not monophyletic and comprises two moderately supported clades. Two samples, 27-816 and 23-7387, do not belong to either clade, with the former supported as sister to all other samples and the latter being sister to the clade comprising the Western and Eastern clusters. The samples representing *P. goodei* (population 7) form a clade within one of the Southern clades that is sister to the lineage comprising other samples from Arizona (populations 2 and 5).

3.3 Niche assessments

We compared the climatic niches occupied by lizards from the Southern genetic cluster (n=44), Western genetic cluster (n=14), and Eastern genetic cluster (n=17). The climatic niche overlap (identity test) was significantly lower between the Southern and Western cluster (D=0.09 p-value=0.02, *I*=0.26 p-value=0.02) and Southern and Eastern cluster (D=0.14 p-value=0.02, *I*=0.31 p-value=0.02), and higher, but not significantly so, between the Western and Eastern clusters (D=0.37 p-value=.24, *I*=0.54 p-value=0.20). The climatic conditions occupied by lizards that belong to the Southern genetic cluster are significantly different from the lizards that belong to Eastern and Western genetic cluster, whereas the Western and Eastern clusters are not significantly different from each other (Figure 3a, 3b; Table 2, S5). The range-break test showed that niche overlap that was observed empirically between the Southern and Western genetic cluster and between the Southern and Eastern genetic cluster is low in comparison with many alternative geographic divisions on the region. However, the empirical overlap is not the lowest possible, which indicates that at least some alternative divisions would result in similar niche overlaps (Figure 3c; Table 2).

3.4 Tests for demographic expansion

Private allele frequency was significantly lower within the Western and Eastern genetic clusters compared to the Southern cluster ($df = 2$, $\chi^2 = 589.51$, $p < 0.001$). However, private allele frequency was not significantly correlated with latitude ($R^2 = 0.209$, $p = 0.06502$). Levels of multilocus heterozygosity (MLH) differed among the three groups (Kruskal-Wallis test: $H = 8.9935$, $df = 2$, $p = 0.01$; Table S6; Figure S4). In contrast with our expectations that heterozygosity would be highest in the Southern genetic cluster, heterozygosity was highest in the Eastern genetic cluster. Additional investigation to determine if the observed results could be due to genetic cluster assignment revealed the same pattern. Levels of MLH were estimated to be 0.15 for the southern cluster, 0.24 for the eastern cluster, and 0.14 for the western cluster; clusters were represented by 6, 5, and 7 samples, respectively. Pairwise comparisons revealed significant differences in heterozygosity between the Eastern and Southern clusters (adjusted $p = 0.0064$) and Eastern and Western clusters (adjusted $p= 0.0106$), but not between the Southern and Western clusters (adjusted $p = 0.4435$). Mean F_{ST} values differed among the three genetic clusters (Kruskal-Wallis test: $H = 12.828$, $df = 2$, $p = 0.001639$; Table S6; Figure S5), with lower mean F_{ST} within the Eastern genetic cluster (i.e., among populations within a cluster; $F_{ST} = 0.0227 - 0.0985$, mean = 0.0573) compared to the Western ($F_{ST} = 0.1801 - 0.4128$, mean = 0.2834) and Southern ($F_{ST} = 0.0199 - 0.4865$, mean = 0.2131) clusters. There were significant differences in F_{ST} between the Eastern and Southern clusters (adjusted $p = 0.0016$) and Eastern and Western clusters (adjusted $p= 0.0010$), but not between the Southern and Western clusters (adjusted $p = 0.0836$). Mean F_{ST} values between genetic clusters were also significantly elevated (Kruskal-Wallis test: $H = 23.054$, $df = 2$, $p = 9.858e\text{-}06$; Table S7; Figure S5), with

higher genetic differentiation between the Southern and Western clusters (F_{ST} = 0.2062 - 0.5855, mean = 0.4314) than the Southern and Eastern clusters (F_{ST} = 0.2340 - 0.5288, mean = 0.3652, adjusted p = 0.0025). The residual genetic distances among neighboring sites support differentiation of populations into three distinct genetic clusters, which contrasts with distances derived from mtDNA (Figure 4a). The ANOVA showed that residual genetic distances are not significantly different within the three genetic clusters (p = 0.396).

Results from the BPP analysis (Figure 4c, Table 1) suggest that there was an increase in effective population size following the split between S2 and S1, E, and W clusters. Subsequent splits between clusters led to decrease in effective population sizes. The largest modern effective population size is inferred for the Southern clusters, with similar, smaller, estimated effective population sizes for the Great Basin clusters. BPP analysis with 3 clusters support a similar pattern with modern population estimates being smaller than ancestral populations (Table 1). However, three cluster analysis suggests that the ancestral EW cluster grew after diverging from the S cluster (Table 1). Like the four-cluster analysis, the largest effective population size is inferred for the Southern cluster, with smaller estimates for the Great Basin clusters.

The population models with Moments suggest that of the 10 models tested for 3 lineages (S, E, W), the best fit was for the model that suggests an initial split between Southern and Great Basin clusters, then another split between Great Basin clusters into the modern clusters with asymmetric migration between all clusters. (Figure 4b, S1; Table S8). Effective population size estimates were similar between all clusters. The Southern and Eastern clusters were estimated to be slightly larger than the Western cluster (Table S8). Migration estimates varied with the greatest amount of migration being from the Western cluster into the Southern cluster and from the Southern cluster into the Eastern cluster. Migration rates from the Eastern cluster were smaller, however, the lowest migration rate estimates were from the Southern cluster into the Western cluster and from the Western cluster into the Eastern cluster (Table S8). Of the 10 models tested for 4 groups (S2, S1, E, W), the best fit was for the model that followed the topology inferred by the ML analysis with asymmetric migration within the Southern and Great Basin clusters but no migration between them (Figure 4b). Effective population size estimates varied, with the largest estimated cluster being S1 and the smallest being W; the S2 and E clusters had similar population size estimates (Table S8). Migration rate estimates differed between the Southern and Great Basin clusters. Estimates suggest that migration is much greater between the Southern clusters than the migration between the Great Basin clusters (Table S8).

3.5 Outlier scans and genome-environment associations

Outlier tests, which identify putatively selected loci that exhibit greater differentiation among populations than expected, were based on 12 PCs that explain 86% of variation in the SNP dataset. Using this approach, we

identified 731 outlier loci. Genome-environment association analysis using RDA, which assesses association between alleles and environmental conditions, identified 292 candidate loci (Figure 5a; Table S9). Candidate loci were correlated with all four environmental variables, 134 were most associated with AMT, 92 with PCQ, 25 with PWQ, and 45 with TS. Within RDA analysis, all four constrained axes were significant (Table S9). The four environmental variables (AMT, TS, PWQ, PCQ) explained 21% of variance within the SNP dataset (Adjusted R-squared = 0.21; Table S10). For the RDA dataset, most loci inferred to be under selection were associated with annual mean temperature (n=134). Latitude alone only explained 8% of variance with 34 identified outliers, suggesting that neutral allele surfing, that is expected to have occurred in the northward direction, cannot account for all the detected outliers.

Twenty-six loci were identified by both *pcadapt* and RDA analyses. These values were greater than what is expected due to random chance, with our permutations only identifying an expected overlap of three loci 8 times, two loci 310 times, and one locus among analyses in 7,630 of 100,000 replicates and no overlap in the remaining replicates (Figure S7). Outlier loci were found dispersed across the genome with 10 loci on chromosome 1, 12 loci on chromosome 2, 2 loci on chromosome 3, 8 loci on chromosome 4, 5 loci on chromosome 6, and 12 loci on the microchromosomes (Figure 5d). Functional annotation with Blast2GO identified 13 candidates that were associated with gene ontology (GO) terms. The remaining 283 candidates were either not associated with GO terms (did not map or annotate) or did not result in any BLASTX results below the E-value threshold of 10^{-5}. Candidates were assigned GO terms associated with all three GO categories (molecular function, biological process, cellular component) and some candidates were assigned multiple GO terms (Table S11). Candidates GO terms included associations with DNA metabolism, binding, transcription, recombination, integration, and RNA-directed DNA polymerase activity (Table S11).

4. Discussion

Climate-imposed selection frequently leads to local adaptation across latitudes (Colautti & Barrett, 2013). However, areas of large climatic differences along latitudinal gradients also correspond to areas with pronounced climate fluctuations through time (Sandel et al., 2011). Fluctuating climate often results in large-scale shifts in species distributions, with some of the most drastic changes having occurred during the glacial-interglacial cycles of the Pleistocene (Hewitt, 1999). As a result, populations throughout the range of some species exhibit different histories; some populations have persisted in place for long periods of time whereas others were established only recently, following a range expansion into new areas (Jezkova et al., 2016, 2015). During large-scale geographic expansions, genetic drift is believed to have a pronounced effect on population

genetic structure through a series of founder effects that result in the decrease of standing variation in the direction of the expansion. This has been demonstrated in patterns of neutral (or nearly neutral) mtDNA variation that shows a significant decrease in genetic diversity within populations and decrease in differentiation among populations (Excoffier et al., 2009a; Jezkova et al., 2016; Mulcahy, 2008). Additionally, allele surfing enables mutations and standing variation present in populations at the expanding range margins to increase in frequency (Excoffier et al., 2009a; Klopfstein, Currat, & Excoffier, 2006). This potentially decreases the adaptive potential of expanding populations and may even decrease the fitness of populations if mutations are deleterious (Burton & Travis, 2008). Allele surfing can also mimic the signatures of selective sweeps, confounding analyses aimed at detecting loci under selection (De Villemereuil, Frichot, Bazin, François, & Gaggiotti, 2014; Excoffier, Hofer, & Foll, 2009b). At the same time, postglacial range shifts often include expansions into new habitats with contrasting climates, habitats, and communities, resulting in exposure to new selection pressures. It is therefore an open question how intraspecific genetic diversity is being shaped in the context of range expansion into novel environmental conditions.

Here we explored intraspecific genetic diversity of a desert lizard that occupies a wide range of environmental conditions and that has experienced northward post-LGM expansion. Our results are consistent with the hypothesis that the overall genetic variation is being affected by the species demographic history as well as selection pressures on populations that occur across divergent environments (Hereford & Winn, 2008; Kolbe, Leal, Schoener, Spiller, & Losos, 2012). Indeed, pronounced genetic differentiation between the Mojave and Great Basin populations indicates that differential environmental conditions in the two deserts have influenced overall population genetic structure. This stands in sharp contrast with genetic structure inferred solely from mtDNA (Jezkova et al., 2016) that is consistent with the expectations of a stepping-stone founder effect proposed for large-scale geographic expansions (Figure 4a; Excoffier et al., 2009a; Hewitt, 2000). Importantly, most Mojave individuals exhibit allelic variation that indicates ancestral variation also shared with both Great Basin genetic clusters. This indicates that allele frequencies of Great Basin populations originated from the standing variation of southern populations rather than new mutations emerging during the expansion, emphasizing the importance of standing variation on a species ability to adapt to novel environments (Barrett & Schluter, 2008; Hedrick, 2013).

Our inference of selection pressures caused by divergent environments is further supported by numerous outlier loci exhibiting high amounts of genetic differentiation and additional loci with significant associations with environmental variables (both temperature and precipitation). Thirteen of these candidate loci were located near genes associated with relevant GO functions. The importance of temperature and precipitation as environmental factors is well characterized, as several studies have identified climate to drive

selection among populations (Bay et al., 2018; Prates, Penna, Rodrigues, & Carnaval, 2018). The four climatic variables used in this study accounted for 21% of variance in genetic structure of *P. platyrhinos*. Although genetic drift and allele surfing along a latitudinal gradient can produce a false association between climate and genetic variation (Excoffier et al., 2009a), it is unlikely that this accounted for all loci significantly associated with climate, as latitude alone only accounted for 8% of genetic variation. The contribution of climate in shaping the genomic structure of *P. platyrhinos* could be indicative of extreme environmental divergence among populations, corroborated by our niche assessments that suggests large climatic differences between Mojave and Great Basin deserts with no geographic barriers separating the Great Basin populations from the southern populations. Although the climatic differences between Western and Eastern Great Basin are not significantly different (and likely do not represent the driving force of genetic differentiation between the two regions, see below), the colder temperatures and higher precipitation in the Eastern Great Basin may impose additional differential selection pressures within the Great Basin (Figure 5b, 3; Fick & Hijmans, 2017). In addition to climate, natural history traits, such low dispersal rates and associated low levels of gene flow, may lead to faster fixation of beneficial alleles in different regions (Ujvari, Dowton, & Madsen, 2008).

The two methods used here to identify candidate loci under selection, outlier scans and genotype-environment associations (GEA), yielded different sets of loci, with only a small subset of loci identified by both tests. This overlap, however, was higher than what would be expected by chance. Moreover, generally low overlap between the two methods is not surprising as each approach searches for loci that may be under selection in a different way. The GEA methods such as RDA look for loci that are significantly associated with environmental predictors, and are able to identify loci that exhibit large differences as well as gradual changes across the landscape due to spatially heterogenous environments, which are common for climate related traits. (François, Martins, Caye, & Schoville, 2016; Le Corre & Kremer, 2012). Conversely, outlier-based approaches look for loci that are significantly more differentiated than the majority (background) of loci, and therefore also have the potential to capture loci that have experienced selective sweeps that may or may not be associated with climate (Capblancq, Luu, Blum, & Bazin, 2018). Simulation-based studies have found that GEAs have in general more power to detect loci that may be under selection than outlier-based approaches, but have higher rates of false positives (Hohenlohe, Hand, Andrews, & Luikart, 2018). Among the different GEA methods, the one used here (i.e., RDA) have been shown to exhibit a superior combination of low false-positive rates and high true positive-rates across scenarios of weak, moderate, and strong selection of individual loci on a landscape. Results were robust to population structure, demographic history (including equilibrium isolation by distance and nonequilibrium isolation by distance from one or two refugia), and sampling design (Forester et al. 2018).

Although selection likely plays a significant role in shaping intraspecific genetic structure of *P. platyrhinos*, the neutral genetic signal of postglacial range expansion is still present in the genomes of Great Basin individuals, supporting the importance of genetic drift and migration in shaping the genetic structure of expanding populations. The Western and Eastern populations represent two distinct genetic clusters, and many outlier loci are unique for one or the other colonization route. This is likely the result of varying standing genetic variation available during the expansion and allele surfing along each colonization route. Three lineage demographic models suggest gene flow between Mojave and Great Basin populations, with asymmetrical bi-directional gene flow along the two expansion routes and more limited gene flow between the Eastern and Western parts of the Great Basin. Contrastingly, four lineage models suggest no gene flow between Mojave and Great Basin populations and asymmetrical bi-directional gene flow within the Great Basin and Mojave. Estimated migration rates in four-lineage models agree with three-lineage models in that migration is lower in the Great Basin than Mojave. This is expected, as the climatic conditions of the central Great Basin are generally unsuitable for *P. platyrhinos*, with the only possible contact in the extreme north. We also found evidence of a decrease in effective population size for each of the colonization routes, consistent with predictions of a range expansion. It is important to note that despite our removal of outlier loci for the demographic analyses, background selection can still influence estimates of population history. Background selection can result in an excess of rare alleles at some loci, similarly to a signal of a recent expansion (Johri et al., 2020).

Patterns of population genetic characteristics across the range of this group further illustrate how large-scale geographic expansions affect genome-wide variation within and among populations. As predicted, the expanded Great Basin populations exhibit a significantly lower number of private alleles in comparison with the stable Mojave populations, consistent with a model in which the variation in expanded populations comprises a subset of variation present in the southern stable populations. Indeed, the fact that *P. goodei*, an apparently distinct lineage of *Phrynosoma*, is nested within the Southern clade of *P. platyrhinos* provides strong support for the long-term diversification of *P. platyrhinos* in the Mojave Desert, and suggests that taxonomic revisions may be necessary. Our genomic data are also consistent with patterns of mtDNA structure in *P. platyrhinos*, which also show rapid decreases of genetic diversity and differentiation (Figure 4a; Jezkova et al., 2016) in expanded populations. The loss of genetic diversity in the direction of expansion has also been suggested by simulation studies (Slatkin & Excoffier, 2012) and documented in multiple empirical studies (Bélouard, Paillisson, Oger, Besnard, & Petit, 2019; Jezkova et al., 2015). Interestingly, the recently expanded populations do not exhibit lower heterozygosity and lower differentiation among populations (as measured by F_{ST} and residual genetic distance). The reason why heterozygosity is highest in Eastern genetic clusters that

represent recently expanded populations is unclear. A possible explanation for this finding is that the colonization front in the eastern Great Basin was wider, which may have facilitated expansion of genetically differentiated individuals. Subsequent contact between these individuals may therefore lead to increased levels of heterozygosity. Additionally, the loss of genetic diversity has been shown to be less severe when dispersal rates increase with the population density because many more migrants from the diverse, high-density regions arrive at the expansion edge (Birzu, Matin, Hallatschek, & Korolev, 2019). Regardless of the mechanism, the high levels of heterozygosity represent positive signs for the genomic 'health' of the expanding populations in terms of fitness and adaptability.

In summary, the current phylogeographic and population genetic structure of *P. platyrhinos* seems to be a result of both divergent environmental conditions and demographic history associated with the species range expansion. Our study highlights the importance of standing variation as a genetic substrate that selection acts on, and alternative routes that selection can take when subsets of this variation are available. Individual heterozygosity is not necessarily reduced during range expansion, which may have important fitness consequences when populations face environmental change. Finally, further interrogation of loci with signals of selection may lead to the discovery of new genes involved in thermal adaptation, genetic regions important for adaptation to novel environments generally, or corroborate the importance of known thermal-related genes (Valero et al., 2014).

Acknowledgments

We acknowledge and thank Dr. Andor Kiss of the Center for Bioinformatics & Functional Genomics (CBFG) at Miami University for instrumentation and computational support. Dr. Jens Mueller of the Research Computing group at Miami University provided invaluable assistance with the Miami RedHawk Cluster. We also thank Dr. Jon Puritz for helpful advice with dDocent, Dr. Brenna Forester for advice with RDA, Dr. Olivier Francois for help with LEA We would like to thank Dr. Jamie Kass, Dr. Mark Miller, Dr. Aaron Ragsdale, and Dr. Dan Warren for their guidance in utilizing their software. We thank Mason Murphy for help proofing this manuscript. This project was supported by Miami University start-up funds to TJ and the National Science Foundation Graduate Research Fellowship Program to KF (Award #2037786).

Data Accessibility

Data and scripts are available from the Dryad Digital Repository https://datadryad.org/stash/dataset/doi:10.5061/dryad.79cnp5hvz. Python scripts to perform demographic

modelling are available at https://github.com/kfarleigh/Moments. Processed and demultiplexed RADseq data are available at NCBI SRA (Bioproject PRJNA743315). The genome assembly is deposited at DDBJ/ENA/Genbank (Bioproject PRJNA685451).

Author Contributions

K.F. and T.J. designed the study, K.F., S.V., C.B., N.F., J.B., N.K., J.H., T.J., D.S., D.C., and A.L. performed analyses, K.F. and T.J. wrote the paper. All authors contributed to writing and editing the manuscript.

References

Akashi, H. D., Cádiz Díaz, A., Shigenobu, S., Makino, T., & Kawata, M. (2016). Differentially expressed genes associated with adaptation to different thermal environments in three sympatric Cuban Anolis lizards. *Molecular Ecology*, *25*(10), 2273–2285. https://doi.org/10.1111/mec.13625

Ayrinhac, A., Debat, V., Gibert, P., Kister, A., Legout, H., Moreteau, B., … David, J. R. (2004). Cold adaptation in geographical populations of *Drosophila melanogaster*: phenotypic plasticity is more important than genetic variability. *Functional Ecology*, *18*(5), 700–706. https://doi.org/10.1111/j.0269-8463.2004.00904.x

Barrett, R. D. H., & Schluter, D. (2008). Adaptation from standing genetic variation. *Trends in Ecology and Evolution*, *23*(1), 38–44. https://doi.org/10.1016/j.tree.2007.09.008

Bay, R. A., Harrigan, R. J., Underwood, V. Le, Gibbs, H. L., Smith, T. B., & Ruegg, K. (2018). Genomic signals of selection predict climate-driven population declines in a migratory bird. *Science*, *359*(6371), 83–86. https://doi.org/10.1126/science.aan4380

Bélouard, N., Paillisson, J., Oger, A., Besnard, A., & Petit, E. J. (2019). Genetic drift during the spread phase of a biological invasion. *Molecular Ecology*, (August), 1–13. https://doi.org/10.1111/mec.15238

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.2307/2346101

Betancourt, J. L., Van Devender, T. R., & Martin, P. S. (1990). *Packrat middens: the last 40,000 years of biotic change*. University of Arizona Press.

Birzu, G., Matin, S., Hallatschek, O., & Korolev, K. S. (2019). Genetic drift in range expansions is very sensitive to density feedback in dispersal and growth. *Ecology Letters*, *22*(11), 1817–1827. https://doi.org/10.1111/ele.13364

Brodie A, Azaria JR, Ofran Y. How far from the SNP may the causative genes be? Nucleic acids research. 2016 Jul 27;44(13):6046-54.

Burri, R. (2017). Interpreting differentiation landscapes in the light of long-term linked selection. *Evolution Letters*, *1*(3), 118-131.

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model

selection. *Sociological Methods and Research*, *33*(2), 261–304. https://doi.org/10.1177/0049124104268644

Burton, O. J., & Travis, J. M. J. (2008). The frequency of fitness peak shifts is increased at expanding range margins due to mutation surfing. *Genetics*, *179*(2), 941–950. https://doi.org/10.1534/genetics.108.087890

Campbell-Staton, S. C., Cheviron, Z. A., Rochette, N., Catchen, J., Losos, J. B., Edwards, S. V., & .... (2017). Winter storms drive rapid phenotypic, regulatory, and genomic shifts in the green anole lizard. *Science*, *357*(6350), 495–498. https://doi.org/10.1126/science.aam5512

Campbell-Staton, S. C., Winchell, K. M., Rochette, N. C., Fredette, J., Maayan, I., Schweizer, R. M., & Catchen, J. (2020). Parallel selection on thermal physiology facilitates repeated adaptation of city lizards to urban heat islands. *Nature Ecology & Evolution*, *4*(4), 652–658. https://doi.org/https://doi.org/10.1038/s41559-020-1131-8

Campbell-Staton, S. C., Bare, A., Losos, J. B., Edwards, S. V, & Cheviron, Z. A. (2018). Physiological and regulatory underpinnings of geographic variation in reptilian cold tolerance across a latitudinal cline. *Molecular Ecology*, *27*(9), 2243–2255. https://doi.org/10.1111/mec.14580

Capblancq, T., Luu, K., Blum, M. G. B., & Bazin, E. (2018). Evaluation of redundancy analysis to identify signatures of local adaptation. *Molecular Ecology Resources*, *18*(6), 1223–1233. https://doi.org/10.1111/1755-0998.12906

Catchen, J., Hohenlohe, P. A., Bassham, S., & Amores, A. (2013). Stacks: an analysis tool set for population genomics. *Molecular Ecology*, *22*(11), 3124–3140. https://doi.org/10.1111/mec.12354

Clark, P. U., Dyke, A. S., Shakun, J. D., Carlson, A. E., Clark, J., Wohlfarth, B., … McCabe, A. M. (2009). The last glacial maximum. *Science*, *325*(5941), 710–714. https://doi.org/10.1126/science.1172873

Colautti, R. I., & Barrett, S. C. H. (2013). Rapid adaptation to climate facilitates range expansion of an invasive plant. *Science*, *342*(6156), 364–366. https://doi.org/10.1126/science.1242121

Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., & Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, *21*(18), 3674-3676. https://doi.org/10.1093/bioinformatics/bti610

Cox, C. L., Stringer, J. F., Moseley, M. A., Chippindale, P. T., & Streicher, J. W. (2018). Testing the geographical dimensions of genetic diversity following range expansion in a North American snake.

Biological Journal of the Linnean Society, *125*(3), 586–599. https://doi.org/10.1093/biolinnean/bly137

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., … Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

De Villemereuil, P., Frichot, É., Bazin, É., François, O., & Gaggiotti, O. E. (2014). Genome scan methods against more complex models: When and how much should we trust them? *Molecular Ecology*, *23*(8), 2006–2019. https://doi.org/10.1111/mec.12705

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., … Lautenbach, S. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, *36*(1), 027–046. https://doi.org/10.1111/j.1600-0587.2012.07348.x

Eaton, D. A. R. (2014). PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics*, *30*(13), 1844–1849. https://doi.org/10.1093/bioinformatics/btu121

Excoffier, L., Hofer, T., & Foll, M. (2009b). Detecting loci under selection in a hierarchically structured population. *Heredity*, *103*(4), 285–298. https://doi.org/10.1038/hdy.2009.74

Excoffier, Laurent, Foll, M., & Petit, R. J. (2009a). Genetic Consequences of Range Expansions. *Annual Review of Ecology, Evolution, and Systematics*, *40*(1), 481–501. https://doi.org/10.1146/annurev.ecolsys.39.110707.173414

Faircloth, B. C., Glenn, T. C., & White, N. D. (2014). *Illumina library prep protocol*.

Fedorka, K. M., Winterhalter, W. E., Shaw, K. L., Brogan, W. R., & Mousseau, T. A. (2012). The role of gene flow asymmetry along an environmental gradient in constraining local adaptation and range expansion. *Journal of Evolutionary Biology*, *25*(8), 1676–1685. https://doi.org/10.1111/j.1420-9101.2012.02552.x

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, *37*(12), 4302–4315. https://doi.org/10.1002/joc.5086

Flouri, T., Jiao, X., Rannala, B., & Yang, Z. (2018). Species Tree Inference with BPP Using Genomic Sequences and the Multispecies Coalescent. *Molecular Biology and Evolution*, *35*(10), 2585–2593. https://doi.org/10.1093/molbev/msy147

Forester, B. R., Jones, M. R., Joost, S., Landguth, E. L., & Lasky, J. R. (2016). Detecting spatial genetic

signatures of local adaptation in heterogeneous landscapes. *Molecular Ecology*, *25*(1), 104–120. https://doi.org/10.1111/mec.13476

Forester, B. R., Lasky, J. R., Wagner, H. H., & Urban, D. L. (2018). Comparing methods for detecting multilocus adaptation with multivariate genotype–environment associations. *Molecular Ecology*, *27*(9), 2215–2233. https://doi.org/10.1111/mec.14584

François, O., Martins, H., Caye, K., & Schoville, S. D. (2016). Controlling false discoveries in genome scans for selection. *Molecular Ecology*, *25*(2), 454–469. https://doi.org/10.1111/mec.13513

Frichot, E., & François, O. (2015). LEA: An R package for landscape and ecological association studies. *Methods in Ecology and Evolution*, *6*(8), 925–929. https://doi.org/10.1111/2041-210X.12382

Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G., & François, O. (2014). Fast and Efficient Estimation of Individual Ancestry Coefficients. *Genetics*, *196*(4), 973–983. https://doi.org/10.1534/genetics.113.160572

Frichot, E., Schoville, S. D., Bouchard, G., & François, O. (2013). Testing for associations between loci and environmental gradients using latent factor mixed models. *Molecular Biology and Evolution*, *30*(7), 1687–1699. https://doi.org/10.1093/molbev/mst063

Frydenberg, J., Hoffmann, A. A., & Loeschcke, V. (2003). DNA sequence variation and latitudinal associations in hsp23, hsp26 and hsp27 from natural populations of *Drosophila melanogaster*. *Molecular Ecology*, *12*(8), 2025–2032.

Garrison, E., & Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. *ArXiv Preprint*.

Glor, R. E., & Warren, D. (2011). Testing ecological explainations for biogeographical boundaries. *International Journal of Organic Evolution*, *65*(3), 673–683. https://doi.org/10.1111/j.1558-5646.2010.01177.x

Goudet, J. (2005). Hierfstat, a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, *5*(1), 184–186. https://doi.org/10.1111/j.1471-8278

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3 . 0. *Systematic Biology*, *59*(3), 307–321. https://doi.org/10.1093/sysbio/syq010

Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H., & Bustamante, C. D. (2009). Inferring the Joint Demographic History of Multiple Populations from Multidimensional SNP Frequency Data. *PLoS Genetics*, *5*(10). https://doi.org/10.1371/journal.pgen.1000695

Hedrick, P. W. (2013). Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation. *Molecular Ecology*, *22*(18), 4606–4618. https://doi.org/10.1111/mec.12415

Hereford, J., & Winn, A. A. (2008). Limits to local adaptation in six populations of the annual plant *Diodia teres*. *New Phytologist*, *178*(4), 888–896. https://doi.org/10.1111/j.1469-8137.2008.02405.x

Hewitt, G. M. (1996). Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*, *58*(3), 247–276. https://doi.org/10.1111/j.1095-8312.1996.tb01434.x

Hewitt, G. M. (1999). Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society*, *68*(1–2), 87–112. https://doi.org/10.1006/bijl.1999.0332

Hewitt, G. M. (2000). The genetic legacy of the quaternary ice ages. *Nature*, *405*(6789), 907–913. https://doi.org/10.1038/35016000

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, *25*(15), 1965–1978. https://doi.org/10.1002/joc.1276

Hoang, D. T., Chernomor, O., Haeseler, A. Von, Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Molecular Biology and Evolution*, *35*(2), 518–522. https://doi.org/10.5281/zenodo.854445

Hoban, S., Kelley, J. L., Lotterhos, K. E., Antolin, M. F., Bradburd, G., Lowry, D. B., … Whitlock, M. C. (2016). Finding the Genomic Basis of Local Adaptation: Pitfalls, Practical Solutions, and Future Directions. *The American Naturalist*, *188*(4), 379–397. https://doi.org/10.1086/688018

Hohenlohe, P. A., Hand, B. K., Andrews, K. R., & Luikart, G. (2018). *Population genomics provides key insights in ecology and evolution*. Spring, Cham.

Iraeta P, Monasterio C, Salvador A, Díaz JA. Mediterranean hatchling lizards grow faster at higher altitude: a reciprocal transplant experiment. Functional Ecology. 2006 Oct 1:865-72.

Jezkova, T., Jaeger, J. R., Oláh-Hemmings, V., Jones, K. B., Lara-Resendiz, R. A., Mulcahy, D. G., & Riddle, B. R. (2016). Range and niche shifts in response to past climate change in the desert horned lizard *Phrynosoma platyrhinos*. *Ecography*, *39*(5), 437–448. https://doi.org/10.1111/ecog.01464

Jezkova, T., Riddle, B. R., Card, D. C., Schield, D. R., Eckstut, M. E., & Castoe, T. A. (2015). Genetic consequences of postglacial range expansion in two codistributed rodents (genus *Dipodomys*) depend on ecology and genetic locus. *Molecular Ecology*, *24*(1), 83–97. https://doi.org/10.1111/mec.13012

Johri, P., Riall, K., Becher, H., Charlesworth, B., & Jensen, J. D. (2020). The impact of purifying and background selection on the inference of population history: Problems and prospects. *BioRxiv*. https://doi.org/10.1101/2020.04.28.066365

Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics*, *27*(21), 3070-3071.

Jouganous, J., Long, W., Ragsdale, A. P., & Gravel, S. (2017). Inferring the Joint Demographic History of Multiple Populations : Beyond the Diffusion Approximation. *Genetics*, *206*(3), 1549–1567. https://doi.org/10.1534/genetics.117.200493

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K., Haeseler, A. Von, & Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods*, *14*(6), 587–589. https://doi.org/10.1038/nmeth.4285

Kamvar, Z. N., Tabima, J. F., & Grunwald, N. J. (2014). Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, *e281*(2), 1–14. https://doi.org/10.7717/peerj.281

Kealoha Freidenburg, L., & Skelly, D. K. (2004). Microgeographical variation in thermal preference by an amphibian. *Ecology Letters*, *7*(5), 369–373. https://doi.org/10.1111/j.1461-0248.2004.00587.x

Klopfstein, S., Currat, M., & Excoffier, L. (2006). The fate of mutations surfing on the wave of a range expansion. *Molecular Biology and Evolution*, *23*(3), 482–490. https://doi.org/10.1093/molbev/msj057

Knowles, L. L., & Richards, C. L. (2005). Importance of genetic drift during Pleistocene divergence as revealed by analyses of genomic variation. *Molecular Ecology*, *14*(13), 4023–4032. https://doi.org/10.1111/j.1365-294X.2005.02711.x

Kolbe, J. J., Leal, M., Schoener, T. W., Spiller, D. A., & Losos, J. B. (2012). Founder effects persist despite

adaptive differentiation: a field experiment with lizards. *Science*, *335*(6072), 1086–1089. https://doi.org/10.1126/science.1209566

Koochekian, N., M.S., Ascanio, A., Farleigh, K., Card, D. C., Schield, D. R., Castoe, T. A., & Jezkova, T. (In review). A chromosome-level genome assembly and annotation of the Desert Horned Lizard, *Phrynosoma platyrhinos*, provides insight into chromosomal rearrangements among reptiles. *GigaScience*.

Kozak, K. H., & Wiens, J. J. (2006). Does niche conservatism promote speciation? A case study in North American salamanders. *Evolution*, *60*(12), 2604–2621. https://doi.org/10.1111/j.0014-3820.2006.tb01893.x

Le Corre, V., & Kremer, A. (2012). The genetic differentiation at quantitative trait loci under local adaptation. *Molecular Ecology*, *21*(7), 1548–1566. https://doi.org/10.1111/j.1365-294X.2012.05479.x

Leaché, A. D., Banbury, B. L., Felsenstein, J., De Oca, A. N. M., & Stamatakis, A. (2015). Short tree, long tree, right tree, wrong tree: New acquisition bias corrections for inferring SNP phylogenies. *Systematic Biology*, *64*(6), 1032–1047. https://doi.org/10.1093/sysbio/syv053

Leaché, A. D., & McGuire, J. A. (2006). Phylogenetic relationships of horned lizards (*Phrynosoma*) based on nuclear and mitochondrial data: Evidence for a misleading mitochondrial gene tree. *Molecular Phylogenetics and Evolution*, *39*(3), 628–644. https://doi.org/10.1016/j.ympev.2005.12.016

Leaché, A. D., Portik, D. M., Rivera, D., Oliver, M., Penner, J., Gvoždík, V., … Fujita, M. K. (2019). Exploring rain forest diversification using demographic model testing in the African foam-nest treefrog *Chiromantis rufescens*. *Journal of Biogeography*, *46*(August), 2706–2721. https://doi.org/10.1111/jbi.13716

Li, H. (2013). Aligning sequence reads, clon sequences and assembly contigs with BWA-MEM. *ArXiv:1303.3997v1*.

Lotterhos, K. E., & Whitlock, M. C. (2014). Evaluation of demographic history and neutral parameterization on the performance of FST outlier tests. *Molecular Ecology*, *23*(9), 2178–2192. https://doi.org/10.1111/mec.12725

Luu, K., Bazin, E., & Blum, M. G. B. (2017). pcadapt: an R package to perform genome scans for selection based on principal component analysis. *Molecular Ecology Resources*, *17*(1), 67–77.

https://doi.org/10.1111/1755-0998.12592

Ma L, Liu P, Su S, Luo LG, Zhao WG, Ji X. Life-history consequences of local adaptation in lizards: Takydromus wolteri (Lacertidae) as a model organism. Biological Journal of the Linnean Society. 2019 Apr 17;127(1):88-99.

Mccormack, J. E., Zellmer, A. J., & Knowles, L. L. (2010). Does niche divergence accompany allopatric divergence in *Aphelocoma* jays as predicted under ecologoical speciation?: Insights from tests with niche models. *Evolution*, *64*(5), 1231–1244. https://doi.org/10.1111/j.1558-5646.2009.00900.x

McKinney, G. J., Larson, W. A., Seeb, L. W., & Seeb, J. E. (2017). RAD seq provides unprecedented insights into molecular ecology and evolutionary genetics: comment on Breaking RAD by Lowry et al.(2016). *Molecular ecology resources*, *17*(3), 356-361.

Miller, M. P. (2005). Alleles In Space (AIS): Computer Software for the Joint Analysis of Interindividual Spatial and Genetic Information. *Heredity*, *96*(6), 722–724. https://doi.org/10.1093/jhered/esi119

Mulcahy, D. G. (2008). Phylogeography and species boundaries of the western North American Nightsnake (*Hypsiglena torquata*): revisiting the subspecies concept. *Molecular Phylogenetics and Evolution*, *46*(3), 1095–1115. https://doi.org/10.1016/j.ympev.2007.12.012

Mulcahy, D. G., Spaulding, A. W., Mendelson, J. R., & Brodie, E. D. (2006). Phylogeography of the flat-tailed horned lizard (*Phrynosoma mcallii*) and systematics of the *P. mcallii-platyrhinos* mtDNA complex. *Molecular Ecology*, *15*(7), 1807–1826. https://doi.org/10.1111/j.1365-294X.2006.02892.x

Munoz, M. M., Stimola, M. A., Algar, A. C., Conover, A., Rodriguez, A. J., Landestoy, M. A., … Losos, J. B. (2014). Evolutionary stasis and lability in thermal physiology in a group of tropical lizards. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1778), 20132433. https://doi.org/10.1098/rspb.2013.2433

Niewiarowski PH, Roosenburg W. Reciprocal transplant reveals sources of variation in growth rates of the lizard Sceloporus undulatus. Ecology. 1993 Oct;74(7):1992-2002.

Nguyen, L., Schmidt, H. A., Haeseler, A. Von, & Minh, B. Q. (2015). IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution*, *32*(1), 268–274. https://doi.org/10.1093/molbev/msu300

Oksanen, J., Blanchet, G. F., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., … Wagner, H. (2016). *vegan:*

Community Ecology Package. R package version 2.3-5.

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PloS One*, *7*(5), e37135. https://doi.org/10.1371/journal.pone.0037135

Pianka, E. R., & Parker, W. S. (1975). Ecology of Horned Lizards: A Review with Special Reference to Phrynosoma platyrhinos. *Copeia*, (1), 141–162. https://doi.org/https://doi.org/10.2307/1442418

Portik, D. M., Leaché, A. D., Rivera, D., Barej, M. F., Burger, M., Hirschfeld, M., … Fujita, M. K. (2017). Evaluating mechanisms of diversification in a Guineo-Congolian tropical forest frog using demographic model selection. *Molecular Ecology*, *26*(19), 5245–5263. https://doi.org/10.1111/mec.14266

Prates, I., Penna, A., Rodrigues, M. T., & Carnaval, A. C. (2018). Local adaptation in mainland anole lizards: Integrating population history and genome–environment associations. *Ecology and Evolution*, *8*(23), 11932–11944. https://doi.org/10.1002/ece3.4650

Puritz, J. B., Hollenbeck, C. M., & Gold, J. R. (2014). dDocent: a RADseq, variant-calling pipeline designed for population genomics of non-model organisms. *PeerJ*, *2*, e431. https://doi.org/10.7717/peerj.431

Puritz, J. B., Matz, M. V., Toonen, R. J., Weber, J. N., Bolnick, D. I., & Bird, C. E. (2014). Demystifying the RAD fad. *Molecular Ecology*, *23*(24), 5937–5942. https://doi.org/10.1111/mec.12965

R Core Team. (2020). *R: A language and environment for statistical computing*. Retrieved from https://www.r-project.org/

Ragland, G. J., & Kingsolver, J. G. (2008). Evolution of thermotolerance in seasonal environments: the effects of annual temperature variation and life-history timing in *Wyeomyia smithii*. *Evolution: International Journal of Organic Evolution*, *62*(6), 1345–1357. https://doi.org/10.1111/j.1558-5646.2008.00367.x

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Systematic Biology*, *67*(5), 901–904. https://doi.org/10.1093/sysbio/syy032

Riehle, M. M., Bennett, A. F., & Long, A. D. (2001). Genetic architecture of thermal adaptation in *Escherichia coli*. *Proceedings of the National Academy of Sciences*, *98*(2), 525–530.

Sandel, B., Arge, L., Dalsgaard, B., Davies, R. G., Gaston, K. J., Sutherland, W. J., & Svenning, J.-C. (2011).

The influence of Late Quaternary climate-change velocity on species endemism. *Science*, *334*(6056), 660–664.

Schoener, T. W. (1968). The Anolis Lizards of Bimini: Resource Partitioning in a Complex Fauna. *Ecology*, *49*(4), 704–726. https://doi.org/10.2307/1935534

Slatkin, M., & Excoffier, L. (2012). Serial founder effects during range expansion: a spatial analog of genetic drift. *Genetics*, *191*(1), 171–181. https://doi.org/10.1534/genetics.112.139022

Stoffel, M. A., Esser, M., Kardos, M., Humble, E., Nichols, H., David, P., & Hoffman, J. I. (2016). inbreedR: an R package for the analysis of inbreeding based on genetic markers. *Methods in Ecology and Evolution*, *7*(11), 1331–1339. https://doi.org/10.1111/2041-210X.12588

Storey, J. D., Bass, A. J., Dabney, A., Robinson, D., & Warnes, G. (2019). *qvalue: Q-value estimation for false discovery control*.

Streicher, J. W., McEntee, J. P., Drzich, L. C., Card, D. C., Schield, D. R., Smart, U., … Castoe, T. A. (2016). Genetic surfing, not allopatric divergence, explains spatial sorting of mitochondrial haplotypes in venomous coralsnakes. *Evolution*, *70*(7), 1435–1449.

Ujvari, B., Dowton, M., & Madsen, T. (2008). Population genetic structure , gene flow and sex-biased dispersal in frillneck lizards (*Chlamydosaurus kingii*). *Molecular Ecology*, *17*(15), 3557–3564. https://doi.org/10.1111/j.1365-294X.2008.03849.x

Valero, K. C. W., Pathak, R., Prajapati, I., Bankston, S., Thompson, A., Usher, J., & Isokpehi, R. D. (2014). A candidate multimodal functional genetic network for thermal adaptation. *PeerJ*, e578. https://doi.org/10.7717/peerj.578

Waltari, E., Hijmans, R. J., Peterson, A. T., Arpad, N. S., Perkins, S. L., & Guralnick, R. P. (2007). Locating Pleistocene Refugia: comparing phylogeographic and ecological niche model predictions. *PloS One*, *2*(7). https://doi.org/10.1371/journal.pone.0000563

Warren, D. L., Glor, R. E., & Turelli, M. (2008). Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. *International Journal of Organic Evolution*, *62*(11). https://doi.org/10.1111/j.1558-5646.2008.00482.x

Warren, D. L., Glor, R. E., & Turelli, M. (2010). ENMTools : a toolbox for comparative studies of environmental niche models. *Ecography*, *1*(September 2009), 607–611. https://doi.org/10.1111/j.1600-

0587.2009.06142.x

Weir, B. S., & Cockerham, C. C. (1984). Esimating F-statistics for the analysis of population structure. *Evolution*, *38*(6), 1358–1370.

Yang, Z. (2015). The BPP program for species tree estimation and species delimitation. *Current Zoology*, *61*(5), 854–865. https://doi.org/10.1093/czoolo/61.5.854

TABLE 1

Table 1 Results from combined BPP runs for 3 cluster (top) and 4 cluster (bottom) analyses. Estimates of population size (θ), divergence times (τ) and log-likelihood (lnL) are represented by mean and 95% confidence interval values. Effective sample size (ESS) represents the number of independent samples in the MCMC. E-Eastern Clade, S-Southern Clade, S_1 Southern Clade, S_2-Southern Clade composed of 6 individuals from localities 31, 34, and 54 (see Fig. 2c), W-Western Clade.

3 cluster				**θ**				**τ**			
	E	S_2	S	W	S_2,S_1, E,W	S,E,W	E,W	S_2,S_1, E,W	S,E,W	E,W	**lnL**
Mean	7.045E-4	-	1.593E-2	6.208E-4	-	5.802E-3	1.290E-2	-	1.079E-3	2.019E-4	-368455.956
2.5% HPD	5.955E-4	-	0.015	5.125E-4	-	5.466E-3	0.010	-	1.014E-3	1.670E-4	-368629.72
97.5% HPD	8.113E-4	-	0.017	7.927E-4	-	6.149E-3	0.016	-	1.147E-3	2.368E-4	-368283.01
ESS	278.900	-	5338.400	132.500	-	447.000	721.200	-	196.600	253.800	181.100
4 cluster		S_2	S_1		S_2,S_1, E,W	S_1,E,W		S_2,S_1, E,W	S_1,E,W		
Mean	6.770E-4	1.114E-2	1.080E-2	6.046E-4	5.281E-3	6.700E-2	4.530E-3	1.140E-3	4.314E-4	2.722E-4	-85823.022
2.5% HPD	4.590E-4	8.422E-3	4.512E-3	4.090E-4	4.652E-3	4.250E-4	3.360E-4	9.970E-4	2.100E-4	1.840E-4	-85886.666

97.5% HPD	8.990E-4	0.015	0.021	8.040E-4	5.918E-3	0.153	0.017	1.274E-3	1.075E-3	3.670E-4	-85756.295
ESS	442.600	9002.500	157.900	525.300	1626.600	471.300	134.400	561.300	102.800	460.400	400.400

TABLE 2

Table 2 Identity and range break test results for each pair of the three genetic clusters using Schoener's "D" and similarity measurement "I" derived from Hellinger's distance. *p*-values are based on the observed similarity value of the genetic cluster pair in comparison to the null distributions of similarity values derived from random individual assignment (identity test) and random division of combined geographic area (range break test).

	Southern x Western		**Southern x Eastern**		**Western x Eastern**	
Test	**observed value**	***p-value***	**observed value**	***p-value***	**observed value**	***p-value***
Identity - D	0.09	*0.02*	0.13	*0.02*	0.37	*0.24*
Range Break - D		*0.18*		*0.37*		*0.84*
Identity - I	0.26	*0.02*	0.29	*0.02*	0.54	*0.20*
Range Break - I		*0.20*		*0.31*		*0.88*

Supporting Information

Figure Legends

Figure 1 Sampling localities of *Phrynosoma platyrhinos* used in this study with locality numbers corresponding to those in Jezkova et al. (2016). The pie charts represent assignment of lizards at each sampling locality to one of the three genetic clusters identified from the SNP dataset: Southern (yellow), Western (light green), Eastern (dark green). The grey outlines correspond to the geographic extents of the four desert regions where *P. platyrhinos* occurs and the dotted outline corresponds to the approximate geographic distribution of *P. platyrhinos*. The background shows mean annual temperature (Hijmans et al., 2005).

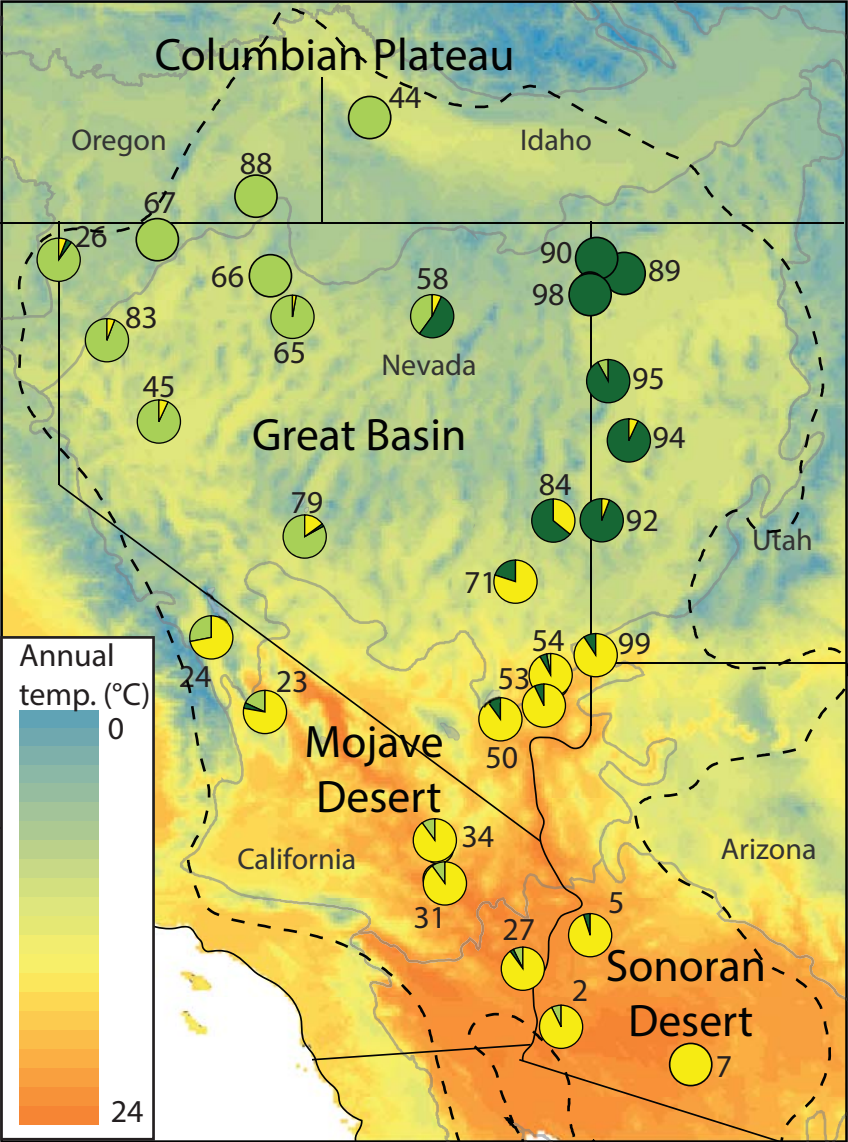
Figure 2 Ancestry matrix built using sNMF (a) and a principal component analysis (PCA) plot (b) for 64 samples of *P. platyrhinos* genotyped for 8,094 SNPs. A maximum likelihood (ML) tree (c) produced with IQ-TREE using a concatenated matrix of 923,866 bp from ipyrad. Numbers along the x-axis of the ancestry matrix correspond to population numbers (see Figure 1), and numbers on the tree correspond to the population number and sample number, separated by a dash. Samples of *P. goodei* are marked with a gray bar. The three genetic clusters identified in the ancestry matrix are used to color-code individual samples in the PCA plot and the ML tree. Individuals that compose the S1 and S2 cluster for the 4 population Moments and BPP analyses are indicated by orange bars and the two samples excluded from these two analyses are indicated by an asterisk. Nodes that are strongly supported on the ML tree based on bootstrap (>95%) and SH-aLRTs (>80%) values are indicated with black circles, whereas those supported by one estimate are indicated with grey circles.

Figure 3. Climatic niche differentiation among genetically differentiated populations of *P. platyrhinos* using (a) principal component analysis (PCA), (b) identity test of Schoener's D and (c) range break test of Schoener's D for each pair of genetic clusters. Dashed lines in (b) and (c) represents observed niche overlap between named clusters, while histograms represent null distributions generated from random assignment of combined individuals and random division of combined geographic area, respectively

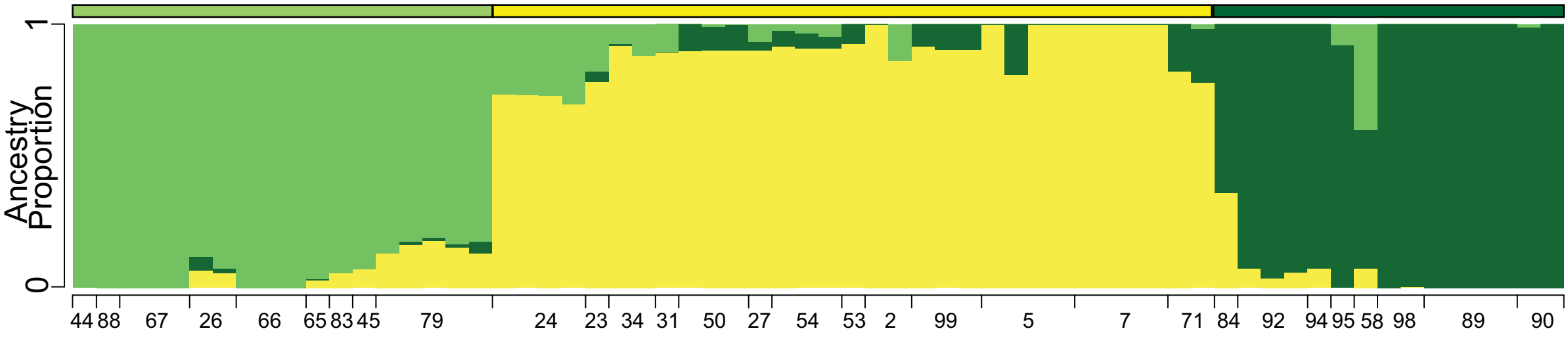
Figure 4 (a) Residual genetic distances among adjacent populations of *P. platyrhinos* derived from an mtDNA dataset (left; 104 localities, 216 individuals, modified from Jezkova et al 2016) and a SNP dataset (right; 30 localities, 64 individuals). The distances have been interpolated across landscape within the range of *P. platyrhino*s. (b) Best performing (lowest AIC) demographic scenarios in the three and four-lineage demographic analyses using Moments. (c) Schematic depiction of changes in effective population sizes estimates (circles) with BPP for the three-cluster and four-cluster models.

Figure 5. Plots from redundancy analysis (RDA) for the first two constrained ordination axes. (a) Outlier loci (color-coded to the climatic predictor) and neutral loci (gray circles) along with the directionality of a

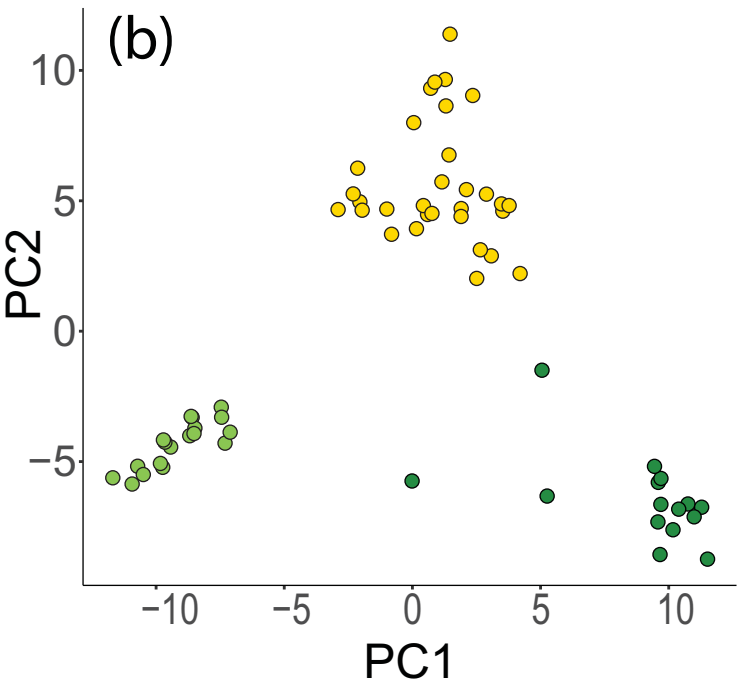
relationship for the four climatic variables (arrows). (b) Relationship between individuals belonging to the three genetic clusters (color-coded circles) and tested environmental variables (arrows). (c) Outlier loci (red; n = 731) identified from the SNP dataset (n=8,094) in an outlier analysis. Loci are plotted according to their respective chromosomes. (d) Chromosomes of *P. platyrhinos* with identified 26 loci that were identified by both *pcadapt* and RDA (red).


Columbian Plateau
Oregon
Idaho
Nevada
Great Basin
Utah
Mojave Desert
California
Arizona
Sonoran Desert
Annual temp. (°C)
0
24
44
88
67
26
66
65
58
83
45
90
89
98
95
94
84
92
79
71
54
99
53
50
24
23
34
31
27
5
2
7

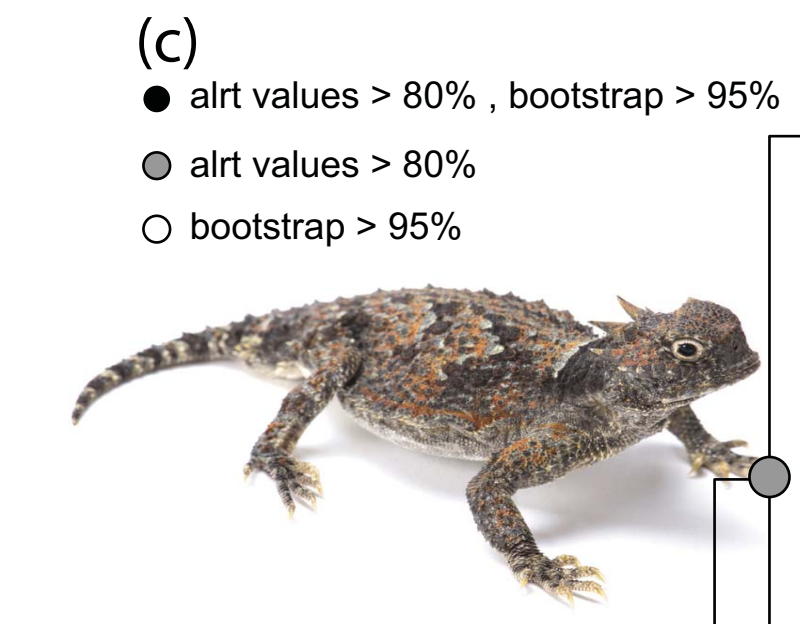
(a)



(b)



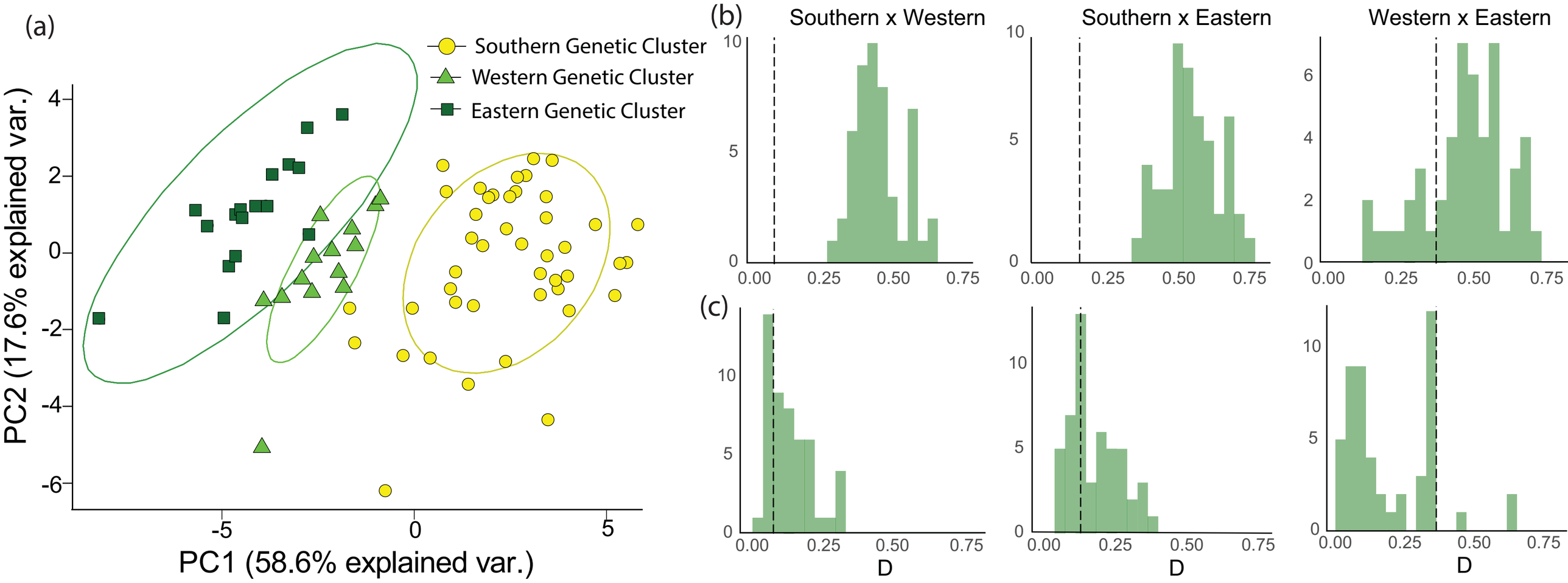
(c)

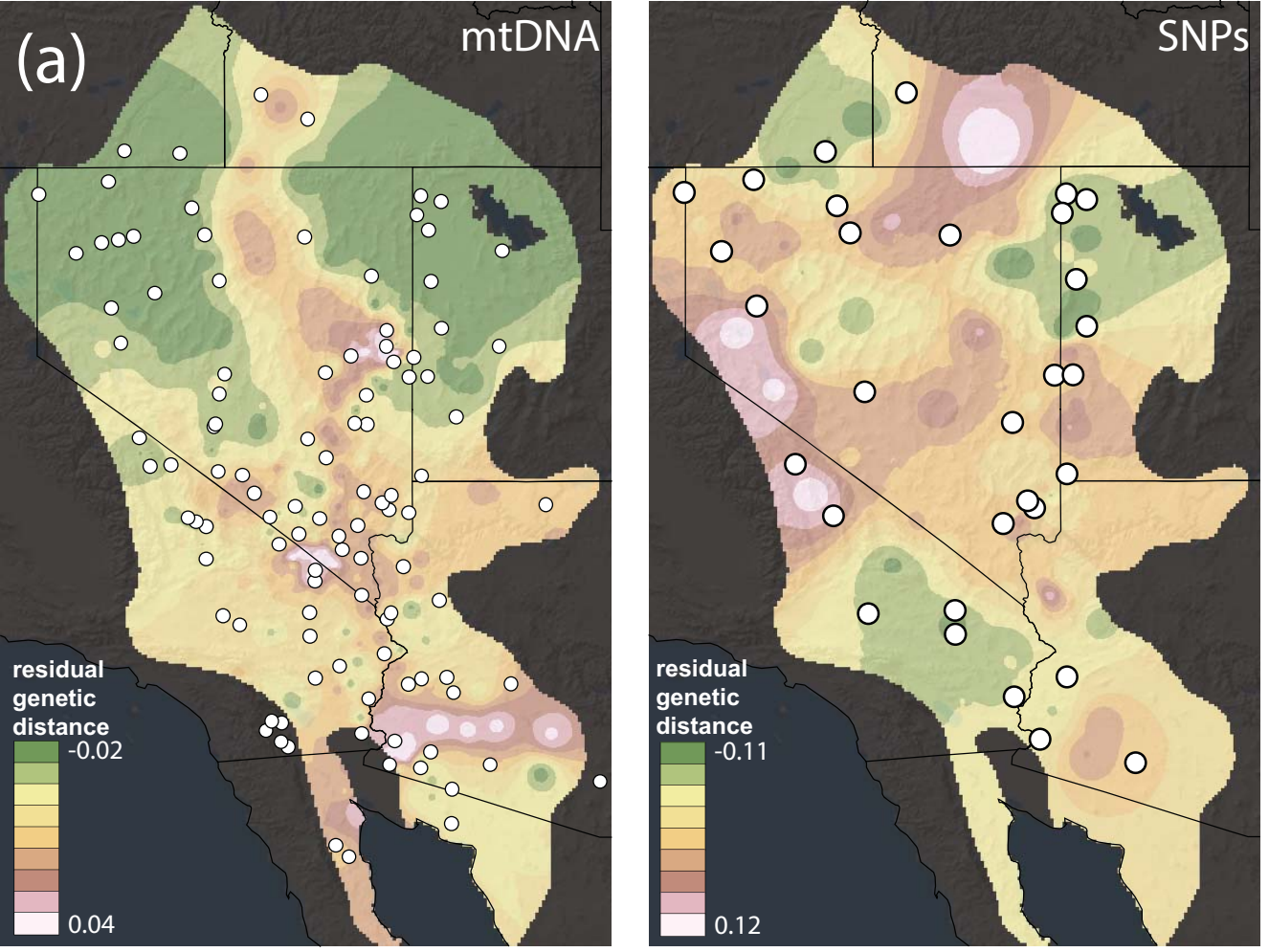


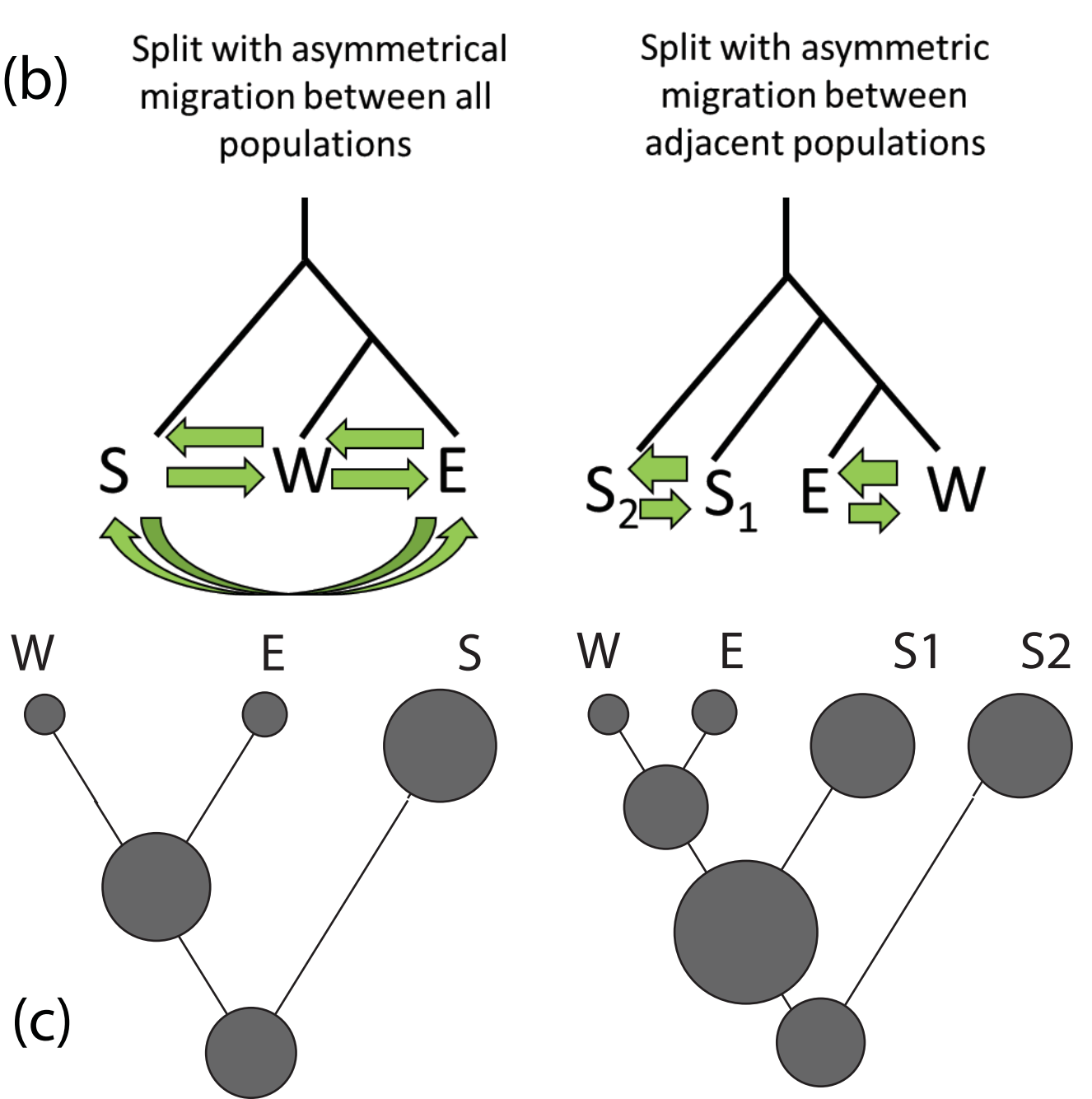
alrt values > 80% , bootstrap > 95%

alrt values > 80%

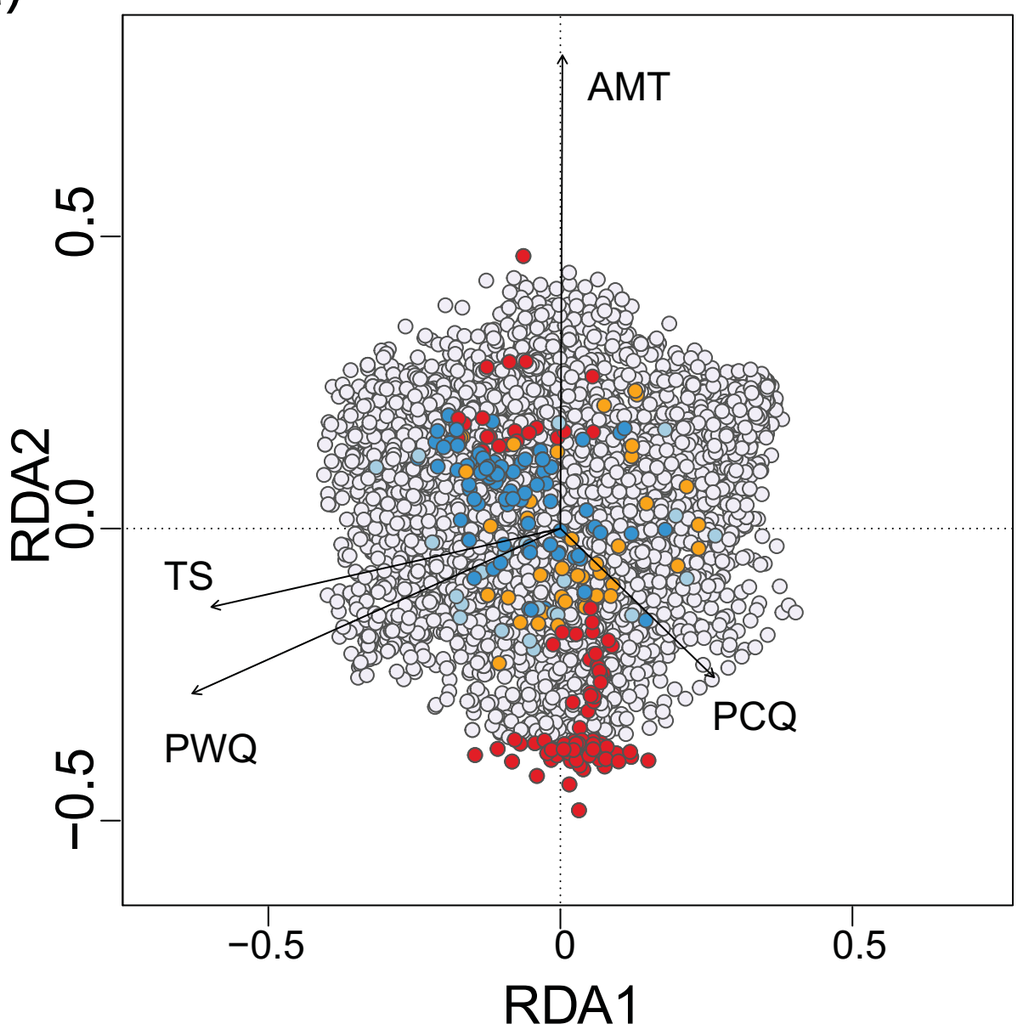
bootstrap > 95%


(a)
Southern Genetic Cluster
Western Genetic Cluster
Eastern Genetic Cluster
PC2 (17.6% explained var.)
PC1 (58.6% explained var.)
(b)
Southern x Western
Southern x Eastern
Western x Eastern
(c)
D
D
D

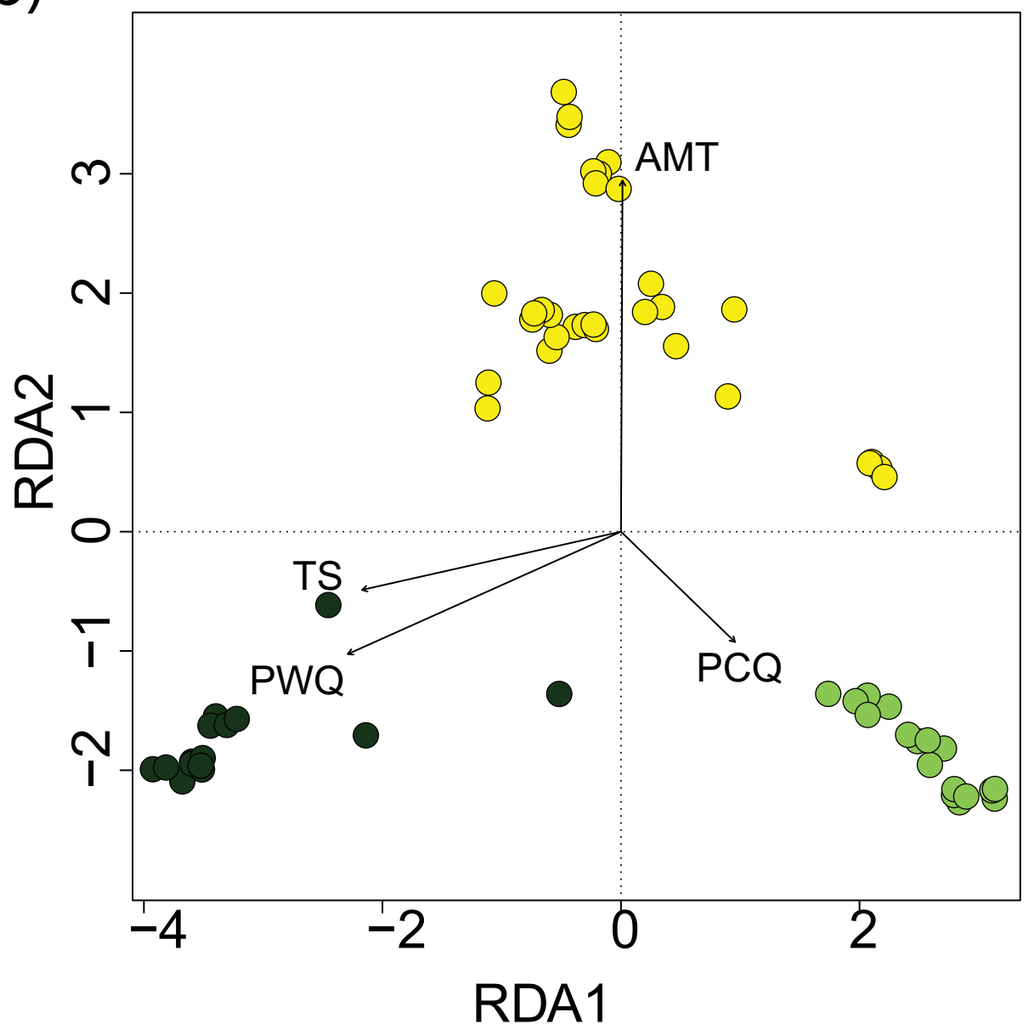

(a)
mtDNA
residual genetic distance
-0.02
0.04
SNPs
residual genetic distance
-0.11
0.12


(b)
Split with asymmetrical migration between all populations
Split with asymmetric migration between adjacent populations
S
W
E
S2
S1
E
W
(c)
W
E
S
W
E
S1
S2

(a)



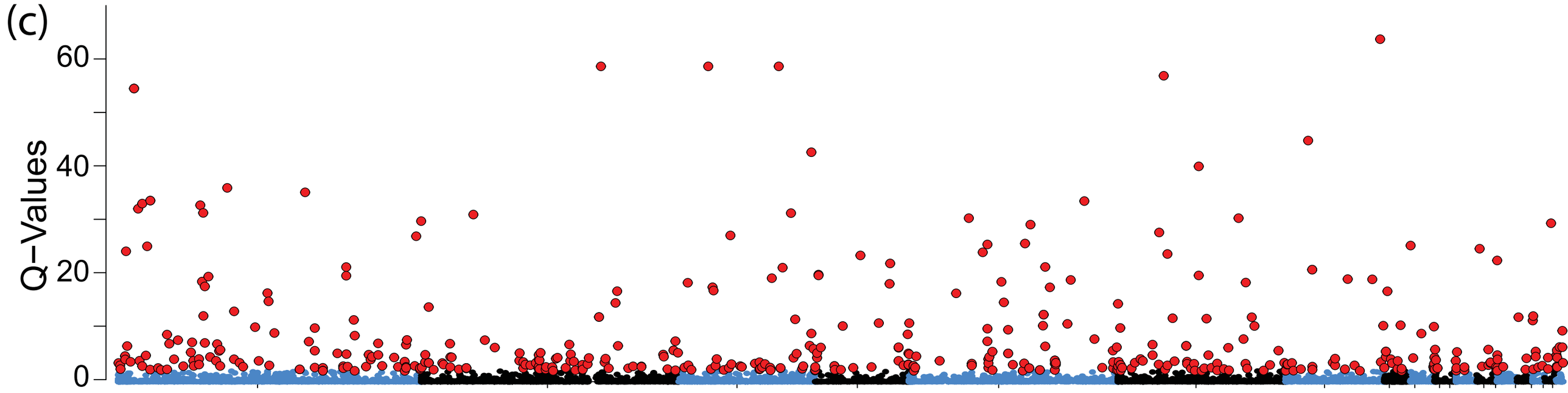
(b)



- Annual Mean Temperature (AMT)
- Temperature seasonality (TS)
- Precipitation of Warmest Quarter (PWQ)
- Precipitation of Coldest Quarter (PCQ)

- Eastern Genetic Cluster
- Southern Genetic Cluster
- Western Genetic Cluster

(c)



(d)