Evolutionary constraints decouple genetic diversity from adaptive capacity under climate change

Tania Hernandez-Hernandez

thernandez@dbg.org

Desert Botanical Gardens

Eddy Mendoza-Galindo

University of California Berkeley

Francisco Molina-Freaner

Univ Nacl Autonoma Mexico

Michiel Pillet

University of Arizona https://orcid.org/0000-0003-3215-8585

Michelle Cabrera-Guerra

Escuela Nacional de Estudios Superiores

Antonio Hernandez-Lopez

Ciencias Agrogenómicas, Escuela Nacional de Estudios Superiores Unidad León, Universidad Nacional Autónoma de México,

Tram Vi

University of Auckland

Piero Foncerrada-Elizondo

Escuela Nacional de Estudios Superiores

Diana Martinez-Lona

Centro de Investigación y de Estudios Avanzados

Article

Keywords:

Posted Date: June 3rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9895616/v1

Additional Declarations: There is **NO** Competing Interest.

Title

Evolutionary constraints decouple genetic diversity from adaptive capacity under climate change

Authors

Eddy Mendoza-Galindo[1,†], Francisco Molina-Freaner[2], Michiel Pillet[3,4,‡], Michelle Cabrera-Guerra[1], Antonio Hernández-López[1], Tram Vi[5], Piero Foncerrada-Elizondo[1], Diana I. Martínez-Lona[6], Tania Hernández-Hernández[4,7,8,*]

Affiliations

[1] *Ciencias Agrogenómicas, Escuela Nacional de Estudios Superiores, Universidad Nacional Autónoma de México, León, Guanajuato, México*
[2] *Departamento de Ecología de la Biodiversidad, Instituto de Ecología, Universidad Nacional Autónoma de México, Hermosillo, Sonora, México*
[3] *Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ, USA*
[4] *International Union for Conservation of Nature, Species Survival Commission, Cactus and Succulent Plants Specialist Group*
[5] *School of Biological Sciences, University of Auckland, New Zealand*
[6] *Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, Irapuato, Guanajuato, México*
[7] *Department of Research, Conservation and Collections; Desert Botanical Garden, Phoenix, AZ, USA*
[8] *School of Life Sciences, Arizona State University, Tempe, AZ, USA*
[†] *current address: Plant & Microbial Biology, University of California, Berkeley, CA, USA*
[‡] *current address: Department of Integrative Biology; University of South Florida, St. Petersburg, FL, USA*
* Corresponding author: thernandez@dbg.org

Abstract

Climate change is expected to drive widespread biodiversity loss, and species persistence may depend on migration, phenotypic plasticity, or genetic adaptation. In long-lived species with constrained life histories, however, limited dispersal and slow demographic turnover restrict these alternatives, increasing reliance on adaptive genetic responses. Under this framework, adaptive capacity is often inferred from levels of standing genetic variation within populations. Genetic diversity is therefore widely assumed to buffer populations against environmental change, yet this assumption remains largely untested.

Here, we integrate population genomics, genome-environment associations, and genomic vulnerability modelling to assess adaptive capacity across the range of a desert foundation species, the saguaro cactus (*Carnegiea gigantea*). We identify two genetically structured lineages occupying distinct climatic niches with contrasting evolutionary trajectories under future climates. Despite higher genetic diversity and stronger genome–environment associations, southern populations exhibit elevated genomic offset and are projected to undergo near-complete loss of suitable habitat. In contrast, northern populations show lower genomic offset but remain constrained by physiological limits and restricted recruitment.

Across populations, environmental conditions explain substantially more variation in performance than genetic factors, indicating a limited contribution of standing genetic variation to persistence. Our results demonstrate that genetic diversity does not reliably predict adaptive capacity under rapid climate change. By showing that populations with greater genetic diversity can nonetheless experience higher genomic mismatch and increased collapse risk, we reveal that diversity-based metrics can misrepresent vulnerability. These findings call for conservation frameworks that explicitly integrate adaptive genomic variation with projected environmental change.

Introduction

Climate change is reshaping species distributions and persistence worldwide, placing increasing emphasis on the capacity of populations to respond to rapidly changing environments (1). To persist under changing climates, plant populations may rely on multiple, non-exclusive pathways, including phenotypic plasticity, the use of microrefugia, range shifts tracking suitable environments, or adaptive genetic change enhancing fitness under novel conditions (2,3). In long-lived species with constrained life histories, however, limited dispersal and slow demographic turnover restrict these alternatives, increasing reliance on adaptive genetic responses. In this context, adaptive capacity has traditionally been inferred from levels of standing genetic variation within populations. Genetic diversity is widely assumed to buffer populations against environmental change, therefore commonly used as a proxy for population resilience and long-term persistence in conservation biology (4,5). However, although the assumption that genetic diversity predicts adaptive potential has been challenged on theoretical and conceptual grounds (6), empirical tests remain scarce, particularly in extreme environments where strong selection may decouple genome-wide diversity from adaptive variation.

Cacti have long been considered emblematic of plant adaptation to extreme arid environments, often viewed as inherently resilient to heat and drought. However, increasing reports of mortality associated with extreme climatic events (7-9) suggest that this perceived resilience may reflect adaptation to historical conditions rather than capacity to withstand rapid climatic shifts. At the same time, cacti are among the world's most threatened plant groups, with roughly one-third of the ca. 1,400-1,800 species assessed by the IUCN (International Union for the Conservation of Nature) classified as threatened with extinction, placing the family among the ten most imperiled groups globally (10). Most cacti are slow-growing, long-lived, and often require decades to reach reproductive maturity (11-13), characteristics that severely constrain migration and the rate of adaptive genetic change. In addition, seedling recruitment, the most vulnerable demographic stage, is typically rare and episodic, occurring only under narrow climatic windows (14), and is highly sensitive to small increases in temperature and decreases in moisture (15-17). As a result, effective population sizes are often substantially smaller than census sizes, and although persistence in microrefugia may buffer short-term declines, it may also reduce genetic diversity and increase long-term extinction risk. Together, these life-history and ecological constraints limit the potential for migration and plastic responses, placing greater reliance on adaptive evolutionary change

Conservation strategies have traditionally prioritized *in situ* protection, while *ex situ* approaches such as seed banks, living collections, cryopreservation, and tissue culture have served primarily as complementary safeguards for restoration and potential reintroduction (18-21). However, rapid climate change introduces substantial uncertainty to *in situ* conservation outcomes, as suitable habitats for many species are projected to shift dramatically or disappear within decades (22-25), increasing the importance of *ex situ* collections as reservoirs of biodiversity and evolutionary potential. In response, conservation frameworks have emphasized capturing high levels of genetic diversity within conservation collections, under the assumption that standing genetic variation enhances adaptive capacity and long term persistence (18, 26-28). Yet recent advances in landscape genomics and genomic offset modelling challenge this assumption. Emerging evidence indicates that populations with substantial genetic variation may still experience high genomic offset, elevated maladaptation risk, and limited evolutionary rescue

potential when future climates diverge strongly from historical selective environments (29-31). These findings suggest that genetic diversity and adaptive resilience represent distinct dimensions of climate vulnerability, and that conservation strategies based solely on maximizing genetic diversity may fail to identify populations most likely to persist. Integrative approaches that explicitly link genomic variation to environmental change are therefore needed to guide conservation under rapid climate shifts.

Among cacti, the saguaro (*Carnegiea gigantea*) stands as both a biological keystone and a cultural symbol of the Sonoran Desert (32,33; Figure 1). Its ecology has been studied for over a century (32-39), yet recent mass mortality events associated with extreme heat and prolonged drought have revealed previously unrecognized physiological limits (7, 40), catalyzed widespread public concern (Table S1), and raised concerns about its long-term persistence. Once regarded as a model of desert resilience, the saguaro now provides a powerful system to examine how climate change can challenge even highly adapted species. Here, we integrate population genomics, genome-environment associations, and genomic vulnerability modelling to assess adaptive capacity across the species' range. We test a central assumption in conservation biology, that genetic diversity predicts adaptive potential, by linking genomic variation with climatic gradients and population performance. Our results reveal a decoupling between genetic diversity and adaptive potential, showing that resilience depends not on the amount of genetic variation alone, but on its alignment with future environmental conditions, with important implications for conservation planning and the management of *ex situ* collections.

Results

Population Structure and Genetic Differentiation

We followed both latitudinal and longitudinal gradients to achieve a comprehensive sampling of saguaros across its range in the Sonoran Desert (Fig. 1), aiming to capture the region's geographical and climatic diversity. The distribution of the saguaro cactus spans a five-fold gradient in mean annual precipitation, from less than 100mm in the northwestern to about 500 mm in the extreme southern edge of its range (41). Although we did not find evidence of discrete and geographically separated populations in the field, the distribution of saguaros is not continuous; large patches devoid of individuals are interspersed with areas of high abundance. We collected 201 wild tissue samples from 30 localities spread across the US and Mexico (Fig. 1, Supplementary Data). For each locality, we estimated a Health-Score, an integrative estimate that describes the phenotypic, reproductive, and demographic evaluation of the wild populations. We genotyped all individuals using a ddRAD-seq approach, generating a total of 9.53×10^8 reads, with an average mapping rate of 97.83 ± 0.94% per individual mapped to the publicly available saguaro genome (42), and each site supported on average by 26.64 ±5.46 reads. After SNP calling and filtering (Figure S1), we identified 46,647 genome-wide SNPs distributed in 41.48% of the genome scaffolds and with a per-sample coverage of 94.14%. We characterized the genetic relationships and population structure among all the saguaros sampled using maximum likelihood (ML) phylogeny, principal component analyses (PCA) as well as ADMIXTURE (Figure 2). All these analyses support two well separated genetic groups, that are coincident with the country of origin of samples, and that we called the North (US) and the South (Mexico) group.

An examination of PCA analyses results shows noticeable differences among the North and South groups (Figure 2a). Both groups have significantly different variances (Levene's test $P < 2.2 \times 10^{-16}$), being that the North group occupies a more restricted area of the multivariate genotypic space suggesting lower genetic diversity (Figure 2a-b). Along the four main principal components, which cumulatively explain 50.99% of the total variance in the data, a multivariate analysis of variance supported significant differences between localities (MANOVA, Wilk's $\lambda = 1.8 \times 10^{-8}$, $P < 2.2 \times 10^{-16}$). Notably, PC1 varied significantly with latitude (linear regression, $P < 2.2 \times 10^{-16}$, R = 0.90; Figure S2) supporting a strong North-South pattern of differentiation. ML Phylogenetic reconstruction (Figure 2c) also supported a North-South segregation. All samples from the South lineage formed a monophyletic clade with 100% bootstrap support, whereas relationships within the northern lineage showed little support, consistent with a rapid expansion into the northern Sonoran Desert (43). The midpoint of the tree also separated North and South groups. An analysis of molecular variance (AMOVA) further supports significant differentiation between these North and South saguaro lineages ($P = 9 \times 10^{-4}$), where 8.9% of the total genetic variation is explained by this group difference. Consistent with recent divergence and ongoing gene flow across the range, genome-wide analysis of genetic divergence with F_{ST} showed low divergence between the groups ($F_{ST} = 0.0106$), and only 0.1125% of all SNPs show high differentiation ($F_{ST} > 0.25$); with a scattered distribution along the genome (Figure S3).

Individual assignments to ancestral population clusters also suggest bipartite population stratification within the saguaro (Figure 2d). At $K = 2$, samples from all localities in the USA (North populations) were overwhelmingly assigned to a single cluster (average = 0.99, range = 0.91-1.00), while although those from Mexico (South populations) had generally intermediate assignments, they tend to represent the second cluster (average = 0.70, range = 0.28-1.00). At K = 5, which showed the lowest average cross-validation error and thus the highest model support (Figure S4), the North-South differentiation was more evident, since samples from North populations were mostly assigned to two clusters, while South populations samples had (with few exceptions) variable assignments to the other three clusters that also include samples in Mexico. In congruence, analyses with other values of K largely reflect substructure within these two groups while maintaining a clear North-South division.

Estimated population parameters reveal contrasting demographic and genetic patterns between the two groups (Table S2), with overall reduced genetic diversity in the North populations. The North group exhibits a lower proportion of heterozygous sites (H North $H_O = 0.1583$, South $H_O = 0.1635$) and reduced nucleotide diversity ($\pi = 0.1862$ vs. 0.1964 in the South). Southern populations harbor 2.15% private alleles that are absent from the North. In contrast, northern populations show a slightly lower inbreeding coefficient (FIS = 0.1503 vs. 0.1685). Both regions display positive Tajima's D values, with a higher value in the North (0.5549 vs. 0.4596). Interestingly, the North group shows a lower mean frequency of deleterious alleles (DL = 0.857 ± 0.593) compared with the South (1.00 ± 0.681), and in addition, we observed stronger linkage disequilibrium in the North group (Figure S5). Together, these patterns may reflect differences in demographic history or the efficacy of purifying selection between regions and suggest that deleterious load and genomic vulnerability do not necessarily track overall genetic diversity.

Demographic reconstructions coincide with the genetic structure and diversity patterns found. Results based on the site frequency spectrum (SFS) reveal that both North and South saguaro lineages experienced population contractions following the Last Glacial Maximum (19-33 kya), with the decline more pronounced in the North (Figure 3). The SFS of both groups is skewed toward low-frequency alleles, consistent with historical bottlenecks and recent expansion signals, particularly in northern populations that show an excess of rare variants and have reached stability similar to a Wright-Fisher population (Fig. 3a). Stairway Plot analyses indicate a continuous decline in effective population size (N_e) after the LGM, and a severe bottleneck after the Penultimate Glacial Period (PGP; 135-109 kya) (Fig. 3b). Present-day N_e estimates are similar (North ≈ 1,225; South ≈ 1,116), suggesting that despite parallel population size reductions, the South experiences stronger genetic drift.

Genetic diversity represented in *ex situ* collections

In addition to wild populations, we sampled tissue from 23 saguaros randomly selected from the National Cactaceae Collection (DBG; Phoenix, AZ, USA) and from 24 urban individuals obtained through collaborations with residents of the Phoenix Metropolitan Valley, who permitted sampling of saguaros growing on their properties. In total, our dataset comprised 248 unique individuals. Although some saguaros in the Phoenix Valley occur naturally, particularly near undeveloped slopes and protected desert areas, a substantial fraction of urban individuals, especially those in landscaped neighborhoods distant from natural habitats, are purchased from nurseries.

Both DBG and Urban saguaro samples clustered within the North lineage (Figure 2a-c), with no representatives of the South group detected either in DBG's Collection or among plants safeguarded by residents of the cities in the Phoenix Valley. The multivariate PCA space shows that while both DBG and Urban samples overlap with the genetic variation of wild North populations (ANOVA $P > 0.05$), Urban saguaros span a broader portion of this space (Figure 2b), a pattern also supported by the ML tree (wider distribution of Urban samples across the Northern lineage, in contrast with a more clustered pattern seen in DBG samples, Figure 2c). Population genetic parameters corroborate these contrasts: Urban saguaros retained higher observed heterozygosity (Ho = 0.1694 vs. 0.1604), more segregating SNPs (83.1% vs. 79.3%), and a greater proportion of private alleles (0.84% vs. 0.27%) than DBG plants, while also exhibiting lower inbreeding (FIS = 0.1402 vs. 0.181) and lower mean frequency of deleterious alleles (DL = 0.875 ±0.579 vs 0.983 ±0.726) (Table S2). Interestingly, several saguaros from the DBG collection and some Urban saguaros represent a unique ancestry in the North group that may represent the ancient wild populations of Papago Park (DBG location) and surrounding area. Notably, the unplanned urban assemblage of *ex situ* saguaros encompasses greater genetic diversity than the curated DBG collection, highlighting the central role of community stewardship in maintaining regional saguaro genetic diversity.

Signatures of adaptive evolution

We examined allele-frequency patterns across the gene space. From the 46,195 SNPs (filtered dataset VCF-2, see Table S3), 21,839 (47.27%) mapped to 5,624 independent annotated genes, primarily within exons (19.5 %) and untranslated regions (2.5 %). We compared Tajima's D per gene between North and South populations (Fig. 3c). While values were broadly correlated (R^2 = 0.42, $p < 2\times10^{-16}$), a subset of genes displayed opposite-sign D values, indicating contrasting

selective pressures between groups. Outlier genes were identified as those with $\Delta\pi > 0.2$, $F_{ST} > 0.25$, and opposite Tajima's *D* signs, yielding 41 candidates (five with higher diversity in the North, 36 in the South; Figure 3c and Table S4). Codon-based analyses revealed that 19 (46 %) harbored at least one site under pervasive positive selection, and 26 (63 %) deviated from neutrality ($d_N/d_S < 0.96$ or > 1.00; Fig. 3d). These findings indicate that a fraction of the saguaro genome bears signatures of selection, consistent with adaptive divergence among regional lineages likely shaped by environmental heterogeneity and climatic stress.

The 41 genes identified as outliers between northern and southern saguaro lineages encompass a diverse set of biological functions, many of which are plausibly related to stress tolerance and cellular maintenance (Table S4). Several candidates are associated with proteostasis and thermal regulation (e.g., *heat-shock protein 90*, ATP-dependent chaperones), redox and metabolic homeostasis (*NAD(P)H-ubiquinone oxidoreductase*), and transcriptional or translational control (*TFIID*, *ribosomal* and *initiation factor* proteins). Others are linked to cytoskeletal remodeling, vesicle transport, and organelle gene expressions (*myosin-kinesin ATPase*, *PPR* and *helicase* proteins), which may contribute to maintaining cellular integrity under environmental stress. Although functional annotation alone cannot establish causality, the recurrence of genes related to protein folding, redox regulation, and membrane or structural maintenance suggests that adaptive divergence in saguaro may involve pathways that enhance physiological stability during episodes of heat and drought. These findings are consistent with polygenic selection acting on multiple cellular systems rather than on a single pathway.

Environmental differentiation and suitability predictions under climate change

Our original georeferenced database comprised 975 saguaros that were visited, photographed, and sampled for tissue or fruit between 2021 and 2024. After removing redundant records (points falling within the same spatial grid), 205 unique field records remained. To expand spatial coverage, we incorporated additional occurrence data from SEINet (Arizona–New Mexico Chapter; https://swbiodiversity.org/seinet/) and iNaturalist (https://www.inaturalist.org), resulting in a final dataset of 30,268 georeferenced saguaro records. Both datasets were manually curated to remove points outside the known species distribution and potential misidentifications. Genetic group identity was assigned using latitude as a geographic proxy. Based on clustering, ADMIXTURE, and ML phylogenetic analyses, we defined the Imuris-Nogales region in northern Sonora, Mexico, as the northernmost limit of the South Group (Figs. 1-2). Records north of Imuris were classified as 'North', and those to the south as 'South'. Although this boundary may introduce minor uncertainty near the contact zone, it provides a robust framework for large-scale geographic analyses.

Principal component analysis (PCA) of all environmental predictors revealed clear environmental niche segregation between the North and South genetic lineages, consistent with their genomic structure (Figure 4a). In the current scenario, the first two principal components, explaining 56% of total variance, were primarily driven by Annual Precipitation (BIO12), Maximum Temperature of the Warmest Month (BIO5), and Temperature and Precipitation Seasonality (BIO15 and BIO4), separating the drier, more temperature-variable North from the comparatively precipitation-variable South. Species distribution models (SDMs) showed high predictive performance (AUC values for the South/North lineages: training = 0.94/0.93, average validation = 0.93/0.91, SD validation = 0.017/0.039), confirming strong climatic signal. For the

South group, the retained predictors included BIO2 (Mean Diurnal Range), BIO3 (Isothermality), BIO8 (Mean Temperature of Wettest Quarter), BIO9 (Mean Temperature of Driest Quarter), HAS_q0.75 (75^{th} percentile of hydrologic year accumulated drought severity), HMD_median (50^{th} percentile of hydrologic year median drought duration), and HMS_mean (Average of hydrologic year mean drought severity), with several interaction terms (i.e. BIO3:BIO8, BIO3:BIO9, BIO3:HAS_q0.75, BIO3:HMD_median). For the North group, the model retained the same core variables plus HAD_q0.75 (75^{th} percentile of hydrologic year accumulated drought duration), and the same interaction terms. The overlap in retained predictors suggests that both groups respond to similar temperature and precipitation dimensions.

Current suitability maps closely match the observed distribution range but reveal contrasting future trajectories (Fig. 3b-c). The North group show relative increases in suitable area with a centroid shift northwestward under future Shared Socioeconomic Pathway (SSP) scenarios SSP1(*Sustainability*) and SSP3 (*Regional Rivalry*) and further northward under SSP5 (*Fossil-Fueled Development*; Figure 3b). The South group, in contrast, experiences sharp contraction of suitable habitat under all SSPs (Fig. 3c). According to our predictions, both under SSP3 and SSP5 the South group will show a complete disappearance of suitable climatic conditions by 2100. To minimize extrapolation bias, all projections were clamped to the calibration range, ensuring conservative estimates under severe drought scenarios. Overall, these results indicate that the two saguaro lineages occupy distinct climatic niches and will likely face markedly different trajectories under future climate change, persistence only through potential spatial shift in the North versus near-total range collapse in the South.

Genetic environment associations and genomic vulnerability

To disentangle the relative influence of genetic and environmental factors on population performance, we quantified the genetic contribution to variation in the Health-Score, an integrative measure of phenotypic, reproductive, and demographic condition that reflects how well populations perform under their local environmental conditions (see full explanation in the Supplementary Methods). Broad-sense heritability ($H^2 = 0.3163 \pm 0.1758$; $P > 0.05$) indicated that genetic factors explain only a limited fraction of variance in Health-Score, with no significant difference between the North and South lineages ($P = 0.1596$). In contrast, Health-Score showed strong associations with multiple environmental predictors (Table S5; FDR-adjusted $P = 3.10\times10^{-12}$–2.07×10^{-7}), identifying the environment as the dominant predictor of saguaro population health.

To test whether genetic lineages occupy distinct environmental conditions, we extracted 25 uncorrelated environmental variables ($r < 0.90$, Table S6) for each individual in our genotyped georeferenced database, and compared North versus South lineage distributions. Environmental differentiation between North and South populations was highly significant across temperature, precipitation, drought, and soil texture variables (Table S6), with the strongest effects observed for temperature annual range (BIO7; FDR-adjusted $P = 2.7\times10^{-41}$, effect size = 2.48) and precipitation seasonality (BIO15; $P = 5.1\times10^{-43}$, effect size = 2.57). Environmental ordination confirmed pronounced niche divergence between groups along PC1 (PC1; $P = 8.6\times10^{-45}$, effect size = 2.66).

Since North and South lineages occupy distinct environmental niches, we next evaluated how environmental gradients shape genomic variation (Figure 5a-c). Redundancy analysis (RDA) showed that 13.98 % of total genomic variance is significantly explained by environmental predictors ($P < 2.2 \times 10^{-16}$; Figure 5a). Marginal permutation RDA tests confirmed that most variables were significantly associated with genetic variation ($P \leq 1 \times 10^{-3}$), each contributing 0.67–1.79 % of total variance (Table S6). When analyzed separately, environmental predictors accounted for 7.25 % of genetic variance in the North group and 13.35 % in the South group ($P < 2.2 \times 10^{-16}$ in both), indicating stronger environmental constraint in southern populations.

Genetic differentiation associated with environmental gradients can result either from neutral demographic processes or from natural selection acting on functional loci. To test for adaptive responses, we performed environmental genome-wide association analyses (GWAS). Observed association patterns deviated from the null expectation based on random simulated phenotypes, supporting biologically meaningful relationships (Figure S6). Seventeen SNPs were significantly associated with environmental variables ($P < 1 \times 10^{-9}$), including drought (1 SNP), temperature (3 SNPs), soil characteristics (3 SNPs), and environmental PCA axes (10 SNPs); none were linked to precipitation. To evaluate the robustness of these associations, we conducted a sensitivity analysis in which 1000 random sets of 17 SNPs were sampled from the full dataset and used to calculate genomic offset under each climate scenario. The observed genomic offset based on the candidate adaptive SNPs consistently exceeded that expected from random SNP sets, indicating that the detected associations capture non-random, environmentally relevant signals (Figure S9).

These 17 SNPs associated with environmental variables were located within or near ten genes spanning diverse functional categories, including metabolic regulation, cytoskeleton organization, membrane transport, protein turnover, transcriptional regulation, RNA modification, and stress response pathways (Table S8a-b). Loci were associated with genes involved in cellular stress and homeostasis, such as enzymes in metabolic pathways, heat-shock–related proteins, vesicle transport components, and ubiquitin-mediated protein turnover. Transcriptional and post-transcriptional regulation was also represented, including squamosa promoter-binding–like transcription factors, pentatricopeptide repeat proteins, and RNA-modifying enzymes (Table S8b). Although some candidate loci mapped to genes of unknown function, the overall functional profile suggests that environmentally associated variation targets multiple biological processes relevant to stress tolerance, cellular maintenance, and adaptive responses; confirming that adaptation to climate and soil conditions is likely polygenic, involving many loci with cumulative effects.

Because the environmental niches of the North and South groups differ under current conditions and show contrasting responses to future scenarios, the observed association between putatively functional polygenic variation and environmental gradients suggests that these groups vary in adaptive potential and are therefore likely to be differentially affected by projected climatic shifts. To quantify this potential disparity, we estimated the genomic offset based on the 17 putatively adaptive SNPs under three alternative future climate scenarios. Genomic offset differed significantly between lineages and among scenarios (Figure 5d; Table S9). The North group exhibited consistently lower offset values, indicating a closer match between current genotypes and future predicted environments. In contrast, the South group displayed markedly

higher and more heterogeneous offset values, suggesting that its populations harbor genotypes less suited to future conditions and are therefore at greater genomic vulnerability to climatic change.

Discussion

The capacity of long-lived plants to persist under rapid climate change depends primarily on their ability to tolerate or adapt to shifting environmental conditions rather than to track suitable habitats through migration (44-47). Here, we show that this adaptive capacity is fundamentally constrained in a desert foundation species. Despite clear genomic structure and signatures of local adaptation, environmental conditions overwhelmingly determine population performance, and adaptive genetic variation is insufficient to buffer populations against projected climatic change. These findings demonstrate that, in long-lived species with limited dispersal and slow demographic turnover, exposure to changing environments rather than escape will dominate responses to climate change. As a result, adaptive capacity, not geographic range shifts, emerges as the central determinant of persistence, challenging the long-standing perception of cacti and other desert plants as inherently resilient to heat and drought.

Genomic structure and environmental divergence in a shallowly differentiated species

Our genome-wide analyses indicate that across saguaro's range, there are two well-differentiated lineages (North vs. South) with shallow divergence (FST ≈ 0.011) but strong structure in multivariate genotype space (PCA/ADMIXTURE) and significant among-locality effects (AMOVA; Fig. 2–5, Table S2). The South lineage harbors more genetic diversity than the North, evidenced by higher observed heterozygosity (H_o = 0.1635 vs 0.1583), higher nucleotide diversity (π = 0.1964 vs 0.1862), and a greater share of private alleles (2.15% absent from the North) (Table S2). Moreover, multiple lines of evidence reveal that saguaros in the South show a stronger genome-environment association, that is also associated with a stronger adaptive signal: selection scans recovered 36 vs 5 candidate genes meeting the gene divergence criteria (Fig. 3c–d; Table S4), RDA attributes nearly twice the genome-wide variance to environmental variables ($R^2adj \approx 0.13$ vs 0.07; Fig. 5a; Table S7), and climatic PCAs as well as SDMs show contrasting patterns for North and South, with no overlap (Fig. 4). Collectively, these patterns and our demographic reconstructions align with previous evidence (43, 48-49), suggesting demographic stability and long-term environmentally driven selection in southern Sonora, Mexico, where saguaro populations appear to have persisted since the Last Glacial Maximum under stable yet heterogeneous climatic regimes.

Shallow genomic divergence is expected in long-lived organisms with extended generation times and ongoing gene flow, as in saguaro, where the persistence of ancestral polymorphisms through incomplete lineage sorting can obscure population differentiation. Despite this, we detect pronounced ecological and genomic structuring between lineages, indicating that adaptive divergence can emerge even under strong constraints imposed by lineage sorting. This pattern suggests that the concurrent action of ancestral allele retention and selection can generate substantial differentiation in both genetic and climatic space and may help explain why columnar cacti have been reported among the fastest-diversifying plant lineages (50), with evolutionary dynamics that combine slow lineage sorting and rapid ecological differentiation. Our results highlight the need for more population-level genomic studies in this group to better

resolve how these processes interact to shape diversification and adaptation in long-lived desert species under strong environmental pressures.

Polygenic adaptation to extreme environments

Current adaptive genomic patterns in saguaro reveal a complex, likely polygenic architecture of adaptation tightly associated with environmental variation. In addition to the potentially adaptive-responding genes found with our gene divergence tests (Table S4), mixed-model GWAS identified 17 SNPs mapping to ten genes associated with temperature, drought and soil variables (Table S8; Fig. 5, Fig. S6). Most of our candidate genes together implicate pathways plausibly mediating heat-drought tolerance, amino acid and organic acid metabolism (e.g. methylmalonate-semialdehyde dehydrogenase), cytoskeletal dynamics (e.g. actin filament-severing protein), proteostasis and stress signaling (e.g. E3 ubiquitin ligase), developmental and acclimatory regulation (e.g. SPL transcription factor), RNA modification/stability (e.g. pseudouridine synthase), redox-related transport processes (e.g. nucleobase–ascorbate transporter), protein translocation for photosynthetic/energetic complexes (e.g. TatB), and organelle gene expression (e.g. PPR protein). In aggregate, although still speculative, these functions suggest that selection in saguaro acts to enhance the acclimatory and stress-response capacity of cellular and organelle systems rather than targeting a single pathway, supporting a polygenic model in which multiple loci of small-to-moderate effect contribute to environmental adaptation, consistent with the observed climate-associated genomic differentiation.

Whole-genome analyses will be essential to fully resolve the functional basis of these adaptations to the extreme environments of the Sonoran Desert. Together, these results demonstrate that even reduced-representation genomic approaches can detect biologically meaningful adaptive signals, consistent with a polygenic architecture underlying climate adaptation.

Genetic diversity does not equal resilience under climate change

A key result of our study is that genetic diversity does not necessarily translate into resilience under future climate change. In spite of the adaptive signal, population performance might be dominated by environment rather than genotype: the composite Health-Score shows limited, non-significant broad-sense heritability ($H^2 \approx 0.32$) but strong associations with temperature, humidity, and drought predictors (Table S5). Forward projections reinforce this asymmetry in adaptive prospects. Lineage-specific SDMs (with clamping to limit extrapolation) predict relative retention and a northwest-northward shift of suitable conditions for the North group, whereas the South group is projected to lose nearly all accessible suitable area under all SSPs by 2071-2100 (Fig. 4b-c). Concordantly, genomic offset estimated from putatively adaptive loci is consistently higher and more heterogeneous in the South (Fig. 5d; Table S9), indicating a larger genotype-environment mismatch under future climates.

We believe that a potential northward range expansion for the North lineage is unlikely to proceed at the pace required to match future climatic shifts, given the species' long generation time and episodic recruitment, the persistence of freezing temperatures in northern areas that exceed saguaro tolerance limits, and the increasing aridity and scarcity of suitable nurse plants along the Mojave Desert area, which are essential for saguaro germination and establishment. Together, our results suggest that while climate-linked selection is detectable, adaptive capacity will be outpaced, most acutely in the South lineage, highlighting a critical decoupling between genetic diversity and future persistence. This finding prioritizes conservation actions that secure

first South-lineage diversity *ex situ* (seed banks, living collections), expand representation in safeguard plantings, identify and protect microrefugia, and evaluate alternatives such as assisted movement or managed gene flow to sustain persistence under accelerating aridification.

Our study shows that even keystone desert plants such as the saguaro have limited adaptive capacity to the accelerating pace of climate change. Despite harboring two genetically distinct lineages with evidence of local adaptation, environmental factors overwhelmingly dictate population performance, and the South lineage, though genetically more diverse, is projected to lose nearly all suitable habitat within decades. These findings refute the perception of cacti as inherently resilient and reveal that long generation times and restricted dispersal constrain evolutionary responses to rapid warming and drought.

Representation gaps in ex situ conservation

Although protected areas and restoration remain the cornerstone of *in situ* conservation (51,52), climate change introduces substantial unpredictability to their effectiveness. Rapid shifts in temperature and precipitation may outpace species' capacity to adapt or migrate, rendering many current protected areas climatically unsuitable within decades (53-57). Emerging strategies like microclimate manipulation and habitat buffering (58), management of microrefugia (59), and dynamic re-planning of protected-area networks (53, 60-61) offer localized mitigation, but are unlikely to secure long-term persistence under accelerating change and realistic constraints on resources for conservation (62, 63). This uncertainty elevates the importance of generating robust *ex situ* strategies (64). Effective *ex situ* conservation must not only safeguard species but also capture the genetic diversity of wild populations, guided by biological and botanical information, genetic data, and institutional collaboration to preserve a species' evolutionary potential (65-70). Beyond genetic representation, our study incorporates a genomic vulnerability assessment to design an *ex situ* sampling framework and prioritize effort for saguaro cactus, ensuring collections preserve the capacity to confront future change.

Our analyses of *ex situ* preserved saguaros revealed notable differences in the genetic diversity represented outside natural populations. Both the DBG National Cactaceae Collection and urban saguaros from the Phoenix Metropolitan Valley clustered within the North lineage, with no representation of the South lineage in either set (Fig. 2a-c). Within the North group, interestingly, urban saguaros captured higher genetic diversity than the curated DBG collection, showing greater observed heterozygosity (H_o = 0.1694 vs. 0.1604), a higher proportion of segregating and private alleles (83.1% vs. 79.3%; 0.84% vs. 0.27%), and lower inbreeding (FIS = 0.1402 vs. 0.181) (Table S2). Although unplanned, urban plantings have inadvertently safeguarded a broader share of northern genetic diversity. This reveals a striking asymmetry: while urban populations unintentionally capture substantial genetic diversity within the North lineage, both formal and informal *ex situ* collections fail to represent the full evolutionary breadth of the species, particularly the South lineage. This unexpected result, together with recent findings from the DBG Urban Saguaro Census showing that most saguaros in the Phoenix Metropolitan Valley grow in private yards (71), highlight that the future of this species depends not only on institutional conservation efforts but also on the community's stewardship.

From genomic vulnerability to conservation action

Our results demonstrate that genetic diversity does not reliably predict adaptive capacity under rapid climate change. Populations with greater genetic diversity can nonetheless experience higher genomic mismatch and greater projected collapse when their adaptive

variation is poorly aligned with future environmental conditions. This decoupling between genetic diversity and resilience reveals a critical limitation of conservation strategies that prioritize diversity-based metrics as proxies for persistence. Instead, our findings show that the evolutionary relevance of genetic variation, its relationship to current and future environments, determines its contribution to survival. By integrating genomic variation with environmental projections, our study provides a framework for identifying populations at greatest risk and for designing conservation strategies that more effectively capture adaptive potential. Our findings directly motivated the establishment of the DBG's Saguaro Initiatives (71), a coordinated suite of programs that translate genomic insights into conservation action. More broadly, these results suggest that for long-lived species facing rapid climate change, evolutionary constraints may limit the capacity of natural populations to respond, even when genetic diversity is high.

Materials and Methods

Study area and sample collection

Because the species' range is roughly split between the United States and Mexico, we collected samples from 15 localities in each country (Fig. 1 and Supplementary Data). Throughout the manuscript, we use the terms *locality* and *population* interchangeably to refer to these collection sites. At each locality, we collected stem tissue from 10 individuals, although in some cases, DNA extraction was not successful. From each plant, we collected approximately 5 square inches of stem tissue, which was immediately frozen in liquid nitrogen. Biodiversity collection remains challenging in several parts of Mexico. Some areas, particularly in the northwestern portion of Sonora, including Puerto Peñasco and its surroundings, were not visited due to safety concerns. We believe the exclusion of these localities does not significantly impact our study, as we achieved strong representation from both the northern (Organ Pipe) and southern (Puerto Libertad) ends of the distribution. Moreover, our results are consistent with previous findings (43).

Health-Score determination

Following DBG guidelines for assessing the health of saguaros in our *ex situ* collection, we developed a protocol to evaluate the general health status of saguaro populations at each collection site (see Supplementary Text). This evaluation was based on five criteria: abundance, plant condition, reproduction, cohort replacement and yield. At each locality, up to four DBG-trained individuals independently scored these criteria using a standardized questionnaire (Supp. Text). To ensure calibration, the group first discussed which populations appeared to be the healthiest and the least healthy based on initial impressions and any discrepancies were resolved through group discussion. Final scores for each locality were averaged to generate an overall health score (Supplementary Data and Fig. 1). To improve the reproduction criterion, we revisited the same populations during the fruiting seasons between 2022 and 2024. At each site, we collected up to ten ripe fruits (if available, since several plants would not produce ten ripe fruits at a time) from randomly selected individuals. The fruits were processed to extract seeds, which were then weighed to estimate the number of seeds per individual. From this, we calculated the average estimated seed count per locality, which served as an indicator of reproductive yield (Supplementary Data).

DNA extraction and genomic sequencing

Genomic DNA was extracted from stem tissue using a modified CTAB protocol. Briefly, frozen tissue was thawed on ice before grinding it with 2X CTAB pH 8 (Tris HCl 1.0M, NaCl 5M, EDTA 0.25M, and 0.4 ml of beta-mercaptoethanol) and 4 µl of RNAse. Samples were kept overnight in a water bath at 65 °C (see Supplementary Text for details). DNA quality and concentration were assessed using Qubit fluorometry and NanoDrop spectrophotometry. Reduced-representation genomic libraries were prepared following a double-digest Restriction-site Associated DNA sequencing (ddRAD-seq) protocol (FLORAGENEX, LLC). Briefly, genomic DNA was digested with combinations of restriction enzymes *PstI*/*MspI*, yielding an estimated 50,000–150,000 fragments per sample. Adapter ligation, size selection, PCR amplification, and purification steps were performed following standard protocols. Library quality was evaluated using agarose gel electrophoresis, Agilent Fragment Analyzer/Bioanalyzer, and quantitative PCR to ensure integrity and appropriate concentration. Sequencing was performed on an Illumina NovaSeq 6000 platform (1 × 100 bp, single-end), allocating one full lane per 95-sample batch. Across runs, we obtained ~325–400 million reads passing filters, corresponding to ~280–375 million indexed reads, with >80% of bases achieving Q30 or higher. Ten samples were sequenced in duplicate as positive control for genetic relatedness.

QC and SNP filter calling

We assessed the quality of raw sequencing reads based on per-base sequence quality, GC content, sequence duplication levels, and adapter contamination, using *FastQC* v0.11.9 (72). For sample demultiplexing, we used *STACKS* v2.53 (73-74) with options: *--inline-null* (inline barcode), *-c* (clean), *-q* (quality), *-r* (rescue). Restriction enzyme options were set as: *-mseI*, *-pstI*. 285 sampled individual processed reads were then mapped to the saguaro SGP5 v1.3 assembly (75) (GenBank accession GCA_002740515.1) using the *mem* algorithm of *bwa* v0.7.18 (76) in single-end mode (default parameters were used). Alignment statistics were calculated with *Samtools* v1.21 (77).

For single-nucleotide polymorphism (SNP) calling, we used *STACKS* (73-74) with default parameters. After this initial SNP calling that resulted in 443,263 SNPs, we filtered SNPs with a read depth between 4 and 250 using *Bcftools* v1.14 (77). We then calculated per-sample and per-site missingness rate (Fig. S1), and imposed a sample missingness rate threshold of 0.6 (samples with more than 60% of missing data were removed) and a site missingness rate coverage threshold of 0.3 (sites with more than 30% of missing data were discarded), to filter low-coverage samples and sites. Of the initial 285 sequenced individuals, 24 were removed due to excessive missing data, leaving a dataset of 261 saguaros and 46,647 variant sites. We generated three different VCF files (VCF-1, VCF-2 and VCF-3) with *ad hoc* filtering and sampling designed to meet the data requirements for further different analysis (Table S3). All VCF filtering was done using *PLINK* v1.90b7 (78).

Population structure and ancestry

We followed three independent approaches to determine population structure: model-free principal component analyses (PCA), model-based ancestry inference with ADMIXTURE analyses and Maximum Likelihood (ML) inference estimation of phylogenetic trees. All the listed analyses were done using the VCF-1 dataset (19,280 SNPs) that included a filter for linkage disequilibrium (Table S3). The PCA was done using the *--pca* tool from *PLINK* with 20

components (78). Differences in the PC loadings were estimated using the R v4.2.2 *stats* functions (79). Ancestry coefficients were estimated using *ADMIXTURE* v1.3.0 (80). We performed 100 replicates with a random starting seed each and 200 bootstraps (-s and -B arguments), and with *K* ranging from 2 to 12. Cross validation error values were collected for each *K* value and repetition. As each random replicate is expected to produce slightly different ancestry hypotheses, we employed the CLUMPAK online server (https://clumpak.evolseq.net) with default parameters to generate consensus ancestry coefficients for each *K* (81). For phylogenetic analyses, we used the *vcf2phylip* v2.0 script (82) to convert the VCF file into PHYLIP format, using IUPAC ambiguous nucleotide codes for heterozygous calls (82). 5316 invariant sites were removed using the Felsenstein and Stamatakis ascertainment bias correction (83) with a script that allowed processing IUPAC ambiguous nucleotides (84). Maximum likelihood phylogenetic inference was done using IQTREE v2.3.6 (85). First, the most fit substitution model was determined using the Model Finder Plus (MFP) tool in IQTREE (85-86). The best overall model was TVM+F+ASC+R6, which was used to infer the ML tree with 1000 bootstraps (-B 1000) and 1000 SH-aLRT test iterations for support values (-alrt 1000) under a safe kernel (-safe).

Demographic inference

We used Site Frequency Spectrum (SFS) to infer the demographic history of the genetic groups. First, we estimated the folded SFS for 50 diploid individuals (100 gene copies) using *snpR* v.2.9.2 package (87) and dataset VCF-4; which includes 19,000 SNPs (Table S3). The folded SFS was used to create the input "blueprint" files for *Stairwayplot* v2.1.1 (88). We ran the *Stairwayplot* algorithm using *java* v1.8.0_44 (89) and the following parameters: 67% of the sites for training step, break points at {6, 12, 24, 48}, 1,000 bootstraps, a mutation rate (μ) of 8.75 x 10^{-8} substitutions per site per generation, and a generation time of 35 years (following 76). As a control, we simulated 1,000 standard coalescent SFSs using *msprime* v1.3.4 (90) using the previous parameters and $Ne = 5{,}000$, and ran *Stairwayplot* as mentioned. Given the number of segregating SNPs in VCF-4, we approximated *L* to 2.1 x 10^6 by using the formula $\theta = 4Ne\mu L$ (91). To remove biases in recent demographic history, we set the present time at 10 generations ago (~350 years ago). To further corroborate the demographic model, we applied our same demographic modeling method to a whole-genome saguaro dataset that includes 20 genomes sequenced at low-coverage (data from 45). After downloading these genomes, we assessed read quality, trimmed low-quality bases and removed adapters using *trimgalore!* v0.6.10 (92). Paired-end reads were aligned to the same reference using *bwa* v0.7.17 (93). Bad-quality alignments were filtered and de-duplicated using *Samtools* v1.16 (77) (*samtools view -Sh -f 0x2 -q 30; samtools fixmate*). Then, we performed a custom SNP-calling pipeline using *Bcftools* v1.10.2 (77). We isolated biallelic SNPs and applied identical MAF and missingness filters as previously mentioned with *vcftools* v0.1.14 (94). The folded SFS was estimated for the remaining 17 samples and 1,138,328 SNPs with *snpR* using 34 gene copies. The demographic inference was performed using *Stairwayplot* as mentioned before.

Genetic diversity and differentiation

To estimate levels of genetic polymorphism, we employed the VCF-2 dataset, which includes 46,195 SNPs (Table S3). This dataset was not filtered by linkage disequilibrium, thereby leveraging all possible genetic variation patterns under the same frequencies and missingness rates. We used the *snpR* (87) to estimate weighted genome-wide summary statistics for: observed

and expected homozygosity (H_O and H_E), nucleotide diversity (π), inbreeding coefficient (F_{IS}), Tajima's *D*, and the percentage of segregating and private sites. We estimated the summary statistics for all samples combined, for each genetic group (South and North), and within the North group, for *ex situ* collections (DBG) and Urban plants separately. Linkage disequilibrium decay was estimated for North and South complete groups only. Linkage blocks with more than four SNPs in a 10kbp span were isolated with the --blocks utility from *PLINK* linkage blocks were used to estimate the linear relationship between genomic distance (bp) and linkage disequilibrium (R^2), the last was estimated with the --r2 function from *PLINK*.

To examine patterns of polymorphism potentially underlying differences in phenotypic diversity, we focused on annotated gene regions. We quantified the presence of deleterious alleles using a deleterious load statistic (*DL*), calculated as the ratio of deleterious nonsynonymous homozygous sites to all segregating sites (95-96; see Eq. 1, Supplementary Text). Deleterious SNPs were annotated with *SnpEff* v5.2-1 (97) using the available *Carnegiea gigantea* genome (42). Variants annotated as "start_lost," "stop_gained," or "stop_lost," expected to disrupt protein function, were classified as deleterious. *DL* was estimated for the same population groupings described above (All samples, North, South, Urban and DBG). Nucleotide diversity (π) and divergence statistics were estimated using the missingness-aware pixy v1.2.10.beta2 (98) method following the gene coordinates in the reference genome (44) and dataset VCF-2 (Table S3). For each gene, we calculated π and F_{ST} between the North and South groups, excluding genes with only a single SNP, which resulted in 5,624 genes. Differences in nucleotide diversity were expressed as $\Delta\pi = \pi_{South} - \pi_{North}$. Watterson's estimator (θ_W) and Tajima's D were calculated manually following Eqs. 2 and 3 in Supplementary Methods.

Identifying candidate genes under selection

To assess whether divergence was driven by neutral processes or by natural selection, we conducted a phylogeny-based test of selection on loci exhibiting strong departures from neutrality. Candidate genes were identified using three criteria (summarized in Fig. S8): (i) a one-tailed *t*-test to detect outliers in the distribution of $\Delta\pi$, retaining genes with FDR (False Discovery Rate)-corrected $P < 1 \times 10^{-3}$; (ii) contrasting signs of Tajima's *D* between South and North lineages (e.g., negative in the South and positive in the North, or vice versa); and (iii) high population differentiation ($F_{ST} > 0.25$; 99). A total of 41 genes met these thresholds (Table S7). For each candidate gene, phased haplotypes were reconstructed using *SHAPEIT* v2.r904 (window size 0.1 Mb) (100), and consensus DNA sequences were extracted with *Bcftools* v1.10.2 (77). Coding sequences were extracted with *gffread* v0.12.7 (101), concatenated, and aligned at the codon level with *PhyKit* v2.0.2 (102) while masking stop codons. We then estimated model-based d_N/d_S ratios and identified codon sites under pervasive selection using the *FEL* method (103) implemented in *HyPhy* v2.5.64 (104). Phylogenetic trees were prepared by branching haplotypes from the same individual as terminal sister tips without additional branch lengths.

Occurrence data and environmental variables

We compiled georeferenced locality records for saguaros from three sources: (1) our own database derived from recent field collections, (2) the SEINet portal (Arizona–New Mexico Chapter; https://swbiodiversity.org/seinet/), and (3) iNaturalist (https://www.inaturalist.org). Records from SEINet and iNaturalist were manually reviewed, and outliers falling outside the

known distribution were excluded. Because saguaros are morphologically distinctive and easily recognized in the field (with no sympatric species of similar appearance except *Pachycereus pringlei* ('cardón') in a restricted portion of the southern range in Mexico) the risk of misidentification is minimal. Southern records, where both species may co-occur, were carefully verified. Although occasional misidentifications cannot be ruled out, their frequency is likely negligible, and we consider our dataset robust and reliable.

Environmental predictors were extracted for all occurrence records, encompassing 38 variables (before removal of correlated variables) standardized to a ~1 km resolution. These included seven soil, nineteen bioclimatic, and twelve drought variables derived from SoilGrids v2.0 (isric.org/explore/soilgrids), CHELSA v2.1 (https://chelsa-climate.org), and the Global Future Drought Layers v1 (105), respectively. Soil variables (pH, cation exchange capacity (CEC), clay, sand, and silt fractions, coarse fragments volumetric fraction (CFVO), and soil organic carbon (SOC)) were derived from SoilGrids at 1 km resolution using the equal-area Goode homolosine projection, which was also applied to all other datasets. Assuming a maximum rooting depth of 60 cm, weighted averages were calculated across depth intervals up to this depth. Soil variables were assumed to remain static for future projections. Climatic data were obtained from the CHELSA repository at 30 arc-second resolution (≈1 km). The dataset included 19 bioclimatic variables for temperature and precipitation for two time periods: 1981–2010 (present) and 2071–2100 (future). Empirical evidence shows that acute heat and drought events have already caused widespread physiological damage and mortality in saguaros and other Sonoran Desert plants (15, 42), while macroecological modeling across nearly a thousand cactus species demonstrates that explicitly incorporating severe drought metrics greatly increases projected extinction risks (5). Together, these findings highlight that forecasts of saguaro persistence must account not only for gradual climate trends but also for the growing impact of severe drought events. In addition to soil and climate data, to include drought events we used four SPEI-based indices of severity and duration, summarized across the same time periods (106). Future projections were based on the MPI-ESM1-2-HR global circulation model (CMIP6) under three Shared Socioeconomic Pathways (SSPs): SSP1-2.6 (sustainability, "Green Road"), SSP3-7.0 (regional rivalry, "Rocky Road"), and SSP5-8.5 (fossil-fueled development, "Highway"). SSP1-2.6 represents a sustainable trajectory, whereas SSP3-7.0 and SSP5-8.5 reflect increasingly emission-intensive scenarios.

Severe drought variables were obtained from the Global Future Drought Layers (105) at 0.25 arc degrees resolution (~25 km), and resampled to 1 km resolution. Monthly time series of the Standardized Precipitation Evapotranspiration Index (SPEI; 106) were used to derive drought characteristics for the hydrologic year (1 October-30 September). SPEI quantifies the balance between precipitation and potential evapotranspiration (PET), thereby integrating temperature effects on drought intensity. A drought event was defined as a sequence of negative SPEI values reaching ≤ -1 at least once, starting on the first negative value and ending when SPEI returned to zero. Drought duration corresponded to the number of months in the event, and severity to the absolute sum of all SPEI values. Annual layers were generated for hydrologic-year accumulated drought severity (HAS), mean drought severity (HMS), accumulated drought duration (HAD), and mean drought duration (HMD). Each set of layers was summarized by its mean, median, and 75th percentile for 1981-2010 and 2071-2100, matching the temporal scale of the climatic data.

Species distribution modeling and climatic suitability projections

To evaluate whether North and South genetic groups are expected to respond differently to climate change, we built separate distribution models for each group (Fig. 4). To minimize sampling bias, occurrence records were spatially thinned using a 10 km radius. The modeling domain for background point selection (10,000 points) and prediction were defined as the minimum convex hull around all occurrences expanded by a 500 km buffer. Within this domain, highly correlated environmental variables ($r > 0.7$) were iteratively removed to reduce overfitting, resulting in a total of 21 variables. For spatial cross-validation, occurrences were first stratified using k-means clustering into 25 clusters, which were then aggregated into five folds. Fold size balance was optimized by maximizing Shannon entropy.

Model fitting was performed with the *ENMeval* v2.0.4 (106) using a cross-validated implementation of the *MaxEnt* algorithm (108-109) based on the package *glmnet* v4.1.8 (110-111). Models were fitted using penalized maximum likelihood. We applied lasso regularization to shrink model coefficients and prevent overfitting, selecting the optimal value of the regularization parameter (λ) through cross-validation. Specifically, we first identified the λ with the lowest cross-validation error, then chose the largest λ within one standard error of this minimum to favor model simplicity. Linear, quadratic, and product feature classes were included, and final models were trained with all data folds using the selected λ value. Model performance was evaluated using cross-validated AUC, a threshold-independent measure of binary classification accuracy.

We created predicted suitability maps for the present and future. Given the extreme changes in the magnitude of the drought variables and consequent risk of extrapolation, we "clamped" predictions. Clamping reduces extrapolation by constraining environmental values for projection to the limit of values observed during model fitting. As before, a 500-km buffer around the occurrences was used. All analyses were performed with *R* v4.3.1 (79). All spatial analyses were performed using the *terra* v1.7-78 (112).

Genetic-environment association

To test whether genetic groups occupy distinct environmental conditions, we analyzed group differences in environmental variables. Variables deviating from normality (Shapiro test, < 0.01) were transformed using the Ordered Quantile normalization in *bestNormalize* v1.9.1 (113). Linear models were fitted with genetic structure (North or South) as a predictor, and significance was assessed by ANOVA using *car* v3.1-3 (114). Effect sizes were estimated with *emmeans* v1.10.3 (115). The Health-Score provides a diagnostic measure of the current condition of saguaro populations at each locality, and was used to evaluate the relative contributions of genetics and environment to population health. Broad-sense heritability (H^2) of the Health Score was estimated from a kinship matrix derived from the VCF-3 dataset (Table S2) using the *make-rel* function in *PLINK* (78). Random-effects linear models were fitted with genetic relatedness as a random factor in *lme4qtl* v0.2.2 (116), and H^2 was calculated as the proportion of variance explained by kinship using the *VarProp* function. The significance of genetic effects was assessed through likelihood ratio tests comparing models with and without the kinship term (*lmrtest* v0.9-40; 119). Since Health-Score values were averaged per locality, the H^2 distribution was estimated by resampling two individuals per locality 1,000 times to reduce inflation from relatedness. Environmental associations with Health-Score were evaluated using mixed-effects

linear models (MELMs; Eq. 4, Supplementary Methods) and tested by ANOVA. Finally, redundancy analysis (RDA) was used to quantify the contribution of environmental variation to genome-wide genetic structure. Fourteen non-correlated variables ($r < 0.70$, Table S7) from the MPI-ESM1-2-HR climate model were retained using *recipes* v1.1.0 (115). RDA was conducted with *vegan* v2.4-10 (116). The proportion of genetic variance explained by the environment was estimated from adjusted R^2, with significance assessed by permutation. Marginal effects of individual variables were tested by ANOVA permutations, and analyses were repeated separately for North and South groups.

Genome-wide association analyses (GWAS) and genomic vulnerability

We linked fine-scale genomic variation with environmental and phenotypic variables through genome-wide association analyses (GWAS) using mixed-effects linear models following (117) and (118) (see details and Eq. 5 in Supplementary Methods). Because ddRADseq datasets often contain missing genotypes and limited linkage information, we analyzed associations directly from observed genotypes (allowing up to 30% missing data per marker) rather than performing imputation. Mixed-effects linear models (MELMs) were fitted with *lme4qtl* (113) for each of the 44,589 variable SNPs in dataset VCF-3 (Table S2), testing both phenotypic variables (Health-Score) and environmental predictors. The latter included the 14 uncorrelated environmental variables previously described, the first two components of environmental PCA (with and without drought variables), and 20 randomly generated variables used as negative controls. In total, 1,738,971 MELMs were evaluated across 39 variables. GWAS results were assessed for biological relevance using Q–Q plots and comparisons of significance distributions against the random controls, with multiple testing corrected by FDR adjustment.

SNPs associated with environmental adaptation were mapped to nearby genes (±2,000 bp) using *Bedtools* v2.30 (119), and functional annotations were obtained with *EggNOG mapper* v2 (120). We modeled the genomic vulnerability of saguaro populations to future climate change using a machine-learning approach with *gradientForest* v0.1-37 (121). Multivariate, non-parametric genotype–environment models were fitted for each Shared Socioeconomic Pathway (SSP) scenario, incorporating the non-redundant climate variables and genotypes from candidate GWAS SNPs (parameters: *ntree = 1000*, *maxLevel = 1.6452*, *corr.threshold = 0.70*). Allele turnover under projected climates was predicted, and genomic offset was calculated as the Euclidean distance between present and future scenarios. To test whether genetic groups differed in vulnerability across SSPs, we fitted regression models with genomic offset as the response variable. Error distributions were selected using *fitdistrplus* v1.2-1 (122), comparing model performance with AIC values and validating residuals with *DHARMa* v0.4.7 (123) and *performance* v0.10.4 (124). The final model, implemented in *lme4qtl* (113) following Eq. 6 (Supplementary Methods), included genetic group and SSP as fixed effects, the kinship matrix as a random effect, and residual error. Fixed effects were evaluated using Type II Wald chi-square tests (*car*; 114), and estimated marginal means, standard errors, and pairwise contrasts between groups within each GCM × SSP combination were extracted using *emmeans* v1.10.3 (112).

References

1. Walther, G.R. et al. Ecological responses to recent climate change. Nature 416, 389–395 (2002).
2. Aitken, S.N. et al. Adaptation, migration or extirpation: climate change outcomes for tree populations. Evol Appl. 1, 95–111 (2008).
3. Parmesan, C. Ecological and evolutionary responses to recent climate change. Annu. Rev. Ecol. Evol. Syst. 37, 637–669 (2006).
4. Frankham, R., Ballou, J.D. & Briscoe, D.A. Introduction to Conservation Genetics. (Cambridge University Press, 2010).
5. Reed, D.H. & Frankham, R. Correlation between fitness and genetic diversity. Conserv. Biol. 17, 230–237 (2003).
6. Teixeira, J.C. & Huber, C.D. The inflated significance of neutral genetic diversity in conservation genetics. Proc. Natl. Acad. Sci. 118, e2015096118 (2021).
7. Wilder, B.T. et al. Plant responses to anomalous heat and drought events in the Sonoran Desert. Glob. Change Biol. 31, e70217 (2025).
8. Gazol, A. et al. Droughts preceding tree mortality events have increased in duration and severity across global drylands. Nat. Commun. 16, 2897 (2025).
9. Bombi, P. et al. Climate change effects on desert ecosystems: A case study on the keystone species of the Namib Desert Welwitschia mirabilis. PLOS ONE 16, e0259767 (2021).
10. Goettsch, B. et al. High proportion of cactus species threatened with extinction. Nat. Plants 1, 15142 (2015).
11. Gibson, A.C. & Nobel, P.S. The Cactus Primer (Harvard University Press, 1986).
12. Nobel, P.S. Environmental Biology of Agaves and Cacti. (Cambridge University Press, 2003).
13. Anderson, E.F. The Cactus Family (Timber, 2001).
14. Godinez-Alvarez, H., Valverde, T. & Ortega-Baes, P. Demographic trends in the Cactaceae. Bot. Rev. 69, 173–201 (2003).
15. Aragón-Gastélum, J.L. et al. Seedling survival of three endemic and threatened Mexican cacti under induced climate change. Plant Species Biol. 32, 92–99 (2017).
16. Aragón-Gastélum, J.L et al. Potential impact of global warming on seed bank, dormancy and germination of three succulent species from the Chihuahuan Desert. Seed Sci. Res. 28, 312–318 (2018).
17. Barrios, D. et al. Combined effect of temperature and water stress on seed germination of four Leptocereus spp. (Cactaceae) from Cuban dry forests. Plant Species Biol. 36, 512–522 (2021).
18. Center for Plant Conservation. CPC Best Plant Conservation Practices to Support Species Survival in the Wild (Center for Plant Conservation, 2019).
19. IUCN/SSC. Guidelines on the Use of Ex Situ Management for Species Conservation, Version 2.0 (IUCN Species Survival Commission, 2014).
20. Heywood, V.H. Plant conservation in the Anthropocene – challenges and future prospects. Plant Divers. 39, 314–330 (2017).
21. Maunder, M. et al. Ex situ methods: a vital but underused set of conservation resources, in Guerrant, E.O., Havens, K. & Maunder, M. Ex situ Plant Conservation: Supporting Species Survival in the Wild (Island, 2004).
22. Bellard, C. et al. Impacts of climate change on the future of biodiversity. Ecol. Lett. 15, 365–377 (2012).
23. Urban, M.C. Climate change extinctions. Science 386, 1123–1128 (2024).
24. Wiens, J.J & Zelinka, J. How many species will Earth lose to climate change? Glob. Change Biol. 30, e17125 (2024).

25. Rubenstein, M.A. et al. Climate change and the global redistribution of biodiversity: substantial variation in empirical support for expected range shifts. Environ. Evid. 12, 1–21 (2023).
26. Hoban S. New guidance for ex situ gene conservation: Sampling realistic population systems and accounting for collection attrition. Biol. Conserv. 235, 199-208 (2019).
27. Hoban, S. et al., Genetic diversity goals and targets have improved, but remain insufficient for clear implementation of the post-2020 global biodiversity framework. Conserv Genet. 24, 181–191 (2023).
28. Griffith, M.P. et al. Ex Situ Conservation of Large and Small Plant Populations Illustrates Limitations of Common Conservation Metrics. Int. J. Plant Sci. 182, 263–276 (2021).
29. Mahony, C.R.et al. Evaluating genomic data for management of local adaptation in a changing climate: A lodgepole pine ase study. Evol. Appl. 13, 116–131 (2020).
30. Shryock, D.F. et al. Harnessing landscape genomics to identify future climate resilient genotypes in a desert annual. Mol. Ecol. 30, 698–717. (2021).
31. van Daele, F.et al. Genomic analyses point to a low evolutionary potential of prospective source populations for assisted migration in a forest herb. Evol. Appl. 15, 2080–2095 (2022).
32. Yetman, D. et al. The Saguaro Cactus: A Natural History (University of Arizona Press, 2020).
33. Bird Jr, W.L. In the Arms of Saguaros: Iconography of the Giant Cactus (University of Arizona Press, 2023).
34. Shreve, F. The rate of growth of the giant cactus. Plant World 13, 235–240 (1910).
35. Steenbergh, W.F & Lowe C.H. Ecology of the Saguaro II: Reproduction, Germination, Establishment, Growth, and Survival of the Young Plant, Scientific Monograph Series No. 8 (U.S. National Park Service, 1977).
36. Drezner, T.D. The keystone saguaro (Carnegiea gigantea, Cactaceae): a review of its ecology, associations, reproduction, limits, and demographics. Plant Ecol. 215, 581–595 (2014).
37. Pierson,E.A & Turner, R.M. An 85-year study of saguaro (Carnegiea gigantea) demography. Ecology 79, 2676–2693 (1998).
38. Pierson, E.A., Turner, R.M. & Betancourt, J.L. Regional demographic trends from long-term studies of saguaro (Carnegiea gigantea) across the northern Sonoran Desert. J. Arid Environ. 88, 57–69 (2013).
39. Orum, T.V., Ferguson, N. & Mihail, J.D. Saguaro (Carnegiea gigantea) mortality and population regeneration in the Cactus Forest of Saguaro National Park: seventy-five years and counting. PLOS ONE 11, e0160899 (2016).
40. Wilder, B. Scorching saguaros: Climate extremes are killing cactuses. Axios, 19 May 2025; https://www.axios.com/local/phoenix/2025/05/19/scorching-saguaros-heat-drought-cactus-deaths.
41. Hultine, K.R. et al. Relationships among climate, stem growth, and biomass δ13C in the giant saguaro cactus (Carnegiea gigantea). Ecosphere 9, e02498 (2018).
42. Copetti, D. et al. An improved genome assembly of the saguaro cactus (Carnegiea gigantea (Engelm.) Britton & Rose). Preprint at

https://www.researchgate.net/publication/370025962_An_improved_genome_assembly_of_the_saguaro_cactus_Carnegiea_gigantea_Engelm_Britton_Rose#citations (2023).
43. Sanderson, M.J. el al. Origin and diversification of the saguaro cactus (Carnegiea gigantea): a within-species phylogenomic analysis. Syst. Biol. 71, 1178–1194 (2022).
44. Renton, M. Plant migration and persistence under climate change in sessile organisms: the case of plants. Ecol. Model. 250, 153–164 (2013).
45. Thurman, L.L., Keller, S.R. & Gougherty, A.V. Persist in place or shift in space? Evaluating the adaptive capacity of plants under climate change. Front. Ecol. Environ. 18, 451–459 (2020).
46. Loreto, F. Climate challenges: can plants adapt in time? Front. Sci. 2, 1522649 (2024).
47. Miryeganeh, M. Plants are sessile organisms that need to adapt to constantly changing environmental conditions. Genes 12, 1185 (2021).
48. Félix-Burruel, R.E. et al. Episodic recruitment in the saguaro cactus is driven by multidecadal periodicities. Ecology 102, e03458 (2021).
49. Félix-Burruel, R.E. et al. Population decline of the saguaro cactus throughout its distribution is associated with climate change. Ann. Bot. 135, 317–328 (2025).
50. Hernández-Hernández, T., Brown, J.W., Schlumpberger, B.O., Eguiarte, L.E. and Magallón, S., 2014. Beyond aridification: multiple explanations for the elevated diversification of cacti in the New World Succulent Biome. New phytologist, 202(4), pp.1382-1397.
51. Heywood, V.H. Conserving plants within and beyond protected areas. Plant Divers. Resour. 40, 136–144 (2018).
52. Bellon, M.R., Gonzalez, C. & Brush, S.B. In situ conservation—harnessing natural and human-derived evolutionary potential. Evol. Appl. 10, 6–27 (2017).
53. Xia, X. et al. Developing long-term conservation priority planning for threatened medicinal plants under climate change. BMC Biol. 20, 153 (2022).
54. Manda, S.et al. Climate change reveals contractions and expansions in the habitat suitability of Vigna crop wild relatives in Benin. Front. Conserv. Sci. 3, 870041 (2022).
55. Qu, Y. et al. Planning priority conservation areas under climate change for six plant species with extremely small populations in China. Nat. Conserv. 26, 55–75 (2018).
56. Civantos-Gómez,M. et al. Climate change conditions the selection of rust-resistant crop wild relatives. Front. Plant Sci. 13, 1010799 (2022).
57. Aragón-Gastélum, J.L. et al. Assessment of the impact of climate change on suitable habitats of melliferous plant species. Climate 5, 44 (2025).
58. Greenwood, S. et al. Using in situ management to conserve biodiversity under climate change. J. Appl. Ecol. 53, 885–894 (2016).
59. Bátori, Z. et al. Managing climate change microrefugia for vascular plants. For. Ecol. Manage. 488, 118989 (2021).
60. D'Aloia, C.C. et al. Coupled networks of permanent protected areas and dynamic conservation areas for biodiversity conservation under climate change. Front. Ecol. Evol. 7, 27 (2019).
61. Weng, Y. et al. Dynamic conservation strategies for protected areas of biodiversity under climate change. Ecol. Indic. 162, 111123 (2024).
62. Walls, S.C. Coping with constraints: achieving effective conservation with limited resources. Front. Ecol. Evol. 6, 24 (2018).
63. Bottrill, M.C. et al. Is conservation triage just smart decision-making? Trends Ecol. Evol. 23, 649–654 (2008).
64. Mounce, R., Smith, P. & Brockington, S.F. Ex situ conservation of plant diversity in the world's botanic gardens. Nat. Plants 3, 795–802 (2017).

65. Griffith, M.P. et al. Can a botanic garden cycad collection capture the genetic diversity in a wild population?. Am. J. Bot. 102, 46–57 (2015).
66. Griffith, M.P. et al. Can a botanic garden metacollection better conserve wild plant diversity? A case study comparing pooled collections with an ideal sampling model. Int. J. Plant Sci. 181, 485–496 (2020)
67. Griffith, M.P.,Meyer, A. & Grinage, A. Global ex situ conservation of palms: living treasures for research and education. Front. For. Glob. Change 4, 711414 (2021).
68. Linsky, J. et al. Integrated plant conservation through the Global Conservation Consortia. Appl. Plant Sci. 12, e11586 (2024).
69. Hoban, S. Taxonomic similarity does not predict necessary sample size for ex situ conservation: a comparison among five genera. Proc. R. Soc. 287, 20200102 (2020)
70. Hoban, S. et al. How many individuals? How many populations?, in Heywood, V. et al Eds. Species Recovery Manual (Botanic Gardens Conservation International, 2018).
71. Abrego-López, C. et al.Saguaro Initiatives at Desert Botanical Garden: Generating tools and resources to save threatened cactus species. Cactus Succul. J. 97, 241–256 (2025).
72. Andrews S. FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc (2010)
73. Catchen, J. et al. Stacks: an analysis tool set for population genomics. Mol. Ecol. 22, 3124–3140 (2013).
74. Rivera-Colón, A.G. & Catchen, J. Population genomics analysis with RAD, reprised: Stacks 2. Methods Mol. Biol. 2498, 99–149 (2022).
75. Copetti, D. et al. Extensive gene tree discordance and hemiplasy shaped the genomes of North American columnar cacti. Proc. Natl. Acad. Sci. U.S.A. 114, 12003–12008 (2017).
76. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760 (2009).
77. Danecek, P. et al. Twelve years of SAMtools and BCFtools. Gigascience 10, giab008 (2021).
78. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 4, 7 (2015).
79. R Core Team, R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Preprint at https://cran.r-project.org/doc/manuals/r-release/fullrefman.pdf (2025).
80. Alexander, D.H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 19, 1655–1664 (2009).
81. Kopelman, N.M. et al.Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. Mol. Ecol. Resour. 15, 1179–1191 (2015).
82. Ortiz, E.M. Vcf2phylip v2.0: Convert a VCF Matrix into Several Matrix Formats for Phylogenetic Analysis (2019)
83. Leaché, A.D. et al. Stamatakis, Short tree, long tree, right tree, wrong tree: New acquisition bias corrections for inferring SNP phylogenies. Syst. Biol. 64, 1032–1047 (2015).
84. Martin, B.T.M. raxml_ascbias: Script for removing or counting invariant sites for the RAxML ascertainment bias corrections (2018
85. Minh, B.Q. et al. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Mol. Biol. Evol. 37, 1530–1534 (2020).
86. Kalyaanamoorthy, S. et al. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods 14, 587–589 (2017).

87. Hemstrom, W. & Jones, M. snpR: User friendly population genomics for SNP data sets with categorical metadata. Mol. Ecol. Resour. 23, 962–973 (2023).
88. Liu, X. & Fu, Y.X. Stairway Plot 2: demographic history inference with folded SNP frequency spectra. Genome Biol. 21, 280 (2020).
89. Oracle Corporation, Java Development Kit (JDK) Version 1.8.0_44 (Oracle, 2015); https://www.oracle.com/java/.
90. Baumdicker, F. et al. Efficient ancestry and mutation simulation with msprime 1.0. Genetics 220, iyab229 (2022).
91. Watterson, G.A. On the number of segregating sites in genetical models without recombination. Theor. Popul. Biol. 7, 256–276 (1975).
92. Krueger, F. Trim Galore!: A wrapper around Cutadapt and FastQC to consistently apply adapter and quality trimming to FastQ files, with extra functionality for RRBS data (Babraham Institute, 2015).
93. Li,H., Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760 (2009).
94. Danecek, P. et al. 1000 Genomes Project Analysis Group, The variant call format and VCFtools. Bioinformatics 27, 2156–2158 (2011).
95. Willi, Y. et al. Accumulation of mutational load at the edges of a species range. Mol. Biol. Evol. 35, 781–791 (2018).
96. Leventhal, L., Ruffley, M.R. & Exposito-Alonso, M. Simulated climate change magnifies genetic vulnerabilities from mutation load and maladaptation. doi**:** https://doi.org/10.1101/2025.09.25.677362
97. Cingolani, P. et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly 6, 80–92 (2012).
98. Korunes, K.L. & Samuk, K. pixy: Unbiased estimation of nucleotide diversity and divergence in the presence of missing data. Mol. Ecol. Resour. 21, 1359–1368 (2021).
99. Wright, S. Evolution and the Genetics of Populations, Volume 4: Variability within and among Natural Populations (University of Chicago Press, 1984).
100. Delaneau, O., Zagury, J.F. & Marchini, J. Improved whole-chromosome phasing for disease and population genetic studies. Nat. Methods 10, 5–6 (2013).
101. Pertea, G. & Pertea, M. GFF utilities: GffRead and Gff Compare. F1000Res. 9, 304 (2020).
102. Steenwyk, J.L. et al. PhyKIT: a broadly applicable UNIX shell toolkit for processing and analyzing phylogenomic data. Bioinformatics 37, 2325–2331 (2021).
103. Pond, K. & Frost, S. Not so different after all: Comparison methods detecting amino acid sites under selection. Mol. Biol. Evol. 22, 1208–1222 (2005).
104. Kosakovsky Pond, S.L.K. et al. HyPhy 2.5-A customizable platform for evolutionary hypothesis testing using phylogenies. Mol. Biol. Evol. 37, 295–299 (2020).
105. Lange, S. ISIMIP3b bias adjustment fact sheet. Preprint at https://www.isimip.org/documents/413/ISIMIP3b_bias_adjustment_fact_sheet_Gnsz7CO.pdf (2021).
106. Vicente-Serrano, S.M., Beguería, S. & López-Moreno, J.I. A multiscalar drought index sensitive to global warming: The standardized precipitation evapotranspiration index. J. Clim. 23, 1696–1718 (2010).

107. Kass, J.M. et al. ENMeval 2.0: Redesigned for customizable and reproducible modeling of species' niches and distributions. Methods Ecol. Evol. 12, 1602–1608 (2021).
108. Phillips, S.J. & Dudík, M. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. Ecography 31, 161–175 (2008).
109. Phillips, S.J., Anderson, R.P. & Schapire, R.E. Maximum entropy modeling of species geographic distributions. Ecol. Modell. 190, 231–259 (2006).
110. Friedman, J., Tibshirani, R. & Hastie, T. Regularization paths for generalized linear models via coordinate descent. J. Stat. Softw. 33, 1–22 (2010).
111. Tay, J.K., Narasimhan, B. & Hastie, T. Elastic net regularization paths for all generalized linear models. J. Stat. Softw. 106, 1–31 (2023).
112. Hijmans, R.J. terra: Spatial data analysis. Preprint at: https://cran.r-project.org/web/packages/terra/terra.pdf ,(2024);
113. Peterson, R.A. Finding optimal normalizing transformations via bestNormalize. The R Journal (2021)
114. Fox, J. & Weisberg, S. An R Companion to Applied Regression. Sage [Preprint] (2019). https://www.john-fox.ca/Companion/.
115. Lenth, R.V. emmeans: Estimated marginal means, aka least-squares means (2024);
116. Ziyatdinov, A. et al. Lme4qtl: Linear mixed models with flexible covariance structure for genetic studies of related individuals. BMC Bioinformatics 19, 68 (2018).
117. Kuznetsova, A., Brockhoff, P.B. & Christensen, R.H.B. lmerTest package: Tests in linear mixed effects models. J. Stat. Softw. 82, 1–26 (2017).
118. Kuhn, M., Wickham, H. & Hvitfeldt, E. recipes: Preprocessing and feature engineering steps for modeling (2025).
119. Oksanen, J. et al. vegan: Community ecology package (2024).
120. Yu, J. et al. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat. Genet. 38, 203–208 (2006).
121. Huang, X. et al. Genome-wide association studies of 14 agronomic traits in rice landraces. Nat. Genet. 42, 961–967 (2010).
122. Quinlan, A.R., & Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842 (2010).
123. Cantalapiedra, C.P. EggNOG-mapper v2: Functional annotation, orthology assignments, and domain prediction at the metagenomic scale. Mol. Biol. Evol. 38, 5825–5829 (2021)

Acknowledgments

We would like to thank Pedro Andrade, José F. Martínez, Alexandra Schuessler, Hal Hoover and Ryan Koeth for their assistance during fieldwork, seed processing, and help during the early stages of this project. We used ChatGPT (OpenAI, 2025) to assist with language editing, grammar correction, and overall improvement of the manuscript's clarity and readability.

Author contributions:

Conceptualization: THH
Methodology: THH, EMG, TV, MP, FMF
Formal analysis: EMG, MP, TV, MCG, AHL, PF, DML
Visualization: EMG, MP, MCG, THH
Supervision: THH
Writing-original draft: THH
Writing-review & editing: all authors
Project administration: THH
Funding acquisition: THH

Data and materials availability: All necessary code to recreate the results in this work will be placed in https://github.com/hernandezt-lab. Raw sequencing data will be made available upon manuscript acceptance.

Figures and Tables

Figure 1. Study area and organism. Saguaro samples were collected along its distribution in the Sonoran Desert region (top left) by visiting 30 localities both in Mexico and the US. Two *ex-situ* collections were also included (the National Cactaceae collection in DBG and Urban saguaros from different neighborhoods in the Phoenix Valley). For each wild locality (for locality names see Supplementary Data), a mean Health-Score (HS) was estimated by integrating phenotypic, reproductive and demographic observations, and it is shown as a scale from dark-red (high HS) to white (low HS).

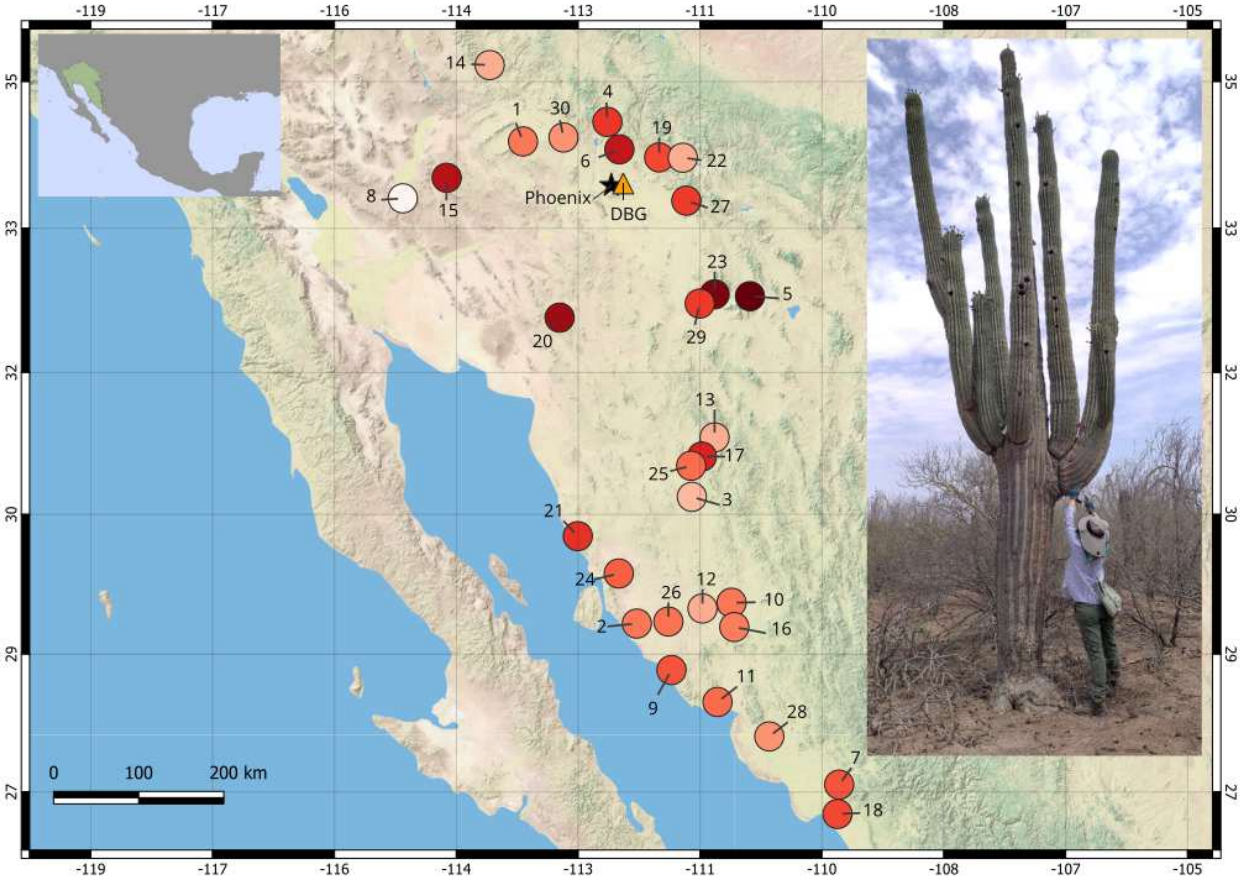


Figure 2. Genetic diversity and population structure of saguaro cactus. (a) Individual loadings dotplot (left) and distributions (right) along the first and second principal components from a PCA for all sequenced individuals and (b) for the North cluster. Wild collected samples are shown in blue, while *ex situ* preserved samples collected in the National Cactus Collection at DBG are shown in yellow and Urban samples collected in local neighborhoods in the Phoenix Metropolitan Valley, AZ are shown in red. (c) Midpoint-divided maximum-likelihood (ML) phylogenetic tree of all saguaro samples. The tree shown is the consensus of 1,000 bootstrap replicates. Nodes with bootstrap support equal to or above 95% are shown with black asterisks. (d) Individual sample assignments to two ($K = 2$), three ($K = 3$), and five ($K = 5$) ancestral populations. Sampling localities are given on the *x*-axis and are ordered in approximate order according to latitude.

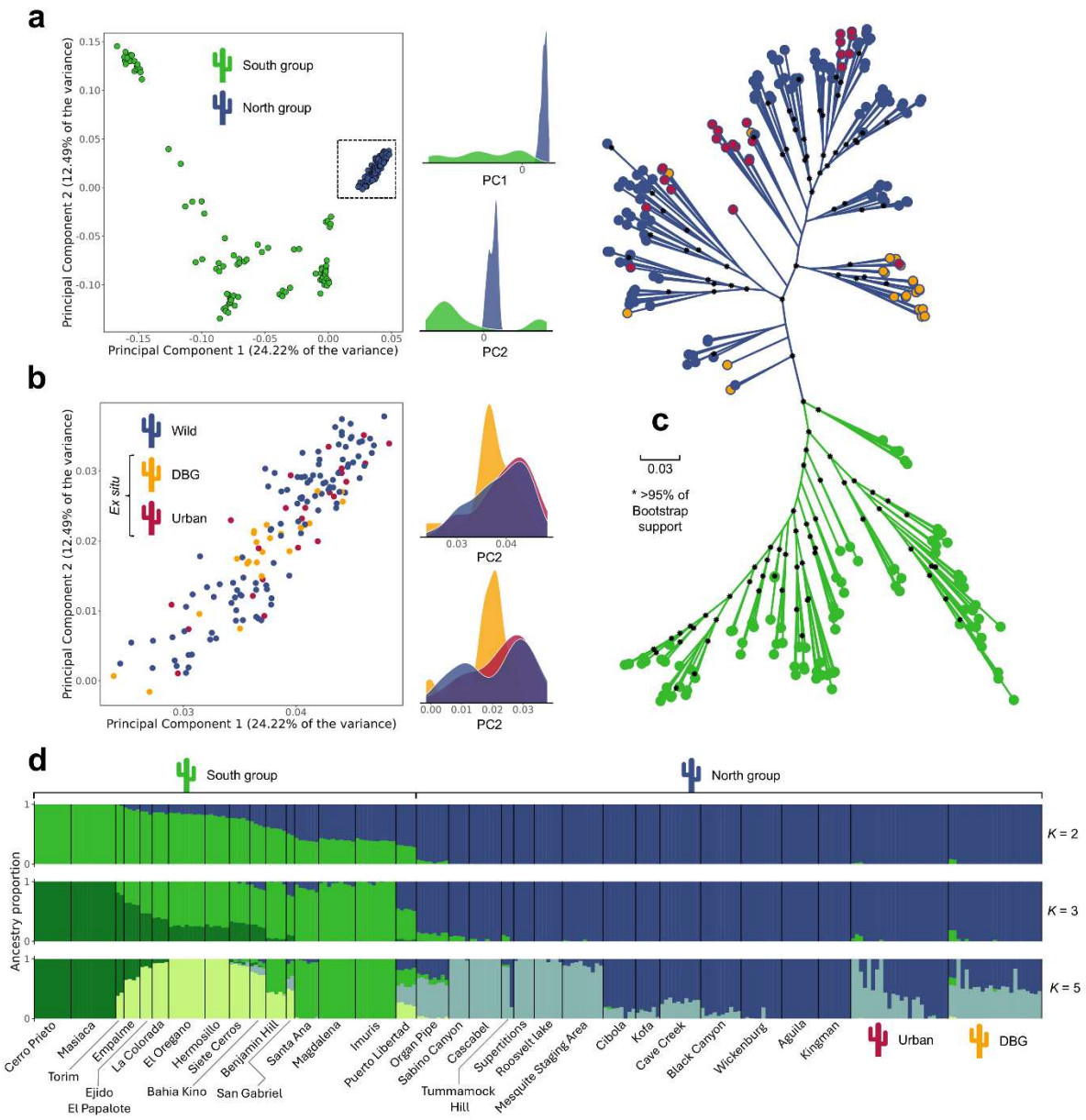

a
South group
North group
Principal Component 2 (12.49% of the variance)
Principal Component 1 (24.22% of the variance)
PC1
PC2
b
Wild
DBG
Urban
Ex situ
Principal Component 2 (12.49% of the variance)
Principal Component 1 (24.22% of the variance)
PC2
PC2
c
0.03
* >95% of Bootstrap support
d
South group
North group
Ancestry proportion
K = 2
K = 3
K = 5
Cerro Prieto
Masiaca
Torim
Empalme
Ejido El Papalote
La Colorada
El Oregano
Hermosillo
Bahia Kino
Siete Cerros
Benjamin Hill
San Gabriel
Santa Ana
Magdalena
Imuris
Puerto Libertad
Organ Pipe
Sabino Canyon
Tummamock Hill
Cascabel
Superstitions
Roosvelt lake
Mesquite Staging Area
Cibola
Kofa
Cave Creek
Black Canyon
Wickenburg
Aguila
Kingman
Urban
DBG

Figure 3. Demographic insights and contrasting selection patterns in North and South saguaro groups. (a) Folded site frequency spectra (SFS) for the North and South saguaro groups. The SFS for a Wright-Fisher population is shown in black bars. Kolmogorov–Smirnov test (KS) *P* values for the comparison against the Wright-Fisher SFS are shown. (b) Historical changes in effective population size (N_e) inferred for each group. The median estimate is shown in black; shaded areas represent uncertainty intervals, with bright colors indicating the 12.5–87.5 percentile range and opaque colors the 2.5–97.5 percentile range. Median and 12.5–87.5 percentile values for the present time (~10 generations ago) are shown. The demographic reconstruction for a simulated Wright-Fisher population is shown in red. (c) Comparison of genetic divergence and diversity in genic regions between the North and South groups; the 41 candidate genes are outlined in black. (d) Codon-based tests of selection across genes with and without constrained synonymous diversity. Each point represents a gene plotted by its gene-wide dN/dS estimate (x-axis). Vertical dashed lines denote thresholds used to assess deviation from neutrality; illustrating the relationship between overall selective regime, site-specific positive selection, and background levels of synonymous constraint.

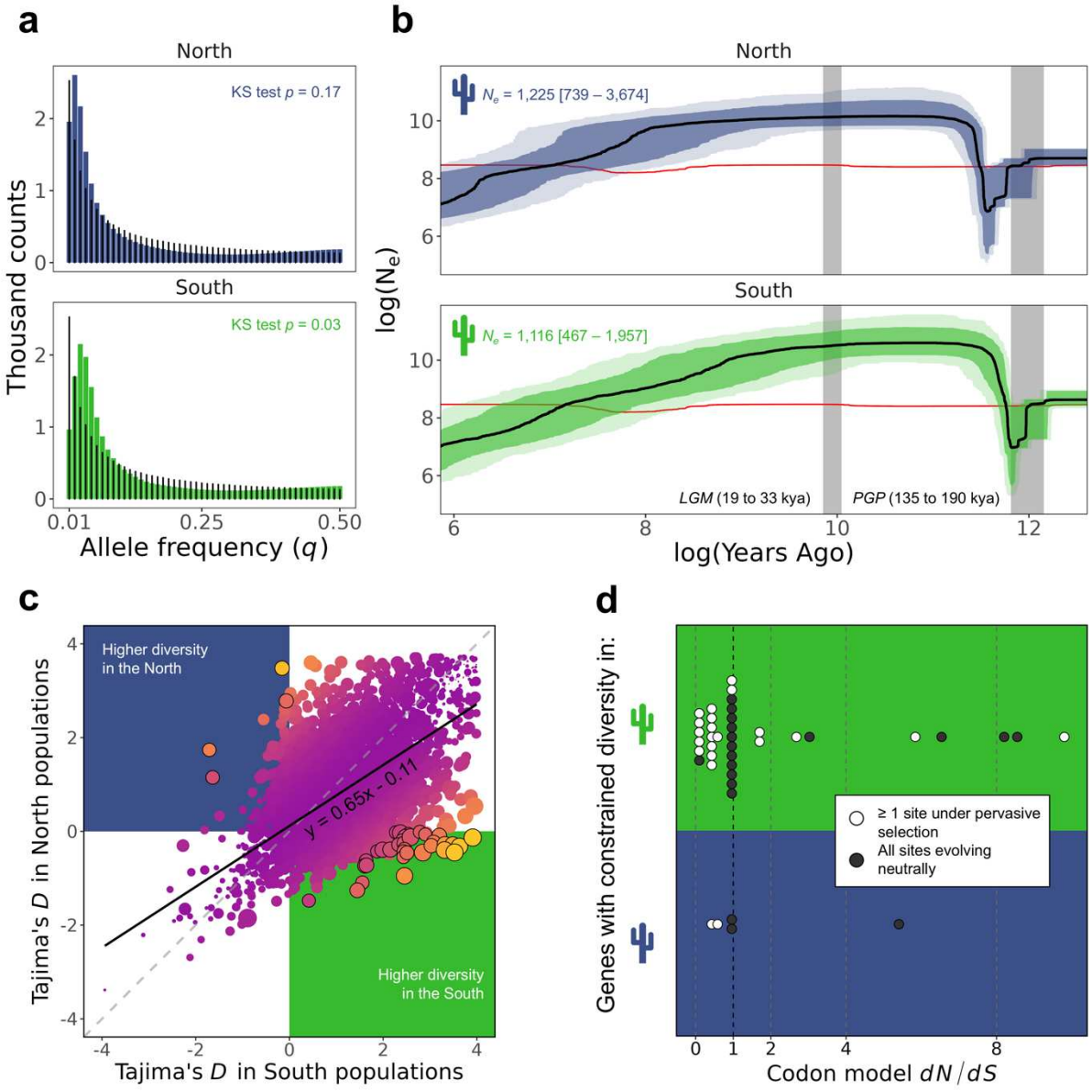

a
North
KS test p = 0.17
South
KS test p = 0.03
Thousand counts
Allele frequency (q)
b
North
Ne = 1,225 [739 – 3,674]
South
Ne = 1,116 [467 – 1,957]
log(Ne)
LGM (19 to 33 kya)
PGP (135 to 190 kya)
log(Years Ago)
c
Higher diversity in the North
Higher diversity in the South
y = 0.65x - 0.11
Tajima's D in North populations
Tajima's D in South populations
d
Genes with constrained diversity in:
≥ 1 site under pervasive selection
All sites evolving neutrally
Codon model dN/dS

Figure 4. Environmental PCA and environmental niche modeling for North and South saguaro groups. (a) PCA of Environmental data. (b) Species distribution models for the Saguaro cactus North lineage, and (c) South lineage. Models for the MPI-ESM1-2-HR GCM are shown.

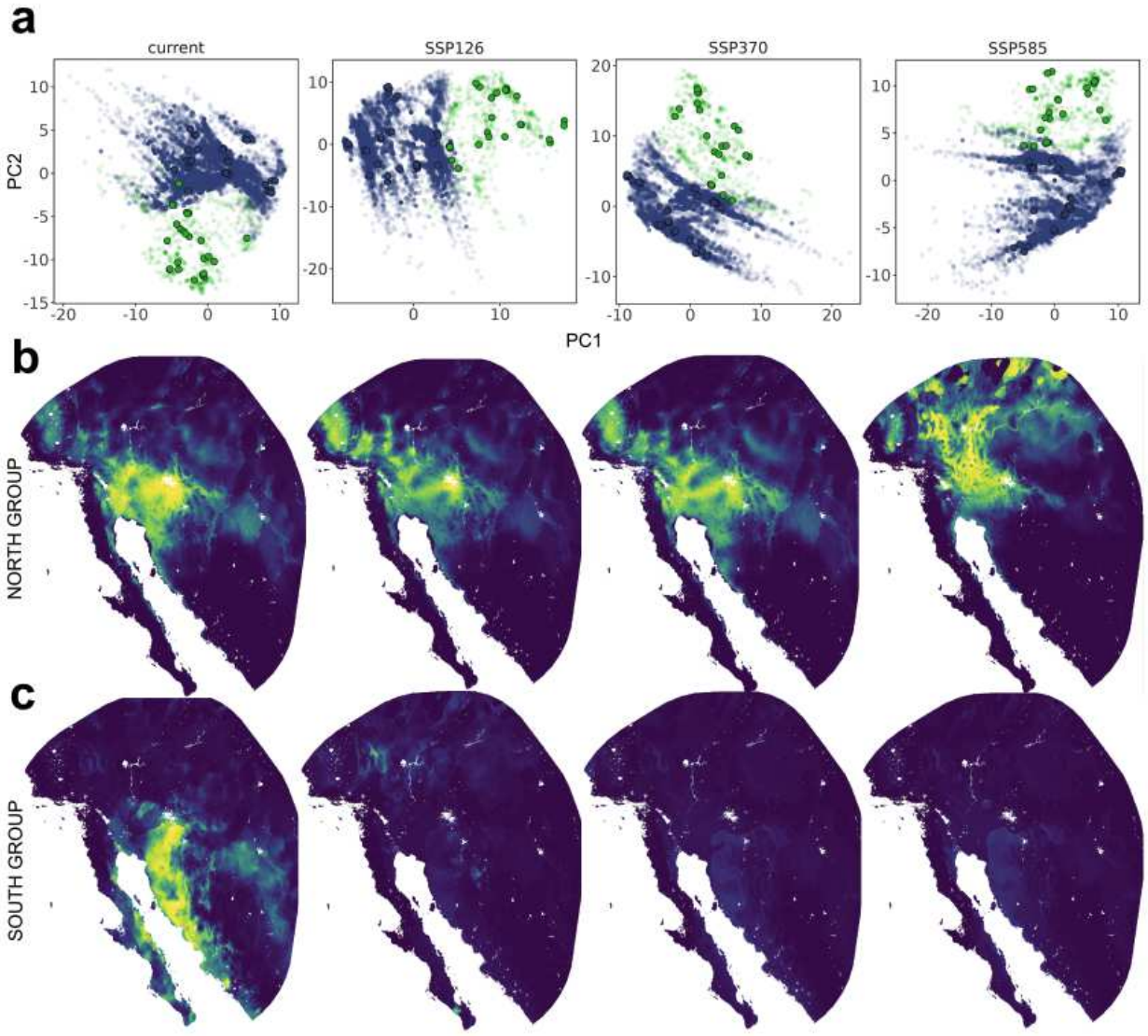
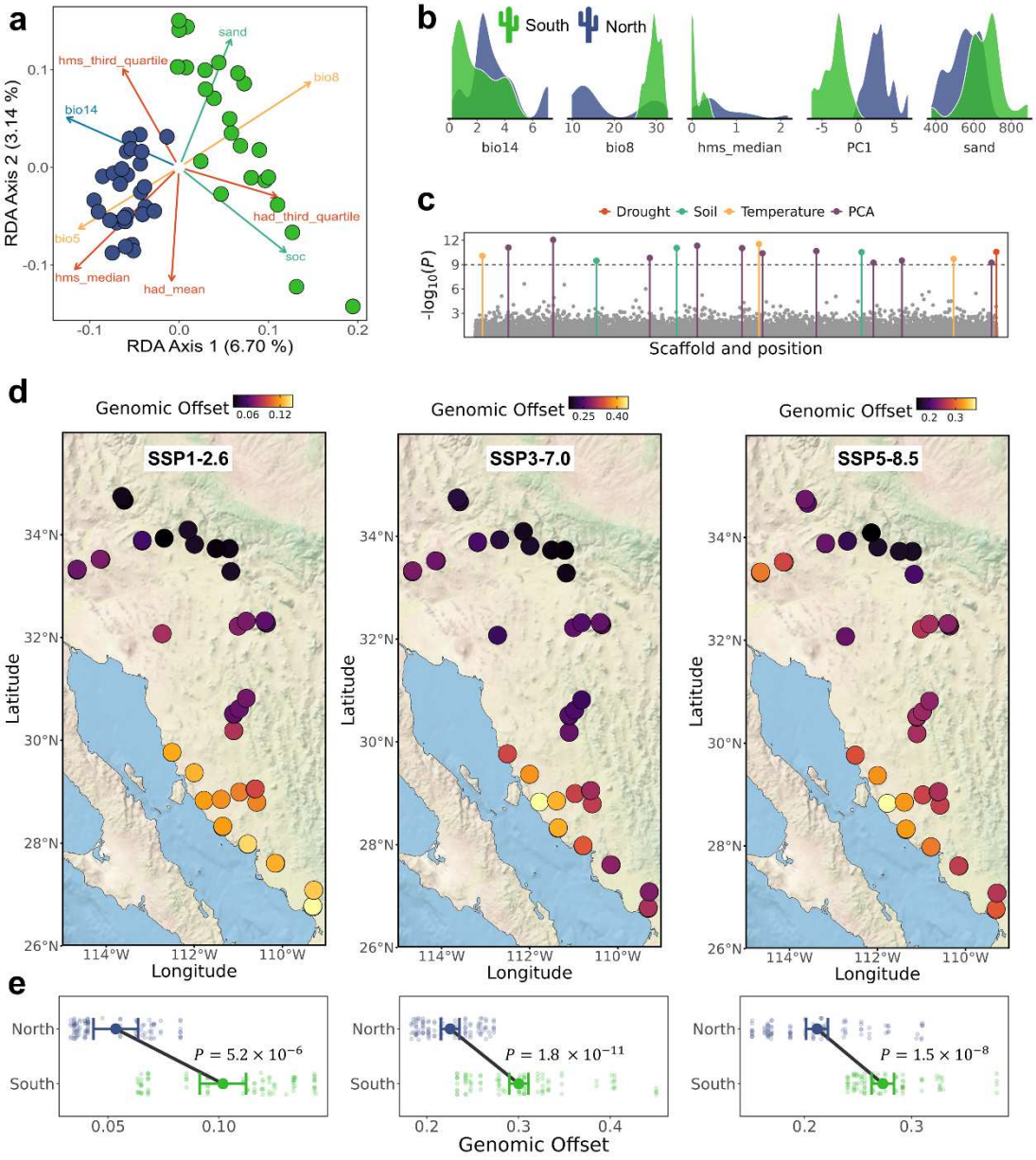


Figure 5. Environmental drivers of adaptive differentiation and projected genomic offset between North and South saguaro populations under future climate scenarios. (a) Redundancy analysis (RDA) showing the relationship between genomic variation and environmental predictors. Arrows indicate the direction and strength of the top explanatory environmental variables. (b) Distributions of North and South genetic groups for representative environmental predictors from each category. (c) Manhattan plot of genome–environment association results; the dotted line denotes the significance threshold (1×10^{-9}). (d) Predicted genomic offset under three future climate scenarios (SSP1-2.6, SSP7-7.0, SSP8-8.5). (e) Distribution of genomic offset estimates for each genetic group across scenarios. Points represent estimated marginal means and bars denote standard errors. *P* values for pairwise group comparisons are shown.



Supplementary Materials

Supplementary Text: Supplementary Methods (Health Score Determination, DNA extraction: modified CTAB protocol, Equations)
Figs. S1 to S8
Tables S1 to S9

Other Supplementary Materials for this manuscript include the following:
Mendoza_etal2026_Supplementary Data_v1.tsv

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Mendozaetal2026SupplementaryDatav3.txt
- Mendozaetal2026SupplMatv6.pdf