Project: Final report 2025: Critically Needed Genetic Assessment of Proposed Endangered Plant at Coronado National Memorial

PMIS 316237

Coronado National Memorial

September 2022

Strategic Plan Goals: GPRA Goal Ia1B

Organizations involved:
Northern Arizona University
U.S. Fish and Wildlife Service

Principal Project Manager

/s/ _____ Date 2022
JoAnn Blalack
Integrated Resources Manager, Coronado National Memorial

/s/ _____ Date 2022
XXXX
Superintendent, National Park of American Samoa

NRSS Project Coordinator

/s/ _____ Date 2022
Katie VinZant
Restoration Ecologist, WASO-NRSS

INTRODUCTION

Assessing genetic diversity in endangered plants

Nearly 40% of vascular plants are considered at risk of extinction, necessitating broadscale efforts to conserve biodiversity (Nic Lughadha et al., 2020)**.** Rare plant conservation planning is predicated on understanding the distribution of genetic variation across the landscape (Dennis et al., 1991, Dibner et al., 2019, Lande & Shannon 1996). Land managers and conservationists apply this knowledge to select conservation targets (Hedrick & Miller, 1992), identify critical corridors to maintain gene flow (Segelbacher et al., 2010), and introduce novel genotypes across the landscape via managed relocation or assisted evolution (Holderegger et al., 2019). The importance of integrating genetic approaches in conservation and land management is increasing, as rapid anthropogenic global change applies novel selective regimes on already imperiled species.

Rare plants are often characterized by small population sizes that shape demographic and evolutionary trajectories and thus conservation approaches. Small populations are subject to demographic and genetic stochasticity that increase the likelihood of random allele loss, potentially leading to non-adaptive shifts in gene frequencies (i.e., genetic drift). When small population size increases inbreeding, individual fitness may decrease due to inbreeding depression via the expression of deleterious recessive alleles (Hamabata et al., 2019). Other factors, like Allee effects, that operate in small populations to reduce *per capita* reproductive rates, reinforce patterns of rarity. Smaller patch sizes of flowering plants are less effective at attracting pollinators due to pollen limitation and diminished visual cues, driving Allee effects while also reducing the potential for genetic rescue from neighboring populations (Anic et al., 2014). These factors — inbreeding depression, Allee effects, and genetic drift — can create a feedback loop termed ‘the extinction

vortex' that iteratively reduces population size, driving populations toward extinction (Nabutanyi & Wittmann 2021).

Rapid climatic and environmental changes can outpace the adaptive potential of rare plants (Jump & Penuelas, 2005), which often possess life history traits associated with low rates of reproduction and slow growth (Adler et al., 2014). Further, if small population size erodes genetically-based phenotypic variation, plant species are less likely to adapt to novel environmental conditions. This is particularly true for dryland plants with restricted ranges that exist at the edge of physiological tolerances to aridity such that even small environmental changes can cause widespread mortality (Harrison et al., 2019, Laikre et al., 2010, Richardson et al., 2023). Conversely, when populations are connected, rather than isolated, gene flow maintains genetic diversity through admixture with populations in other selective environments. For this reason, rare species conservation often focuses on maintaining genetic diversity both within and across populations and may target maintenance of connectivity, particularly with populations that possess distinct genetic signatures, and thus could represent unique adaptations to the environment, or populations with high levels of genetic diversity that serve as a repository of adaptive phenotypic variation.

If climate change and anthropogenic influences continue as projected, pre-established patterns of local adaptation may render species maladapted to current sites. Species comprised of locally adapted populations may be particularly vulnerable to climate change, since these species have site-specific adaptations and thus are often specialized to a narrower range of conditions relative to wide-spread, generalist species. In this way, restoration and conservation initiatives that favor the "local is best" philosophy may unintentionally reinforce patterns of lower fitness by introducing seeds maladapted to novel climates (Hereford, 2009; Sgrò et al., 2011). Local

adaptation is widely thought to be commonplace in large plant populations, but a meta-analysis found that it is less common for small plant populations (Leimu & Fischer, 2008). Identifying patterns of adaptation can help inform conservation strategies, particularly in the case of interventions to increase population sizes by introducing non-local individuals. Admixture with non-local populations may increase performance by reducing inbreeding depression and increase adaptive potential by introducing novel alleles on which selection can act. However, the opposite could be true, if outbreeding depression, a reduction in fitness due to introduction of non-local maladaptive genotypes, lowers performance.

Pectis imberbis and the Sky Islands

Arid and semiarid ecosystems, here broadly defined as systems that receive 500 millimeters or less of precipitation annually (*herein, referred to as 'arid' systems for brevity*), are projected to experience significant impacts on biodiversity due to global temperature increases and precipitation variability (Salguero-Gómez et al., 2012). Arid ecosystems support numerous endemic and rare species, yet few studies have investigated the genetic diversity and structure of range-restricted plants in the Southwest (Harrison et al., 2019; Jones et al., 2021; Wolfe et al., 2016). This is particularly true in the Sky Island region of southern Arizona. Located at the northern edge of the Madrean Archipelago, this mountainous region is typified by high levels of biodiversity, as several distinct vegetation classes coincide. Precipitous peaks, which host a variety of community-types, emerge from desert scrub and grasslands, naturally increasing isolation among communities located on different mountain ranges.

Pectis imberbis A. Gray (Asteraceae) is an endangered perennial forb native to the Sky Islands, with a historical range stretching from the Santa Rita mountains of Southeastern Arizona down to southern Sonora in Mexico. Within this region, *P. imberbis* occurs in oak woodland/savanna habitats. Original decline of this species was believed to be driven by

introduction of domesticated livestock by European settlers. Notably, however, significant declines including the extirpation of at least 9 populations in the US, and potentially all of the 8 populations located in MX, occurred since the 1970s after cattle grazing was significantly reduced. Recent studies suggest climate change may interact with a variety of stressors, including herbivory by both livestock and wild ungulates and occurrence of non-native species, to drive decline. Of the nine remaining *P. imberbis* populations, only three exhibit population growth rates above replacement level - all located in or adjacent to the Coronado National Memorial (CORO). Moreover, 80% of all known individuals occur within the CORO. This pattern could, in part, be explained by resilience fostered by connectivity and positive impacts on reproduction and genetic diversity afforded by large population size.

PROBLEM STATEMENT

The CORO supports most of all known *P. imberbis* individuals and contains two of the three populations that exhibit positive population growth, with about two-thirds of the park designated as critical habitat. This spatial concentration of viable populations renders the species particularly vulnerable, especially since the Southwest is experiencing dramatic changes in fire frequency and severity, climate, and vegetative community composition as non-native grasses invade ecosystems. One catastrophic fire, severe drought or other disturbance may imperil most of the known *P. imberbis* individuals. Loss of the CORO populations could potentially reduce genetic diversity below a threshold for recovery, since most individuals occur in this small area. Immediate action is needed to shift the trajectory of *P. imberbis* towards recovery by preemptively addressing threats through strategic, targeted management and risk reducing measures. Actions including the protection of genetically based adaptive phenotypic variation and establishment of populations in novel areas within the CORO or beyond the CORO park boundary are urgently needed prior to

further population and genetic loss.

APPROACH

To describe patterns of genetic variation to guide conservation and management of *P. imberbis*, we employed reduced representation genomic sequencing (RADSeq). Advanced genomic tools like RADSeq provide high-resolution insights into genetic diversity, population structure, and adaptive capacity of rare and endangered species. RADSeq can be used to associate population structure with environmental predictors, offering quantitative insights into the biological underpinnings of adaptive capacity or genetic diversity (Beever et al., 2016; Cao et al., 2020). RADSeq employs restriction-site enzymes to produce libraries of single-nucleotide-polymorphisms (SNPs) that allows for the quantification of genetic diversity and population structure coefficients such as expected heterozygosity (He), pairwise genetic differentiation values (F_{ST}), and nucleotide diversity (π) (Schmidt & Jensen 2000, Hevroy et al. 2018, Jones et al 2021). Moreover, generating SNPs allows for the implementation of Redundancy Analyses (RDAs), which visualize the context of genetic differentiation as it relates to climatic and habitat drivers of genetic variation (Capblancq et al., 2018). As a complement to RADSeq and RDAs, common gardens explicitly test for adaptive differentiation among populations and can be used to understand the relationship between particular traits and fitness, as well as the selective environments that maximize performance (Clausen et al., 1941). Together, SNPs and common gardens can provide insight to genetic structure and performance.

SPECIFIC OBJECTIVES

To improve *P. imberbis* conservation and management, we identified the following five objectives (*Note: These objectives appear in different sequence than in the Detailed Implementation Plan (DIP)*):

1) Complete comprehensive population counts and GIS mapping of beardless chinchweed on NPS property to inform:

a. Decisions on where to conduct genetic testing and restoration efforts;

b. USFWS Species Recovery Plan;

c. Management actions to reduce threats to individuals/populations.

We successfully updated population estimates for most known populations in CORO. Initial assessments indicate that the CORO supports at least 4883 individuals, around 80% of all known *P. imberbis* plants. Locations of populations were provided to the State Heritage Program by Martha Sample in 2024. We updated point cloud information for individual plants and used their coordinates to assess fine scale environmental associations with genetic diversity and survival.

2) Identify beardless chinchweed habitat vital to the species survival within CORO to guide management activities.

Using Species Distribution Models (SDMs), we identified the following predictors of suitable habitat. The methodology and results from SDM modeling is found here: Species Distribution Models for Pectis imberbis, a Rare Plant Species in Southeastern Arizona - ScienceBase-Catalog. This accompanies the USGS data release (*Appendix A*).

3) When conducting populations surveys and collecting genetic samples from populations, collect seed for *ex situ* and *in situ* restoration efforts.

We collected seed from all known populations of *P. imberbis* according to protocols developed by Seeds of Success. We used collections to establish an experimental restoration garden, also used to test for patterns of adaptive differentiation. All remaining accessions were given to the Museum of Northern Arizona.

4) Identify genetically distinct/diverse populations for prioritized conservation.

Using Small Nucleotide Peptide (SNP) markers, we identified three genetically distinct populations and used this information to assess common garden performance for different genetic groups. We also derived estimates of gene flow between these populations and compared migration surfaces to an ensemble species distribution model to inform *ex situ* restoration efforts, specifically with suggestions to increase genetic connectivity between populations using a steppingstone planting strategy.

5) Develop a protocol for beardless chinchweed *ex situ* and *in situ* conservation/restoration

Connectivity appears important for supporting positive population growth and recovery following disturbance. Moreover, we did not detect evidence of local adaptation, though sample size was limited. Together, these findings suggest the need to establish populations within areas of high habitat suitability that connect populations of *P. imberbis*, supporting active gene flow among populations. Greater details are provided in the report below.

The following report outlines methods and results in further detail and contextualizes the results in a discussion section.

METHODS

Habitat modeling – *See Appendix A*.

Quantitative genetics methods

DNA Extraction and Sequencing

In summer 2022, we collected foliar samples from 281 individuals across 32 subpopulations of *P. imberbis* representing its known range in this region (Souther et al., 2022). We defined subpopulations as clusters of plants at least 150 meters from conspecifics or separated by other significant natural or manmade barriers. We randomly selected approximately 10 fresh leaves per individual, which were then desiccated and frozen at -80° C until extraction. We sourced the sampled individuals from sites in proximity to three different Sky Island Mountain ranges (the Huachuca mountains, Santa Rita mountains, and Atascosa-Pajarito mountains), and recorded sample coordinates for each population cluster. Specifically, we extensively sampled *P. imberbis* at the following five sites: Coronado National Memorial (CNM), Anne Tank Wash (ATW), Scotia Canyon (SCO), O'Donnell Canyon (ODC), Santa Rita mountains (WSP & MCC), and Pena Blanca (PBC).

We then prepared leaf material samples in Qiagen Extraction tubes (Qiagen, Valencia, CA, USA) and shipped to the University of Minnesota Genomics facility for DNA extraction and Next-Generation genomic sequencing. Following DNA extraction, the sequencing facility created 2 pools of inline barcoded SbfI RAD-seq libraries. Each pool was sequenced on separate lanes of a NovaSeq S1 1 x 100-base pair run (Illumina inc., UK), generating reads 91 base-pairs long at 26X coverage. For each pool, over 750 M pass filter reads for each pool were generated, totaling 94 GB of raw data. Mean phred quality scores were 30 or higher and sequencing produced an average of 5 million reads for each sample. Adapter content for each sample ranged from 2-8% and thus downstream analysis proceeded without the removal of adapters. However, DNA extraction yielded low amounts of material, thus optimization techniques of sequencing between two pools of data led to strong lane effects. As a result, we used only one pool of data (n=141) with greater sequencing

depth for the downstream analysis to maximize statistical power. The higher quality sequencing pool we selected for the final population genetic analysis retained representative subpopulations from all broader populations of interest, representing three separate sky island vicinities. Following filtering steps with a cutoff of 40% maximum missing data per individual, we excluded 3 additional individuals, resulting in a total of 137 individuals across 17 subpopulations and 6 broader populations

SNP calling and filtering using Stacks

We analyzed Fastq files containing RAD tags using the Stacks (Version 2.66) (Catchen et al. 2013) de novo pipeline. We used the denovo_map.pl pipeline wrapper to assemble orthologous tags into stacks, generate a catalog of putative RAD loci, and transpose the data to reorient it by locus (Dang et al 2022), and call SNPs (single nucleotide polymorphisms). To maintain more stringent SNP calling we assigned a value of 2 to all three primary stacks parameters: minimum depth of coverage required to create a stack (m), maximum distance allowed between stacks (M) and number of mismatches allowed between stacks (n) (Paris et al., 2017). Following catalog assembly, we used dynamic filtering steps to ensure higher confidence SNP calls were incorporated into the final matrix. We then used the output files from sstacks (produced in denovo.map_pl) to re-run filtering steps outside of the populations program in Stacks. We used the VCFtools program package (Version 0.1.16) to filter SNPs, outputting .012 snp file matrices that we then imported into R to produce PCAs with custom code (Faske et al 2021). The filtering steps retained only biallelic SNPs containing sequencing data for 40% of the samples and minor allele frequencies of 0.05%. After filtering, we retained 337 SNPs across 137 individuals and used them to assess population and landscape genetic variation.

Common garden methods

Seed sampling and site selection

In the summer of 2022, we collected seeds from 25 subpopulations of *P. imberbis* ranging from the southern Huachuca Mountains to the Santa Rita Mountains (**Fig. 1-2; Table 1**). We defined subpopulations as groups of plants at least 150 meters from other groups or separated by other significant natural or manmade barriers. Seeds were collected following the Seeds of Success (SOS) protocol to limit negative ramifications on populations (Seeds of Success 2021). These protocols include collecting no more than 10% of the seeds at any given population, distributing collections across individuals, and minimizing disturbance during the collection process.

Germination and growth in greenhouse setting

During the spring of 2023, we initiated propagation of seeds in the NAU Research Greenhouse for later out-planting to the common garden site. We placed over 5000 seeds on a sandy substrate (quikrete premium play sand) under a 20-watt germination grow light and misted with water once a day to increase moisture content. we monitored seeds every 1-2 days until cotyledons emerged, after which they were immediately transplanted to a Ray Leech Super Cell 10 1.5in x 8.25 in cone-tainer (0.15L) filled with soil (1:1:1 mixture of sphagnum peat moss, horticultural perlite, and coarse vermiculite) and topped with slightly compacted, moistened sand. Transferred cotyledons were recorded as successful germinants and individual cone-tainers were watered to saturation every other day on average.

Common garden out-planting

Seedlings were monitored and transplanted over a 16-week period, with transplant date noted for each seedling. We monitored germination and seedling survival for five months until the

eventual out-planting of individuals at CNM in August 2023 during the monsoon season (**Table 2, 3)**. A 10 x10 m plot was established near local park facilities. To prepare the site, invasive grasses (*Eragrostis lehmannii* and *Eragrostis curvula*) were pulled by hand. The garden was fenced to prevent herbivory. Soil was intentionally disturbed with shovels and pickaxes prior to planting to break up rocky soil and allow for easier transfer of soil cones into the ground without damaging roots of seedlings. Seedlings were randomly assigned to positions within the array spaced approximately 50-100 cm apart in the 100m^2 garden (N = 181, **Fig. 3**). In addition to survival, the following traits were measured every two weeks for two months, followed by a final visit one year after planting: height (cm), stem number, crown diameters, total leaf count, and number of buds or flowers.

Statistical methods

Population structure and genetic diversity calculations

To assess between-population genetic differentiation, we converted genetic variation calculated from SNP frequencies into principal components and mapped them in multivariate space to visualize population genetic clustering (Jombart et al., 2010). Without *a priori* knowledge of population genetic patterns, we ran a K-means clustering algorithm using the '*find.clusters*' and '*kmeans*' functions in the R stats package (v 3.6.2) (Truelove et al., 2015). We assigned three centroids based on minimum Bayesian information criterion (BIC) to the mapped populations to investigate population structure (Hartigan & Wong, 1979; Milano et al., 2020). We performed a dispersion analysis to test for significant differences in dispersion values for three K-means clusters in PCA space. After assigning groups and testing for dispersion, we conducted a PERMANOVA test with the '*adonis2*' function in the Vegan package (Evans et al., 2023) (permutations = 10,000) to test for significant differences between three genetic clusters determined from kmeans. To

further determine genetic distances between populations at a higher resolution, we calculated pairwise genetic differentiation (FST) among 17 subpopulations and 6 populations using the hierfstat R package (Goudet & Jombart, 2020, Faske et al. 2021). To compare Fst to another estimate of between-population genetic diversity, we calculated Nei's D estimates of genetic diversity using custom code following Faske et al. (2021). Finally, we estimated migration of alleles between populations using Wright's Island Model of Migration, using FST as an estimated parameter (Wright 1951, Slatkin, 1985).

For within population genetic diversity estimates, we used expected heterozygosity and nucleotide diversity as the two primary indicators to assess relative genetic variability for each population. We calculated expected heterozygosity (He) at the subpopulation level using custom code again following Faske et al. (2021). We also calculated coefficients of nucleotide diversity (π) and Inbreeding (Fis) in the Stacks populations program for 17 subpopulations. We also summarized He and π values at the population level. To facilitate comparisons between populations, we adjusted expected heterozygosity (He) values for differences in sample size using a linear regression model. This adjustment was made because larger sample sizes typically exhibit higher heterozygosity, potentially biasing estimates (Sopniewskwe & Catullo, 2024). We corrected the He values by the standard error obtained from the regression, allowing for more accurate comparisons across populations. We similarly adjusted nucleotide diversity. These corrections enhance the reliability of the genetic diversity estimates and adjust the differences observed among populations to more accurately reflect true biological variation, rather than artifacts of sampling bias or population size discrepancies (Khatrwe & Burt, 2019).

Estimated Effective Migration Surfaces (EEMS)

We used SNP genetic distances paired with geographical distances across populations to estimate effective migration surfaces across the region of study (EEMS; Petkova et al., 2016). For assessing potential barriers to gene flow between Sky Islands and into Mexico, we included all 137 individuals and the associated high-quality SNPs that were filtered for population genetic analyses to visualize similarities between population structure and estimated gene flow. The 337 high-quality SNPs extracted for the population structure analyses were used to calculate a genetic distance matrix with the bed2diff_v1 program (Petkova et al., 2016). Demes were calculated in a habitat grid that extended with a buffer range around the sampled populations not only to avoid border effects for the Markov chain simulation, but also to provide insight into potential migration surfaces into Mexico where the species distribution and genetic diversity is less well understood. We set the input parameter value for the runeems_snps function (Petkova et al., 2016) to calculate migration through 300 and 400 demes across the range and ran the EEMS Markov chain Monte Carlo algorithm with 1 million burn-in iterations and 12 million sampling iterations in attempts to converge to a stationary probability distribution of migration. We inspected the results after multiple runs and adjusted parameters based on convergence success after each run (Herman et al 2022). We then used the reemsplots2 package in R to visualize the results of the program, plotting effective migration rates (m) and effective genetic diversity rates (Q) on a $\log_{10}$ scale after inspection of correlation between observed versus expected genetic dissimilarity (Jones et al 2021).

Adaptation - Isolation by distance and environment

Topographical, climatic, and edaphic variables were combined into a site-level matrix in which values were averaged on the subpopulation level and used to visualize environmental influences in multivariate space based on principal components. Due to the high levels of

multicollinearity between bioclimatic variables, a variance inflation factor analysis using the CAR package (Fox & Weisberg 2019) was used to remove factors (layers) with high colinearity (VIF > 5) (Jiang et al., 2019). This resulted in a combination of four topographic and climatic variables to visualize clusters of climatic influence across groups. These factors included climatic factors, annual mean temperature and isothermality, and topographic variables, slope and elevation.

In addition to an RDA, we assessed the influence of geographic and environmental isolation on the considered populations. The combination of four environmental variables was used to produce environmental distance matrices using a Canberra transformation, while geographic distances among populations were calculated Haversine distances (Jiang et al 2019, Faske et al 2021). To associate genomic data with these landscape distances, we used pairwise Nei's Diversity and assessed isolation-by-distance (IBD, Wright 1943) and isolation-by-environment (IBE, Wand & Bradurd, 2014) using Mantel tests (Mantel, 1967).

Adaptation - Redundancy analysis using SNP data environmental predictors

We combined spatial data, site-level data, and genomic data to investigate whether and to what extent population genetic structure of *P. imberbis* populations was predicted by climate and environmental factors. We extracted climate and environmental factors from 1-m digital elevation model datasets (USGS 2021) overlaid on point features of individual *P. imberbis* locations. Following coordinate projection corrections (WGS 1984), we used the raster calculator tool in ArcGis Pro to calculate slope and aspect values at the 1m level for all plants tracked from 2021-2023 in wild populations used for seed sourcing and genetic sampling. Additionally, we imported WorldClim's 19 bioclimatic variable raster layers for the years 1970-2000 and performed the same value by point feature extraction for every plant (Fick & Hijmans, 2017). The USGS SSURGO database was used to assign soil orders to each population location, and grazing allotment maps

from the Bureau of Land Management (BLM) were used to classify population locations as grazed versus non-grazed. We also collected five (10 cm depth) soil cores at each of 30 subpopulation clusters using a 2-cm diameter stainless steel soil sampler. Samples from each subpopulation location were homogenized then analyzed for pH using a Fisherbrand benchtop pH-meter. Subsamples of soil were dried, ground, then analyzed using continuous-flow isotope ratio mass spectrometry to measure total carbon and nitrogen (C:N). Dried samples were additionally weighed out to 50 g and used to calculate soil texture via particle size analysis using the hydrometer method (Gavlak et al. 2005).

We then used a Redundancy Analysis (RDA) with the R Vegan package (Oksanen et al, 2015) to examine relationships between landscape variation and population genetic structure. We used a predictive approach (Mac Nally, 2000) to attempt to maximize the genetic variance explained by selected environmental predictors (Capblancq & Forester, 2021). We began with a set of 15 biologically relevant landscape variables (selected from WorldClim and raster datasets) and used a variable selection approach using the ordiR2step forward selection function in Vegan to reduce the number of variables while retaining as much variance as possible (VIF < 2.5). Final variables used in the RDA were annual mean temperature (°C), precipitation seasonality (%), slope (°), soil order, percent sand (%), percent silt (%), ecological response unit type (grassland or woodland), and grazing allotment type (grazing or no grazing). Following variable selection and constrained ordination, we used an ANOVA permutation test with 1000 steps to assess the significance of the global model. Finally, we employed variance partitioning to quantify the amount of variation attributed to each variable, partitioning climate, habitat, and soil variables separately. We plotted genetic relationships between grouped populations ("meta-populations") in multivariate space as predicted by the most suitable landscape variables as discerned by the variance partitioning and permutation tests. Finally, we multiplied the total variance explained by the RDA

by the average F_{ST} value calculated from SNP frequencies to estimate an overall percentage of genetic variation explained by the selected environmental variables (Meirmans 2023).

Garden survival and growth across populations

We used Generalized Linear Models (GLMs) to test for differences in performance metrics among populations, specifying 'Population' as a fixed effect and adjusting the error term to correspond with response variable data type (i.e., survival = binomial distribution; growth = Gaussian distribution). To explore fitness patterns, we related vital rate information for each population to the difference between the garden site and the home-site of the population for key climate variables using regression analysis.

Mapping gene flow and habitat suitability to identify steppingstone restoration sites

Following calculations of estimated effective migration surfaces and species distribution models with bioclimatic variables, we matched raster spatial references in ArcGIS pro and ran the Raster Calculator tool to visualize gene flow corridors. We produced maps in which areas of low log(migration) rates were prioritized and multiplied by SDM suitability values (*methods for development of SDM described above*).

We then extracted the ensemble SDM at values of 800 or higher (equating to 80% suitability or higher) and used the Raster to Polygon tool to create a homogenous layer for 80-100% suitability. We then selected the map of low migration rates, upon which we layered the ensemble SDM polygon. Overtop of these layers, we added in points for the four primary populations of *P. imberbis* that encircle the Huachuca Mountain range, adding in a 1-km buffer to each point to visualize potential pollinator foraging distances. Finally, we created a point feature layer and manually added interspaced points with 1-km buffers along the highest suitability surfaces at intervals of roughly 2-km to produce a map of potential "stepping stone" restoration garden

locations, which would allow for the development of a future gene flow corridor connect CNM populations to Scotia Canyon and O'Donnell Canyon populations, both of which contain higher rates of nucleotide diversity and expected heterozygosity relative to other populations.

RESULTS

Sequencing

RADSeq determination of population structure

Reduced dimensionality of SNP frequencies across individual samples demonstrated evidence of population structure across sampled populations of *P. imberbis* **(Fig. 4)**. In multivariate space, the 137 individual samples and 337 quality-filtered SNPs displayed evidence of three distinct population clusters: Coronado National Memorial, the western edge of the Huachuca Mountains (with Peña Blanca included as an outlier), and the Santa Rita (WASP and MCCM1) populations ($F = 98.718$, $p < 0.001$). Dispersion between k-mean clusters was not significantly different, passing assumptions for a PERMANOVA test. A PCA highlighting population means showed distinct separation between CNM and Santa Rita **(Fig. 5),** and when rotated on its axis, the PCA of genetic relationships reflected the physical geography of the sampled populations (**Fig. 6,7).** For population genetic diversity estimates, pairwise FST values for 6 aggregated populations ranged from 0.077 to 0.421 and Nei's diversity coefficient ranged from 0.026 to 0.343. The global FST value for all pairwise groups were 0.21 (**Fig 8).** When calculated at the subpopulation level, CNM Blue Waterfall showed distinctly high Fst values, which was likely an artifact of low sample size compared to other populations ($n = 2$), and thus was removed from the final figure comparing Fst between populations to emphasize the high genetic dissimilarity of subpopulations outside of the memorial (**Fig. 9).** Overall, the four CNM subpopulations included in the fine-scale pairwise

Fst figure shows that CNM subpopulations and not noticeably genetic distinct from one another, with Montezuma Wash and Maintenance Shed populations demonstrating the highest within-CNM pairwise Fst value of 0.237.

Genetic diversity

Gene flow coefficients ranged from 0.343 to 2.978, with a mean of 1.21. For within population genetic diversity estimates for 17 subpopulations, expected heterozygosity ranged from 0.09 to 0.25, with a global He of 0.20 (**Table 4, 5)**. Nucleotide diversity (π) for 17 subpopulations ranged from 0.04 to 0.10, with a mean of 0.07 (**Table 4, 5)**. Nucleotide diversity was highest in Scotia Canyon when adjusted using the standard error of a linear regression model, followed by ATW and O'Donnell Canyon (**Fig. 9)**. Gene flow was highest between CNM and ATW (Nm = 2.978), whereas Scotia Canyon had lower values of 0.926 between ATW and 0.903 between CNM **(Table 5)**. At the same time, Scotia Canyon demonstrated the highest inbreeding coefficient (0.09) whereas CNM and Pena Blanca had the lowest (0.06 & 0.02, respectively) (**Table 5)**.

Estimated Effective Migration Surfaces

The EEMS resulted in positive relationships between expected and observed genetic dissimilarity in simulations with both 300 and 400 predefined demes, with runs of 400 demes resulting in higher R^2 values for dissimilarities between pairs of sampled demes, indicating a stronger fit for observed versus predicted values for the 400 deme models (Jones et al. 2021, Petkova et al. 2016) (**Table 7, 8; Fig. 10)**. MCMC simulations of migration rates and genetic diversity through EEMS mirrored the geography of the region and demographic patterns of the species, with low effective migration surfaces coinciding with Sky Island edges (**Fig. 11)**. Specifically, migration resistance was evident on the western edge of the Huachuca Mountains and through the Patagonia Mountains.

The pattern of relatively low effective migration coincided with locations of populations thought to be extirpated **(Fig. 12)**, which suggests that this region is less conducive to dispersal and establishment. Scotia and O'Donnell Canyon sites, where populations are declining, appeared in the center of the low migration pathway. Simulations with the courser 300 deme parameter showed greater migration connectivity between Coronado National Memorial and the neighboring Anne Tank Wash populations, but increasing the deme count to 400 showed a distinct reduction in log(m) coinciding with the geographic feature of Montezuma pass **(Fig. 12).** Genetic diversity calculations from EEMS showed a distinct limitation of Q diversity rates in the most distant, isolated population of Peña Blanca. Populations across the Huachuca Mountains to the Santa Rita Mountains showed similar rates of diversity, with Santa Rita and O'Donnell Canyon populations demonstrating higher diversity than the larger populations of Anne Tank Wash and CNM **(Fig. 13, 14).** When multiplied by the SDM surface produced from the bioclimatic variables, two maps were produced emphasizing that the western edge Huachuca mountains leading into the Santa Ritas has high overlap in low migration rates and highly suitable habitat, whereas the eastern edge of the Huachucas is an important region with high migration likelihood overlapping high habitat suitability (**Fig. 12).** When compared to the USGS ensemble model and overlaid with existing populations and simulated restoration steppingstone garden sites, it would require approximately 8 sites to connect Scotia Canyon with CNM populations, given a 1 km pollinator flight buffer. Connecting O'Donnell Canyon to CNM would require a wider buffer and a minimum of 4 steppingstone populations (**Fig. 15).**

Adaptation

Adaptation - Genetic variation

The constrained ordination approach (RDA) showed genetic variation in the context of environmental variables (**Fig. 17)**. The VIF was less than 2.5 for all variables. The first RDA axis accounted for 25.07% of the variation, the second accounted for 15.56% of the variation, and the third accounts for 14.43% of the variation. Together, the first three axes account for 55.06% of the normalized genetic variation (Ji et al., 2022). All variables used in final RDA explained 10.60% of the variation. Partitioning the variance by climate variables (precipitation seasonality and annual mean temperature) explained 2.9% of the variation, habitat variables (slope, grazing pressure, and ERU) explained 4.4%, and soil variables (percent sand, percent silt, and soil order) explained 3.07%. The total proportion of constrained variance explained by the model was 10.66% ($F < 0.001$). Multiplied by the average F_{ST} value of 0.21, the model, which included an array of 8 climatic and habitat variables, explains 2.24% of the overall genetic variation determined by SNP frequencies. The mantel tests did not result in significant isolation by distance or environment (**Fig. 16),** but the RDA provides finer resolution when assessing the effect of isolation and environment on genetic dissimilarity.

Adaptation - Common gardens

Of 5,036 seeds sourced from 25 subpopulations observed in the germination chamber over 3 months, a total of 435 seeds germinated (8.6% germination rate overall). Germination rates by subpopulation ranged from 0% to 30%, with the "Maintenance Shed" population from CNM resulting in the highest germination rate of 30% (**Table 4, 5**).

A total of 174 germinants that were transplanted from the germination chamber into cones survived past the initial transplant phase. Surviving transplants at the time of transport to CNM represented 18 subpopulations. Of the174 germinants that were transported to CNM and planted in the common garden, a total of 65 survived the first 8 weeks, with an overall survival rate of 37% for the same season. One year later, only 19 plants had survived, with an overall survival rate of 10.9%. Overall, individuals from the Northern Huachucas (Scotia Canyon and O'donnell Canyon) had the highest survival rates (54%, total planted = 19), whereas locally sourced populations from CNM had the second highest survival rates (32%, total planted = 113) **(Table 5, Fig. 18, 19)**. More environmentally and geographically distant populations (from the Santa Rita Mountains and Ann Tank Wash complex) demonstrated the lowest survival rates, with survival rates of 28% (total planted = 9) and 22% (total planted = 33), respectively (**Table 5**). Analysis of survival and growth only included data from the first 8 weeks, as monitoring efforts were limited to this timeframe.

Results from the generalized linear mixed models did not suggest a strong home-site advantage (or strong patterns of local adaptation) (**Table 6**). The generalized linear mixed model used to predict survival rates as a response to various genetic and population predictors yielded mixed results, and the top three models presented in Table 10 retain a minimum of one significant predictor variable. Overall, the null model was the best fit for the data (AIC = 17.9) compared to the global model (AIC =23.1). Starting with the global model (7 predictor variables), the only significant predictor variable in was initial height at planting ($p < 0.05$*). The model with the second lowest AIC retained significance for initial height at planting ($p < 0.05$*) as well as environmental distances ($p < 0.05$*). In the model with the lowest AIC (excluding the null model), initial planting height was significant ($p < 0.05$*) and environmental distance was marginally significant. In each model, environmental distance had a positive relationship with survival ratio, and initial height at planting retained a negative relationship with survival ratio.

DISCUSSION

Genetic diversity and climate change in arid landscapes

Sky Islands are natural laboratories to examine the consequences of rapid climate change on endemic species, given the vulnerability of this arid region to shifting climate and the naturally fragmented habitat that mimics anthropogenic land use change, yet few studies have focused on this region (Love et al 2023). This study represents the first use of both genomic analysis and a common garden to investigate population structure and performance of an endemic rare plant in the Madrean Archipelago. Findings from this study demonstrate how the population genetic structure of the rare endemic plant *P. imberbis* is sensitive to fine-scale shifts in distribution, environmental factors, and population sizes. Moreover, the work completed here advances our understanding of the current status of this endangered species by supporting comprehensive counts, mapping initiatives and habitat assessments. Taken together, we use these data to direct conservation and management suggestions to support the recovery of *P. imberbis.*

Status of Pectis imberbis *at CORO (Objective 1)*

We counted and mapped *P. imberbis* occurring at CORO and in the known peripheral populations. We located a total of 4883 individuals within CORO and produced hand-drawn polygons (in ArcGis Pro) of approximated population "envelopes" categorized and visualized by population density. We initially defined subpopulations as groups of plants at least 150 meters from other groups or separated by other significant natural or manmade barriers. Based on this definition, we counted 63 individuals at the Blue Waterfall site, 2004 at Cave Trail, 805 at the Maintenance Shed Back site, 1081 at Maintenance Shed Front site, 57 at Montezuma Wash, and 873 at West Joe's Canyon.

Overall, the results of the genomic analysis suggest that there is distinct population

structure among subpopulations of *P. imberbis*, indicating that the small remaining range (100 km) of the monitored populations is variable enough to hinder gene flow. However, whether between-population genetic variation is due to geographic separation (genetic drift), differences in local adaptation, or population size is less clear. Results from the k-means clustering analysis reflected the geographic separation of at least two of these regions, demonstrating significant differences between populations from the Huachuca Mountains and the Santa Rita Mountains. Moreover, subtle differences in population spatial structure and size could be a driver of within-population diversity at finer scales. At coarser scales, results from the redundancy analysis of environmental variables suggest that slope and precipitation seasonality are two strong predictors of genetic clustering. Environmental distances and genetic groups did not significantly predict performance in the common garden, but non-local individuals demonstrated marginally higher survival rates, contrary to the home-site advantage hypothesis. However, given relatively higher germination rates, survival rates, expected heterozygosity values, and a relatively lower inbreeding coefficient, we suggest that the Maintenance Shed Backside population in CORO is a genetically and reproductively robust population with high potential for seed sourcing and out-planting in future restoration efforts, and should be considered as a source of genetic material to outsource to steppingstone gardens.

Pectis imberbis habitat (Objective 2) – See Appendix A

Seed and foliar sample collection (Objective 3)

We successfully collected both seeds and foliar material across *P. imberbis* populations. Given the distinct genetic makeup of several peripheral populations (ie. Santa Rita, O'Donnell Canyon, and Scotia Canyon), seed collections should target these populations before environmental conditions, isolation and other factors reduce seed production further. Alternatively, populations

within Coronado, especially those with both robust population sizes and relatively high heterozygosity (Maintenance Shed and Lower Cave Trail) should be prioritized as seed sources for enhancing genetic connectivity outside of the memorial. These conclusions should be considered carefully, and actions should consider revision over time given fluctuating climate conditions and habitat loss. Notably, the genome of *P. imberbis* is large (appx. 13m bp), rendering genetic analysis with current prices not only expensive and computationally intensive, but also subject to statistical biases due to the nature of sequencing an unusually large genome.

Patterns of genetic variation (Objective 4)

CNM populations were unexpectedly low in both nucleotide diversity and expected heterozygosity, despite their size. Inbreeding coefficients produced from Stacks indicate that aside from Pena Blanca (whose values are likely by small samples sizes), CNM populations had relatively lower average inbreeding coefficient (Fis) values compared to Anne Tank Wash, Scotia Canyon, O'Donnell Canyon, and the Santa Rita populations. Estimates of genetic diversity indicate that relative to population size, nucleotide diversity was highest in Scotia Canyon when adjusted using the standard error of a linear regression model. O'Donnell Canyon, on the other hand, showed higher levels of expected heterozygosity when similarly adjusted. Scotia Canyon also showed higher levels of inbreeding, perhaps indicating some degree of genetic drift, which is supported by the lower migration values on the EEMS surface**.** In the case of Scotia Canyon, whose population growth rates have been consistently below replacement, a combination of high inbreeding and high nucleotide diversity may show that this population was historically connected to populations in and around CNM but have diverged through isolation and selfing over many generations. Inbreeding may be a strategy the species uses to counteract isolation, thus the existence of high inbreeding values in conjunction with low replacement levels may indicate

inbreeding depression, while higher nucleotide diversity may be a relic of connectivity to more genetically diverse populations. Both Scotia Canyon and O'Donnell Canyon persist to the northwest of CNM populations, with patches of suitable habitat but low migration rates connecting them to the memorial. Given the higher genetic diversity of Scotia and O'Donnell Canyon, an increase in connectivity may facilitate an influx, or restoration, of genetic diversity to CNM populations in the future, buffering the populations from a slow decline in genetic diversity.

Environmental predictors of success

Given evidence of population structure across sampled populations of *P. imberbis*, the investigation into how environmental factors drive population structure demonstrates how climate and topographical variables contribute to genetic clustering. The constrained ordination of environmental variables significantly predicted population genetic variation as a response to climatic and topographic variability. However, the landscape factors used in the RDA explained only 2.24% of the genetic variation amongst populations as determined by SNP frequencies, indicating the model would benefit from more fine scale abiotic and biotic data. Overall, the RDA reflects and further delineates clustering patterns determined by the PCA and k-means clustering designations. As compared to a Mantel IBD/IBE test in this system, the RDA effectively detects various landscape influences on genetic variation, albeit with coarse resolution (Capblancq & Forester, 2021). Specifically, the model indicates a moderate influence from precipitation seasonality, slope, soil type, and presence of grazing. The directionality of the ordinated linear regression arrows estimates the impact of these environmental factors, providing insight into potential adaptive variation across groups (Ji et al 2022). Take precipitation seasonality (a measure of variance) as an example – variable precipitation patterns may be acting as a selective gradient across populations (Capblancq et al., 2018), in which CNM populations are experiencing slightly

less drastic shifts in precipitation perhaps due to regulating effects of the valley they grow in. Conversely, O'Donnell Canyon and Scotia Canyon populations persist in more exposed habitats on the western face of the Huachuca Mountains, where monsoonal rain patterns may be more variable. As a result, selective pressures on populations outside of CNM may favor offspring that are able to survive more drastic fluctuations in seasonal precipitation or temperature. In another example, slope and grazing may have interacting effects, in which populations that grow on steeper slopes may be protected from grazing.

The variables used in the RDA were selected due to their relevance to climate-change metrics (in addition to their filtering through covariance and VIF thresholds). Climate-change related factors such as drought, increasing temperatures, shifting vegetation communities, excessive erosion due to flooding, and habitat loss are some of many threats that may affect *P. imberbis* across its entire range (USFWS 2023). Considering the genetic divergence of the monitored populations, as evidenced by high F_{ST} and low gene flow values, local adaptation may contribute to the resilience of this endangered species. Based on limited knowledge of its physiology, *P. imberbis* may be well-adapted to handle drought to a certain extent, but demographic observations have shown dramatically increased seedling output and survival during wetter years (Souther et al, 2022). For example, a molecular phylogeny of the *Pectis* genus determined that like many of the 90 species within the genus, *P. imberbis* utilizes C4 photosynthesis pathways as is common with many Xeric-adapted plants (Hansen et al 2019). This may indicate that *P. imberbis* is resilient to fluctuations in drought conditions on an individual level. Moreover, as a perennial herb, other physiological mechanisms should be considered when assessing varying capacities for local adaptation across genetic groups, including seed sets, resprout ability, and protection from browse (Souther et al 2022).

Challenges of common garden

The common garden trial demonstrated that *P. imberbis* can be successfully cultivated in a greenhouse and transplanted to the field, but achieving enough viable plants for impactful restoration requires substantial effort. Multiple stages in the process pose risks of damage or stress to the plants. The small size and fragility of the seeds make transplanting germinated individuals into cones particularly challenging. To mitigate this, direct seeding into cones and growing plants in a large growth chamber is recommended. Adult plants also exhibit structural fragility, with instability in stems increasing the likelihood of damage during transfer from the greenhouse to the transplant site.

Establishing the 10 × 10-meter garden required intensive manual labor, including fence construction, removal of invasive species, breaking compacted soil with pickaxes, digging planting holes in rocky terrain, planting cones with fragile root systems, and installing an irrigation system. Despite these efforts, the survival rate one year after planting was only 10.9% for the original seedlings that were produced from 5000 seeds in the greenhouse efforts. Scaling up greenhouse operations to improve survivorship would necessitate a significant expansion of greenhouse capacity and considerable resources for germinant monitoring.

Given these constraints, we recommend that future restoration trials incorporate an experimental design focused on direct seeding at various *in situ* restoration sites. Targeting zones of high habitat suitability and incorporating minor environmental gradients may enhance out-planting survival while reducing resource inputs. Moreover, if these sites are planned along the suggested gene flow corridor with seeds from successful and genetically diverse populations, this trial could result in the facilitation of enhanced connectivity and a continuum of genetic diversity that could enhance the adaptive capacity of future generations of the species.

Protocol for beardless chinchweed ex situ *and* in situ *conservation/restoration (Objective 5)*

Ex situ conservation and considerations for recovery planning

We suggest that there are 3 critical genetic groupings which should be protected through both conservation and restoration efforts. The results from our analysis of the restriction-site associated DNA sequencing of 6 grouped populations indicate that while there is considerable overlap in genetic similarity between all 6 populations, overall, the k-means clustering indicated significant differences between three clusters consisting of 1) Coronado National Memorial subpopulations, 2) Anne Tank Wash, Scotia Canyon, O'Donnell Canyon, and Pena Blanca subpopulations, and 3) Santa Rita subpopulations, which consists of Wasp Canyon and McCleary Canyon. Notably, due to the loss of samples during sequencing, we could not examine patterns of genetic diversity within the CNM at the resolution planned in our initial sampling scheme, though we did manage to capture reliable information on four of the largest subpopulations. Applying Nei's Genetic Distance limits for distinguishing sub-populations (0.05 – 0.15) and populations (> 0.15), or Wright's definition of Fst > 0.25 as "very high differentiation," we identify Pena Blanca and Scotia Canyon as the most distinct populations on average, with Anne Tank Wash representing the highest degree of among-subpopulation distinctiveness. According to Wright's classical definition of fixation index (Fst), every population sampled would be considered highly differentiated from one another. However, a more cautious interpretation would be to assess differentiation at the relative scale of the study organism (Bhatia et al, 2013). If we use this approach, subpopulations within CNM are relatively similar to one another, with plants within the Montezuma Wash subpopulation showing the highest degree of distinction within the Memorial. Assuming isolation by distance does exist to some extent despite lack of clear statistical evidence, we see that Pena Blanca falls in line with

the assumption that smaller, more geographically distant populations are more genetically distinct.

Referencing the clustering PCA results, these groupings are based entirely on overlapping loci at neutral genetic markers, and do not consider differences in local adaptation or survivorship capacities across groups. Our investigations into local adaptation did not produce clear evidence of local adaptation, suggesting that perhaps despite genetic dissimilarities, *P. imberbis* distributions may be driven primarily by strict habitat requirements rather than genetic differences. As such, we suggest the risks of introducing novel genetic material across populations is outweighed by the benefit of facilitating a continuum of genetic diversity. Specifically, we suggest implementing trials of a gene flow "corridor" that will connect CNM populations to populations at the western and northern range of the Huachuca Mountains through a series of directly seeded, steppingstone restoration gardens. We developed a map to demonstrate this concept, but further discussion of funding, land ownership, and natural resources is necessary to determine the most feasible locations (within high suitability SDM polygons) for such steppingstone gardens.

In situ conservation planning

Habitat mapping indicates that habitat suitability is diminishing as a function of climate change. One possibility is to remove co-occurring stressors to increase resilience of this species. Known stressors include herbivory and competition. Fencing to exclude herbivores including native ungulates may be warranted. Additionally, methods of reducing co-occurring vegetation such as prescribed burns or invasive species pulls may also improve *P. imberbis* performance.

As climate changes in the Sky Islands, species are challenged to track the climatic conditions to which they are adapted, once species are 'stranded' at the peaks of within this

mountain range. Our maps of shifts in habitat suitability indicate that suitable habitat has shifted northward towards areas that likely have different soils, slopes, and other abiotic characteristics, relative to current *P. imberbis* habitat. Methods of accelerating evolutionary adaptation, such as the approach described above to increase connectivity, or by directly enhancing gene flow *via* the introduction of individuals from other populations to increase genetic diversity or warm- or dry-adapted genotypes, may increase the likelihood of persistence.

We found high levels of genetic isolation of populations at the edge of *P. imberbis*'s current range. A cross-population demographic study of *P. imberbis* found that population growth rates of these populations are below replacement levels and speculated that they are on a trajectory to extirpation, in part driven by rarity which reinforces population decline (Souther et al. 2022). One possible solution is to increase connectivity across populations, possibly by reintroducing populations in suitable habitat across the species' range. Increasing connectivity could increase the likelihood of genetic and demographic rescue, critical in this highly heterogeneous environment, characterized by disturbance events like drought that can result in stochastic population loss. However, it is unclear whether reestablished populations would be sufficient to bolster cross-population gene flow in the continental "islands" of the Madrean archipelago. The SDM presented here, indicated that of the suite of 43 predictor variables and 11 hectares used to train the ensemble models, the final model predictions identified 5 key predictor variables (spring precipitation, elevation, solar radiation, EPA ecoregion, and USGS/SWRG data) and a potential suitable habitat extent of 35,505 hectares. When compared to the EEMS migration surfaces produced in this study, a narrow band of suitable habitat overlaps a region of what is determined to be an area of historically limited migration. One approach to enhancing genetic connectivity could be to establish restoration garden sites along intervals of suitable habitat to allow for pollen transfer between Scotia and Anne Tank Wash populations. Unexpectedly, the former was shown to have higher

nucleotide diversity, suggesting its potentially higher adaptive capacity despite inbreeding. If Scotia has been undergoing genetic drift, this may provide it with genetic input from the largest populations near CNM.

An alternative approach would be to directly introduce individuals or pollen from outside populations to increase genetic diversity, but such interventions are not without risk. If genetic signatures reflect underlying adaptations to local conditions, introduction of non-local genotypes may result in outbreeding depression. By defining population clusters and barriers to gene flow this study revealed broad gene flow patterns, yet it does not determine the influence of local adaptation on survival – a critical next step for conservation planning. As a counterpoint to reestablishing new populations, rapid climate change may render formerly suitable habitats maladaptive (Souther & McGraw, 2011, Souther et al. 2012). Thus, interventions, such as increasing connectivity and or introduction of propagules from other populations may be warranted even when adaptive differentiation is detected. In cases where local conditions are changing, the benefits of greater genetic diversity and population numbers outweigh the risks of outbreeding depression (Souther & McGraw 2014).

CONCLUSIONS

This study described patterns of gene flow and isolation of *P. imberbis,* an endangered plant species in the Sky Island Region. We found evidence of three distinct population clusters across the populations we sampled, but the degree to which these groups are influenced by genetic drift or local adaptation is less clear. Additionally, we identified a distinct reduction of estimated effective migration rates, which we used as an indicator of gene flow, from the largest population at Coronado National Memorial (CNM) to all other populations. Populations in both the Huachuca

and Santa Rita Mountains had higher genetic diversity than that of the more distant, isolated Peña Blanca population. The calculations of nucleotide diversity, expected heterozygosity, and estimates of Q diversity from EEMS supported this finding. Future studies should attempt to corroborate gene flow and migration patterns directly through observations of seed dispersal strategies.

Tests for adaptive differentiation were limited due to low germination rates of *P. imberbis*. A reciprocal transplant experiment would bolster understanding of local adaptation across populations, as well as the role of inbreeding depression in influencing phenotypic variation and fitness – a necessity to fully understand the risks and benefits of establishing new populations outside of current population sites. Patterns of genetic isolation and population decline documented here serves as an early warning of the threats facing *P. imberbis* and other species in the Madrean Archipelago as anthropogenic change progresses. In this naturally fragmented habitat, as connectivity erodes, the performance and adaptive capacity of this species—and many others within these biologically rich desert islands—are at risk of collapse. Identifying conservation pathways to maintain connectivity and support demographic and genetic resilience, is critical for preserving the diversity of this unique landscape.

FIGURES

Figure 1: Population locations

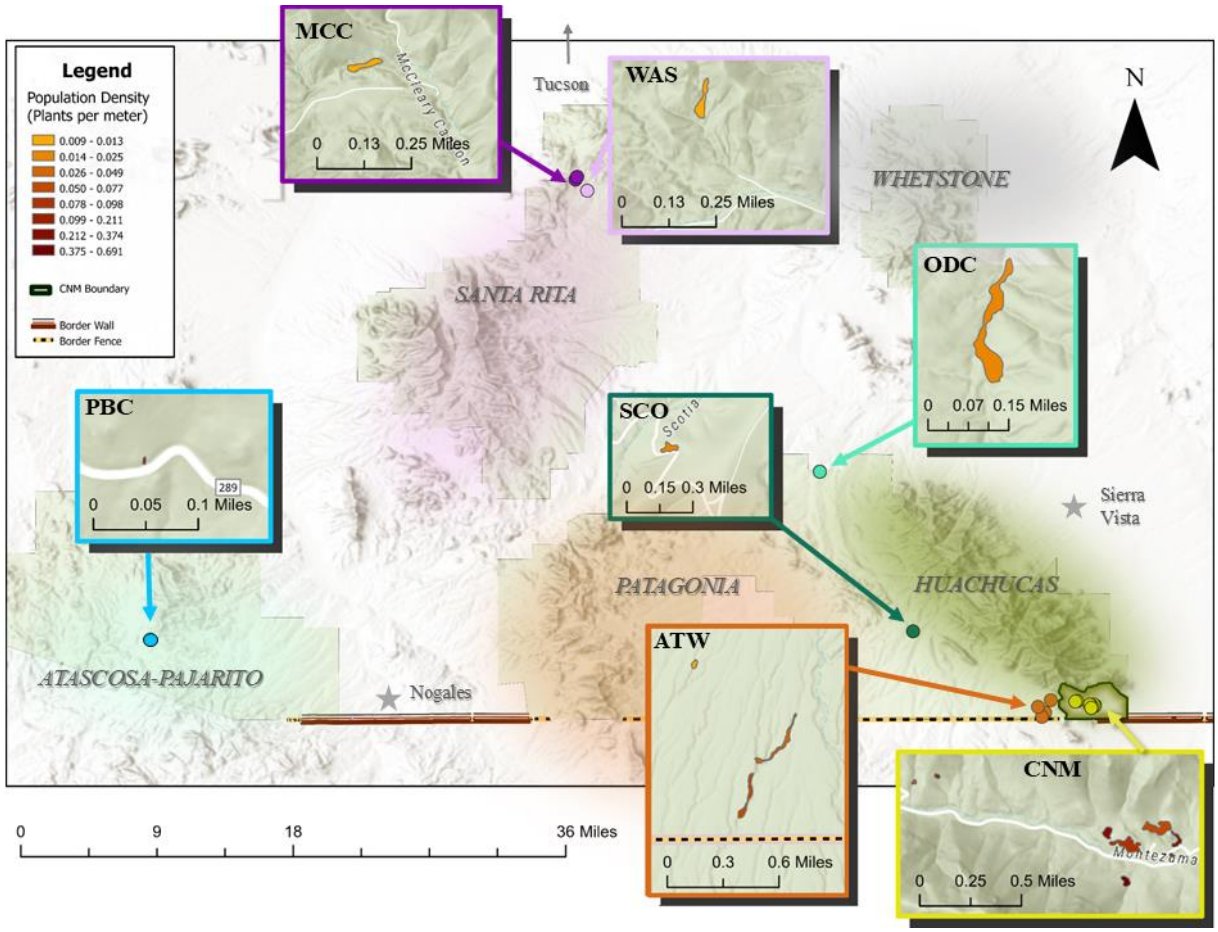


Figure 1: Detailed overview of all populations considered in this study. Boxes are color coded with similar colors used for figures in genetic analyses. Sky Island ranges are highlighted in individual colors to emphasize gaps and isolation between populations. Polygons are drawn around population clusters and the area of each polygon was used to estimate population density (plants per meter). Extirpated populations are not shown on this figure.

Figure 2: Population sizes

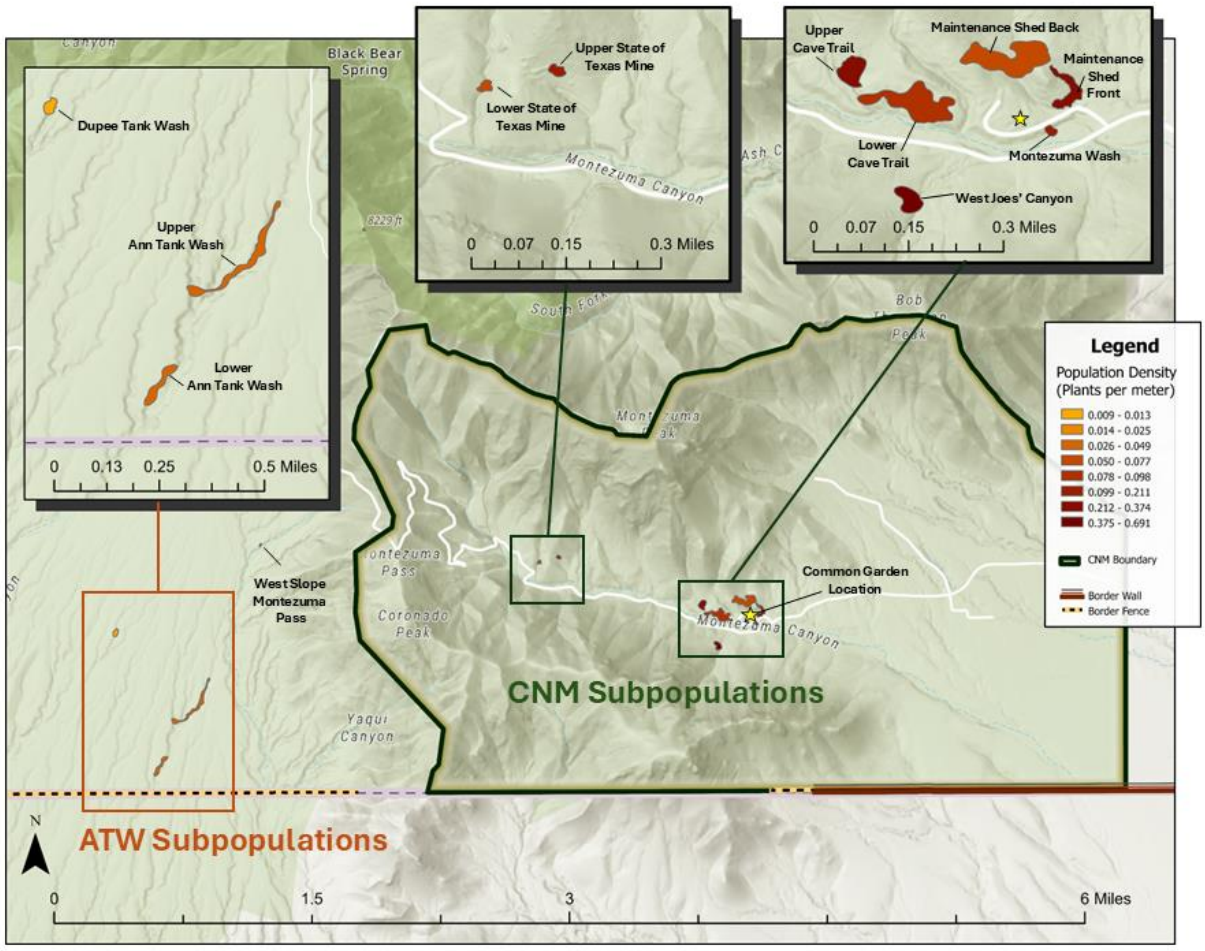


Figure 2: High detail of CNM and Ann Tank Wash *P. imberbis* populations, emphasizing polygons for each group with colors indicating population density within each polygon.

Figure 3: Common garden seed source and structure

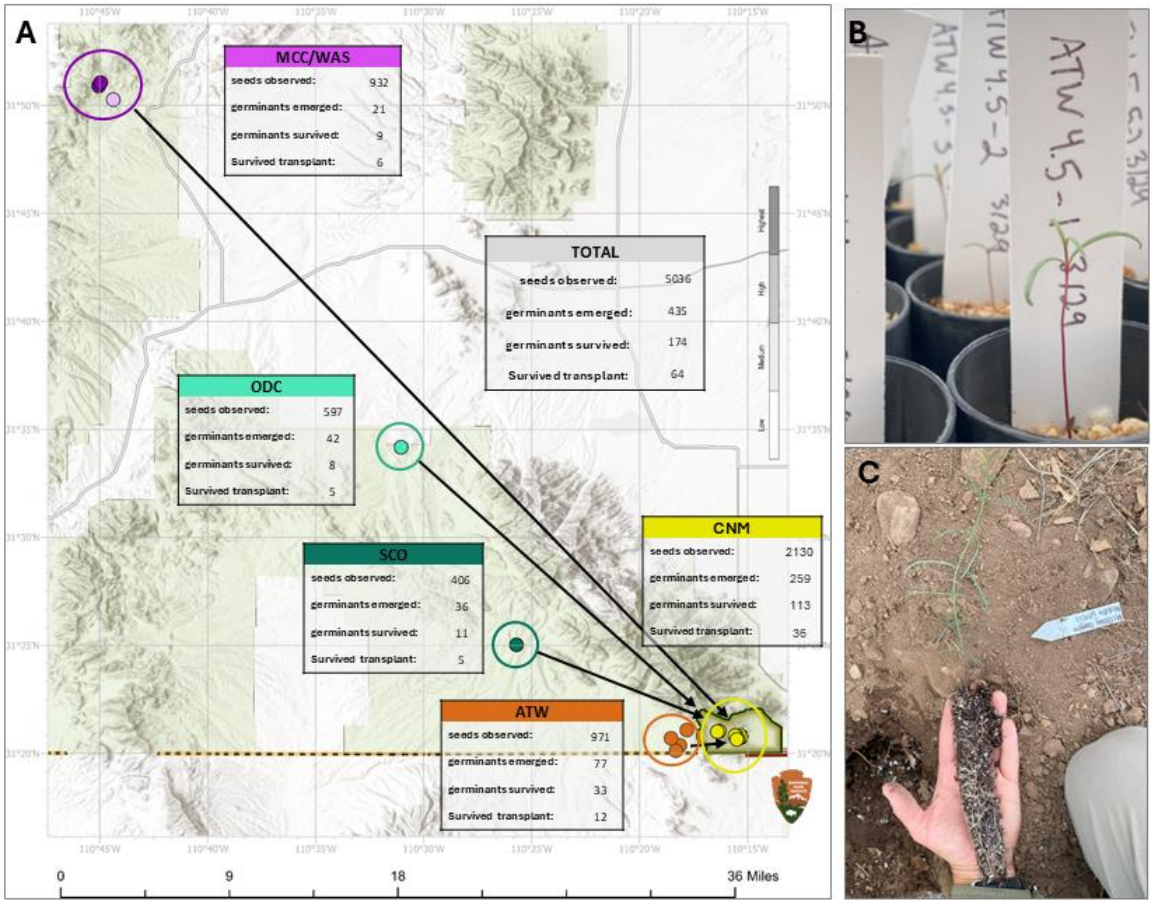


Figure 3: A) Map of seed sources for common garden experiment, planted in Coronado National Memorial. Nested tables provide an overview of germination and transplant data. "Survived transplant" indicates how many individual *P. imberbis* plants were planted in the garden site and survived until the first revisit two weeks later. Populations at Peña Blanca (PB) were too small to collect seeds from without risking interference to species reproduction. Longest distance between populations (CNM to MCC) is approximately 72 km. For purposes of common garden performance analysis, all populations outside of CNM were considered non-local, whereas all populations within CNM were considered local. B) Photo of germinants in greenhouse that were successfully transplanted into a growing cone. C) Photo of an individual that survived transfer from germination chamber to growing cone and eventually made it to the common garden.

Figure 4: RadSEQ sampling

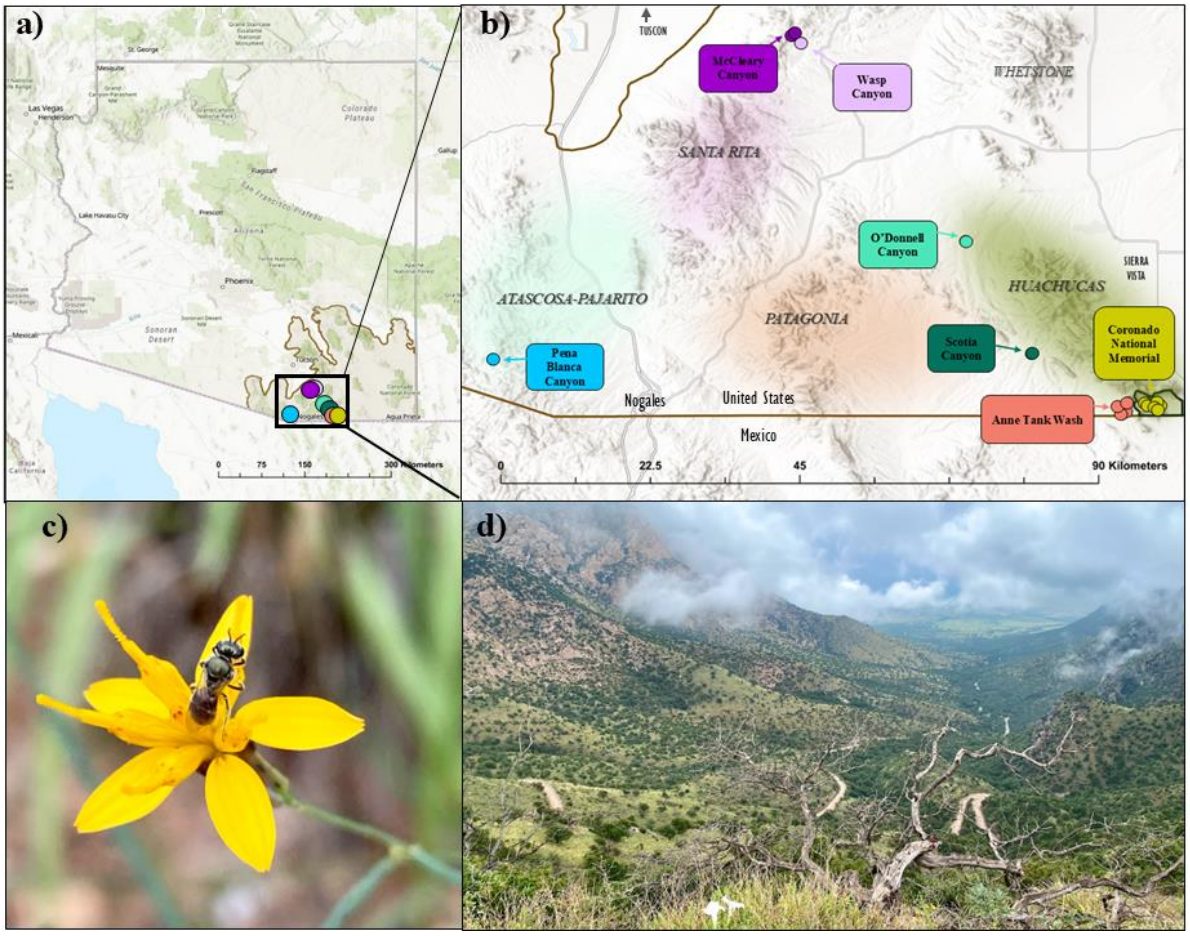


Figure 4: a) Map of populations sampled for SNP RadSEQ analysis. b) Overview of sampled populations with respect to individual Sky Island ranges, which are highlighted in different colors.

Figure 5: PCA with k-means genetic clusters

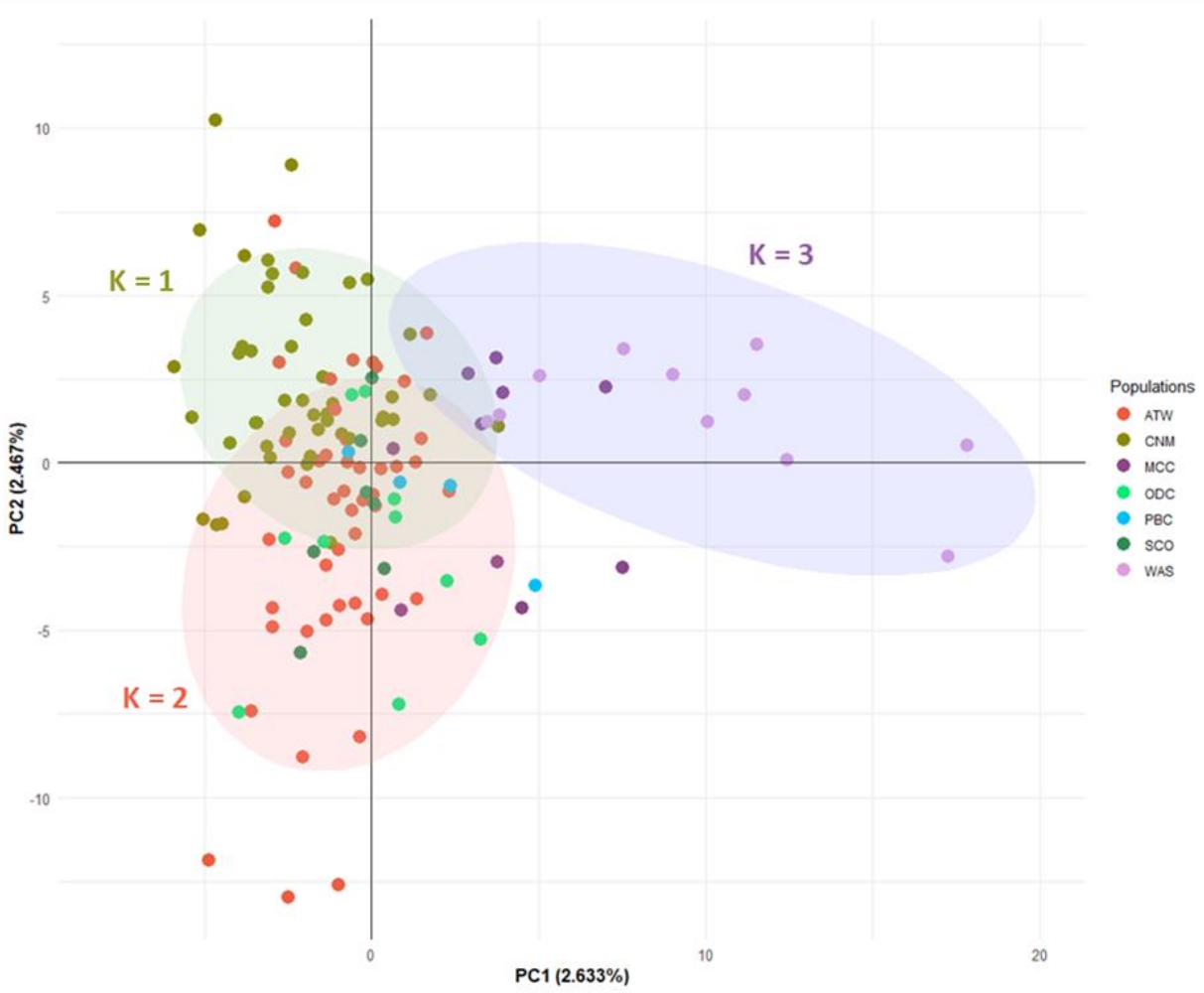


Figure 5: PCA with k-means genetic clusters derived from genetic (SNP) variation. Permanova results provide evidence of significant differences between k-means clusters. Population genetic clusters broadly reflect geographic distinctions between populations (Figure 1).

Figure 6: PCA SNP genetic differences

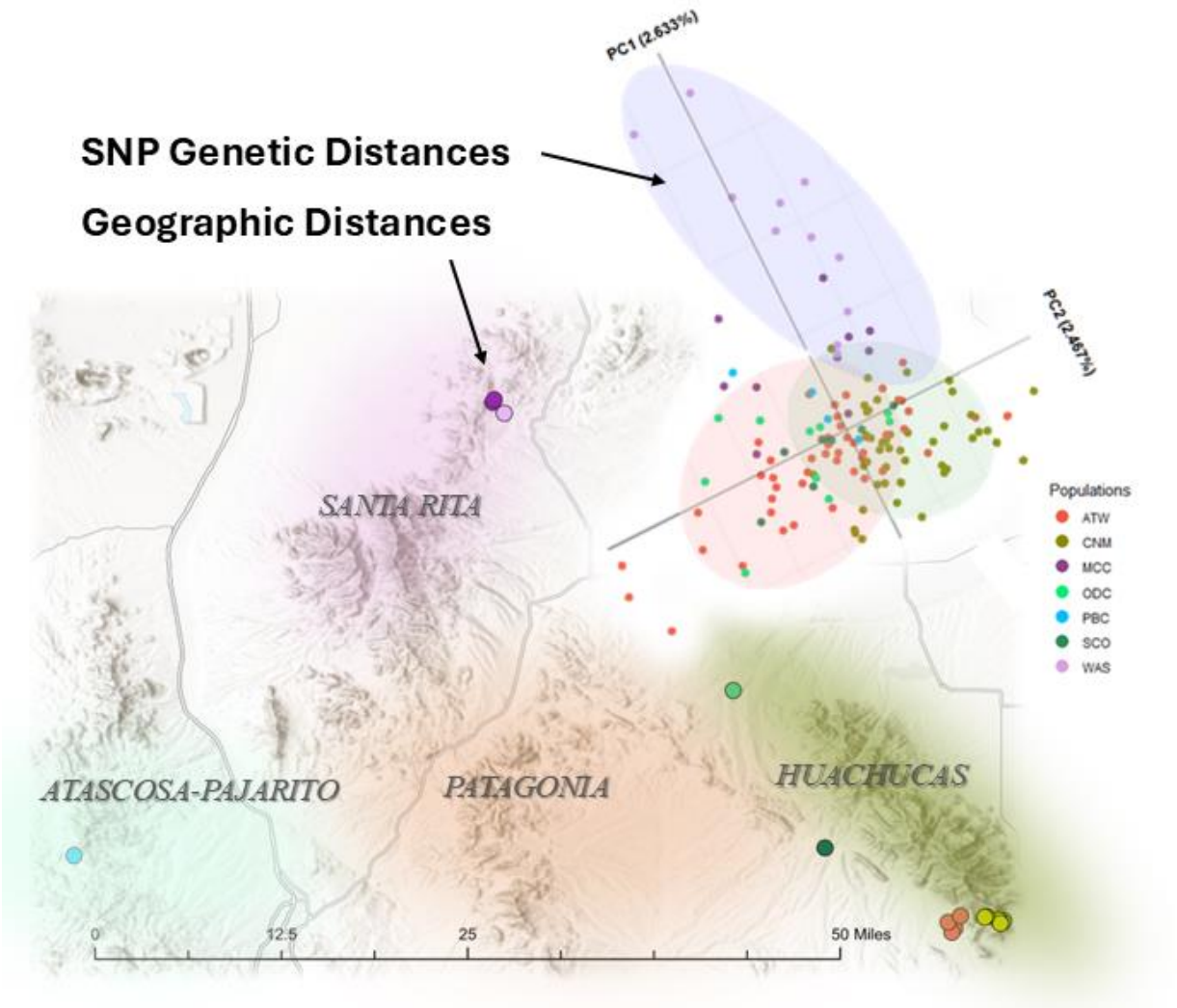


Figure 6: PCA of SNP genetic distances rotated and overlaid onto map of geographic differences between populations. Genetic distances mirror geography of sampled populations. Points on the rotated SNP PCA represent individual samples but are color-coded to match the populations shown on the map. Populations on western edge of the Huachuca mountains demonstrate dissimilarity from the nearby CNM populations.

Figure 7: Firework plots of genetic dissimilarity

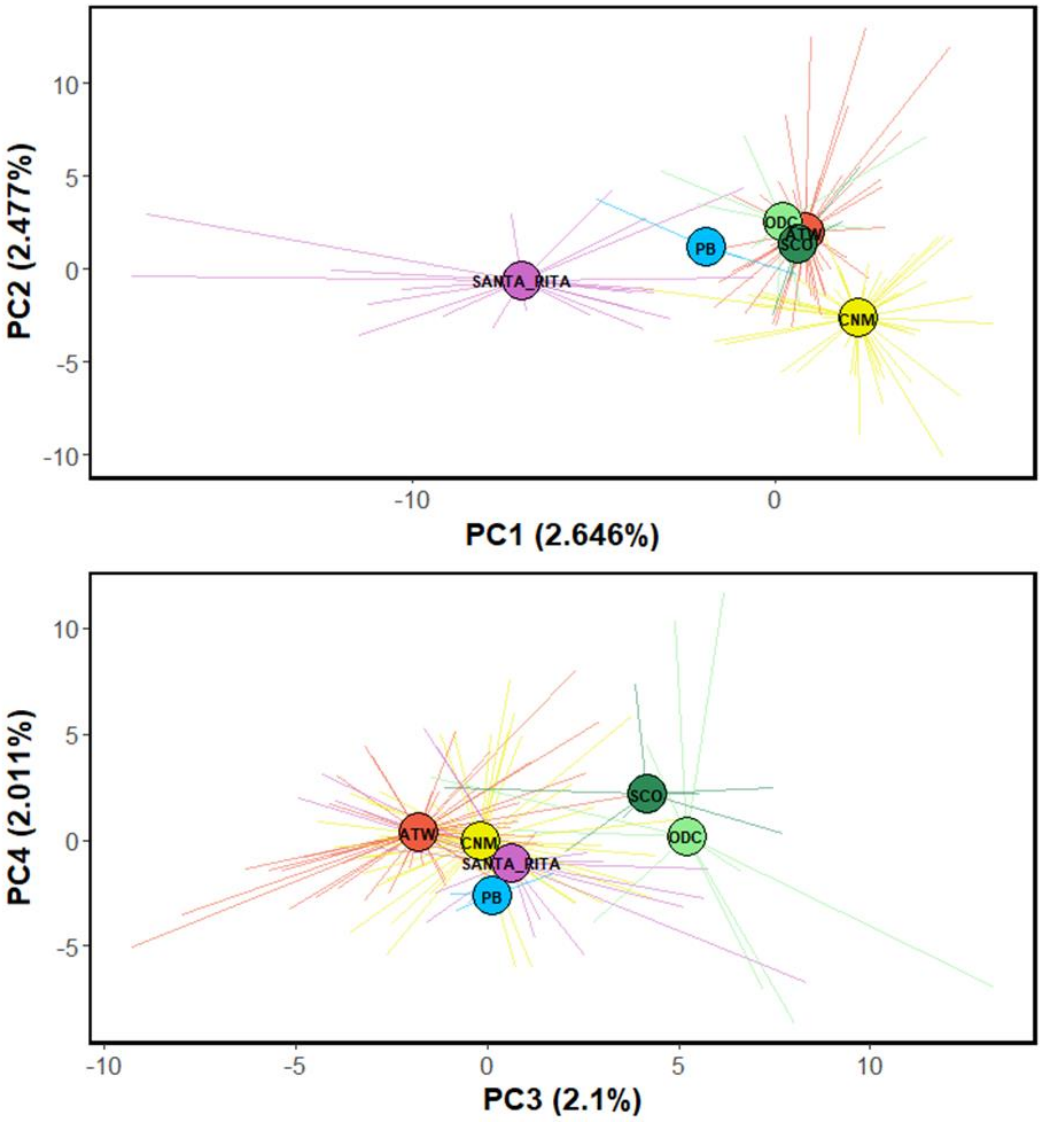


Figure 7: Firework plots of first 4 PCA axes showing weighted means of genetic dissimilarity for each population in multivariate space. Santa Rita includes Wasp Canyon and McCleary Canyon. Similar clustering patterns emerge as compared to figure 7, in which CNM is largely its own genetic group, followed by a 2nd cluster containing Scotia Canyon, Anne Tank Wash, O'Donnell Canyon, and Pena Blanca, followed by a 3rd cluster largely consisting of Santa Rita populations.

Figure 8: Fst and Nei's genetic diversity pairwise comparison

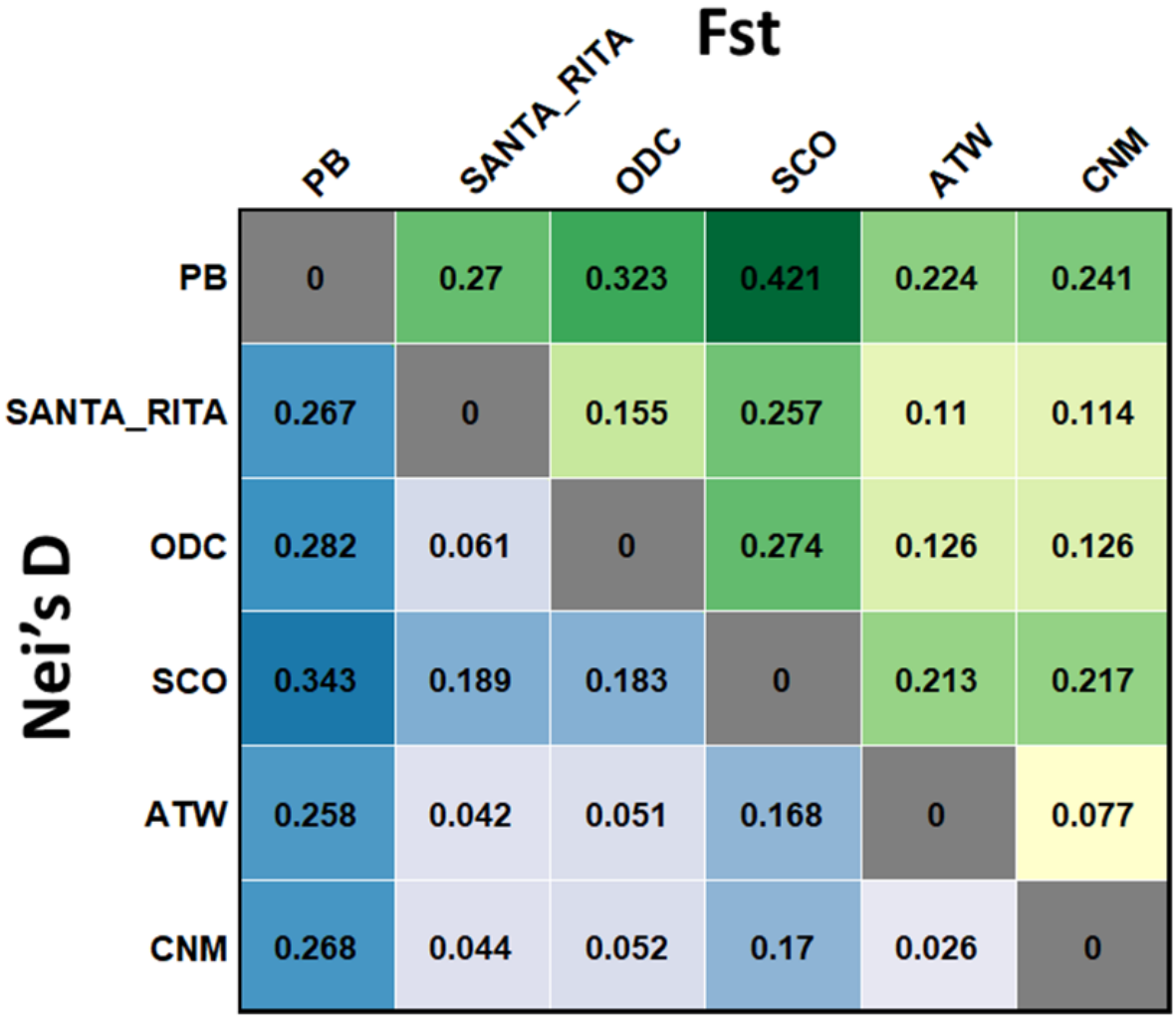


Figure 8: Averages Fst and Nei's genetic diversity pairwise plot for each of the 6 sampled populations in genetic analysis. Scotia Canyon is emphasized in dark green with the highest pairwise genetic diversity when compared to Pena Blanca, but also as a whole, with surprising genetic dissimilarity from ATW and CNM despite its proximity.

Figure 9: Pairwise Fst calculations

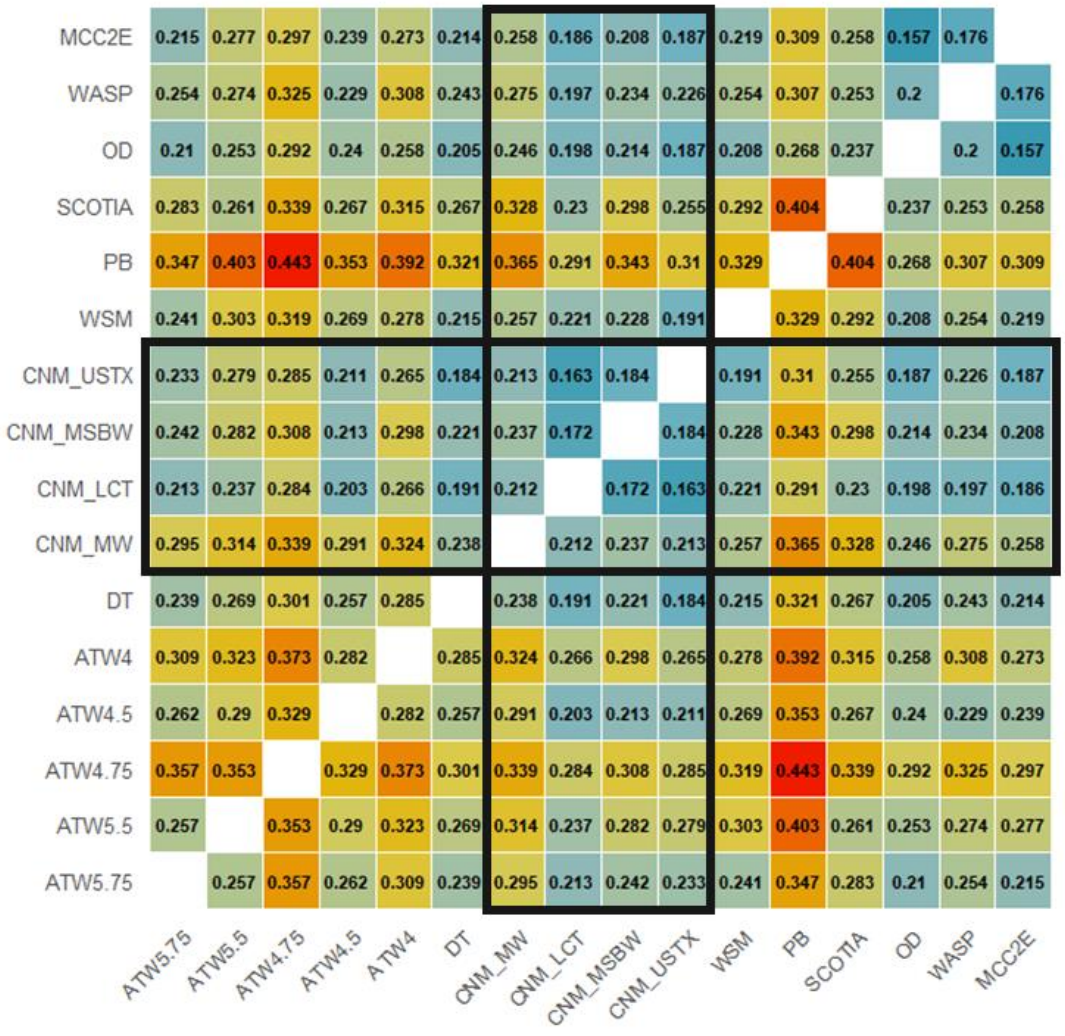


Figure 9: Pairwise Fst values calculated by subpopulation with four CNM subpopulations emphasized in black outlines. CNM Blue Waterfall was removed from this calculation of Fst values due to low sample size, which led to potentially skewed estimates of genetic differentiation. When included, CNM Blue Waterfall showed high genetic differentiation, when excluded, values of genetic relationships were more plausible within and between CNM subpopulations.

Figure 10: Effective migration rates

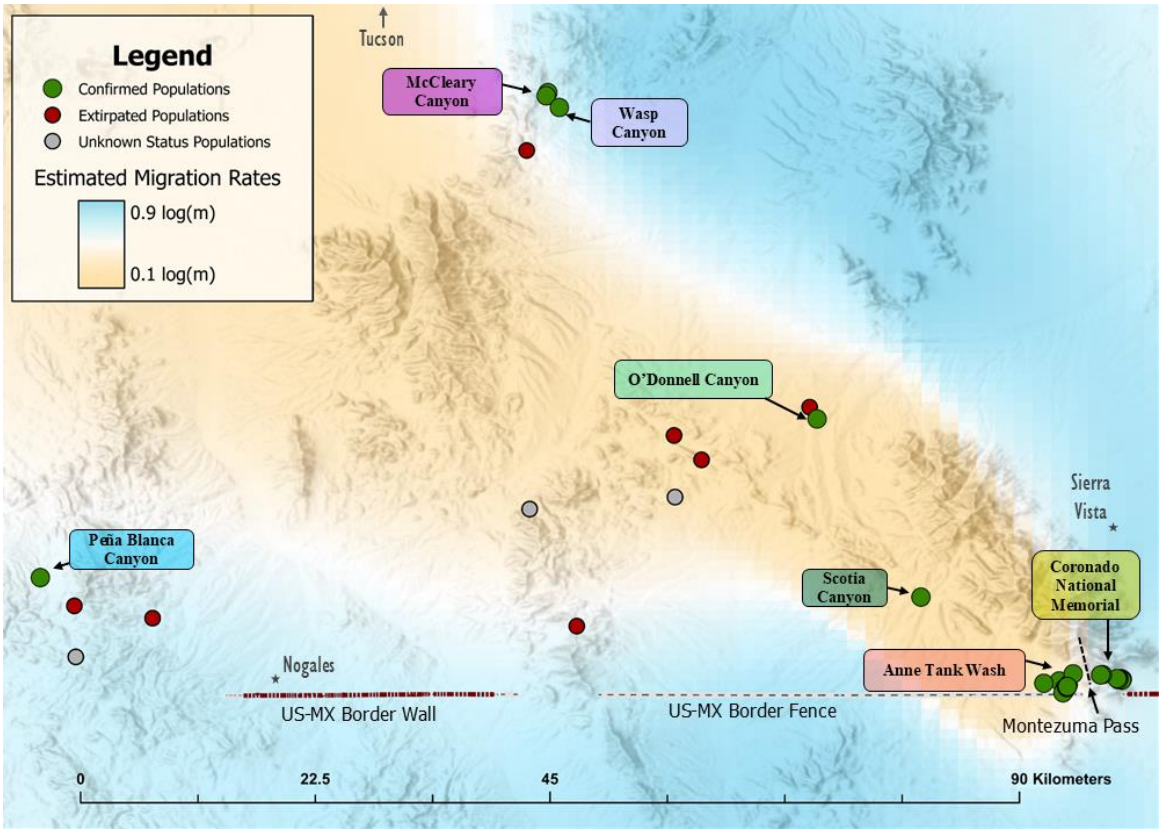


Figure 10: Estimated effective migration rates with respect to confirmed, extirpated, and unknown status populations of *P. imberbis.* Orange zones represent lower migration rates, and blue represents higher historical gene flow.

Figure 11: EEMS migrations and genetic diversity surfaces

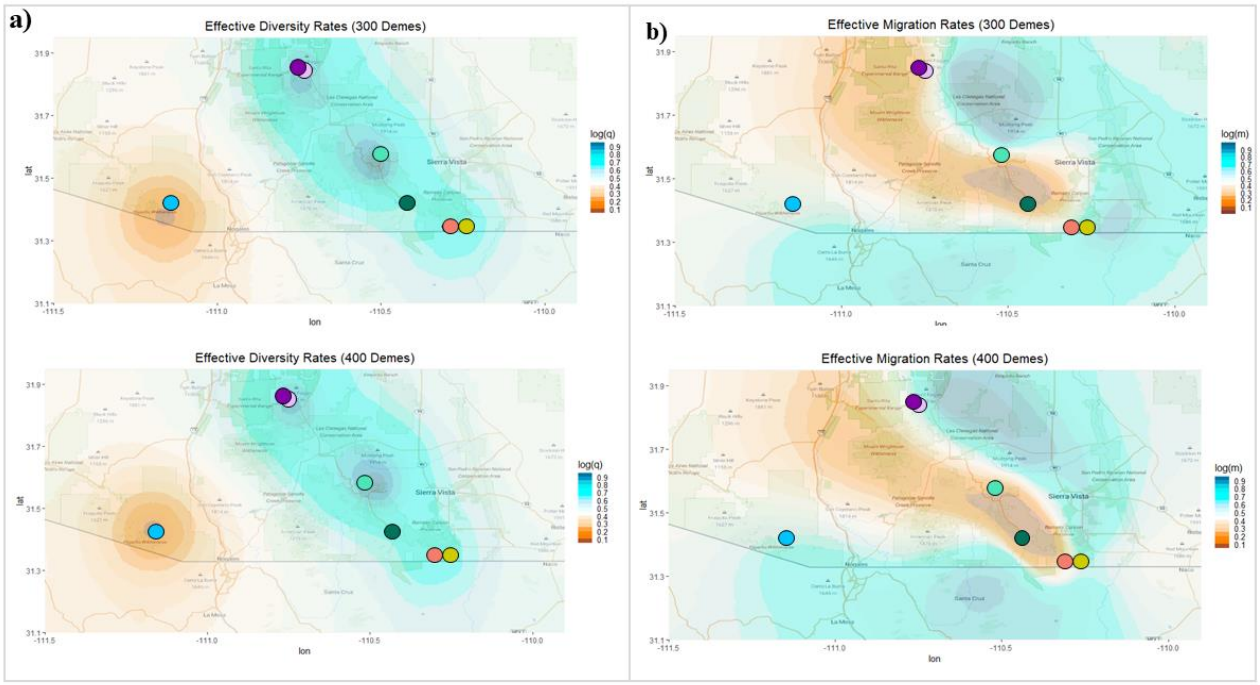


Figure 11: Comparison of deme parameter values (300 vs 400) as it affects EEMS migrations and genetic diversity surfaces.

Figure 12: Migration surfaces and habitat suitability

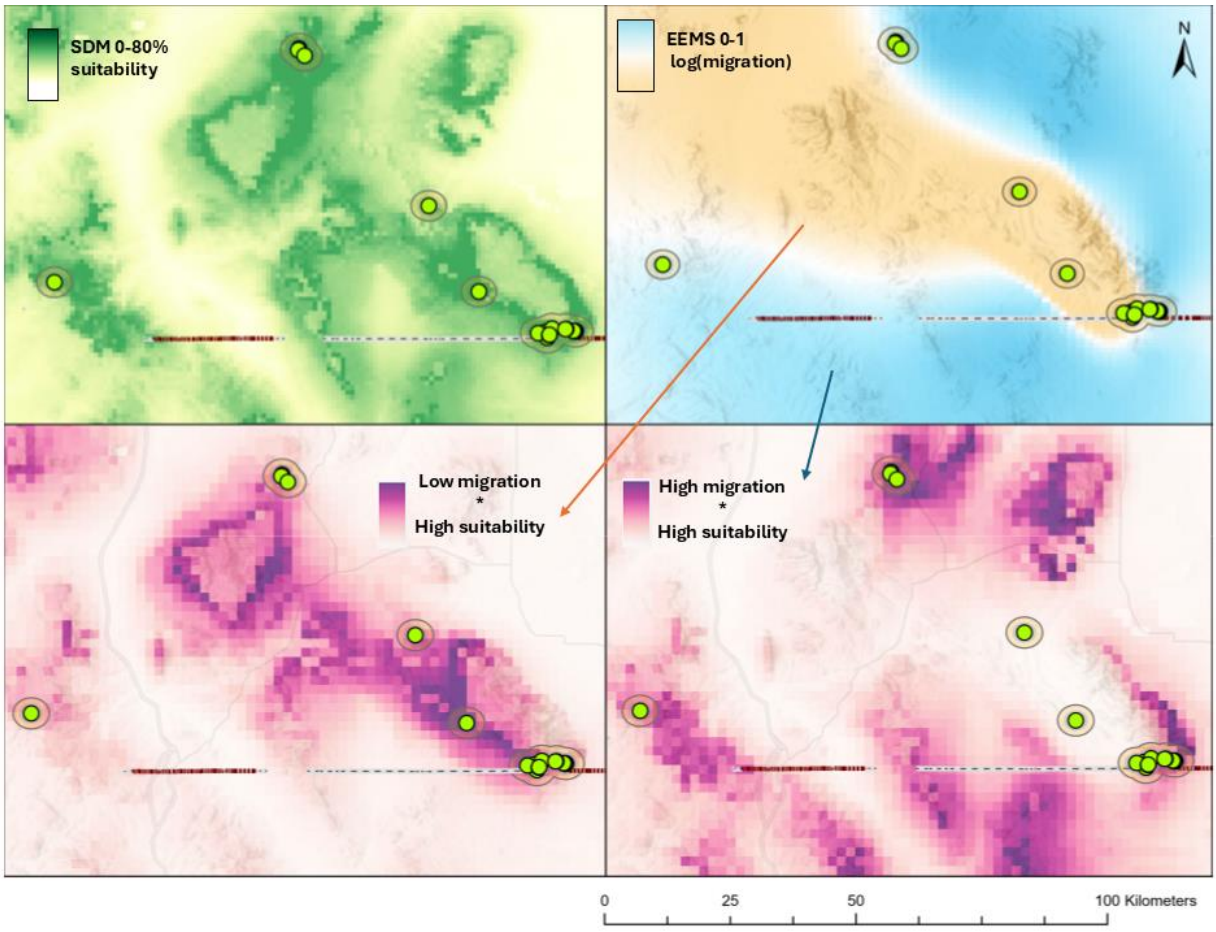


Figure 12: Green dots represent *P. imberbis* population samples in genetic analysis, each with a 3km buffer to emphasize pollinator flight distance. (**Top left):** Coarser scale SDM model using WorldClim bioclimatic variable to predict habitat suitability. **(Top Right)**: Estimated Effective Migration Rates (EEMS) highlighting areas of low (orange) and high (blue) gene flow. **(Bottom left)**: Raster calculation weighting low migration EEMS multiplied by SDM surface to emphasize regions of high habitat suitability and low gene flow potential. **(Bottom right):** Raster calculation weighting low migration EEMS multiplied by SDM surface to emphasize regions of high habitat suitability and high gene flow potential.

Figure 13: Nucleotide diversity

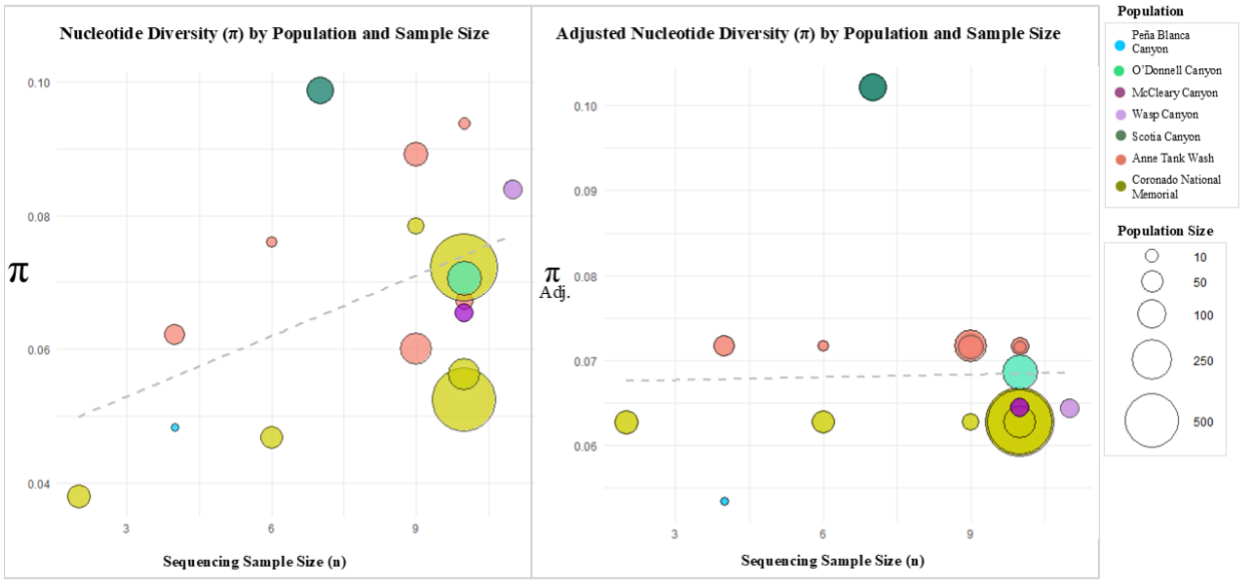


Figure 13. Subpopulation estimates of nucleotide diversity (Π) from Stacks populations output. Subpopulations are represented as individual circles color-coded to their population and scaled to match the estimated size of the subpopulation based on demographic surveys. (Left) Raw output Π values with linear regression line showing relationship between sample size, population size, and nucleotide diversity. (Right) Π values are weighted by sample size and population size using a multiple linear regression to account for sample size and estimated population size, uniform standard error applied. Scotia Canyon is an outlier when visualized with adjusted nucleotide diversity values.

Figure 14: Expected heterozygosity

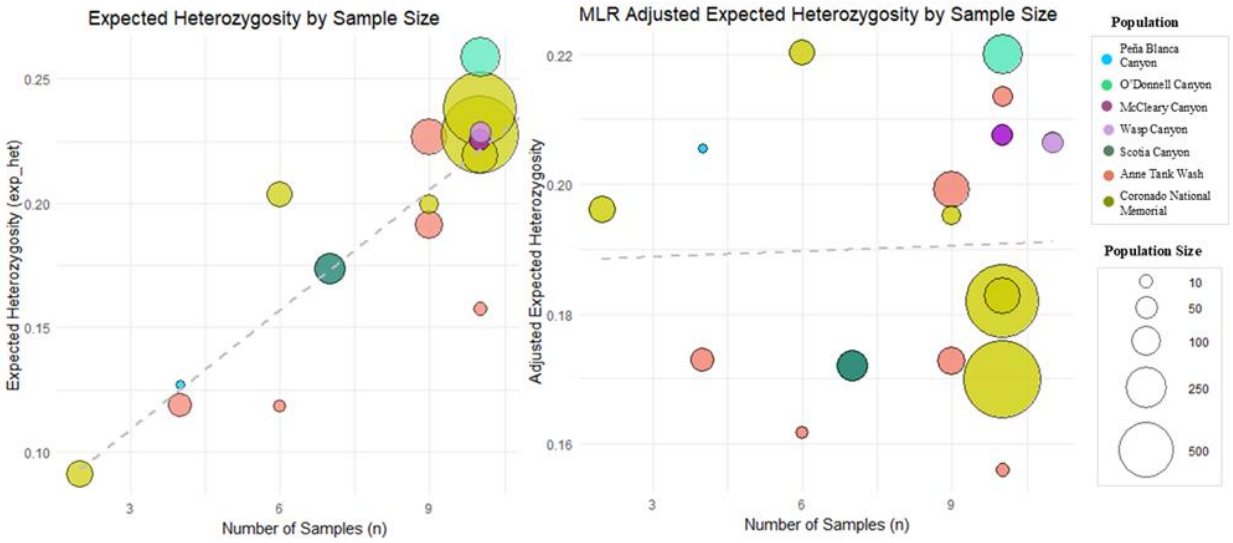


Figure 14: Subpopulation estimates of expected heterozygosity. Subpopulations are represented as individual circles color-coded to their population and scaled to match the estimated size of the subpopulation based on demographic surveys. **(Left)** Raw output expected heterozygosity values with linear regression line showing relationship between sample size, population size, and expected heterozygosity. (**Right)** Expected heterozygosity values are weighted by sample size and population size using a multiple linear regression to account for sample size and estimated population size, uniform standard error applied. O'Donnell Canyon demonstrates higher diversity when visualized with adjusted expected heterozygosity.

Figure 15: Steppingstone populations

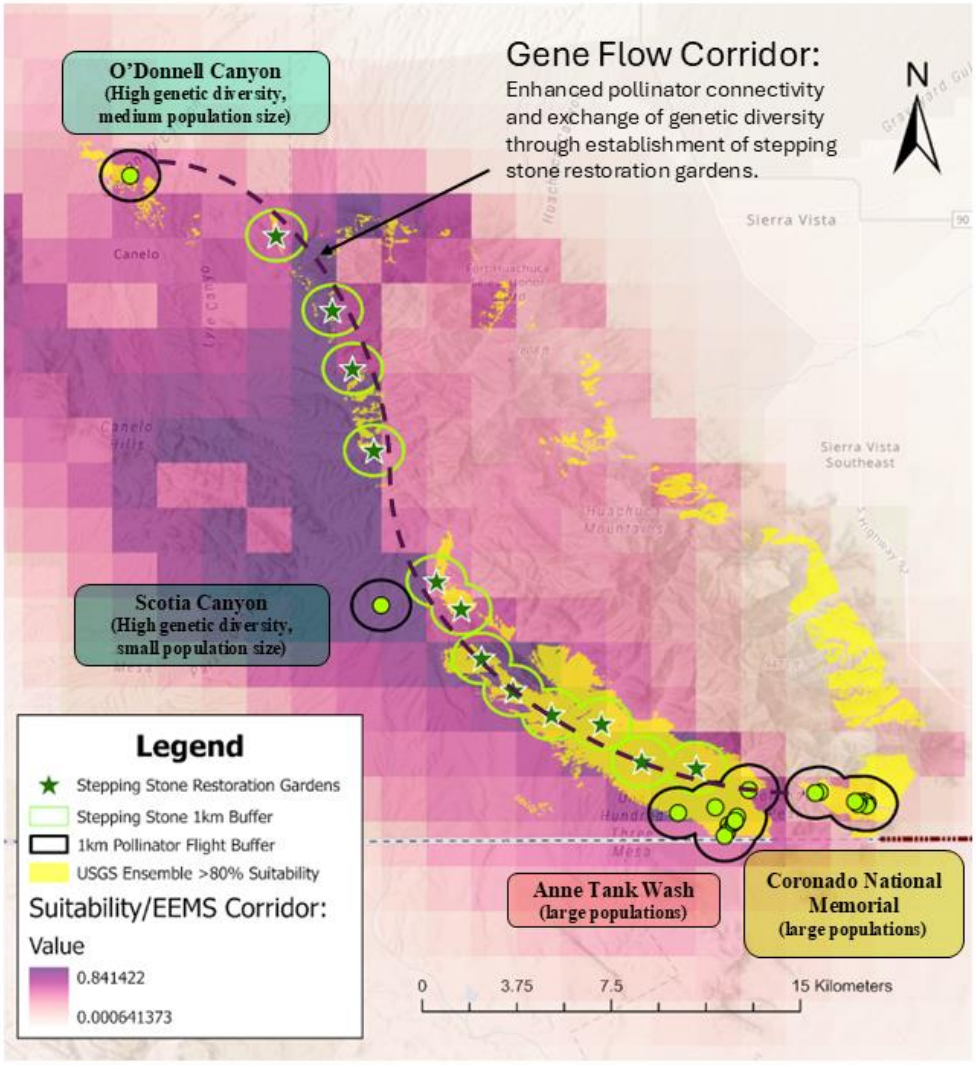


Figure 15: Conceptual map illustrating "Steppingstone" restoration plan to restore genetic connectivity from CNM populations to northern, isolated and more genetically diverse populations. The background raster layer was calculated by weighting low migration surfaces and multiplying it by high habitat suitability to generate promising gene flow corridors. The yellow layer represents the high suitability habitat extracted from the USGS ensemble model. Future restoration gardens with small patches of outplanted *P. imberbis,* as demonstrated in this map, may provide the necessary genetic and demographic buffer to assist in the recovery of this species.

Figure 16**:** Mantel Correlogram of geographic vs. environmental distances

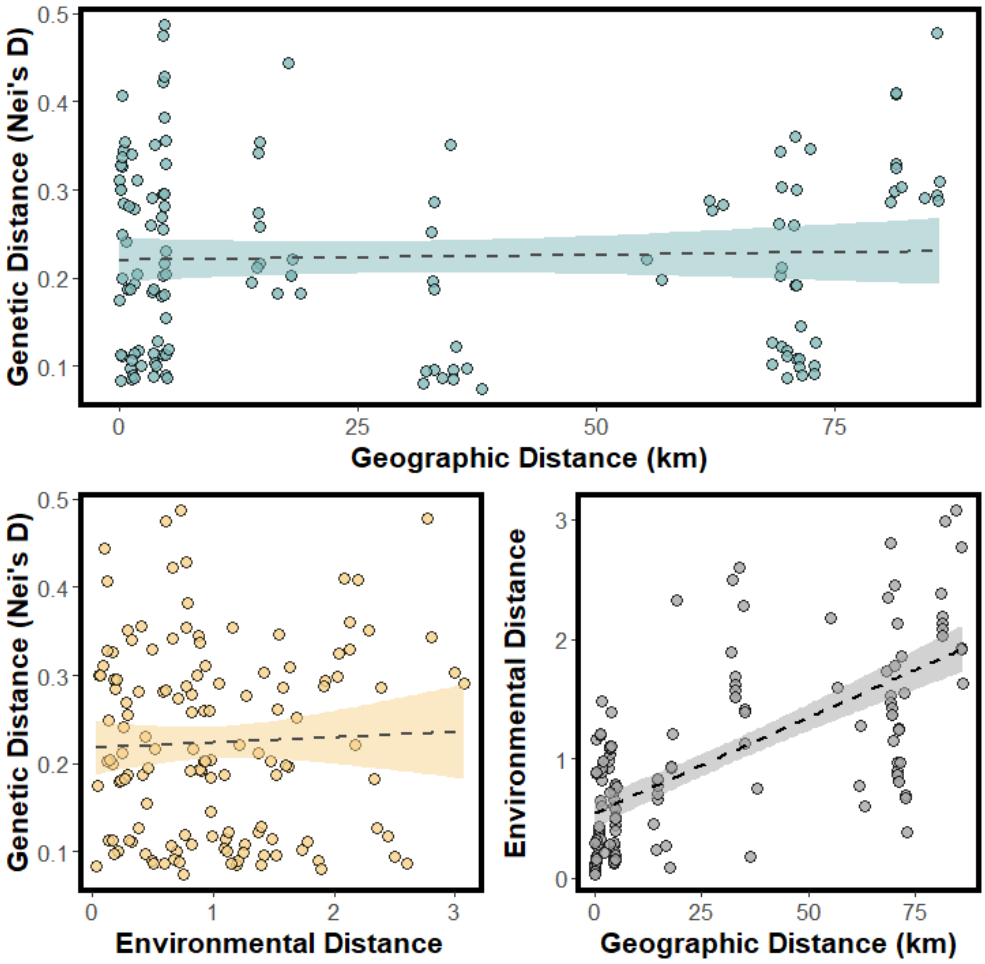


Figure 16: Mantel Correlogram of geographic vs. environmental distances. Shows a slightly positively correlated relationship between geographic distance and environment for the same populations and variables represented by the two PCA plots (*P < 0.001). As geographic distance increases between groups, so does environmental dissimilarity for a combination of four variables used, but the slope is not a strong positive relationship. Mantel test does not provide support for isolation by distance or isolation by environment.

Figure 17: Redundancy analysis

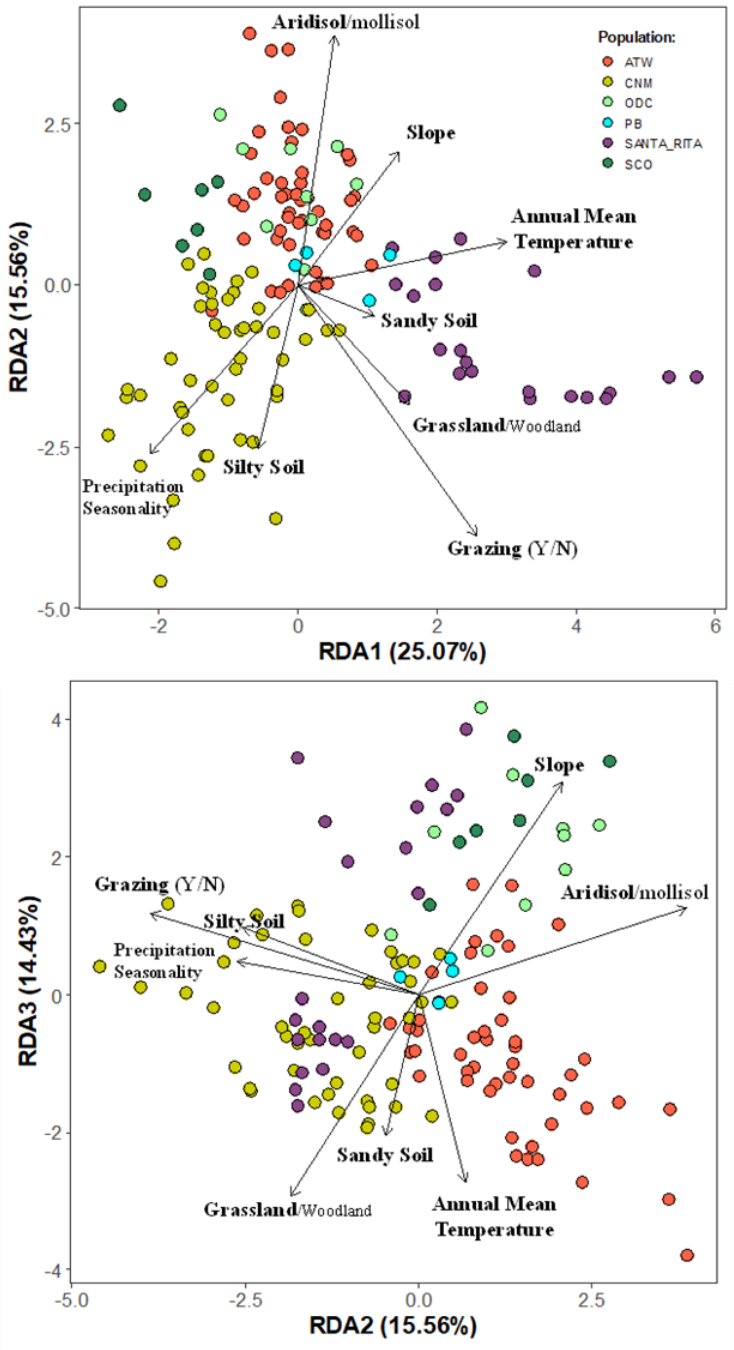


Figure 17: Redundancy analysis showing predictive relationship between 6 landscape variables and population genetic structure, calculated from SNP frequencies. Magnitude of arrows were scaled for better visibility. Total variance explained 10.6%. The RDA highlights nuances of local adaptation not shown by mantel tests.

Figure 18: Growth rate comparison from common garden experiment

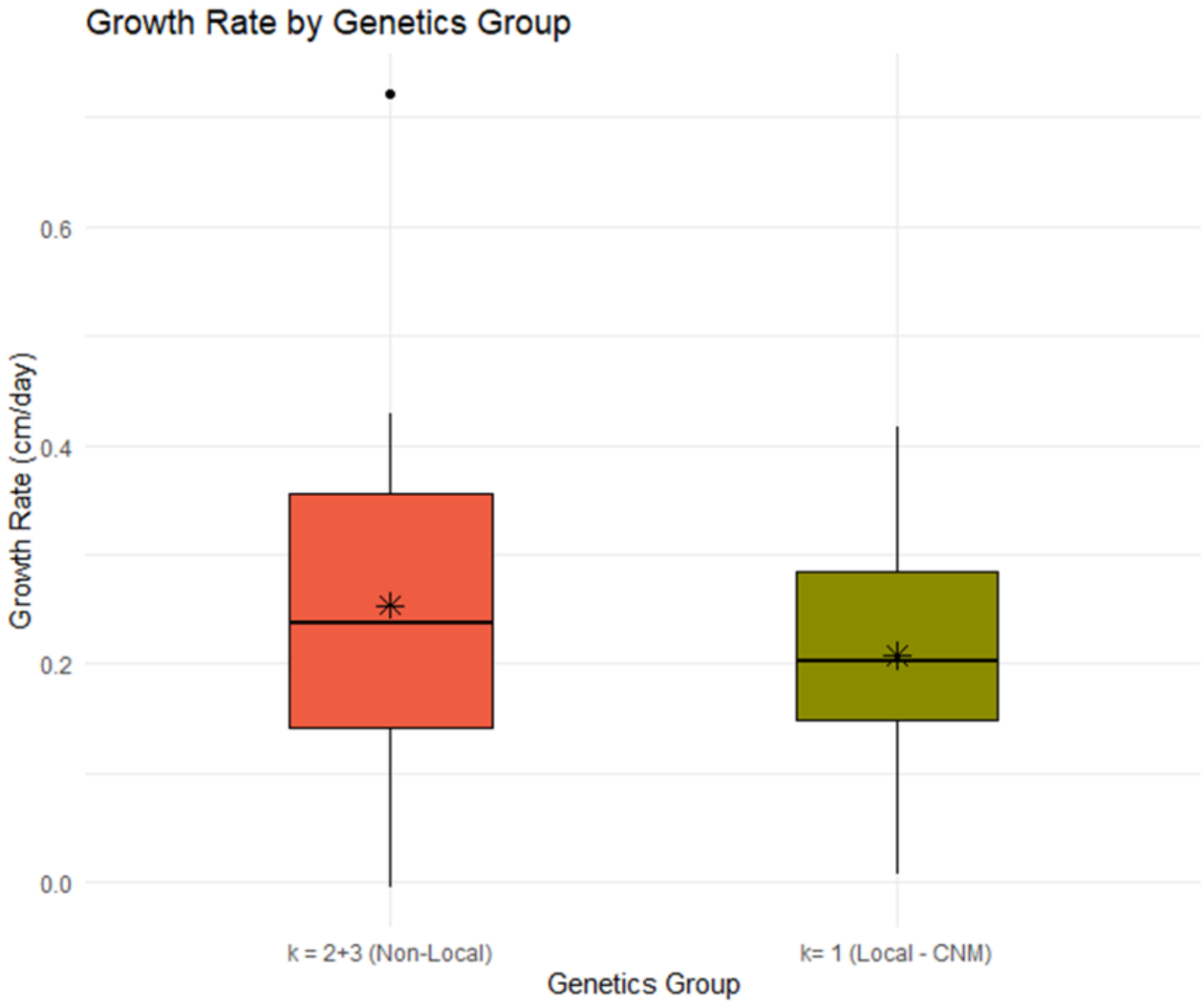


Figure 18: Comparison of growth rates (cm/day) from locally sourced versus non-locally sourced seeds. Local populations broadly correlate to a distinct genetic group, attributed to populations within CNM. Non-local group combined k = 2 and k = 3 genetic groups (figure 3) to increase sample size comparisons for local versus non-local (n = 19 for both groups) . No significant differences between groups were detected by t-test.

Figure 19: Survivorship from common garden experiment

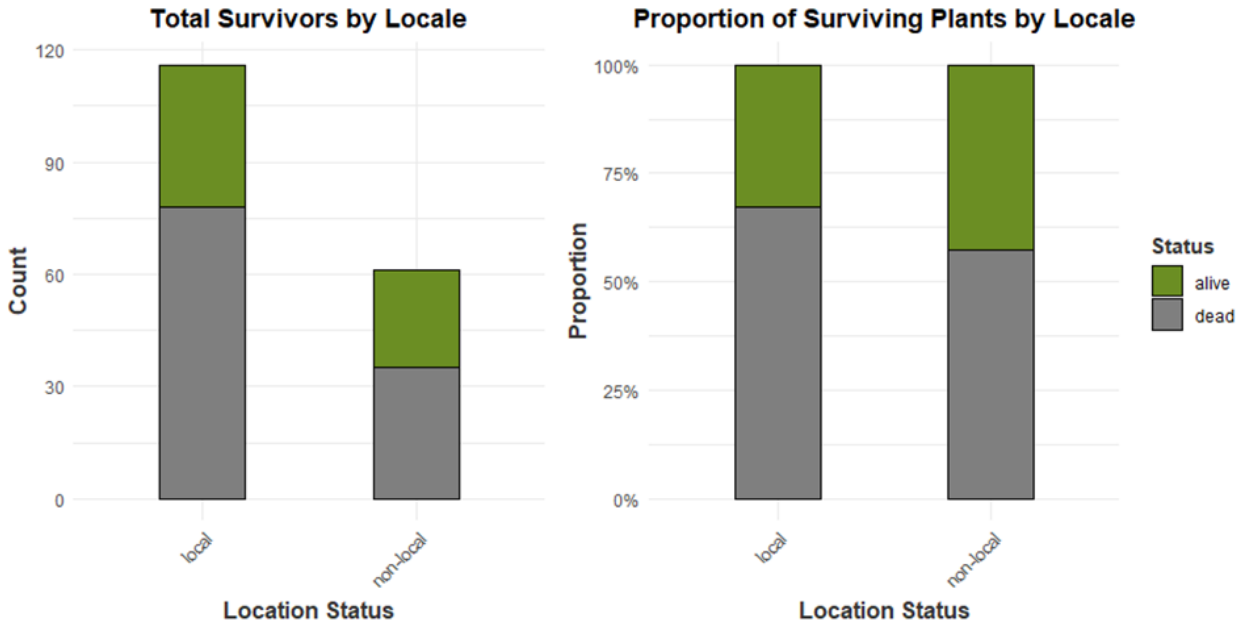


Figure 19: Comparison of survival rates from locally sourced versus non-locally sourced seeds. Local populations broadly correlate to a distinct genetic group, attributed to populations within CNM. Non-local group combined k = 2 and k = 3 genetic groups (figure 4) to increase sample size comparisons for local versus non-local. (Left) Total number of surviving individuals in common garden 42 days after out-planting from greenhouse. (Right) Relative proportion of surviving individuals to individuals that did not survive full 42 observation period.

TABLES

Table 1: Characterization of *P. imberbis* sites

Table 1: Characterization of *P. imberbis* sites, including information on positionality, biotic interactions, climate (those that were included in the final model), disturbance agents, and soils. Two subpopulations were discussed after soil data was processed, so that information is not included here. Climate variables are represented using bioclimatic variable codes: bio11 (Mean temperature of the coldest quarter), bio16 (Precipitation of the wettest quarter), bio4 (Temperature seasonality, standard deviation × 100), bio19 (Precipitation of the coldest quarter), bio6 (Minimum temperature of the coldest month), bio13 (Precipitation of the wettest month), and bio17 (Precipitation of the driest quarter).

		Location			Biotic					
Population	Subpopulation	Elevation (m)	Slope (Degree)	Aspect (Degree)	Proportion browsed	Proportion bare ground	Proportion perennial grass	Canopy cover (%)	Invasive presence (0/1)	Evidence of cattle disturbance (0/1)
CNM Vistor Center	Blue Waterfall									
	Cave Trail	1740	13	186	0.35	0.76	0.23	15	1	0
	Maintenance Shed Back	1740	15	84	0.01	0.79	0.19	15	1	0
	Maintenance Shed Front	1740	11	162	0.03	0.93	0.07	25	1	0
	Montezuma Wash	1740	32	127	0.09	0.95	0.05	1	1	0
	West Joe's Canyon									

		Historic climate						
		Bio4	Bio6	Bio11	Bio13	Bio16	Bio17	Bio19
CNM Vistor Center	Blue Waterfall							
	Cave Trail	647.015747	0.5	8.5	122	275	32	123
	Maintenance Shed Back	647.015747	0.5	8.5	122	275	32	123
	Maintenance Shed Front	647.015747	0.5	8.5	122	275	32	123
	Montezuma Wash	647.015747	0.5	8.5	122	275	32	123
	West Joe's Canyon							

		Near current climate						
		Bio4	Bio6	Bio11	Bio13	Bio16	Bio17	Bio19
CNM Vistor Center	Blue Waterfall							
	Cave Trail	670.900024	1.70000005	10.1999998	121	280	34	119
	Maintenance Shed Back	670.900024	1.70000005	10.1999998	121	280	34	119
	Maintenance Shed Front	670.900024	1.70000005	10.1999998	121	280	34	119
	Montezuma Wash	670.900024	1.70000005	10.1999998	121	280	34	119
	West Joe's Canyon							

		Future climate						
		Bio4	Bio6	Bio11	Bio13	Bio16	Bio17	Bio19
CNM Vistor Center	Blue Waterfall							
	Cave Trail	665.799988	7	14.6999998	182	400	36	140
	Maintenance Shed Back	665.799988	7	14.6999998	182	400	36	140
	Maintenance Shed Front	665.799988	7	14.6999998	182	400	36	140
	Montezuma Wash	665.799988	7	14.6999998	182	400	36	140
	West Joe's Canyon							

		Soils					Disturbance			
Population	Subpopulation	Soil pH	Soil Carbon	Soil Nitrogen	Soil Carbon to Nitrogen	USDA soil classification	Number of wildfires (N)	Time since fire (years)	Distance from roads (m)	Evidence of cattle disturbance (0/1)
CNM Vistor Center	Blue Waterfall									
	Cave Trail	6.49	1.96	0.15	13.26	mollisol	1	10	138	0
	Maintenance Shed Back	NA	1.32	0.08	17.36	mollisol	1	10	93	0
	Maintenance Shed Front	7.95	3.02	0.22	13.87	mollisol	1	10	35	0
	Montezuma Wash	7.57	NA	NA	NA	mollisol	1	10	19	0
	West Joe's Canyon									

Table 2: Greenhouse seed count and germination success

Table 2: Greenhouse seed count and germination success by population and sub-population.

Population	Sub-Population	Total Seeded	Germinants Transplanted	Germination Rate
CNM	Backside East Maintenance Shed	315	58	18.41%
	Backside West Maintenance Shed	148	44	29.73%
	Montezuma Wash	247	37	14.98%
	Lower Cave Trail	369	23	6.23%
	Upper Cave Trail	280	43	15.36%
	State of Texas Mine (Upper)	590	39	6.61%
	Lower State of Texas Mine	181	15	8.29%
ATW	Ann Tank Wash 1	85	1	1.18%
	Ann Tank Wash 1.5	69	0	0.00%
	Ann Tank Wash 2	104	9	8.65%
	Ann Tank Wash 2.5	79	0	0.00%
	Ann Tank Wash 3	114	0	0.00%
	Ann Tank Wash 3.5	60	0	0.00%
	Ann Tank Wash 4	75	8	10.67%
	Ann Tank Wash 4.5	80	19	23.75%
	Ann Tank Wash 4.75	25	5	20.00%
	Ann Tank Wash 5.5	105	17	16.19%
	Ann Tank Wash 5.75	70	7	10.00%
	Dupee Tank Wash	105	11	10.48%
SCO	Scotia Canyon	406	36	8.87%
ODC	O'Donnell Canyon grouped	597	42	7.04%
SANTA RITA	McCleary Canyon 2 West	72	0	0.00%
	McCleary Canyon 2 Middle	167	8	4.79%
	McCleary Canyon 2 East	97	0	0.00%
	Wasp Canyon 1 Lower	432	13	3.01%
	Wasp Canyon 2	164	0	0.00%

Table 3: Greenhouse seeding and germination

Table 3: Summary of greenhouse seeding and germination by CNM and non-CNM populations.

Common Garden Status	Population	(n) Seeded	Germination Rate
Local	CNM	2130	14.23%
Non-Local	ATW	866	8.41%
	SCO	406	8.87%
	ODC	597	7.04%
	SANTA RITA	932	1.56%
	Garden Locality	**Total**	**Average Germ. Rate**
	Local	2130	14.23%
	Non-Local	2801	6.47%

Table 4: SNP sampling and estimated population sizes

Table 4: Naming conventions of populations and subpopulations with associated geographic data, SNP sampling data, and estimated population sizes.

Population	Sub-Population	Sky Island	latitude	longitude	n	Pop Size
Ann Tank Wash (ATW)	Ann Tank Wash 4	Huachucas	31.3387	-110.3043	4	39
	Ann Tank Wash 4.5	Huachucas	31.3378	-110.3046	9	69
	Ann Tank Wash 4.75	Huachucas	31.3363	-110.3048	6	6
	Ann Tank Wash 5.5	Huachucas	31.3354	-110.3054	10	8
	Ann Tank Wash 5.75	Huachucas	31.3352	-110.3056	9	149
	Dupee Tank Wash	Huachucas	31.3450	-110.3091	10	25
Coronado National Memorial (CNM)	Blue Waterfall	Huachucas	31.3504	-110.2580	2	63
	Lower Cave Trail	Huachucas	31.3463	-110.2580	10	914
	Backside West Maint. Shed	Huachucas	31.3473	-110.2563	10	805
	Montezuma Wash	Huachucas	31.3457	-110.2551	6	57
	State of Texas Mine (Upper)	Huachucas	31.3506	-110.2716	10	149
	West Slope Montezuma Pass	Huachucas	31.3514	-110.2968	9	21
O'Donnell Canyon (ODC)	O'Donnell Canyon	Canelo hills	31.5705	-110.5173	10	186
Pena Blanca Canyon (PBC)	Peña Blanca	Atascosa	31.4096	-111.1571	4	5
(SANTA RITA)	McCleary Canyon 2 East	Santa Ritas	31.8484	-110.7505	10	31
	Wasp Canyon 1 Lower	Santa Ritas	31.8377	-110.7396	11	26
Scotia Canyon (SCO)	Scotia Canyon	Huachucas	31.4170	-110.4281	7	93

Table 5: Genetic diversity estimates

Table 5: Within population genetic diversity estimates by population and subpopulation (Inbreeding coefficient (Fis), number of private alleles, nucleotide diversity, expected heterozygosity). Estimated population size and number of individuals samples (n) included.

Pop	SubPop	Fis	Priv. Alleles	π	ExpHet	Est. Pop Size	n
ATW	ATW4	0.05014	841	0.06213	0.11851	39	4
ATW	ATW4.5	0.07723	3864	0.08913	0.19123	69	9
ATW	ATW4.75	0.05001	2083	0.07594	0.11825	6	6
ATW	ATW5.5	0.08416	4404	0.09377	0.15747	8	10
ATW	ATW5.75	0.04213	1689	0.0601	0.22675	149	9
CNM	CNM_BW	0.00418	256	0.038	0.09104	63	2
CNM	CNM_LCT	0.04911	2352	0.07213	0.22764	914	10
CNM	CNM_MSBW	0.0284	1082	0.05248	0.2381	805	10
CNM	CNM_MW	0.01875	783	0.0468	0.20351	57	6
CNM	CNM_USTX	0.03558	1621	0.05621	0.21901	149	10
ATW	DT	0.05461	1829	0.06729	0.22864	25	10
SANTA_RITA	MCC2E	0.04397	1754	0.06547	0.22515	31	10
ODC	OD	0.05829	2322	0.07066	0.25891	186	10
PBC	PB	0.02261	1033	0.04839	0.1269	5	4
SCOTIA	SCOTIA	0.09112	4504	0.09869	0.17368	93	7
SANTA_RITA	WASP	0.06296	3116	0.08402	0.22986	26	11
CNM	WSM	0.04964	2549	0.07839	0.19954	21	9

Table 6: Gene flow values

Table 6: Pairwise gene flow values and corresponding pairwise geographic distances. Minimum and maximum values in bold.

Location1	Location2	Distance (km)	Gene Flow (Nm)
ATW	CNM	**3.94762**	**2.978931**
ATW	ODC	32.76821	1.7398234
ATW	PB	81.31793	0.8647602
ATW	SANTA_RITA	69.97324	2.0250395
ATW	SCO	14.58167	0.9259179
CNM	ODC	34.34753	1.7379988
CNM	PB	**84.96148**	0.7856759
CNM	SANTA_RITA	71.36325	1.9402392
CNM	SCO	17.18108	0.9036743
ODC	PB	63.32326	0.5247913
ODC	SANTA_RITA	37.22198	1.3652179
ODC	SCO	19.07466	0.6613792
PB	SANTA_RITA	62.07508	0.6749724
PB	SCO	69.26742	**0.343365**
SANTA RITA	SCO	56.14196	0.7242905

Table 7: Geographic distance by population

Table 7: Mean F_{ST}, mean Gene Flow, and mean geographical distance by population.

Population	Mean F_{ST}	Mean Gene Flow (Nm)	Mean Distance (km)
CNM	0.16	1.67	42.36
ATW	**0.15**	**1.71**	40.52
SCO	0.28	0.71	**35.25**
ODC	0.20	1.21	37.35
SANTA RITA	0.18	1.35	59.36
PB	**0.30**	**0.64**	**72.19**
Mean:	**0.21**	**1.21**	**47.84**

Table 8: Pairwise gene flow between populations

Table 8: Pairwise gene flow average values between populations.

Nm (Gene Flow)	PB	SANTA_RITA	ODC	SCO	ATW	CNM
PB	0	0.6749724	0.524791	**0.343365**	0.86476	0.785676
SANTA_RITA		0	1.365218	0.724291	2.02504	1.940239
ODC			0	0.661379	1.739823	1.737999
SCO				0	0.925918	0.903674
ATW					0	**2.978931**
CNM						0

Table 9: Common garden germination rates

Table 9: Results of germination and survival for individuals out-planted into CNM garden.

Population	Population	(n) Germinants	(n) Planted	(n) Survivors	Survival Ratio
CNM	Backside East Maintenance Shed	58	11	2	18.18%
	Backside West Maintenance Shed	44	10	5	50.00%
	Montezuma Wash	37	12	4	33.33%
	Lower Cave Trail	23	18	5	27.78%
	Upper Cave Trail	43	22	8	36.36%
	State of Texas Mine (Upper)	39	13	5	38.46%
	Lower State of Texas Mine	15	27	7	25.93%
ATW	Ann Tank Wash 1	1	0	0	0.00%
	Ann Tank Wash 1.5	0	0	0	0.00%
	Ann Tank Wash 2	9	8	2	25.00%
	Ann Tank Wash 2.5	0	0	0	0.00%
	Ann Tank Wash 3	0	0	0	0.00%
	Ann Tank Wash 3.5	0	0	0	0.00%
	Ann Tank Wash 4	8	1	1	100.00%
	Ann Tank Wash 4.5	19	11	6	54.55%
	Ann Tank Wash 4.75	5	3	0	0.00%
	Ann Tank Wash 5.5	17	6	2	33.33%
	Ann Tank Wash 5.75	7	3	0	0.00%
	Dupee Tank Wash	11	1	1	100.00%
SCO	Scotia Canyon	36	11	5	45.45%
ODC	O'Donnell Canyon grouped	42	8	5	62.50%
SANTA RITA	McCleary Canyon 2 West	0	0	0	0.00%
	McCleary Canyon 2 Middle	8	5	2	40.00%
	McCleary Canyon 2 East	0	0	0	0.00%
	Wasp Canyon 1 Lower	13	4	4	100.00%
	Wasp Canyon 2	0	0	0	0.00%

Table 10: Common garden survival rates

Table 10: Common garden survival rates for local and non-local plants.

Common Garden Status	Population	(n) Planted	Survival Rate
Local	CNM	113	32.86%
Non-Local	ATW	33	26.07%
	SCO	11	45.45%
	ODC	8	62.50%
	SANTA RITA	9	28.00%
	Garden Locality	**Total**	**Average Survival Rate**
	Local	113	32.86%
	Non-Local	61	40.51%

Table 11: Predictors of common garden survivorship

Table 11: Summary of top three generalized linear mixed models for *P. imberbis* survival ratio in common garden. Includes AIC, beta coefficient estimates, standard errors, z-values, and *p*-values. Population source was a random effect in each model. Top model is lowest AIC while maintaining significance for at least one predictor variable.

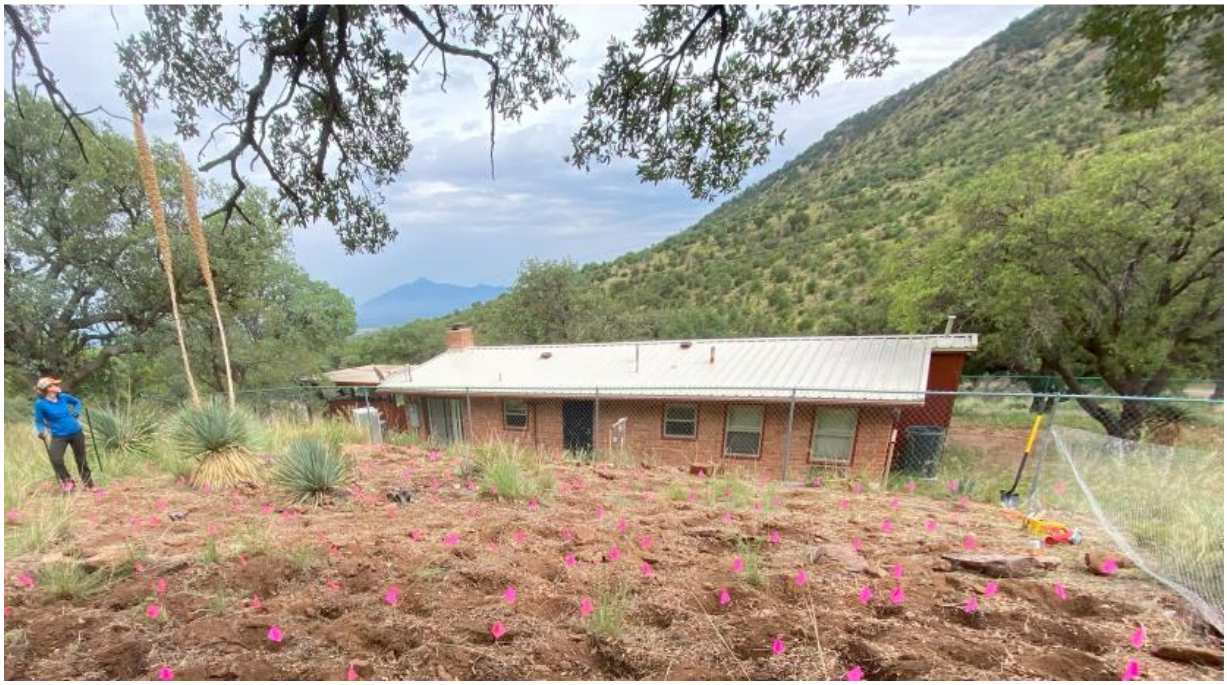
AIC	Predictor Variable	Est.	Std. Error	z	*p*-value
20.7	**(Intercept)**	**0.46**	**0.08**	**6.03**	**<0.001*** **
	Environmental Distance	**0.21**	**0.12**	**1.84**	**0.06 .**
	Genetic Cluster	**0.16**	**0.12**	**1.32**	**0.19**
	Source Population Size	**-0.05**	**0.10**	**-0.57**	**0.57**
	Germination Rate	**0.16**	**0.12**	**1.35**	**0.18**
	Initial Planting Height	**-0.21**	**0.11**	**-1.97**	**0.04 ***
22.2	**(Intercept)**	**0.46**	**0.08**	**6.15**	**<0.001*** **
	Nucleotide Diversity (II)	**-0.07**	**0.10**	**-0.23**	**0.46**
	Environmental Distance	**0.24**	**0.12**	**2.01**	**0.04 ***
	Genetic Cluster	**-0.06**	**0.14**	**1.52**	**0.13**
	Source Population Size	**-0.06**	**0.10**	**-0.60**	**0.54**
	Germination Rate	**0.20**	**0.13**	**1.55**	**0.12**
	Initial Planting Height	**-0.24**	**0.11**	**-2.13**	**0.03 ***

23.1	**(Intercept)**	**0.46**	**0.07**	**6.37**	**<0.001*** **
	Nucleotide Diversity (II)	**-0.03**	**0.10**	**0.33**	**0.74**
	Environmental Distance	**0.21**	**0.11**	**1.73**	**0.08 .**
	Genetic Cluster	**0.28**	**0.15**	**1.88**	**0.06 .**
	Difference % Sand	**0.12**	**0.12**	**1.03**	**0.30**
	Source Population Size	**0.03**	**0.13**	**0.27**	**0.79**
	Germination Rate	**0.22**	**0.13**	**1.78**	**0.08 .**
	Initial Planting Height	**-0.23**	**0.11**	**2.08**	**0.04 ***

PICTURES TAKEN DURING THIS PROJECT

Pectis imberbis **seedlings.**

***Pectis imberis* individuals prior to transplant to common gardens.**



Common gardens at CORO.

CITATIONS

Adler, P. B., Salguero-Gomez, R., Compagnoni, A., Hsu, J. S., Ray-Mukherjee, J., Mbeau-Ache, C., et al. (2014). Functional traits explain variation in plant life history strategies. *Proceedings of the National Academy of Sciences* 111, 740–745. doi: 10.1073/pnas.1315179111

Anic, V., C. A. Henríquez, S. R. Abades, and R. O. Bustamante. "Number of Conspecifics and Reproduction in the Invasive Plant Eschscholzia Californica (Papaveraceae): Is There a Pollinator-Mediated Allee Effect?" *Plant Biology* 17, no. 3 (2015): 720–27. https://doi.org/10.1111/plb.12293.

Bhatia, Gaurav, Nick Patterson, Sriram Sankararaman, and Alkes L. Price. "Estimating and Interpreting FST: The Impact of Rare Variants." *Genome Research* 23, no. 9 (September 2013): 1514–21. https://doi.org/10.1101/gr.154831.113.

Birzu, Gabriel, Sakib Matin, Oskar Hallatschek, and Kirill S. Korolev. "Genetic Drift in Range Expansions Is Very Sensitive to Density Dependence in Dispersal and Growth." *Ecology Letters* 22, no. 11 (2019): 1817–27. https://doi.org/10.1111/ele.13364.

Bontrager, Megan, and Amy L. Angert. "Gene Flow Improves Fitness at a Range Edge under Climate Change." *Evolution Letters* 3, no. 1 (2019): 55–68. https://doi.org/10.1002/evl3.91.

Borokini, Israel T., Kelly B. Klingler, and Mary M. Peacock. "Life in the Desert: The Impact of Geographic and Environmental Gradients on Genetic Diversity and Population Structure of Ivesia Webberi." *Ecology and Evolution* 11, no. 23 (2021): 17537–56. https://doi.org/10.1002/ece3.8389.

Bridle, Jon R., and Timothy H. Vines. "Limits to Evolution at Range Margins: When and Why Does Adaptation Fail?" *Trends in Ecology & Evolution* 22, no. 3 (March 2007): 140–47. https://doi.org/10.1016/j.tree.2006.11.002.

Burnett, N. P., Badger, M. A., and Combes, S. A. (2021). Wind and route choice affect performance of bees flying above versus within a cluttered obstacle field. doi: 10.1101/2021.10.08.463704

Buchmann, S. L. (1994). Diversity and importance of native bees from the Arizona/Mexico Madrean archipelago. In L. F. DeBano, P. F. Ffolliott, A. Ortega-Rubio, G. J. Gottfried, R. H. Hamre, & C. B. Edminster (Eds.), *Biodiversity and management of the Madrean archipelago: The sky islands of southwestern United States and northwestern Mexico* (pp. 301–310). Fort Collins, CO: U. S. Department of Agriculture, Forest Service, Rocky Mountain Forest and Range Experiment Station.

Busch, J. W. "Inbreeding Depression in Self-Incompatible and Self-Compatible Populations of Leavenworthia Alabamica." *Heredity* 94, no. 2 (February 2005): 159–65. https://doi.org/10.1038/sj.hdy.6800584.

Catchen, Julian, Paul A. Hohenlohe, Susan Bassham, Angel Amores, and William A. Cresko. "Stacks: An Analysis Tool Set for Population Genomics." *Molecular Ecology* 22, no. 11 (June 2013): 3124–40. https://doi.org/10.1111/mec.12354.

Castillo, William J., Laura A. Burkle, and Carsten F. Dormann. "Dynamics of a Plant–Pollinator Network: Extending the Bianconi–Barabási Model." *Applied Network Science* 9, no. 1 (December 2024): 1–12. https://doi.org/10.1007/s41109-024-00636-0.

Cheng, Jin, Huixia Kao, and Shubin Dong. "Population Genetic Structure and Gene Flow of Rare and Endangered Tetraena Mongolica Maxim. Revealed by Reduced Representation Sequencing." *BMC Plant Biology* 20, no. 1 (August 26, 2020): 391. https://doi.org/10.1186/s12870-020-02594-y.

Chi, Katherine, and Brenda Molano-Flores. "Can Floral Display Size Compensate for Allee Effects Caused by Low Population Abundance and Density in Synthyris Bullii (Plantaginaceae), a Rare Species?" *American Journal of Botany* 101, no. 3 (2014): 428–36. https://doi.org/10.3732/ajb.1300457.

Dang, Zhenhua, Jiabin Li, Yanan Liu, Miaomiao Song, Peter J. Lockhart, Yunyun Tian, Miaomiao Niu, and Qinglang Wang. "RADseq-Based Population Genomic Analysis and Environmental Adaptation of Rare and Endangered Recretohalophyte Reaumuria Trigyna." *The Plant Genome* n/a, no. n/a (n.d.): e20303. https://doi.org/10.1002/tpg2.20303.

Dennis, Brian, Patricia L. Munholland, and J. Michael Scott. "Estimation of Growth and Extinction Parameters for Endangered Species." *Ecological Monographs* 61, no. 2 (1991): 115–43. https://doi.org/10.2307/1943004.

Dibner, Reilly R., Megan L. DeMarche, Allison M. Louthan, and Daniel F. Doak. "Multiple Mechanisms Confer Stability to Isolated Populations of a Rare Endemic Plant." *Ecological Monographs* 89, no. 2 (2019): e01360. https://doi.org/10.1002/ecm.1360.

Ellstrand, Norman C., and Diane R. Elam. "Population Genetic Consequences of Small Population Size: Implications for Plant Conservation." *Annual Review of Ecology, Evolution, and Systematics* 24, no. Volume 24, 1993 (November 1, 1993): 217–42. https://doi.org/10.1146/annurev.es.24.110193.001245.

Evans, Simone A., Dennis F. Whigham, Ida Hartvig, and Melissa K. McCormick. "Hybridization in the Fringed Orchids: An Analysis of Species Boundaries in the Face of Gene Flow." *Diversity* 15, no. 3 (March 2023): 384. https://doi.org/10.3390/d15030384.

Faske, Trevor M., Alison C. Agneray, Joshua P. Jahner, Lana M. Sheta, Elizabeth A. Leger, and Thomas L. Parchman. "Genomic and Common Garden Approaches Yield Complementary Results for Quantifying Environmental Drivers of Local Adaptation in Rubber Rabbitbrush, a Foundational Great Basin Shrub." *Evolutionary Applications* 14, no. 12 (2021): 2881–2900. https://doi.org/10.1111/eva.13323.

Fishbein, M. and P. Warren. 1994. Population studies of sensitive plants of the Lone Mountain Ecosystem Management Area, Coronado National Forest, Arizona.

Frye, Christopher T., and Maile C. Neel. "Benefits of Gene Flow Are Mediated by Individual Variability in Self-Compatibility in Small Isolated Populations of an Endemic Plant Species." *Evolutionary Applications* 10, no. 6 (2017): 551–62. https://doi.org/10.1111/eva.12437

Gathmann, Achim, and Teja Tscharntke. "Foraging Ranges of Solitary Bees." *Journal of Animal Ecology* 71 (September 6, 2002): 757–64. https://doi.org/10.1046/j.1365-2656.2002.00641.x.

Gamba, Diana, and Nathan Muchhala. "Pollinator Type Strongly Impacts Gene Flow within and among Plant Populations for Six Neotropical Species." *Ecology* 104, no. 1 (2023): e3845. https://doi.org/10.1002/ecy.3845.

Gentili, Rodolfo, Aldo Solari, Martin Diekmann, Cecilia Duprè, Gianna Serafina Monti, Stefano Armiraglio, Silvia Assini, and Sandra Citterio. "Genetic Differentiation, Local Adaptation and Phenotypic Plasticity in Fragmented Populations of a Rare Forest Herb." *PeerJ* 6 (June 13, 2018): e4929. https://doi.org/10.7717/peerj.4929.

Gompert, Zachariah, and C. Alex Buerkle. "What, If Anything, Are Hybrids: Enduring Truths and Challenges Associated with Population Structure and Gene Flow." *Evolutionary Applications* 9, no. 7 (April 26, 2016): 909–23. https://doi.org/10.1111/eva.12380.

Goudet, J. and Jombart, T. (2020) Hierfstat: Estimation and Tests of Hierarchical F-Statistics. R Package Version 0.5-7. https://CRAN.R-project.org/package=hierfstat

Greenleaf, Sarah S., Neal M. Williams, Rachael Winfree, and Claire Kremen. "Bee Foraging Ranges and Their Relationship to Body Size." *Oecologia* 153, no. 3 (September 2007): 589–96. https://doi.org/10.1007/s00442-007-0752-9.

Hannah, Lee. "Chapter 4 - Ecological, Evolutionary, and Biogeographic Implications of Climate Change." In *Climate Change Biology (Third Edition)*, edited by Lee Hannah, 77–94. Academic Press, 2022. https://doi.org/10.1016/B978-0-08-102975-6.00004-2.

Hamabata, T., Kinoshita, G., Kurita, K., Cao, P.-L., Ito, M., Murata, J., Komaki, Y., Isagi, Y., & Makino, T. (2019). Endangered island endemic plants have vulnerable genomes. *Communications Biology*, *2*(1), Article 1.

Hartigan, J. A., and M. A. Wong. "Algorithm AS 136: A K-Means Clustering Algorithm." *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 28, no. 1 (1979): 100–108. https://doi.org/10.2307/2346830.

Hartl, Daniel L., and A. G. Clark. "Principles of Population Genetics, 4th Edition". *Journal of Heredity*, Volume 98, Issue 4, July 2007, Page 382. (2007). https://doi.org/10.1093/jhered/esm035

Hedrick, P. W., & Miller, P. S. (1992). Conservation genetics: techniques and fundamentals. Ecological Applications, 2(1), 30-46

Herman, Rachael W, Benjamin M Winger, Donna L Dittmann, and Michael G Harvey. "Fine-Scale Population Genetic Structure and Barriers to Gene Flow in a Widespread Seabird (Ardenna Pacifica)." *Biological Journal of the Linnean Society* 137, no. 1 (September 1, 2022): 125–36. https://doi.org/10.1093/biolinnean/blac091

Hedrick, Philip W., and Philip S. Miller. "Conservation Genetics: Techniques and Fundamentals." *Ecological Applications* 2, no. 1 (1992): 30–46. https://doi.org/10.2307/1941887.

Hofmann, Michaela M., Andreas Fleischmann, and Susanne S. Renner. "Foraging Distances in Six Species of Solitary Bees with Body Lengths of 6 to 15 Mm, Inferred from Individual Tagging, Suggest 150 m-Rule-of-Thumb for Flower Strip Distances." *Journal of Hymenoptera Research* 77 (June 29, 2020): 105–17. https://doi.org/10.3897/jhr.77.51182.

Holderegger, Rolf, Niko Balkenhol, Janine Bolliger, Jan O. Engler, Felix Gugerli, Axel Hochkirch, Carsten Nowak, Gernot Segelbacher, Alex Widmer, and Frank E. Zachos. "Conservation Genetics: Linking Science with Practice." *Molecular Ecology* 28, no. 17 (September 2019): 3848–56. https://doi.org/10.1111/mec.15202.

Honnay, Olivier, and Hans Jacquemyn. "Susceptibility of Common and Rare Plant Species to the Genetic Consequences of Habitat Fragmentation." *Conservation Biology* 21, no. 3 (2007): 823–31. https://doi.org/10.1111/j.1523-1739.2006.00646.x.

Jombart, Thibaut, Sébastien Devillard, and François Balloux. "Discriminant Analysis of Principal Components: A New Method for the Analysis of Genetically Structured Populations." *BMC Genetics* 11, no. 1 (October 15, 2010): 94. https://doi.org/10.1186/1471-2156-11-94.

Jones, Matthew R., Daniel E. Winkler, and Rob Massatti. "Demographic Modeling Informs Functional Connectivity and Management Interventions in Graham's Beardtongue." *Conservation Genetics* 22, no. 6 (December 1, 2021): 993–1003. https://doi.org/10.1007/s10592-021-01392-9.

Jump, Alistair S., and Josep Peñuelas. “Running to Stand Still: Adaptation and the Response of Plants to Rapid Climate Change.” *Ecology Letters* 8, no. 9 (2005): 1010–20. https://doi.org/10.1111/j.1461-0248.2005.00796.x.

Khatri, Bhavin S., and Austin Burt. “Robust Estimation of Recent Effective Population Size from Number of Independent Origins in Soft Sweeps.” *Molecular Biology and Evolution* 36, no. 9 (September 1, 2019): 2040–52. https://doi.org/10.1093/molbev/msz081.

Kling, M. M., and Ackerly, D. D. (2021). Global wind patterns shape genetic differentiation, asymmetric gene flow, and genetic diversity in trees. *Proc. Natl. Acad. Sci. U.S.A.* 118, e2017317118. doi: 10.1073/pnas.2017317118

Kwak, Manja M., Odilia Velterop, and Jelte van Andel. “Pollen and Gene Flow in Fragmented Habitats.” *Applied Vegetation Science* 1, no. 1 (1998): 37–54. https://doi.org/10.2307/1479084.

Laikre, Linda, Fred W. Allendorf, Laurel C. Aroner, C. Scott Baker, David P. Gregovich, Michael M. Hansen, Jennifer A. Jackson, et al. “Neglect of Genetic Diversity in Implementation of the Convention of Biological Diversity.” *Conservation Biology* 24, no. 1 (2010): 86–88.

Lande, Russell, and Susan Shannon. “The Role of Genetic Variation in Adaptation and Population Persistence in a Changing Environment.” *Evolution* 50, no. 1 (1996): 434–37. https://doi.org/10.2307/2410812.

Li, Yao, Amol C. Shetty, Chanthap Lon, Michele Spring, David L. Saunders, Mark M. Fukuda, Tran Tinh Hien, et al. “Detecting Geospatial Patterns of Plasmodium Falciparum Parasite Migration in Cambodia Using Optimized Estimated Effective Migration Surfaces.” *International Journal of Health Geographics* 19, no. 1 (April 10, 2020): 13. https://doi.org/10.1186/s12942-020-00207-3.

Luijten, Sheila H., Angelo Dierick, J. Gerard, B. Oostermeijer, Léon E. L. Raijmann, and Hans C. M. Den Nijs. “Population Size, Genetic Variation, and Reproductive Success in a Rapidly Declining, Self-Incompatible Perennial (Arnica Montana) in The Netherlands.” *Conservation Biology* 14, no. 6 (2000): 1776–87. https://doi.org/10.1111/j.1523-1739.2000.99345.x.

McGreevy, Thomas J., Nicholas G. Crawford, Pierre Legreneur, and Christopher J. Schneider. “Influence of Geographic Isolation and the Environment on Gene Flow among Phenotypically Diverse Lizards.” *Heredity*, September 12, 2024, 1–14. https://doi.org/10.1038/s41437-024-00716-y.

McGreevy, Thomas J., Sozos Michaelides, Mihajla Djan, Mary Sullivan, Diana M. Beltrán, Bill Buffum, and Thomas Husband. “Location and Species Matters: Variable Influence of the Environment on the Gene Flow of Imperiled, Native and Invasive Cottontails.” *Frontiers in Genetics* 12 (September 29, 2021). https://doi.org/10.3389/fgene.2021.708871.

McPherson, Guy R., and Jake F. Weltzin. “Disturbance and Climate Change in United States/Mexico Borderland Plant Communities: A State-of-the-Knowledge Review.” *Gen. Tech. Rep. RMRS-GTR-50. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 24 p.* 050 (2000). https://doi.org/10.2737/RMRS-GTR-50.

Mhemmed, Gandour, Hessini Kamel, and Abdelly Chedly. “Does Habitat Fragmentation Reduce Genetic Diversity and Subpopulation Connectivity?” *Ecography* 31, no. 6 (2008): 751–56. https://doi.org/10.1111/j.1600-0587.2008.05622.x.

Milano, Elizabeth R., Margaret R. Mulligan, Jon P. Rebman, and Amy G. Vandergast. “High-Throughput Sequencing Reveals Distinct Regional Genetic Structure among Remaining Populations of an Endangered Salt Marsh Plant in California.” *Conservation Genetics* 21, no. 3 (June 1, 2020): 547–59. https://doi.org/10.1007/s10592-020-01269-3.

Nazareno, Alison G., and L. Lacey Knowles. “There Is No ‘Rule of Thumb’: Genomic Filter Settings for a Small Plant Population to Obtain Unbiased Gene Flow Estimates.” *Frontiers in Plant Science* 12 (October 14, 2021). https://doi.org/10.3389/fpls.2021.677009.

Nic Lughadha, Eimear, Steven P. Bachman, Tarciso C. C. Leão, Félix Forest, John M. Halley, Justin Moat, Carmen Acedo, et al. “Extinction Risk and Threats to Plants and Fungi.” *PLANTS, PEOPLE, PLANET* 2, no. 5 (2020): 389–408. https://doi.org/10.1002/ppp3.10146.

Paris, Josephine R., Jamie R. Stevens, and Julian M. Catchen. “Lost in Parameter Space: A Road Map for Stacks.” *Methods in Ecology and Evolution* 8, no. 10 (2017): 1360–73. https://doi.org/10.1111/2041-210X.12775.

Petkova, Desislava, John Novembre, and Matthew Stephens. “Visualizing Spatial Population Structure with Estimated Effective Migration Surfaces.” *Nature Genetics* 48, no. 1 (January 2016): 94–100. https://doi.org/10.1038/ng.3464.

Pornon, André, Christophe Andalo, Monique Burrus, and Nathalie Escaravage. “DNA Metabarcoding Data Unveils Invisible Pollination Networks.” *Scientific Reports* 7, no. 1 (December 4, 2017): 16828. https://doi.org/10.1038/s41598-017-16785-5.

Segelbacher, G., Cushman, S. A., Epperson, B. K., Fortin, M. J., Francois, O., Hardy, O. J., ... & Manel, S. (2010). Applications of landscape genetics in conservation biology: concepts and challenges. Conservation genetics, 11, 375-385.

Slatkin, Montgomery, and Naoyuki Takahata. “The Average Frequency of Private Alleles in a Partially Isolated Population.” *Theoretical Population Biology* 28, no. 3 (December 1, 1985): 314–31. https://doi.org/10.1016/0040-5809(85)90032-2.

Sopniewski, Jarrod, and Renee A. Catullo. “Estimates of Heterozygosity from Single Nucleotide Polymorphism Markers Are Context-Dependent and Often Wrong.” *Molecular Ecology Resources* 24, no. 4 (2024): e13947. https://doi.org/10.1111/1755-0998.13947

Souther, S., Lechowicz, M. J., & McGraw, J. B. (2012). Experimental test for adaptive differentiation of ginseng populations reveals complex response to temperature. *Annals of Botany*, *110*(4), 829-837.

Souther, S., & McGraw, J. B. (2014). Synergistic effects of climate change and harvest on extinction risk of American ginseng. *Ecological Applications*, *24*(6), 1463-1477.

Souther, S., & McGraw, J. B. (2011). Evidence of local adaptation in the demographic response of American ginseng to interannual temperature variation. *Conservation Biology*, *25*(5), 922-931.

Souther, Sara, James B. McGraw, John D. Souther, and Donald M. Waller. "Effects of altered climates on American ginseng population dynamics." *Population Ecology* 64, no. 1 (2022): 47-63.

Souther, Sara, Martha Sample, Genevieve Conley, and Clare Aslan. “Demographic Analysis of the Endangered Plant Pectis Imberbis Highlights Tradeoffs between Deer Browse and Interspecific Competition.” *Natural Areas Journal* 42, no. 3 (July 2022): 230–41. https://doi.org/10.3375/22-3.

Souther, S., Sample, M., and Aslan, C. (2024). Interacting stressors drive landscape variation in demographic response of the endangered plant, Pectis imberbis (Gray). *Journal of Arid Environments* in press.

Tackenberg, Oliver, Peter Poschlod, and Susanne Bonn. “Assessment of Wind Dispersal Potential in Plant Species.” *Ecological Monographs* 73, no. 2 (2003): 191–205. https://doi.org/10.1890/0012-9615(2003)073[0191:AOWDPI]2.0.CO;2.

Truelove, Nathan K., Kim Ley-Cooper, Iris Segura-García, Patricia Briones-Fourzán, Enrique Lozano-Álvarez, Bruce F. Phillips, Stephen J. Box, and Richard F. Preziosi. “Genetic Analysis Reveals Temporal Population Structure in Caribbean Spiny Lobster (*Panulirus Argus*) within Marine Protected Areas in Mexico.” *Fisheries Research* 172 (December 1, 2015): 44–49. https://doi.org/10.1016/j.fishres.2015.05.029.

U.S. Fish and Wildlife Service. 2023. Beardless Chinchweed (Pectis imberbis) Final Recovery Plan. Southwest Region, Albuquerque, NM, USA. 37 pp.

Warzecha, Daniela, Tim Diekötter, Volkmar Wolters, and Frank Jauker. “Intraspecific Body Size Increases with Habitat Fragmentation in Wild Bee Pollinators.” *Landscape Ecology* 31, no. 7 (September 1, 2016): 1449–55. https://doi.org/10.1007/s10980-016-0349-y.

Wang, Z.-F., Lian, J.-Y., Ye, W.-H., Cao, H.-L., Zhang, Q.-M., and Wang, Z.-M. (2016). Pollen and seed flow under different predominant winds in wind-pollinated and wind-dispersed species Engelhardia roxburghiana. *Tree Genetics & Genomes* 12, 19. doi: 10.1007/s11295-016-0973-3

Wilson, N.R., 2024, Species Distribution Models for Pectis imberbis, a Rare Plant Species in Southeastern Arizona: U.S. Geological Survey data release, https://doi.org/10.5066/P13VMRBC.

Willi, Yvonne, Marco Fracassetti, Olivier Bachmann, and Josh Van Buskirk. “Demographic Processes Linked to Genetic Diversity and Positive Selection across a

Species' Range." *Plant Communications* 1, no. 6 (November 2020): 100111. https://doi.org/10.1016/j.xplc.2020.100111.

Wright, S. "The Genetical Structure of Populations." *Annals of Eugenics* 15, no. 4 (March 1951): 323–54. https://doi.org/10.1111/j.1469-1809.1949.tb02451.x.

Wright, Sewall. "The Interpretation of Population Structure by F-Statistics with Special Regard to Systems of Mating." *Evolution* 19, no. 3 (1965): 395–420. https://doi.org/10.2307/2406450.

Yanahan, A. D., & Moore, W. (2019). Impacts of 21st-century climate change on montane habitat in the Madrean Sky Island Archipelago. *Diversity and Distributions*, *25*(10), 1625–1638. https://doi.org/10.1111/ddi.12965

Zhou, Chengchuan, Shiqi Xia, Qiang Wen, Ying Song, Quanquan Jia, Tian Wang, Liting Liu, and Tianlin Ouyang. "Genetic Structure of an Endangered Species Ormosia Henryi in Southern China, and Implications for Conservation." *BMC Plant Biology* 23, no. 1 (April 26, 2023): 220. https://doi.org/10.1186/s12870-023-04231-w.

APPENDIX A.

Results of the Species Distribution Model for *P. imberbis*. The complete methodology and results from SDM modeling is found here: Species Distribution Models for Pectis imberbis, a Rare Plant Species in Southeastern Arizona - ScienceBase-Catalog.

Ensemble Species Distribution Modeling for *Pectis imberbis*, an Endangered Plant Species

Natalie R. Wilson[1]

Affiliations:
1: U.S. Geological Survey, Western Geographic Science Center, 520 N. Park Ave, Tucson, AZ 85719, USA

Corresponding Author: Natalie R. Wilson, nrwilson@usgs.gov

Abstract

Species distribution models (SDMs) can be used in rare species conservation to locate previously unknown populations and identify sites for reintroduction or translocation. We applied SDM to a recently listed plant species, *Pectis imberbis*, which is found in the Madrean Archipelago region of southern Arizona, United States, and northern Mexico. We developed 4 different predictor datasets using correlation analysis, feature selection, and variable importance analysis. We used presence-pseudoabsence data and applied 5 different modeling algorithms. The resulting models were evaluated based on k-fold cross validation. High performing individual models were then included in ensemble model creation using the ensemble mean, ensemble median, and committee averaging methods. Ensemble models were consistent in predictions across known populations of *P. imberbis*, and all ensemble models identified areas beyond currently occupied habitat which could be ecologically suitable for this rare species. We evaluated these ensemble models based on calibration metrics, intra-model variance, consistency across known populations, and data coverage, and chose the model best suited to meet the stated conservation goals. Our results show that the model that performs the best is the committee averaging method based binary component models created with a threshold optimized for the area under the curve of the receiver operating characteristic curve calibration metric and based on 5 predictor variables: spring precipitation, elevation, solar radiation, U.S. Environmental Protection Agency ecoregion, and U.S. Geological Survey geologic data.

Introduction

Pectis imberbis A. Gray (Figure 1) is an herbaceous perennial member of the Heliantheae tribe in the Asteraceae family. It was recently listed as endangered under the Endangered Species Act (U.S. Fish and Wildlife Service 2021). The historical range for *P. imberbis* extends from southeastern Arizona, United States (U.S.), into northern sections of Sonora and Chihuahua, Mexico. However, within this range, populations are naturally fragmented, being constrained to disjointed populations in the mid-elevation grasslands, oak savannas, oak woodlands, and pine oak juniper woodlands on the slopes of the mountain ranges of the Madrean Archipelago ecoregion (U.S. Fish and Wildlife Service 2020). These populations have been in decline, with several populations extirpated in both the U.S. and Mexico (Sánchez Escalante et al. 2019; U.S. Fish and Wildlife Service 2020). These declines have been attributed to increased drought frequency and severity, competition from non-native plants, disturbance from road maintenance, livestock grazing, and wild ungulate (Coues deer; *Odocoileus virginianus* couesi Coues and Yarrow 1875) grazing (Phillips III et al. 1982; U.S. Fish and Wildlife Service 2020; Souther et al. 2022). The U.S. Fish and Wildlife Service (2020) declared that *P. imberbis* "needs multiple resilient

populations distributed widely across its range to maintain its persistence into the future and avoid extinction," and suggested several measures for the conservation of *P. imberbis*, including locating additional plants and establishing new populations at appropriate sites. These two measures can be assisted through the application of species distribution modeling.

Species distribution models (SDMs) combine known species presence data with environmental data to explore ecological questions, inform conservation efforts, and support land management decisions

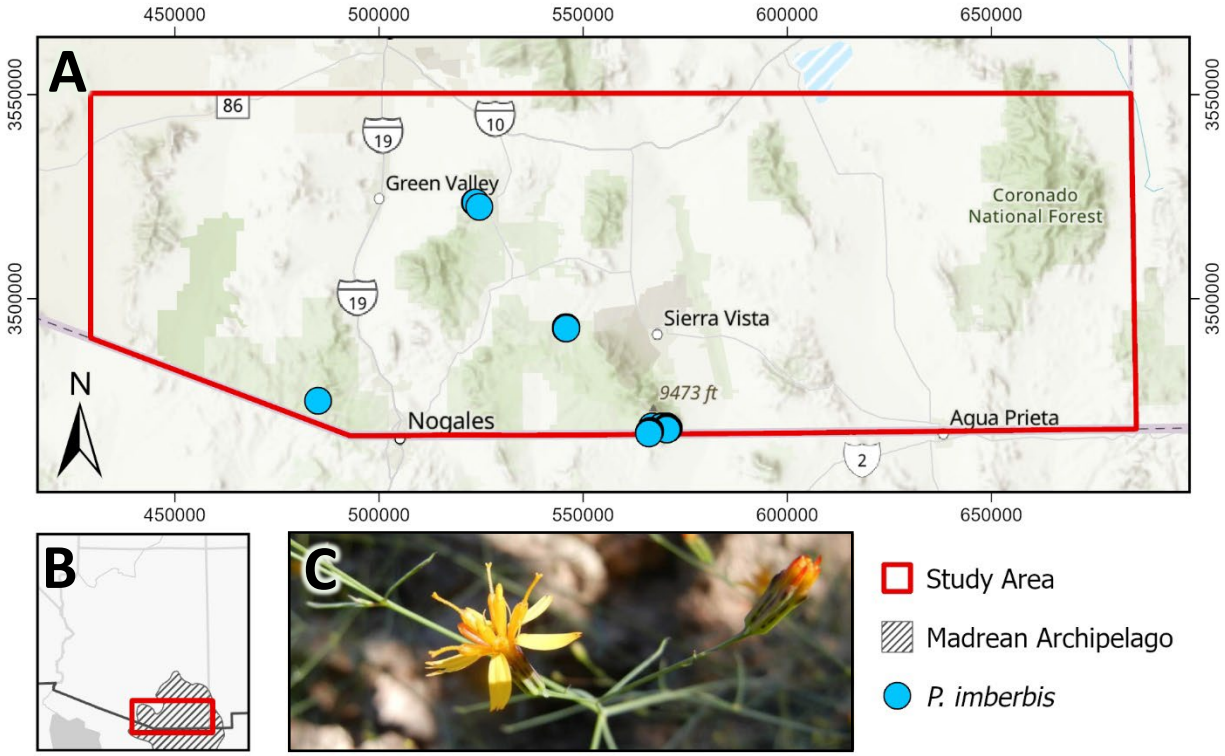


Figure 1. A: Study area for P. imberbis species distribution models and known populations. Projection: NAD83 UTM 12N. B: Location of study area in Arizona and extent of Madrean Archipelago. C: P. imberbis in flower (photograph by Teague Embrey, accessed through SEINet Portal Network 2024). Basemaps by ESRI (ESRI 2024a, 2024b).

(Pecchi et al. 2019; Sillero et al. 2021). Ensemble SDMs combine multiple single models to reduce the impact of individual model errors and capture a more robust prediction of the overall central tendency (Hao et al. 2019; Hysen et al. 2022; Sillero et al. 2021). Ensemble and single SDMs have been used to predict the impact of anthropogenic disturbances on species distribution, analyze the impact of climate change on species distribution, assess species richness across a landscape, find previously unknown populations of rare species, and identify areas for translocation (Manel et al. 2001; Rathore & Sharma 2023). The latter two applications may apply to the conservation of *P. imberbis*. SDMs have repeatedly aided in locating new populations of rare plant species across taxa and habitats for over a decade (Williams et al. 2009; McCune 2016; Borokini et al. 2023). Despite the frequent use of SDMs to guide reintroductions and translocations of rare species (Tomlinson et al. 2020; Meyer et al. 2022; Rusconi et al. 2022; Elliott et al. 2024), the effectiveness of these applications has not been adequately evaluated (Questad et al. 2014; Draper et al. 2019). This lack of evaluation likely results a from a general absence of long-term monitoring data and not from shortcomings of SDMs (Ruiz-Jaen & Mitchell Aide 2005; Bellis et al. 2024).

Here we present the development of SDMs for the endangered species *P. imberbis*. This research can be used to inform endangered species conservation by land management agencies, particularly for locating previously unknown populations and identifying general locations for reintroductions or translocations. This research also explores the efficacy of different algorithms for species distribution modeling.

Methods

Study Area

Historical populations of *P. imberbis* spanned a broad area, extending 60 km north of the international border into southeastern Arizona and over 400 km south into eastern Sonora and western Chihuahua, Mexico (U.S. Fish and Wildlife Service 2020). Much of this extent overlaps with the Madrean Archipelago ecoregion, which is globally recognized for its biodiversity and productivity (DeBano & Ffolliott 1994; López-Hoffman & Quijada-Mascareñas 2012; Devender et al. 2012). The Madrean Archipelago is characterized by tall mountain ranges separated by low valleys, creating a unique assemblage of vegetation communities (Baker Jr et al. 1994; Felger & Wilson 1994) in which *P. imberbis* inhabits mid-elevation grasslands, savannas, and woodlands (U.S. Fish and Wildlife Service 2020). For model development, we extended our study area from the western edge of the Madrean Archipelago to the eastern Arizona border and 25 km north of the northern-most known *P. imberbis* occurrence (Figure 1). We constrained our study area to north of the international border due to management agency jurisdiction limits and the lack of cohesive predictor data extending into Mexico. Within our study area, extant populations of P. imberbis exist in 4 areas (U.S. Fish and Wildlife Service 2020; Souther et al. 2022). The largest and best studied populations occur within and west of Coronado National Memorial in the southern Huachuca Mountains. There is one population in the Canelo Hills, which are between the Huachuca Mountains and the Patagonia Mountains. There are a few populations in the northern Santa Rita Mountains, and one population in the southern Atascosa Mountains.

Presence Data

Presence data were based on field observations from 2020 to 2022 provided by collaborators at Northern Arizona University (NAU) (M. Sample, personal communication, 11/21/2023). We filtered points and polygons to one observation per 30 m^2 raster cell resulting in 123 presence points (Sillero et al. 2021).

Background Point Data

P. imberbis is a cryptic species that is challenging to detect without species-specific training, systematic sampling, and sampling sensitive to the plant's phenology, making confirmed absence observations difficult (National Park Service 2014). For species without adequate absence data, background points (also referred to as pseudo-absence points) can be used to characterize the study area for SDM (Li et al. 2011; Barbet-Massin et al. 2012; Hysen et al. 2022). Model algorithms can be sensitive to the ratio of background points to presence points (i.e., prevalence) and require specific ratios for optimal performance (Barbet-Massin et al. 2012; Hysen et al. 2022). We created different background point datasets for predictor variable selection, initial model testing, and final modeling runs. Predictor variable selection methods are not sensitive to prevalence like modelling algorithms, so we created a large set of background points in ArcGIS Pro (ESRI 2023) by randomly distributing 10,000 points in the study area with a minimum separation of 100 m between each background point and an exclusion buffer of 300 m around presence points. For initial model testing, we created two sets of background points for different

predictor variable sets using the biomod2 package (Thuiller et al. 2024) in R (R Core Team 2024). Each set consisted of 40 sets of 250 randomly distributed points which we then combined into sets of 500, 1,000, 2,000, and 5,000 points as well as a full set of 10,000 points. Initial model testing consisted of exploring different model parameters such as different algorithms, validation techniques, and replicate numbers to optimize coding and test computation time prior to the final modeling runs. For final modeling runs, we used the sets of 250, 500, 5000 and 10000 created for initial model testing.

Predictors and Feature Selection

We assembled a total of 43 predictor variables consisting of topographic, edaphic, geologic, climatic, vegetative, land cover, and remote sensing data, all of which we considered to be potentially influential for characterizing suitable *P. imberbis* habitat (Table 1). All predictor variable data were converted to raster data, clipped to the area of interest, resampled to 30m, projected to NAD83 UTM 12N, and snapped to the elevation dataset grid in ArcGIS Pro. Although many predictor variables were initially considered, a subset of variables was selected to avoid overfitting and limit complications during modeling due to collinearity between variables (Fielding & Bell 1997; Dormann et al. 2013; Sillero et al. 2021). We used correlation analysis, zero and near-zero variance analysis, linear dependency analysis, and feature selection via the R packages caret (Kuhn 2023), ggcorrplot (Kassambara 2023), and Boruta (Kursa & Rudnicki 2022) for this task. This first round of predictor variable selection was based on presence data and a set of background points developed specifically for feature selection. After this first round of predictor variable selection, two geologic variables were still included: Southwest Regional Gap Analysis Project surficial geological formation (SWRG; RS/GIS Laboratory, College of Natural Resources, Utah State University 2004a) and USGS State Geologic Map Compilation geological unit (USGS; Horton 2017). These geological variables were similar in their correlation with other variables and of similar importance; neither was clearly superior. To explore which geological data was better for our analysis, separate predictor datasets were formed: one with USGS data and one with SWRG data. This resulted in 2 sets of 10-predictor variables (Table 1): eight continuous variables which were spring precipitation, minimum temperature, elevation, solar radiation, available water storage, soil mean pH, soil mean organic matter, tassel cap brightness for summer, and two categorical variables, EPA level IV ecoregion, US Environmental Protection Agency 2013, and either USGS geology or SWRG geology. After initial model testing, these variables were further reduced to 5 predictors via variable importance analysis in the biomod2 package in R: spring precipitation, elevation, solar radiation, EPA ecoregion, and either USGS or SWRG geologic data. Final modeling runs were performed on a total of 4 predictor datasets (Table 1): 10-predictor and 5-predictor variable datasets, with either SWRG or USGS geologic data included.

Single Model Building

Ensemble modeling aggregates multiple single models which are created using a wide variety of model algorithms (Pecchi et al. 2019; Valavi et al. 2022; Borokini et al. 2023). The biomod2 package uses two broad categories of algorithms: regression-based algorithms and classification or tree-based algorithms. Of the regression-based algorithms, we used generalized linear models (GLM) and generalized additive models (GAM). From the classification-based algorithms, we chose boosted regression trees (also referred to as generalized boosting model [GBM]), random forests (RF), and classification tree analysis (CTA). We also included multiple adaptive regression splines (MARS) and maximum entropy (Maxent) in the initial modeling phase, but they were excluded from the main analysis. We excluded MARS because of long processing times, even when parallel processing was forced, and we excluded Maxent because it did not run when parallel processing was used. We used different sets of background points for the different categories to maximize model performance (Barbet-Massin et al. 2012; Valavi et al. 2022). For

the classification-based algorithms, we used 40 sets of 250 background points and 20 sets of 500 background points. For the regression-based algorithms we used 2 sets of 5000 background points and 1 set of the full 10,000 background points.

Due to the limited amount of presence data, we did not retain a fully independent set of evaluation data apart from calibration data. Instead, we used k-fold cross validation (Fielding & Bell 1997; Borokini et al. 2023; Sillero et al. 2021). In k-fold cross-validation with 5 folds, the data are split into 5 independent partitions. For one run, 4 partitions are used to calibrate the model while the remaining partition is used to evaluate the model. This split is repeated 5 times, allowing each partition to be used to evaluate a model calibrated on the other partitions. We created 10 replicates of 5-fold cross validation partitions per presence-background dataset. We used these to fit 10 replicates of each of the 5 model algorithms over the presence-background datasets. Ideally, this would result in 50 models per algorithm per presence-background dataset, or 9300 models per predictor dataset.

After the models were created, we evaluated them for inclusion in ensemble model development. Evaluation metrics differ in a variety of ways (Lawson et al. 2014; Jarnevich et al. 2015); they can be threshold dependent or independent, they can have different sensitivities to prevalence, and they can measure model performance in terms of discrimination, the ability to discriminate between presence and absence locations, or calibration, the ability of a model to correctly predict the presence of a species. Often, the best approach to model evaluation is to use multiple evaluation metrics (Jarnevich et al. 2015; Sillero et al. 2021; Soley-Guardia et al. 2024). For the evaluation of single models, we calculated 4 metrics: Cohen's kappa statistic (K, Fielding & Bell 1997), the area under the curve of the receiver operating characteristic curve (ROC, Fielding & Bell 1997), the True Skill Statistic (TSS, Allouche et al. 2006), and the Boyce Index (BI, Boyce et al. 2002). We also compared the sensitivity to specificity of models during evaluation. Our initial selection criteria consisted of BI and K scores > 0.8 and a model sensitivity that was greater than the model's specificity. Because computation time was a concern, we limited the number of models per algorithm to 100 for each predictor dataset. If there were more than 100 models that met our initial selection criteria, we increased our required BI and K scores until the number of models selected was close to but not below 100. From this we randomly selected 100 models for ensemble model creation.

All model fitting, cross validation, and model evaluation was performed using the biomod2 package in R. Models were fitted with the biomod2 "bigboss" model parameters. Other packages used included terra (Hijmans 2024), ggplot2 (Wickham et al. 2024), and doParallel (Daniel et al. 2022). All scripting was done in RStudio (Posit Team 2024).

Ensemble Model Building

Ensemble models combine the predictions from multiple single models in multiple ways (Araujo & New 2007; Marmion et al. 2009) and 4 ensemble modelling methods are included in the biomod2 package in R. Ensemble mean and ensemble median apply the stated statistic to the continuous prediction outputs of the component models. Ensemble weighted mean weighs the component models based on the quality of a model, as determined by an evaluation metric that mean. *C*ommittee averaging, applies a threshold to the continuous output of the individual component models to create a binary prediction and then sums these binary predictions. We did not use the weighted mean method because the high quality of the chosen models would have resulted in little difference between model weights. We did apply the ensemble mean, ensemble median, and committee averaging model methods. Because we ran committee averaging with thresholds based on all four of the single model evaluation metrics, we

generated 6 ensembles per predictor dataset: mean, median, and four committee averaging results. We also computed the coefficient of variation for the model components of each predictor dataset. We combined models across all background point sets and cross validation folds. No data were held back for evaluation or validation; this will result in overestimations of model performance during ensemble model evaluation. We used K, ROC, and TSS metrics to evaluate the ensemble model predictions. All ensemble model building and evaluation was performed using the biomod2 package in R.

Results

Single Model Evaluation

During the single model fitting, several GLM and GAM models failed for each predictor variable dataset (Table 2) due to k-fold partitions of background points not capturing adequate data for analysis of the categorical variables, particularly classes with a small areal extent. Model validation metrics were variable across predictor datasets and among themselves (Figure 2). ROC and TSS were so high across predictor datasets and algorithms that they were uninformative as metrics. K was informative for regression-based algorithms, GLM and GAM, but not for the classification algorithms. BI had the widest range across algorithms and predictor datasets and therefore had the most impact on model selection for inclusion in ensemble model building. After applying the criteria that BI and K must be greater than 0.8 of the regression-based algorithms, 15-20% of GLM models and 0 GAM models remained. The classification-based algorithms (GBM, RF, CTA) were all similarly affected by the BI, K criteria with 35-50% of models remaining. The criteria that sensitivity must be greater than specificity further reduced model numbers. GLM was reduced to 0-2% of the original fitted models, GBM was reduced to 1-10%, RF reduced to 8-25%, and CTA reduced to 19-40%. For all predictor datasets, RF and CTA were further reduced by increasing the BI and K criteria until the number of models remaining was close to but not below 100. From these, 100 models were randomly selected for inclusion in ensemble model building. GMB models were similarly reduced for one predictor dataset (SWRG-10v).

Ensemble Model Evaluation

For each of the 4 predictor datasets, 6 ensemble methods were applied: ensemble mean, ensemble median, and committee averaging with thresholds based on BI, K, TSS, or ROC. This resulted in 24 ensemble models overall. These 24 ensemble models were evaluated with K, TSS, and ROC (Figure 3); BI was not available for ensemble model evaluation in the biomod2 package. These metrics were applied to the calibration data with no data held back for validation or evaluation. This must be kept in mind when considering the following comparisons. These metrics were expected to be overestimated, and all were high. However, similar to the single model evaluations, TSS and ROC were particularly high, and the low variance between ensemble model outputs made these metrics uninformative for model evaluation. K values were also high but there was some variance between models. K indicates that the 10-predictor variable datasets (-10v) performed slightly better than the 5-predictor datasets created using variable importance analysis (-5v-5v). K also suggests that there was little difference between the performance of the models using the USGS geologic data and the models using SWRG geologic data, despite the difference in number of classes for each variable (26 and 8, respectively). However, there was a difference in performance between ensemble model methods, with ensemble median consistently underperforming compared to the others.

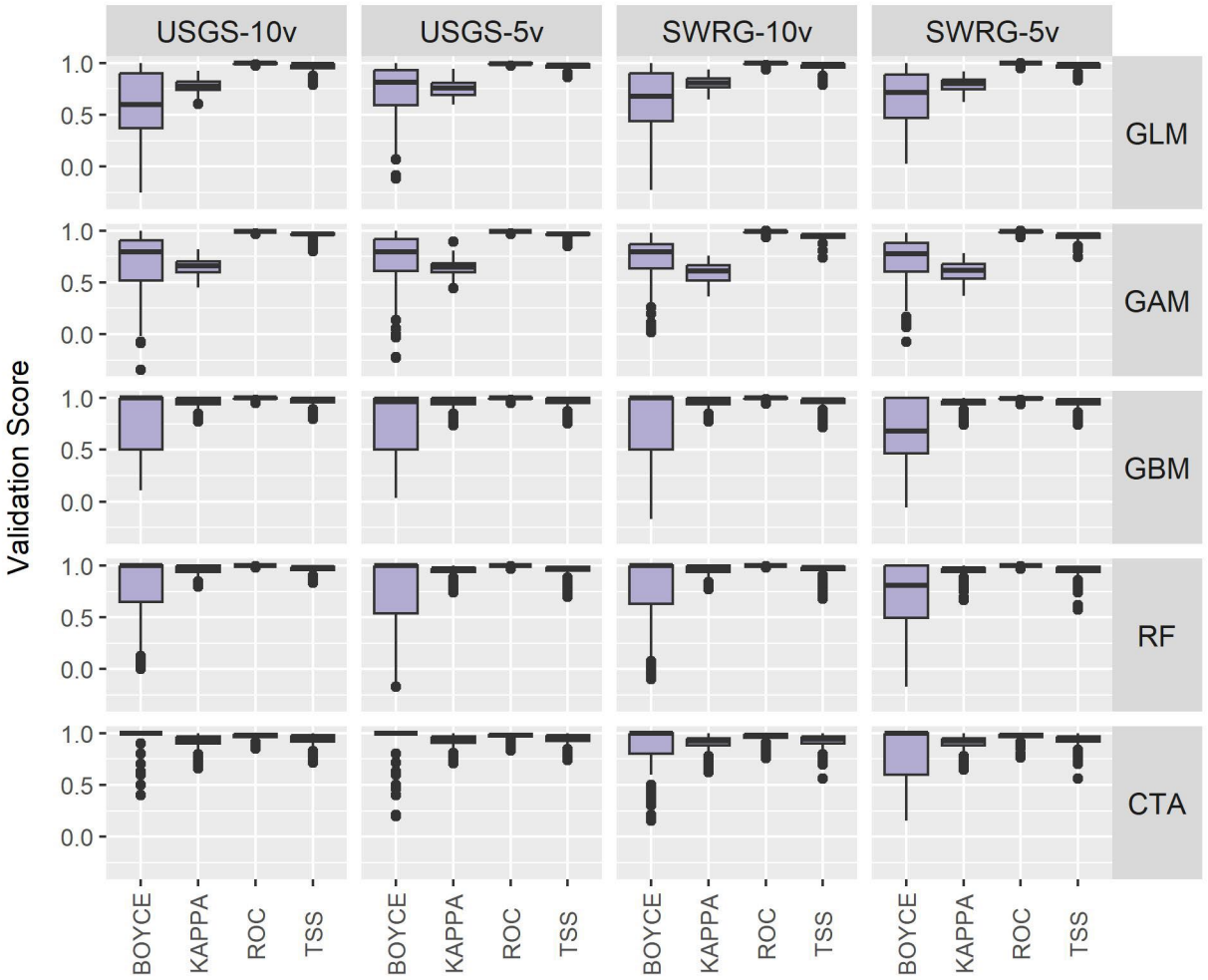


Figure 2: Evaluation metrics for models based on model algorithm (y-axis, right) and predictor dataset (x-axis, top). Each evaluation metric (x-axis, bottom) was applied to k-fold (here, k=5) cross-validation data. Algorithms are generalized linear model (GLM), generalized additive model (GAM), generalized boosted model (GBM), random forest (RF), and classification tree analysis (CTA). Predictor datasets are the 10-predictor dataset with USGS geologic data (USGS-10v), the 5-predictor dataset with USGS geologic data (USGS-5v), the 10-predictor dataset with SWRG geologic data (SWRG-10v), and the 5-predictor dataset with SWRG geologic data (SWRG-5v). Evaluation metrics are Cohen's kappa statistic (KAPPA), the area under the receiver operating characteristic curve (ROC), the True Skill Statistic (TSS), and the Boyce Index (BOYCE).

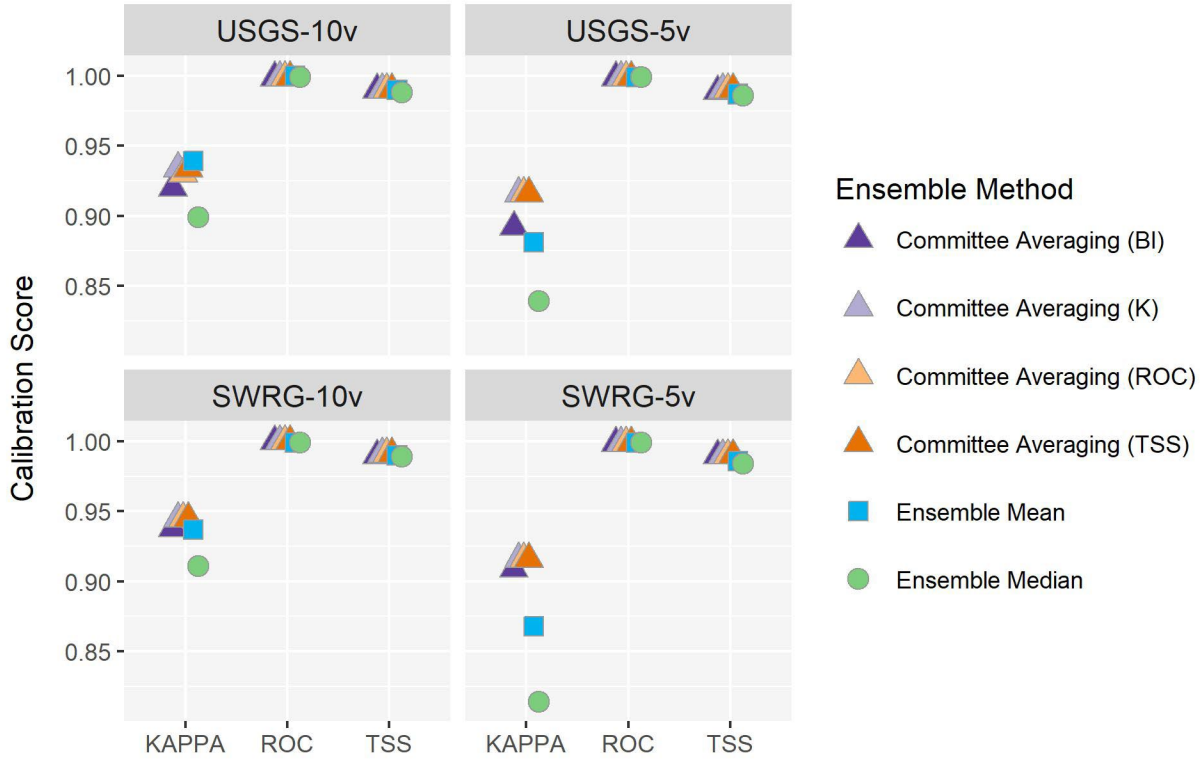


Figure 3. Evaluation metrics for the 6 ensemble models by predictor dataset. Algorithms are committee averaging of models with a threshold applied which was optimized for the Boyce Index (BI), committee averaging of models with a threshold applied which was optimized for Cohen's kappa statistic (K), committee averaging of models with a threshold applied which was optimized for the area under the receiver operating characteristic curve (ROC), committee averaging of models with a threshold applied which was optimized for the True Skill Statistic (TSS), ensemble mean, and ensemble median. Predictor datasets are the 10-predictor dataset with USGS geologic data (USGS-10v), the 5-predictor dataset with USGS geologic data (USGS-5v), the 10-predictor dataset with SWRG geologic data (SWRG-10v), and the 5-predictor dataset with SWRG geologic data (SWRG-5v). Evaluation metrics are Cohen's kappa statistic (KAPPA), the area under the receiver operating characteristic curve (ROC), the True Skill Statistic (TSS).

Geographic Extent

For each of the remaining ensemble models, we created a presence/absence binary raster based on the threshold value that optimized each evaluation metric. This step created 3 binary models for each continuous model, resulting in 72 binary models. For each predictor dataset, we identified the binary model with the minimal and maximal predicted extent of *P. imberbis*, which we present in Figure 4 with the corresponding continuous models in Figure 5. The smallest extent from all ensemble models was 2,204 hectares predicted by the ensemble mean model using the 10-predictor dataset with the USGS geologic data (USGS-10v) with the threshold optimized for K applied (USGS-10v_EMmean_binKAPPA). The largest extent from all models was 35,505 hectares predicted by the ensemble median model using the 5-predictor dataset with the SWRG geologic data (SWRG-5v) with the threshold optimized for TSS applied (SWRG-5v_EMmedian_binTSS). The 10-predictor dataset models have smaller predicted extents because of areas of missing data due to the lack of data for available water storage, an edaphic variable (Figures 4 & 5).

Most known populations of *P. imberbis* were consistently identified correctly across all 72 ensemble model binary rasters (Figure 6). All populations consisted of multiple pixels, which created for a range of predicted presence for each population. In the Huachuca Mountains, populations in Coronado National Memorial were consistently identified as presence, while 2 populations were identified as 40–80% presence by 5 models and as 100% absence by 2 models. The populations to the west of Coronado National Memorial in the Huachuca Mountains were both correctly identified as 100% presence for all models. Of the two populations in the northern Santa Rita Mountains, one was consistently identified as 100% presence by all ensemble models while the other was identified as less than 100% presence by 3 models and as complete absence by 1 model. The one population in the Canelo Hills was identified as less than 100% presence by 5 models and as total absence by 2 models. There was one population, in the Atascosa Mountains, which despite being used as training data, was consistently identified as total absence in all ensemble models.

All ensemble model binary rasters identified predicted areas of presence for *P. imberbis* beyond current known populations, indicating additional areas of ecological suitable habitat for this rare species (Figures 7 & 8). In Coronado National Memorial, the areas between known populations were identified as suitable for *P. imberbis* by 90% of the ensemble model binary rasters and this habitat extended along the lower south, southeast, and east-facing slopes of Montezuma Peak. Additionally, the southwestern slopes of Huachuca Mountains from the Mexican border and Yaqui Canyon to the Lone Mountain were consistently identified as suitable habitat by more than 90% of the models. In the Canelo Hills, the area identified as suitable habitat by 90% of the models extended along the drainages northwest and southeast of the known population. In the northern Santa Rita Mountains, the areas between the known populations in McCleary Canyon and Wasp Canyon, are identified as good habitat by over 90% of the models. Other areas identified as suitable habitat by over 90% of the ensemble model binary rasters were along the southeast-facing slopes of Enzenberg Canyon, the southwestern-facing slopes of upper Fish Canyon, and the south-facing slopes of Mansfield Canyon. Although there is one extant population in the Atascosa Mountains, ensemble model predictions of suitable habitat had a high degree of spatial variability and there were no areas where more than 50% of models agreed.

Ensemble model binary rasters also identified areas of suitable habitat in areas with no current or historical populations of *P. imberbis*; however, these areas tended to be less consistent than those in mountain ranges where there are extant populations (Figures 7 & 8). The Dragoon Mountains had areas identified as suitable habitat by over 90% of binary models, particularly the south-facing slopes of Middlemarch Canyon in the southern end of the range, and Fourr Canyon in the northern end. Areas identified in over 50% of models extended from the upper south-facing slopes of Stronghold Canyon West, across Cochise Stronghold, to the south-facing slopes between Stronghold Canyon East, Grapevine Canyon, and Noonan Canyon. In the Sierrita Mountains, areas identified as suitable habitat by over 90% of models included the south-facing slopes of upper Tascuela Canyon and the south-facing slopes above upper Ash Creek. Areas identified in over 50% of models included the southwest-facing slopes of Samaniego Peak and the south-facing slopes of drainages along Ox Frame Canyon. While the Chiricahua Mountains, the Whetstone Mountains, The Baboquivari Mountains, the area of the Rincon Mountains included in the study area, and the Patagonia Mountains (excluding the border with Canelo Hills) did have areas identified as suitable habitat by individual ensemble model binary rasters, the areas were varied with limited areas consistent across half the models.

We also analyzed the coefficient of variation (CV) for the component models used for each predictor dataset (Figure 9). CV compares the predictions of component models to each other; a low CV indicates agreement between models while a high CV indicates disagreement (Figure 9).

All ensemble model outputs, both continuous and binary, as well as coefficient of variation rasters are available in Wilson (2024).

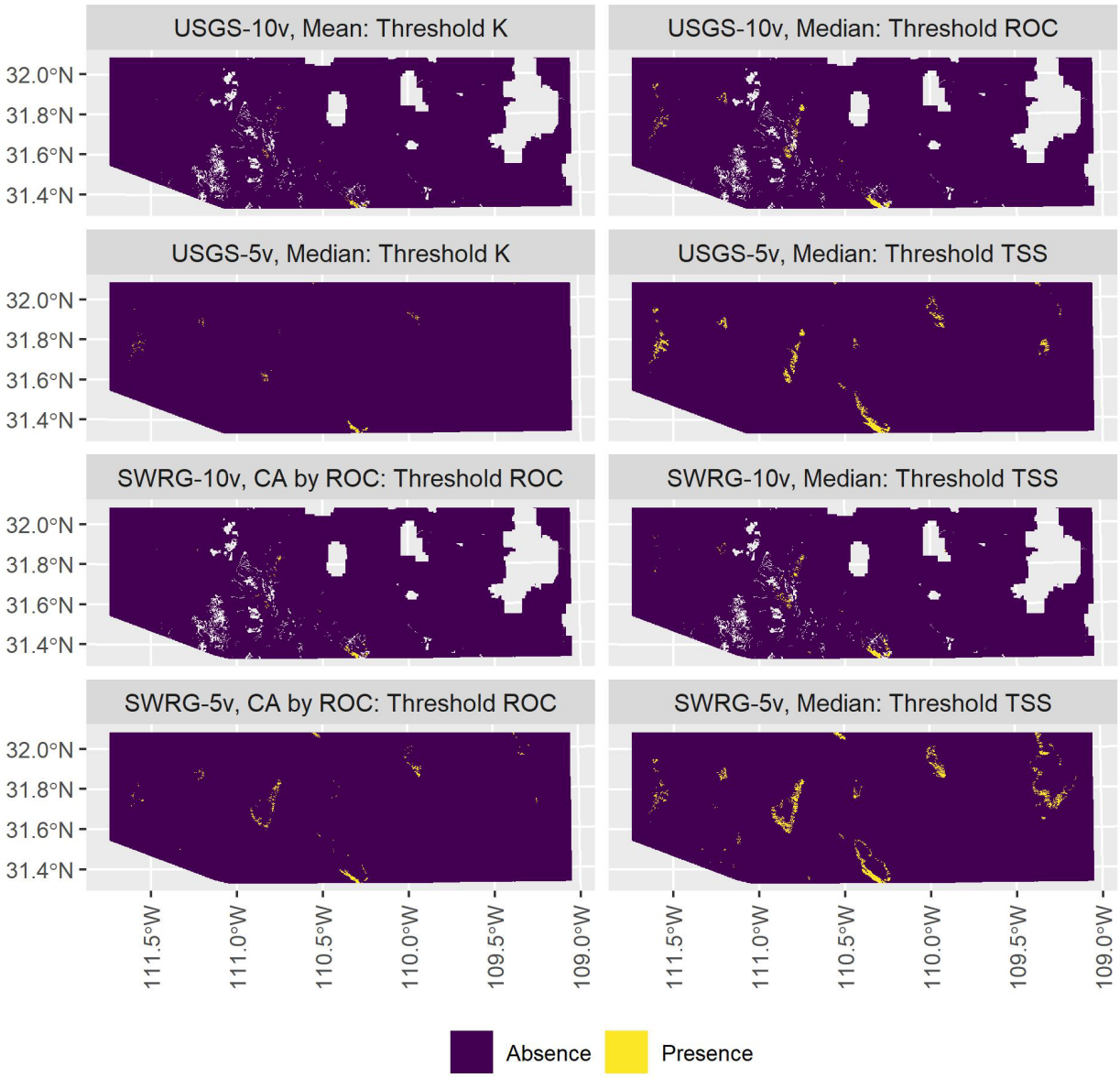


Figure 4. Minimal (left column) and maximal (right column) binary model (presence/absence) predicted extents of P. imberbis for each predictor dataset (USGS-10v, USGS-5v, SWRG-10v, and SWGR-5v). Ensemble methods included committee averaging (CA), ensemble mean, and ensemble median. Large patches of missing data in the first and third rows are the result of missing data in the available water storage soil data in the 10-variable predictor datasets. Datasets and threshold metrics abbreviations same as Fig. 2 & 3.

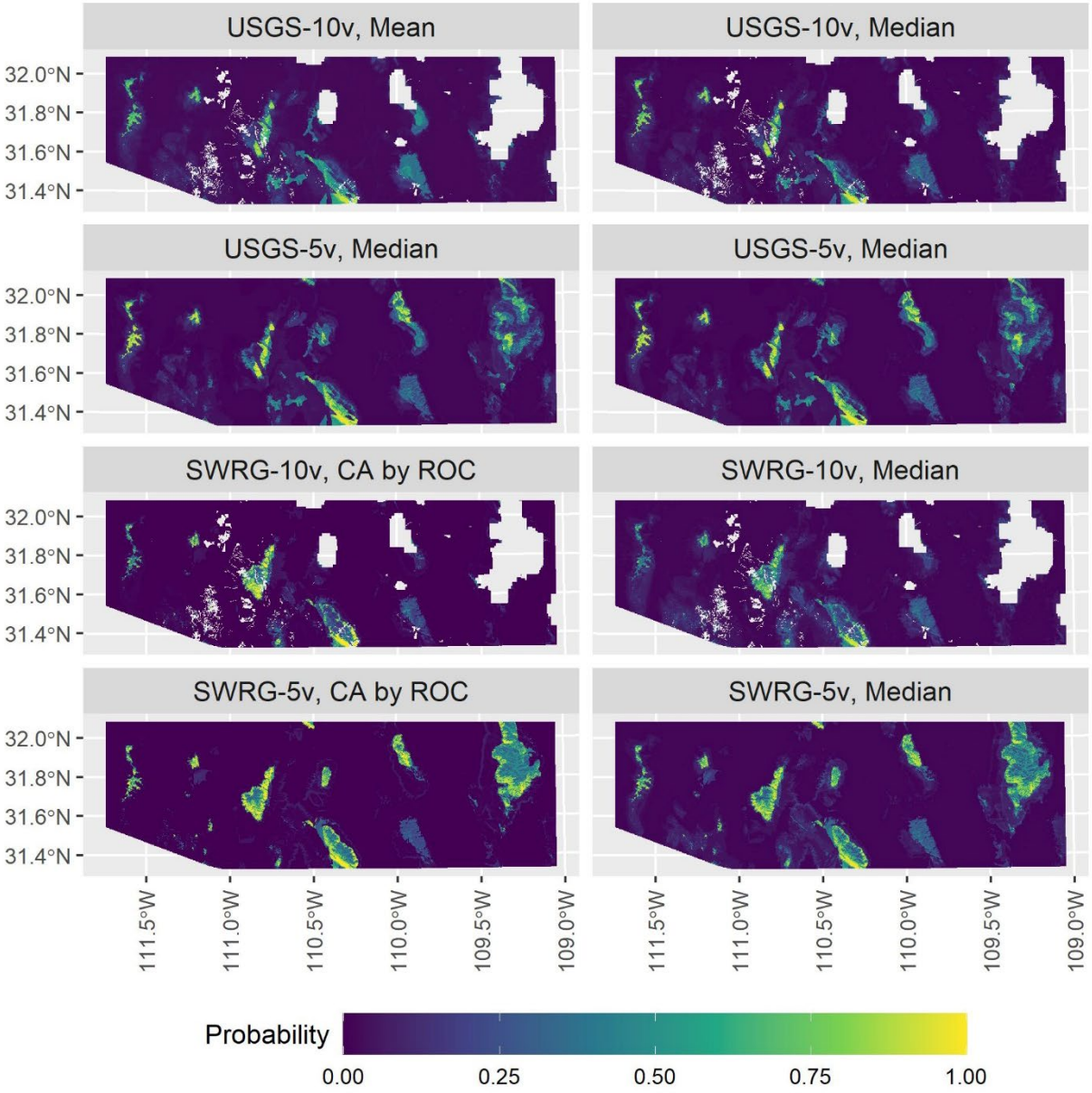


Figure 5. Continuous models corresponding to the binary models in Figure 4. Ensemble methods included committee averaging (CA), ensemble mean, and ensemble median. Large patches of missing data in the first and third rows are the result of missing data in the available water storage soil data in the 10-variable predictor datasets. Datasets and threshold metrics abbreviations same as Fig. 2 & 3.

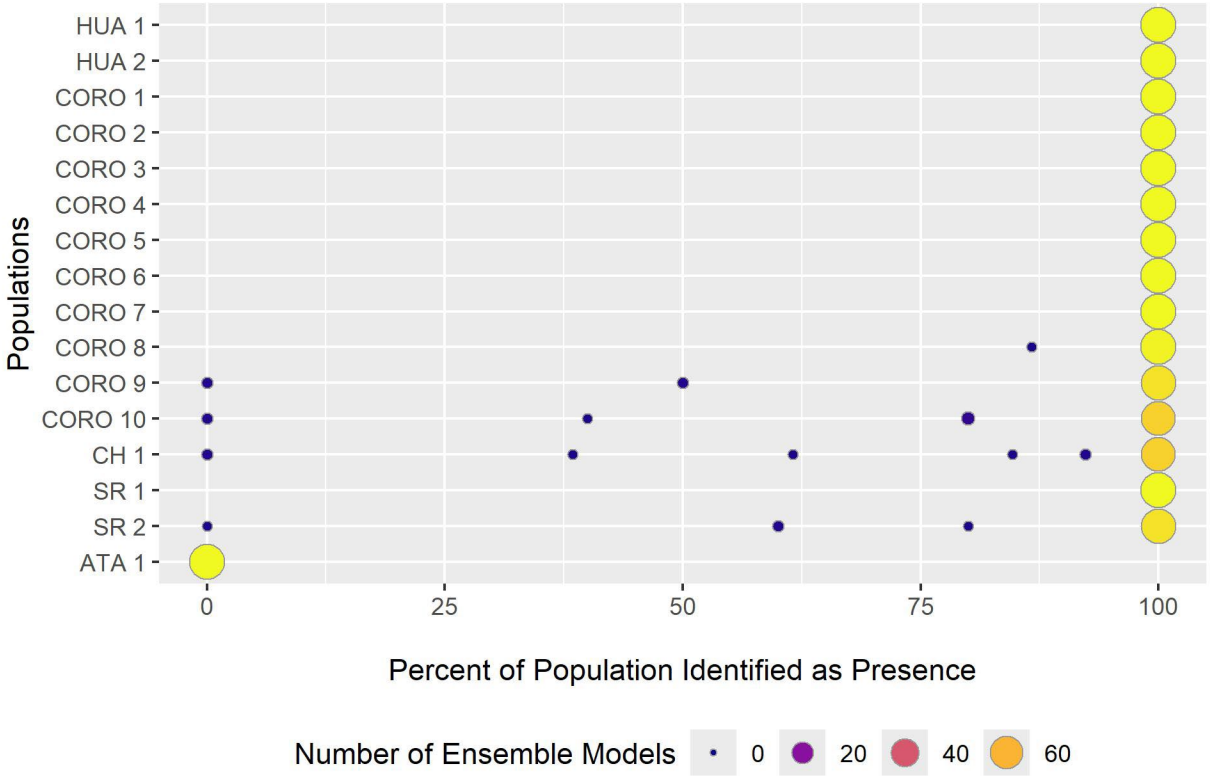


Figure 6. Percent of extant populations identified as presence by ensemble model binary rasters. Populations were based on data provided by collaborators at Northern Arizona University and numbered by general area: populations in Coronado National Memorial in the Huachuca Mountains (CORO), in the Huachuca Mountains west of Coronado National Memorial (HUA), in the Santa Rita Mountains (SR), in the Canelo Hills (CH), and in the Atascosa Mountains (ATA).

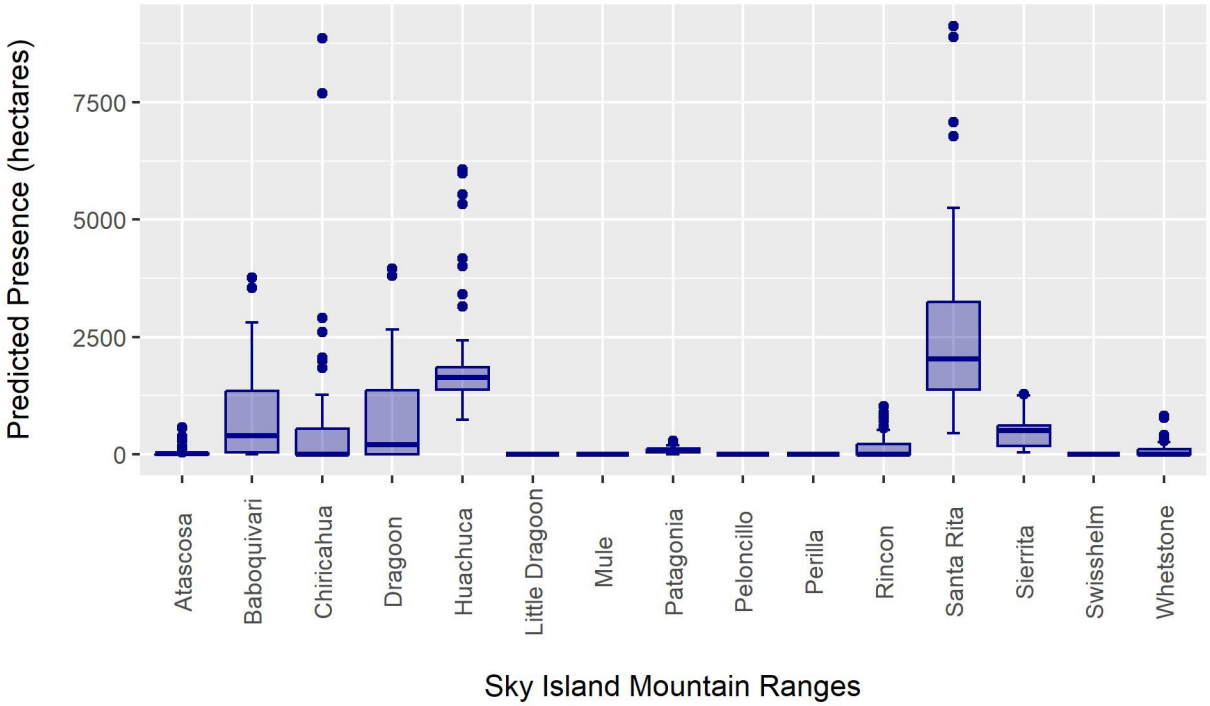


Figure 7: Area extent of predicted presence for P. imberbis for all ensemble model binary rasters by mountain range complex in the study area. Areas of predicted presence outside of known populations indicate areas of areas of ecological suitable habitat. The Canelo Hills are considered part of the Patagonia Mountains.

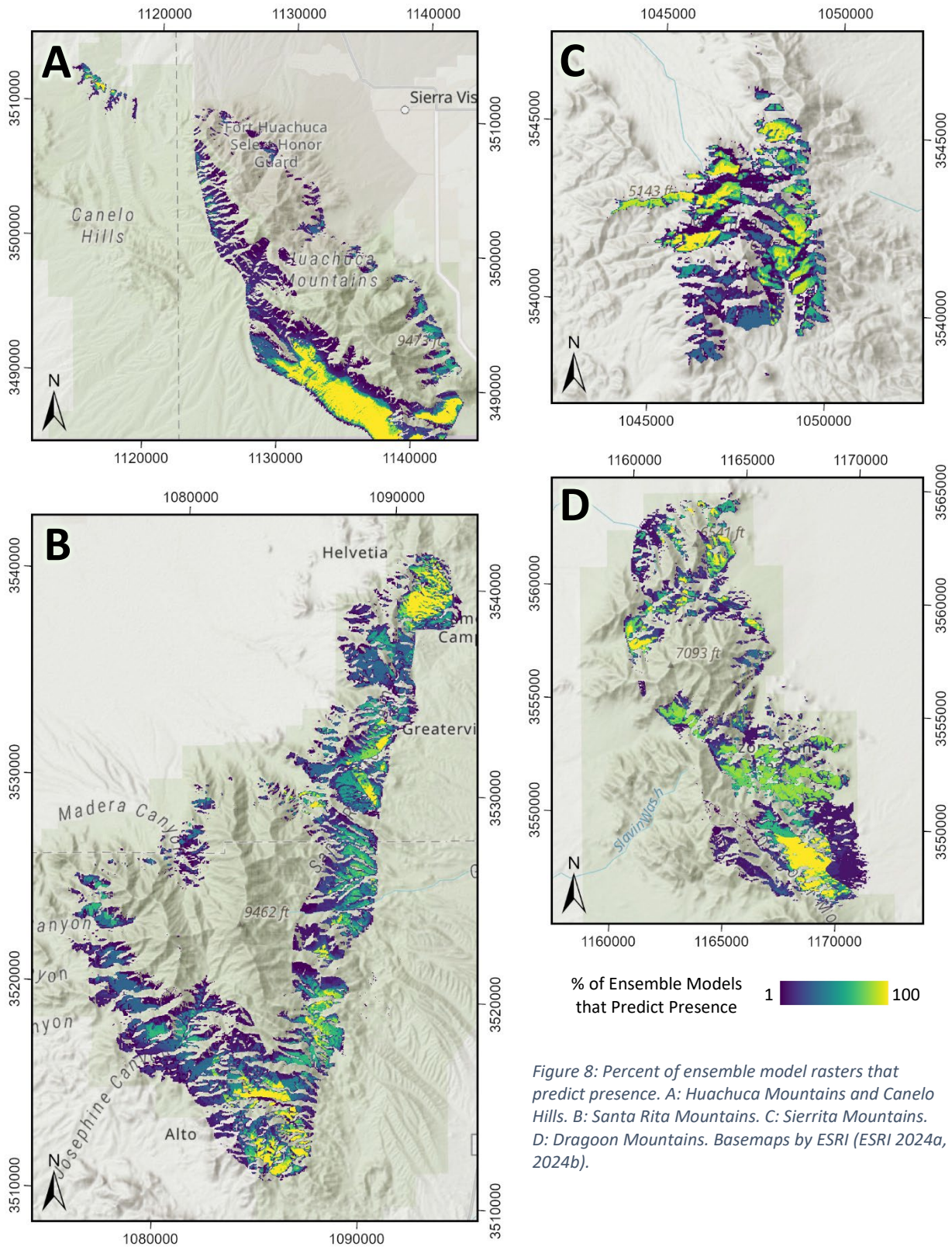


Figure 8: Percent of ensemble model rasters that predict presence. A: Huachuca Mountains and Canelo Hills. B: Santa Rita Mountains. C: Sierrita Mountains. D: Dragoon Mountains. Basemaps by ESRI (ESRI 2024a, 2024b).

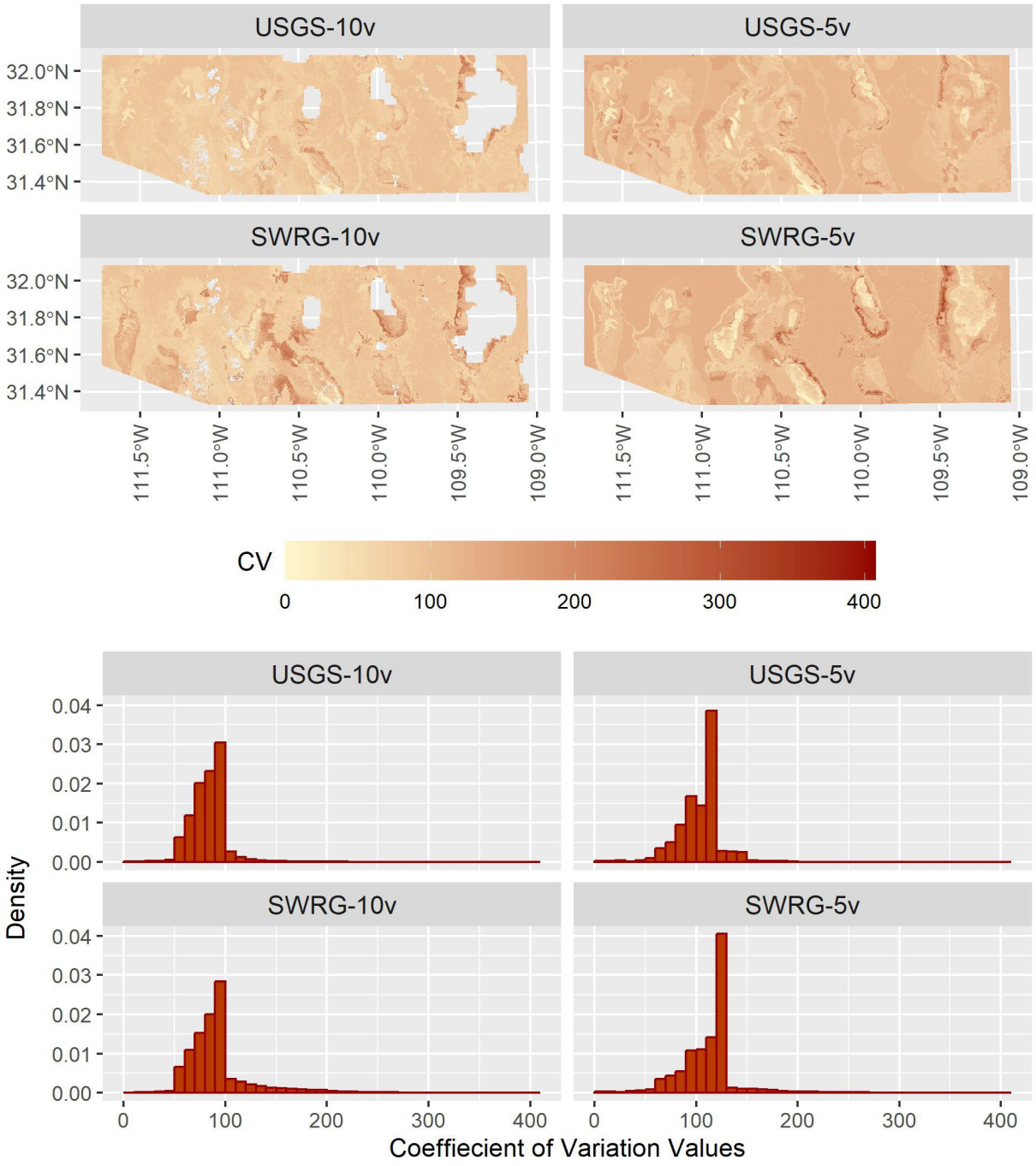


Figure 9. Top: Coefficient of variation (CV) for component models used in ensemble model building for each predictor dataset. Low values indicate areas of agreement on predicted presence between component models, high values indicate areas of disagreement between models. Bottom: Histograms of coefficient of variation for each predictor dataset. Datasets abbreviations same as Fig. 2 & 3.

Discussion

Ensemble SDMs were created for the endangered species *P. imberbis* to locate previously unknown populations and identify sites for reintroduction or translocation. Ensemble models were consistent in predictions across known populations of *P. imberbis,* and all ensemble models identified areas beyond currently occupied habitat that could be ecologically suitable for *P. imberbis*. The models were trained on populations covering 11 hectares, but the final ensemble model predictions ranged from 2,200 to 35,500 hectares of suitable habitat. The southwestern slopes of the Huachuca Mountains from the Mexican border to Lone Mountain and the lower south, southeast, and east-facing slopes of Montezuma Peak were most consistently identified as suitable habitat. However, 80% of the training data was located the southern Huachuca Mountains, and the observed consistency is likely due to the spatial bias of the presence points. This is a common issue when modeling rare species with skewed distributions, but results can still be used if this limitation is considered (Lomba et al. 2010; Jarnevich et al. 2015).

To build our models, we used a suite of 43 predictor variables of different types (e.g., climatic, edaphic, topographic, etc.). However, some ecological variables were not included because data are not available or because they were removed in the variable selection process. Biotic interactions can be major drivers of species distribution (Jarnevich et al. 2015; Soley-Guardia et al. 2024) and for *P. imberbis*, wild ungulate grazing and co-occurring vegetation have been shown to influence population growth (Souther et al. 2022). Deer population data was not included in the model but site-specific knowledge of deer populations and grazing patterns could be considered when using the model to identify specific areas at a reintroduction or translocation site. Although total herbaceous cover, shrub cover, and average shrub height variables (Table 1) were initially included, they were removed during initial feature selection. Instead, they could be retained during model development. Another practice when using SDMs for species reintroduction/translocation is including microsite analysis (Tomlinson et al. 2020; Elliott et al. 2024) which has been shown to improve outcomes of those efforts particularly in arid and semi-arid regions (Questad et al. 2014; Draper et al. 2019).

During model development, evaluation of models was complicated by consistently high and similar metric values. Metrics that focus on discrimination, such as ROC and TSS, weigh errors of commission and omission equally and are affected by prevalence of the study species (Leroy et al. 2018; Konowalik & Nosol 2021). Kappa is a calibration metric that is also affected by prevalence (Lawson et al. 2014; Pecchi et al. 2019), but Kappa varied more across models than ROC or TSS in this study. TSS, K, and BI are all threshold dependent metrics and the choice of threshold can have major implications on error rates and types of errors (Fielding & Bell 1997; Lawson et al. 2014). However, BI is a calibration metric design for presence-only data and is not affected by prevalence (Hirzel et al. 2006; Sillero et al. 2021); therefore, BI had the most impact on model selection for inclusion in ensemble model building. Apart from weaknesses inherent in model evaluation metrics, our generally high metric values suggest possible overfitting of our models to the data and overly complex models. This can cause issues with errors of commission excluding areas of possible new populations or suitable reintroduction/translocation sites (Soley-Guardia et al. 2024). However, given the spatial constraint of known populations, we chose to emphasize model sensitivity, or ability to identify true-positives, over concern about false-negatives; which is a common decision when modeling rare species (Buechling & Tobalske 2011; McCune 2016; Meyer et al. 2022).

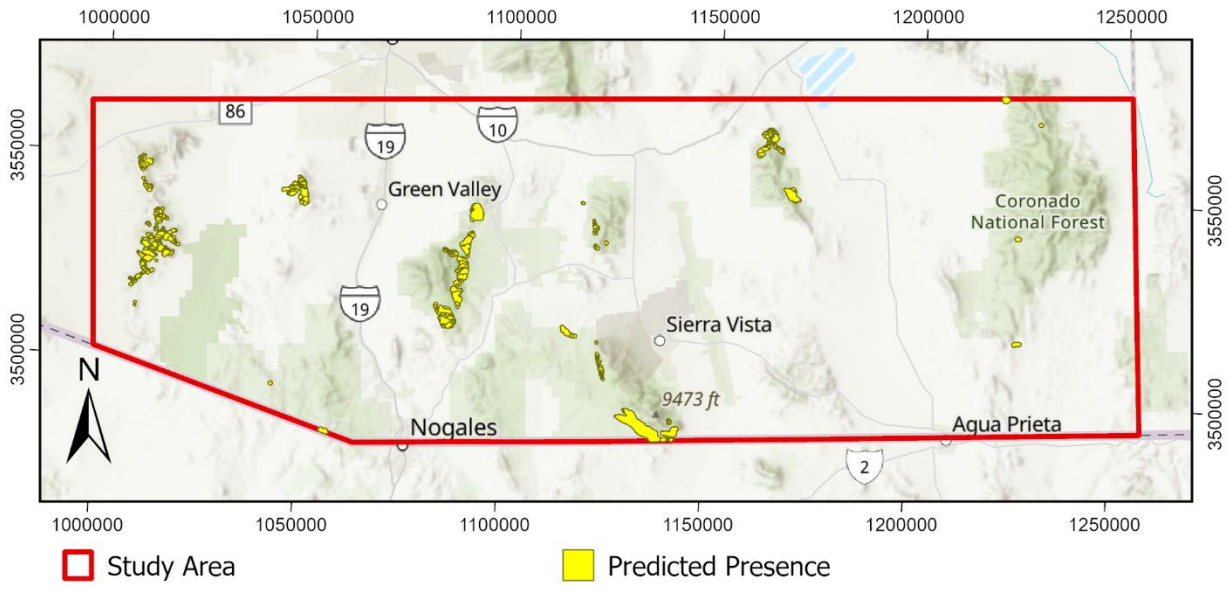


Figure 10. Area of predicted presence of P. imberbis by the ensemble committee averaging model based on the ROC metric using the 5-predictor dataset with USGS geologic data and a threshold based on optimizing the True Skill Statistic (USGS-5v_EMcaByROC_binTSS). Area is not to scale; it was increased for better visualization across the study area. Basemaps by ESRI (ESRI 2024a, 2024b).

Conclusions

We developed and compared a series of species distribution models for the rare species *Pectis imberbis* to help identify previously unknown populations and areas for translocations. We assessed the coefficient of variation between component models, compared evaluation metrics for separate ensemble models, and considered the planned application. All ensemble models outputs are available in Wilson (2024), and the best model from these analyses for this application is the ensemble committee averaging based on the ROCmetric using the 5-predictor dataset with USGS geologic data (data release raster: USGS-10v_EMcaByROC.tif) and a threshold based on optimizing the True Skill Statistic (data release raster: USGS-5v_EMcaByROC_binTSS.tif, Figure 10). These models can help inform conservation of *P. imberbis*.

Acknowledgements

This research was conducted by the U.S. Geological Survey (USGS), Western Geographic Science Center, Aridlands Water Harvesting Study with support from the USGS Land Change Science Program under the USGS Core Science Systems Mission Area. Funding was provided by U.S. Fish and Wildlife Service Science Support Partnership Program and Quick Response Program Grant G21AC10137. We would like to thank our partners at Northern Arizona University, Sara Souther, Clare Aslan, and Marsha Sample, for sharing their data and their dedication to rare species conservation. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Tables

Table 1. All predictor variables considered for species distribution modelling for *Pectis imberbis*. Variables included in the final modelled datasets are labeled with "Y". USGS-10v is the 10-predictor dataset using the USGS geologic data; USGS-5v is the 5-predictor dataset using the USGS geologic data developed with variable importance analysis. SWRG-10v is the 10-predictor dataset using the SWRG geologic data; SWRG-10v is the 5-predictor data using the SWRG geologic data developed with variable importance analysis.

			Predictor Datasets			
Type	**Predictor Name**	**Description and Source**	USGS-10v	USGS-5v	SWRG-10v	SWRG-5v
Topographic	Elevation	NASA Shuttle Radar Topography Mission Version 3 void-filled digital elevation model (DEM, NASA JPL 2013).	Y	Y	Y	Y
	Slope	Calculated from DEM using ArcGIS Pro slope tool.				
	Aspect	Calculated from DEM data using ArcGIS Pro aspect tool.				
	Northness	Calculated from aspect data using the formula: cos(aspect*pi/180).				
	Eastness	Calculated from aspect data using the formula: sin(aspect*pi/180).				
	Solar Radiation	Calculated from DEM using ArcGIS Pro solar radiation tool.	Y	Y	Y	Y
	Curvature Curvature Planar Curvature Profile	Calculated from DEM using ArcGIS Pro curvature tool.				
Climatic	Annual Precipitation	PRISM climate data (PRISM Climate Group 2022) 30-year normal (1991 – 2020) annual total precipitation.				
	Spring Precipitation	Calculated by summing PRISM 30-year normal monthly precipitation for February to May (Bodner & Robles 2017).	Y	Y	Y	Y
	Maximum Monthly Temperature	Maximum of the PRISM 30-year normal monthly average of daily maximum temperature.				
	Average Minimum Temperature	Minimum of the PRISM 30-year normal monthly average of daily minimum temperature.	Y		Y	
	Average Mean Temperature	PRISM 30-year normal average of daily mean temperature.				

Edaphic	Mean Bulk Density, 0 - 100 cm	Weighted mean calculated from POLARIS soil data (Chaney et al. 2016) retrieved using the XPolaris package in R.				
	Mean Percent Organic Matter, 0 - 100 cm	Weighted mean calculated from POLARIS soil data retrieved using the XPolaris package in R.	Y		Y	
	Mean pH, 0 - 100 cm	Weighted mean calculated from POLARIS soil data retrieved using the XPolaris package in R.	Y		Y	
	Mean Percent Clay, 0 - 100 cm	Weighted mean calculated from POLARIS soil data retrieved using the XPolaris package in R.				
	Mean Percent Sand, 0 - 100 cm	Weighted mean calculated from POLARIS soil data retrieved using the XPolaris package in R.				
	Mean Percent Silt, 0 - 100 cm	Weighted mean calculated from POLARIS soil data retrieved using the XPolaris package in R.				
	Soil Class	Soil Survey Geographic database (Soil Survey Staff, Natural Resources Conservation Service, United States Department of Agriculture 2021) soil classification (map unit key).				
	Dominant Drainage Class	SSURGO dominant drainage class.				
	Available Water Storage, 0 - 100 cm	SSURGO available water storage.	Y		Y	
Geologic	Geological Unit	USGS State Geologic Map Compilation database (Horton 2017) geological unit.	Y	Y		
	Generalized Lithology	USGS SGMC generalized lithology class.			Y	Y
	Surficial Geological Formations	Southwest Regional Gap Analysis Project (SWReGAP) Geology data (RS/GIS Laboratory, College of Natural Resources, Utah State University 2004a).				

Remote Sensing	Median NDVI Monsoon	Median Normalized Difference Vegetation Index (NDVI) for August - October from 2013 - 2023 derived from Landsat 8 Level 2 Collection 2 Tier 1 Surface Reflectance imagery (LC08; Earth Resources Observation and Science Center 2020) using Google Earth Engine (GEE; Gorelick et al. 2017).		
	Median NDVI Summer	Median NDVI for May - June from 2013 - 2023 derived from LC08 imagery using GEE.		
	Median Brightness Monsoon	Median brightness from the tasseled cap transformation (Crist & Kauth 1986) applied to LC08 imagery for August - October from 2013 – 2023 using GEE.		
	Median Brightness Summer	Median brightness from the tassel cap transformation applied to LC08 imagery for May - June from 2013 - 2023.	Y	Y
	Median Greenness Monsoon	Median greenness from the tassel cap transformation applied to LC08 imagery for August - October from 2013 - 2023.		
	Median Greenness Summer	Median greenness from the tassel cap transformation applied to LC08 imagery for May - June from 2013 - 2023.		
	Median Wetness Monsoon	Median wetness from the tassel cap transformation applied to LC08 imagery for August - October from 2013 - 2023.		
	Median Wetness Summer	Median wetness from the tassel cap transformation applied to LC08 imagery for May - June from 2013 - 2023.		
Miscellaneous	SWRG Land Cover	SWReGAP land cover data (RS/GIS Laboratory, College of Natural Resources, Utah State University 2004b).		
	Land Form	SWReGAP land form data (RS/GIS Laboratory, College of Natural Resources, Utah State University 2004c).		

Ecoregion	EPA Level 4 Ecoregion (US Environmental Protection Agency 2013).	Y	Y	Y	Y
NLCD Land Cover	National Land Cover Database (NLCD) 2021 land cover (U.S. Geological Survey & Dewitz 2023).				
Annual Herbaceous Percent Cover	NLCD Rangeland Condition Monitoring Assessment and Projection (RCMAP) 2016 annual herbaceous percent cover (U.S. Geological Survey 2020).				
Bare Ground Percent Cover	NLCD RCMAP 2016 bare ground percent cover.				
Total Herbaceous Percent Cover	NLCD RCMAP 2016 total (annual and perennial) herbaceous percent cover.				
Shrub Percent Cover	NLCD RCMAP 2016 shrub percent cover.				
Average Shrub Height	NLCD RCMAP 2016 average shrub height in centimeters.				

Table 2. Number of models completed during single model runs and number of models selected for inclusion in ensemble modelling per algorithm and predictor dataset.

		Models Completed per Predictor Dataset				Models Selected for Ensemble Building per Predictor Dataset			
Algorithm	**Models Possible**	USGS-10v	USGS-vI	SWRG-10v	SWRG-vI	USGS-10v	USGS-vI	SWRG-10v	SWRG-vI
GLM	150	129	129	137	137	2	1	2	0
GAM	150	129	129	137	137	0	0	0	0
GBM	3000	3000	3000	3000	3000	92	43	100	28
RF	3000	3000	3000	3000	3000	100	100	100	100
CTA	3000	3000	3000	3000	3000	100	100	100	100
Total	9300	9258	9258	9274	9274	294	244	302	228

Literature

Allouche O, Tsoar A, Kadmon R (2006) Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). Journal of Applied Ecology 43:1223–1232 https://doi.org/10.1111/j.1365-2664.2006.01214.x

Araujo M, New M (2007) Ensemble forecasting of species distributions. Trends in Ecology & Evolution 22:42–47 https://doi.org/10.1016/j.tree.2006.09.010

Baker Jr MB, DeBano LF, Ffolliott PF (1994) Hydrology and watershed management in the Madrean Archipelago. In: Biodiversity and management of the Madrean archipelago: The sky islands of southwestern United States and northwestern Mexico. DeBano, LF, Ffolliott, PF, Ortega-Rubio, A, Gottfried, GJ, Hamre, RH, & Edminster, CB, editors. General Technical Report Vol. RM-GTR-264 USDA Forest Service, Rocky Mountain Research Station, Tucson, Arizona pp. 329–337. https://doi.org/10.2737/RM-GTR-264

Barbet-Massin M, Jiguet F, Albert CH, Thuiller W (2012) Selecting pseudo-absences for species distribution models: how, where and how many? Methods in Ecology and Evolution 3:327–338 https://doi.org/10.1111/j.2041-210X.2011.00172.x

Bellis J, Albrecht MA, Maschinski J, Osazuwa-Peters O, Stanley T, Heineman KD (2024) Advancing the science and practice of rare plant conservation with the Center for Plant Conservation Reintroduction Database. Applications in Plant Sciences 12:e11583 https://doi.org/10.1002/aps3.11583

Bodner GS, Robles MD (2017) Enduring a decade of drought: Patterns and drivers of vegetation change in a semi-arid grassland. Journal of Arid Environments 136:1–14 https://doi.org/10.1016/j.jaridenv.2016.09.002

Borokini I, Nussear K, Petitpierre B, Dilts T, Weisberg P (2023) Iterative species distribution modeling results in the discovery of novel populations of a rare cold desert perennial. Endangered Species Research 50:47–62 https://doi.org/10.3354/esr01218

Boyce MS, Vernier PR, Nielsen SE, Schmiegelow FKA (2002) Evaluating resource selection functions. Ecological Modelling 157:281–300 https://doi.org/10.1016/S0304-3800(02)00200-4

Buechling A, Tobalske C (2011) Predictive Habitat Modeling of Rare Plant Species in Pacific Northwest Forests. Western Journal of Applied Forestry 26:71–81 https://doi.org/10.1093/wjaf/26.2.71

Chaney NW, Wood EF, McBratney AB, Hempel JW, Nauman TW, Brungard CW, Odgers NP (2016) POLARIS: A 30-meter probabilistic soil series map of the contiguous United States. Geoderma 274:54–67 https://doi.org/10.1016/j.geoderma.2016.03.025

Crist EP, Kauth RJ (1986) The Tasseled Cap De-mystified. Photogrammetric Engineering and Remote Sensing 52:81–86

Daniel F, Microsoft Corporation, Weston S (2022) doParallel: Foreach Parallel Adaptor for the 'parallel' Package. v 1.0.17 https://CRAN.R-project.org/package=doParallel

DeBano LF, Ffolliott PF (1994) The sky island conference: looking back, looking ahead. In: Biodiversity and management of the Madrean archipelago: The sky islands of southwestern United States and northwestern Mexico. DeBano, LF, Ffolliott, PF, Ortega-Rubio, A, Gottfried, GJ, Hamre, RH, & Edminster, CB, editors. General Technical Report Vol. RM-GTR-264 USDA Forest Service, Rocky Mountain Research Station, Tucson, Arizona pp. 1–5. https://doi.org/10.2737/RM-GTR-264

Devender TRV, Avila-Villegas S, Emerson M, Turner D, Flesch AD, Deyo NS (2012) Biodiversity in the Madrean Archipelago of Sonora, Mexico. In: Merging science and management in a rapidly changing world: Biodiversity and management of the Madrean Archipelago III. Gottfried, GJ, Ffolliott, PF, Gebow, BS, Eskew, LG, & Collins, LC, editors. Proceedings Vol. RMRS-P-67 U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Tucson, Arizona pp. 10–16. https://www.fs.usda.gov/rm/pubs/rmrs_p067.pdf

Dormann CF, Elith J, Bacher S, Buchmann C, Carl G, Carré G, Marquéz JRG, Gruber B, Lafourcade B, Leitão PJ, Münkemüller T, McClean C, Osborne PE, Reineking B, Schröder B, Skidmore AK, Zurell D, Lautenbach S (2013) Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography 36:27–46 https://doi.org/10.1111/j.1600-0587.2012.07348.x

Draper D, Marques I, Iriondo JM (2019) Species distribution models with field validation, a key approach for successful selection of receptor sites in conservation translocations. Global Ecology and Conservation 19:e00653 https://doi.org/10.1016/j.gecco.2019.e00653

Earth Resources Observation and Science (EROS) Center (2020) Landsat 8-9 Operational Land Imager / Thermal Infrared Sensor Level-2, Collection 2 [dataset]. U.S. Geological Survey. https://doi.org/10.5066/P9OGBGM6. Accessed 8/24/23

Elliott CP, Tomlinson S, Lewandrowski W, Miller BP (2024) Species distribution and habitat attributes guide translocation planning of a threatened short-range endemic plant. Global Ecology and Conservation 51:e02915 https://doi.org/10.1016/j.gecco.2024.e02915

ESRI (2023) ArcGIS Pro. v. 3.2.2 https://www.esri.com/en-us/home

ESRI (2024a) Human Geography Map. Accessed: 6/21/24 https://www.arcgis.com/home/item.html?id=3582b744bba84668b52a16b0b6942544

ESRI (2024b) World Topographic Map. Accessed: 6/21/24 https://basemaps.arcgis.com/arcgis/rest/services/World_Basemap_v2/VectorTileServer

Felger RS, Wilson MF (1994) Northern Sierra Madre Occidental and its Apachian outliers: a neglected center of biodiversity. In: Biodiversity and management of the Madrean archipelago: The sky islands of southwestern United States and northwestern Mexico. DeBano, LF, Ffolliott, PF, Ortega-Rubio, A, Gottfried, GJ, Hamre, RH, & Edminster, CB, editors. General Technical Report Vol. RM-GTR-264 USDA Forest Service, Rocky Mountain Research Station, Tucson, Arizona pp. 36–51. https://doi.org/10.2737/RM-GTR-264

Fielding AH, Bell JF (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. Environmental Conservation 24:38–49 https://doi.org/10.1017/S0376892997000088

Gorelick N, Hancher M, Dixon M, Ilyushchenko S, Thau D, Moore R (2017) Google Earth Engine: Planetary-scale geospatial analysis for everyone. Big Remotely Sensed Data: tools, applications and experiences 202:18–27 https://doi.org/10.1016/j.rse.2017.06.031

Hao T, Elith J, Guillera-Arroita G, Lahoz-Monfort JJ (2019) A review of evidence about use and performance of species distribution modelling ensembles like BIOMOD Serra-Diaz, J, editor. Diversity and Distributions 25:839–852 https://doi.org/10.1111/ddi.12892

Hijmans RJ (2024) terra: Spatial Data Analysis. v. 1.7-71 https://CRAN.R-project.org/package=terra

Hirzel AH, Le Lay G, Helfer V, Randin C, Guisan A (2006) Evaluating the ability of habitat suitability models to predict species presences. Ecological Modelling 199:142–152 https://doi.org/10.1016/j.ecolmodel.2006.05.017

Horton JD (2017) The State Geologic Map Compilation (SGMC) Geodatabase of the Conterminous United States. https://doi.org/10.5066/F7WH2N65

Hysen L, Nayeri D, Cushman S, Wan HY (2022) Background sampling for multi-scale ensemble habitat selection modeling: Does the number of points matter? Ecological Informatics 72:101914 https://doi.org/10.1016/j.ecoinf.2022.101914

Jarnevich CS, Stohlgren TJ, Kumar S, Morisette JT, Holcombe TR (2015) Caveats for correlative species distribution modeling. Ecological Informatics 29:6–15 https://doi.org/10.1016/j.ecoinf.2015.06.007

Kassambara A (2023) ggcorrplot: Visualization of a Correlation Matrix using 'ggplot2'. v.0.1.4.1 https://CRAN.R-project.org/package=ggcorrplot

Konowalik K, Nosol A (2021) Evaluation metrics and validation of presence-only species distribution models based on distributional maps with varying coverage. Scientific Reports 11:1482 https://doi.org/10.1038/s41598-020-80062-1

Kuhn M (2023) caret: Misc functions for training and plotting classification and regression models. v. 6.0-94 https://cran.r-project.org/package=caret

Kursa MB, Rudnicki WR (2022) Boruta: Wrapper Algorithm for All Relevant Feature Selection. v 8.0.0 https://CRAN.R-project.org/package=Boruta

Lawson CR, Hodgson JA, Wilson RJ, Richards SA (2014) Prevalence, thresholds and the performance of presence–absence models Freckleton, R, editor. Methods in Ecology and Evolution 5:54–64 https://doi.org/10.1111/2041-210X.12123

Leroy B, Delsol R, Hugueny B, Meynard CN, Barhoumi C, Barbet-Massin M, Bellard C (2018) Without quality presence–absence data, discrimination metrics such as TSS can be misleading measures

of model performance. Journal of Biogeography 45:1994–2002 https://doi.org/10.1111/jbi.13402

Li W, Guo Q, Elkan C (2011) Can we model the probability of presence of species without absence data? Ecography 34:1096–1105 https://doi.org/10.1111/j.1600-0587.2011.06888.x

Lomba A, Pellissier L, Randin C, Vicente J, Moreira F, Honrado J, Guisan A (2010) Overcoming the rare species modelling paradox: A novel hierarchical framework applied to an Iberian endemic plant. Biological Conservation 143:2647–2657 https://doi.org/10.1016/j.biocon.2010.07.007

López-Hoffman L, Quijada-Mascareñas A (2012) Madrean Sky Islands, North America. In: Climate and Conservation. Hilty, JA, Chester, CC, & Cross, MS, editors. Island Press/Center for Resource Economics, Washington, DC pp. 217–226.

Manel S, Williams HC, Ormerod SJ (2001) Evaluating Presence-Absence Models in Ecology: The Need to Account for Prevalence. Journal of Applied Ecology 38:921–931 http://www.jstor.org/stable/827232

Marmion M, Parviainen M, Luoto M, Heikkinen RK, Thuiller W (2009) Evaluation of consensus methods in predictive species distribution modelling. Diversity and Distributions 15:59–69 https://doi.org/10.1111/j.1472-4642.2008.00491.x

McCune JL (2016) Species distribution models predict rare species occurrences despite significant effects of landscape context Baraloto, C, editor. Journal of Applied Ecology 53:1871–1879 https://doi.org/10.1111/1365-2664.12702

Meyer J-Y, Pouteau R, Vincent F (2022) Assessing habitat suitability for the translocation of Ochrosia tahitensis (Apocynaceae), a critically endangered endemic plant from the island of Tahiti (South Pacific). Journal for Nature Conservation 68:126198 https://doi.org/10.1016/j.jnc.2022.126198

NASA JPL (2013) NASA Shuttle Radar Topography Mission Global 1 arc second V003. https://doi.org/10.5067/measures/srtm/srtmgl1.003

National Park Service (2014) Rare Species Memo: Pectis imberbis at Coronado National Memorial. https://www.regulations.gov/document/FWS-R2-ES-2018-0104-0025

Pecchi M, Marchi M, Burton V, Giannetti F, Moriondo M, Bernetti I, Bindi M, Chirici G (2019) Species distribution modelling to support forest management. A literature review. Ecological Modelling 411:108817 https://doi.org/10.1016/j.ecolmodel.2019.108817

Phillips III AM, Phillips BG, Green III LT, Mazzoni J, Brian N (1982) Status Report: Pectis imberbis. Museum of Northern Arizona, Flagstaff, Arizona. https://www.regulations.gov/document/FWS-R2-ES-2018-0104-0026

Posit Team (2024) RStudio: Integrated development environment for R. v. 2023.12.1.402 http://www.posit.co/

PRISM Climate Group (2022) PRISM climate data. Accessed: 2/2/23. https://prism.oregonstate.edu

Questad EJ, Kellner JR, Kinney K, Cordell S, Asner GP, Thaxton J, Diep J, Uowolo A, Brooks S, Inman-Narahari N, Evans SA, Tucker B (2014) Mapping habitat suitability for at-risk plant species and its implications for restoration and reintroduction. Ecological Applications 24:385–395 https://doi.org/10.1890/13-0775.1

R Core Team (2024) R: a language and environment for statistical computing. v. 4.3.3 https://www.R-project.org/

Rathore MK, Sharma LK (2023) Efficacy of species distribution models (SDMs) for ecological realms to ascertain biological conservation and practices. Biodiversity and Conservation 32:3053–3087 https://doi.org/10.1007/s10531-023-02648-1

RS/GIS Laboratory, College of Natural Resources, Utah State University (2004a) Southwest Regional Gap Analysis Project: 1:500,000 Scale Geology for the Southwestern U.S. https://swregap.org/data/geology/

RS/GIS Laboratory, College of Natural Resources, Utah State University (2004b) Southwest Regional Gap Analysis Project: Digital Landcover Dataset for the Southwestern United States. https://swregap.org/data/landcover/

RS/GIS Laboratory, College of Natural Resources, Utah State University (2004c) Southwest Regional Gap Analysis Project: Ten Class DEM Derived Landform for the Southwest United States. https://swregap.org/data/landform/

Ruiz-Jaen MC, Mitchell Aide T (2005) Restoration success: How is it being measured? Restoration Ecology 13:569–577 https://doi.org/10.1111/j.1526-100X.2005.00072.x

Rusconi O, Broennimann O, Storrer Y, Le Bayon R, Guisan A, Rasmann S (2022) Detecting preservation and reintroduction sites for endangered plant species using a two-step modeling and field approach. Conservation Science and Practice 4:e12800 https://doi.org/10.1111/csp2.12800

Sánchez Escalante JJ, León JPC, Cruz-Zagasta JO (2019) Surveys for *Cirsium wrightii* and other rare plants (*Graptopetalum bartramii*, *Pediomelum pentaphyllum*, *Pectis imberbis*, *Leucosyris blepharophylla*, and *Eryngium sparganophyllum*) in northeastern Sonora and northern Chihuahua, Mexico. Universidad de Sonora, Herbario, Hermosillo, Sonora, México https://www.regulations.gov/document/FWS-R2-ES-2018-0104-0026

SEINet Portal Network (2024). Accessed: 6/21/24 https://swbiodiversity.org/seinet/taxa/index.php?tid=2326&taxauthid=1&clid=0

Sillero N, Arenas-Castro S, Enriquez-Urzelai U, Vale CG, Sousa-Guedes D, Martínez-Freiría F, Real R, Barbosa AM (2021) Want to model a species niche? A step-by-step guideline on correlative ecological niche modelling. Ecological Modelling 456:109671 https://doi.org/10.1016/j.ecolmodel.2021.109671

Soil Survey Staff, Natural Resources Conservation Service, United States Department of Agriculture (2021) Soil Survey Geographic (SSURGO) database. Accessed: 4/1/22. https://sdmdataaccess.sc.egov.usda.gov/

Soley-Guardia M, Alvarado-Serrano DF, Anderson RP (2024) Top ten hazards to avoid when modeling species distributions: a didactic guide of assumptions, problems, and recommendations. Ecography 2024:e06852 https://doi.org/10.1111/ecog.06852

Souther S, Sample M, Conley G, Aslan C (2022) Demographic Analysis of the Endangered Plant Pectis imberbis Highlights Tradeoffs Between Deer Browse and Interspecific Competition. Natural Areas Journal 42:230–241 https://doi.org/10.3375/22-3

Thuiller W, Georges D, Gueguen M, Engler R, Breiner F, Lafourcade B, Patin R, Blancheteau H (2024) biomod2: Ensemble Platform for Species Distribution Modeling. v. 4.2-5 https://CRAN.R-project.org/package=biomod2

Tomlinson S, Lewandrowski W, Elliott CP, Miller BP, Turner SR (2020) High-resolution distribution modeling of a threatened short-range endemic plant informed by edaphic factors. Ecology and Evolution 10:763–777 https://doi.org/10.1002/ece3.5933

US Environmental Protection Agency (2013) Level IV Ecoregions of Arizona. Accessed: 2/3/23. https://www.epa.gov/eco-research/level-iii-and-iv-ecoregions-continental-united-states

U.S. Fish and Wildlife Service (2021) Endangered and Threatened Wildlife and Plants; Endangered Status for the Beardless Chinchweed and Designation of Critical Habitat. Federal Register 86:31830–31868

U.S. Fish and Wildlife Service (2020) Species status assessment report for *Pectis imberbis* (beardless chinchweed) Version 2.0. U.S. Fish and Wildlife Service, Southwest Region, Tucson, Arizona https://www.regulations.gov/document/FWS-R2-ES-2018-0104-0024

U.S. Geological Survey Landsat Collection 2 Level 2 Surface Reflectance. https://www.usgs.gov/landsat-missions/landsat-collection-2-surface-reflectance

U.S. Geological Survey (2020) Rangeland Condition, Monitoring, Assessment, and Projection (RCMAP) fractional component data. https://doi.org/10.5066/P9MJVQSQ

U.S. Geological Survey, Dewitz J (2023) National Land Cover Database (NLCD) 2021 Land Cover Conterminous United States. https://doi.org/10.5066/P9JZ7AO3

Valavi R, Guillera-Arroita G, Lahoz-Monfort JJ, Elith J (2022) Predictive performance of presence-only species distribution models: a benchmark study with reproducible code. Ecological Monographs 92:e01486 https://doi.org/10.1002/ecm.1486

Wickham H, Chang W, Henry L, Pedersen TL, Takahashi K, Wilke C, Woo K, Yutani H, Dunnington D, van den Brand T (2024) ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics. v. 3.5.1 https://cran.r-project.org/package=ggplot2

Williams JN, Seo C, Thorne J, Nelson JK, Erwin S, O'Brien JM, Schwartz MW (2009) Using species distribution models to predict new occurrences for rare plants. Diversity and Distributions 15:565–576 https://doi.org/10.1111/j.1472-4642.2009.00567.x

Wilson NR (2024) Species Distribution Models for Pectis imberbis, a Rare Plant Species in Southeastern Arizona. U.S. Geological Survey data release. https://doi.org/10.5066/P13VMRBC