Conservation Genetics
https://doi.org/10.1007/s10592-019-01212-1

RESEARCH ARTICLE

The conservation genomics of the endangered distylous gypsophile *Oreocarya crassipes* (Boraginaceae)

James I. Cohen[1]

Received: 26 April 2019 / Accepted: 6 August 2019

Abstract

The Chihuahuan Desert includes many endemic angiosperm species, some having very restricted geographic ranges. One of these species is *Oreocarya crassipes* (I. M. Johnst.) Hasenstab & M. G. Simpson, an endangered distylous gypsophile from the Trans-Pecos region in southern Brewster County, Texas, USA. The species is known from 10 populations, and this small number of populations, human development in the area, a distylous breeding system, and edaphic requirements threaten the long-term viability of the species. Using both hundreds of single nucleotide polymorphisms identified via tunable genotyping-by-sequencing (tGBS) and 10 microsatellite loci, patterns of genetic diversity, demography, selection, and migration were examined for 192 individuals from four populations of *O. crassipes*. From the sampled individuals, two populations (clusters) were identified via multiple methodologies and with both types of data. With SNP data, population substructure was further resolved among one of these populations to identify two distinct groups of individuals. Multiple individuals recognized as having mixed ancestry, along with *F*st values and AMOVA results, provide evidence of genetic exchange among populations, which is less common for gypsophiles than non-gypsophiles, and the rate of migration among populations has been increasing recently. The *F*st values for *O. crassipes* are more similar to those of other rare species than to other gypsophiles. Additionally, while distyly specifically does not necessarily impact the population genetics of the species, allogamy, which is facilitated by distyly, seems to have played a role in the genetic structure of *O. crassipes*.

Keywords Boraginaceae · Chihuahuan Desert · Conservation genomics · Distyly · Genotyping-by-sequencing · Gypsum

Introduction

The Chihuahuan Desert spans much of northwestern Mexico and the southwestern United States, and the desert climate and landscape have influenced the flora of the area, particularly in terms of endemism. Of the more than 3200 angiosperm species in the flora, ca. 25% are endemic (Villarreal-Quintanilla et al. 2017). This high level of endemism may be due to the old age of the desert allowing ample time for speciation (i.e., millions years old [Moore and Jansen 2007]), the presence of diverse habitats (Muldavin 2002; Michaud et al. 2013), edaphic and other ecological specialization (Powell and Turner 1977; Moore et al. 2014), and frequency of disturbance (Hernandez and Barcenas 1995; Kerley and Whitford 2000). The plant family Boraginaceae includes the third largest number of endemic species in the Chihuahuan Desert, 34 (Villarreal-Quintanilla et al. 2017). One of these endemic species, *Oreocarya crassipes* (I. M. Johnst.) Hasenstab & M. G. Simpson, the Terlingua Creek Cat's-Eye, is a perennial, primarily allogamous species restricted to a small geographic area in the Trans-Pecos region in southern Brewster County, Texas, USA just north of Big Bend National Park, and according to the United States Fish and Wildlife Service, the species is endangered (U. S. Fish and Wildlife Service 1993). In its approximately 300 km^2 geographic range, *O. crassipes* is known from 10 populations (although presently some may be historic), and all of the known populations are on private land (Warnock 2012). Although more than 10,000 individuals are currently known (up from 4500 ca. 20 years ago), the species

---

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10592-019-01212-1) contains supplementary material, which is available to authorized users.

✉ James I. Cohen
jcohen@kettering.edu

[1] Applied Biology, Kettering University, 1700 University Ave., Flint, MI 48504, USA

has ecological (gypsophily) and reproductive (distyly) requirements that make its long-term persistence challenging. Given the small number and size of populations of *O. crassipes*, along with the species' distylous breeding system and edaphic specialization to gypsum, understanding the genetic diversity and patterns of migration will be critical to developing the most appropriate conservation measures for the species.

Oreocarya crassipes grows in a habitat known as moonscape, an appropriate term due to the relatively barren landscape composed of clay flats with high levels of gypsum ($CaS0_4 \cdot 2H_2O$). In *O. crassipes*, the adaptation to high gypsum levels may allow the plants to colonize areas that have greater moisture than nearby soils or to reduce competition from other species of plants due to difficulties for seedling growth and establishment in gypsum soil (Meyer and Garcia-Moya 1986; Meyer 1986). *Oreocarya crassipes* also produces quantities of pyrrolizidine alkaloids that are ca. 10–50 times greater than reported from other species (Williams et al. 2011). These large amounts are hypothesized to provide protection from herbivores and microbes as well as increase osmoregulation, all of which can be useful adaptations to survival in xeric environments (Williams et al. 2011).

In addition to adaptations to the harsh moonscape environment, *O. crassipes* exhibits distyly, an uncommon plant breeding system (Barrett and Shore 2008; Cohen 2014). Distyly is characterized by two floral morphs present in a population. In the long-style (LS) morph, the stigmas are positioned above the anthers, and in the short-style (SS) morph, the stigmas are situated below the anthers. The anthers of one morph are at the same height as the stigmas in the other morph, a condition known as reciprocal herkogamy, which should result in effective intermorph pollination as well as reduced interference between the sexual organs within the flower (Barrett 1992; Barrett and Shore 2008; Cohen 2010). Along with this morphological component, most distylous species include a self- and intramorph incompatibility system, and this results in only intermorph crosses (i.e., between sexual organs of the same height) producing legitimate offspring (Barrett 1992; Barrett and Shore 2008; Cohen 2010). Although, in the related species *Oreocarya flava* A. Nelson, self- and intramorph crosses also produce offspring (Casper 1985), and LS:SS morph ratios in the populations of *O. flava* are close to 1:1. The equal morph ratio suggests that another mechanism (e.g., late-acting self- and intramorph incompatibility or selective abortion) may be responsible for maintaining isoplethy, which is the theoretical LS:SS ratio in distylous populations given the mechanism of inheritance of the breeding system (Ganders 1979; Barrett et al. 2000). Consequently, both morphs must be present within populations for efficient pollination and for legitimate offspring to be produced. This can be a considerable concern in smaller populations, particularly for endangered species, where one morph can come to dominate the population (e.g., Van Rossum et al. 2006).

The edaphic specialization of *O. crassipes* to gypsum can result in inhospitable areas between gypsum patches increasing the difficulty of the establishment of new populations, and this, along with the distylous breeding system, present challenges for gene flow among populations, increasing among-population genetic variation. To better study the impact of gypsophily, distyly, and small populations on the population genetics of *O. crassipes*, multiple types of genetic markers are available (e.g., microsatellites, AFLPs, RFLPs, SNPs) (Nybom 2004). While microsatellites have proven quite useful for population genetic studies in *Oreocarya* Greene (Bresowar and McGlaughlin 2015), researchers examining non-model species can currently take advantage of next-generation sequencing technology to generate hundreds to thousands of single nucleotide polymorphisms (SNPs) from across the genome. One method to identify SNPs is tunable genotyping-by-sequencing (tGBS) (Ott et al. 2017). tGBS uses a digestion with two restriction enzymes (e.g., NspI and BfuCI), followed by selective PCR, to reduce much of the genome. Short reads are subsequently sequenced and used to identify polymorphic loci among samples (Ott et al. 2017). Examining results from analyses of different genetic markers (e.g., SNPs and microsatellites), each with different types and rates of mutation, can allow for a comprehensive examination of genetic diversity, population structure, and patterns of selection.

Human activities, such as mining, off-road vehicle traffic, and grazing, and disease, along with the restricted geographic range, small number of individuals, edaphic specialization, and distylous breeding system, result in threats and challenges to the long-term viability of *O. crassipes* (Williams et al. 2011; Warnock 2012). The objective of the present study is to investigate the patterns of genetic diversity, demography, selection, and migration among four populations of *O. crassipes* in order to inform conservation efforts to ensure long-term viability and genetic diversity, for a unique species adapted to a harsh xeric environment. The hypothesis of the present study is that while *O. crassipes* inhabits a small geographic range, given the presence of edaphic specializiation and distyly, population structure exists, particularly between the northern and southern populations, and migration occurs throughout the geographic range of the species.

Methods

Material

In March and April 2014, leaf material was collected from individuals of four populations of *O. crassipes*, the Field

Lab (F, an alternative energy and sustainable living field laboratory, http://thefieldlab.blogspot.com), with 999 individuals identified, and three populations from the O2 Ranch (O2-1, O2-2, and O2-3), with ca. 8700 individuals identified among the populations (Fig. 1) (Warnock 2012) and a greater number in O2-1 than O2-2 and O2-3. The Field Lab and O2-Ranch populations are ca. 17 km apart. The latitude and longitude and morph (i.e., LS or SS) were recorded for each collection. Leaf material was placed in silica gel and after being at ambient temperature for less than a week, stored at 4 °C. Vouchers specimens were deposited at the University of Michigan Herbarium (MICH).

DNA sequencing

Leaf material for 192 samples from the four populations (87 F, 58 O2-1, 31 O2-2, 16 O2-3, sampling scheme representative of a diversity of geographic distances of plants able to be collected within and among populations [intrapopulation distances of individual plants range from approximately 1–800 m]) was sent to the biotechnology company data2bio (https://www.data2bio.com) for DNA isolation and tunable genotyping-by-sequencing (tGBS) to generate single nucleotide polymorphisms (SNPs) (Ott et al. 2017) using four runs on an Ion Proton (Rothberg et al. 2011). Low quality reads and bases (PHRED score < 15) were removed, and consensus sequences were generated from the resulting reads. The consensus sequences were used for mapping the high-quality reads for each sample using the alignment program GSNAP with the following parameters: ≤ 2 mismatches per 36 base pairs (bp) and fewer than five bases for every 75 bp as tails (Wu and Nacu 2010). Two SNP datasets were produced based on different criteria for missing data: LMD10 with 238 SNPs and up to 10% missing data and LMD50 with 17,143 SNPs and up to 50% missing data. For both SNP datasets, the minor allele frequency had to be present in at

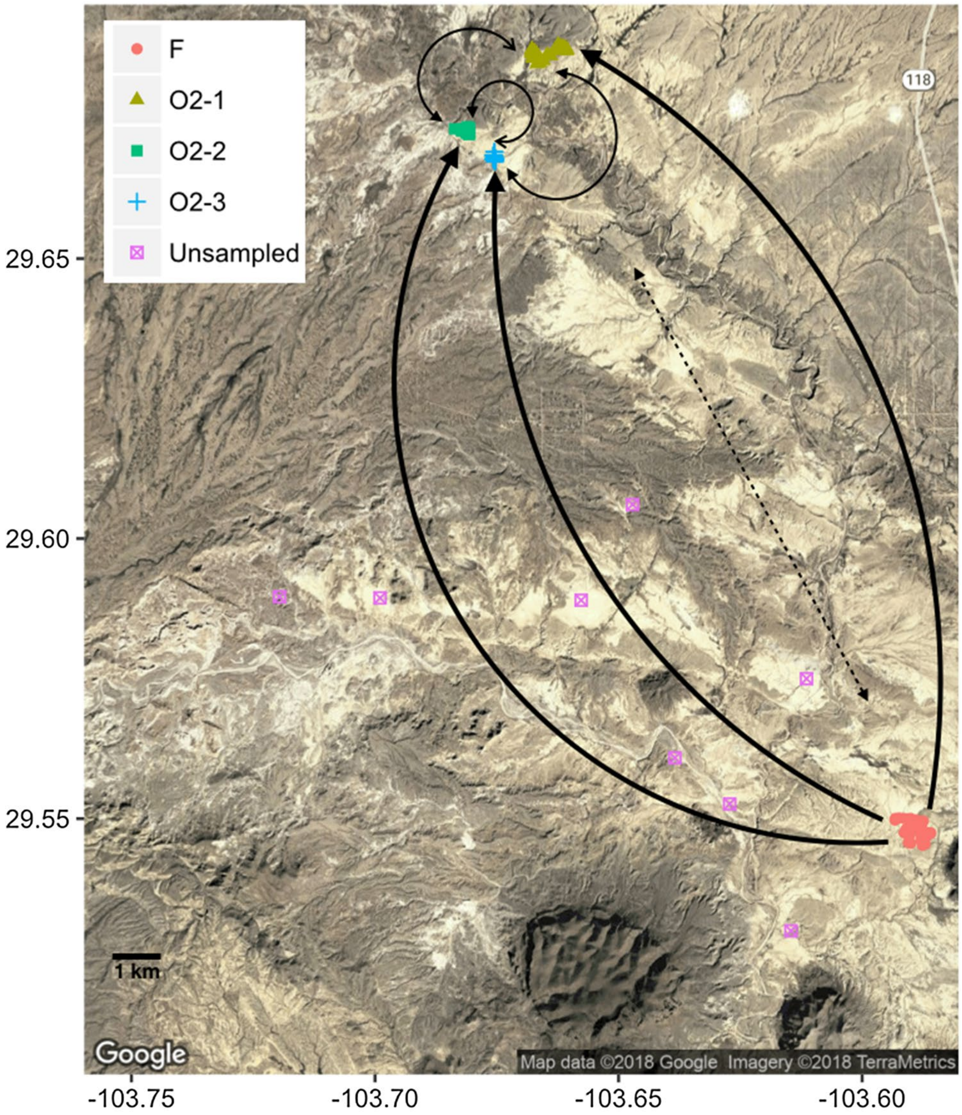

Fig. 1 Map of four studied populations of *Oreocarya crassipes*. F is Field Lab (red circles) and three O2-Ranch populations are O2-1 (olive-green triangles), O2-2 (green squares), and O2-3 (blue crosses). Unsampled populations represented by pink squares with Xs. Latitude and longitude included on axes. Lines and arrows represent patterns of migration between populations. Thick lines denote greater migrants than thin lines. Thick lines also denote a high number of migrants with SNPs, but approximately 10% of that number with microsatellites. Thin lines with closed arrows have similar numbers of migrants with SNP and microsatellites, and those with open arrows have approximately 33% fewer migrants with microsatellites than with SNPs. The dashed line represents patterns of migration with two populations, F and O2 Ranch, and includes equal numbers of migrants. (Color figure online)

least two individuals and 30% of the aligned reads at the position, with a minimum of five reads at a site.

Along with tGBS, leaf material for 184 samples from the four populations (83 F, 55 O2-1, 30 O2-2, 16 O2-3) was sent to the biotechnology company Eurofins BioDiagnostics (https://www.eurofinsus.com/biodiagnostics/) for DNA isolation and microsatellite amplification and sequencing using an ABI 3730XL, with a GeneScan 500LIZ size standard. Ten microsatellite loci that were identified by Bresowar and McGlaughlin (2014) for species of *Oreocarya* were amplified. The length of the amplified region for each locus was scored, by Eurofins BioDiagnostics, for each sample, using GeneMapper software (Applied Biosystems, Foster City, CA, USA). The resulting microsatellite data included one locus that was monomorphic (ORBA 102) and one that appeared to have duplicated in *O. crassipes* (ORPA 521); therefore, the final data set included two alleles arbitrarily selected for the duplicated locus.

Based on preliminary analyses of the putative number of populations (i.e., with fastSTRUCTURE and CloudSTRUCTURE, see analyses in Genetic Diversity and Population Structure section below), the Field Lab population appears genetically isolated from O2-Ranch populations, and the O2-Ranch populations may act as one to three populations. Therefore, two other matrices were constructed, one with three populations (Field Lab, O2-1, and a population consisting of O2-2 and O2-3) and one with two populations (Field Lab and the three O2-Ranch populations), and analyzed along with the four-population matrix. The datasets generated and analyzed in the present study are available at the Dryad repository (XXX).

Genetic diversity and population structure

For investigations of Hardy–Weinberg equilibrium in Arlequin (Excoffier and Lischer 2010), a markov chain of 1,000,000 steps with 100,000 steps as burnin (dememorisation) was used, for tests of population pairwise *F*st, 10,000 permutations for significance, 1000 permutations for the Mantel test were conducted, and 10,000 permutations were undertaken for Analysis of Molecular Variance (AMOVA). In Hierfstat (Goudet 2005; De Meeûs and Goudet 2007) and Genepop (Raymond 1995; Rousset 2008), default settings were used to calculate observed and expected heterozygosity measurements and *F*-statistics. Population pairwise *G*-statistics and Jost's *D* were determined with Genodive (Meirmans and Van Tienderen 2004), and *G*-statistics per population were analyzed using 999 permutations. GenAlEx (Peakall and Smouse 2006) was utilized to generate basic statistics, such as allele numbers, observed and expected heterozygosity, and the fixation index, for the microsatellite data (Table S1).

To better understand population structure, fastSTRUCTURE (Raj et al. 2014), CloudSTRUCTURE (Pritchard et al. 2000; Falush et al. 2003; Wang and Ware 2014), and MavericK (Verity and Nichols 2016) were employed for the SNP and microsatellite data, respectively. For fastSTRUCTURE analyses, K = 1–12 were analyzed for the two SNP datasets, using the Discovery environment on CyVerse (www.cyverse.org), and ten replicates were run for each K, with a convergence criterion of 0.000001, a simple prior, and 100 test sets for cross-validation. For CloudSTRUCTURE, K = 1–12 were analyzed, using the Discovery environment on CyVerse, for the microsatellite datasets using 1,000,000 steps and 500,000 burnin, with alpha and lambda of 1, and with or without various combinations of admixture and the use of prior sampling information (i.e., locprior). Ten replicates were run for each K. The CLUMPAK main pipeline (Kopelman et al. 2015), which employs CLUMPP (Jakobsson and Rosenberg 2007) and DISTRUCT (Rosenberg 2004), was utilized to organize, cluster, and visualize the results of independent fastSTRUCTURE and CloudSTRUCTURE analyses, via 10,000 permutations of the LargeKGreedy algorithm. MavericK analyses were run for K = 1–12 with 10 replications per K, for the microsatellite and LMD10 datasets, with and without admixture, using the following parameters for each replicate: 100,000 steps and 10,000 burnin for markov chain monte carlo sampling and an alpha of 1, 1000 steps and 100 burnin, with 20 rungs, for thermodynamic integration (TI) sampling, and 100 expectation–maximization repeats. These analyses were conducted using Structure Threader (Pina-Martins et al. 2017) on the Kettering University High-Performance Computing Cluster (KUHPC). Results from fastSTRUCTURE and CloudSTRUCTURE were analyzed using Structure Harvester (Earl 2012) and CLUMPAK (Kopelman et al. 2015), and the most appropriate value of K was two for both the method of Evanno et al. (2005) and Ln Pr(X|K). Additionally, the marginal likelihood that maximizes model complexity from fastSTRUCTURE and the TI values from MavericK were noted.

To examine the impact of geography on population structure, tess3r (Caye et al. 2017) was employed using the alternating projected least squares method, for K = 1–12. Results for each K were visualized with bar graphs and maps. With tess3r, a cross-entropy plot was used to identify the most appropriate K value for each of the datasets analyzed. Along with explicit tests for population structure, the SNP datasets were examined using principal component analyses (PCA) of genetic covariance, using SNPRelate (Zheng et al. 2012), and PCA as well as correspondence analyses (CA) were conducted with Adegenet (Jombart and Ahmed 2011) for the three datasets.

In addition to analyses with pre-defined population clusters, discriminant analysis of principal components (DAPC)

(Jombart et al. 2010; Jombart and Ahmed 2011) was used to investigate the most appropriate number of clusters based on the data, without the use of prior population assignation. For the LMD10, LMD50, and microsatellites datasets, five, three, and nine clusters were recognized as optimal, respectively, based on the Bayesian Information Criterion (BIC), with 40, 60, and 18 PCs retained, respectively, resulting from the DAPC cross-validation to identify the most appropriate number of PCs to use.

Isolation-by-distance analyses were undertaken with Adegenet (Jombart and Ahmed 2011) and the Isolation by Distance Web Service (IBDWS) (Jensen et al. 2005). In Adegenet, separate Mantel tests, for the SNP and microsatellite matrices, were conducted for populations and for individuals, and each Mantel test included 999 simulated replicates. Only the microsatellite data were analyzed using the Isolation by Distance Web Service, and *F*st, Slatkin's Similarity Index, and Rousset's distance were each measured with 30,000 randomizations.

Patterns of selection, migration, and demography

Loci under selection were analyzed with BayeScan (Foll and Gaggiotti 2008) and LEA (Frichot and François 2015). For each dataset, 100,000 iterations were run, with a burnin of 50,000 iterations, a thinning interval of 10, and a sample size of 5000, and for each analysis, 20 pilot runs were conducted, each with 5000 steps. Loci under selection were visualized in R (R Core Team 2013) using *F*st values and a false discovery rate of 0.05. For LEA, the snmf function was employed to identify population structure and to recognize outlier loci, those being loci with an adjusted *P* value of 0.05 or less. Given that the results from BayeScan are more conservative than those from LEA (i.e., many fewer loci identified as under selection) and that *O. crassipes* lacks an annotated genome, the putative identity and function of the loci identified as under selection using BayeScan were examined with Blast2GO (Conesa et al. 2005), using blastx.

Bottleneck (Piry et al. 1999) and Migrate-N (Beerli 2009; Beerli and Palczewski 2010) were employed to investigate the demographic history, patterns of migration, and population sizes in *O. crassipes*. The LMD10 and microsatellite matrices were analyzed with Bottleneck, using 1000 replications and the IAM, TPM, and SMM models in order to identify loci that differ from the mutation-drift equilibrium under the various models.

For Migrate-N, two groupings of populations were analyzed—four populations (F, O2-1, O2-2, and O2-3) and two populations (F and O2 Ranch [O2-1, O2-2, and O2-3])—with the LMD10 and microsatellite datasets. For Migrate-N with SNP data, Bayesian analyses were undertaken using the following initial parameters: theta and migration equal to *F*st, a constant mutation rate across loci, theta (θ) as *F*st, θ and migration with uniform priors, four markov chains for 40,000,000 steps (40,000 recorded steps every 1000 steps) with a burnin of 5000 steps, and heated swapping (1, 1.5, 3, and 1,000,000 as heating scheme). For analyses with the microsatellite data, the same parameters were used, but the four markov chains ran for 80,000,000 steps (80,000 recorded steps every 1000 steps) with a burnin of 10,000 steps. For each of the population groupings, three Migrate-N analyses were run, one with equal migration among the four populations, one with migration only from F to the O2 populations, and one with migration only from the O2 populations to F. For the latter two analyses, migration was allowed between pairs of O2-Ranch populations. For each analysis, skyline plots were produced, and along with values of θ and the mutation-scaled migration rate (M), the effective number of migrants was determined using the equation $Nm = \theta M/4$ (Yuan et al. 2014). All Migrate-N analyses were run on the CIPRES server (www.phylo.org).

Identification of conservation and management units

To identify appropriate units for management and conservation, the method of Funk et al. (2012) was employed. This method involves three steps: (1) delineating evolutionary significant units (ESUs) using all loci, (2) recognizing management units (MUs) via non-outlier loci, and (3) identifying adaptive variation among management units through the use of outlier loci. The first step was described above with analyses in fastSTRUCTURE (Raj et al. 2014) and CLUMPAK (Kopelman et al. 2015) as well as with PCA and DAPC with Adegenet (Jombart and Ahmed 2011). For steps two and three, utilizing the same methods, four SNP datasets based on the LMD50 dataset—outlier and non-outlier loci from BayeScan (Foll 2012) and from LEA (Frichot and François 2015)—were constructed and analyzed to identify population structure. For DAPC, 80 PCs were optimal, based on cross-validation tests, for all of the datasets apart from the one that included the BayeScan loci under selection, for which 40 PCs was identified as optimal. The LMD10 and microsatellite datasets were not included due to their lack of outlier loci. Based on the results of these analyses of population clusters, ESUs, MUs, and MUs with similar adaptations were identified.

Results

Genetic diversity and population structure

For SNP data, observed heterozygosity ranged from 0.099 in F for LMD10 to 0.244 for O2-3 for LMD50 (Table 1). In general, increasing SNPs resulted in greater observed

Table 1 Observed (Ho) and expected (He) heterozygosity and theta (θ) for four studied populations, for SNP datasets and 10 microsatellites

Dataset	Population	Ho	He	θ
LMD10	Field lab	0.099 ± 0.095	0.104 ± 0.095	0.088
	O2-1	0.126 ± 0.115	0.129 ± 0.107	0.114
	O2-2	0.149 ± 0.149	0.140 ± 0.118	0.092
	O2-3	0.185 ± 0.149	0.179 ± 0.126	0.095
LMD50	Field lab	0.146 ± 0.146	0.172 ± 0.150	0.172
	O2-1	0.167 ± 0.154	0.188 ± 0.150	0.175
	O2-2	0.215 ± 0.180	0.222 ± 0.157	0.170
	O2-3	0.244 ± 0.186	0.253 ± 0.153	0.170
Microsatellites	Field Lab	0.368 ± 0.226	0.418 ± 0.195	1.628
	O2-1	0.376 ± 0.238	0.421 ± 0.186	1.624
	O2-2	0.411 ± 0.247	0.424 ± 0.211	1.617
	O2-3	0.427 ± 0.275	0.451 ± 0.174	1.617

Table 2 Weir and Cockerham's *F*st (above diagonal) and Nei's *F*st (below diagonal) for SNP datasets and 10 microsatellites

	Field lab	O2-1	O2-2	O2-3
LMD10				
Field lab	–	0.054	0.067	0.070
O2-1	0.052	–	0.046	0.049
O2-2	0.066	0.048	–	0.029
O2-3	0.069	0.052	0.029	–
LMD50				
Field lab	–	0.078	0.099	0.097
O2-1	0.077	–	0.056	0.059
O2-2	0.099	0.056	–	0.029
O2-3	0.098	0.060	0.029	–
Microsatellites				
Field lab	–	0.125	0.152	0.082
O2-1	0.124	–	0.007	0.028
O2-2	0.151	0.007	–	0.018
O2-3	0.080	0.028	0.017	–

heterozygosity. Expected heterozygosity was close to the observed heterozygosity, with most differences being less than 0.01 (Table 1). Based on AMOVA analyses in Arlequin, the majority of the variation, over all loci, was within individuals (81–93%), with only small percentages of variation among populations (5–10%) or among individuals within populations (1–10%).

Values for Nei's and Weir and Cockerham's *F*st are listed in Table 2. All pairwise *F*st values were statistically significant at $P < 0.01$, except for comparisons with microsatellite loci between O2-1 and O2-2 and between O2-2 and O2-3. The results from Hedrick's *G*st and Jost's *D* showed similar results to those from *F*st analyses (Table S2). In general, *F*is values are slightly greater than 0, ranging from 0.02 to 0.15 (Table 1). LMD10 had two populations, O2-2 and O2-3, in which *F*is values were slightly negative, and microsatellite data resulted in positive *F*is values.

The methods of Evanno et al. (2005), Ln Pr(X|K), and TI integration identified the optimal number of clusters (K) as two (Figs. S1, S2) (one for analyses of LMD10 with admixture), and this was also the case for the model complexity that maximized the marginal likelihood from fastSTRUCTURE for LMD50 and some analyses of LMD10; although, for most analyses with LMD10, this value was one. With K = 2 for SNP data, most individuals of F were assigned to one population, and most individuals from the O2 Ranch were assigned to another (Fig. 2). With the microsatellite data, less mixed ancestry was seen between the two clusters. With K = 3 or higher, three clusters became apparent in most of the SNP datasets; however, these results were sometimes restricted to minor clusters. The greater number of clusters resolved a distinction between F, O2-1, and O2-2 and O2-3, and mixed ancestry was still evident. Some substructure was observed based on analyses with microsatellite data and the admixture model, but the clusters were not as defined as with the SNP data (Fig. 2). The SNP and microsatellite data produced similar results (i.e., number of clusters and ancestry), with two differences. One is that most SNP data identified three discernable clusters with K = 3 or higher. The second is that, with microsatellite data, the O2-3 population has more mixed ancestry with F than occurred with the SNP data, and the other O2-Ranch populations have less. With the microsatellite data, F is recognized as much more genetically homogenous than is the case with the SNP data (Fig. 2). Without admixture included in the model, F and O2-Ranch populations are two distinct clusters, at least until K = 8, and

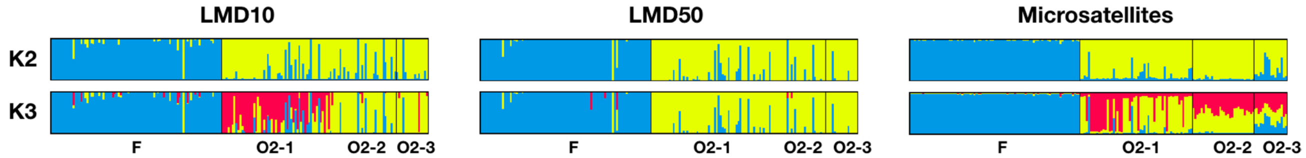

Fig. 2 Bar graphs of population assignment for K = 2 and 3, based on analyses of fastSTRUCTURE and CloudSTRUCTURE, for LMD10 and LMD50 and 10 microsatellites; F is Field Lab and three O2-Ranch populations are O2-1, O2-2, and O2-3

without the use of prior sampling information, population structure becomes fairly homogenous at K = 4 and higher. Analyses with MavericK and tess3r provided similar results to those from fastSTRUCTURE and CloudSTRUCTURE, (Figs. 2, S5).

The results from the SNPrelate and Adegenet PCA also recognize three clusters corresponding to F, O2-1, and O2-2 and O2-3 (Fig. 3a–c). Clusters from LMD50 are more distinct, from each other, than those of LMD10. Even though three clusters are apparent, multiple individuals from each of the clusters group together in the middle of the PCA graphs. The PCA with the microsatellite data results in much less differentiation among populations, with little distinction between the three O2-Ranch populations.

The results of the DAPC are similar for the SNP datasets, with F, O2-1, and O2-2 and O2-03 resolved as clusters, including two clusters recognized for individuals of F and of O2-1 (Fig. 3d–f). With the SNP data, however, some individuals show mixed ancestry, as is the case with other methods. The microsatellite data recognized nine clusters, and while some clusters are composed only of individuals from F or of O2-Ranch populations, this is not the case for all clusters. This results in less distinction among the identified clusters; however, clusters composed primarily of individuals from one population that group together are most distinct from other clusters (e.g., clusters 1, 5, and 7 of Fig. 3f).

Analyses of isolation-by-distance (IBD) from the IBDWS demonstrate that as distance increases between pairs of populations, *F*st and genetic distance increase as well (Fig. S6). The three pairs of O2-Ranch populations are more similar to each other than any is to F, and this difference in *F*st or genetic distance tends to follow a linear trend from closer to more distant populations. When IBD is investigated with Mantel tests, population differentiation, based on distance, is not significantly different. However, the use of individuals, as opposed to populations, results in statistically significant isolation among populations ($P < 0.001$). Two clusters of individuals are identified by the Mantel test, which, again, is similar to results from other analyses of population structure. Analyses of each population demonstrate that genetic distance increases with greater geographic distance, with the exception of microsatellite data and O2-3, in which a

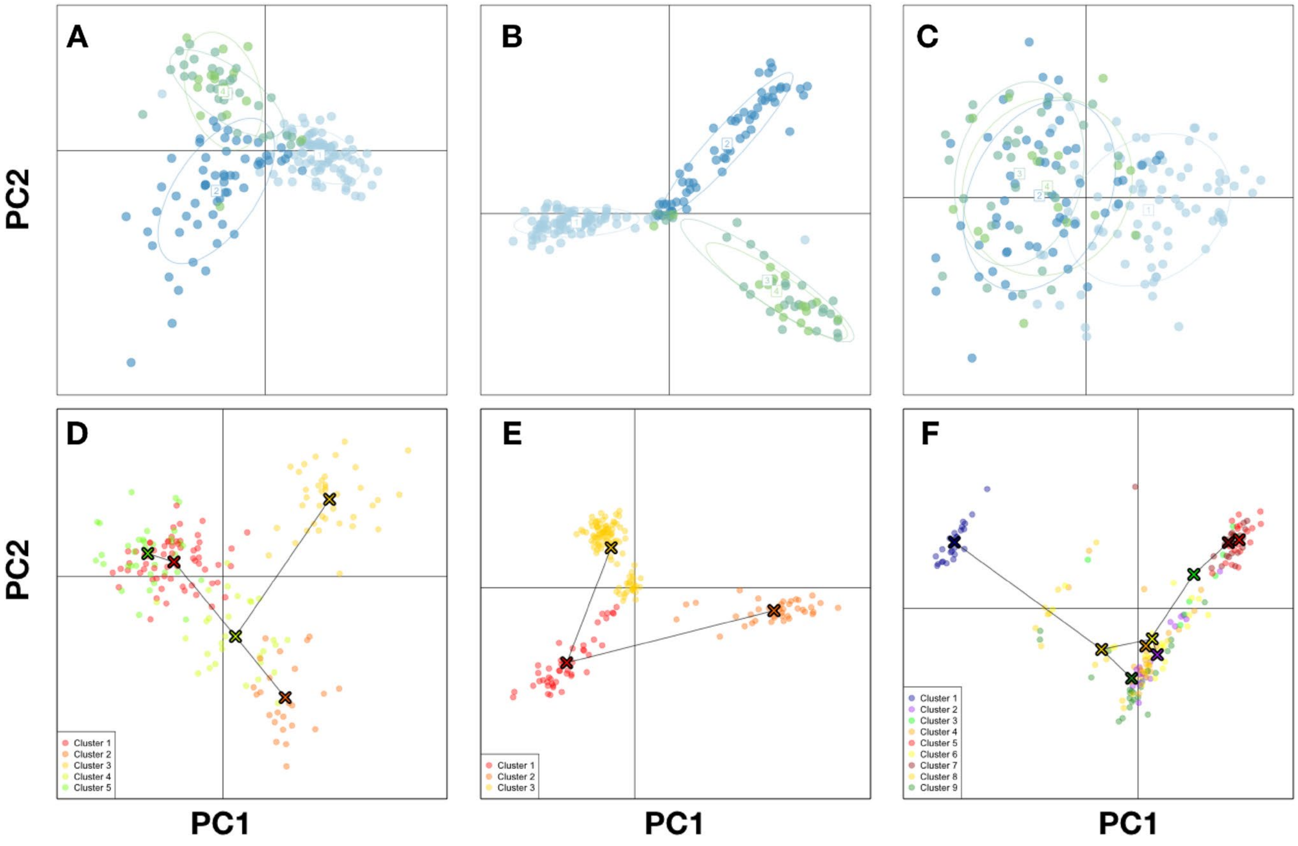

Fig. 3 PCA graphs of first and second eigenvectors based on **a** LMD10, **b** LMD50, and **c** microsatellite datasets. DAPC graphs of first and second eigenvectors based on **d** LMD10, **e** LMD50, and **f** microsatellite datasets. 1 is Field Lab and three O2-Ranch populations are 2, O2-1, 3, O2-2, and 4, O2-3. Ellipses are 95% confidence ellipses. X denotes centers of clusters, and lines demonstrate connections between clusters

very slight decrease is observed (Figs. S7–9). Of the four populations, only O2-1 has the observed correlation outside of the simulated values.

Patterns of selection, migration, and demography

With four populations, analyses with BayeScan resulted in zero and 103 SNPs under selection for LMD10 and LMD50, respectively. No microsatellite loci were under selection. With fewer populations (e.g., O2-2 and O2-3 treated as one population), the number of loci under selection was less for LMD50, with only four SNPs identified when inferring two populations (F and all O2-Ranch populations). Analyses with LEA identified much larger numbers of SNPs under selection. With the LMD10 and LMD50 datasets, 137 and 5299 SNPs were recognized as under selection, respectively, for four recognized population clusters.

The 103 SNPs identified as under selection in the LMD50 dataset, via BayeScan, are present among 90 genes, all of which are ca. 160 bp in length. The putative identity of 24 of these genes could be determined via blastx. The most common Gene Ontology (GO) terms were DNA metabolic processes, transport, and cellular component organization for biological processes; extracellular region and mitochondrion for cellular component; and nucleic acid binding, hydrolase activity, and nucleotide binding for molecular function.

Analyses of genetic bottlenecks based on SNP data resulted in more loci than expected as heterozygote deficient (Tables S3, S4), but only the O2-Ranch populations were identified as being shifted based on the mode-shift test. The opposite pattern is observed with the microsatellite data, with more loci found as being in excess of heterozygotes than deficient. Most of these differences were not statistically significant, but F was recognized as having a shifted distribution based on the mode-shift test. This was the only population not found to have a shifted distribution with SNP data, and these different results are similar to the situation with *F*st values between the SNP and microsatellite data.

Analyses of SNP and microsatellite data, with Migrate-N, identified similar patterns of migration, but the number of populations employed resulted in different preferred models of migration, as determined by the model probabilities and log Bayes factors. With four populations, the preferred model was one of migration from F to the O2-Ranch populations and with migration possible among O2-Ranch populations (Fig. 1). In contrast, with only two populations, bidirectional migration between the two populations was preferred, and this was the case with both types of data. Under the unidirectional F-to-O2-Ranch model, θ of F was approximately 50% greater than that for the O2-Ranch populations (0.07 vs 0.045). Based on SNP data, approximately 50–100% more migration occurred from F to the O2-Ranch populations than was taking place among the three O2-Ranch populations; however, with microsatellites, migration was two to seven times as much among the O2-Ranch populations than between F and any O2-Ranch population. With only two populations, values of θ are relatively equal between the two populations (0.06 for SNPs and 0.95 for microsatellites). The rate of migration is also approximately equal between the two, although the rate was greater with the use of SNP data than with microsatellites (Fig. 1, Table S5).

Identification of conservation and management units

Results of the analyses of the multiple datasets allow for the identification of populations that can be used for conservation and management units. Results of fastSTRUCTURE analyses for the two datasets that include loci not under selection have an optimal K = 2, based on the model complexity that maximized the marginal likelihood from fastSTRUCTURE, and this results in F distinguished from the O2-Ranch populations (Fig. 4a). For these non-outlier-loci datasets derived from LEA and BayeScan analyses, two and three clusters were identified, respectively, as optimal based on DAPC. These clusters correspond to F and O2-Ranch populations and primarily to F, O2-1, and O2-2 and O2-3, respectively, and the PCA produces similar results (Fig. 4c, e).

Results of the fastSTRUCTURE analyses for the two datasets that include loci under selection have an optimal K = 3, based on the model complexity that maximized the marginal likelihood, with F, O2-1, and O2-2 and O2-3 distinguished from each other (Fig. 4b). For the outlier-loci datasets derived from LEA and BayeScan analyses, three and six clusters were identified, respectively, as optimal based on DAPC. The larger number of clusters from the DAPC, based on BayeScan loci, is similar to that of the model components used to explain the structure of data, ranging from five to nine, and more clusters allow for the recognition of greater population substructure among the three O2-Ranch populations. These clusters correspond to F, O2-1, and O2-2 and O2-3, and to F and various permutations of the O2-Ranch populations, respectively. With the BayeScan-selected-loci dataset, O2-2 and O2-3 are divided into three clusters, and O2-1 is separated in two, with one also including individuals from O2-2 to O2-3. While the PCA results from the loci-under-selection datasets are similar to those based on loci not under selection, one notable difference involves greater separation between O2-2 and O2-3 with the former compared to the latter (Fig. 4d, f).

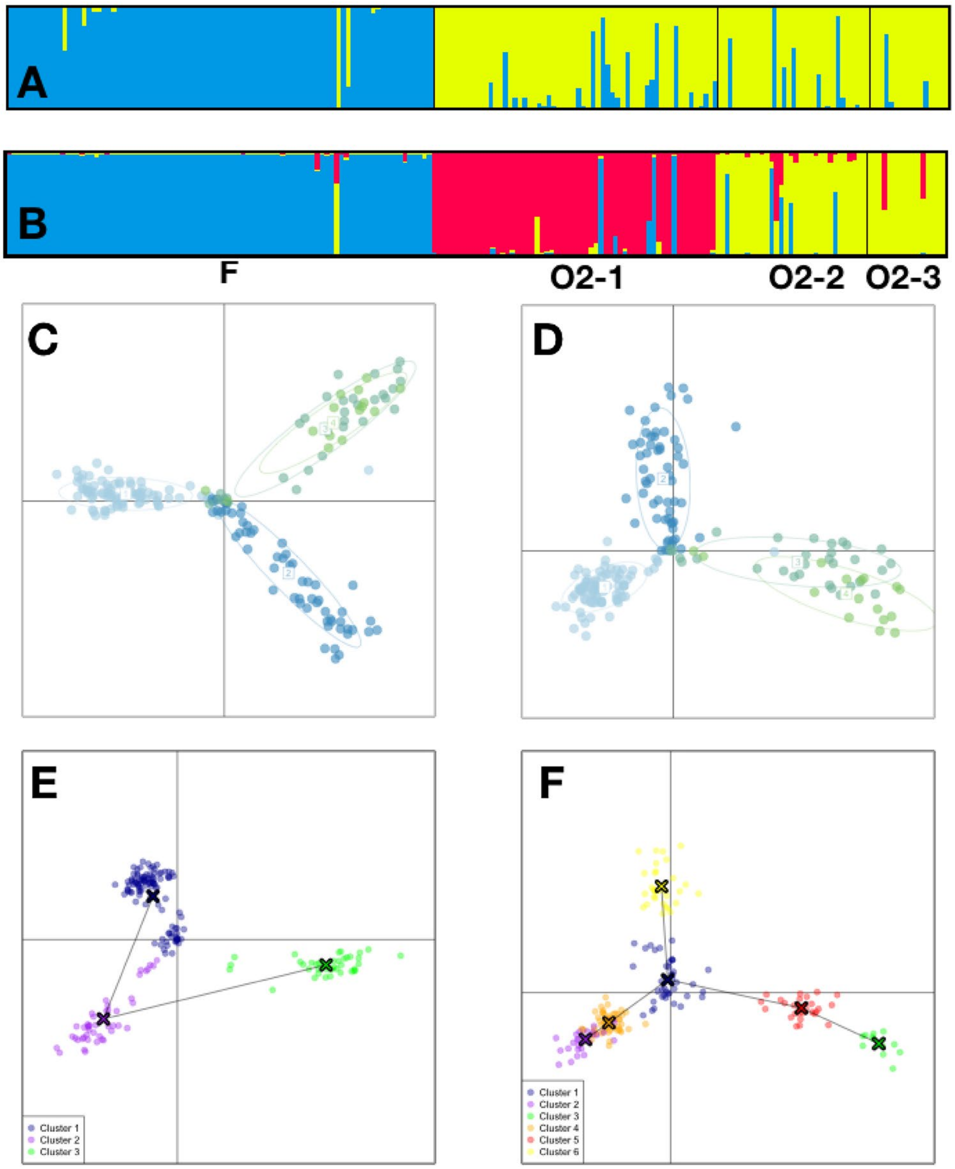

Fig. 4 Bar graphs of population assignment, based on analyses of fastSTRUCTURE, for BayeScan genes not under selection (**a**) and under selection (**b**), PCA graphs of first and second eigenvectors for BayeScan genes not under selection (**c**) and under selection (**d**). 1 is Field Lab and three O2-Ranch populations are 2, O2-1, 3, O2-2, and 4, O2-3. DAPC graphs of first and second eigenvectors based on BayeScan genes not under selection (**e**) and under selection (**f**). X denotes centers of clusters, and lines demonstrate connections between clusters

Discussion

Genetic diversity and population structure

For the four sampled populations of *O. crassipes*, two distinct clusters of individuals were recognized via multiple methodologies and datasets. These clusters correspond to the Field Lab (F) population and the three O2-Ranch populations, and these results are congruent with the hypothesis of the present study concerning population structure between the northern and southern ends of the geographic range of *O. crassipes*. Furthermore, with most of the SNP datasets, three distinct clusters were resolved, suggesting population substructure among the O2-Ranch populations, with O2-1 distinguished from O2-2 to O2-3 (Table 2, Figs. 2, 3). While three populations were not identified as the optimal number by multiple methods for determining the most appropriate K, results from various analyses provide evidence of sufficient variation in the species to recognize three groups among the four sampled populations. Substructure among the O2-Ranch populations was only clearly identified with the use of SNP data, not microsatellite data (Figs. 2, 3). In a study on the phylogeography of Crucian carp (*Carrassius carrassius* L.), Jeffries et al. (2016) recovered a similar result when comparing these two types of data suggesting that increased population substructure with SNP data is applicable across kingdoms. The lack of identification of population substructure with microsatellite data may be due to the low number of microsatellite loci included in the study, the mutation rate of these loci being too high to adequately distinguish substructure (Coates et al. 2009; Fischer et al. 2017), which is also supported by the larger values of θ from the microsatellites than from the SNPs, or different numbers of individuals included among the populations.

A small number of individuals show mixed ancestry, and this is more prevalent from SNPs than from microsatellites, which may be due to the smaller number of microsatellite loci employed compared to SNPs or to different patterns of mutation between the two types of loci. With the microsatellite data, more individuals from the O2-Ranch populations are resolved to have genetic similarity to individuals from F than the opposite pattern. Regardless, these analyses provide evidence of genetic exchange among the populations, which appears to have helped ensure diversity and lack of inbreeding depression within populations of *O. crassipes*. Although small to moderate amounts of inbreeding (*F*is) were identified in the populations, given the inhospitable Big Bend habitat and edaphic specialization of *O. crassipes*, more inbreeding within populations would likely be expected. This is the case for other gypsum species that have greater population differentiation, and these include *Arctomecon humilis* Coville (Papaveraceae) and *Boleum asperum* Desv. (Cruciferae), which are from different areas of the angiosperm phylogeny than *O. crassipes* (Moore et al. 2014). However, the habitat issues for *O. crassipes* do not seem to inhibit migration, a result further supported by the vast majority of the genetic variation identified as within individuals.

While the presented results may suggest that the populations were more contiguous until recently, the edaphic requirements for the species along with the patchy gypsum soil make this possibility less likely. Additionally, *O. crassipes* is a long-lived, allogamous perennial, and this type of life history results in less population differentiation than alternative life histories (Nybom 2004). Collectively, these opposing processes could create conditions under which the majority of the genetic variation in *O. crassipes* is within individuals, but distinct populations are still recognized.

Unfortunately, it was not possible to sample individuals from all of the populations of the species, and the F and O2-Ranch populations represent the second southernmost and most northern, respectively, known populations, with at least two populations in between F and the O2 Ranch (Fig. 1) (Warnock 2012). These intermediate populations could allow for a stepping-stone model of migration between the two extreme ends of the geographic range of the species. Future studies involving individuals from these populations would help to further understand the relationship across the geographic range of the species.

Recognizing patterns of population structure of *O. crassipes* can assist with *in* and ex situ conservation efforts by identifying patterns of genetic diversity within and among populations. Following the methodology of Funk et al. (2012), two evolutionary significant units (ESUs) were identified based on the use of all loci for analyses, and these ESUs correspond to F and to the O2-Ranch populations (Fig. 2). These two clusters were also recovered with the use of non-outlier loci; therefore, these two resolved populations should be treated as distinct management units (MUs) for *O. crassipes*. Floral morphological differences were also found between the F and O2-Ranch populations (Cohen et al. In review). With the use of outlier loci, multiple groups feature adaptive differentiation among the MUs. While F is considered to be one MU with its own adaptive features, O2-1 is separate from O2-2 to O2-3 as a distinctive cluster with its own adaptive characteristics (Fig. 4). Identification of three adaptive groups is consistent with fastSTRUCTURE and PCA analyses of the outlier loci; although, the DAPC results can provide even greater population substructure among the O2-Ranch population. It is useful to identify this greater variation within the population, but given the geographic separation of F, O2-1, and O2-2 and O2-3, it seems prudent to recommend that these three groups be used for in situ conservation management. For ex situ conservation efforts, it would be useful to sample individuals of F and, possibly more importantly, broadly among the three O2-Ranch populations to ensure the greatest amount of genetic diversity and adaptive variation.

Patterns of selection, migration, and demography

A limited number of SNPs (at most 0.6%), but no microsatellite loci, were recognized as being under selection. The lack of an annotated genome for *O. crassipes* makes it challenging to fully identify the function of the genes under selection; however, the Blast2GO analyses shed a small amount of light on this subject (Conesa et al. 2005). Of the genes that could be identified, one is a putative homologue of E3 ubiquitin-protein ligase, and another is resolved as a putative pentatricopeptide repeat-containing protein. Researchers have identified both genes as important for responding to abiotic stress (Lee and Kim 2011; Liu et al. 2016). These types of genes would be quite useful for adaptation to and success in the xeric environment *O. crassipes* inhabits. Additional data on patterns of selection and differential expression of plants growing on and off gypsum soil would provide further information on genes that help plants adapt to these edaphic conditions.

With SNP data, the O2-Ranch populations were identified as having undergone recent genetic bottlenecks, and with microsatellite data, this was the case for F. While these results may seem incompatible, these two different types of markers appear to be capturing genetic bottlenecks at different, but still relatively recent, times (Yuan et al. 2014), and skyline plots of θ for the four populations resolve similar patterns. These bottlenecks could be due to the increasing severity of aridity in the region during the past 6000 years (Elias and Van Devender 1990), with the populations being impacted at different times. These bottlenecks may be one reason that geographically close populations are recognized as distinct, even with putative migration between them. It is useful to note that it would not have been possible to identify genetic bottlenecks in the various populations if only one type of marker was employed, and this provides an additional, compelling reason to, if possible, include in analyses multiple types of markers, each with a different mutation rate.

As hypothesized, migration occurs among the populations. Similar patterns of migration among populations were recovered with both SNP and microsatellite data, but the rates of migration between F and the O2-Ranch populations were greater with the former than the latter (Table S5). Results from microsatellites appear more in line with the geographic proximity of the studied populations, with greater migration among the O2-Ranch populations than between these populations and F. However, given the lack of sampling of populations in between F and the O2 Ranch, it may be possible that migration is more equal between these populations than the present study can identify should a stepping-stone model involving unsampled, intermediate populations be occurring (Fig. 1).

It is notable that the microsatellite data resolved more genetic similarity between F and O2-3, a result not recovered with SNPs. Given the preferred direction of migration, the microsatellite data, which has a higher rate of mutation than SNPs, may be capturing a record of recent migration from F to O2-3, resulting in the different *F*st values from the two types of data. In all pairs of populations, migration has been increasing recently, suggesting that there may have been barriers to gene flow across the limited geographic range of the species; however, the current, larger barriers would tend to be inhospitable habit with non-gypsum soil between populations rather than tall or deep physical barriers. Since the last ice age, the climate of the Big Bend region has warmed and become drier (Elias and Van Devender 1990; Herbert 2006), and these environmental conditions may have helped facilitate increased gene flow among populations of *O. crassipes*.

One important question remains concerning genetic material moving among populations: is this via pollen, fruits (nutlets), or both? While the particular insect pollinators continue to remain undiscovered, floral visitors include Lepidoptera and Hymenoptera (Warnock 2012; pers. obs.). These insects are able to travel among populations of *O. crassipes* (Ratnieks and Shackleton 2015), which would facilitate pollen dispersal, even if longer-distance pollen dispersal occurs infrequently. Wind dispersal of nutlets could result in longer-distance dispersal than via pollinators (Casper and Wiens 1981; Maddox and Carlquist 1985; Casper 1988), but the dispersal distance of the nutlets of *O. crassipes* has not been studied. These two possibilities can be examined in more detail to better elucidate patterns of migration among the populations and to identify if a stepping-stone model for migration is taking place.

Impact of breeding system and edaphic specialization on *O. crassipes*

The presence of distyly does not appear to impact the population structure of *O. crassipes*. This could be due to the relatively even morph ratio identified within and among the studied populations (Cohen et al. In review). Shibayama and Kadono (2007) found that in the distylous *Nymphoides indica* (L.) Kuntze (Menyanthaceae), populations with more equal morph ratios had greater genetic diversity due to shorter distance for pollinators to travel between morphs for legitimate pollination events, and skewed morph ratios decrease the likelihood of effective pollination between compatible morphs, an Allee effect (Stephens and Sutherland 1999). Other rare distylous species, such as *Primula merrilliana* Schltr. (Primulaceae) (Shao et al. 2009) and *Amsinckia grandiflora* (A. Gray) Kleeb. ex Greene (Pavlik et al. 1993), that have been studied have larger geographic ranges than *O. crassipes*, and this makes useful comparisons challenging; however, studies of genetic diversity in populations of the distylous species *Pulmonaria officinalis* have identified *F*st values similar to those in the present study (Meeus et al. 2012). Therefore, the studied populations of *O. crassipes* are as genetically distinct as a more widespread relative. The observed genetic patterns may not be the result of distyly specifically, but could be due to the facilitation of allogamy by distyly. Given the necessity of outcrossing among individuals of allogamous species, these species tend to have less population differentiation than autogamous species, and this is observed in *O. crassipes* (Nybom 2004).

In addition to its unique breeding system, *O. crassipes* is adapted to gypsum soils. Researchers have utilized various molecular markers, primarily allozymes and AFLPs, to study the population genetics of approximately a dozen gypsophyles from across the angiosperms (Moore et al. 2014), and the observed and expected values of heterozygosity for *O. crassipes,* based on SNPs, are similar to those from other gypsophiles and rare species, such as *A. humilis* (ca. 0.1–0.15) (Gitzendanner and Soltis 2000; Cole 2003; Moore et al. 2014), with these values from microsatellites slightly higher (Nybom 2004). In contrast, the *F*st values for *O. crassipes* (ca. 0.5–0.15) are much lower, both for pairwise comparisons and for the species, than other gypsophiles (0.42). Indeed, the *F*st values, for both types of data, for *O. crassipes* are much lower than for all other studied gypsophiles, apart from *Ferula loscosii* Willk. (Apiaceae) (Moore et al. 2014), a species that also has similar a R^2 value from IBDWS analyses to those from *O. crassipes* (0.5–0.87, depending on the type of analysis). The *F*st values in *O. crassipes* are more similar to those from other rare or endemic species (ca. 0.2) (Gitzendanner and Soltis 2000; Cole 2003; Moore et al. 2014). These *F*st values provide evidence of greater exchange of genetic material in *O. crassipes* than in other gypsum species, and this increased gene flow could be due to the small geographic range of the species, the possibility of a stepping-stone model of dispersal, the efficient pollination due to distyly, the long-lived perennial life form, or others.

Bresowar and McGlaughlin (2015) utilized 11 microsatellite loci to investigate the population genetics of two related, gypsophilic species of *Oreocarya*, *O. paradoxa* A. Nelson and *O. revealii* W. A. Weber & R. C. Wittmann, both of which have greater geographic ranges than *O. crassipes*. In their study of 21 populations of the two species, Bresowar and McGlaughlin (2015) identified two clusters (K), each corresponding to one of the species. The populations of these two species have larger *F*is values (0.3–0.5) than those of *O. crassipes* (0.05–0.12), providing evidence for greater inbreeding in the two species. It is notable to observe greater population structure in *O. crassipes* than in *O. paradoxa* and *O. revealii*; although, this may be due to the recent divergence of the two species (Bresowar and McGlaughlin 2015) or the use of many SNPs in the present study because the results from the microsatellites did not allow for the recognition of population substructure in *O. crassipes* (Figs. 2, 3, 4). Collectively, the results from the present study inform conservation and management of *O. crassipes* and paint a hopeful picture for the long-term persistence and viability of the species, at least from the perspective of genetics, migration, and demography.

Acknowledgements D. Garcia, J.-M. Choi, M. Williams, and B. Warnock provided wonderful assistance with field work. J. Wells, B. Gardiner, and H. Mills and the O2 Ranch allowed access to property for sampling plants, and the project would not have been possible without their cooperation. C. Ritzi and students identified floral visitors. L. G. Ruane and four reviewers provided helpful comments on the manuscript. A. M. Powell and C. D. Kellogg supported and encouraged the successful completion of the project. Funding for the project came from four sources: primarily from a traditional Section 6 Grant from the Texas Parks and Wildlife (TX E-160-R), from Texas A&M International University and Kettering University, and from the National Science Foundation Major Research Instrumentation Program (Award Number 1725938) for the KUHPC.

References

Barrett SC (1992) Evolution and function of heterostyly. Springer Science & Business Media, Berlin

Barrett S, Shore J (2008) New insights on heterostyly: comparative biology, ecology and genetics. Self-incompatibility in flowering plants. Springer, Berlin, pp 3–32

Barrett SC, Jesson LK, Baker AM (2000) The evolution and function of stylar polymorphisms in flowering plants. Ann Bot 85:253–265

Beerli P (2009) How to use MIGRATE or why are Markov chain Monte Carlo programs difficult to use. Popul Genet Anim Conserv 17:42–79

Beerli P, Palczewski M (2010) Unified framework to evaluate panmixia and migration direction among multiple sampling locations. Genetics 185:313–326

Bresowar GE, McGlaughlin ME (2014) Characterization of microsatellite markers isolated from members of Oreocarya (Boraginaceae). Conserv Genet Resour 6:205–207

Bresowar GE, McGlaughlin ME (2015) Morphological and genetic discrepancies in populations of *Oreocarya paradoxa* and *O. revealii*: the impact of edaphic selection on recent diversification in the Colorado Plateau. Am J Bot 102:1647–1658

Casper BB (1985) Self-compatibility in distylous *Cryptantha flava* (Boraginaceae). New Phytol 99:149–154

Casper BB (1988) Post-dispersal seed predation may select for wind dispersal but not seed number per dispersal unit in *Cryptantha flava*. Oikos 52(1):27–30

Casper BB, Wiens D (1981) Fixed rates of random ovule abortion in *Cryptantha flava* (Boraginaceae) and its possible relation to seed dispersal. Ecology 62:866–869

Caye K, Jay F, Michel O, Francois O (2017) Fast inference of individual admixture coefficients using geographic data. bioRxiv. https://doi.org/10.1101/080291

Coates BS, Sumerford DV, Miller NJ, Kim KS, Sappington TW, Siegfried BD, Lewis LC (2009) Comparative performance of single nucleotide polymorphism and microsatellite markers for population genetic analysis. J Hered 100:556–564

Cohen JI (2010) "A case to which no parallel exists": the influence of Darwin's different forms of flowers. Am J Bot 97:701–716

Cohen JI (2014) A phylogenetic analysis of morphological and molecular characters of Boraginaceae: evolutionary relationships, taxonomy, and patterns of character evolution. Cladistics 30:139–169

Cohen JI, Rodriguez H, Hutcheson H (In review) Distyly in *Oreocarya crassipes* (Boraginaceae), an endangered species. Southwest Nat

Cole CT (2003) Genetic variation in rare and common plants. Annu Rev Ecol Evol Syst 34:213–237

Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21:3674–3676

De Meeûs T, Goudet J (2007) A step-by-step tutorial to use HierFstat to analyse populations hierarchically structured at multiple levels. Infect Genet Evol 7:731–735

Earl DA (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour 4:359–361

Elias SA, Van Devender TR (1990) Fossil insect evidence for late Quaternary climatic change in the Big Bend region, Chihuahuan Desert, Texas. Quat Res 34:249–261

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14:2611–2620

Excoffier L, Lischer HE (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour 10:564–567

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164:1567–1587

Fischer MC, Rellstab C, Leuzinger M, Roumet M, Gugerli F, Shimizu KK, Holderegger R, Widmer A (2017) Estimating genomic diversity and population differentiation–an empirical comparison of microsatellite and SNP variation in Arabidopsis halleri. BMC Genom 18:69

Foll M (2012) Bayescan v2. 1 user manual. Ecology 20:1450–1462

Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. Genetics 180:977–993

Frichot E, François O (2015) LEA: an R package for landscape and ecological association studies. Methods Ecol Evol 6:925–929

Funk WC, McKay JK, Hohenlohe PA, Allendorf FW (2012) Harnessing genomics for delineating conservation units. Trends Ecol Evol 27:489–496

Ganders FR (1979) The biology of heterostyly. NZ J Bot 17:607–635

Gitzendanner MA, Soltis PS (2000) Patterns of genetic variation in rare and widespread plant congeners. Am J Bot 87:783–792

Goudet J (2005) Hierfstat, a package for R to compute and test hierarchical F-statistics. Mol Ecol Resour 5:184–186

Herbert J (2006) Future climate change scenarios for Big Bend National Park, Texas. Southwest Geogr 10:36–51

Hernandez HM, Barcenas RT (1995) Endangered cacti in the Chihuahuan Desert: I. Distribution patterns. Conserv Biol 9:1176–1188

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23:1801–1806

Jeffries DL, Copp GH, Lawson Handley L, Olsén KH, Sayer CD, Hänfling B (2016) Comparing RADseq and microsatellites to infer complex phylogeographic patterns, an empirical perspective in the Crucian carp, *Carassius carassius*, L. Mol Ecol 25:2997–3018

Jensen JL, Bohonak AJ, Kelley ST (2005) Isolation by distance, web service. BMC Genet 6:13

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

Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet 11:94

Kerley GI, Whitford WG (2000) Impact of grazing and desertification in the Chihuahuan Desert: plant communities, granivores and granivory. Am Midl Nat 144:78–91

Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I (2015) Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. Mol Ecol Resour 15:1179–1191

Lee J-H, Kim WT (2011) Regulation of abiotic stress signal transduction by E3 ubiquitin ligases in Arabidopsis. Mol Cells 31:201–208

Liu J-M, Zhao J-Y, Lu P-P, Chen M, Guo C-H, Xu Z-S, Ma Y-Z (2016) The E-subgroup pentatricopeptide repeat protein family in Arabidopsis thaliana and confirmation of the responsiveness PPR96 to abiotic stresses. Front Plant Sci 7:1825

Maddox JC, Carlquist S (1985) Wind dispersal in Californian desert plants: experimental studies and conceptual considerations. Aliso 11:77–96

Meeus S, Honnay O, Jacquemyn H (2012) Strong differences in genetic structure across disjunct, edge, and core populations of the distylous forest herb *Pulmonaria officinalis* (Boraginaceae). Am J Bot 99:1809–1818

Meirmans PG, Van Tienderen PH (2004) GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. Mol Ecol Resour 4:792–794

Meyer SE (1986) The ecology of gypsophile endemism in the eastern Mojave Desert. Ecology 67:1303–1313

Meyer S, Garcia-Moya E (1986) Gypsum grasslands of northern San Luis Potosi, Mexico: patterns of vegetation, soil, and seasonal moisture availability. Occasional Papers of the Chihuahuan Desert Research Institute, Alpine

Michaud G, Monger H, Anderson D (2013) Geomorphic-vegetation relationships using a geopedological classification system, northern Chihuahuan Desert, USA. J Arid Environ 90:45–54

Moore MJ, Jansen RK (2007) Origins and biogeography of gypsophily in the Chihuahuan Desert plant group Tiquilia subg. Eddya (Boraginaceae). Systematic Botany 32:392–414

Moore MJ, Mota JF, Douglas NA, Olvera HF, Ochoterena H (2014) The ecology, assembly and evolution of gypsophile floras. In: Rajakaruna N, Boyd RS, Harris T (eds) Plant ecology and evolution in harsh environments. Nova Science Publishers, New York, pp 97–128

Muldavin EH (2002) Some floristic characteristics of the northern Chihuahuan Desert: a search for its northern boundary. Taxon 51:453–462

Nybom H (2004) Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. Mol Ecol 13:1143–1155

Ott A, Liu S, Schnable JC, Yeh C-TE, Wang K-S, Schnable PS (2017) tGBS® genotyping-by-sequencing enables reliable genotyping of heterozygous loci. Nucleic Acids Res 45:e178–e178

Pavlik BM, Nickrent DL, Howald AM (1993) The recovery of an endangered plant. I. Creating a new population of *Amsinckia grandiflora*. Conserv Biol 7:510–526

Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol Ecol Resour 6:288–295

Pina-Martins F, Silva DN, Fino J, Paulo OS (2017) Structure_threader: an improved method for automation and parallelization of programs structure, fastStructure and MavericK on multicore CPU systems. Mol Ecol Resour 17:e268–e274

Piry S, Luikart G, Cornuet J (1999) BOTTLENECK: a computer program for detecting recent reductions in the effective population size using allele frequency data. J Hered 90:502–503

Powell AM, Turner B (1977) Aspects of the plant biology of the gypsum outcrops of the Chihuahuan Desert. In: Transactions of a symposium on biological resources of the Chihuahuan desert US Department of the Interior, National Park Service, Transactions Proceedings Series, pp 315–325

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959

R Core Team (2013) R: a language and environment for statistical computing. R Core Team, Vienna

Raj A, Stephens M, Pritchard JK (2014) fastSTRUCTURE: variational inference of population structure in large SNP data sets. Genetics 197:573–589

Ratnieks FL, Shackleton K (2015) Does the waggle dance help honey bees to forage at greater distances than expected for their body size? Front Ecol Evol 3:31

Raymond M (1995) GENEPOP: population genetics software for exact tests and ecumenism. Vers. 1.2. J Hered 86:248–249

Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. Mol Ecol Resour 4:137–138

Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, Leamon JH, Johnson K, Milgrew MJ, Edwards M (2011) An integrated semiconductor device enabling non-optical genome sequencing. Nature 475:348

Rousset F (2008) genepop'007: a complete re-implementation of the genepop software for Windows and Linux. Mol Ecol Resour 8:103–106

Service USFaW (1993) Draft Terlingua Creek Cat's-eye (Cryptantha crassipes) Recovery plan. U.S. Fish and Wildlife Service, Austin

Shao J-W, Chen W-L, Peng Y-Q, Zhu G-P, Zhang X-P (2009) Genetic diversity within and among populations of the endangered and endemic species Primula merrilliana in China. Biochem Syst Ecol 37:699–706

Shibayama Y, Kadono Y (2007) Reproductive success and genetic structure of populations of the heterostylous aquatic plant *Nymphoides indica* (L.) Kuntze (Menyanthaceae). Aquat Bot 86:1–8

Stephens PA, Sutherland WJ (1999) Consequences of the Allee effect for behaviour, ecology and conservation. Trends Ecol Evol 14:401–405

Van Rossum F, De Sousa SC, Triest L (2006) Morph-specific differences in reproductive success in the distylous Primula veris in a context of habitat fragmentation. Acta Oecol 30:426–433

Verity R, Nichols RA (2016) Estimating the number of subpopulations (K) in structured populations. Genetics 115:180992

Villarreal-Quintanilla JA, Bartolomé-Hernández JA, Estrada-Castillón E, Ramírez-Rodríguez H, Martínez-Amador SJ (2017) El elemento endémico de la flora vascular del Desierto Chihuahuense. Acta Bot Mex 18:65–96

Wang L, Ware D (2014) CloudSTRUCTURE: infer population STRU CTURE on the cloud. arXiv preprint arXiv:14084081

Warnock BJ (2012) Final report—population biology, habitat description and delineation and conservation of Terlingua Creek Cat's-eye (*Cryptantha crassipes*). Texas Parks and Wildlife, Texas, p 29

Williams MT, Warnock BJ, Betz JM, Beck JJ, Gardner DR, Lee ST, Molyneux RJ, Colegate SM (2011) Detection of high levels of pyrrolizidine-N-oxides in the endangered plant *Cryptantha crassipes* (Terlingua Creek Cat's-eye) using HPLC-ESI-MS. Phytochem Anal 22:532–540

Wu TD, Nacu S (2010) Fast and SNP-tolerant detection of complex variants and splicing in short reads. Bioinformatics 26:873–881

Yuan N, Sun Y, Comes HP, Fu CX, Qiu YX (2014) Understanding population structure and historical demography in a conservation context: population genetics of the endangered kirengeshoma palmata (hydrangeaceae). Am J Bot 101:521–529

Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS (2012) A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics 28:3326–3328