Final Report

Genetic diversity within *Sclerocactus cloverae* Heil & Porter based on ddRAD-seq: the genetic basis for subspecies recognition.

A study of *Sclerocactus cloverae* subsp. *brackii* Heil & Porter, R-E-D Code: 2-1-3, State of New Mexico Endangered, BLM Sensitive, USFWS Species of Concern; New Mexico Heritage Program S1; NatureServe Global Status: G3T2, imperiled, National Status: N2.

Principal investigator: J. Mark Porter, Researcher, Rancho Santa Ana Botanic Garden, 1500 N. College Ave., Claremont, CA 91711; (909) 625-8767; jmporter@rsabg.org, j.mark.porter@cgu.edu; Associate Professor of Botany, Botany Program, Claremont Graduate University.

Co-principal investigator: Arnold Clifford, Carrizo Mountain Herbarium, Beclabito, New Mexico 87420

Laboratory contributors: Nicolás Medina, Ph.D. Candidate, Claremont Graduate University, Rancho Santa Ana Botanic Garden, 1500 N. College Ave., Claremont, CA 91711; nmedina@rsabg.org.

Sandra Namoff, Ph.D. Candidate, Claremont Graduate University, Rancho Santa Ana Botanic Garden, 1500 N. College Ave., Claremont, CA 91711; snamoff@gmail.com.

Nick Jensen, Ph.D. Candidate, Claremont Graduate University, Rancho Santa Ana Botanic Garden, 1500 N. College Ave., Claremont, CA 91711; njensen@rsabg.org.

Dylan Cohen, Ph.D. Candidate, Claremont Graduate University, Rancho Santa Ana Botanic Garden, 1500 N. College Ave., Claremont, CA 91711; dcohen@rsabg.org.

Field contributors: Daniella Roth, Endangered Plant Program Coordinator, EMNRD-Forestry Division; Daniela.Roth@state.nm.us

Zoe Davidson, State Botanist, Bureau of Land Management, Santa Fe, NM; zdavidson@blm.gov

Robert Sivinski, RCS Southwest, Santa Fe, NM; bsivinski@cybermesa.com

John Kendall, Bureau of Land Management, Farmington Field Office, j01kenda@blm.gov

Institutional Contact: Dr. Lucinda McDade, Director, Rancho Santa Ana Botanic Garden, 1500 N. College Ave., Claremont, CA 91711; (909) 625-8767; lmcdade@rsabg.org

Executive Summary:

Sequence-based genotyping, or double-digest restriction site associated DNA sequencing (ddRAD-seq), a genomic scale high-throughput DNA sequencing technology has been used to investigate genetic diversity within Brack's cactus (*Sclerocactus cloverae* subsp. *brackii*) and between it and its close relative Clover's cactus (*S. cloverae* subsp. *cloverae*). We generated two genetic data sets, using different DNA assembly pipelines, and describe two data sets produced using optimal parameters in *STACKS* and *ipyrad*, composed of 61 and 1,944 loci, respectively. Both data sets reveal that *S. cloverae* is a genetically structured species, but that structure is inconsistent with the two currently recognized taxonomic subspecies, *brackii* and *cloverae*. We show that if *S. cloverae* is assumed to represent two genetic groups, the populations from which the types of both *S. cloverae* subsp. *brackii* and *S. cloverae* subsp. *cloverae* were taken, represent the same genetic group.

Even so, there exists significant structure and genetic differentiation within *S. cloverae*. In particular, the southern populations represent a genetically differentiated group, which may require management as a distinctive unit. This genetic distinction is evident in both data sets produced by ddRAD sequences and all analyses.

There is sufficient evidence that *S. cloverae* is a genetic lineage apart from both *S. parviflorus* and *S. whipplei*, composed of semi-isolated populations which undergo episodic (and sometimes high rates of) gene flow, and therefore, we consider it appropriate to treat it at the rank of species. There is, however, some evidence of hybrid introgression and additivity in populations north of Kirtland, at the Hogback and southwest of San Ysidro, New Mexico. This introgression appears to involve genetically isolated elements of *S. parviflorus*; however, broader sampling in *S. parviflorus* is needed to clarify these relationships.

Important findings:

1. Brack's cactus is not genetically distinct from Clover's cactus, in spite of the morphological differences and should not be afforded taxonomic status.

2. Clover's cactus (both former subspecies combined) displays a high degree of genetic differentiation across its range and populations in the southern portion of its range should be considered a distinct genetic element.

3. The Southern Genetic Element may require individual management, separate from management of the northern populations of Clover's cactus.

4. Clover's cactus (both former subspecies combined) is a species, autonomous from *Sclerocactus whipplei* and *S. parviflorus*. There is evidence of hybridization between *S. cloverae* and *S. parviflorus*, but no evidence of hybridization with *S. whipplei*.

Objective: The purpose of this study is to characterize genetic variation within *Sclerocactus cloverae*, in order to determine if there is genetic divergence and population structure supporting or refuting the presence of two subspecies, corresponding to *S. cloverae* subsp. *cloverae* and *S. cloverae* subsp. *brackii*. Further, we have included populations of *S. parviflorus* from New Mexico and Utah, as well as those of *S. whipplei* from Arizona. At one time, *S. cloverae* was considered to be conspecific with these taxa. This study will evaluate the degree of divergence between *S. cloverae* and two species closely related species.

Background: Accurate species and subspecies delineations and distributions are critical for effective communication concerning biological organisms, for sound research and key to successful management and recovery efforts for threatened and endangered species. Knowing whether different organisms represent different species, whether they represent distinct subspecies, or whether they are part of an unstructured, common (but morphologically variable) gene pool is essential. Unfortunately, in some genera, species boundaries and subspecies delineation are complex and debated by different taxonomists.

Sclerocactus Britt. & Rose (Cactaceae) represents such a group. Because of the instability of nomenclature and variability of perceived species boundaries, it remains significant management challenge for federal agencies and a continuing frustration to biologists (Heil & Porter 1994; 2003). The genus is composed of cylindroid cacti with apical flowers and formidable hooked spines in about half of the species. Much of the continuing problem with the genus centers on our poor understanding of species boundaries and genetic differentiation within species, leading to changes in taxonomies every time new data are presented. This has resulted in a bewildering nomenclature, from the conflicting interpretations by the various taxonomists who have worked with this group over the years, and a general difficulty in recognizing species and subspecies.

Of particular interest here is *Sclerocactus cloverae* Heil & Porter (Fig. 1), which incorporates two subspecies: subsp. *cloverae*, the larger, slow-developing phase, associated with pinyon-juniper woodlands; and subsp. *brackii* Heil & Porter, the smaller

Figure 1. General form and morphology of the two subspecies of *Sclerocactus cloverae*. *Sclerocactus cloverae* subsp. *brackii* (A.) is diminutuve and often lacking hooked central spines, here growing at the type locality, near Bloomfield, New Mexico. Subspecies *cloverae* (B.) is often large and cylindrical, with robust hooked central spines, here growing near Navajo Dam, San Juan Co., New Mexico. Photos by Robert Sivinski.

paedomorphic phase associated with Nacimiento Formation badlands (Heil & Porter 1994). When first discovered, what is now referred to as *S. cloverae* was thought to represent two varieties of the yellow-flowered *S. whipplei* (Engelm. & J.M. Bigelow) Britton & Rose, from Arizona. The two varieties were named *S. whipplei* var. *heilii*

Castetter, P. Pierce & K. H. Schwerin and *S. whipplei* var. *reevesii* Castetter, P. Pierce & K.H. Schwerin (Castetter et al. 1976). They differed primarily in spine traits: var. *heilii* has more spines (16-17) per areole than var. *reevesii* (13–15), but var. *reevesii* has the spines more densely packed because of the smaller tubercles. Both of these entities are currently treated as *S. cloverae* subsp. *cloverae*. Heil and Porter (1994) described *S. cloverae* subsp. *brackii*, considering it different and distinct from any of the previously named varieties. They suggested that this subspecies was a unique element within *S. cloverae*, reaching first flowering at a much younger age and while the stem was very small, relative to the typical subspecies. Morphologically, *S. cloverae* subsp. *cloverae* differs in having a larger number of central spines (6–9), one of them always hooked, whereas, *S. cloverae* subsp. *brackii* has fewer central spines (4–5) and at least some of the areoles lack hooked spines. Moreover, *S. cloverae* subsp. *brackii* appeared to be restricted to badlands of the Nacimiento Formation, while *S. cloverae* subsp. *cloverae* occurred in pinyon-juniper woodlands of the San Jose and Kirtland/Fruitland Formations.

Little was known about the natural distributions of *Sclerocactus* species in northwestern New Mexico at the time of the description of *S. cloverae* subsp. *brackii*. Indeed, in total only eight collections are cited for the two subspecies, combined. Since the original description, the range of *S. cloverae* subsps. *brackii* and *cloverae* have become better characterized (Fig. 2). This has, in major part, been a consequence of increased field survey efforts related to activity in oil exploration in the region. The current knowledge concerning the distribution of *S. cloverae*, including subspp. *brackii* and *cloverae*, suggests that it is restricted to a small region of northwestern New Mexico, largely confined to the Nacimiento Fm. or closely proximate to it. At the global scale this represent a highly endemic species.

There has been some degree of controversy over how *Sclerocactus cloverae* should be treated in classification. As previously noted, it was originally treated as part of *S. whipplei* (Castetter et al. 1976). More recently, Hochstätter has similarly treated *S. cloverae* as part of *S. whipplei* at the rank of subvariety (Hochstätter 1997). However, *S. cloverae* has also been included under *S. parviflorus* Clover & Jotter var. *intermedius*

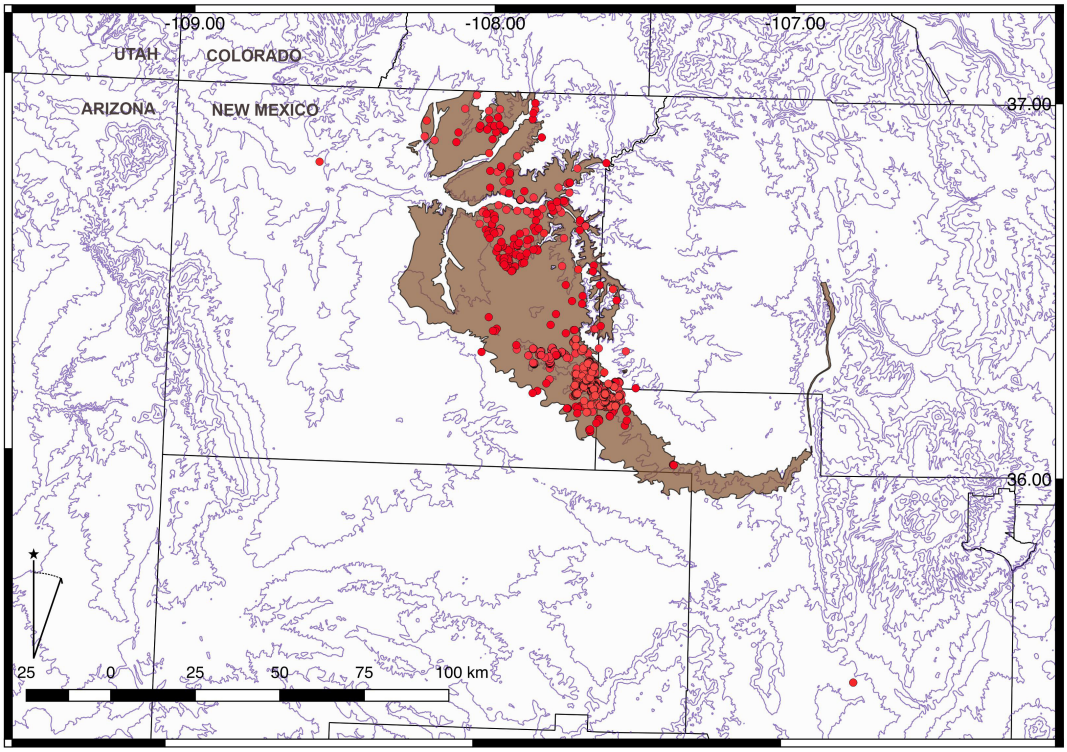


Figure 2. Documented distribution of *Sclerocactus cloverae* (red dots) and its relationship to the Nacimiento Formation in northwestern New Mexico.

(Peebles) Woodruff & L. Benson by Benson (1982; see also CITES 2015). In several recent revisions and floras, *S. cloverae* has been treated at the species rank (Heil & Porter 1994; 2003; Porter & Prince 2011; Heil et al. 2013). This situation is not only confusing to the users of the taxonomies but have very different and conflicting implications with respect to land management and conservation.

Here we address the genetic distinctness of *Sclerocactus cloverae* subsp. *brackii*, with respect to subsp. *cloverae*. It is critical to recognize that the patterns of variation below the rank of species are very different than the expected patterns between species. The species-status of a group of organisms, as here viewed, is based upon reproductive isolation between other groups (Mayr 1942; 1982; Dobzhansky 1959; 1970), and as a consequence, this results in evolutionarily independent lineages (sensu de Queiroz 1998; 1999; 2005; 2007). As lineages remain reproductively isolated from one another (but

connected by gene flow within each respective lineage), genetic divergence occurs. Hey and Pinho (2012; see also Shaffer and Thomson 2007) have shown that under an isolation-with migration model of populations, population genetic parameters (e.g., F_{ST} and time to common ancestry) can discriminate interspecific relationships from intraspecific relationships. These parameters reflect on the degree and permeability of reproductive isolation and provide a basis for objectively accepting or rejecting hypotheses of species-status. These parameters can be estimated using microsatellite markers (e.g., Hey and Pinho 2012) or restriction site associated single nucleotide polymorphism (SNPs) (Davey et al. 2011; Catchen et al. 2013).

Below the species level, intraspecific relationships can be more complicated. Because subspecies are subdivisions within a single species, by definition they experience some degree of regular gene flow. However, generally it is accepted that subspecies represent genetically differentiable and geographically cohesive elements within a species. Recent studies have shown that both microsatellite markers and SNP surveys based on restriction site associated DNA sequencing (RAD-seq) have the capacity to support the recognition of subspecies (e.g., Benestan et al. 2015; Meredith et al. 2007; Mullen et al. 2009; Shaffer and Thomson 2007) or reject it (e.g., Mudrik et al. 2015).

To characterize genetic variation within *Sclerocactus cloverae*, determine if there is genetic divergence and population structure supporting or refuting the presence of two subspecies, corresponding to *S. cloverae* subsp. *cloverae* and *S. cloverae* subsp. *brackii*, and to characterize divergence between *S. cloverae* and two species closely related species, We have performed genomic sequencing to develop a SNP assay of *Sclerocactus cloverae*, including subspecies *cloverae* and *brackii*, using the double digest restriction site associated DNA sequencing (ddRAD-seq) protocol. This method has generated several thousand genetic markers from throughout the genome. We have used these markers to investigate the genetic diversity within *Sclerocactus cloverae*, in order to determine if genetic diversity is differentiated into two primary groups, corresponding to subspp. *cloverae* and *brackii*. Alternatively, genetic differentiation may be related to the

geographic distance between populations, but uncorrelated with the morphological traits (i.e., isolation by distance).

Previous Research—Investigations into species and subspecies relationships of *Sclerocactus*, utilizing evidence beyond comparative morphology, did not begin appearing in the literature until 2000. Early studies focused largely on species relationships, inferred from chloroplast DNA sequences. Even so, these studies revealed a close relationship among *Sclerocactus cloverae*, *S. parviflorus* and *S. whipplei* (Porter et al. 2000; Baker & Porter 2016). Population genetic studies employing amplified fragment length polymorphism (AFLP; Porter et al. 2012) and microsatellite loci (Schwabe et al. 2013; 2015) have demonstrated that *Sclerocactus* in the Uintah Basin of Utah and those near Mesa, Colorado, both formerly treated as *S. glaucus* (K. Schum.) L.D. Benson, represent two genetically isolated species. Moreover, within both of these groups there is significant genetic structure correlating to geography, consistent with the existence of subspecies. These genetic data have decisively resolved long-standing problems associated with species and subspecies boundaries, as well as hybridization in *S. glaucus*.

Methods

Sampling – Members of *Sclerocactus cloverae* were sampled in San Juan, Sandoval and Rio Arriba Counties, New Mexico (Table 1; Supplemental Figure 1) by Robert Sivinski, Zoe Davidson, Daniella Roth, Arnold Clifford, John Kendall and J. Mark Porter. Samples of *Sclerocactus* were also collected in Arizona, Colorado, Nevada and Utah, by Arnold Clifford and J. Mark Porter, representing *S. cloverae* and closely related species. We have sampled 13 populations of subsp. *cloverae* and 9 populations of subsp. *brackii*. At each population we sampled floral materials of 12–15 individuals. Approximately 0.5 g of either floral or fruit pericarpal material will be collected and placed into silica gel for desiccation. Because *S. cloverae* has been suggested to be conspecific with both *S. whipplei* and *S. parviflorus*, we have included a minimal sampling of these species. In total, with 294 individuals were sampled, representing 41 populations (Table 1).

Table 1. Collections sampled for ddRAD-seq analysis of *Sclerocactus cloverae*. Provided are collection (voucher) numbers for specimens housed at Rancho Santa Ana Botanic Garden Herbarium (RSA), the species or subspecies represented (Taxon), sample size (N), latitude and longitude, in decimal degrees (dd). Four sampled populations are of questionable identity (?) and likely represent hybrid introgression.

Collection number	Taxon	N	Latitude (dd)	Longitude (dd)
15541	*S. cloverae* subsp. *brackii*	11	36.26091	-107.60397
15542	*S. cloverae* subsp. *brackii*	12	36.24627	-107.6775
15543	*S. cloverae* subsp. *brackii*	12	36.6755	-107.9936
15544	*S. cloverae* subsp. *brackii*	12	36.66246	-107.98161
15545	*S. cloverae* subsp. *brackii*	12	36.22981	-107.65862
15546	*S. cloverae* subsp. *brackii*	12	36.18053	-107.58088
15549	*S. cloverae* subsp. *brackii*	11	36.60201	-107.9549
15550	*S. cloverae* subsp. *brackii*	12	36.8002	-107.96571
15565	*S. cloverae* subsp. *brackii*	11	36.78273	-107.93587
15540	*S. cloverae* subsp. *cloverae*?	12	35.49054	-106.86588
15547	*S. cloverae* subsp. *cloverae*	11	36.22504	-107.70009
15548	*S. cloverae* subsp. *cloverae*	12	36.30623	-108.01028
15552	*S. cloverae* subsp. *cloverae*	11	36.79575	-108.22879
15554	*S. cloverae* subsp. *cloverae*	12	36.88889	-108.11309
15558	*S. cloverae* subsp. *cloverae*	12	36.89978	-108.04034
15562	*S. cloverae* subsp. *cloverae*	12	36.94509	-107.86337
15563	*S. cloverae* subsp. *cloverae*	12	36.96433	-107.8503
15564	*S. cloverae* subsp. *cloverae*	12	36.89849	-107.95638
15567	*S. cloverae* subsp. *cloverae*	12	36.68642	-107.78829
15568	*S. cloverae* subsp. *cloverae*	12	36.63826	-107.69786
15569	*S. cloverae* subsp. *cloverae*	12	36.81831	-107.61647
NKrt	*S. cloverae* subsp. *cloverae*?	7	36.77452	-108.34609
15511	*S. johnsonii*	1	36.08715	-114.07664
15585	*S. nyensis*	3	37.641364	-117.50128
15590	*S. nyensis*	2	38.026345	-117.233187
15596	*S. nyensis*	3	38.202297	-116.185646
15521	*S. parviflorus* subsp. *intermedius*	1	38.640545	-110.667806
HB	*S. parviflorus* subsp. *intermedius*?	3	36.8237	-108.48263
YJ	*S. parviflorus* subsp. *intermedius*?	6	37.33252	-109.03328
15530	*S. parviflorus* subsp. *parviflorus*	1	37.919286	-110.387569
15535	*S. parviflorus* subsp. *terre-canyonae*	1	36.89009	-111.60133
15539	*S. polyancistrus*	1	34.730457	-117.321284
15574	*S. polyancistrus*	1	37.369192	-118.023333

Collection number	Taxon	N	Latitude (dd)	Longitude (dd)
15601	*S. spinosior* subsp. *blainei*	3	37.7807	-114.46717
15520	*S. spinosior* subsp. *spinosior*	3	38.63219	-112.17932
SL	*S. whipplei*	3	37.11421	-109.51098
15522	*S. wrightiae*	1	38.596288	-110.699541
15523	*S. wrightiae*	1	38.388989	-110.893287
15527	*S. wrightiae*	3	38.26377	-110.79127

DNA extraction and sequencing – Upon return to the laboratory the silica gel was replaced twice, at 2–4-day intervals (as determined by color change in indicator), insuring that the tissue samples dried thoroughly.

Graduate student Nicolas Medinas, Nick Jensen, Sandy Namoff and Dylan Cohen extracted DNA from the samples employing a modification of the Doyle and Doyle (1987; see also Griffith and Porter 2006) CTAB-protocol. Our procedure differed by the addition of steps for the enzymatic digestion of RNAs, using RNAse A; proteins, using proteinase K; and mucilage/polysaccharides, using pectinase (17389 Sigma-Aldrich) prior to DNA precipitation. DNA concentrations were standardized to 20ng/µl and at least 1µg of total DNA, using the Qubit 2.0 Fluorometer (Invitrogen, Inc.). Samples were prepared and shipped to Floragenex (4725 Village Plaza Loop, Suite 200, Eugene, OR 97401). Reduced representation genomic DNA sequencing libraries were prepared following Truong et al. (2012), using the sequence-based genotyping (SBG) method, a double-digest restriction site associated sequencing (ddRAD-seq) method. This procedure employed two restriction enzymes, MseI and PstI. Libraries were single-end sequenced, 100 bases in length, on the Illumina HiSeq platform.

ddRAD-seq SNP analyses –Bioinformatics analyses utilized the *STACKS* (Catchen et al. 2011; 2013) pipeline to assemble RAD-tags from populations, as well as call SNPs, genotypes, and haplotypes of these individuals. We have employed computational resources provided by Cyverse (www.cyverse.org; Goff et al. 2011; Merchant et al. 2016) for these assemblies. Pipeline execution followed the protocols outlined by Rochette and Catchen (2017), including procedures for parameter selection. The samples were de-multiplexed, sorted and binned into individual samples using *process_radtags*. Poor

quality reads were removed, adapters sequences and restriction sites were trimmed from the sequences. Samples with a low number of reads were excluded from subsequent analyses. We applied both *de novo* locus assembly, employing *ustacks*, and referenced assembly, employing *pstacks* and using the saguaro (*Carnegiea gigantea* [Engelm.] Britton & Rose) genome as a reference (Copetti et al. 2017). Pipeline programs *cstacks*, *sstacks* and *populations* were used for both *de novo* and referenced assemblies. Multiple runs were performed varying parameters that control maximum distance (in nucleotides) permitted between stacks (*ustacks* -M), number of mismatches allowed between sample loci (*cstacks* -n), minimum depth of coverage required to create a stack (*ustacks* -m and *pstacks* -m), minimum number of populations a locus must be present in to process a locus (*populations* -p), and minimum percentage of individuals in a population required to process a locus for that population (*populations* -r). The referenced assemblies used *bowtie2* vers. 2.2.9 (Langmead and Salzber 2012) to assemble the demultiplexed, and cleaned sequences against the saguaro genome scaffold. In addition, we have employed *ipyrad* (Eaton 2014; http://ipyrad.readthedocs.io) to conduct an alternative *de novo* assembly of RAD-tags. As was the case for *STACKS*, multiple runs were performed in which parameter settings were varied to determine the sensitivity of results to these differences.

Standard population parameters, including observed and expected heterozygosity (H_O and H_E, respectively) were calculated by *STACKS* and *ipyrad*. Genetic distance measures and Analysis of Molecular Variance (AMOVA) were estimated using GenAlEx (Peakall and Smouse 2012; 2006). We used *STRUCTURE* vers. 2.3.4 (Pritchard et al. 2000; Falush et al. 2003; Hubisz et al. 2009) and *Lea* (Frichot and Francois 2015) to implement a Bayesian *K*-clustering algorithm for detecting population structure and to assign individuals to populations, based upon their multi-locus genotypes. In these analyses, we have assumed that all individuals are diploids and two alleles have been scored for each individual. The assumed number of populations used in analyses ranged from K=1–10 in *structure*, and the set of 10 analyses were replicated five times using *ParallelStructure* (Besnier and Glover 2013) on the CIPRES Science Gateway (Miller et al. 2011) and summarized using the web version of *Structure Harvester* (Earl and von Holdt 2012). The most likely *K*-score was evaluated using $\ln(K)$, $\Pr(X|K)$ and the ΔK

method (Evanno et al. 2003) and consideration of F_{ST} values. Although the ΔK method is the most commonly used method currently, it has been shown to over-represent K= 2 in both empirical and simulation studies (Janes et al. 2017). The assumed number of populations used in analyses ranged from K=1–30 in *Lea* (Frichot and Francois 2015), under the R framework (R Core Team 2013), with 10 replicates for each value of K. K-values were contrasted using the cross-entropy criterion, which evaluates the quality of fit of the model to the genotypic data using cross-validation (Alexander and Lange 2011; Frichot et al. 2014).

A direct comparison of DNA sequences using maximum likelihood estimation of phylogenetic relationships, using RaxML (Stamatakis 2014), and Bayesian phylogenetic posterior probability inference, using MrBayes (Huelsenbeck and Ronquist 2001; Ronquist and Huelsenbeck 2003), was performed at the CIPRES Science Gateway (Miller et al. 2011). The concatenated matrix of 51 populations by all sampled loci was employed in both analyses. Alternatively, these data were phylogenetic relationships were estimated under a coalescent model using the program SVDquartets v.1.0 (Chifman and Kubatko 2014). These data were further analyzed using a neighbor-net analysis (Bryant and Moulton 2003) of uncorrected pairwise nucleotide differences, using SplitsTree4 (Huson and Bryant 2006). This is a network analysis that graphically represents conflicting information in the data.

Results

The DNA extraction procedures were successful in providing high molecular weight, mucilage-free DNAs, ideal for library preparation. Single-end, Illumina sequencing of 100 bases per read produced 906,683,708 reads (mean per sample= 3,073,504.09), across three lanes, each multiplexed with 95 samples and a control. Demultiplexing and sequence cleaning resulted in a set of individuals with a mean sampling of 10,944.8 loci (standard deviation= 2,358.58). However, locus sampling ranged from a minimum of 429 loci (sample HB3) to a maximum of 23,236 (sample 15544_G). Lower coverage samples did not show any bias toward specific populations or taxa (see Supplemental Figure 2).

Estimation of the most appropriate parameters to be employed in the *STACKS* pipeline revealed that variation in the -*M* and -*n* parameters had little influence on sample coverage (Table 2). While there is no absolute criterion for selecting parameters (Rochette and Catchen 2017), we note that there is a stabilization in the relationship between the number of genotyped loci and the number of loci present in at least 80% (r80) of the samples when *M*=*n*= 6 (Supplemental Fig. 3). Therefore, we have employed the *M*=*n*= 6 analyses of the full data set, using *STACKS* in this report. This assembly produced 3770 loci, of which 61 were found in 80% or more of the individuals. The low number of loci (i.e., 3770) appears to be a consequence of filtering a large number of loci with poor reads at the restriction sites (compare with *ipyrad* assembly below). Mean observed heterozygosity (H_{obs}) across all populations (at variable sites) is 0.0573 (mean H_{exp}= 0.0603) and across all sites, including invariant sites is 0.00278 (mean H_{exp}= 0.0029). A summary of diversity statistics at the population level, based on Shannon-Wiener Index of multilocus genotypic diversity (Shannon 2001), Stoddart and Taylor Index of multilocus genotypic diversity (Stoddart and Taylor 1988), Simpson's λ Index (Simpson 1949) and Nei's unbiased gene diversity (Nei 1978), as calculated using GenAlEx and *popprR* 2.7.1 (Kamvar et al. 2014, under the R framework: R Core Team 2017), are provided in Table 2. In general, populations of *Sclerocactus cloverae* contain similar amounts of genetic variation. Populations HB, NKrt and YJ have less diversity. In addition, linkage disequilibrium statistics, E_5 evenness index (Pielou 1975; Ludwig and Reynolds 1988; Grünwald et al. 2003), index of association (Brown, Feldman and Nevo, 1980; Manard Smith et al. 1993) and the standardized index of association ($\bar{r}_d$, Agapow and Burt 2001) are provided in Table 3. These statistics show that the ten populations of *S. cloverae* display similar levels of multilocus genetic diversity and similar degrees of heterozygosity. In contrast, there are some differences in linkage disequilibrium, as measured by I_A and r_d. Three populations of questionable identity (HB, NKrt and YJ) display somewhat different measures of heterozygosity and linkage disequilibrium; however, the low sampling at these populations may be producing a bias. Analysis of molecular variance (AMOVA) indicates that there is significant population differentiation in *S. cloverae*, rejecting the null hypothesis that there is a single,

Table 2. Summary of the influence on varying parameters -*M* (maximum nucleotide differences between alleles) and -*n* (maximum nucleotide differences between loci) on assemblies using *STACKS* 2.0 β9 (Catchen et al. 2011; 2013; http://catchenlab.life.illinois.edu/stacks/). Here we contrast a subset of 12 individuals, analyzed using nine different parameter settings, using the mean per-sample coverage and standard deviation of this estimate (Per-sample coverage stdev). Compared are the number of genotyped loci (Genotyped loci) of all 12 individuals as well as those loci that are present in at least 80% of the individuals (r80 loci). We also compare the number of SNPs (r80 SNPs) and loci lacking variation (Invariant r80 loci) in those loci that are present in at least 80% of the individuals.

M=*n* value	Mean per-sample coverage	Per-sample coverage stdev	Genotyped loci	r80 loci	r80 SNPs	Invariant r80 loci
1	21.9	3.5	856	64	66	19
2	22	4.2	764	68	88	18
3	21.7	4.3	729	69	95	18
4	21.8	4.3	701	70	107	17
5	21.9	4.5	680	71	118	17
6	22	4.4	669	72	125	17
7	22	4.5	664	72	123	17
8	22.1	4.7	659	74	132	17
9	22	4.9	650	74	132	17

undifferentiated gene pool (Φ_{PT}= 0.7282; p= 0.001; F_{ST}= 0.6962; p= 0.001). AMOVA, calculated using GenAlEx, infers that 73% of the variance in the data is accounted by differentiation among populations, while 27% is accounted for by variability within populations (Table 4). Estimates of migration rate (*Nm*) among populations displays disparity depending on the specific method used for the estimation and the scope of the estimate. If a single migration rate is assumed between all populations of *S. cloverae*, then the rate is 0.109 (1 migrant every 10 generations) based on estimates using F_{ST} (Wright 1931; Slatkin and Barton 1989) and 0.585 (1 migrant every ca. 2 generations) based on estimates using private alleles (Slatkin and Barton 1989). However, a comparison between the type locality of Brack's cactus (15543) and a population northeast of Bloomfield (15565) is estimated with a rate is 0.118 (1 migrant every 10 generations) based on estimates using F_{ST} and 1.329 (1 migrant every generation) based

Table 3. Genetic diversity of populations of *Sclerocactus cloverae*, as measured by the Shannon-Wiener Index of multilocus genotypic diversity (*H*, Shannon 2001), the Stoddart and Taylor Index of multilocus genotypic diversity (*G*, Stoddart and Taylor 1988), Simpson's Index (λ, Simpson 1949) Nei's unbiased gene diversity (H_{exp}, Nei 1978). In addition, we report linkage disequilibrium statistics E_5 evenness index (E_5, Pielou 1975; Ludwig and Reynolds 1988; Grünwald et al. 2003), index of association (I_A, Brown, Feldman and Nevo, 1980; Smith et al. 1993) and the standardized index of association ($\bar{r}_d$, Agapow and Burt 2001). Population sample size (N) and estimated number of multilocus genotypes (eMLG) are described. Population numbers correspond with those described in Table 1.

Population	N	eMLG	*H*	*G*	λ	E_5	H_{exp}	I_A	$\bar{r}_d$
15540	12	10	2.48	12	0.917	1	0.0553	0.9705	0.011156
15541	12	10	2.48	12	0.917	1	0.0509	1.7011	0.022688
15542	12	10	2.48	12	0.917	1	0.0641	1.1477	0.010532
15543	12	10	2.48	12	0.917	1	0.0665	1.3591	0.014933
15544	12	10	2.48	12	0.917	1	0.0737	3.6812	0.030398
15545	12	10	2.48	12	0.917	1	0.0546	-0.0494	-0.000439
15546	12	10	2.48	12	0.917	1	0.0578	0.9387	0.009216
15547	11	10	2.4	11	0.909	1	0.0617	1.3206	0.011931
15549	13	10	2.56	13	0.923	1	0.0587	0.571	0.006891
15549	10	10	2.3	10	0.9	1	0.069	0.7544	0.007161
15550	12	10	2.48	12	0.917	1	0.0815	1.9463	0.019525
15552	11	10	2.4	11	0.909	1	0.0768	1.0502	0.010632
15554	12	10	2.48	12	0.917	1	0.0746	1.4701	0.01584
15558	12	10	2.48	12	0.917	1	0.0601	0.869	0.010979
15562	12	10	2.48	12	0.917	1	0.061	2.0057	0.0279
15563	12	10	2.48	12	0.917	1	0.0536	0.962	0.014107
15564	12	10	2.48	12	0.917	1	0.0797	8.2225	0.069827
15565	11	10	2.4	11	0.909	1	0.0911	1.177	0.009787
15567	12	10	2.48	12	0.917	1	0.0539	1.0806	0.015981
15568	12	10	2.48	12	0.917	1	0.0543	0.3562	0.003277
15569	12	10	2.48	12	0.917	1	0.0522	1.9516	0.022528
HB	3	1	0	1	0	NaN	0	NaN	NaN
NKrt	7	7	1.95	7	0.857	1	0.0943	17.2864	0.166675
YJ	6	6	1.79	6	0.833	1	0.0895	1.7676	0.01468
Total	264	10	5.56	258	0.996	0.99	0.1089	2.7969	0.007081

on estimates using private alleles. A summary of pairwise *Nm* values using estimates based on F_{ST} is provided in Supplemental Table 1.

Table 4. Analysis of molecular variance (AMOVA) of 20 populations of *Sclerocactus cloverae* and four populations of *Sclerocactus* that are close to *S. cloverae* but show additivity based on *STRUCTURE* analyses (putative hybrids), conducted using GenAlEx 6.5 (Peakall and Smouse 2006; 2012). The AMOVA contrasts 264 individuals and 804 SNP loci, generated using *STACKS* (Catchen et al. 2013), with parameters *-n*= *-M*= 6, and *-r*= 80. We report the degrees of freedom (**df**), sums of squares (**SS**), mean squares (**MS**), estimated variance (**Est. Var.**), as well as the percentage of variance explained (**%**). Lastly, we report the Φ_{PT} value and the probability (*p*) of an equal or more extreme value of Phi.

Source	df	SS	MS	Est. Var.	%
Among Populations	23	99303.308	4317.535	380.314	73%
Within Populations	240	34073.541	141.973	141.973	27%
Total	263	133376.848		522.287	100%

Φ_{PT}	0.728	*p*= 0.001

STACKS assembled loci infer a high degree of structure among populations of *Sclerocactus cloverae*, based upon analyses using *structure* (Pritchard et al. 2000) and *LEA* (Frichot and Francois 2015). The Evanno method (ΔK, Evanno et al. 2005; Earl and von Hold 2012) infers that K= 2 is the optimal clustering level (Fig. 3A, B; Fig. 4). At K= 2 the genetic groups have F_{st} values of 0.5451 and 0.3305 and a log likelihood of -127,572.5. As previously noted, the ΔK method has been shown to over-represent K= 2 in both empirical and simulation studies (Janes et al. 2017). It is evident that the log likelihood score increases to -77,276.7 and the mean value of F_{ST} increases to 0.74346 with K= 11 in the *structure* analyses (Fig. 3A & C). We suggest that optimal clustering is achieved when within-group variance is minimized, and between-group variance is maximized. This would occur when F_{ST} values are at a maximum and estimated heterozygosity (H_{exp}) is at a minimum. This occurs at K= 11 (Fig. 3C). By contrast, *LEA*, which uses cross-entropy criterion to evaluate the quality of fit of the model to the genotypic data, using cross-validation (Alexander and Lange 2011; Frichot et al. 2014), inferred a much higher K value (Fig. 3D; Fig. 5). There is some "plateauing" at K=10, but final "plateauing" begins at K= 21 and the minimum is at K= 25. This later value is not particularly helpful, as only 24 populations are included in these analyses. The structure of the genetic data is evident in the principle coordinates analysis of these data (Fig. 6). It

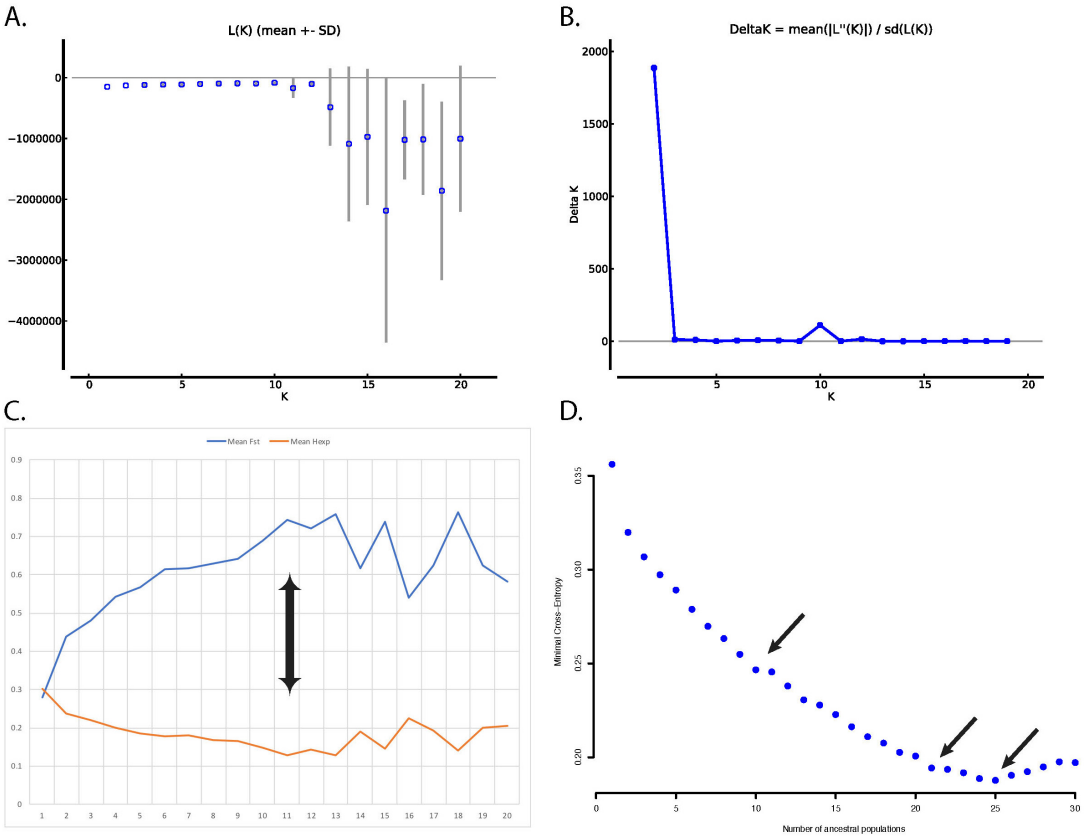


Figure 3. Comparison of *STRUCTURE* 2.3.4 analyses, using *K*-values ranging from 1 to 20 and 30, based on *STACKS* assembly of loci. The change in log likelihood with respect to *K* is shown in panel A; change in Δ*K* with respect to *K* is shown in panel B; the difference in mean F_{ST} and mean expected heterozygosity (Hexp) with respect to *K* is shown in panel C (the arrow indicates a maximum of the difference); minimum cross-entropy with respect to *K* , based on analyses using *Lea*, is shown in panel D. The three arrows indicate regions of plateauing and minimum value.

is important to note that the two axes displayed account for only 21.88% of the variance in the data. This indicates that these data contain considerably more structure than in represented in the two dimensions of Fig. 6. In this figure we also show the inconsistency between these genetic clusters and the named taxa, *S. cloverae* subsp. *brackii* (circles) and *S. cloverae* subsp. *cloverae* (triangles).

The STACKS pipeline generates a DNA sequence data file that can be analyzed using phylogenetic methods. It summarizes each population by selecting those nucleotide sites of the filtered loci that are fixed within populations but differ among the

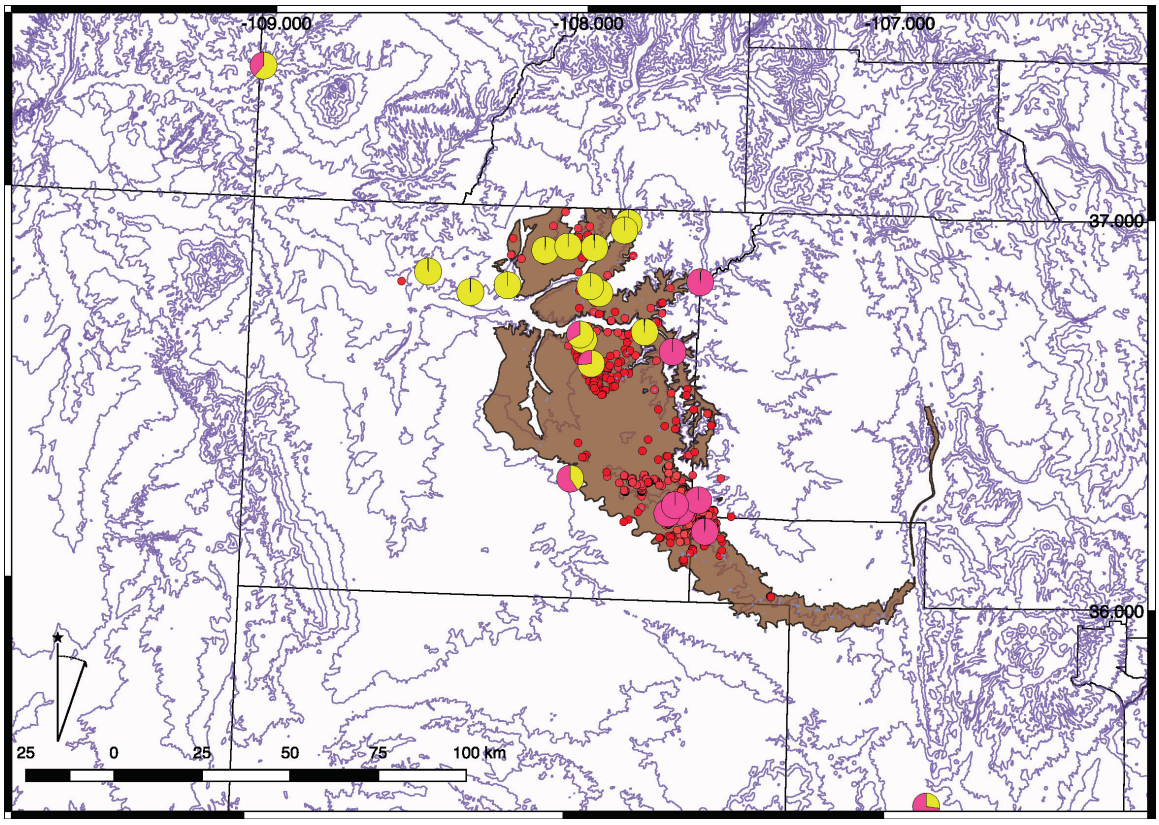


Figure 4. *STRUCTURE* 2.3.4 results at *K*= 2, for 25 populations of *Sclerocactus*. Populations are color-coded, yellow: northern genetic cluster; magenta: southern genetic cluster. Known sites of *S. cloverae* are shown as small, red dots and the Nacimiento Formation is shown in brown.

populations. For the parameters assigned to the data (i.e., loci present in at least 80% [r80] of the samples when *M*=*n*= 6), none of the loci are retained. If the *-r* parameter is relaxed to include all loci present in 50% or more of the individuals (i.e., r50), then 2,465 loci are retained, but there is considerable missing data. Maximum likelihood analysis of these 2,465 loci, using the GTR model with gamma distributed rates in RaxML, produced a population-level estimate of phylogenetic relationships (Fig. 7) that implies no distinction between *Sclerocactus cloverae* subspp. *cloverae* and *brackii*. It also infers that *S. cloverae* is not closely related to *S. whipplei*, as suggested by some previous authors (Castetter et al. 1976; Hochstätter 1997). The relationship between *S. parviflorus* and *S. cloverae* is more complicated because *S. parviflorus* is polyphyletic in this estimate. Even so, *S. cloverae* is inferred to be an independent lineage.

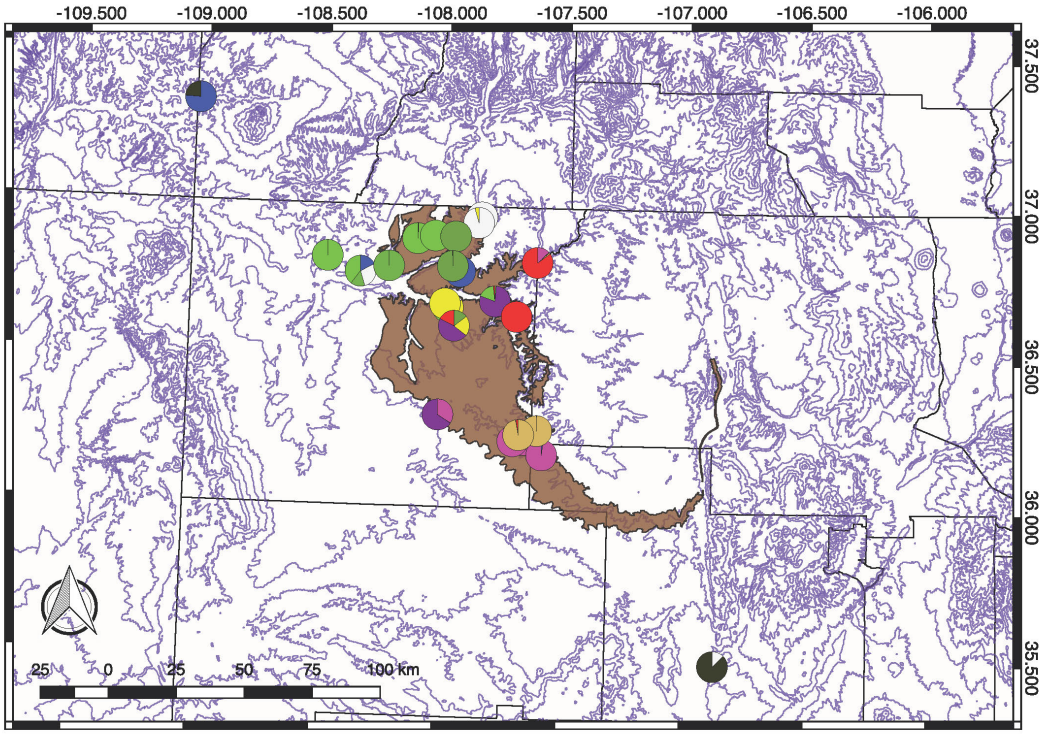


Figure 5. *STRUCTURE* 2.3.4 results at *K*= 11, for 25 populations of *Sclerocactus*. Populations are color-coded based upon group membership. The Nacimiento Formation is shown in brown.

Parameter estimation using *ipyrad* (http://ipyrad.readthedocs.io) focused largely on varying parameters 14 (clustering threshold) and 21 (minimum samples per locus). Increasing the cluster threshold increases the number of loci discovered, while increasing the minimum number of samples per locus decreases the number of loci retained in the final data set (Table 5). Here we report on results with parameters 14= 0.87 and 21= 0.85 (Table 5, IV), which recovered 1944 loci found in 85% or more of the samples, of which 134 were invariant. Mean estimated heterozygosity (H_o) across all individuals sampled is 0.0113, and across all individuals of *Sclerocactus cloverae* is 0.0115. Mean observed heterozygosity (H_{obs}) across all populations (at all sites) is 0.0078 (mean H_{exp}= 0.0114). As observed in the *STACKS* assembly, the *ipyrad* assembly shows that populations of *Sclerocactus cloverae* contain similar amounts of genetic variation (Table 6), as measured by Shannon-Wiener Index of multilocus genotypic diversity (*H*), Stoddart and Taylor Index of multilocus genotypic diversity (*G*), Simpson's λ Index and Nei's

Figure 6. Principle coordinates analysis of ddRAD genetic data for 20 populations of *Sclerocactus cloverae*, representing subspecies *brackii*, *cloverae* and three population of uncertain (unknown) affiliation, suspected represent hybrid introgression. Populations are color coded and different taxa use different shapes. Each axis parenthetically includes the percent of variance explained.

unbiased gene diversity (H_{exp}). Populations HB, NKrt, YJ and SL have less diversity, but also have smaller sample sizes. In addition, linkage disequilibrium statistics, E_5 evenness index, index of association (I_A) and the standardized index of association ($\bar{r}_d$) indicate that significant linkage is detected in populations 15542, 15546, 15548, 15550, 15552, 15554, 15563, 15565, 15567 and 15568 (Table 6). Analysis of molecular variance (AMOVA) indicates, again, that there is significant population differentiation in *S. cloverae*, calculated using *poppr* and GenAlEx (Φ_{ST} = 0.2962; p= 0.001 and Φ_{PT}= 0.1301; p= 0.001, respectively), inferring that 30% of the variance in the data is

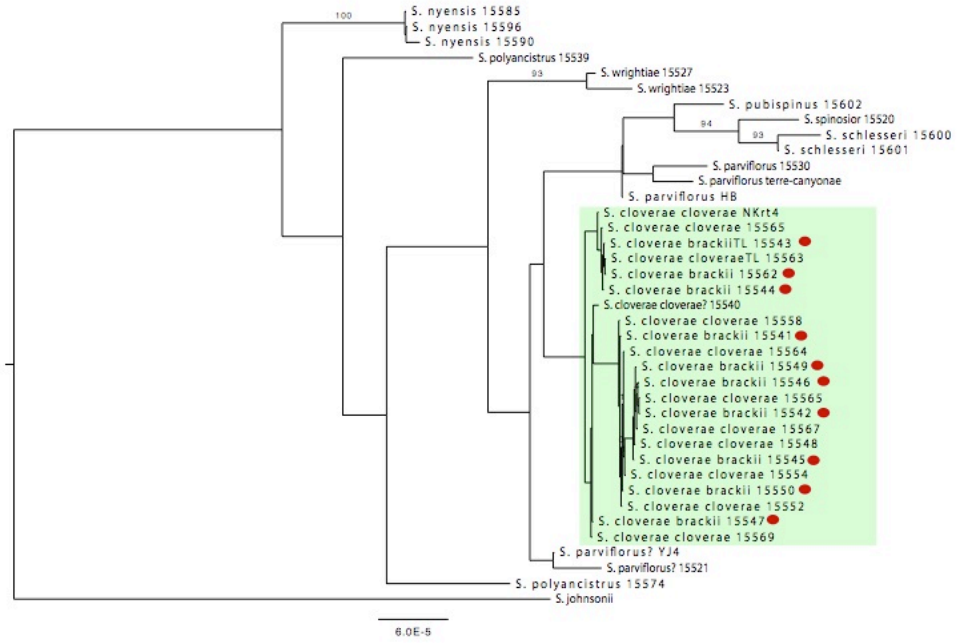


Figure 7. Maximum likelihood estimation of phylogenetic relationships among populations of *Sclerocactus cloverae* and eight other species of the genus, based on RaxML analyses of loci assembled using *STACKS*. Red ovals indicate those populations corresponding to the morphological form referred to as Brack's cactus. The green box encloses those populations identified as the species *S. cloverae*.

Table 5. Parameters applied to *ipyrad* assemblies I through VI, including maximum read error (phred Q score offset), maximum pairwise distance between alleles in a locus (clustering threshold) and minimum number of samples possessing a locus (Min # samples per locus). Characteristics of the assemblies are described, including the number of loci recovered prior to filtering (Prefiltered loci), the number of loci removed by specific filters (i.e., Duplicates, Max SNPS, Max shared H, Min sample and Max alleles) and the final number of loci retained (Filtered loci).

	I	II	III	IV	V	VI
Phred Q score offset	33	33	43	43	33	33
Phred Q score offset	33	33	43	43	33	33
Clustering threshold	0.85	0.9	0.87	0.9	0.87	0.87
Min # samples per locus	4	4	219	251	263	277
Prefiltered loci	193982	219871	200298	219932	200298	200298
Duplicates	3110	4574	3700	4637	3700	3700
Max SNPS	13607	12426	2137	1109	907	522
Max shared H	5732	9735	871	764	536	392
Min sample	69052	80879	192110	215054	200276	200276
Max alleles	29567	37939	32995	38003	32995	32995
Filtered loci	88807	99249	2911	1944	11	11

Table 6. Summary statistics describing 13 populations of *Sclerocactus cloverae* subsp. *cloverae* and 10 populations of *S. cloverae* subsp. *brackii*. Estimated sample size (N), number of private alleles, observed heterozygosity (H_O), expected heterozygosity (H_E), F_{IS} and nucleotide diversity, as measured by π, are reported.

Sample Number	N	Private alleles	H_O	H_E	F_{IS}	π
15541	10.1508	7	0.0264	0.0406	0.0399	0.043
15542	10.2205	15	0.0217	0.0359	0.0413	0.0379
15543	9.9244	23	0.0302	0.0447	0.0531	0.0474
15544	9.7511	23	0.0259	0.0453	0.0654	0.0481
15545	10.2838	15	0.0237	0.0418	0.0582	0.0443
15546	10.3037	14	0.0198	0.0379	0.0577	0.0401
15547	9.6097	17	0.0257	0.0387	0.0438	0.041
15549	9.2729	14	0.0271	0.0392	0.0365	0.0417
15550	9.9635	11	0.0236	0.0408	0.0549	0.0432
15562	9.9436	18	0.0265	0.0366	0.0355	0.0388
15540	10.3872	18	0.0161	0.0287	0.0399	0.0303
15548	10.1571	9	0.0246	0.0368	0.0405	0.039
15552	9.1016	18	0.0243	0.0431	0.057	0.046

Sample Number	N	Private alleles	H_O	H_E	F_{IS}	π
15554	9.7404	16	0.0292	0.0456	0.054	0.0484
15558	9.9874	20	0.0262	0.0431	0.0549	0.0455
15563	9.8542	11	0.0274	0.0447	0.0524	0.0473
15564	9.6194	9	0.0325	0.0451	0.046	0.0478
15565	9.4257	16	0.0303	0.0413	0.0411	0.0438
15567	10.026	19	0.0207	0.03	0.0339	0.0319
15568	9.5093	14	0.0255	0.0349	0.0312	0.037
15569	10.0866	12	0.0256	0.0416	0.0508	0.0441
NKrt1	5.4903	16	0.0242	0.0367	0.0365	0.0406
YJ1	5.035	15	0.0321	0.0411	0.0312	0.0463

accounted by differentiation among populations, while 70% is accounted for by variability within populations.

Like the assembly generated by *STACKS*, the *ipyrad* assembly displays disparity depending on the specific method of estimation of migration (*Nm*). If a single migration rate is assumed between all populations of *S. cloverae*, then the rate is 1.876 (nearly 2 migrants per generation) based on estimates using F_{ST} (Wright 1931) and 0.493 (1 migrant every ca. 2 generations) based on estimates using private alleles (Slatkin and Barton 1989). A comparison between the type locality of Brack's cactus (15543) and a population northeast of Bloomfield (15565) is estimated with a rate is 4.286 (ca. 4 migrants per generation) based on estimates using F_{ST} and 1.137 (1 migrant every generation) based on estimates using private alleles. Supplemental Table 1 summarizes the pairwise *Nm* values using estimates based on F_{ST}.

STRUCTURE analyses, based on the 1944 loci assembled using *ipyrad*, implies complex, or at least hierarchic structure of populations. The Evanno method (ΔK) indicates that the best fit is at K= 2 clusters. This is the same as observed with assemblies using *STACKS*. Within-group genetic variance is minimized, and between-group genetic variance is maximized also when K= 2 clusters (Fig. 8A; Fig. 9). By contrast, using *Lea*, the minimum cross-entropy is achieved when K= 11 clusters (Fig. 8B). These results are similar to those found using *STACKS* in that both K= 2 and K= 11 are indicated as important clustering thresholds, using alternative criteria for selecting K values.

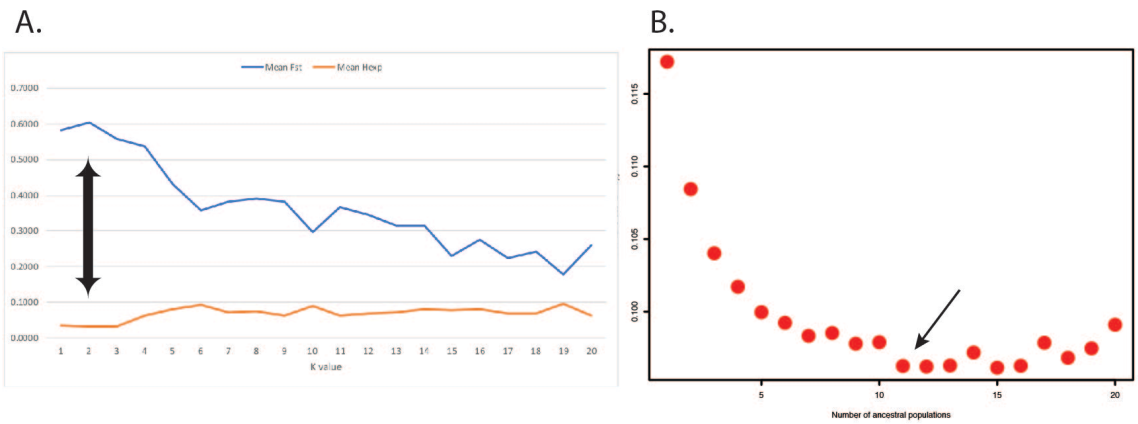


Figure 8. Comparison of *STRUCTURE* 2.3.4 analyses, using *K*-values ranging from 1 to 20, based on *ipyrad* assembly of loci. The difference in mean F_{ST} and mean expected heterozygosity (Hexp) with respect to *K* is shown in panel A. The arrow indicates a maximum of the difference at *K*= 2. Minimum cross-entropy with respect to *K*, based on analyses using *Lea*, is shown in panel B. The arrow indicates the minimum value at *K*= 11.

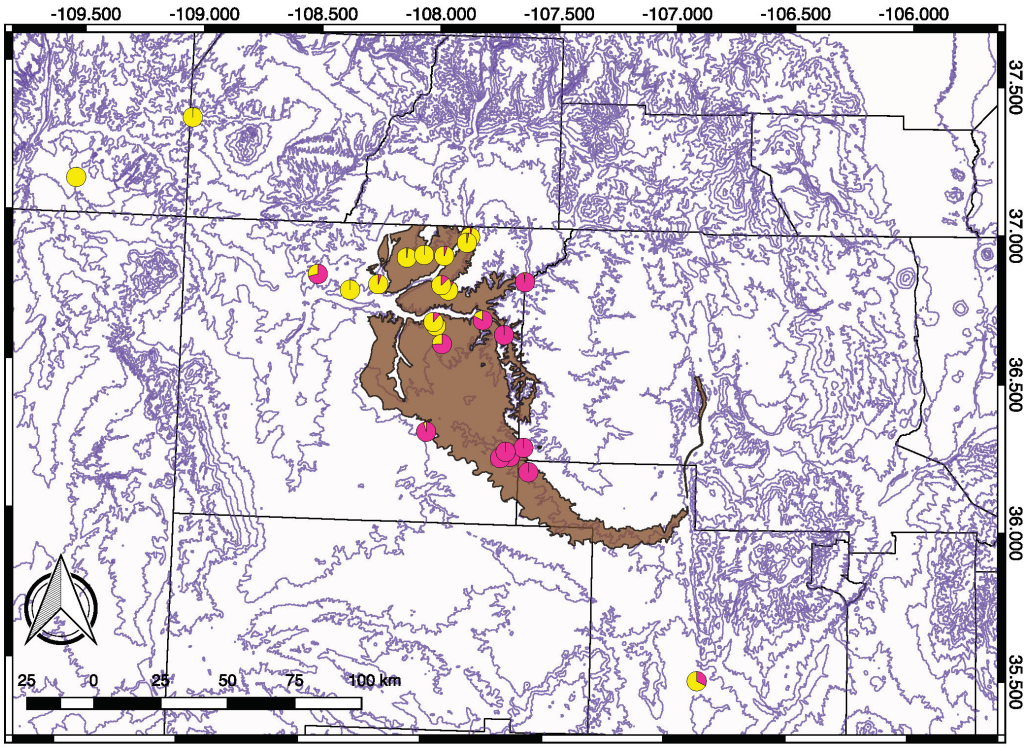


Figure 9. *STRUCTURE* 2.3.4 results at *K*= 2, for 25 populations of *Sclerocactus*, employing the *ipyrad* assembly. Populations are color-coded, yellow: northern genetic cluster; magenta: southern genetic cluster. The Nacimiento Formation is shown in brown.

The *ipyrad* pipeline generates a nucleotide sequence data file for phylogenetic analysis on a per individual basis, rather than summarizing populations. The full data set, which included outgroup, had 1944 loci with a total of 19,514 variable SNPs and totaled 175,049 nucleotides. For the parameters here selected, the maximum likelihood (ML) tree has a log likelihood of -710839.31, based on RaxML, using the GTR model with gamma distributed rates in (Fig. 10). The ML tree is very similar to the posterior probability tree from MrBayes, the branch order differing only at a few unsupported nodes. Similar to the results from the *STACKS* assembly, there is strong support for the isolation of *Sclerocactus cloverae* away and apart from *S. whipplei*. Populations of *Sclerocactus* from near Kirtland, New Mexico, from the Hogback, New Mexico and from Yellow Jacket, Colorado form an unsupported clade sister to *S. cloverae* (*S. parviflorus* is the sister to this group, a population from near San Ysidro and the remainder of *S. cloverae*). Similarly, an anomalous population from near San Ysidro, New Mexico (population 15540) is highly distinctive and sister to the remaining populations of *S. cloverae*. Populations of *S. cloverae* are monophyletic and are composed of several well supported clades. There is a primary division in *S. cloverae* separating populations of the Animas Valley and west, as well as populations from Aztec, New Mexico south to lower Kutz Canyon, from the remainder of the sampled populations. The former group includes the type localities of both *S. cloverae* and *S. cloverae* subsp. *brackii*. This implies that *S. cloverae* subsp. *brackii* and *S. cloverae* subsp. *cloverae* are taxonomically the same lineage. Within the clade representing the eastern and southern populations of S. cloverae is a strongly supported clade representing the populations occurring on the Nacimiento Formation between Nageezi and Counselor. These populations display a higher degree of intermixing, suggesting either a higher degree of gene-flow or higher degrees of lineage sorting as a consequence of recency of common ancestry. This is a contrast to other populations, which tend to be monophyletic, implying the absence of gene-flow.

Discussion

The genetic basis for subspecies status of Brack's cactus—Brack's cactus was recognized as distinct element of *Sclerocactus cloverae* based on the morphological characteristics of reproductive individuals of a population on the Nacimiento Formation

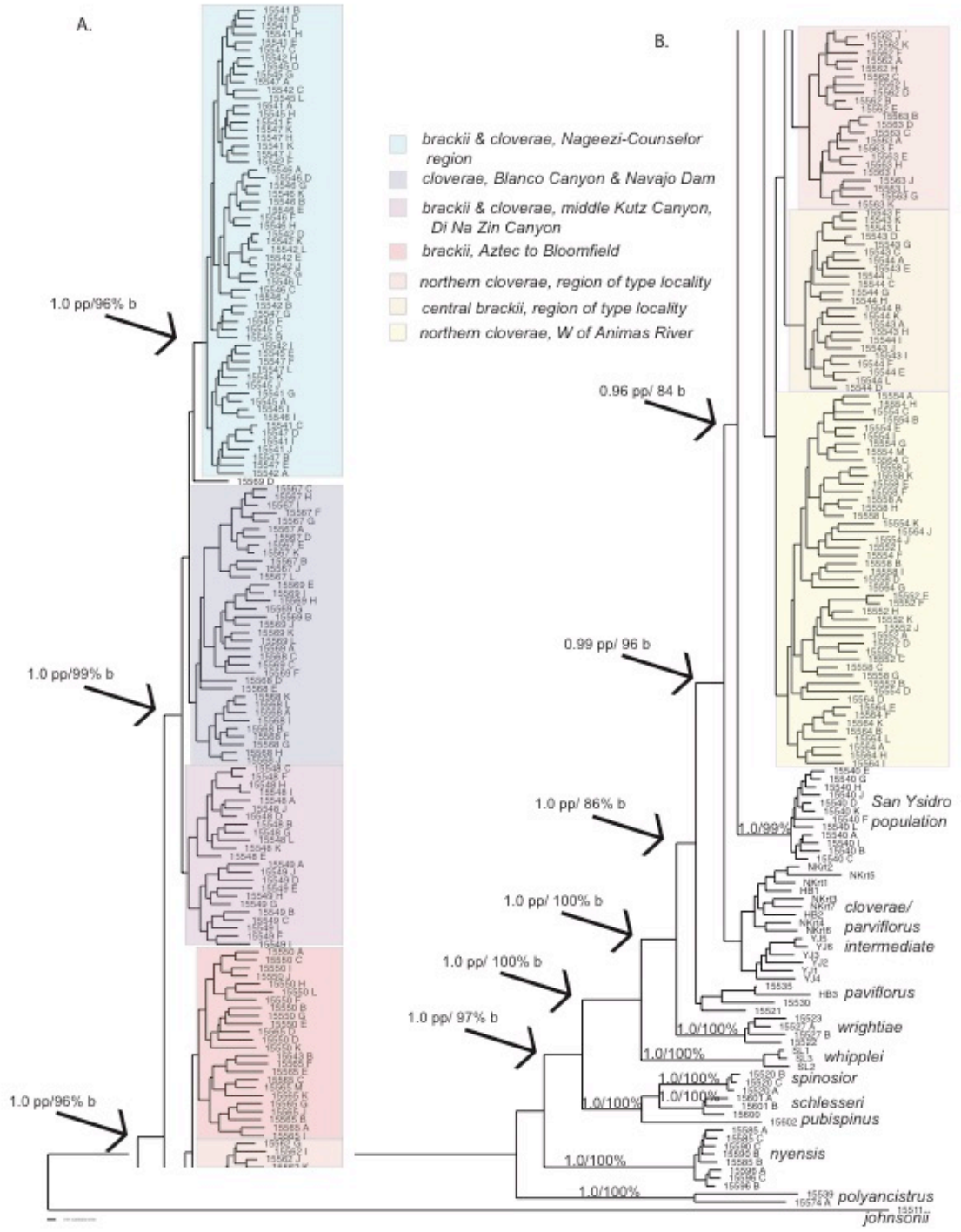


Figure 10. Maximum likelihood estimation of phylogenetic relationships among populations of *Sclerocactus cloverae* and eight other species of the genus, based on RaxML and MrBayes analyses of loci assembled using *ipyrad*. Posterior probabilities (PP) and bootstrap percentages (b) are reported for clades of interest. Inset B is a continuation of A.

south of Bloomfield, New Mexico. Individuals in this population possessed very small stems and a preponderance of individuals that lacked hooked central spines. Following the description of Brack's cactus, this morphotype has been found nearly throughout the range of *S. cloverae*. The lack of geographic contiguity of *S. cloverae* subsp. *brackii* is problematic because it is expected that a genetically-based subspecies would display geographic continuity. This concern is well founded. Genetic *K*-clustering analyses of data sets including 61 and 1914 loci, using *structure* and *Lea*, fail to find support for the two subspecies of *S. cloverae* (Figs. n and n). Nor do the PCoA analyses of genetic data (Fig. n), which indicate that the two morphotypes are not genetically distinct from one another. Genetic relationships are indeed related to geography and populations display an overall pattern of isolation by distance.

The general lack of support by ddRAD sequences for the distinction between subspecies *brackii* and *cloverae* is particularly evident in the maximum likelihood and Bayesian analyses (Fig. n), which show that some populations of *Sclerocactus cloverae* subsp. *cloverae* (i.e., population 15558) share closer ancestry with populations of subsp. *brackii* (i.e., population 15559), than with other populations of subsp. *cloverae* (i.e., population 15557). From a taxonomic point of view, it is critical to recognize that the population from which the type specimen of subsp. *brackii* came is found in the same genetic grouping as the population from which the type specimen of subsp. *cloverae* was taken, based on *STRUCTURE* (*K*= 2; Fig. n) and RaxML/MrBayes (Fig. n) analyses. This implies that, by definition (i.e., based on type material), subsp. *brackii* is the same taxon as subsp. *cloverae*. The only conclusion that can be drawn is that Brack's cactus is synonymous with Clover's cactus and therefore should have no taxonomic standing.

How could morphological evidence have led to such a contrary conclusion regarding subspecies status? We suggest that the morphological differences observed among populations of *Sclerocactus cloverae* are real. However, the basis of those differences is not genetic difference between populations, but instead a difference caused by development and the age of populations. As noted by Heil and Porter (1994), the morphology of *S. cloverae* changes dramatically with age. Young, mostly pre-reproductive individuals have small, more globose stems, a fewer number of spines and

the notable absence of central, hooked spines. As plants age, the stems become cylindrical, the number of spines increases (both radial and central spines) and a prominent hooked central spine is evident in each areole. Brack's cactus was believed to represent a paedomorphic form, retaining the juvenile stem and spine morphology when reaching reproductive maturity (Heil and Porter 1994). This may not to be the case, or at least is not the entire explanation. We believe that rather than a paedomorph, we are observing an age-stratified population (i.e., populations in which all individuals are approximately the same age).

It has been noted in a number of species of *Sclerocactus* that periodic infestation of cactus weevil (*Gerstaeckeria sp*.) or cactus longhorn beetle (*Moneilema semipunctatum*) occurs, i.e., *S. mesae-verdae*, *S. wetlandicus*, *S. wrightiae*, *S. brevihamatus* subsp. *tobuschii* (Kass 2001; Coles et al. 2012). This has been observed in *S. cloverae* as well (Porter, pers. obs.). A consequence of particularly extensive infestations of cactus longhorn beetle is the death of nearly all mature, large-stemmed individuals. Following this mortality, seed germination results in a nearly even-aged stand of *S. cloverae*. In addition, another possible consequence of cactus longhorn beetle infestation is flowering and reproduction at an earlier age. Since larger plants appear to be preferentially targeted by beetles, reproducing earlier may be selectively advantageous. The sooner the plant reproduces the greater the probability of avoiding predation while maintaining fitness. While not demonstrated to be a selective agent, it is certainly a plausible scenario. The combination of even aged populations and flowering while the stem retains juvenile morphology results in populations with a common morphology, corresponding to the Brack's cactus form. However, in time and in the absence of beetle infestation, the population will eventually take on the morphology of Clover's cactus. Here again, the largely even aged stand results in the preponderance of individuals with a common morphology.

Genetic differentiation in *Sclerocactus cloverae*—While there is no genetic merit for taxonomic status of Brack's cactus, this does not mean that *Sclerocactus cloverae* is genetically uniform across its relatively small range. In spite of its endemism, AMOVAs indicate that there is significant genetic differentiation among populations. This

differentiation is evident as well in the *STRUCTURE* analyses (Figs. x and x), which tend to display evidence of hierarchic structure. Strong evidence exists for genetic divergence of at least two primary groups: one composed of more southern populations, centering on populations on the Nacimiento Formation of southeastern San Juan and adjacent Sandoval and Rio Arriba Counties; while the other, is more northern in distribution mostly north of the San Juan River (with the exception of populations near the type locality of Brack's cactus). Mean Φ_{ST} among the southern populations of the Nageezi to Counselors region (populations 15541, 15543, 15554–7) is 0.017 (*ipyrad*) and 0.029 (*STACKS*), slightly less than within-group F_{ST} values observed in *S. glaucus* (F_{ST}= 0.058–0.072; McGlaughlin and Ramp Neale 2017), using microsatellite loci. These values are lower than mean Φ_{ST} between the southern populations and populations at the type localities of Brack's cactus and Clover's cactus, 0.087 and 0.126 (*ipyrad* and *STACKS*, respectively). McGlaughlin and Ramp Neale (2017) reports between group (northern vs. southern *S. glaucus* groups) F_{ST}= 0.121–0.122. *F*-statistics based on *STRUCTURE* analyses imply an even higher degree of differentiation of these groups, F_{ST}= 0.3305 based on the *STACKS* assembly and F_{ST}= 0.5086 based on the *ipyrad* assembly.

While the genetic divergence of the populations in the Nageezi to Counselors region might merit taxonomic recognition, the morphological basis for such a change is lacking. Population in this southern region display the range of morphological forms spanning Brack's cactus and Clover's cactus. This is precisely the same degree of morphological variation observed in among the northern populations. We must assume that the genetic differences are associated with non-morphological traits, e.g., physiological differences. Even though we believe that it is not productive to name the southern populations as a new subspecies, it is critical to recognize that these populations are genetically distinct from those in the north. McGlaughlin and Ramp Neale (2017) has suggested that the northern populations group of *S. glaucus* warranted individual management considerations, apart from *S. glaucus* in the Gunnison River and Grand Valley regions. This may also be true for the southern populations of *S. cloverae*. This genetic difference is particularly important if plants are removed from their native populations to be transplanted into other populations or at different sites. Without experimental evidence to the contrary, it is safer to assume that genetic mixing of these

two genetic groups could have negative fitness consequences. The genetic differences we observe suggests to us that transplants should remain within genetic groupings. That is, plants from northern populations should remain in the north and those from southern population should remain in the south.

As noted above, genetic differentiation appears to be somewhat hierarchical. This is particularly true of the loci identified by the *STACKS* assembly and analyzed using *STRUCTURE*. While Evanno's ΔK method identifies $K = 2$ as the preferred clustering, mean F_{ST} values increase with increasing K values up through, at least, $K = 11$. Further, the minimum cross entropy method used by *Lea* selects $K > 20$. This basically indicates that nearly all populations sampled can be genetically discriminated. This is also evident in the ML phylogenetic tree (Fig. x), in which most of the populations are recovered as monophyletic groups. It seems unlikely that each population is being fixed by selection; however, high mortality rates, driven by cactus weevil infestation, could allow genetic drift to affect change. Again, our ability to detect these differences likely is a consequence of the large number of loci that are sampled using ddRAD methods.

Relationships of *Sclerocactus cloverae, S. whipplei* and *S. parviflorus*—*Sclerocactus clover*** was originally recognized as a pair of varieties (vars. *heilii* and *reevesii*) of *S. whipplei*. Heil and Porter (1994) were the first to treat this taxon at the species rank and apart from *S. whipplei*. These authors point out that *S. cloverae* possesses a suite of traits that could be argued to align it with either *S. parviflorus* or *S. whipplei*. At the same time, the combination of traits found in *S. cloverae* did not permit unambiguous affiliation with either. Hochstätter (1997) has treated *S. cloverae* as a subvariety of *S. whipplei* (i.e., *S. whipplei* subvar. *aztecia* Hochstätter), but more recently as a subspecies (Hochstätter 2007: *S. whipplei* subsp. *cloverae* [K.D. Heil & J.M. Porter] Hochstätter). Our data has relevance to this debate. The ML tree (Fig. 10) reveals significant support for the placement of *S. whipplei* as the sister to a clade that includes *S. wrightiae* Woodruff & Benson, *S. parviflorus* and *S. cloverae*. This relationship leads to the conclusion that *S. cloverae* should not be included within *S. whipplei* unless *S. wrightiae* and *S. parviflorus* are also included. This, we believe, provides the basis to reject the treatment of *S. cloverae* as an element (e.g., subspecies) of *S. whipplei*. The phylogeny also infers that *S.*

cloverae is most closely related to *S. parviflorus*. Samples representing *S. parviflorus* subsps. *parviflorus*, *intermedius*, *terre-canyonae* and one individual from the Hogback in New Mexico, form a clade. However, the relationships between *S. cloverae* and *S. parviflorus* require further investigation. Several populations were included (15540, NKrt, YJ and HB) that are morphologically intermediate between *S. cloverae* and *S. parviflorus*. They are anomalously placed in the phylogeny as a grade between *S. cloverae* and *S. parviflorus*. It is unclear if these populations represent part of a hybrid zone between *S. cloverae* and *S. parviflorus*, or more isolated lineages of *S. cloverae*. It should be pointed out that some of these populations do display patterns of additivity, but no more than do some populations of *S. cloverae*. Moreover, some populations, i.e., 15540, generally do not display a signature of additivity. Because several relevant species, e.g., *S. glaucus* (K. Schum.) L.D. Benson, *S. wetlandicus* Hochstätter and *S. brevispinus* K.D. Heil & J.M. Porter have not been sampled in these analyses, placement of *S. parviflorus* could change with their inclusion. Even so, at this point it appears that *S. parviflorus* is the closest related species to *S. cloverae*.

Conclusions— Sequence-based genotyping data, representing 3,770 loci (which filtered to 61–74 loci) using the *STACKS* pipeline, and 193,982–219,932 loci (which filtered to 11–99,249 loci) using the *ipyrad* pipeline, were used to investigate genetic diversity and subspecies status in *Sclerocactus cloverae*. We describe two data sets produced using optimal parameters in *STACKS* and *ipyrad*, composed of 61 and 1,944 loci, respectively. These data show that *S. cloverae* is a genetically structured species, but that structure is inconsistent with the two currently recognized taxonomic subspecies, *brackii* and *cloverae*. We show that the populations from which the types of both *S. cloverae* subsp. *brackii* and *S. cloverae* subsp. *cloverae* were taken, represent the same genetic group. This renders Brack's cactus a synonym of Clover's cactus.

At the same time, there exists significant structure and genetic differentiation within *S. cloverae*. In particular, the southern populations represent a genetically differentiated group, which may require management as a distinctive unit. This genetic distinction is evident in both data sets produced by ddRAD sequences and regardless of the degree of structure (i.e., $K = 2$ or $K = 11$). While it is not the purpose of this report to

summarize conservation need of *S. cloverae*, we note that the overall range of the two genetic units combined is comparable to that of *S. glaucus* (K. Schum.) L.D. Benson, *S. mesae-verdae* (Boissev. & C. Davidson) L.D. Benson and *S. wrightiae* L.D. Benson, three species currently listed as threatened or endangered. Indeed, even with the combination of the two subspecies, *S. cloverae* is a rare and endemic species. These data further indicate that *S. cloverae*, originally described as two varieties of *S. whipplei* (vars. *heilii* and *reevesii*), is more closely related to *S. parviflorus* than to *S. whipplei*.

There is sufficient evidence that *S. cloverae* is a genetically lineage composed of semi-isolated populations which undergo significant gene flow, and therefore, we consider it appropriate to treat it at the rank of species. Even so, there is evidence of introgression and additivity in populations north of Kirtland, at the Hogback and southwest of San Ysidro, New Mexico. This introgression appears to involve genetically isolated elements of *S. parviflorus*; however, broader sampling in *S. parviflorus* is needed to clarify these relationships. While there is some evidence of hybrid introgression at in a few peripheral populations involving *S. cloverae* and *S. parviflorus*, there is no evidence of hybridization involving *S. whipplei*.

Literature Cited

Alexander, D.H., and K. Lange. 2011. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics*. 2011 Jun 18; 12: 246. doi: 10.1186/1471-2105-12-246.

Baker, M. A., and J. M. Porter. 2016. The use of multivariate analyses of morphological characters and DNA analyses for assessing the taxonomic ranking of rare plant taxa: an example comparing populations of *Echinomastus erectocentrus* var. *acunensis* with those of its relatives *E. erectocentrus* var. *erectocentrus* and *E. johnsonii*. In: Endangered Species: Threats, Conservation and Future Research Nova Science Publishers, Inc. Hauppauge, NY.

Beerli, P. 2005. MIGRATE: Documentation and Program, Part of LAMARC. Version 2.0. http://evolution.gs.washington.edu/lamarc.html.

Benestan, L., T. Gosselin, C. Perrier, B. Sainte-Marie, R. Rochette and L. Bernatchez. 2015. RAD genotyping reveals fine-scale genetic structuring and provides powerful population assignment in a widely distributed marine species, the American lobster (*Homarus americanus*). *Molecular Ecology* 24: 3299-315.

Benson, L.D. 1982. Cacti of the United States and Canada. Stanford Univ Press. 1044 p.

Besnier, F., and K.A. Glover. 2013. ParallelStructure: A R package to distribute parallel runs of the population genetics program STRUCTURE on multi-core computers. PLOS One https://doi.org/10.1371/journal.pone.0070651

Brown, A.H.D., M.W. Feldman and E. Nevo. 1980. Multilocus structure of natural populations of *Hordeum spontaneum*. *Genetics* 96: 523–536.

Bryant, D., and V. Moulton 2004. Neighbor-Net: An agglomerative method for the construction of phylogenetic networks. *Molecular Biology and Evolution* 21: 255–265, https://doi.org/10.1093/molbev/msh018.

Castetter, E.F., P. Pierce, and K.H. Schwerin. 1976. A new cactus species and two new varieties from New Mexico. Cactus and Succulent Journal (Los Angeles) 48: 77–82.

Catchen, J., A. Amores, P. Hohenlohe, W. Cresko and J. H. Postlethwait. 2011. Stacks: building and genotyping loci de novo from short-read sequences. G3: Genes, Genomes and Genetics 1: 171–182.

Catchen, J., S. Bassham, T. Wilson, M. Currey, C. O'Brien, Q. Yeates and W.A. Cresko. 2013. The population structure and recent colonization history of Oregon three-spine stickleback determined using restriction-site associated DNA-sequencing. *Molecular Ecology* 22: 2864–2883.

Chifman J., and L. Kubatko. 2014. Quartet inference from SNP data under the coalescent model. *Bioinformatics* 30: 3317–3324.

CITES 2015

Coles, J.J., K.L. Decker and T.S. Naumann. 2012. Ecology and population dynamics of *Sclerocactus mesae-verdae* (Boissev. & C. Davidson) L.D. Benson. Western North American Naturalist. 72: 311–322.

Copetti, D., A. Búrquez, E. Bustamante, J. L. M. Charboneau, K. L. Childs, L. E. Eguiarte, S. Lee, T. L. Liu, M. M. McMahon, N. K. Whiteman, R. A. Wing, M. F. Wojciechowski, and M. J. Sanderson. 2017. Extensive gene tree discordance and hemiplasy shaped the genomes of North American columnar cacti. *Proceedings of the National Academy of Sciences of the United States of America* 114: 12003–12008.

Davey, J. W., P.A. Hohenlohe, P.D. Etter, J.Q. Boone, J.M. Catchen and M.L, Blaxter. 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics* 12: 499–510.

de Queiroz, K. 1998. The general lineage concept of species, species criteria, and the process of speciation. In: Endless Forms: Species and Speciation (eds. Howard SJ, Berlocher SH), pp. 57–78. Oxford University Press, New York.

de Queiroz, K. 1999. The general lineage concept of species and the defining properties of the species category. Pages 49–89 in Species: New interdisciplinary essays (R. A. Wilson, ed.). MIT Press, Cambridge, Massachusetts.

de Queiroz, K. 2005. Ernst Mayr and the modern concept of species. *Proc. Natl. Acad. Sci. USA* 102: 6600–6607.

de Queiroz, K. 2007. Species concepts and species delimitation. *Syst. Biol*. 56: 879-886.

Dobzhansky T. 1939 Genetics of natural populations. IV. Mexican and Guatemalan populations of *Drosophila pseudoobscura*. Genetics 24: 391–412.

Dobzhansky, T. 1970. Genetics of the evolutionary process. Columbia University Press, New York.

Doyle, J.J., and J.L. Doyle. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. 19: 11-15.

Durand E.Y., N. Patterson, D. Reich and M. Slatkin. 2011. Testing for ancient admixture between closely related populations. Mol. Biol. Evol. 28: 2239–2252.

Earl, D.A. and B.M. von Holdt. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* 4: 359-361. doi: 10.1007/s12686-011-9548-7

Eaton, D.A. 2014. PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. Bioinformatics 30:1844–1849. doi: 10.1093/bioinformatics/btu121.

Eaton, D.A.R., and R.H. Ree. 2013. Inferring Phylogeny and Introgression using RADseq Data: An Example from Flowering Plants (Pedicularis: Orobanchaceae). Syst. Biol. 62: 689–706.

Evanno, G., S. Regnaut and J. Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* 14: 2611–2620. doi:10.1111/j.1365-294X.2005.02553.x

Excoffier, L., G. Laval, and S. Schneider. 2005. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* 1: 47-50.

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

Frichot, E., and O. François. 2015. LEA: An R package for landscape and ecological association studies. *Methods Ecol Evo*l 6: 925–929.

Frichot, E., F. Mathieu, T. Trouillon, G. Bouchard and O. François. 2014. Fast and efficient estimation of individual ancestry coefficients. *Genetics* 196: 973–983.

Goff, S.A., et al., 2011. The iPlant Collaborative: Cyberinfrastructure for plant biology. *Frontiers in Plant Science* 2 (2011), doi: 10.3389/fpls.2011.00034.

Griffith, M. P. and J. M. Porter. 2003. Back to basics: a simple method of DNA extraction for mucilaginous cacti. *Bradleya* 21: 126-128.

Grünwald, N.J., S.B. Goodwin, M.G. Milgroom and W.E. Fry. 2003. Analysis of genotypic diversity data for populations of microorganisms. *Phytopathology* 93: 738–746.

Heil, K. D. and J. M. Porter. 1994. *Sclerocactus* (Cactaceae): A revision. *Haseltonia* 2: 20-46.

Heil, K. D. and J. M. Porter. 2003. Sclerocactus, pp. 197–206. In Flora of North America Editorial Committee [eds.]. Flora of North America north of Mexico, vol. 4. New York, USA, and Oxford, UK.

Hey, J., and C. Pinho. 2012. Population genetics and objectivity in species diagnosis. *Evolution* 66: 1413–1429.

Hochstätter, F. 1997. The genus *Sclerocactus* (Cactaceae)---Part 4. *Cact. Succ. J. Gr. Brit*. 15: 74-81.

Hochstätter, F. 2005. The Genus *Sclerocactus* Vol.1, published by author, Manheim.

Hubisz, M. J., D. Falush, M. Stephens, and J. K. Pritchard. 2009. Inferring weak population structure with the assistance of sample group information. Mol. Ecol. Resour. 9: 1322–1332.

Huelsenbeck, J.P., and F. Ronquist; MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17: 754–755, https://doi.org/10.1093/bioinformatics/17.8.754

Huson, D. H. and D. Bryant. 2006. Application of Phylogenetic Networks in Evolutionary Studies. *Molecular Biology and Evolution* 23:254-267.

Janes, J.K., J.M. Miller, J.R. Dupuis, R.M. Malenfant, J.C. Gorrell, C.I. Cullingham and R.L. Andrew. 2017. The K = 2 conundrum. Molecular Ecology 26: 3594–3602.

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

Kass, R. 2001. Mortality of Endangered Wright Fishhook Cactus (*Sclerocactus wrightiae*) by an Apuntia-Borer Beetle (Cerambycidae: *Moneilema semipunctatum*). *Western North America Naturalist* 61: 495-497.

Langmead, B., and S. Salzberg. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods* 9: 357-359.

Ludwig, J.A., and J.F. Reynolds. 1988. Statistical ecology: A primer in methods and computing. Wiley-Interscience.

Manard Smith, J., N.H. Smith, M. O'Rourke and B.G. Spratt. 1993. How clonal are bacteria. *Proceedings of the National Academy of Sciences* 90: 4384–4388.

Mayr, E. 1942. Systematics and the Origin of Species. Columbia Univ. Press, New York.

Mayr, E. 1982. Speciation and macroevolution. *Evolution* 36: 1119–1132.

McGlaughlin, M., and J. Ramp Neale. 2017. Genetic structure of *Sclerocactus glaucus* (Colorado Hookless Cactus) in Western Colorado, Status Report – November, 2017. Unpubl. report

Merchant, N., et al. 2016. The iPlant Collaborative: Cyberinfrastructure for enabling data to discovery for the life sciences. PLOS Biology (2016), doi: 10.1371/journal.pbio.1002342.

Meredith, E. P., J. A. Rodzen, J. D. Banks, R. Schaefer, H. B. Ernest, T. R. Famula and B. P. May. 2007. Microsatellite analysis of three subspecies of elk (*Cervus elaphus*) in California. *Journal of Mammalogy* 88:801-808.

Miller, M.A., W. Pfeiffer and T. Schwartz. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Proceedings of the Gateway Computing Environments Workshop (GCE), 14 Nov. 2010, New Orleans, LA pp 1–8.

Mudrik, E. A., T. A. Kashentseva, P. S. Redchuk and D. V. Politov. 2015. Microsatellite variability data confirm low genetic differentiation of western and eastern subspecies of common crane *Grus grus* L. (Gruidae, Aves). *Molecular Biology* 49: 260-266.

Mullen, L. M., S. N. Vignieri, J. A. Gore and H. E. Hoekstra. 2009. Adaptive basis of geographic variation: genetic, phenotypic and environmental differences among beach mouse populations. *Proceedings of the Royal Society of London B* 276: 3809-3818. DOI: 10.1098/rspb.2009.1146.

Nei, M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* 89: 583–590.

Peakall, R., and P.E. Smouse 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* 6: 288–295.

Peakall, R., and P.E. Smouse 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics* 28: 2537–2539.

Pielou, E.C. 1975. Ecological diversity. Wiley New York.

Porter, J. M., J. Cruse-Sanders, L. M. Prince, and R. Lauri. 2012. Species status of *Sclerocactus brevispinus*, *S. wetlandicus*, and *S. glaucus*: inferences from morphology, chloroplast DNA sequences, and AFLP markers. *Aliso* 30: 69-83.

Porter, J. M., M. S. Kinney, and K. D. Heil. 2000 Relationships between *Sclerocactus* and *Toumeya* (Cactaceae) based on chloroplast *trnL–trnF* sequences. *Haseltonia* 7: 8–23.

Pritchard, J.K., M. Stephens, and P. Donnelly. 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 Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Rochette, N.C., and J.M. Catchen 2017. Deriving genotypes from RAD-seq short-read data using Stacks. *Nature Protoc.* 2017 Dec;12(12):2640-2659. doi: 10.1038/nprot.2017.123. Epub 2017 Nov 30.

Ronquist, F., and J.P. Huelsenbeck. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19: 1572–1574, https://doi.org/10.1093/bioinformatics/btg180

Saitou, N., and M. Nei. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4: 406-425.

Schwabe, A.L., A.R. Hubbard, J.R. Neale, and M.E. McGlaughlin. 2013. Microsatellite loci development for rare Colorado *Sclerocactus* (Cactaceae). *Conservation Genetics Resources* 5: 69-72.

Schwabe, A.L., J.R. Neale, and M.E. McGlaughlin. 2015. Examining the genetic integrity of a rare endemic Colorado cactus (*Sclerocactus glaucus*) in the face of hybridization threats from a close and widespread congener (*Sclerocactus parviflorus*). *Conservation Genetics* 16: 443-457.

Shaffer, H.B., and R.C. Thomson. 2007. Delimiting species in recent radiations. *Systematic Biology* 56: 896–906.

Shannon, C.E. 2001. A mathematical theory of communication. ACM SIGMOBILE Mobile Computing and Communications Review 5: 3–55.

Simpson, E.H. 1949. Measurement of diversity. *Nature* 163: 688.

Slatkin, M., and N.H. Barton. 1989. A comparison of three indirect methods for estimating average levels of gene flow. *Evolution* **43:** 1349-1368.

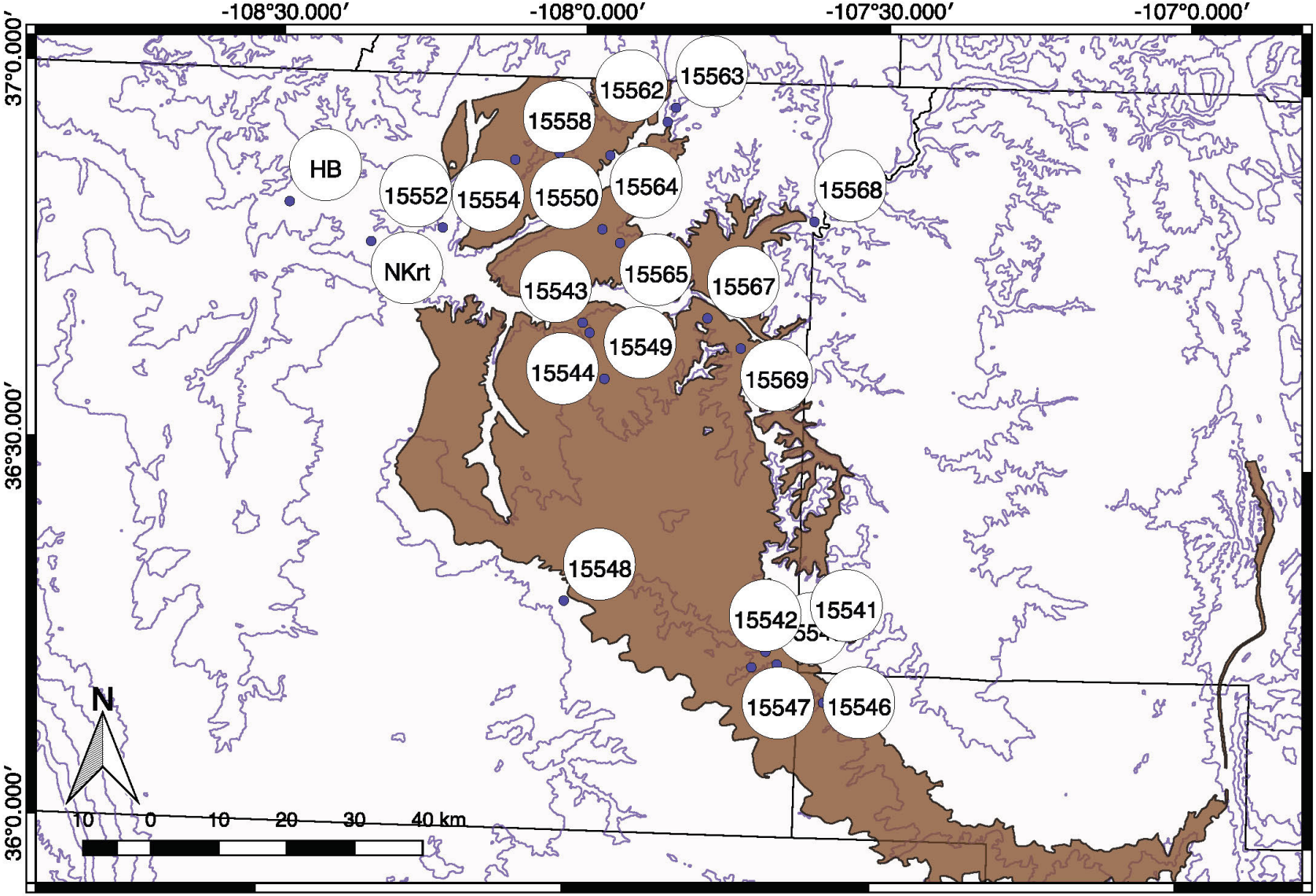
Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30: 1312-1313, doi: 10.1093/bioinformatics/btu033

Stamatakis, A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22, 2688–2690.

Stoddart, J.A., and J.F. Taylor. 1988. Genotypic diversity: Estimation and prediction in samples. *Genetics* 118: 705–711.

Truong, H.T, A.M Ramos, F. Yalcin, M. de Ruiter, H.J.A. van der Poel, K.H.J. Huvenaars, R.C.J. Hogers, L.J.G. van Enckevort, A. Janssen, N.J. van Orsouw and M.J.T. van Eijk. 2012. Sequence-based genotyping for marker discovery and co-dominant scoring in germplasm and populations. PLOS One 7(5): e37565, https://doi.org/10.1371/journal.pone.0037565

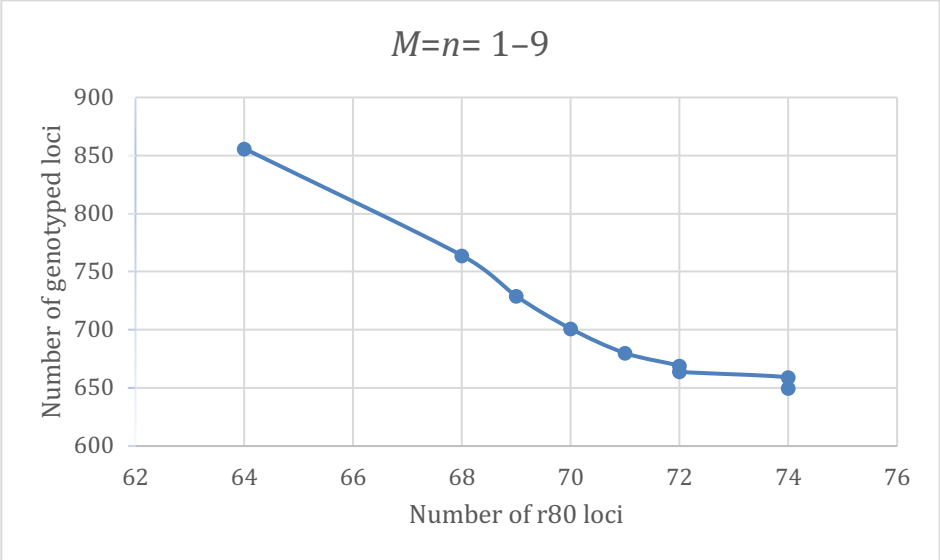
Wright, S. 1931. Evolution in Mendelian populations. *Genetics* **16:** 97-159.



Supplemental Figure 1. Location and identity (population number) of sampled sites of *Sclerocactus cloverae*. Hidden number is 15545.

Supplemental Figure 2. Distribution of loci captured by SBG/dd-RAD sequencing across all samples.

Supplemental Figure 3. Relationship between the number of genotyped loci recovered by *STACKS* assembly and the number of loci found in 80% or more of the individuals in the sample with -*M* (maximum nucleotide differences between alleles) and -*n* (maximum nucleotide differences between loci parameters ranging from one to nine.



Supplemental Table 2. Pairwise population migration rate estimates (*Nm*), using F_{ST} values (Wright 1931; Slatkin and Barton 1989). Values based on F_{ST} estimates from *ipyrad* assembly, above the diagonal; values based on F_{ST} estimates from STACKS assembly, below the diagonal.

	15540	15541	15542	15543	15544	15545	15546	15547	15548	15549	15550	15552	15554	15558	15562	15563	15564	15565	15567	15568	15569	HB	NKrt	YJ	
15540	—	1.349	1.260	1.358	1.371	1.282	1.093	1.192	1.281	1.282	1.159	1.356	1.404	1.336	1.174	1.258	1.664	1.079	1.346	1.234	1.545	0.259	1.019	0.887	15540
15541	0.137	—	9.096	2.814	3.252	30.090	9.083	48.320	4.738	4.720	2.612	3.226	3.043	3.414	2.821	2.465	5.285	2.785	3.677	4.002	5.396	0.279	1.179	1.091	15541
15542	0.113	0.210	—	2.615	2.971	32.006	8.300	15.236	3.705	3.860	2.640	2.534	2.559	3.133	2.180	2.151	3.923	2.475	2.853	3.000	3.553	0.251	1.024	0.946	15542
15543	0.107	0.115	0.134	—	n.a.	2.850	2.051	2.488	2.395	3.273	6.101	4.127	5.675	6.253	5.331	5.027	8.600	4.286	3.838	2.167	3.365	0.317	1.219	1.083	15543
15544	0.112	0.129	0.143	0.209	—	3.043	2.162	2.890	2.778	3.976	8.062	5.190	6.457	7.044	6.396	6.142	15.354	5.088	3.950	2.367	3.560	0.369	1.452	1.333	15544
15545	0.100	0.130	0.179	0.141	0.131	—	19.100	n.a.	5.106	5.267	2.689	2.689	2.910	3.625	2.502	2.291	4.539	2.931	4.110	4.397	5.239	0.240	1.016	0.977	15545
15546	0.118	0.172	0.208	0.123	0.117	0.205	—	14.752	3.057	3.274	1.989	2.075	1.998	2.258	1.886	1.743	3.243	2.135	2.540	2.919	2.808	0.231	1.008	0.918	15546
15547	0.133	0.184	0.184	0.127	0.128	0.237	0.224	—	4.453	4.659	2.449	2.314	2.538	2.869	2.263	2.078	4.429	2.811	3.950	4.436	5.227	0.237	1.019	0.912	15547
15548	0.101	0.127	0.135	0.136	0.124	0.123	0.127	0.142	—	5.081	2.463	2.542	2.593	2.959	2.437	2.200	3.974	2.427	3.381	3.203	3.761	0.264	1.076	1.038	15548
15549	0.107	0.118	0.124	0.162	0.187	0.139	0.128	0.129	0.139	—	3.620	3.727	3.648	4.580	3.863	3.272	12.103	4.528	4.639	3.632	4.499	0.333	1.565	1.407	15549
15550	0.080	0.084	0.090	0.136	0.153	0.091	0.079	0.085	0.110	0.159	—	4.550	5.261	5.476	6.989	6.625	15.442	6.062	2.956	2.020	2.967	0.302	1.194	1.106	15550
15552	0.092	0.115	0.101	0.136	0.145	0.088	0.092	0.095	0.107	0.162	0.161	—	13.152	7.565	5.200	6.165	14.462	2.993	3.053	1.849	3.102	0.417	1.772	1.641	15552
15554	0.088	0.097	0.095	0.109	0.138	0.081	0.077	0.091	0.103	0.128	0.104	0.142	—	21.344	6.513	7.391	34.991	3.710	3.831	1.969	3.688	0.365	1.497	1.471	15554
15558	0.075	0.094	0.088	0.102	0.106	0.072	0.071	0.077	0.084	0.103	0.113	0.122	0.143	—	5.534	5.112	38.774	4.885	3.877	2.324	3.788	0.301	1.286	1.276	15558
15562	0.087	0.104	0.085	0.123	0.142	0.087	0.078	0.092	0.115	0.134	0.135	0.132	0.101	0.093	—	16.312	28.700	5.284	2.906	1.908	3.073	0.345	1.294	1.141	15562
15563	0.080	0.101	0.093	0.123	0.165	0.079	0.074	0.089	0.111	0.120	0.129	0.129	0.119	0.092	0.158	—	17.985	3.543	2.965	1.675	2.695	0.440	1.650	1.336	15563
15564	0.097	0.107	0.104	0.133	0.163	0.107	0.087	0.109	0.146	0.158	0.175	0.151	0.132	0.129	0.144	0.124	—	12.498	5.975	3.393	5.003	0.488	2.679	2.786	15564
15565	0.092	0.106	0.107	0.118	0.141	0.089	0.100	0.108	0.107	0.178	0.138	0.160	0.118	0.108	0.108	0.112	0.134	—	3.522	2.701	3.356	0.247	1.072	0.975	15565
15567	0.084	0.093	0.102	0.125	0.138	0.088	0.083	0.105	0.108	0.140	0.116	0.121	0.114	0.091	0.098	0.111	0.140	0.125	—	3.435	5.952	0.282	1.216	1.175	15567
15568	0.096	0.130	0.137	0.093	0.110	0.125	0.139	0.135	0.093	0.123	0.093	0.097	0.081	0.077	0.082	0.067	0.116	0.107	0.098	—	4.271	0.211	0.925	0.861	15568
15569	0.119	0.126	0.119	0.142	0.143	0.108	0.116	0.141	0.155	0.168	0.124	0.117	0.114	0.100	0.104	0.107	0.124	0.132	0.133	0.121	—	0.257	1.069	1.024	15569
HB	0.039	0.044	0.047	0.055	0.066	0.039	0.036	0.044	0.043	0.049	0.052	0.063	0.062	0.044	0.055	0.045	0.074	0.051	0.045	0.035	0.042	—	18.972	0.574	HB
NKrt	0.083	0.087	0.100	0.092	0.107	0.082	0.085	0.088	0.085	0.109	0.100	0.108	0.102	0.088	0.087	0.080	0.122	0.131	0.096	0.087	0.079	0.078	—	8.203	NKrt
YJ	0.109	0.095	0.108	0.098	0.138	0.093	0.098	0.117	0.091	0.155	0.119	0.114	0.101	0.097	0.089	0.084	0.145	0.193	0.106	0.111	0.104	0.046	0.173	—	YJ