Scientific
Research
Publishing

New Interpretations about Clonal Architecture for the Sour Pitaya (*Stenocereus gummosus*, Cactaceae), Arising from Microsatellite Markers of *de novo* Isolation and Characterization

Oscar Adrián Lozano Garza[1], José L. León de la Luz[1], Susana Favela Lara[2], Francisco J. García de León[1*]

[1]Laboratorio de Genética para la Conservación, Centro de Investigaciones Biológicas del Noroeste, La Paz, Mexico
[2]Departamento de Ecología, Facultad de Ciencias Biológicas, Universidad Autónoma de Nuevo León, San Nicolás de los Garza, Mexico
Email: edrian.lozano@gmail.com, jlleon04@cibnor.mx, sfavela@gmail.com, *fgarciadl@cibnor.mx

Abstract

The use of molecular markers has improved the ecological and evolutionary research in the case of clonal reproduction species, allowing the identification of boundaries among clonal modules (ramets), genetic individuals (genets), and populations. Microsatellite markers were developed for *Stenocereus gummosus*, a columnar cactus with both sexual and clonal reproduction which is distributed in the Sonoran Desert, Mexico. 454-Pyrosequencing reads were analyzed to detect microsatellite markers. Forty primer pairs were screened to look for polymorphism. Nine loci were genotyped for two *S. gummosus* localities. Sampling strategy was intended to avoid collecting a genetic individual more than once, considering that clonal architecture for this cactus was previously deduced as clumped. Polymorphic loci exhibited low allele count, ranging from 2 - 7 (mean of 3.78 ± 0.62 SE); mean heterozygosity values were 0.221 and 0.234 H_O and 0.408 and 0.306 H_E, with F_{IS} of 0.383 and 0.299, for peninsular and continental localities respectively. Unexpectedly, some multilocus genotypes were found repeated within locality, which were assumed as clones since data was evaluated as sufficient (clonal richness R of 0.966 and 0.897). These results were different from those previously reported: the distribution of clones might as well be intermingled, having a minimum ramet dispersion distance of 30 m. This characteristic was also consistent with

---

*Corresponding author.

the recent colonization proposed for this and other arid lands plants. A wider genetic neighborhood, due to clone dispersion might affect diversity indexes while increasing the chance of geitonogamy and mating among relatives. The markers isolation and its population characterization allowed addressing new questions about *S. gummosus* ecology, clonal reproduction and reproductive biology.

Keywords

Cactaceae, Clonal Architecture, Microsatellites, Pyrosequencing, Stenocereus

1. Introduction

The use of molecular markers has increased the opportunity to address ecological and evolutionary questions about complex lifestyle organisms. Models originally developed for exclusively sexual reproduction life forms are currently implemented for molecular research on organisms capable of clonal reproduction [1]. An adequate identification of clonal membership is a crucial step for studying populations of clonal organisms, and a task for which highly polymorphic microsatellites have shown to gather enough statistical power [2].

Sexual reproduction is believed to be more important than clonal propagation in columnar cacti [3], despite of several mechanisms of clonal fragmentation and growth have been described for this group [4]-[7]. *Stenocereus gummosus* (Engelm. ex. K. Brandegee) A.C.Gibson & K.E.Horak, is not an exception, since the energy expenditure for flower and fruit development, along with its highly viable seed production, is quite remarkable [8] [9], in contrast to the low seed setting registered for *S. eruca* [10], a species derived from *S. gummosus* [6].

Stenocereus gummosus is a columnar cactus with shrubby growth form of decumbent branching stems (**Figure 1(a)** and **Figure 1(b)**) [11], and it's notable as a co-dominant species in most plant communities in the Baja California Peninsula. The species is endemic to the Sonoran Desert [9]. It exhibits a disjunct distribution between Baja California Peninsula and a restricted coastal area of mainland Mexico, with presence in the Gulf islands [12]. Sphingidae moths and small beetles of the Nitidulidae and Cantharidae families mainly pollinate *S. gummosus* self-incompatible nocturnal flowers (**Figure 1(c)** and **Figure 1(d)**) [9] [13].

Previous allozyme evaluations found heterozygote deficiency under Hardy-Weinberg expectation, for populations throughout the distribution of *S. gummosus* [13], with mating among relatives, genetic subdivision, homozygote advantage, clonal structure, and geographic variations in the mating system, were proposed as possible causes for large inbreeding coefficient (F_{IS}) levels; and of these hypotheses, only the clonal structure have been directly evaluated. The proportion of clonal growth was determined as less frequent than the one from sexual recruitment [14], with R ranging from 0.633 to 0.959 (R estimated according to Dorken and Eckert [15]). This spatial distribution of identical genotypes was only explored for the population with the highest proportion of unique genotypes, where the distance between ramets was less than 9 meters, describing a clumped distribution of the clonal units, according to Charpentier [16].

Nonetheless, it has been pointed out that the clonal architecture may have strong impact on geitonogamy (pollination among flowers of the same plant either from the same or different ramet) [4] [16], a characteristic that is difficult to discriminate from outcrossing events among relatives [17]. Both of these reproductive events may affect F_{IS} and interpretations about genetic population structure.

When clumped clonal architecture occurs for species with self-incompatible flowers, a sampling strategy can be designed to avoid collecting ramets of the same genetic individual, taking advantage of multilocus genotype (MLG) and clonal patch extension. With this strategy, the effect of geitonogamy can be excluded from F_{IS}, which may result exclusively from the effect of outcrossing among relatives, provided that the species is self-incompatible and that the distance among the samples collected exceeds the genetic neighborhood size driven by pollinator's dispersal limitation [4] [7] [18] [19].

Besides *S. gummosus*, other columnar cacti's clonal identification has been accomplished through allozyme and RAPD markers (e.g. *Lophocereus schottii*, *Ferocactus robustus*, *S. eruca*) [4] [7] [20]. These markers have been criticized because of their low resolution for identifying clonal membership, having low polymorphism as a compromising characteristic [2]. Parra *et al.* [21] determined the clonal identity of *Stenocereus pruinosus* individuals, using 4 cross amplification microsatellite markers originally described for *Polaskia chichipe* [22]. Al-

though these markers may be adequate for the evaluation of clonal identity, their available number and its allelic diversity may not provide enough statistical power to that end [23]. This issue could be solved by designing species or genus specific microsatellite markers.

The focus of this study is to isolate and describe specific hypervariable neutral markers and to test its statistical power for clonal membership analysis. This will provide the possibility to assess the causes of the high F_{IS}, which are previously found by allozyme markers in *Stenocereus gummosus*. These loci will be useful for further genetic diversity estimations, clonal group determination through MLG, and the measure of population structure in *S. gummosus* and other members of the genus.

2. Material and Methods

2.1. Plant Material and Sample DNA Preparation

A sample of 40 *S. gummosus* individuals from Bahía Sargento (29.212369˚N, 112.202189˚W), and 30 individuals from Piedra Blanca (28.229661˚N, 113.159389˚W) was collected for population characterization. The first locality was chosen nearby the one from previous analysis for spatial clonal distribution by Molina Freaner and Clark-Tapia [14], allocated on mainland Mexico; the second locality was intended to represent the Peninsula populations, which may be reproductively isolated from the Sonora ones since the Gulf of California permanently separates them. Each sampled individual was distanced on-site at least for 30 - 35 m from the rest, leaving the unsampled individuals in between. Assuming in a clumped clonal architecture, as found by Molina Freaner and Clark-Tapia [14], each individual sample might belong to a different genet. The collected individuals describe two inverted semicircles to opposite sides of each coordinate previously specified. Sampled tissue fragments were taken from the stem chlorenquima to be stored in 96˚ ethanol on 2 mL tubes, and deposited at the tissue collection of the Laboratorio de Genética para la Conservación (CIBNOR, La Paz, BCS, Mexico). The source specimens were left living on-site, for which no voucher specimen was collected.

Genomic DNA was extracted using an optimized protocol of de la Cruz *et al*. [24]. A sample of 50 ng of RNA-free, high quality DNA from an individual collected at El Comitán (24.135385˚N, 110.42504˚W); voucher specimen HCIB27832 deposited at the Herbarium of the CIBNOR (La Paz, Baja California Sur, Mexico) was

used for shotgun sequencing by using a 454 GS-FLX Titanium instrument (Roche Applied Sciences, Indianapolis, Indiana, USA) at the UCSC Genome Sequencing Center (Santa Cruz, California, USA); other seven barcoded species were included in the same plate.

2.2. Microsatellites Identification and Primer Design

Msatcommander [25] was used to filter sequences among 454 reads containing perfect microsatellites with 2 - 6 bp repeat units, counting at least five repeats for di-nucleotides and four for tri-, tetra-, penta-, and exa-nucleotides. Primers were designed for filtered microsatellite hit results using Primer3 software [26] embedded in QDD [27] when hits fulfilled the following parameters: (1) 50˚C - 70˚C melting temperatures with a 2˚C maximum difference between paired primers; (2) PCR product of 90 - 320 bp in length; (3) GC content > 40%; and (4) A primer length between 17 and 27 nucleotides. Default parameter values were maintained in Primer3 for primer self-complementarities and complement between pairs. Favoring larger motifs (repetition numbers) and lower melting temperature differs between primers [28], 40 primer pairs were chosen for PCR amplification screening and polymorphism test. The names for the loci were designated with the prefix "Sgum" and a consecutive number (1 - 40). Polymorphic loci are listed in **Table 1**, describing microsatellite motifs, PCR conditions, and fluorescent dye-labels (Applied Biosystems, Foster City, California, EUA) incorporated to forward primer sequences.

2.3. PRC Procedures

PCR reactions were carried out in 10 μL volume, using an iCycler thermocycler (BIORAD Laboratories, Hercules, California, USA), containing approximately 45 ng of genomic DNA, $1 \times$ PCR buffer (20 mM Tris-HCl, pH 8.4, 50 mM KCl), 0.25 mM of each dNTP, 0.4 μM of each primer, 100 mM of BSA (New England Biolabs, Beverly, Massachusetts, USA), and 0.3 - 0.4 units of *Taq*DNA polymerase (Invitrogen, Carlsbad, California, USA), testing 1.5, 2.0 and 2.5 mM concentrations of $MgCl_2$; reactions for Sgum24 also included 5% glycerol. Forward primers for polymorphic loci were fluorescent-labeled (**Table 1**). Cycling conditions were conducted as

Table 1. Summary data for 9 polymorphic microsatellite loci amplified by PCR for *Stenocereus gummosus*: forward (F) and reverse (R) primer pairs, fluorescent dye-labeled forward primer, product and repeat sizes as found on 454 reads, and repeat motif. The allelic range found (Range), optimized magnesium chloride concentration ($MgCl_2$), and annealing temperatures (Tm) are indicated for each locus.

GenBank accession no.	Locus	PRC primers sequence	Product size [repeat size] (bp)	Range (bp)	Repeat motif	$MgCl_2$ (mM)	T_m (˚C)	Dye[a]
JX512845	Sgum05	F: TCGTAAAACGAATTGAGGGC R: GCTTTTAGCGAGGATTCCCT	122 [28]	111 - 119	$(GGGA)_1 (TGGA)_5 (CGGA)_1$	1.5	58	NED
JX512846	Sgum06	F: ATGAGATCGTGTGGGTCCAA R: ACAGAATCTCATGTGGCAGC	90 [24]	74 - 92	$(AG)_{12}$	2.0	61	NED
JX512847	Sgum12	F: CAGGTTCAGGGGTGGAGTAA R: GGATGCGGCTATAGGAGTGT	110 [10]	105 - 107	$(AC)_5$	2.0	58	VIC
JX567729	Sgum24	F: CCCCAAAGCGGACAATACC R: TGACATACATGTTGGGCACTT	130 [60]	120 - 130	$(AC)_1 (AT)_5 (AC)_{13} [(AT)_2 GT]_3 (AT)_2$	1.5	55	6FAM
JX512848	Sgum25	F: ATTTTGTTTTATATCTGTGTTCATTC R: AGACGCTTCAATTGTAACAAGG	101 [20]	97 - 99	$(GT)_{10}$	2.5	55	6FAM
JX512849	Sgum29	F: CGTCTCATTTGCTGTGCTCAGAT R: CTGTGTGGGGACGAGAGCAG	96 [25]	80 - 100	$(ACCAG)_5$	2.5	64	VIC
JX512850	Sgum31	F: TAACGCGCCGTCTTCAATTC R: TCCTTGCTTTAGCAGGATTCCC	142 [34]	134 - 142	$(GGAT)_5 (GGAC)_1 (GGAT)_1 GG (GGAC)_1$	2.0	64	PET
JX512851	Sgum36	F: AGCACATAACATATGGAACACGACA R: AAGTTAGTCCAATAACCCATCTCA	142 [26]	134 - 156	$(CA)_1 T (CA)_2 A (CA)_9$	2.0	58	NED
JX512852	Sgum39	F: CCTGCTGGCCACTTCATGTC R: ATCGCTCAAATCGACCACCA	111 [18]	106 - 108	$(GA)_9$	2.5	55	6FAM

a. Fluorescent dye-labels (Applied Biosystems) on forward primers: NED (yellow), VIC (green), 6FAM (blue) and PET (red).

follows: 95˚C for 5 min as initial denaturation, 35 cycles at 94˚C for 60 s each, 35 s at the locus-specific anneal-ing temperature (**Table 1**), extension at 72˚C for 15 s, and a final extension at 72˚C for 15 min. Amplification products were visualized in ABI 310-3730 PRISM® (Life Technologies, Carlsbad, California, USA), using LIZ600 as size standard (Applied Biosystems).

2.4. Data Analysis

Allele assignation was automated with GeneMapper 4.1 (Applied Biosystems). The minimum peak height ac-ceptance was set to 100 Relative Fluorescence Units (RFU), and the minimum height ratio between peaks for heterozygotes to one third. All peaks assigned were manually checked for allele shape (**Figure 2**).

To know if the set of loci genotyped provides enough power to discriminate all MLG present in the sample, a Monte Carlo procedure, resampling 5000 permutations, incrementing loci number for each locality was imple-mented in GENCLONE 2.0 [23] [29], and its corresponding results were plotted using the R package ggplot2 [30]. Clones were identified among samples by analyzing the MLG composition, calculating the genotypic richness index (R) and identifying one-allele difference MLG. Both identical MLG and one-allele difference MLG (23 individuals) were confirmed by PCR re-amplification and visualization as described above to reduce the probability of scoring errors. Of the individual genotypes, 10% were reassessed starting from DNA extrac-tion to approach an error rate. The probability of identity (p_{id}) trough all loci (average probability that two unre-lated individuals, drawn from the same randomly mating population, will by chance have the same multilocus genotype) was estimated using GenAlEx 6.5 [31]. Repeated MLG samples were assumed to correspond to ra-mets from the same individual, thus excluding repetitions for the following estimations.

Presence of null alleles, scoring errors due to stuttering and allelic drop out was explored implementing MICRO-CHEKER software [32]. Frequency for null alleles was computed through 5000 bootstrap resampling steps over loci for global fixation index (F_{ST}) values by the Expectation Maximization algorithm [33] with FreeNA software [34]. Effective alleles number (A_E) and inbreeding coefficient (F_{IS}) were estimated with standard me-thods implemented in GenAlEx 6.5 [31]. Expected and observed heterozygosity (H_E and H_O, respectively), dev-iations from the Hardy-Weinberg equilibrium, and linkage disequilibrium between markers, were calculated or tested using default parameters in GENEPOP 4.2 [35]. Benjamini & Hotchberg [36] False Discovery Rate cor-rection was used for multiple hypothesis tests.

3. Results and Discussion

3.1. Microsatellite Identification, Primer Design and PCR Screening

Genomic sequencing has been chosen over other techniques for microsatellite markers development due to the high sequencing throughput and the possibility of selecting microsatellites and motif length. The resulting num-ber of markers and the polymorphic loci proportion are expected to be higher with this technique [37]. Unfortu-nately, no previous information of any microsatellite development technique for neither the *Stenocereus gum-mosus* species nor genus is available to compare against the results reported here.

The 454 pyrosequence reaction resulted in 42,554 reads with a mean length of 311.11 nucleotides, where 1539 (3.6%) contained a microsatellite motif, regardless of their suitability for primer design. A total of 179 reads (11.6% of microsatellite loci found) were suitable for primer design. Out of the 40 chosen loci for PCR and polymorphism screening, 28 were monomorphic, 10 were polymorphic, and 2 exhibited inconsistent ampli-fication patterns. Only polymorphic loci with more than 50% of the genotypes successfully recorded (**Table 1**) were used for population characterization. The polymorphic loci proportion (30%) was greater than reported for genomic libraries (15% and 21%, **Table 2**), but lower than enriched genomic libraries (53%, 91% and 70%, **Table 2**) and cross amplification (57%, **Table 2**); these differences may rely on the species reproduction system, demographic issues, mutation rates on microsatellite loci, specific genomic composition, or even on the markers isolation technique implemented [37] [38].

3.2. Population Characterization

Micro-Checker software [32] detected homozygote excess in 4 loci for Piedra Blanca locality (Sgum05, Sgum12, Sgum31 and Sgum36) and in 3 loci for Bahía Sargento (Sgum05, Sgum31 and Sgum36), scoring error due to stuttering were only predicted to happen in both localities for Sgum05 and Sgum31, and no signal of allelic
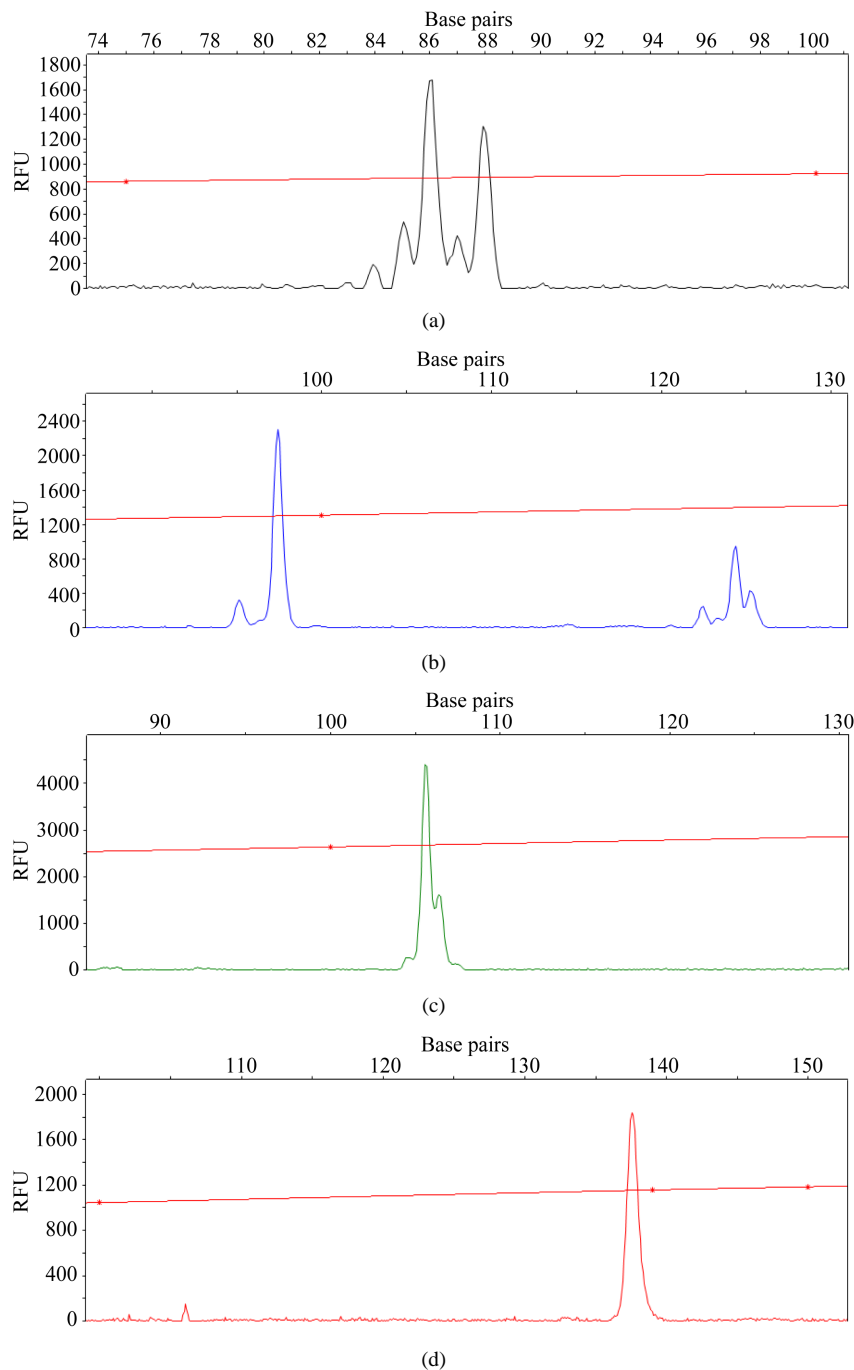
Figure 2. Electropherogram examples as ploted by GeneMapper 4.1 (Applied Biosystems). The red stared horizontal line stands for the size standard assignation. The *x* and *y* axex corresponds to base pair length for the PCR products and the Relative Florescence Units (RFU), respectively. (a) NED dyed Sgum06 (black line) heterocygote for 86 and 88 bases; (b) 6FAM dyed Sgum25 (left) and Sgum24 (right) (blue line) homocygotes for 97 and 124 bases, respectively; (c) VIC dyed Sgum12 (green line) homocygote for 105 bases; (d) PET dyed Sgum31 (red bottom line) homocygote for 138 bases.

dropout was detected. The null alleles frequencies calculated for two loci within Piedra Blanca (0.33 for Sgum05 and 0.27 for Sgum31), and one for Bahía Sargento (0.21 for Sgum05), are likely to hinder population differentiation analysis with F_{ST}, and may overestimate F_{IS} [34]; the rest of the calculated frequencies were left

Table 2. Number and percentage of polymorphic loci (Polym.), genetic diversity measured as mean allele number per locus (A), observed (H_O) and expected (H_E) heterozygosity, ± its respective Standard Error (SE) of six characterizations (a, b, c, d, e and *S. gummosus*) and one population structure (f) studies, all them using microsatellite loci on Cactaceae species. Species, tribe, geographic region where the study has been carried out, the technique use for marker isolation, the number of populations analyzed (Populations), and the sample size (Sample) are specified.

Species	Tribe	Geographic region	Technique	Populations	Sample	Polym.	$A \pm SE$	$H_O \pm SE$	$H_E \pm SE$
Ariocarpus bravoanus[a]	Thelocactineae	S. L. Potosi, Mexico	Genomic Lib.	1	20	8 (15%)	5.60 ± 0.39	0.575 ± 0.038	0.512 ± 0.025
Echinocactus grusonii[b]	Cacteae	S. L. Potosi, Mexico	Genomic Lib.	1	30	12 (21%)	3.33 ± 0.45	0.550 ± 0.039	0.509 ± 0.029
Mammilaria crucigera[c]	Cacteae	Tehuacán, Mexico	Enriched Lib.	2	40	8 (53%)	8.88 ± 1.18	0.320 ± 0.035	0.759 ± 0.025
Pilosocereus machrisii[d]	Cacteae	Caatinga, Brazil	Enriched Lib.	2	51	10 (91%)	4.41 ± 0.66	0.466 ± 0.034	0.561 ± 0.033
Polaskia chichipe[e]	Pachycereeae	Tehuacán, Mexico	Enriched Lib.	1	45	7 (70%)	4.57 ± 0.70	0.417 ± 0.032	0.501 ± 0.038
Stenocereus pruinosus[f]	Pachycereeae	Tehuacán, Mexico	Cross Amp.	3	61	4 (57%)	4.75 ± 0.61	0.670 ± 0.013	0.717 ± 0.004
Stenocereus gummosus	Pachycereeae	Sonora, Mexico	Genomic Seq.	2	70	9 (30%)	3.78 ± 0.62	0.228 ± 0.023	0.357 ± 0.022

a Characterized samples from Hughes *et al*. [42]; b from Hardesty *et al*. [41]; c from Solórzano *et al*. [44]; d from Pérez *et al*. [43]; e from Otero-Arnaiz *et al*. [22]; f wild populations from Parra *et al*. [21].

aside since they were very small (mean frequency of 0.069 ± 0.02 S.E.). When PCR amplification and product visualization were repeated, 5 alleles assignment were modified for the loci Sgum05, Sgum06 and Sgum24 (error ratio of 0.018 accounting for 270 PCR reactions repeated); these scoring errors were always identified as mistakes at the electrophoresis peak reading for allele size assignment, which allowed improving the automated analysis. An increment on sample size [39], or a PCR primers redesign based on individuals' sequences for those loci [40], could be useful to avoid bias due to null alleles on further population genetics studies. In this case, for which the goal was to determine clonal membership, achieving precision and reproducibility on genotyping lets know that the loci are useful, provided that null alleles and stuttering behave evenly at the population level. Significant linkage disequilibrium was detected between Sgum05 and Sgum31 when analyzed globally over both localities individuals. This condition persisted only in Piedra Blanca when sample was analyzed by locality.

The number of alleles per locus (A) among polymorphic loci ranged from 2 - 7 (**Table 3**), H_O ranged from 0.000 (Sgum25 in Piedra Blanca, and Sgum05 in Bahía Sargento) to the maximums of 0.552 and 0.556 (Sgum06), and H_E from 0.034 (Sgum39) and 0.055 (Sgum25) to 0.673 (Sgum36) and 0.557 (Sgum06) (**Table 3**), with mean values of 0.221 and 0.234 for H_O and 0.408 and 0.306 for H_E, in Piedra Blanca and Bahía Sargento respectively. The mean F_{ST} was 0.082 ± 0.39 S.E. Significant deviations from Hardy-Weinberg Equilibrium (HWE) were detected in 4 loci in Piedra Blanca (Sgum05, Sgum12, Sgum31 and Sgum36), and three in Bahía Sargento (Sgum05, Sgum31 and Sgum36); deviations are due to heterozygote deficiency in all cases (**Table 3**), but this departure from HWE may as well be null alleles artifact, as predicted by Micro-Checker software.

3.3. Clonal Membership and Geitonogamy

Both diversity indexes, mean allele number for polymorphic loci ($A = 3.78 \pm 0.062$ SE), and expected and observed heterozygosity ($H_O = 0.228 \pm 0.047$ SE, $H_E = 0.357 \pm 0.048$ SE), reached the lowest value when compared to other cacti (**Table 2**), having mean values (±SD) of $A = 5.05 \pm 1.84$, $H_O = 0.461 \pm 0.153$, and $H_E = 0.559 \pm 0.138$ [21] [22] [41]-[44]. These low diversity indexes are intuitively explained in terms of inbreeding coefficient on a low effective population size (N_e) scenario [45], coherent to the F_{IS} of 0.608 ± 0.025 SE previously reported [13], albeit inconsistent with the Molina Freaner and Clark-Tapia [14] conclusion, where sexual recruitment is found as more frequent than clonal growth. The average F_{IS} here reported (0.341 ± 0.122 SE) may be less biased by selection than allozyme results [46], but null alleles can still be artificially increasing it in a small magnitude in respect to the actual inbreeding [47]. Though attenuated, the inbreeding coefficient is yet remarkable, so that reproduction system

Table 3. Total number of samples analyzed (N) and number of alleles per locus (A) found through both populations are listed first; population sample size (n), effective alleles number (A_E), inbreeding coefficient (F_{IS}), observed heterozygosity (H_O) and expected heterozygosity under Hardy-Weinberg Equilibrium (H_E) at Piedra Blanca and Bahía Sargento are specified. The H_E and F_{IS} estimations do not apply for loci with only 1 allele, depicted with (–).

Locus	N	A	Piedra Blanca				Bahía Sargento			
			n	A_E	F_{IS}	H_O / H_E[a]	n	A_E	F_{IS}	H_O / H_E[a]
Sgum05	65	3	29	2.176	0.938	**0.034 | 0.550**	36	1.246	1.000	**0.000 | 0.200**
Sgum06	65	7	29	2.424	0.078	0.552 | 0.598	36	2.219	0.003	0.556 | 0.557
Sgum12	62	2	27	1.957	0.633	**0.185 | 0.498**	35	1.471	−0.057	0.343 | 0.325
Sgum24	64	4	29	1.811	0.093	0.414 | 0.456	35	1.866	−0.156	0.543 | 0.471
Sgum25	65	2	29	1.000	–	0.0 | –	36	1.057	−0.015	0.056 | 0.055
Sgum29	65	4	29	1.597	0.095	0.345 | 0.381	36	1.405	0.242	0.222 | 0.292
Sgum31	61	3	25	1.903	0.838	**0.080 | 0.484**	36	1.213	0.846	**0.028 | 0.178**
Sgum36	65	7	29	2.951	0.492	**0.345 | 0.673**	36	1.885	0.479	**0.250 | 0.476**
Sgum39	65	2	29	1.035	−0.018	0.034 | 0.034	36	1.246	0.449	0.111 | 0.200
Mean over all loci	64.1	3.78	28.3	1.873	0.383	0.221 | 0.408	35.8	1.512	0.299	0.234 | 0.306

a. Significant heterozygote deficiency after Benjamini and Hotchberg [36]. False Discovery Rate correction ($p < 0.05$) is indicated in bold numbers.

aspects like pollination syndrome, mating among relatives and the possibility of geitonogamy, cannot be discarded as causes for these low diversity indices.

The MLG resampling procedure shown that 7 loci were enough to determine clonal membership for Piedra Blanca locality (**Figure 3(a)**), but 8 or 9 loci might be required in the case of Bahía Sargento (**Figure 3(b)**); the p_{id} values of each locality accounting the 9 loci were 2×10^{-4} and 2.6×10^{-3}, respectively. Piedra Blanca accounted 29 unique genotypes with $R = 0.966$, and 36 unique genotypes were detected for Bahía Sargento with $R = 0.897$. These values slightly increase the ones found by Molina Freaner and Clark Tapia [14] with allozyme markers, where the maximum was $R = 0.959$ for a locality nearby Bahía Sargento from this study. One-allele difference MLG within locality were found only for Bahía Sargento, with 12 individuals into at least 2 putative multilocus lineages (clonal ramets carrying a somatic mutation identified as one-allele difference) [23], where 2 of the 3 repeated MLG of that locality are members. This finding provides a scenario that is supposed to be eliminated by collecting individuals distanced by at least 30 meters, in which ramets are now somehow able to spread away from the source plant, farther than it has been previously found [13].

Identical MLG of *Lophocereus schottii* individuals, distanced in field by 200 or more meters, have been defined as long-distance transportation events, where detached stems are dispersed downstream by floodwaters [7]. This clone dispersion mechanism is likely to occur along the distribution range of *S. gummosus* (shared with *L. schotti*), species mostly found on alluvial, arroyo margins and rocky hillsides [8] [12]. The latter results in an intermingled distribution of the clones and a subsequent increase of the genetic neighborhood [7], where ramets disperse not only to the space surrounding the source specimen but, in this case, also for at least 30 meters away. Intermingled clonal architecture is a characteristic that gives a clonal plant the capability to spread more quickly [48]. It has been proposed that *L. schottii* and *S. gummosus* populations along the Baja California Peninsula have occupied their current distribution range through recent colonization events, reaching such large areas [13] [49].

This scenario implies that no discrimination can be made about the effect of geitonogamy from that of endogamy, because the genetic neighborhood size was not exceeded by the sampling strategy followed [17] [19]. Though the impact of geitonogamy is expected to be less severe as clones dispersion capability increases on insect pollinated plants [16], it may increment the probability of mating among relatives as it increases neighborhood size.

A population analysis of wider geographical range, employing these markers, might be useful to enhance accuracy for linkage disequilibrium tests. These 9 isolated microsatellite loci are the first identified and characte-
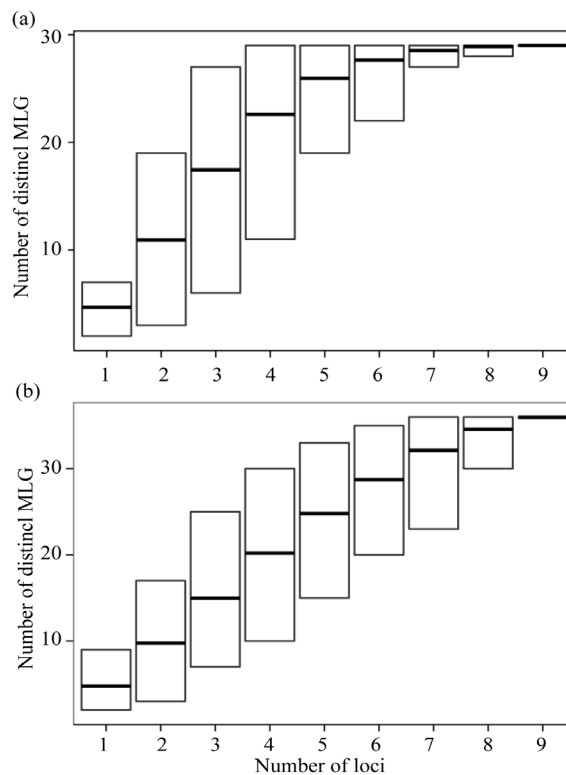
Figure 3. Boxplot describing the genotypic resolution for the 9 loci characterized in each population data set: (a) Piedra Blanca; and (b) Bahía Sargento. The box edges indicate the minimum and the maximum, while the inner line shows the mean number of distinct MLG found for *x* loci.

rized ones for the genus *Stenocereus*. The availability of these molecular tools allows assessing questions integrating population genetics, ecology, conservation, and reproductive biology for *S. gummosus* and for other species phylogenetically related.

Aknowledgements

References

[1] Mandujano, M.C. (2007) La clonalidad y sus efectos en la biología de poblaciones. In: Eguiarte, L.E., Sousa, V. and Aguirre, X., Eds., *Ecología Molecular*, Instituto Nacional de Ecología, SEMARNAT, México Distrito Federal, México, 215-250.

[2] Arnaud-Haond, S., Alberto, F., Teixeira, S., Procaccini, G., Serrão, E.A. and Duarte, C.M. (2005) Assessing Genetic Diversity in Clonal Organisms: Low Diversity or Low Resolution? Combining Power and Cost Efficiency in Selecting Markers. *Journal of Heredity*, **96**, 434-440. http://dx.doi.org/10.1093/jhered/esi043

[3] Pimienta-Barrios, E. and del Castillo, R.F. (2002) Reproductive Biology. In: Nobel, P.S., Ed., *Cacti*: *Biology and Uses*, University of California Press, Berkeley, 75-90.

[4] Carrillo-Angeles, I.G., Mandujano, M.C. and Golubov, J. (2011) Influences of the Genetic Neighborhood on Ramet Reproductive Success in a Clonal Desert Cactus. *Population Ecology*, **53**, 449-458. http://dx.doi.org/10.1007/s10144-010-0254-1

[5] Dubrovsky, J.G. and North, G.B. (2002) Root Structure and Function. In: Nobel, P.S., Ed., *Cacti*: *Biology and Uses*,

University of California Press, Berkeley, 41-56.

[6] Gibson, A.C. (1989) The Systematics and Evolution of the Subtribe Stenocereinae 7. The Macharocerei of *Stenocereus*. *Cactus and Succulent Journal*, **61**, 104-112.

[7] Parker, K.C. and Hamrick, J.L. (1992) Genetic Diversity and Clonal Structure in a Columnar Cactus, *Lophocereus schottii*. *American Journal of Botany*, **79**, 86-96. http://dx.doi.org/10.2307/2445201

[8] León de La Luz, J.L., Domínguez-Cadena, R., Cruz-Estrada, M. and Rodríguez-Estrella, R. (1995) Reproductive Phenology of *Stenocereus gummosus* (Engelm.) Gibson & Horak. Implications for Its Cultivation. *Genetic Resources and Crop Evolution*, **42**, 61-67. http://dx.doi.org/10.1007/BF02310685

[9] León de La Luz, J.L. and Domínguez-Cadena, R. (1991) Evaluación de la reproducción por semilla de la Pitahaya agria (*Stenocereus gummosus*) en Baja California Sur, México. *Acta Botánica Mexicana*, **14**, 75-87.

[10] Clark-Tapia, R., Mandujano, M.C., Valverde, T., Mendoza, A. and Molina-Freaner, F. (2005) How Important Is Clonal Recruitment for Population Maintenance in Rare Plant Species?: The Case of the Narrow Endemic Cactus, *Stenocereus eruca*, in Baja California, México. *Biological Conservation*, **124**, 123-132. http://dx.doi.org/10.1016/j.biocon.2005.01.019

[11] Vázquez-Sánchez, M., Terrazas, T. and Arias, S. (2012) El hábito y la forma de crecimiento en la tribu Cacteae (Cactaceae, Cactoideae). *Botanical Sciences*, **2**, 97-108.

[12] Turner, R.M., Bowers, J.E. and Burgess, T.L. (1995) Sonoran Desert Plants: An Ecological Atlas. University of Arizona Press, Tucson, 504.

[13] Clark-Tapia, R. and Molina-Freaner, F. (2003) The Genetic Structure of a Columnar Cactus with a Disjunct Distribution: *Stenocereus gummosus* in the Sonoran Desert. *Heredity*, **90**, 443-450. http://dx.doi.org/10.1038/sj.hdy.6800252

[14] Molina Freaner, F. and Clark-Tapia, R. (2005) Clonal Diversity and Allelic Relationships between Two Closely Related Species of Columnar Cacti from the Sonoran Desert: *Stenocereus eruca* and *Stenocereus gummosus*. *International Journal of Plant Sciences*, **166**, 257-264. http://dx.doi.org/10.1086/427486

[15] Dorken, M.E. and Eckert, C.G. (2001) Severely Reduced Sexual Reproduction in Northern Populations of a Clonal Plant, *Decodon verticillatus* (Lythraceae). *Journal of Ecology*, **89**, 339-350. http://dx.doi.org/10.1046/j.1365-2745.2001.00558.x

[16] Charpentier, A. (2002) Consequences of Clonal Growth for Plant Mating. *Evolutionary Ecology*, **15**, 521-530. http://dx.doi.org/10.1023/A:1016057503722

[17] Eckert, C.G. (2000) Contributions of Autogamy and Geitonogamy to Self-Fertilization in a Mass-Flowering, Clonal Plant. *Ecology*, **81**, 532-542.

[18] Carrillo-Angeles, I.G. and Mandujano, M. del C. (2011) Patrones de distribución espacial en plantas clonales. *Boletín de la Sociedad Botánica de México*, **89**, 1-18.

[19] Loveless, M.D. and Hamrick, J.L. (1984) Ecological Determinants of Genetic Structure in Plant Populations. *Annual Review of Ecology and Systematics*, **15**, 65-95. http://dx.doi.org/10.1146/annurev.es.15.110184.000433

[20] Clark-Tapia, R., Alfonso-Corrado, C., Eguiarte, L.E. and Molina-Freaner, F. (2005) Clonal Diversity and Distribution in *Stenocereus eruca* (Cactaceae), a Narrow Endemic Cactus of the Sonoran Desert. *American Journal of Botany*, **92**, 272-278. http://dx.doi.org/10.3732/ajb.92.2.272

[21] Parra, F., Casas, A., Peñaloza-Ramírez, J.M., Cortés-Palomec, A.C., Rocha-Ramírez, V. and González-Rodríguez, A. (2010) Evolution under Domestication: Ongoing Artificial Selection and Divergence of Wild and Managed *Stenocereus pruinosus* (Cactaceae) Populations in the Tehuacan Valley, Mexico. *Annals of Botany*, **106**, 483-496. http://dx.doi.org/10.1093/aob/mcq143

[22] Otero-Arnaiz, A., Cruse-Sanders, J., Casas, A. and Hamrick, J.L. (2004) Isolation and Characterization of Microsatellites in the Columnar Cactus: *Polaskia chichipe* and Cross-Species Amplification within the Tribe Pachycereeae (Cactaceae). *Molecular Ecology Notes*, **4**, 265-267. http://dx.doi.org/10.1111/j.1471-8286.2004.00642.x

[23] Arnaud-Haond, S., Duarte, C.M., Alberto, F. and Serrão, E.A. (2007) Standardizing Methods to Address Clonality in Population Studies. *Molecular Ecology*, **16**, 5115-5139. http://dx.doi.org/10.1111/j.1365-294X.2007.03535.x

[24] de la Cruz, M., Ramirez, F. and Hernandez, H. (1997) DNA Isolation and Amplification from Cacti. *Plant Molecular Biology Reporter*, **15**, 319-325. http://dx.doi.org/10.1023/A:1007428818078

[25] Faircloth, B.C. (2008) Msatcommander: Detection of Microsatellite Repeat Arrays and Automated, Locus-Specific Primer Design. *Molecular Ecology Resources*, **8**, 92-94. http://dx.doi.org/10.1111/j.1471-8286.2007.01884.x

[26] Rozen, S. and Skaletsky, H. (2000) Primer3 on the WWW for General Users and for Biologist Programmers. *Methods in Molecular Biology*, **132**, 365-386.

[27] Meglécz, E., Costedoat, C., Dubut, V., Gilles, A., Malausa, T., Pech, N. and Martin, J.-F. (2010) QDD: A User-Friendly Program to Select Microsatellite Markers and Design Primers from Large Sequencing Projects. *Bioinformatics*, **26**, 403-404. http://dx.doi.org/10.1093/bioinformatics/btp670

[28] Gardner, M.G., Fitch, A.J., Bertozzi, T. and Lowe, A.J. (2011) Rise of the Machines—Recommendations for Ecologists When Using Next Generation Sequencing for Microsatellite Development. *Molecular Ecology Resources*, **11**, 1093-1101. http://dx.doi.org/10.1111/j.1755-0998.2011.03037.x

[29] Arnaud-Haond, S. and Belkhir, K. (2007) Geneclone: A Computer Program to Analyse Genotypic Data, Test for Clonality and Describe Spatial Clonal Organization. *Molecular Ecology Notes*, **7**, 15-17. http://dx.doi.org/10.1111/j.1471-8286.2006.01522.x

[30] Wickham, H. (2009) ggplot2: Elegant Graphics for Data Analysis. Springer, New York, 202.

[31] Peakall, R. and Smouse, P.E. (2012) GenAlEx 6.5: Genetic Analysis in Excel. Population Genetic Software for Teaching and Research—An Update. *Bioinformatics*, **28**, 2537-2539. http://dx.doi.org/10.1093/bioinformatics/bts460

[32] Van Oosterhout, C., Hutchinson, W.F., Wills, D.P.M. and Shipley, P. (2004) MICRO-CHECKER: Software for Identifying and Correcting Genotyping Errors in Microsatellite Data. *Molecular Ecology Notes*, **4**, 535-538. http://dx.doi.org/10.1111/j.1471-8286.2004.00684.x

[33] Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977) Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society*, **39**, 1-38.

[34] Chapuis, M.-P. and Estoup, A. (2007) Microsatellite Null Alleles and Estimation of Population Differentiation. *Molecular Biology and Evolution*, **24**, 621-631. http://dx.doi.org/10.1093/molbev/msl191

[35] Rousset, F. (2008) GENEPOP'007: A Complete Re-Implementation of the Genepop Software for Windows and Linux. *Molecular Ecology Resources*, **8**, 103-106. http://dx.doi.org/10.1111/j.1471-8286.2007.01931.x.

[36] Benjamini, Y. and Hochberg, Y. (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society*, *Series B*, **57**, 289-300.

[37] Abdelkrim, J., Robertson, B., Stanton, J.-A. and Gemmell, N. (2009) Fast, Cost-Effective Development of Species-Specific Microsatellite Markers by Genomic Sequencing. *BioTechniques*, **46**, 185-192. http://dx.doi.org/10.2144/000113084

[38] Ellegren, H. (2004) Microsatellites: Simple Sequences with Complex Evolution. *Nature Reviews Genetics*, **5**, 435-445. http://dx.doi.org/10.1038/nrg1348

[39] Bonin, A., Bellemain, E., Bronken Eidesen, P., Pompanon, F., Brochmann, C. and Taberlet, P. (2004) How to Track and Assess Genotyping Errors in Population Genetics Studies. *Molecular Ecology*, **13**, 3261-3273. http://dx.doi.org/10.1111/j.1365-294X.2004.02346.x

[40] Shaw, P.W., Pierce, G.J. and Boyle, P.R. (1999) Subtle Population Structuring within a Highly Vagile Marine Invertebrate, the Veined Squid *Loligo forbesi*, Demonstrated with Microsatellite DNA Markers. *Molecular Ecology*, **8**, 407-417. http://dx.doi.org/10.1046/j.1365-294X.1999.00588.x

[41] Hardesty, B.D., Hughes, S.L., Rodriguez, V.M. and Hawkins, J.A. (2008) Characterization of Microsatellite Loci for the Endangered Cactus *Echinocactus grusonii*, and Their Cross-Species Utilization. *Molecular Ecology Resources*, **8**, 164-167. http://dx.doi.org/10.1111/j.1471-8286.2007.01913.x

[42] Hughes, S.L., Rodriguez, V.M., Hardesty, B.D., Bárcenas Luna, R.T., Hernández, H.M., Robson, R.M. and Hawkins, J.A. (2008) Characterization of Microsatellite Loci for the Critically Endangered Cactus *Ariocarpus bravoanus*. *Molecular Ecology Resources*, **8**, 1068-1070. http://dx.doi.org/10.1111/j.1755-0998.2008.02157.x

[43] Pérez, M.F., Téo, M.F., Zappi, D.C., Taylor, N.P. and Moraes, E.M. (2011) Isolation, Characterization, and Cross-Species Amplification of Polymorphic Microsatellite Markers for *Pilosocereus machrisii* (Cactaceae). *American Journal of Botany*, **98**, e204-e206. http://dx.doi.org/10.3732/ajb.1100033

[44] Solórzano, S., Cortés-Palomec, A.C., Ibarra, A., Dávila, P. and Oyama, K. (2009) Isolation, Characterization and Cross-Amplification of Polymorphic Microsatellite Loci in the Threatened Endemic *Mammillaria crucigera* (Cactaceae). *Molecular Ecology Resources*, **9**, 156-158. http://dx.doi.org/10.1111/j.1755-0998.2008.02422.x

[45] Freeland, J.R. (2006) Molecular Ecology. John Wiley & Sons Ltd., West Sussex, 400.

[46] Schlötterer, C. (2004) The Evolution of Molecular Markers—Just a Matter of Fashion? *Nature Reviews Genetics*, **5**, 63-69. http://dx.doi.org/10.1038/nrg1249

[47] De Meeûs, T., Lehmann, L. and Balloux, F. (2006) Molecular Epidemiology of Clonal Diploids: A Quick Overview and a Short DIY (Do It Yourself) Notice. *Infection*, *Genetics and Evolution*, **6**, 163-170. http://dx.doi.org/10.1016/j.meegid.2005.02.004

[48] Wong, S., Anand, M. and Bauch, C.T. (2011) Agent-Based Modelling of Clonal Plant Propagation across Space: Recapturing Fairy Rings, Power Laws and Other Phenomena. *Ecological Informatics*, **6**, 127-135. http://dx.doi.org/10.1016/j.ecoinf.2010.11.004

[49] Nason, J.D., Hamrick, J.L. and Fleming, T.H. (2002) Historical Vicariance and Postglacial Colonization Effects on the Evolution of Genetic Structure in *Lophocereus*, a Sonoran Desert Columnar Cactus. *Evolution*: *International Journal of Organic Evolution*, **56**, 2214-2226. http://dx.doi.org/10.1111/j.0014-3820.2002.tb00146.x