Full title: Nuclear and chloroplast DNA sequence variation in Yucca brevifolia (Joshua tree)

Short title: DNA sequence variation in Yucca brevifolia (Joshua tree)

Laura E. Vann[1,2], Amy Toulson-Wimmer, Laura A. Katz[2,3], Jeffrey Ross-Ibarra[1,4,5] , Robert B. Merritt[2],

1) Department of Plant Sciences, University of California-Davis, Davis, California 95616, USA

2) Department of Biological Sciences, Smith College, Northampton, Massachusetts 01063, USA

3) Department of Organismic and Evolutionary Biology, University of Massachusetts-Amherst, Amherst, Massachusetts 01002, USA

4) Genome Center, University of California-Davis, Davis, California 95616, USA

5) Center for Population Biology, University of California-Davis, Davis, California 95616, USA

Correspondence should be addressed to Jeffrey Ross-Ibarra (rossibarra@ucdavis.edu)

ABSTRACT

To understand further the population structure of the Mojave Desert endemic *Yucca brevifolia* (Agavaceae), we investigated DNA sequence polymorphism in a 4kb fragment of the gene for the cytosolic phosphoglucose isomerase (*PgiC*) and from a 300bp fragment within the chloroplast *trnD-trnT* spacer region. Consistent with previous studies, we find extensive shared ancestral variation in both nuclear and chloroplast DNA sequence, and the geographic distribution of diversity suggests that this species experienced range contraction and expansion following the Pleistocene glaciation. Additionally, sequence variation shows no clear pattern associated with PGI allozyme allelic class, chloroplast haplotype, or proposed subspecies.

INTRODUCTION

Yucca brevifolia (Agavaceae) is a plant species endemic to the Mojave Desert, inhabiting areas of Arizona, California, Nevada, and Utah. Commonly known as the Joshua tree, this visibly striking species -- known for its height and distinctive branching pattern -- has a patchy distribution across the Mojave Desert, with an elevation range of 500-2100m and a 20˚-100˚F isotherm range [1]. A gradual warming of temperatures since the last glacial minimum has likely contributed to the current patchy distribution of *Y. brevifolia*, and the discovery of Joshua tree fossils at lower elevations and latitudes outside of the current distribution suggests that the species once had a more continuous distribution throughout the Mojave and even the Sonoran Desert [2], [3]. Due to limited seed-dispersal, climatic warming, and habitat loss, populations of this species continue to shrink and recruitment of new individuals is rarely seen [4], [5], [6].

Previous studies suggest that fluctuations in population size of plant species typically leave discrete geographic patterns of variation in present day populations, with refugia represented by distinct haplotype groups [7], [8]. However, few studies have focused on species with discontinuous distributions, and investigation into the phylogeography of *Y. brevifolia* is thus a useful addition to understanding the evolutionary history and present day genetic structure of plants with discontinuous populations. Furthermore, trends observed in *Y. brevifolia* may represent general phylogeographic patterns of southwestern plant species, and will aid in understanding genetic consequences associated with population shifts in response to climate change. With respect to Joshua tree, a clearer understanding of phylogeography and population structure will be valuable

in future conservation and management efforts as populations of this species continue to shrink.

Fragmentation and range shift of populations of *Y. brevifolia* following the Pleistocene glaciation have likely affected the genetic structure of the species, and several studies have investigated genetic variation within *Y. brevifolia* in hopes of elucidating population structure and divergence [9]. Previous research investigating allozyme variation in the enzyme phosphoglucose isomerase (PGI) has shown that geographic barriers may permit differentiation in allozyme allele frequencies between some populations of Joshua trees [10], [9], while studies investigating chloroplast (cpDNA) variation suggest extensive shared ancestral variation [11].

Here, we examine DNA sequence data from a 4kb fragment of the nuclear *PgiC* gene as well as a 300bp fragment of the *trnD-trnT* spacer region in the chloroplast of *Y. brevifolia* to elucidate population structure and phylogeography of this species. We test the hypotheses that sequence relationships will reflect restricted gene flow as previously suggested [10], [9], and investigate whether the combined effects of geographic isolation, genetic drift, population expansion and contraction, and non-neutral forces have left signatures in the patterns of nucleotide variation in nuclear and chloroplast DNA sequence.

MATERIALS AND METHODS

Sampling

Samples used in this study were collected from one to three subpopulations (noted with letters A, B, and C) from a total of five populations across the species distribution; *Yucca brevifolia* ssp. *brevifolia*: Joshua Tree National Park (JT) and Western Mojave (WM),

and *Yucca brevifolia* ssp. *jaegeriana*: Mojave National Preserve (MP), Arizona (AZ), and Utah (UT)(Figure 1). For sampling details see Toulson and Merritt (2008) [9]. Samples from this study were previously investigated for glucose-6-phosphate isomerase allozyme (PGI) variation in Toulson and Merritt (2008)[9]. The cytosolic phosphoglucose isomerase (*PgiC*) gene codes for the PGI protein, and for *PgiC* sequencing we subsampled individuals from Toulson and Merritt (2008), looking at individuals with either a 3 or 4 PGI allozyme allele. For *trnD-trnT* sequencing we subsampled individuals from Toulson and Merritt (2008) at random. Outgroup samples were taken from either the Smith College Botanic Gardens (Northampton, MA), personal collections (L.Vann), or were collected from Joshua Tree National Park (Joshua Tree, CA). Including out-group samples, 86 and 108 individuals were used for *PgiC* and *trnD-trnT* sequencing, respectively. (Table S1).

DNA Extraction, Amplification and sequencing

We removed the top layer of tissue from the leaf, ground this tissue twice in liquid nitrogen, and attempted to remove all fibrous leaf matter prior to extraction. We then extracted total genomic DNA was using the Qiagen DNeasy Plant Mini Kit (Qiagen Inc., Valencia, California, USA) protocol with an extra ethanol precipitation step.

Initial primers used for *PgiC* amplification were designed from a previously isolated fragment of *PgiC* cDNA from *Y. brevifolia* that had been aligned and annotated for intron/exon boundaries using *Dioscorea PgiC* as a reference sequence [12]. DNA amplification was performed in 40 μL reactions using Phusion High Fidelity DNA Polymerase (Finnzymes, Inc., Woburn, MA) and gene specific primers 805X (5'-GGA TGG GTT GTT GGT CGC TAT AGT-3') and 1037X (5'-CCT GAG AGT ATG GCA GTA TGG CCA-3'), or

degenerate primer 93.4 from Terauchi *et al*. (1997) [12] and 1037X (Table S2). We visualized resulting amplification products on a 1.0% SeaKem agarose gel (Lonza Biosciences, Inc.), and then extracted and cleaned products using a Millipore Ultrafree-DA Centrifugal Unit (Millipore, Billerica, MA). Cleaned products were then cloned using the TOPO Zero Blunt Cloning Kit with chemically competent cells (Invitrogen, Carlsbad, California). In order to assess alleles for each individual, 8-12 clones per individual were miniprepped in Invitrogen's 96-well format following protocol B and three to five clones per individual were screened using gene specific primers 805X and 1037X and Phusion High Fidelity DNA Polymerase (40 μl reactions). We cleaned amplification products using Microclean (The Gel Company, San Francisco, CA) and sequenced them at the Nucleic Acid Facility at Penn State University to screen for quality and polymorphisms.

We initially chose two individuals from each of the five populations sampled to survey for chloroplast DNA sequence variation. We amplified and sequenced the following regions of the chloroplast: *trnH-trnF*, *aspH-atpI*, *trnD-trnT*, *petA-psbE*, *rsp18-clpp*, and *petD-petB.* Amplification reactions were performed in 15 μL reactions, using FastStart Taq DNA Polymerase (Roche Applied Science) and primers from Grivet *et al*. (2001) [13], with products cleaned using the QuickStep 2 PCR Purification Kit (Edge Biosystems). We then used a BigDye (Applied Biosystems) terminator sequencing reaction, and sequenced cleaned products on an ABI 3100 sequencer. Only ~1kb the intergenic spacer *trnD-trnT* showed any variation and was chosen for further sequencing. To specifically amplify the polymorphic sites within the *trnD-trnT* region, we designed a new 5' internal primer, *IntD*, (5'-GATAGGATTCCTTCAGTTCAGGC-3') located 715bp from the *trnD* (3') primer (Table S3). We then amplified and sequenced this region in 108 individuals, 26-47 per population

(Table S1) using the above protocol. Additionally, to determine the ancestral state in this region we amplified and sequenced homologous regions from related species of *Yucca* (Table S1).

Data analysis

We edited sequences manually using SeqMan (DNAStar, Inc.) and aligned them using Geneious Pro v.5.4.2 (Biomatters) for both *PgiC* and the *trnD-trnT* region. For *PgiC* alignments we created consensus sequences for each clone surveyed per individual (*e.g*. 8-12), and aligned these sequences. From these 8-12 sequences we identified either one or two alleles per individual based on sequence differences and exported these consensus sequences to Geneious Pro v.5.4.2 where we aligned them by eye and removed singleton SNPs and large unalignable regions of microsatellites. Because we did not consistently recover both alleles for all individuals within our sample set, we randomly chose one allele per individual for all further analyses (86 sequences in total).

For both nuclear and chloroplast sequence data we calculated standard diversity statistics including number of haplotypes; haplotype diversity; number of segregating sites, average pairwise difference between sequences, Watterson's estimator of the population mutation rate [14], and Tajima's D for each subspecies (ssp. *brevifolia* and ssp. *jaegeriana*), and for the sample set as a whole using programs from the Libsequence C++ library [15]. For *PgiC* sequence data we examined peaks of polymorphism using a sliding window analysis of nucleotide diversity implemented in DNAsp, with the default settings of window size of 50bp and step of 10 [16]. In addition to our *Y. brevifolia PgiC* sequence data we included homologous sequence data from related species of Yucca (Table S1). Additionally

we calculated pairwise and global F_{ST} using the *PgiC* data and examined isolation by distance using a Mantel test implemented in the R package ade4 [17].

RESULTS

Sequence Diversity

We investigated sequence variation in a 4kb fragment of the nuclear *PgiC* gene in 86 individuals of Joshua tree, *Yucca brevifolia* (Table S1). The majority of our sequence data is from within introns, with only 232 bp of coding domain. We recovered 151 single nucleotide polymorphisms (SNPs) within the *PgiC* data set, including three synonymous substitutions and 148 intronic SNPs. We identified 62 different sequence variants (haplotypes) in the *PgiC* data set, with haplotype number per population ranging from a minimum of 11 (MP) to a maximum of 17 (AZ and JT) (Table 1A). Haplotype diversity for the entire *PgiC* dataset is 0.83, with the greatest haplotype diversity found in the westernmost populations, WM and JT.

We also sequenced a 300bp region from the *trnD-trnT* chloroplast spacer region from 108 individuals (Table S1). Sequence of the *trnD-trnT* spacer region revealed only two SNPs and a single bas pair insertion that allowed us to identify three haplotypes. Haplotype 1, the ancestral haplotype, is found at very low frequency in AZ and WM, and is fixed in both MP and UT; haplotype 2 is found at high frequency in AZ and is fixed in JT, while haplotype 3 is only found in WM and is at high frequency (Table 1B, Table 2). Patterns of cpDNA haplotype diversity across the landscape are similar to those of our nuclear *PgiC* data, with the greatest haplotype diversity (0.50) in the western populations, JT and WM (*Y. brevifolia* ssp. *brevifolia*) and lower levels of haplotype diversity (0.41) in the

eastern populations AZ, MP, and UT (*Yucca brevifolia* ssp. *jaegeriana*). For *PgiC*, average pairwise diversity, θ_π, and Watterson's estimator of theta, θ_W, across all populations are 0.0064 and 0.0091, respectively, though there is much variation between populations and between proposed subspecies. Both θ_π and θ_W are higher in ssp. *brevifolia* (populations JT and WM) than in ssp. *jaegeriana* (populations AZ, MP, and UT), with JT having the highest values of both θ_π and θ_W (θ_π = 0.0122 and θ_W =0.0114) and UT having the lowest (θ_π = 0.0018 and θ_W = 0.0025) (Table 1A, Table 1B). Tajima's D is strongly negative in all populations except JT (D = 0.31), indicating an excess of low frequency variants in AZ, MP, UT, and WM, and a paucity of low frequency variants in JT. Using sliding window analysis we identified high peaks of θ_π between *PgiC* sequences around bp 1000 and bp 1800 of this region, both of which are within intron 12 (Figure S1).

Population structure

We examined evidence of population differentiation by calculating F_{ST} both globally and for all pairs of populations. Among all populations sampled, we found an F_{ST} of 0.052 (Table 2). Values of pairwise F_{ST} varied between populations, with no significant differentiation found between AZ, MP, and UT, and very low values of F_{ST} between the western populations, JT and WM. Levels of differentiation between subspecies (F_{CT} = 0.052) is equivalent to differentiation among all populations sampled (F_{ST} = 0.057). Pairwise F_{ST} increased with geographic distance between populations, but this observed isolation by distance was not statistically significant (Figure S2).

DISCUSSION

Population Structure and Geographic Distribution of Variation

To examine population structure in *Y. brevifolia* we calculated F_{ST} for all populations, between subspecies, and between all possible pairs of populations. Observed F_{ST} between populations of the same subspecies was lower than between subspecies, suggesting reduced gene flow between subspecies in the process of diverging or that populations have been geographically isolated in the past. This finding is in contrast to that of Toulson and Merritt (2008) [9], who found significant differentiation in allozyme allele frequencies at the PGI locus between populations and even among subpopulations within a single geographic locality. Their findings suggested that geographic barriers may be significant enough to prevent gene flow across long distances within the Mojave, as well as on a more localized scale. However, we find no evidence of strong genetic differentiation at the *PgiC* locus, and sharing of chloroplast haplotypes observed here and in previous work [18] further supports a lack of genetic differentiation between populations.

We examined sequence diversity in both our *PgiC* and *trnD-trnT* datasets in order to elucidate the demographic history of this species. Both data sets reveal a pattern of decreasing variation starting from the southwestern corner of the species distribution to the northeastern areas, which is consistent with previous allozyme and chloroplast data [9], [18]. These geographic patterns within our data are in support with the predicted distribution of diversity following postglacial range expansions; glacial refugia are proposed to be diversity hotspots, while peripheral populations should be fixed for specific haplotype lineages [8]. It is believed that following the Pleistocene, Joshua tree populations expanded north and east from a Southern refugia in the Colorado River Valley and were

abundant throughout both the Mojave and Sonoran Deserts during the Holocene. Following this expansion once continuous populations of Joshua trees became fragmented, and have continued to fragment and shrink as they move upwards in elevation and north in latitude [6]. This trend has been observed in other desert species, and phylogeographic studies of two Mojave and Sonoran Desert species also point to a southwestern Pleistocene refugia for desert flora [7]. Fehlberg and Ranker (2009) [7] found that *Encelia farinosa* has higher levels of genetic diversity near southwestern Arizona and Baja California Sur, and similar investigations into the phylogeography of the columnar cactus *Lophocereus schotti* also point to the northern Baja peninsula and Southwestern CA as a glacial refugia [19].

The effects of genetic drift on small populations may also contribute to the lower levels of diversity in the eastern populations. The eastern populations of *Y. brevifolia* are smaller and discontinuous, and it seems likely that rare alleles may have been lost as populations of this species contracted and shifted upwards in elevation in response to climatic warming. Similarly, the western populations (JT and WM) may retain greater levels of diversity because they are part of a larger, more continuous population of Joshua trees, thus the effects of drift and the likelihood of fixation of alleles are less.

Another explanation for the lower levels of diversity in AZ, UT, and MPA in the *PgiC* data set could be due to differences in sample size. We only sampled 14-22 individuals per population, and only individuals that had either PGI allozyme allele 3 or 4. If we increased and broadened our sampling, we may find additional sequence variants for this 4kb region in eastern populations. Furthermore, trends observed at one, or even a few loci may not be representative of genome-wide patterns.

Polymorphism within Yucca

There has been much disagreement over the taxonomy of this species, and it is commonly believed that there are two subspecies of Joshua tree. Subspecies division has been based on differences in morphology, specifically floral traits [20] and differences in distribution within the Mojave Desert [21]. However, the genetic basis underlying subspecies division remains to be elucidated, and previous molecular studies suggest both shared ancestral polymorphism [18], and gene flow between sympatric populations of the proposed subspecies [22]. Our data support these findings, and do not provide evidence of substantial genetic differentiation between subspecies of Joshua tree. All three chloroplast haplotypes are found in representative populations from both proposed subspecies *brevifolia* (JT and WM) and subspecies *jaegeriana* (AZ, MP, and UT), and values of F_{CT} between subspecies do not suggest genetic differentiation. However, given the long generation time of these plants and the relatively young age of Yucca (estimated to be 6-10 million years old [18], it is unlikely that one or even a few loci will show strong genetic differentiation between morphologically distinct subspecies.

Moreover, the ancestral nature of many of the sequence variants we observe at the *PgiC* locus is consistent with previous studies. The three morphologically based sections within Yucca (Sarcocarpa, Clistocarpa, and Chaenocarpa) show weak resolution using microsatellite and chloroplast data, which may be a reflection of the young age of the genus Yucca, consistent with shared ancestral polymorphism [11]. If this is the case it is likely that many loci will reflect this pattern.

Signatures of ancestral polymorphism in both *PgiC* and *trnD-trnT* data may also represent hybridization and introgression between Yucca species. Many species of Yucca

are thought to hybridize in nature and can be crossed in greenhouse settings; however, the only documented introgression has been from *Y. schidigera* into *Yucca baccata* [23]. Recurrent hybridization and introgression events between Yucca species could account for shared polymorphisms between species; however, we think it unlikely that hybridization and introgression explain the patterns observed in our data for two reasons. First, while hybridization may occur naturally between Yucca species that share moth pollinators, *Yucca brevifolia* has species-specific moth pollinators that are not known to pollinate any other species of Yucca, and cross-pollination events leading to hybridization would be unlikely in nature. Second, none of the species of Yucca examined in this study have overlapping distributions. The species *Y. brevifolia* and *Y. schidigera* both grow in the Mojave Desert, *Y. glauca, Y. angustissima,* and *Y. baileyi* are abundant in the Midwest US, and species such as *Y. filamentosa* and *Y. aloifolia* are restricted to the Southeastern USA.

A final explanation for patterns of variation at the *PgiC* locus of this species may be that selection is maintaining polymorphism at the *PgiC* locus of Yucca. Balancing selection is known to maintain long-term neutral variation in populations, or species, and could account for the transpecific polymorphism at this locus [24]. Variation in PGI on both the allozyme and nucleotide level has been extensively studied in many species, both plant and animal, and it is proposed that selection plays a critical role in maintaining this variation [25]. For example, Filatov and Charlesworth (1999) [26] suggest that balancing selection maintains two divergent haplotype groups at *PgiC* in multiple *Leavenworthia* species, and using sliding window analyses they show that an amino acid change (Asn/Lys) in exon 8 of *PgiC* in this species is linked to peaks of polymorphism within and around intron 12.

The results of our sliding window analyses show high peaks of π around bp1000 and bp1800 of this region, both of which are within intron 12 (Figure S1). It is possible that these peaks of diversity are linked to polymorphisms within the coding domain, such as amino acid changes that correspond to allozyme alleles. In this study we only captured 232 bp of coding domain of the *PgiC* gene, and additional sequencing of the rest of the coding domain of *PgiC* and investigation into nucleotide variation in additional unlinked nuclear markers will be necessary to test this hypothesis and disentangle ancestral polymorphism and the potential for balancing selection in Yucca.

CONCLUSIONS

We investigated DNA sequence variation in the nuclear *PgiC* locus and the *trnD-trnT* spacer region of the chloroplast in 86 and 108 individuals, respectively, from 5 populations of Joshua trees across the Mojave Desert. We find little evidence for restricted gene flow between populations, though this may reflect the long generation time of this species. Furthermore, our sequence diversity results support previous studies suggesting that this species has experienced expansion from a Pleistocene glacial refugium, with subsequent population fragmentation. Finally, we do not find strong genetic evidence supporting the division of this species into two genetically distinct subspecies. Further investigation into genome-wide patterns of diversity to elucidate the extent of gene flow and local adaptation will be necessary if we are to understand evolution and divergence in this species.

Acknowledgements:

The authors would like to thank the National Park Service (NPS), specifically Joshua Tree National Park, Mojave National Preserve, and Tasha La Doux for help and cooperation with field collections.

Figure 1. **Map of sampling sites.** The Mojave Desert is shaded in light grey, with the distribution of *Y. brevifolia* in dark grey. Sample sites are as follows: *Y. brevifolia* ssp. *jaegeriana* : AZ (Arizona), MP (Mojave National Preserve), UT (Utah); *Y. brevifolia* ssp. *brevifolia*: JT (Joshua Tree National Park), WM (Western Mojave).

Figure S1. Sliding window plot of π across the *PgiC* region. Plot of sliding window analysis of π across the *PgiC* region of focus showing peaks of diversity around base pair 1000 and bp 1800. Zero on the plot refers to base pair 805 of the coding domain of *PgiC*, and base pair 4,000 on the plot is synonymous with base pair 1037 of the coding domain. Green rectangles indicate the position of exons within the sequenced region while black lines represent introns. Individuals in Table S1 used for sliding window analysis have an * by their name.

Figure S2. Plot of $F_{ST}/(1-F_{ST})$ and distance (km) between populations. Plot showing increasing genetic differentiation (F_{ST}) with increasing distance (km) between populations of *Y. brevifolia*. Values used for the plot are found in Table S4.

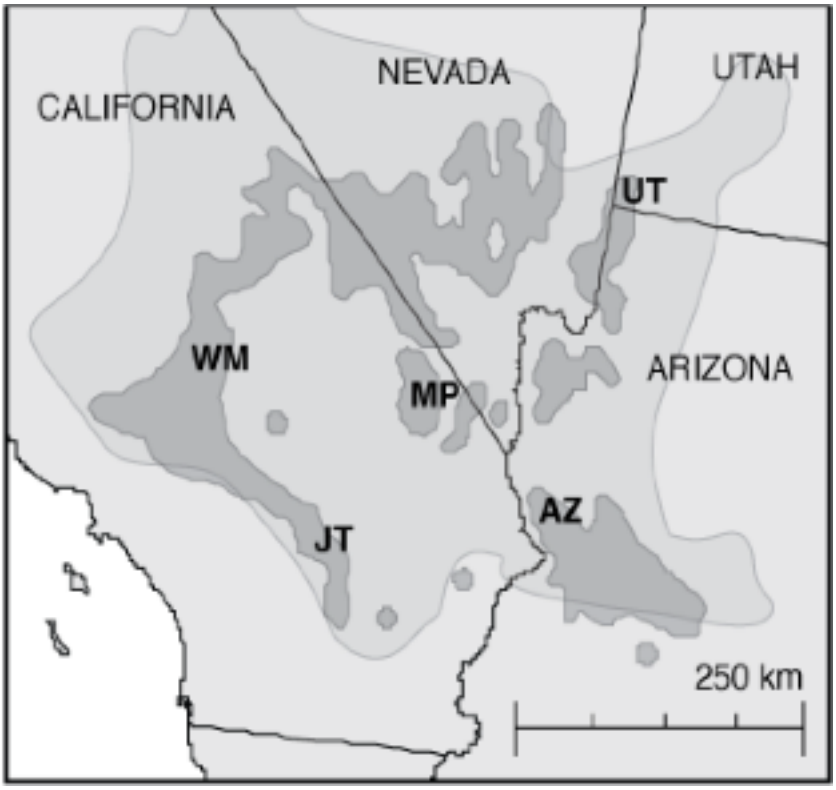

Figure 1. **Map of sampling locations.**

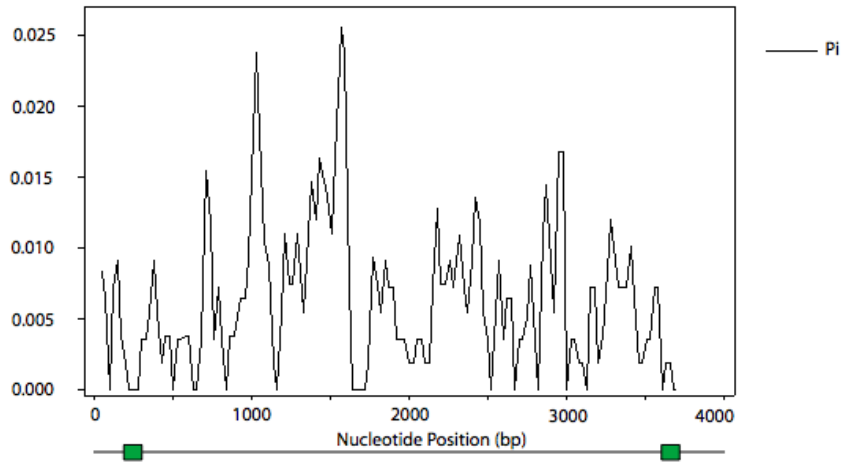

Figure S1. Sliding window plot of π across the *PgiC* region.

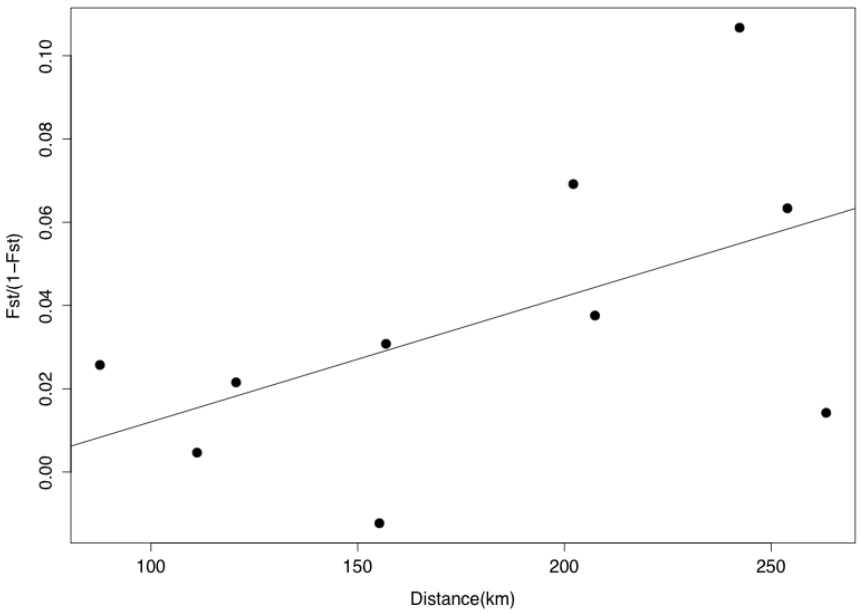

Figure S2. Plot of $F_{ST}/(1-F_{ST})$ and distance (km) between populations.

Table 1. Diversity statistics for the *PgiC* data set (A), and the *trnD-trnT* spacer region (B); note that haplotype statistics do not include information on insertion deletion polymorphisms.

A. Diversity statistics for *PgiC*

Population	Number of Haplotypes	Haplotype Diversity	Number of Segregating Sites	θ_π	θ_W	Tajima's D
AZ	17	0.8874	76	0.0033	0.0064	-2.0611
MP	11	0.9333	105	0.0059	0.0099	-1.9570
UT	12	0.8787	30	0.0018	0.0025	-1.1736
ssp. *jaeg*	41	0.8242	135	0.0035	0.0088	-2.2452
JT	17	0.9542	134	0.0122	0.0114	0.3094
WM	12	0.9780	118	0.0078	0.0107	-1.2520
ssp. *brev*	24	0.9365	139	0.0103	0.0100	0.1364
All	62	0.8268	155	0.0064	0.0091	-1.0283

B. Diversity statistics for *trnD-trnT* region

Population	Number of Haplotypes	Haplotype Diversity	Number of Segregating Sites	θ_π	θ_W	Tajima's D
AZ	2	0.0769	1	0.0003	0.0012	-1.1556
MP	1	0.0000	0	0.0000	0.0000	N/A
UT	1	0.0000	0	0.0000	0.0000	N/A
ssp. *jaeg*	2	0.4114	1	0.0018	0.0009	1.2682
JT	1	0.0000	0	0.0000	0.0000	N/A
WM	1	0.0000	0	0.0000	0.0000	N/A
ssp. *brev*	2	0.4989	0	0.0022	0.0009	1.7777
All	2	0.4638	1	0.0021	0.0008	1.7150

Table 2. Chloroplast haplotype frequencies across populations and subpopulations

		Chloroplast Haplotype		
Population	**Subpopulation**	**1**	**2**	**3**
AZ	AZA	0.100	0.900	0.000
	AZB	0.000	1.000	0.000

	AZC	0.000	1.000	0.000
	Total Population	*0.038*	*0.962*	*0.000*
MP	MPA	1.000	0.000	0.000
	MPC	1.000	0.000	0.000
	Total Population	1.000	0.000	0.000
UT	UTA	1.000	0.000	0.000
	UTB	1.000	0.000	0.000
	UTC	1.000	0.000	0.000
	Total Population	*1.000*	*0.000*	*0.000*
var. *jaegeriana*		0.716	0.284	0.000
JT	JTA	0.000	1.000	0.000
	JTB	0.000	1.000	0.000
	JTC	0.000	1.000	0.000
	Total Population	*0.000*	*1.000*	*0.000*
WM	WMA	0.000	0.000	1.000
	WMB	0.000	0.000	1.000
	WMC	0.889	0.000	0.111
	Total Population	*0.340*	*0.000*	0.660
var. *brevifolia*		0.200	0.413	0.388
All		0.470	0.345	0.185

Table 3. F_{ST} calculations for *PgiC* data

	AZ	**JT**	**MP**	**UT**	**WM**
AZ	-	-	-	-	-
JT	0.065	-	-	-	-
MP	0.000	0.025	-	-	-
UT	0.014	0.096	0.030	-	-
WM	0.036	0.005	0.021	0.060	-
Subspecies (F_{CT})	0.052				
All populations	0.057				

Table S1. Sample set for the *trnD-trnT* and the *PgiC* data

Population	**Individual**	**Chloroplast Haplotype**	**PGI Allozyme Genotype**	**Sequenced *PgiC***
AZ	AZA01	2	3/3	Y
	AZA06	2	3/3	N
	AZA08	2	3/3	Y
	AZA13	2	3/3	Y
	AZA15*	2	3/3	Y
	AZA24	2	3/3	N
	AZA26*	2	3/3	Y
	AZA32	1	3/3	N
	AZA33	2	3/3	Y
	AZA38	2	3/3	Y
	AZB06	2	3/4	Y
	AZB24	2	3/5	N
	AZB25	2	3/3	N
	AZB31	2	3/3	N
	AZB35	2	3/3	N
	AZB36	2	3/3	Y
	AZB41	2	3/3	N
	AZB43*	2	3/3	Y
	AZC06	2	3/3	Y
	AZC13	2	3/4	Y
	AZC17	2	3/3	Y
	AZC18*	2	3/3	Y
	AZC19	2	3/5	N
	AZC32	2	3/3	N
	AZC40	2	3/5	N
	AZC55	2	N/A	N

JT	JTA06	2	3/3	Y
	JTA09	2	3/4	N
	JTA10	2	2/3	N
	JTA13	2	3/4	N
	JTA15*	2	3/3	Y
	JTA21	2	2/3	N
	JTA29	2	3/4	Y
	JTA31	2	2/4	N
	JTA32	2	3/5	N
	JTA38	2	3/4	N
	JTA39	2	5/5	N
	JTB02	2	3/3	Y
	JTB03	2	3/3	N
	JTB10*	2	3/3	Y
	JTB12	2	3/3	N
	JTB16	2	3/4	N
	JTB24	2	3/4	Y
	JTB26*	2	3/4	Y
	JTB31	2	5/5	N
	JTB35	2	3/4	N
	JTB39*	2	3/3	Y
	JTB47	2	4/5	N
	JTC03	2	3/5	N
	JTC06	2	3/4	N
	JTC13	2	2/3	N
	JTC15*	2	4/4	Y
	JTC29	2	4/4	Y
	JTC30	2	4/4	N
	JTC34	2	3/3	Y
	JTC35	2	4/5	N
	JTC37	2	3/4	N
	JTC40	2	4/4	Y
	JTC42	2	4/4	N
MP	MPA01*	1	4/4	Y
	MPA04	1	3/4	Y
	MPA09	1	5/5	N
	MPA12	1	3/3	Y
	MPA16	1	3/3	N
	MPA19	1	3/3	N
	MPA22	1	3/3	N
	MPA24	1	3/4	Y
	MPA25	1	3/4	Y
	MPA31	1	3/4	Y

	MPA37	1	3/4	Y
	MPA40	1	3/4	N
	MPA43	1	3/3	Y
	MPA47	1	3/3	N
	MPA48	1	3/3	Y
	MPC03*	1	3/4	Y
	MPC05	1	3/5	N
	MPC13	1	3/4	Y
	MPC21	1	3/3	N
	MPC22	1	3/4	Y
	MPC23*	1	3/3	Y
	MPC27	1	3/4	N
	MPC29	1	3/5	N
	MPC36	1	3/3	Y
	MPC39	1	3/3	N
	MPC40	1	3/3	Y
	MPC42	1	3/4	N
	MPC43	1	3/3	N
UT	UTA05*	1	3/3	Y
	UTA07	1	3/3	Y
	UTA12	1	3/3	Y
	UTA19	1	3/3	Y
	UTA20	1	1/3	N
	UTA27	1	3/3	N
	UTA33	1	3/4	N
	UTA34	1	3/4	N
	UTA36	1	3/3	Y
	UTA42	1	3/3	Y
	UTA43	1	3/3	Y
	UTA44	1	3/3	Y
	UTB11*	1	3/3	Y
	UTB18	1	3/4	Y
	UTB19	1	1/1	N
	UTB20	1	4/4	Y
	UTB21	1	3/3	Y
	UTB30	1	1/3	N
	UTB35	1	3/3	Y
	UTB37	1	3/3	N
	UTB39	1	3/3	N
	UTB40	1	3/5	N
	UTB41	1	4/4	N
	UTB46	1	3/4	N
	UTC08	1	4/5	N

	UTC09	1	1/3	N
	UTC11	1	3/3	Y
	UTC13	1	3/4	Y
	UTC18	1	4/5	N
	UTC27*	1	3/4	Y
	UTC28	1	3/3	Y
	UTC38	1	3/3	N
	UTC46	1	3/3	Y
	UTC48	1	3/3	N
WM	WMA01	3	4/5	N
	WMA02	3	3/4	Y
	WMA04	3	3/4	N
	WMA05	3	3/3	N
	WMA07	3	3/3	N
	WMA08	3	3/3	N
	WMA09	3	3/4	N
	WMA10	3	3/4	N
	WMA11*	3	4/4	Y
	WMA12	3	3/3	Y
	WMA14	3	3/3	N
	WMA17	3	3/4	N
	WMA19	3	4/4	N
	WMA20	3	3/3	N
	WMA21	3	3/4	Y
	WMA22	3	3/4	N
	WMB02	3	3/4	N
	WMB03	3	3/4	N
	WMB05*	3	3/4	Y
	WMB06	3	3/4	Y
	WMB10	3	4/4	Y
	WMB11*	3	4/4	Y
	WMB12	3	4/4	N
	WMB14	3	4/4	N
	WMB16	3	3/4	Y
	WMB17	3	3/4	N
	WMB18	3	3/4	N
	WMB20	3	3/3	N
	WMB24	3	3/4	N
	WMC01*	1	3/4	Y
	WMC02	1	3/4	N
	WMC04	1	3/4	N
	WMC05	1	3/3	Y
	WMC07	1	3/3	Y

	WMC08	1	3/5	N
	WMC09	1	3/4	N
	WMC10*	3	4/4	Y
	WMC15	1	3/4	N
	WMC16	1	3/3	N
	WMC17	1	3/4	Y
	WMC18	3	4/4	N
	WMC19	1	3/3	N
	WMC20	1	3/4	N
	WMC21	1	3/4	N
	WMC22	1	3/3	N
	WMC23	1	4/4	N
	WMC24	1	3/4	N
Y. aloifolia	Smith College	N/A	N/A	Y
Y. angustissima	Smith College	N/A	N/A	Y
Y. baileyi	Smith College	N/A	N/A	Y
Y. filamentosa	Smith College	N/A	N/A	Y
Y. filamentosa	L.Vann	N/A	N/A	Y
Y. filamentosa	L.Vann	N/A	N/A	Y
Y. glauca	Smith College	N/A	N/A	Y
Y. nana	Smith College	N/A	N/A	Y
Y. schidigera	JOTR	N/A	N/A	Y
Y. schidigera	JOTR	N/A	N/A	Y
Y. schidigera	JOTR	N/A	N/A	Y

Individuals with * by their name were included in sliding window analysis of π.

Table S2. Sequences of primers used for amplification and sequencing

Primer	**Sequence (5'-3')**
trnH	GGG TTG CCC GGG ACT CGA AC
trnS	CGC CGC TTT AGT CCA CTC A
atpH	CCA GCA GCA ATA ACG GAA GC
trnD	ACC AAT TGA ACT ACA ATC CC
petA	GCA TCT GTT ATT TTG GCA CA
rps18	GCT CGT ATT TTA TCT TGT TAC C
petB	CTA TCG TGG RAC CGT TAC WGA GGC T
trnK	CAA CGG TAG AGT ACT CGG CTT TTA
trnR	ATT GCG TCC AAT AGG ATT TGA A
atpI	ATA GGT GAA TCC ATG GAG GG
psbE	TAC CTT CCC TAT TCA TTG CG
clpp	AAC CTG CTA GTT CTT WTT AT
IntD	GAT AGG ATT CCT TCA GTT CAG GC
petD	CAA AYG GAT AYG CAG GTT CAC C
805X	GGA TGG GTT GTT GGT CGC TAT AGT
1037X	CCT GAG AGT ATG GCA GTA TGG CCA
1037X	CCT GAG AGT ATG GCA GTA TGG CCA
Intron7-	ATC CAA TGC GTT AGT ACA GGC
Intron1+	AAG TCA AGC CTG TAC TAA CGC
Intron2+	CGG TAG CAA AGG ACA GAA TGC
Intron3+	CAA GTC TTG AGT AAT AAT TGC
Intron4+	CAA TGT TGA TAG ACC ATT GGC
Intron5-	CAA GTC TAT TAA ACA TCG TGG
Intron4-	AAC CCG ATG TGT ATG ACC TGC
Intron2-	CCA GGA TTC TAT TTC TTC TCG
Intron5+	AAT GCA TCC ACC ACA TCC TGC
Intron3-	GTT ACC ACA TAT CCT GAT GGC
Intron7+	TTT GTT TGG TGA GAT GGG AGC
1236-	CAA GTA GAC AGG CTG CTG AGT
1265-	CAT CGT GGT TGC TCA CAA TCT
1025+	CAT ACT CTC AGG CGC TAG AGA

Table S3. Summary of chloroplast regions surveyed

Region	Length (bp)	Number of Polymorphic sites
trnH-trnF	600	0
aspH-atpI	495	0
trnD-trnT	1000	1 insertion, 1 Substitution
petA-psbE	430	0
rps18-clpp	880	0
petD-petB	885	0

Table S4. F_{ST} and distance (km) between populations

	Populations	F_{ST}	Distance (km)
AZ	JT	0.065	202.150
	MP	0.000	155.240
	UT	0.014	263.320
	WM	0.036	207.400
JT	MP	0.025	87.650
	UT	0.096	242.370
	WM	0.005	111.140
MP	UT	0.030	156.840
	WM	0.021	120.580
UT	WM	0.060	253.940

1. Smith SD, Harstock TL, Nobel PS (1983) Ecophysiology of *Yucca brevifolia*, an arborescent monocot of the Mojave Desert. Oecologia 60: 10-17.
2. Grayson DK (1993) The desert's past: a natural prehistory of the Great Basin. Washington D.C.: Smithsonian Institution Press. 365 p.
3. Laudermilk M, Munz PA (1934) Plants in the dung of *Northrotherium* from Gypsum Cave, Nevada. Paleontological Contributions 453: 29-37.

4. Comanor PL, Clark WH (2000) Preliminary growth rates and a proposed age-form classification for the Joshua tree, *Yucca brevifolia* (Agavaceae). Haseltonia 7: 37-45.
5. Lenz LW (2001) Seed dispersal in *Yucca brevifolia* (Agavaceae) - present and past, with consideration of the future of the species. Aliso 20: 61-74.
6. Cole KL, Ironside K, Eischeid J, Garfin G, Duffy PB et al. (2011) Past and ongoing shifts in Joshua tree distribution support future modeled range contraction. Ecological Applications 21: 137-149.
7. Fehlberg SD, Ranker TA (2009) Evolutionary history and phylogeography of *Encelia farinosa* (Asteraceae) from the Sonoran, Mojave, and Peninsular Deserts. Molecular Phylogenetics and Evolution 50: 326-335.
8. Comes HP, Kadereit (1998) The effect of Quaternary climatic changes on plant distribution and evolution. Trends in Plant Science 3: 432-438.
9. Toulson AT, Merritt RB (2008) Glucose-6-phosphate isomerase variation and genetic structure in *Yucca brevifolia* (Agavaceae). Madrono 55: 284-290.
10. Toulson AT (2003) Phosphoglucose isomerase variation and genetic structure in Joshua tree (*Yucca brevifolia*) populations. Biological Sciences BA:
11. Smith CI, Pellmyr O, Althoff DM, Balcazar-Lara M, Leebens-Mack J et al. (2008) Pattern and timing of diversification in *Yucca* (Agavaceae): specialized pollination does not escalate rates of diversification. PNAS 275: 249-258.
12. Terauchi R, Terachi T, Miyashita NT (1997) DNA polymorphism at the *Pgi* locus of a wild yam, *Dioscorea tokoro*. Genetics 147: 1899-1914.

13. Grivet D, Heinze B, Vendramin GG, Petit RJ (2001) Genome walking with consensus primers: application to the large single copy region of chloroplast DNA. Molecular Ecology Notes 1: 345-349.
14. Watterson GA (1975) On the number of segregating sites in genetical models without recombination. Theoretical Population Biology 7: 256-276.
15. Thornton K (2003) Libsequence: a C++ class library for evolutionary genetic analysis. Bioinformatics 19: 2325-2327.
16. Rozas J, Rozas R (2009) DNAsp version 5.0: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25: 1451-1452.
17. Thiolouse J, Chessel D, Doledec S, Olivier JM (1997) ADE-4: a mutlivariate analysis and graphical display software. Statistics and Computing 7: 75-83.
18. Smith CI, Godsoe WKW, Tank S, Yoder JB, Pellmyr O (2008) Distinguishing coevolution from covicariance in an obligate pollination mutualism: asynchronous divergence in Joshua tree and its pollinators. Evolution 62: 2676-2687.
19. Nason JD, Hamrick JL, Fleming TH (2002) Historical vicariance and postglacial colonization effects on the evolution of genetic structure in *Lophocereus*, a Sonoran Desert columnar cactus. Evolution 56: 2214-2226.
20. Godsoe W, Yoder JB, Smith CI, Pellmyr O (2008) Coevolution and divergence in the Joshua tree/Yucca moth mutualism. The American Naturalist 171: 816-823.
21. McKelvey SD (1938) Yuccas of the Southwestern United States I. Jamaica Plain, MA: Arnold Arboretum Harvard University.

22. Starr TN, Gadek KE, Yoder JB, Flatz R, Smith CI (2013) Asymmetric hybridization and gene flow between Joshua trees (Agavaceae: *Yucca*) reflect differences in pollinator host specificity. Molecular Ecology 22: 437-499.
23. Hanson M (1993) Dispersed unidirectional introgression from *Yucca schidigera* into *Yucca baccata* (Agavaceae). Botany PhD:
24. Charlesworth D (2006) Balancing selection and its effects on sequences in nearby genome regions. PLoS Genetics 2: 0379-0384.
25. Gillespie JH (1991) The Causes of Molecular Evolution. New York, NY: Oxford University Press. 336 p.
26. Filatov DA, Charlesworth D (1999) DNA polymorphism, haplotype structure and balancing selection in the Leavenworthia *PgiC* locus. Genetics 153: 1423-1434.