Phylogeny and Cytogeography of the North American Creosote Bush (Larrea tridentata, Zygophyllaceae)

Authors: Robert G. Laport, Robert L. Minckley, and Justin Ramsey
Source: Systematic Botany, 37(1) : 153-164
Published By: The American Society of Plant Taxonomists
URL: https://doi.org/10.1600/036364412X616738
Phylogeny and Cytogeography of the North American Creosote Bush
(\textit{Larrea tridentata}, Zygophyllaceae)

Robert G. Laport, Robert L. Minckley, and Justin Ramsey

Department of Biology, University of Rochester, 213 Hutchison Hall, River Campus, Rochester, New York 14627-0211 U. S. A.
1Author for correspondence (rob.laport@gmail.com)

Abstract—The North American creosote bush (\textit{Larrea tridentata}, Zygophyllaceae) is a widespread and ecologically dominant taxon of North American warm deserts. The species is comprised of diploid, tetraploid, and hexaploid populations, and touted as a classical example of an autopolyplid taxonomic complex. Here we use flow cytometry and DNA sequence data (non-coding cpDNA and nuclear ribosomal DNA) to evaluate spatial and evolutionary relationships among cytotype races, as well as the origins of the species from its South American ancestors. We find the geographic distribution of North American cytotypes to be highly structured, with limited co-occurrence within populations. Diploids reside only in the Chihuahuan Desert, as reported in previous biotic surveys, but tetraploid and hexaploid populations interdigitate along the margins of the Sonoran and Mojave Deserts. In phylogenetic analyses, North American plants comprise a monophyletic grouping that is sister to the South American diploid species, \textit{L. divaricata}. North American populations exhibit genetic signatures of rapid demographic expansion, including a star-shaped genealogy, unimodal distribution of pairwise haplotype differences, and low genetic structure. Nonetheless, polyploid cytotypes are consistently distinguished from diploid cytotypes by a cpDNA indel character, suggesting a single origin of tetraploidy in the species. These findings suggest a recent origin of the North American creosote bush via long distance dispersal, with establishment of polyploid populations accompanying its rapid spread through the Northern Hemisphere.

Keywords—cpDNA, chromosome evolution, molecular systematics, nrDNA, phylogeography.

The genus \textit{Larrea} comprises long-lived shrubs that are widespread across New World desert regions. Four species inhabit xeric regions of South America (\textit{Larrea divaricata} Cav., \textit{Larrea cuneifolia} Cav., \textit{Larrea nitida} Cav., \textit{Larrea ameghinoi} Cav.) with one species in North America (\textit{Larrea tridentata} (DC.) Coville). \textit{Larrea tridentata} (creosote bush) is the most widespread shrub in the American southwest and northern Mexico, often occurring in monospecific stands across thousands of square kilometers (Benson and Darrow 1944; Turner et al. 1995). Creosote bush is a food resource for many animals, including numerous specialist herbivores and pollinators, and regarded as both a defining element and a keystone species of North American warm deserts (Wells and Hunziker 1976; Mabry et al. 1977).

\textit{Larrea tridentata} is also considered to be a classical example of a polyploid complex (Lewis 1980). Chromosome counts reveal that the species is composed of morphologically-similar diploid (2\textit{n} = 2\textit{x} = 26), tetraploid (2\textit{n} = 4\textit{x} = 52), and hexaploid (2\textit{n} = 6\textit{x} = 78) cytotypes (Yang 1967, 1968, 1970; Yang and Lowe 1968; Barbour 1969). Early sampling indicated that diploid populations occur in the Chihuahuan Desert (western Texas, southern New Mexico, southeastern Arizona, northeastern Mexico), tetraploid populations occur in the Sonoran Desert (south-central Arizona, southeastern California, north-central Mexico, and Baja California), and hexaploid populations occur in the Mojave Desert (northwestern Arizona, southeastern California, southern Nevada, and southwestern Utah). However, recent studies, which made use of guard cell measurements to infer ploidy level on living plants and fossils, suggest that cytotype distributions are geographically complex and temporally dynamic (Hunter et al. 2001). The occurrence of polyploidy in South America is limited to \textit{L. cuneifolia}, a putative allopolyploid species (Hunziker et al. 1977).

Cytogenetic studies and isozyme analyses indicate that the North American creosote bush is autoploidyd (Sternberg 1976; Wells and Hunziker 1976; Hunziker et al. 1977; Poggio et al. 1989; Cortes and Hunziker 1997). Although there has been little discussion of applying species or subspecies nomenclature to the autoploidyd “races” of \textit{L. tridentata}, the species is sometimes regarded as conspecific to the morphologically similar species, and presumed progenitor, \textit{L. divaricata}, which occurs in Argentina, Chile, and southwestern Peru (Raven 1963, 1972; Porter 1974; Hunziker et al. 1977). Moreover, a narrowly-distributed subspecific taxon, the dune creosote bush (\textit{Larrea tridentata} var. arenaria L. D. Benson), has recently been recognized (Benson and Darrow 1981; Felger 2000). Unfortunately, there are limited molecular data available to evaluate taxonomic boundaries in the genus \textit{Larrea} or to test biogeographical hypotheses regarding its unusual amphiboreal distribution. Allozyme studies suggest recent divergence between \textit{L. tridentata} and South American taxa but provide minimal resolution of species relationships (Cortes and Hunziker 1997). Similarly, non-coding cpDNA is consistent with the monophyll of North American creosote bush but provides little statistical support or resolution of species relationships (Sheahan and Chase 1996, 2000; Lia et al. 2001).

Here we present an in-depth analysis of cytotype distributions in \textit{L. tridentata} based on flow cytometry, and perform phylogenetic and population genetic analyses based on non-coding cpDNA and nrDNA, to address the following questions: (1) What is the geographic distribution of cytotype races in relation to the Chihuahuan, Sonoran, and Mojave deserts? (2) Do cytotypes co-occur regionally and/or within populations? (3) Is the North American \textit{L. tridentata} monophyletic and evolutionarily distinct from South American \textit{Larrea} species? (4) Is the subspecific taxon, \textit{L. tridentata} var. arenaria, distinguished phylogenetically from other North American creosote bush populations? And (5) is the population structure within and among North American populations consistent with recent demographic expansion and in situ polyploid formation?
Materials and Methods

Sampling—We sampled individuals from 92 populations of *L. tridentata* throughout the southwestern U. S. A. (Texas, New Mexico, Arizona, California, and Nevada) and northern Mexico (Baja California, Chihuahua, Coahuila, and Sonora), a broad geographic region that encompasses diploid, tetraploid, and hexaploid populations (Appendix 1). The narrowly-distributed *L. tridentata var. arenaria* was sampled from the Algodones Dunes region of southern California (Appendix 1). Each population consisted of 5–50 randomly-selected, individually marked plants. To maximize the probability of sampling different genetic individuals (gene samples) we collected from plants that were at least 10 m apart. For each individual, GPS coordinates and elevation were recorded. Young vegetative tissues (leaves and stems) were harvested from each individual plant during the spring seasons of 2007–2009 and thereafter stored on silica gel. Outgroup sampling included several South American *Laurea* species (*L. divaricata, L. cuneifolia,* and *L. nitida*) as well as *Guaiacum couleri* A. Gray, which were provided by T. Van Devender (Arizona-Sonora Desert Museum, Tucson, Arizona) and J. Takara (Rancho Santa Ana Botanic Garden, Claremont, California). Voucher specimens for sampled populations are deposited at the Rancho Santa Ana Botanic Garden Herbarium, Claremont, California (Appendix 1).

Ploidy Determinations—We used flow cytometry to infer DNA content and ploidy of all collected specimens. For each sample, 10–15 leaves were ground individually in 90% ethanol, and extracted using 25 ml of Qiagen Lysis Buffer (Qiagen Inc., Valencia, California). The DNA concentration of each sample was measured fluorometrically in a UV-Vis spectrophotometer (Beckman Coulter DU 640). The genome size for *L. tridentata* was calculated using *L. arenaria* (hexaploid) as internal control. Samples were centrifuged for 1 min at 10,000 g and resuspended in 490 μl of buffer with 10 mg/ml propidium iodide solution and 0.1% RNase A (QIAGEN Inc., Valencia, California). Concurrent with resuspension, 2.5 μl of prepared trout erythrocyte nuclei (BioSure Controls, Grass Valley, CA, USA) was added to each sample as an internal standard. Samples containing conspicuous debris were passed through a second filter. All samples were run on a FACS Calibur flow cytometer (B-D Biosciences, San Jose, California) in the Cell Sorting Facility at the University of Rochester School of Medicine, Rochester, New York. To determine ploidy level, we analyzed the relative fluorescence (FL2-A) of each sample, summarized as a frequency histogram using CellQuest Pro-Software (version 5.2.1; B-D Biosciences). We calculated 2C DNA contents of all specimens by multiplying 5.2 pg by the ratio between the sample peak and control peak.

DNA Extraction, Amplification, and Sequencing—We isolated genomic DNA of 2-3 plants from 40 *L. tridentata* populations as well as outgroup specimens using the QIAGEN DNeasy plant kit (QIAGEN). For each sample, 20–200 mg of silica-preserved tissue was used. We used five non-coding chloroplast (cpDNA) regions (*psba-trnH* intergenic spacer, *rpl23-trnl* intergenic spacer, *rpl16* intron, *rho2-trnC* intergenic spacer, *petN-trnC* intergenic spacer) using universal primers (Shaw et al. 2005, 2007). We performed 25 μl PCR reactions with 2 μl genomic DNA, 0.125 μl of each primer (100 mM stock, 0.5 μM final concentration), 0.5 μl dNTP mix (10 mM final concentration), 2.5 μl standard Taq reaction buffer (10x) and 0.1 μl standard Taq polymerase (5,000 units/ml; New England Biolabs, Ipswich, Massachusetts). Thermocycling and clean-up were performed as described by Ramsey et al. (2008). In addition to the cpDNA regions, we amplified the nuclear ribosomal external transcribed spacer using the primers 18S-ETS and Ast-1 (Markos and Baldwin 2001), and one copy of the nuclear ITS (photosynthetic) region (Markos and Baldwin 2001), for a representative sample of 17 ingroup and outgroup specimens. We used the primary/secondary primer set for the ITS region with the addition of 2 μl of primer (100 mM stock, 0.2 μM final concentration), 0.5 μl 5× ultrapure betaine (USB, Cleveland, Ohio), 2.5 μl Big Dye buffer (Applied Biosystems, Foster City, California) and 0.5 μl Big Dye (version 3.1; Applied Biosystems). Sequencing reactions were performed in forward and reverse directions. Products were purified using Montage™ vacuum plates (Millipore, Billerica, Massachusetts) and sequenced on an ABI 3730 capillary sequencer (University of Rochester Functional Genomics Center, Rochester, New York).

High quality sequence data were obtained for cpDNA and nrDNA regions in all sampled taxa. Sequences were manually edited using Sequencer™ (version 4.1); Gene Codes, Ann Arbor, Michigan) and manually aligned using MacClade (Maddison and Maddison 2002). For ETS and ITS sequences, we occasionally observed secondary peaks beneath primary peaks in the DNA chromatograms. These secondary peaks could represent sequencing artifacts and/or DNA polymorphism. Because of the small number of base pair positions with secondary peaks, and the uncertain cause of their appearance, we always used the primary/dominant peak for calling base pairs. We retained indels for analysis when inserted/deleted regions did not contain microsatellites, which we defined as tandems of simple sequence repeated ten or more times (Hughes and Queller 1993). All DNA sequences used in this study can be found in GenBank (Appendix 1).

Phylogenetic Analyses—Phylogenetic analyses of cpDNA and nrDNA were performed with a representative sample of 17 ingroup and outgroup specimens, which included three ploidy levels of *L. tridentata* (DC.) Coville and *L. tridentata var. arenaria*. Sequence data were analyzed using maximum parsimony with PAUP* (version 4.0b10; Swofford 2003) and Bayesian approaches with MrBayes (version 3.1b; Ronquist and Huelsenbeck 2003). Partition homogeneity tests (incongruence length difference tests as implemented in PAUP*) did not suggest discordance among noncoding cpDNA regions (*p = 0.274*) or ETS and ITS (*p = 0.412*) so all analyses were based on the combined data sets. Parsimony analyses were conducted with 2,000 random addition tree-bisection-reconnection searches, with a starting tree generated by stepwise addition and gaps treated as fifth bases; bootstrap support for each node was evaluated with 10,000 replicates. Four runs of Metropolis-coupled Markov Chain Monte Carlo analyses were implemented in MrBayes with ten million generations, two runs of four chains each, sampling every 1,000 generations, and a heating factor of 0.2. Convergence was assessed when the average standard deviation in split frequencies between the two parallel runs was less than 0.01 (Glar and Laport 2010). A GTR + I model of evolution was fit to the cpDNA data and a GTR + G model was fit to the ITS data using hierarchical likelihood ratio test (hLRT, Felsenstein 1988) and Akaike information criterion (Akaike 1974) in MrModeltest version 1.1b (Nylander 2002). Nexus files and trees can be found in TreeBASE (study number S13145).

Network and Population Genetic Analyses—Network and population genetic analyses were conducted on 112 sequences from the 40 North American *L. tridentata* populations. Haplotype networks were constructed using TCS (version 1.21; Clement et al. 2000). We used the Arlequin software package (version 2.0; Schneider et al. 2000) to perform analysis of molecular variance (AMOVA) with two population grouping schemes: geographic (grouped by desert region) and cytotype (grouped by ploidy level). Desert designations were based on floristic analyses by Shreve (1942), Benson and Darrow (1944), and McLaughlin (1986). We performed 1,000 permutations to assess significance of variance components. We evaluated the population history of North American *L. tridentata* using mismatch distributions, which are unimodal in evolutionary lineages undergoing demographic expansion (Slatkin and Hudson 1991; Rogers and Harpending 1992). Using Arlequin, we compared the observed frequency of pairwise differences to an expected distribution generated from a population expansion model with parametric bootstrapping (Schneider and Excoffier 1999). We also estimated demographic expansion parameters t (modal value of mismatch distribution), q0 (2Nμ before population expansion), and q1 (2Nμ after population expansion). Tajima’s D, a statistic describing the frequency of alleles at a locus (Tajima 1989), was also calculated.

Results

Flow Cytometry—Analyses of *L. tridentata* revealed a trimodal distribution of DNA contents that corresponded to values expected for a polyploid series (Fig. 1). The inferred mean 2C DNA values were 0.93 pg (diploid; range 0.76–1.19 pg), 1.84 pg (tetraploid; range 1.53–2.14 pg), and 2.79 pg (hexaploid; range 2.38–3.17 pg). Specimens of *L. tridentata var. arenaria* had DNA content values typical of tetraploid *L. tridentata* (mean = 1.79 pg, range 1.61–1.98 pg). There was...
no overlap of DNA content values between diploid, tetraploid, and hexaploid cytotypes, suggesting that flow cytometry can unambiguously determine the ploidy level of the North American creosote bush (Fig. 1).

Cytotype Distribution—Diploid cytotypes occurred primarily in the eastern portion of the study area (Chihuahuan Desert), tetraploid cytotypes in the central portion (Sonoran Desert), and hexaploid cytotypes in the western portion (Mojave Desert) (Figs. 2, 3). The boundary between diploid and tetraploid L. tridentata distributions corresponded closely to the traditionally defined boundary between the Chihuahuan and Sonoran Deserts, based on floristic affinities. In contrast, tetraploid and hexaploid cytotypes exhibited spatially complex distributions in Arizona and eastern California. Several hexaploid populations were found in central Arizona, a region considered typical of the Sonoran Desert floristic province (Figs. 2, 3). The average elevation differed significantly between diploid populations (mean = 1,045 m, range 658–1,540 m), tetraploid populations (mean = 513 m, range 52–1,078 m), and hexaploid populations (mean = 234 m, range 54–658 m) (ANOVA, F3,80 = 5.827, p < 0.001). Populations of L. tridentata var. arenaria were found to be geographically limited to low elevation areas in the Algodones Dunes of southeastern California (mean = 55 m, range 10–124 m) (Figs. 2, 3).

Despite the large number of diploid, tetraploid, and hexaploid plants sampled over the study region (N = 775 specimens in 92 populations), only two populations were found where cytotypes co-occurred, the vast majority of populations were comprised entirely of individuals of a single cytotype. In two tetraploid populations, a single hexaploid was identified out of 25 sampled individuals (populations AZ08-AA and AZ08-AE, Appendix 1). We did not find plants with DNA content values expected for F1 intercytotype hybrids (i.e. triploids and pentaploids; Fig. 1).
Phylogenetic Analyses—The combined cpDNA dataset totaled 3,211 base pairs, including 56 variable and 23 parsimony-informative sites. Most of the five non-coding regions had relatively similar numbers of informative sites (aligned base pairs, variable sites, and informative sites: psbA-trnH, 449, 6, 3; rpl32-trnL, 674, 20, 8; rpl16 intron, 875, 13, 7; rpoB-trnC, 841, 14, 4; petN-trnC, 374, 1, 1). Fifty-one unique haplotypes were recovered from the 112 aligned L. tridentata sequences, with two major haplotypes shared by 44 individuals (Table 1). The most common observed haplotype differences involved single nucleotide polymorphisms; however, 21 indels of varying size were also observed, including a three base pair indel at position 648 in rpoB-trnC that distinguished polyploid cytotypes from diploids (Table 1; Figure 4). Inclusion of cpDNA data from outgroup taxa added 107 parsimony-informative sites and five unique haplotypes.

The ETS/ITS dataset contained 946 nucleotide base pairs, including 24 variable and 10 parsimony informative sites. The ETS and ITS had numbers of informative sites similar to the cpDNA dataset (aligned base pairs, variable sites, and informative sites; ETS, 343, 7, 2; ITS, 603, 17, 8). Inclusion of ETS and ITS data from outgroup taxa added 78 parsimony-informative sites and five unique haplotypes.

Heuristic search of the cpDNA dataset generated a single most parsimonious tree (length = 613 steps; CI = 0.992; RI = 0.974); Bayesian analysis produced an identical tree topology (Fig. 5). There was strong statistical support for the monophyly of North American L. tridentata (congruent nodes with ≥ 99% bootstrap support and ≥ 99% posterior probabilities), which was sister to the South American diploid species L. divaricata (0.313% sequence divergence = 10 indel and base pair differences among 3,247 characters). Heuristic search of the ETS/ITS dataset generated 184 most parsimonious trees (length = 331 steps; CI = 0.936; RI = 0.860); Bayesian analysis produced a tree topology consistent with the heuristic consensus. There was modest statistical support for the monophyly of the North American L. tridentata (convergent nodes with 66% bootstrap support and 55% posterior probabilities; 1.27% sequence divergence = 12 indel and base pair differences among 946 characters). However, the monophyly of L. divaricata and L. tridentata was strongly supported (Fig. 5).

Population Genetic Analyses—Network analysis of L. tridentata cpDNA haplotypes revealed complex genealogical relationships involving two major haplotypes (separated by a single mutational step) and 49 minor haplotypes (separated from major haplotypes by one or a few mutational...
<table>
<thead>
<tr>
<th>Haplotype</th>
<th>psbA</th>
<th>rpl32</th>
<th>rpl16</th>
<th>rpoB</th>
<th>petN</th>
<th>intron</th>
<th>trnC</th>
<th>trnC</th>
<th>trnL</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>-</td>
<td>C</td>
<td>A</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>G</td>
<td>T</td>
<td>A</td>
</tr>
<tr>
<td>b</td>
<td>-</td>
<td>C</td>
<td>A</td>
<td>A</td>
<td>T</td>
<td>T</td>
<td>G</td>
<td>T</td>
<td>A</td>
</tr>
<tr>
<td>c</td>
<td>-</td>
<td>C</td>
<td>A</td>
<td>A</td>
<td>T</td>
<td>T</td>
<td>G</td>
<td>T</td>
<td>A</td>
</tr>
<tr>
<td>d</td>
<td>-</td>
<td>C</td>
<td>A</td>
<td>A</td>
<td>T</td>
<td>T</td>
<td>G</td>
<td>T</td>
<td>A</td>
</tr>
<tr>
<td>e</td>
<td>-</td>
<td>C</td>
<td>A</td>
<td>A</td>
<td>T</td>
<td>T</td>
<td>G</td>
<td>T</td>
<td>A</td>
</tr>
<tr>
<td>f</td>
<td>-</td>
<td>C</td>
<td>A</td>
<td>A</td>
<td>T</td>
<td>T</td>
<td>G</td>
<td>T</td>
<td>A</td>
</tr>
<tr>
<td>g</td>
<td>-</td>
<td>C</td>
<td>A</td>
<td>A</td>
<td>T</td>
<td>T</td>
<td>G</td>
<td>T</td>
<td>A</td>
</tr>
<tr>
<td>h</td>
<td>-</td>
<td>C</td>
<td>A</td>
<td>A</td>
<td>T</td>
<td>T</td>
<td>G</td>
<td>T</td>
<td>A</td>
</tr>
<tr>
<td>i</td>
<td>-</td>
<td>C</td>
<td>A</td>
<td>A</td>
<td>T</td>
<td>T</td>
<td>G</td>
<td>T</td>
<td>A</td>
</tr>
<tr>
<td>j</td>
<td>-</td>
<td>C</td>
<td>A</td>
<td>A</td>
<td>T</td>
<td>T</td>
<td>G</td>
<td>T</td>
<td>A</td>
</tr>
<tr>
<td>k</td>
<td>-</td>
<td>C</td>
<td>A</td>
<td>A</td>
<td>T</td>
<td>T</td>
<td>G</td>
<td>T</td>
<td>A</td>
</tr>
<tr>
<td>l</td>
<td>-</td>
<td>C</td>
<td>A</td>
<td>A</td>
<td>T</td>
<td>T</td>
<td>G</td>
<td>T</td>
<td>A</td>
</tr>
<tr>
<td>m</td>
<td>-</td>
<td>C</td>
<td>A</td>
<td>A</td>
<td>T</td>
<td>T</td>
<td>G</td>
<td>T</td>
<td>A</td>
</tr>
<tr>
<td>n</td>
<td>-</td>
<td>C</td>
<td>A</td>
<td>A</td>
<td>T</td>
<td>T</td>
<td>G</td>
<td>T</td>
<td>A</td>
</tr>
<tr>
<td>o</td>
<td>-</td>
<td>C</td>
<td>A</td>
<td>A</td>
<td>T</td>
<td>T</td>
<td>G</td>
<td>T</td>
<td>A</td>
</tr>
<tr>
<td>p</td>
<td>-</td>
<td>C</td>
<td>A</td>
<td>A</td>
<td>T</td>
<td>T</td>
<td>G</td>
<td>T</td>
<td>A</td>
</tr>
<tr>
<td>q</td>
<td>-</td>
<td>C</td>
<td>A</td>
<td>A</td>
<td>T</td>
<td>T</td>
<td>G</td>
<td>T</td>
<td>A</td>
</tr>
<tr>
<td>r</td>
<td>-</td>
<td>C</td>
<td>A</td>
<td>A</td>
<td>T</td>
<td>T</td>
<td>G</td>
<td>T</td>
<td>A</td>
</tr>
<tr>
<td>s</td>
<td>-</td>
<td>C</td>
<td>A</td>
<td>A</td>
<td>T</td>
<td>T</td>
<td>G</td>
<td>T</td>
<td>A</td>
</tr>
<tr>
<td>t</td>
<td>-</td>
<td>C</td>
<td>A</td>
<td>A</td>
<td>T</td>
<td>T</td>
<td>G</td>
<td>T</td>
<td>A</td>
</tr>
<tr>
<td>u</td>
<td>-</td>
<td>C</td>
<td>A</td>
<td>A</td>
<td>T</td>
<td>T</td>
<td>G</td>
<td>T</td>
<td>A</td>
</tr>
<tr>
<td>v</td>
<td>-</td>
<td>C</td>
<td>A</td>
<td>A</td>
<td>T</td>
<td>T</td>
<td>G</td>
<td>T</td>
<td>A</td>
</tr>
<tr>
<td>w</td>
<td>-</td>
<td>C</td>
<td>A</td>
<td>A</td>
<td>T</td>
<td>T</td>
<td>G</td>
<td>T</td>
<td>A</td>
</tr>
<tr>
<td>x</td>
<td>-</td>
<td>C</td>
<td>A</td>
<td>A</td>
<td>T</td>
<td>T</td>
<td>G</td>
<td>T</td>
<td>A</td>
</tr>
<tr>
<td>y</td>
<td>-</td>
<td>C</td>
<td>A</td>
<td>A</td>
<td>T</td>
<td>T</td>
<td>G</td>
<td>T</td>
<td>A</td>
</tr>
</tbody>
</table>

2012 LAPORT ET AL.: PHYLOGENY OF CREOSOTE BUSH 157
The major haplotypes (a, b) occurred at frequencies of 24.1% and 15.2%, respectively; frequencies of minor haplotypes were 3.6% (c, d), 2.7% (e–h), 1.8% (i–m), and 0.9% (n–y’) (Table 1; Fig. 4). While haplotypes were widely shared among tetraploid and hexaploid populations (major haplotype b and associated minor haplotypes), diploid populations harbored a distinct group of haplotypes (major haplotype a and associated minor haplotypes) (Table 1; Fig. 4). Haplotypes recovered from the sand-dune endemic L. tridentata var. arenaria were broadly distributed throughout the network, and similar or identical to those of tetraploid and hexaploid cytotypes of L. tridentata (Fig. 4).

AMOVA indicated that the majority of cpDNA sequence variation in L. tridentata was found within populations (65.8 and 67.1%) rather than among populations (4.5 and 10.9%) (Table 2). Variance was nonetheless associated significantly with population groupings defined by ploidy or desert province (22.0 and 29.7%; \( p < 0.001 \)) (Table 2). Pairwise differences among cpDNA haplotypes exhibited a unimodal distribution (mean = 2.3 differences) that did not differ significantly from a stepwise population expansion model for a haploid locus (\( p < 0.001 \)) (Fig. 6). Estimates of demographic parameters and confidence intervals (\( a = 0.05 \)) were as follows: modal value, \( t \) = 2.383 [1.510–2.799]; 2Nu before expansion, \( q_0 \) = 0 [0–1.201]; 2Nu after population expansion, \( q_1 \) = 7,730.000 [32.281–9,097.188]. The observed value of Tajima’s D was -2.432. Coalescent simulations based on a neutral model (no recombination, constant population size) had a 95% confidence interval of [-1.595, 1.863]. Thus, the excess occurrence of minor cpDNA haplotypes in North American creosote bush was unlikely to have been recovered by chance (\( p < 0.001 \)).

Mismatch distributions and estimates of demographic parameters and confidence intervals (\( a = 0.05 \)) were also calculated for each cytotype individually. For diploids, the mean number of pairwise differences = 1.3, Tajima’s D was -2.226 ([−1.724, 1.921], \( p = 0.002 \)); for tetraploids the mean number of pairwise differences = 2.4, Tajima’s D was -2.362 ([−1.763, 1.836], \( p < 0.001 \)); for hexaploids the mean number of pairwise differences = 2.4, Tajima’s D was -1.917 ([−1.772, 1.739], \( p = 0.016 \)).

**Discussion**

**Cytogeography**—The cytotypes of L. tridentata span large areas of western North America, and occur in topographically complex terrain (Figs. 2, 3). We found that the distribution of diploid, tetraploid, and hexaploid populations generally correspond to the Chihuahuan, Sonoran, and Mojave Deserts, as reported in classical studies (Yang 1967, 1968, 1970; Yang and Lowe 1968; Barbour 1969). The boundary between diploid and tetraploid cytotypes seems particularly well defined, only two populations of diploids were found further west than the easternmost populations of tetraploids, for example. Populations of tetraploids and hexaploids, however, were broadly mixed throughout central and western Arizona. Hexaploid populations were found further east than previously reported, and
we observed sympatric occurrence of tetraploid and hexaploid cytotypes in two study sites (Fig. 2).

Cytotype distributions may reflect historical barriers that formerly separated the chromosomal races, but have since been obscured. For example, the floristic endemism of the Chihuahuan and Sonoran Deserts suggests their formation has occurred independently and over an extended evolutionary period (Betancourt et al. 1990). In contrast, there is considerable disagreement about the floristic distinctiveness of the Mojave and Sonoran Deserts (McLaughlin 1986; Turner 1994). Even in the western portion of the range, however, co-occurrence of cytotypes was rarely observed within single populations. This finding may reflect minority cytotype exclusion, a frequency-dependent selection caused by reproductive barriers between ploidy levels (Hagberg and Ellerström 1959; Levin 1975, 1983). It is also proposed, however, that the cytotypes of L. tridentata have alternate phenological, life-history, and physiological attributes, and thus may be “pre-adapted” to occur in different environmental conditions (Lumaret 1988; Thompson and Lumaret 1992; De Soyza et al. 1997; Husband and Schemske 1998; Ignace and Huxman 2009; Maherali et al. 2009).

**Origin of North American Larrea**—Phylogenies generated from cpDNA and nrDNA support the monophyly of L. tridentata (Fig. 5). North American Larrea is sister to the South American L. divaricata, confirming the close evolutionary relationship of these taxa inferred from morphological, cytogenetic, and allozyme studies. Colonization of North America appears to have happened on a single occasion, or perhaps on a few occasions from closely-related South American sources, but further sampling of South American populations is needed to resolve these possibilities. The

### Table 2. Analysis of molecular variance for cpDNA haplotypes recovered from North American L. tridentata, with population groupings based on geography (desert) and ploidy level.

<table>
<thead>
<tr>
<th>Analysis</th>
<th>Percentage of variation</th>
<th>Fixation index</th>
<th>Statistical significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Groups defined by geography</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Among groups</td>
<td>22.0</td>
<td>$\Phi_{CT} = 0.22$</td>
<td>$p &lt; 0.001$, df = 2</td>
</tr>
<tr>
<td>Among populations within groups</td>
<td>10.9</td>
<td>$\Phi_{SC} = 0.14$</td>
<td>$p &lt; 0.001$, df = 37</td>
</tr>
<tr>
<td>Within populations</td>
<td>67.1</td>
<td>$\Phi_{ST} = 0.33$</td>
<td>$p &lt; 0.001$, df = 72</td>
</tr>
<tr>
<td>Groups defined by ploidy level</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Among groups</td>
<td>29.7</td>
<td>$\Phi_{CT} = 0.30$</td>
<td>$p &lt; 0.001$, df = 2</td>
</tr>
<tr>
<td>Among populations within groups</td>
<td>8.1</td>
<td>$\Phi_{SC} = 0.08$</td>
<td>$p &lt; 0.001$, df = 37</td>
</tr>
<tr>
<td>Within populations</td>
<td>65.8</td>
<td>$\Phi_{ST} = 0.34$</td>
<td>$p &lt; 0.001$, df = 72</td>
</tr>
</tbody>
</table>
taxonomic status of *L. tridentata* vs. *L. divaricata* is complicated by the species’ allopatric distributions and partial reproductive compatibility in experimental crosses (Yang et al. 1977, 2000). However, because of the DNA sequence differences observed between *L. tridentata* and *L. divaricata*, and the absence of polyploid populations in *L. divaricata* (Hunziker et al. 1977), we favor recognition of the taxa as separate species rather than as conspecific. Nonetheless, sequence divergence between *L. tridentata* and *L. divaricata* is low, approximately 0.313% (10 changes among 3,198 base pairs) for cpDNA and 1.27% (12 changes among 946 base pairs) for ITS/ITS.

Because of its ecological dominance in North American deserts, there has been longstanding interest in dating the arrival of creosote bush to northern latitudes (Barbour 1969; Hunziker et al. 1972, 1977; Betancourt et al. 1990; Lia et al. 2001). Individual clones in the Mojave Desert have been estimated to be several thousand years old (Sternberg 1976; Vasek 1980), and macrofossils recovered from packrat (*Neotoma*) middens date *L. tridentata* to the last glacial maximum (ca. 18,700 ybp; Betancourt et al. 1990; Hunter et al. 2001), while palynological evidence in lake sediments of Death Valley dates to 47,000–109,000 ybp (Woolfendon 1996; Bader 2000). Most botanists suspect that creosote bush has occurred in North American deserts far longer than is documented by the fossil record, however, and have turned to molecular data to estimate divergence time between *L. tridentata* and *L. divaricata* (Hunter et al. 2001; Lia et al. 2001).

DNA substitution rates in flowering plants are on the order of 0.1–0.3% per million years for non-coding cpDNA (Wolfe et al. 1987; Muse 2000) and 0.1–0.8% per million years for non-coding nrDNA (ITS; Kay et al. 2006). Sequence divergence reported here suggests that colonization of North America occurred in the Pleistocene or late Pliocene (cpDNA pairwise divergence = 0.313%, 0.52–1.57 mya; ITS pairwise divergence = 0.672%, 0.42–3.36 mya). This timeframe is younger than suggested by Lia et al. (2001; 4.2–8.4 mya) based on the chloroplast gene *rbcL* and similar to that of Cortes and Hunziker (1997; 0.6–1.2 mya) based on allozymes. An alternate approach to dating the amphitropical dispersal of *Larrea* focuses on North American herbivore and pollinator groups that specialize on creosote bush as a food resource. For example, gall midges of the *Asphondylia auripila* group (Cecidomyiidae) are *L. tridentata* specialists (Waring and Price 1989) and were recently shown by Joy and Crespi (2007) to represent an adaptive radiation via divergence in host feeding location. Based on a mtDNA divergence rate of 1.1–1.2% per million years (Tamura 1992; Brower 1994), the observed COI pairwise divergence between *Larrea*-specialist and non-*Larrea* specialists of *Asphondylia* (9.3%) puts a lower bound of 3.88–4.23 mya for the arrival of creosote bush in North America (Joy and Crespi 2007).

Rates of molecular evolution are too variable to provide precise estimates of divergence times in the absence of fossil calibrations, and botanists may never know the exact time that *Larrea* arrived in North America. Nonetheless, estimates based on a variety of data sources and criteria suggest a relatively recent origin between the mid-Pleistocene and late Pliocene. This conclusion is also supported by population genetic analyses (see below).

**Phylogeography and Genetic Structure**—The phylogeography of *L. tridentata* is suggestive of recent and rapid demographic expansion from a small founding population. We found only two common cpDNA haplotypes (a and b) that together comprised 39.3% of all recovered sequences (Fig. 4). Most cpDNA haplotypes were unique or shared by a few individuals (Table 1) and, consistent with prior work in diploid *L. tridentata* (Duran et al. 2005), there was little genetic structure associated with either geography or population boundaries (Table 2). Genealogical relationships of cpDNA haplotypes were star-shaped (Figs. 4, 5) and exhibited the unimodal DNA mismatch distribution (Fig. 6) that is characteristic of expanding populations (Rogers and Harpending 1992).

The clearest evidence of population genetic structure was associated with ploidy, as a tri-nucleotide indel in the *rpoB-trnC* intergenic spacer distinguished diploid cytotypes from tetraploid and hexaploid cytotypes (Fig. 4). In contrast, we found no consistent sequence differences between tetraploid and hexaploid cytotypes. These findings suggest that tetraploidy evolved a single time among the North American creosote bush, but that hexaploids either (i) have formed recurrently in tetraploid populations; (ii) have extensively hybridized with their tetraploid progenitors; or (iii) have had insufficient time to develop characteristic DNA sequences (Soltis and Soltis 1999; Soltis et al. 2007). The rate of hexaploid formation in tetraploid populations far exceeds the rate of tetraploid formation in diploid populations, and as a rule, pentaploid hybrids are more fertile than triploid hybrids (Ramsey and Schemske 1998, 2002; Ramsey 2007). Thus, the paucity of sequence differences between tetraploid and hexaploid cytotypes is not unexpected.

One complication to the aforementioned interpretation is that the *rpoB-trnC* indel appears to be a derived insertion: South American species (*L. cuneifolia, L. divaricata*, and *L. nitida*) and polyploid cytotypes of the North American *L. tridentata* share the deletion (Table 1). It is unclear if the occurrence of the *rpoB-trnC* deletion in polyploid *L. tridentata* represents a character reversion, i.e. backwards mutation to the ancestral state, as exhibited by the South American *L. divaricata*, or if the indel was polymorphic in ancestral populations of *L. tridentata* with the insertion subsequently fixing only in the diploid lineage. Furthermore, two tetraploid individuals have the insertion (Fig. 4). One of these
individuals (haplotype a’, Fig. 4) was sampled near diploid populations (AZ07-S, Appendix 1) and was probably generated by intercytotype gene flow or neopolyploidy. However, the second individual (haplotype a, Fig. 4) was sampled near Hermosillo, Mexico (SN07-B, Appendix 1), several hundred kilometers from known diploid populations. Backward mutation may be the best explanation for the existence of this individual. Additional population sampling of L. divaricata and diploid L. tridentata may clarify the origin of the rpoB-trnC indel character. Regardless of its history, the indel may be a useful marker for evaluating hybridization and introgression among modern-day diploid and polyploid creosote bush populations.

Larrea tridentata var. arenaria—This subspecific taxon was defined on the basis of morphology and endemism to the Algodones Dunes, there was no prior investigation of its ploidy level or evolutionary history (Benson and Darrow 1981; Felger 2000). Our analyses reveal var. arenaria to be tetraploid and to harbor cpDNA and ITS/ITS haplotypes identical to those of tetraploid and hexaploid L. tridentata (DC.) Coville (Table 1; Fig. 4). Larrea tridentata var. arenaria resides at the western range boundary of tetraploid L. tridentata (DC.) Coville, and moreover, the nearest populations of L. tridentata (DC.) Coville to the Algodones Dunes are hexaploid (Fig. 2).

In a few sites, tetraploid L. tridentata var. arenaria and hexaploid L. tridentata (DC.) Coville occur in sympatry (R. Laport unpubl. data). Nonetheless, specimens of L. tridentata var. arenaria are distinguished from L. tridentata (DC.) Coville by a combination of stem, leaf, and floral traits, and are readily identifiable in the field (R. Laport pers. obs.). Based on these findings, we suggest that L. tridentata var. arenaria had a recent origin as a dune-adapted ecotype of the more widespread L. tridentata (DC.) Coville, and warrants continued recognition at the subspecific level.

Acknowledgments. We thank S. Laport, M. Laport, J. Ng, M. Castiglione, and M. Strangas for assistance in the field; E. Fox, A. Green, O. Hardy, E. Reiss, L. Hatemi, H. Pullman, L. Widener, and T. Ramsey for assistance in the laboratory; and T. Van Devender (Arizona-Sonora Desert Museum) and J. Takara (Rancho Santa Ana Botanic Garden) for provision of outgroup samples. A. Green, J. Ng, and T. Ramsey provided helpful comments on a draft of this manuscript. This research was supported by an NSF DDIG grant (DEB-1010738), a Torrey Botanical Society fellowship, and a Botanical Society of America student research grant to RL, as well as an NSF CAREER grant (DEB-0953551) and a University of Rochester Provost Award for Multidisciplinary Research to JR.

Literature Cited
Larrea tridentata (DC.) Coville var. arenaria (4x); U. S. A.: California: Imperial Co.: CA07-A (32° 44' N, 114° 53' W), R. Laport 766262; JF266819, JF266820; JF267052, JF267053; JF266942, JF266943; JF267169, JF267170, JF267171, JF267172; JF267290, JF267291, JF267292.

Larrea divaricata Cav. (2x); U. S. A.: Arizona: Pima Co.: DIVA (Cultivated at University of Arizona), R. Laport 766263; JF266821; JF267055; JF266938, JF267174; JF267298, JF267299, JF267300.

Larrea cuneifolia Cav. (4x); U. S. A.: Arizona: Pima Co.: CUNE (Cultivated at University of Arizona), R. Laport 766263; JF266821; JF267055; JF266938, JF267174; JF267298, JF267299, JF267300.

Larrea nitida Cav. (2x); U. S. A.: California: Los Angeles Co.: RSABG (Cultivated at Rancho Santa Ana Botanic Gardens), R. Laport 766264; JF266824, JF267056; JF266940, JF267173, JF267291, JF267294, JF267295, JF267296, JF267297.

Guaiacum coulteri A. Gray (–); U. S. A.: Arizona: Pima Co.: GUAIA (Cultivated at Arizona-Sonora Desert Museum), R. Laport 766266; JF266823, JF267058, JF266941; JF267175; JF267292; JF267293.