

# Asymmetric hybridization and gene flow between Joshua trees (*Agavaceae: Yucca*) reflect differences in pollinator host specificity

TYLER N. STARR,\* KATHERINE E. GADEK,\* JEREMY B. YODER,†‡ RAMONA FLATZ\* and CHRISTOPHER I. SMITH\*

\*Department of Biology, Willamette University, 900 State Street, Salem, OR 97301 USA, †Department of Biological Sciences, University of Idaho, Moscow, ID 83844, USA

## Abstract

The angiosperms are by far the largest group of terrestrial plants. Their spectacular diversity is often attributed to specialized pollination. Obligate pollination mutualisms where both a plant and its pollinator are dependent upon one another for reproduction are thought to be prone to rapid diversification through co-evolution and pollinator isolation. However, few studies have evaluated the degree to which pollinators actually mediate reproductive isolation in these systems. Here, we examine evidence for hybridization and gene flow between two subspecies of Joshua tree (*Yucca brevifolia brevifolia* and *Yucca brevifolia jaegeriana*) pollinated by two sister species of yucca moth. Previous work indicated that the pollinators differ in host specificity, and DNA sequence data suggested asymmetric introgression between the tree subspecies. Through intensive sampling in a zone of sympatry, a large number of morphologically intermediate trees were identified. These included trees with floral characters typical of *Y. b. jaegeriana*, but vegetative features typical of *Y. b. brevifolia*. The opposite combination—*Y. b. brevifolia* flowers with *Y. b. jaegeriana* vegetative morphology—never occurred. Microsatellite genotyping revealed a high frequency of genetically admixed, hybrid trees. Coalescent-based estimates of migration indicated significant gene flow between the subspecies and that the direction of gene flow matches differences in pollinator host fidelity. The data suggest that pollinator behaviour determines the magnitude and direction of gene flow between the two subspecies, but that specialized pollination alone is not sufficient to maintain species boundaries. Natural selection may be required to maintain phenotypic differences in the face of ongoing gene flow.

**Keywords:** angiosperms, coevolution, hybridization, population genetics—empirical, speciation, species interactions

Received 24 May 2012; revision received 27 September 2012; accepted 2 October 2012

## Introduction

There are at least 200 000 species of angiosperms (Scotland & Wortley 2003), which together account for more than 15% of the 1.2 million described species on Earth (Mora *et al.* 2011). Despite their extraordinary diversity,

the angiosperms are geologically young, having originated in the mid-Jurassic to early Cretaceous (Bell *et al.* 2010). The mechanisms underlying their explosive radiation have long been a puzzle for evolutionary biologists (Friedman 2009), but many have argued that the reliance of flowering plants on specialized pollinators represents a key innovation that promoted their diversification (Darwin 1876; Grant 1949; Burger 1981; Armbruster & Muchhala 2009; but see Stebbins 1981). Several lines of evidences suggest an association between specialized pollination and plant diversity. The

Correspondence: Christopher Smith, Fax: 1 503 375 5425;

E-mail: csmith@willamette.edu

‡Present address: Department of Plant Biology, University of Minnesota, St Paul, MN USA

diversification of flowering plants was coincident in time with the diversification of many pollinating insect groups (Grimaldi 1999), and plants with specialized pollination are significantly more speciose than generalists (Dodd *et al.* 1999; Sargent 2004).

It is unclear, however, why this association between specialized pollination and species diversity exists. The simplest explanation may be that specialized pollination increases the rates of speciation by promoting reproductive isolation between plants that rely on different pollinators, whether during initial speciation or following secondary contact after divergence in allopatry. Pollinators might act to prevent or reduce gene flow between incipient species if there is a high degree of host specificity, or different pollinators might impose divergent selection on different hosts. A third possibility is that divergent selection and reduced gene flow might act together. However, as Armbruster & Muchhala (2009) argued in a recent review, there are several possible alternatives to the hypothesis that specialized pollinators increase the rates of speciation. Specialized pollination could instead act to reduce extinction rates by allowing successful pollination at low population densities. Pollinator specialization could even be a side effect of high species diversity itself if competition for pollinators by closely related species results in resource partitioning and character displacement (Armbruster & Muchhala 2009).

Obligate pollination mutualisms, such as those between yuccas and yucca moths and between figs and fig wasps where both the plant and pollinator are dependent on one another for reproduction, are often cited as a case where pollinator isolation and selection are particularly likely to produce rapid diversification (Kiestler *et al.* 1984). Within these systems, there is often a high degree of specificity, with each host pollinated by one or several closely related species of pollinator, and each pollinator species specialized on one or a few species of host (Janzen 1979; Weiblen 2002; Pellmyr 2003; but see Machado *et al.* 2005). Plants and animals involved in obligate pollination mutualism also often exhibit matching between morphological features of the flowers and their pollinators (Weiblen 2004; Godsoe *et al.* 2008), which could indicate ongoing co-evolution between them [though Nuismer *et al.* (2010) suggest alternative explanations]. Finally, some mathematical models suggest that the combination of mutual dependence, specificity and reciprocal natural selection is likely to produce rapid differentiation between allopatric populations (Kiestler *et al.* 1984; but see Yoder & Nuismer 2010).

Despite the strong expectations of rapid diversification and speciation in obligate pollination mutualisms, there have been few intensive investigations examining

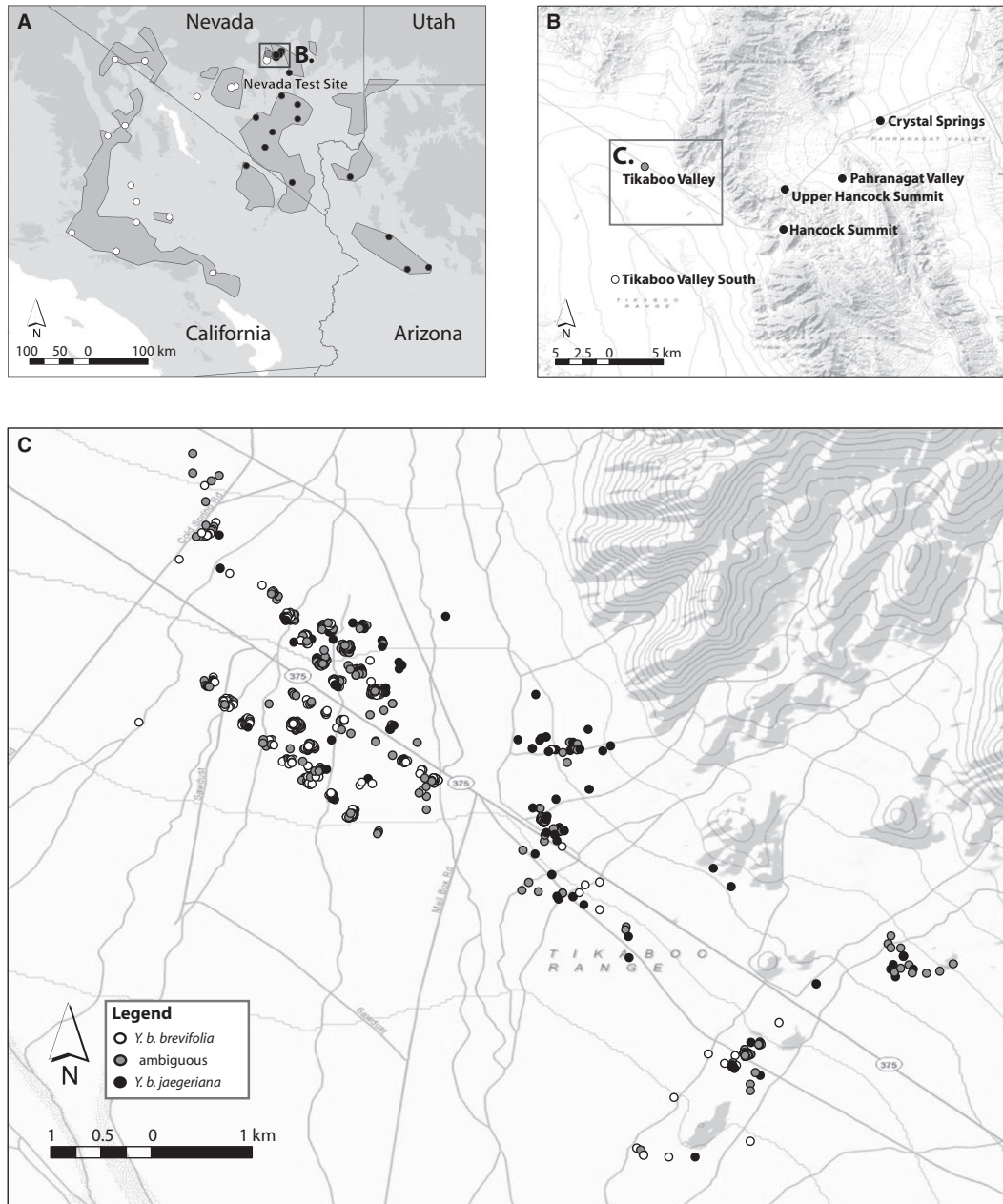
the role of pollinator isolation in speciation within these systems. Recent work on pollination in Joshua trees (*Yucca brevifolia*) suggests that co-evolution with their pollinator might have produced distinct morphotypes that may be different species (Lenz 2007; Godsoe *et al.* 2008). Joshua trees, like all yuccas, produce bisexual flowers that are pollinated exclusively by yucca moths (Lepidoptera: Prodoxidae). The moths lay their eggs inside the yucca flower using a blade-like ovipositor to penetrate the style and then deliberately deposit pollen onto the stigmatic surface using uniquely derived, tentacle-like mouth parts. This active pollination ensures that the flower will produce seeds to nourish the moth's caterpillars (Pellmyr 2003). Across their range, Joshua trees are associated with two species of pollinators, *Tegeticula syntheticca* in the west and *Tegeticula antithetica* in the east (Pellmyr & Segraves 2003; Godsoe *et al.* 2008), which differ significantly in body size and in the length of the female's ovipositor (Pellmyr & Segraves 2003). Populations of trees associated with each pollinator differ significantly in vegetative traits (Fig. 1), but statistically the greatest difference is in the length of the stylar canal—the path through which the female yucca moth inserts her ovipositor when laying eggs (Godsoe *et al.* 2008). On the basis of these differences, some taxonomists have suggested that the pollinator-associated subspecies should be recognized as distinct species (Lenz 2007). The validity of this designation is not yet certain, and here, we conservatively refer to the two morphotypes as subspecies: *Yucca brevifolia brevifolia*, pollinated by *T. syntheticca*, and *Yucca brevifolia jaegeriana* pollinated by *T. antithetica*.



**Fig. 1** *Yucca brevifolia brevifolia* (left) and *Yucca brevifolia jaegeriana* (right) growing in sympatry in Tikaboo Valley, Nevada. Image is reversed from its original orientation for consistency of presentation. *Y. b. brevifolia* is typically arborescent, with fewer branches and longer leaves and first branching point at more than 1 m in height. *Y. b. jaegeriana* is typically shrub-like, with more branches, shorter leaves and first branching at a height of <1 m.

Previous studies have also argued that pollinator isolation may have caused the differentiation of the two Joshua tree subspecies (Godsoe *et al.* 2008; Smith *et al.* 2008), but there is as yet only fragmentary evidence supporting this suggestion, and it is unclear whether

the subspecies are reproductively isolated. The two pollinators and tree subspecies occur primarily in allopatry (Fig. 2), but they occur in sympatry at one site in central Nevada. In the zone of sympatry, the two subspecies flower in synchrony and pollinator specificity



**Fig. 2** Location of study sites. Labelled populations were sampled for both morphology and genetic data. (A) Range of the Joshua tree (shaded) showing the location of populations sampled for studies of morphology. Background shading shows elevation, with a contour interval of 1500 m. Populations of *Yucca brevifolia brevifolia* (pollinated by *Tegeticula synthetica*) are shown in white, and populations of *Y. b. jaegeriana* (pollinated by *T. antithetica*) are shown in black. Box indicates the location of detailed maps. Joshua tree range is after Godsoe *et al.* (2009). (B) Distribution of study localities within Tikaboo Valley, Nevada; boxed area shows location of hybrid zone. Contour interval is 100 m. (C) Distribution of sampled trees. Each point is shaded according to the logistic regression score based on vegetative morphology. Contour interval is 50 m.

breaks down. At this site, *T. synthetica* visits both *Y. b. brevifolia* and *Y. b. jaegeriana*, although it shows preference for its native host and seems never to successfully produce larvae on *Y. b. jaegeriana* (Smith *et al.* 2009). *T. antithetica*, on the other hand, visits both tree subspecies indiscriminately in the zone of sympatry and successfully oviposits on both tree types (Smith *et al.* 2009).

Finally, prior work in this system has identified suggestions of asymmetric, ongoing gene flow between the two tree subspecies. Analysis of DNA sequence data from the chloroplast (which is maternally inherited through seeds) finds that although there is effectively no gene flow from *Y. b. jaegeriana* into *Y. b. brevifolia*, there appears to be ongoing movement of chloroplast genomes from *Y. b. brevifolia* into *Y. b. jaegeriana* (Smith *et al.* 2008). While the very low rate of sequence evolution in the chloroplast makes it difficult to distinguish contemporary gene flow from ancestral polymorphism, it is intriguing that this pattern of asymmetrical plastid introgression mirrors the differences in pollinator behaviour described above. Thus, Smith and colleagues (2009) speculated that the lower host fidelity in *T. antithetica* caused a chloroplast capture event, where nuclear gene flow through pollen from *Y. b. jaegeriana* into *Y. b. brevifolia* has resulted in the introgression of the *Y. b. brevifolia* chloroplast genome into the *Y. b. jaegeriana* genetic background.

Here, we evaluate the extent of reproductive isolation between *Y. b. brevifolia* and *Y. b. jaegeriana* and test the hypothesis that differences in host fidelity between *T. antithetica* and *T. synthetica* have produced asymmetric rates of nuclear gene flow between the two Joshua tree subspecies. Using morphological and genetic data, we show that where the two pollinator species co-occur, a large, statistically significant fraction of trees appear to be of hybrid origin. In addition, genetic analyses indicate higher rates of nuclear gene flow from *Y. b. jaegeriana* into *Y. b. brevifolia*, which may explain the introgression of *Y. b. brevifolia* chloroplast genomes into *Y. b. jaegeriana*. Finally, we discuss the relative roles that moth behaviour and natural selection may play in mediating species boundaries in this system.

## Methods

To identify the potential hybrids between *Yucca brevifolia brevifolia* and *Yucca brevifolia jaegeriana*, we completed extensive surveys of trees in Tikaboo Valley, NV, where the two pollinator species occur in sympatry, and adjoining 'allopatric' populations where only *Tegeticula antithetica* or *Tegeticula synthetica* occur (Fig. 2). We collected vegetative, floral and genetic data from trees within these

populations and used a suite of statistical and population genetic methods to determine whether morphologically intermediate or genetically admixed individuals exist and whether they occur with greater frequency where the pollinators co-occur.

## Sampling

In May of 2009, and April of 2010, we established 38 study plots arrayed along eight linear transects. Transects were established haphazardly, in an effort to sample both the maximum density of trees and variation in altitude and soil type within the valley. Within each plot, we collected vegetative data from all trees that were at least 1 m in height or had more than one terminal branch. Because the most informative vegetative characters for distinguishing the tree subspecies—height at first branching and number of branches—are characteristics of mature trees, short, unbranched trees could not be unambiguously assigned to a particular subspecies and were therefore excluded. Each tree was measured for five vegetative traits that have been shown to differ significantly between the two Joshua tree subspecies (Godsoe *et al.* 2008): the length and width of one leaf, number of branches, tree height and height at first branching. A small (~1 g) piece of leaf tissue was sampled from each tree, labelled as to the tree of origin and stored in liquid nitrogen pending DNA extraction and genotyping.

In both study years, flowering frequency was very low (~3%), so to obtain data on floral morphology, we supplemented the data collected using the stratified design above with an opportunistic sample that included any trees with flowers encountered during extensive searches of the contact zone. For these trees, we measured the five vegetative traits (above) and collected 3–4 flowers. Flowers were stored in 70% ethanol for subsequent dissection. From each flower, we measured eight traits known to differ significantly between tree subspecies: petal length, petal width, petal diagonal length, pistil length, pistil width, style length, style base width and style midpoint width (Godsoe *et al.* 2008). Two floral measurements (petal diagonal length and style midpoint width) varied significantly between observers and were excluded from further morphological analysis.

Together all surveys and data collection yielded vegetative data from 662 trees, floral data from 117 trees and matched floral and vegetative data from 91 trees within the Tikaboo Valley contact zone. (In some analyses, 13 trees with missing observations were eliminated leaving 78 trees total.) These data were combined with vegetative and floral morphology from 298 trees sampled outside the contact zone at study sites across the range of *Y. brevifolia s. l.*, including previously

published data from 207 trees (Godsoe *et al.* 2008) and unpublished from 81 trees.

#### *Analysis of morphological data*

Vegetative and floral data sets were analysed by logistic regression to classify trees as *Y. brevifolia brevifolia*, *Y. b. jaegeriana* or 'ambiguous'. From each data set, a subset of trees from populations outside the contact zone were selected at random to serve as training data; this subset included 60% of the trees from each subspecies. For the remaining 40% of trees from allopatric *Y. b. brevifolia* and *Y. b. jaegeriana* populations and for all trees from the Tikaboo Valley contact zone, the logistic regression score (i.e. the probability of assignment as *Y. b. brevifolia*) was calculated based on these training data. Trees with logistic regression scores of  $<0.05$  were classified as *Y. b. jaegeriana*; trees with logistic regression scores of  $>0.95$  were classified as *Y. b. brevifolia*. Trees with scores  $\geq 0.05$  and  $\leq 0.95$  were considered ambiguous. This criterion allowed us to identify both potential hybrids and aberrant individuals of each subspecies and thus to statistically compare the frequency of morphologically ambiguous trees in allopatry and in sympatry.

Multicollinearity is a well-known problem in multiple linear regression, and collinearity among predictor variables may also lead to unstable estimates of model parameters in logistic regression (Aguilera *et al.* 2006). However, the standard remedies for this problem—dropping one or more correlated variables from the analysis or combining variables into a single index using principal component analyses—may also have undesirable statistical consequences (O'Brien 2007) and should therefore be applied only when necessary. To evaluate multicollinearity in our morphological data, we calculated the variance inflation factors (VIF), a measure of the degree of multicollinearity in a least-squares regression, for each of our predictor variables. VIF of  $>10$  are generally considered to denote a harmful degree of collinearity (Menard 1995). In the vegetative data, no VIF values were  $>2$ , and the average VIF was 1.181 (see Supporting information). In the floral data, VIF values ranged from 1.43 to 3.13, with an average of 2.05. These values suggest low levels of multicollinearity. We therefore conducted the logistic regression on the raw data and did not eliminate correlated variables; regression coefficients and significance values for each predictor variable are provided as Supporting information.

To address the question of whether there is an excess of morphologically ambiguous trees in the contact zone, we counted the number of trees classified as ambiguous from Tikaboo Valley and from among the trees sampled from allopatric populations that had not previously

been used as training data to develop the logistic regression. We then used Fisher's exact test to determine whether the number of ambiguous trees within the contact zone was significantly greater than expected, given the frequency with which ambiguous trees occurs in populations outside the zone of sympatry.

#### *Genetic analysis*

To evaluate the population genetic evidence of admixture between tree subspecies, we genotyped a subset of the trees surveyed for morphological data (above). From within Tikaboo Valley, 255 trees were selected for genotyping, including all trees with floral data and a subset of trees with only vegetative data. To compare genetic variation in the contact zone with that seen in allopatric populations, we genotyped an additional 42 *Y. b. brevifolia* and 77 *Y. b. jaegeriana* trees from five populations adjoining the contact zone (Fig. 2). Whole genomic DNA was extracted from leaf tissue samples using Qiagen DNeasy Plant Mini Kits (Qiagen GmbH, Hilden, Germany). Prior work showed that although there is a range-wide pattern of chloroplast introgression from *Y. b. brevifolia* into *Y. b. jaegeriana*, within Tikaboo Valley and several adjoining populations all trees carry the same cpDNA haplotype (Smith *et al.* 2008). Thus, to estimate the rates of gene flow and introgression within this zone of sympatry, each sample was genotyped for ten microsatellite loci described previously (Flatz *et al.* 2011). Multiplex reaction conditions followed those described in Flatz *et al.* (2011). PCRs were analysed using capillary electrophoresis on a 3130XL Genetic Analyzer (Perkin-Elmer, Applied Biosystems, Foster City, CA, USA), using LIZ 500 size standard (Perkin-Elmer, Applied Biosystems). Resulting electropherograms were visualized and converted to diploid genotypes using automated allele calling implemented GeneMarker v. 4.0 (SoftGenetics LLC, State College, PA, USA). All automated genotyping calls were re-checked manually. Two loci described by Flatz *et al.* (2011), yb05 and yb20, were excluded from this study because they could not be reliably scored.

Prior to completing analyses of population structure and admixture, genotype frequencies within subpopulations at each locus were tested for departures from Hardy-Weinberg equilibrium in GenoDive version 2.0b21 (Meirmans & Van Tienderen 2004) using a heterozygosity-based method (Nei's  $G_{TS}$ ) and 1000 permutations to test for significance. After Bonferroni correction, six loci showed significant departures from HWE within the Tikaboo Valley hybrid zone (see Supporting information). However, because this population is a contact zone between the two subspecies, population structure is expected to result in reductions

in heterozygosity (Wahlund 1928; Hartl & Clark 1989). Excluding the contact zone, two loci (yb12, and yb21) showed significant departures from HWE in at least one subpopulation and one (yb08) showed departures in three subpopulations. To evaluate the consequence of including these data, we estimated the probability of assignment using STRUCTURE (see below) either including or excluding each of these loci in turn. The results of analyses that excluded each of these loci were qualitatively indistinguishable from those that included them, and parameter estimates were on average 95% correlated with analyses that included or did not include these loci. We therefore included all ten loci that could be scored reliably in subsequent analyses.

To evaluate the extent of population structure within and between tree subspecies, we analysed the microsatellite data using an analysis of molecular variation (AMOVA), excluding the contact zone, implemented in GenoDive version 2.0b21 (Meirmans & Van Tienderen 2004). Significance was tested using 999 permutations.

To estimate the rates of hybridization and admixture between subspecies, we assigned trees to populations based on multilocus microsatellite genotypes using Bayesian clustering algorithms implemented in STRUCTURE version 2.3.3 (Pritchard *et al.* 2000). STRUCTURE uses multilocus genetic data to estimate either the proportion of genetic material ( $q$ ), that is, derived from each population or the posterior probability that an individual belongs to a given population, given the data and assuming  $K$  distinct populations. We first estimated the proportion of each individual's ancestry belonging to different populations using a simple admixture model. We next estimated the probability of assigning each individual to a population, given the data, using prior population information from allopatric *Y. b. brevifolia* or *Y. b. jaegeriana* populations to assist in clustering and setting the MIGRPRIOR parameter (the probability that an individual in the pure populations is an immigrant) to 0.001. In both analyses, we used a burn-in period of 20 000 steps followed by a 200 000 step MCMC, independent allele frequencies and the admixture model for individuals. Analyses were conducted assuming one to five clusters ( $K = 1-5$ ), repeating each analysis three times to evaluate convergence between runs. To estimate the best-fit number of distinct populations ( $K$ ), given the data, we calculated the values for  $\Delta K$ , as described by (Evanno *et al.* 2005). As suggested by Vähä & Primmer (2006), a threshold  $q$ -value of 0.2 was used to separate pure from hybrid individuals, maximizing the accuracy of assignment. Thus, an admixed individual with at least 20% of its genetic material from both subspecies was classified as a hybrid.

Finally, to evaluate the magnitude and direction of gene flow between tree subspecies, we estimated the per-generation migration rates ( $Nm$ ) between *Y. b. brevifolia* and *Y. b. jaegeriana* in Tikaboo Valley, using MIGRATE-N version 3.1.6 (Beerli & Felsenstein 1999, 2001). Additionally, following the approaches of Scasciulli *et al.* (2010) and Field *et al.* (2011), we also estimated the migration rates between the two subspecies and hybrids, allowing us to evaluate whether hybrids serve as a 'bridge' for gene flow between parental populations. Individuals were assigned to populations (*Y. b. brevifolia*, *Y. b. jaegeriana* or hybrids) based on the results of the STRUCTURE analysis (above), assuming that trees where  $0.2 < q < 0.8$  represent truly admixed individuals. We completed two different analyses, either considering only nonadmixed individuals or including hybrids as a third population. We jointly estimated the parameters  $\theta$  (equal to four times the effective population size, times the per-locus neutral mutation rate) and the asymmetric migration rates ( $M = m/\mu$ ) between each population using a Bayesian search strategy (Beerli 2006). For the sake of comparison, we also calculated the actual number of migrants moving between populations ( $Nm$ ) by multiplying the modal values of  $M$  by  $\theta/4$ , using the mode of the posterior distribution of  $\theta$ . For each locus, the search employed two 10 000 000-generation MCMCs sampling every 100 generations after an initial 100 000-generation burn-in. Migration rate priors used a uniform prior from zero to 2000 with a mean of 500. Because mutation rates in microsatellites are typically orders of magnitude greater than for sequence data and because initial analyses indicated that the default priors were forcing artificially low estimates of  $\theta$ , we used an exponential distribution with a maximum of 100 and a mean of 50. The analysis was completed using two independent runs, and chain convergence was assessed by comparing the correlation between parameter estimates between independent runs. The overall correlation between runs was  $>0.981$ .

#### Comparisons between data sets

We evaluated the agreement among genetic, vegetative, and floral data with respect to the assignment of individual trees, using parametric and nonparametric statistics. First, we extracted principal components from the vegetative and floral data. Principal component loadings are provided as Supporting information. We then compared statistical associations between the first principal components and the posterior probability of assignment as *Y. b. brevifolia* as estimated from the genetic data using STRUCTURE. Because the probability of assignment is constrained to values between zero and one, we

arcsine-transformed these data. Correlations between data sets were assessed using a linear regression, treating the STRUCTURE assignments as the dependent variable. Thus, this analysis evaluated how well each morphological data set predicts genetic assignment. We compared the power of each of these data sets to predict genetic assignment using delta AIC scores.

Next, we used the logistic regression scores (above) to categorize trees as either *Y. b. brevifolia* or *Y. b. jaegeriana*, based on either floral or vegetative data. For the purpose of this comparison, we considered trees with a logistic regression score  $\geq 0.5$  to be *Y. b. brevifolia* and scores  $< 0.5$  to be *Y. b. jaegeriana*. We then used chi-squared tests to evaluate whether vegetative and floral data disagreed more frequently in the sympatric population than in allopatric populations and whether these data disagreed more frequently for hybrid trees than for nonadmixed trees (as determined by STRUCTURE, above).

## Results

All data analyses revealed strong evidence for hybridization between subspecies. The analyses of both vegetative and floral data revealed a significant excess of

**Table 1** Classification of trees based on morphological data from sympatric and allopatric populations

	Allopatry	Sympatry
Vegetative data		
Ambiguous	63	454
Unambiguous	45	206
Floral data		
Ambiguous	10	27
Unambiguous	84	90

Trees were scored for five vegetative measurements and six floral measurements, and trees were assigned as *Y. b. brevifolia*, *Y. b. jaegeriana* or ambiguous using a logistic regression. Trees with logistic regression scores of  $< 0.05$  or  $> 0.95$  were considered unambiguous, and trees with logistic regression scores between 0.05 and 0.95 were considered ambiguous. The frequency of trees from each category (unambiguous and ambiguous) in each population (sympatric vs. allopatric) were compared using Fisher's exact test.

**Table 2** AMOVA results

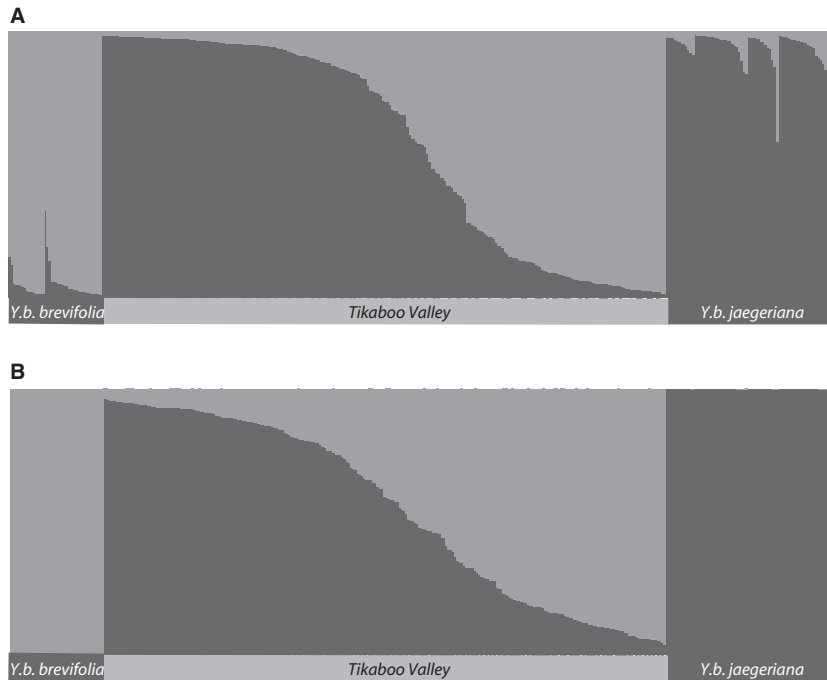
Source of variation	Nested in	Proportion of variation explained	F-stat	F-value	SE	P-value
Within individuals	–	0.730	$F_{IT}$	0.270	0.060	–
Among individual	Populations	0.061	$F_{IS}$	0.077	0.029	0.028
Among populations	Tree subspecies	0.071	$F_{SC}$	0.083	0.028	0.001
Subspecies	–	0.137	$F_{CT}$	0.137	0.058	0.062

Trees were genotyped for 10 microsatellite loci, and genetic variation was partitioned into hierarchically nested sources of variance using GenoDive version 2.0b21. Significance of population structure was assessed using 999 permutations of the data set.

morphological intermediates in sympatry ( $P = 0.036$  for vegetative data,  $P = 0.023$  for floral data; Table 1). Despite the large number of morphological intermediates within the contact zone and in allopatry, the genetic data suggest that the two tree subspecies are differentiated; AMOVA revealed that a large fraction of the genetic variation (13.7%) assort among tree subspecies. Although this result is marginally nonsignificant ( $P = 0.062$ ), differences between tree subspecies account for more of the variation than either difference between populations within subspecies or differences between individuals within populations (Table 2).

The results of the STRUCTURE analysis varied depending on whether prior population information was used (Fig. 3). Without incorporating this information, 5 of 119 trees from allopatric populations were identified as potential hybrids (i.e.  $0.2 < q < 0.8$ ), while within Tikaboo Valley, 59 of 255 trees were identified as hybrids, 76 were identified as *Yucca brevifolia brevifolia* and 120 were identified as *Yucca brevifolia jaegeriana*. When incorporating prior population information, no trees from allopatric populations were scored as hybrids. Within Tikaboo Valley, 94 trees were identified as hybrids (i.e.  $0.2 < q < 0.8$ ), 64 were identified as *Y. b. brevifolia* and 97 were identified as *Y. b. jaegeriana*. Regardless of whether prior population information is used, the optimal value for  $\Delta K$  was achieved at  $K = 2$ .

Estimates of gene flow suggest significant asymmetry in rates of introgression between the two subspecies. Estimates of per-generation numbers of migrants generated using MIGRATE-N were generally high—on the order of 1–100 migrants per generation—though for many migration parameters, the 95% confidence interval includes zero (Fig. 4). However, both analyses (either considering only nonadmixed individuals or including hybrids as a third population) revealed substantial asymmetry in the direction of gene flow, with much greater gene flow moving from *Y. b. jaegeriana* into *Y. b. brevifolia*, than vice versa. The highest rates of gene flow were from *Y. b. jaegeriana* trees into *Y. b. brevifolia* in the comparison including only nonadmixed individuals (63.44 migrants per generation) and from *Y. b. jaegeriana* into hybrids in the comparison including



**Fig. 3** Results of STRUCTURE analysis for  $K = 2$  (A) assuming no prior population information. (B) using population information for parental populations. Each bar represents one individual; shading reflects probability of assignment to (A), or proportion of genetic material from (B), *Y. b. brevifolia* (light grey) and *Y. b. jaegeriana* (dark grey). Trees are ordered by population of origin from west to east and then by probability of assignment.

hybrids as a third population (45.22 migrants per generation). The lowest rates of gene flow were from *Y. b. brevifolia* into *Y. b. jaegeriana* (0.28 migrants per generation in the contrast excluding hybrids; 0.04 migrants when including hybrids) and from hybrids into *Y. b. jaegeriana* (0.06 migrants).

Comparisons between data sets revealed significant agreement between vegetative, floral, and genetic assignments overall. The principal components of variation derived from the vegetative and floral data sets were weakly correlated ( $R^2 = 0.144$ ), but this effect was highly significant ( $P < 0.001$ ). Similarly, both vegetative and floral PCAs were good predictors of the genetic probabilities of assignment to tree subspecies estimated by STRUCTURE (vegetative  $R^2 = 0.158$ ,  $P < 0.001$ ; floral  $R^2 = 0.133$ ,  $P < 0.001$ ). Floral data were a significantly better predictor of genetic assignments overall ( $\Delta AIC = 111.862$ ).

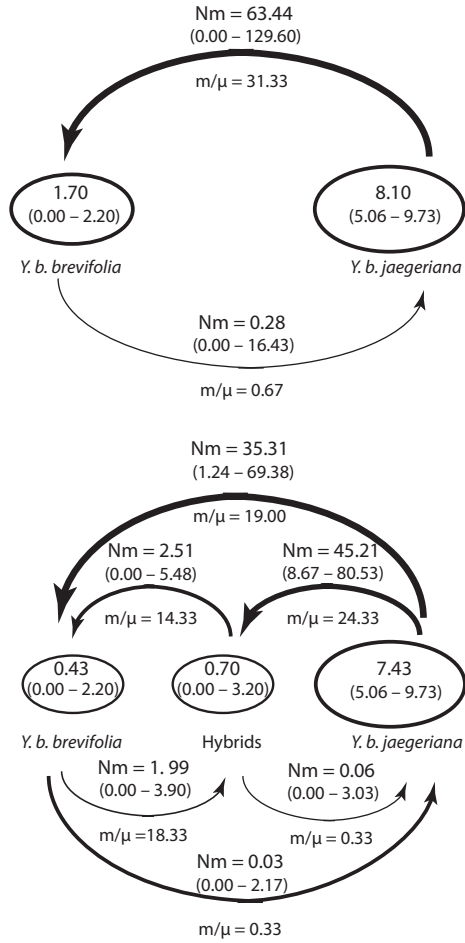
However, comparing the subspecies assignments also revealed a surprising number of conflicts (Fig. 5). Relative to allopatric populations, in Tikaboo Valley, there were significantly more trees where vegetative and floral data predicted different subspecies assignments ( $P = 0.004$ ;  $\chi^2 = 8.370$ ; d.f. = 1). However, genetically admixed trees were not more likely to show conflicts between vegetative and floral assignments ( $P = 0.148$ ;  $\chi^2 = 2.090$ ; d.f. = 1) (Table 3). This could indicate that some hybrid trees were incorrectly classified as nonadmixed, but it is also possible that selection within Tikaboo Valley produces different floral morphology than in the allopatric regions (see below).

## Discussion

Both morphological and genetic data suggest the existence of hybrid individuals within the zone of sympatry in Tikaboo Valley, and the genetic data suggest appreciable rates of backcrossing between hybrid and parental trees. However, rates of gene flow differ markedly between tree subspecies, with significantly greater gene flow occurring from *Yucca brevifolia jaegeriana* into *Yucca brevifolia brevifolia* than vice versa. These differences are consistent with previous measures of gene flow based on chloroplast DNA sequence data (Smith *et al.* 2008) and with the lower host specificity in *T. antithetica* (Smith *et al.* 2009). Thus, pollinator behaviour, rather than intrinsic features of tree biology, plays a primary role in mediating gene flow between tree subspecies.

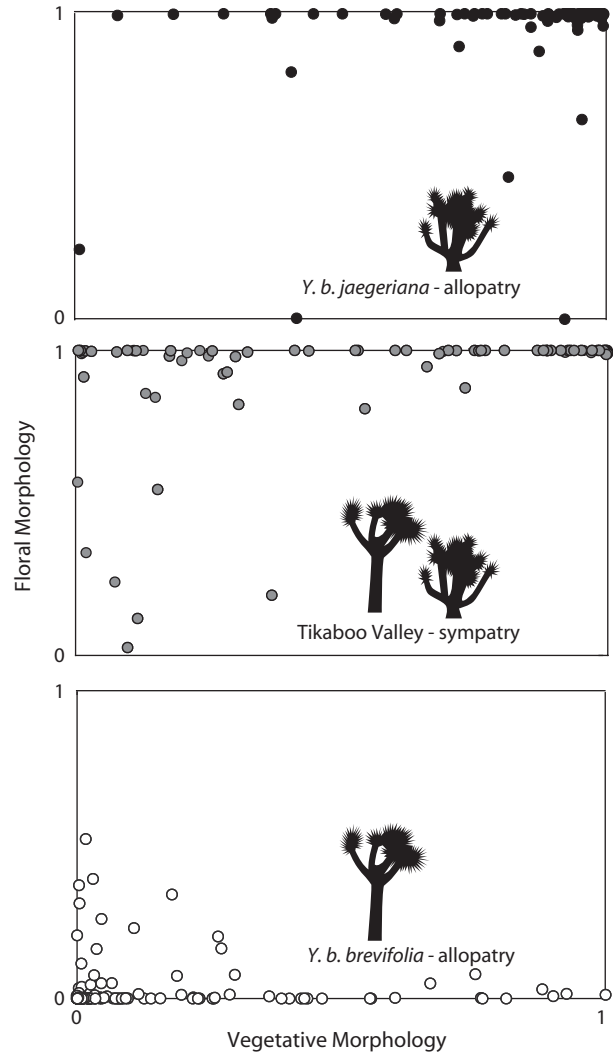
Naturalists have argued for more than a century that pollinators play a key role in generating species diversity within angiosperms (Burger 1981; Hodges & Arnold 1995; Dodd *et al.* 1999; Grimaldi 1999; Sargent 2004; Friedman 2009), either by preventing gene flow between incipient species (Kieffer *et al.* 1984) or by selecting for differences in floral morphology (Fulton & Hodges 1999; Bradshaw & Schemske 2003; Whittall & Hodges 2007), or both (Grant 1949). This idea remains controversial however. Some fossil evidence suggests that associations between angiosperms and insects evolved considerably earlier than the periods of most rapid angiosperm diversification (Pellmyr 1992), and recent modelling results suggest that co-evolution with





**Fig. 4** Estimates of the asymmetric migration rates and  $\theta$  among trees within Tikaboo Valley, as inferred by MIGRATE-N considering either only nonadmixed individuals (top) or including hybrids as a third population (bottom). Trees were assigned to populations using STRUCTURE. Migration rates are expressed as the average number of migrants moving between populations per generation ( $4 Nm$ ) above the arrows and as  $M (m/\mu)$  below the arrows; parentheses list 95% posterior probability distribution for each parameter. The size of arrows and ovals is proportional to parameter estimates.

pollinators should constrain, rather than promote, diversification (Yoder & Nuismer 2010). Finally, Armbruster & Muchhala (2009) have argued that the correlation between specialized pollination and plant species diversity might also be explained by competition for pollinators and resource partitioning among already diverse lineages of plants. Our results show that asymmetric rates of gene flow between incipient species match differences in pollinator behaviour and, thus, that pollinator isolation might indeed play a significant role in generating plant species diversity. However, the results also suggest a potentially important role for natural selection in maintaining species boundaries.



**Fig. 5** Comparisons of tree subspecies assignments based on vegetative and floral morphology, by geography. Trees were scored for the probability of assignment as *Yucca brevifolia jaegeriana* using a logistic regression. Graphs show trees from allopatric populations of *Y. b. jaegeriana* (top panel), *Y. b. brevifolia* (bottom panel) and from the zone of sympatry in Tikaboo Valley (middle panel). A subset (60%) of trees from allopatric populations served as training data to classify the remaining trees.

Prior work indicates that where they occur in sympatry with their respective hosts, *Tegeticula syntheticca* and *T. antithetica* show a breakdown in host specificity. Both pollinator species visit both tree subspecies, but there are substantial differences in the degree of specificity and in the outcomes of these visits. While *T. syntheticca* primarily visits its native host, *T. antithetica* visits both tree subspecies indiscriminately. Moreover, *T. antithetica* frequently produces larvae on its non-native host, suggesting that pollinator visitation results in successful oviposition and pollination. In contrast, *T. syntheticca*

**Table 3** Comparisons between vegetative and floral morphology, by location and by hybrid status, as determined by STRUCTURE

	Agree	Disagree
Across range		
Allopatry	46	6
Sympatry	50	28
Within Tikaboo Valley		
Hybrid	17	15
Nonadmixed	33	13

Trees were classified as either *Y. b. brevifolia* (regression score  $>0.5$ ) or *Y. b. jaegeriana* (regression score  $\leq 0.5$ ) with a logistic regression using either vegetative or floral data sets and tallied as to whether these produced concordant subspecies assignments.

larvae are never reared from *Y. b. jaegeriana*, suggesting that *T. synthetica* moths are either unable to oviposit on their non-native host or are unable to pollinate them. In the current study, the highest rates of gene flow were from *Y. b. jaegeriana* into hybrids and from hybrids into *Y. b. brevifolia*. In contrast, the lowest rates of gene flow were into *Y. b. jaegeriana* (Fig. 4). These patterns match what would be predicted if pollinator behaviour determines gene flow between tree subspecies: pollination of *Y. b. brevifolia* trees by *T. antithetica* moths should result in gene flow from *Y. b. jaegeriana* (via pollen) and *Y. b. brevifolia* (via ovules) into hybrids. However, if *T. synthetica* moths only successfully pollinate *Y. b. brevifolia*, then there will be fewer opportunities for gene flow into *Y. b. jaegeriana*. This scenario is also consistent with the chloroplast capture hypothesis described in prior studies (Smith *et al.* 2009).

While the data presented here seem to implicate pollinator host specificity as the primary mechanism shaping species boundaries in this system, the high frequency of hybridization may also suggest a role for natural selection. The genetic data indicate that approximately one-third of all trees within Tikaboo Valley were of hybrid origin, and the estimates of gene flow between populations indicate that the equivalent of sixty migrants per generation move between the two tree subspecies. One migrant per generation is considered sufficient to prevent divergence through genetic drift (Wright 1978). Thus, strong selection should be required to maintain the morphological differentiation between tree types.

Ecological factors extrinsic to the pollination system might exert divergent selection on the two tree subspecies, but niche comparisons suggest that the abiotic environments inhabited by each tree subspecies are not meaningfully different from one another (Godsoe *et al.* 2009). Furthermore, a more fine-scale analysis of

Tikaboo Valley in this same study found that the location of the hybrid zone does not coincide with any measurable transition in the abiotic environment. It is of course possible that the abiotic environments inhabited by each subspecies may differ in some other, unmeasured, variable, but such an argument is purely *ad hoc*.

Instead, pollinators might maintain morphological differences between subspecies by exerting selection on floral morphology, in addition to determining the extent of gene flow between tree subspecies. Within the yucca–yucca moth interaction, the moths both pollinate the plants, and their larvae consume a portion of the seed crop. In Joshua trees, feeding by yucca moth larvae may destroy more than 25% of the seeds produced by a single fruit (C. Licata and C. Smith unpublished), which probably represent a significant fitness cost for the plants. Previous studies have speculated that differences in style length between *Y. b. brevifolia* and *Y. b. jaegeriana* reflect selection acting to reduce ovules lost to feeding by moths (Godsoe *et al.* 2008; Smith *et al.* 2009).

The idea that differences in floral morphology between the two tree subspecies represent adaptations to pollinators remains speculative. Direct measures of selection acting on floral morphology would be required to test this hypothesis, and such measurements are beyond the scope of the current study. However, comparisons between vegetative, floral and genetic data from the trees in Tikaboo Valley offer some suggestive insights. Within Tikaboo Valley, a significant number of trees were assigned to different subspecies depending on whether vegetative or floral data were used to infer subspecies status (Table 3). The conflicts between the data sets show a striking asymmetry, however. Of the 32 trees with conflicting floral and vegetative assignments, all are assigned as *Y. b. brevifolia* based on vegetative morphology, but as *Y. b. jaegeriana* based on floral morphology (Fig. 5). It is tempting to conclude that the conflicts between vegetative and floral data may represent local adaptation by *Y. b. brevifolia*. Given that *T. antithetica* visits both tree subspecies, one might imagine that selection favours ‘jaegeriana-like’ floral features in both tree subspecies. In contrast, because *T. synthetica* visits *Y. b. brevifolia* almost exclusively, selection for ‘brevifolia-like’ features in *Y. b. jaegeriana* is absent.

While these suggestions are alluring, there are reasons for caution in interpreting these results. It is currently unclear how long the two moth species have existed in sympatry and whether the current hybrid zone might represent secondary contact following initial divergence in allopatry. Joshua trees, like many temperate organisms, underwent significant range changes following the end of the last glacial period. These changes include localized extinctions in the former southern

portions of the range (Cole *et al.* 2011) and may have involved some slight northernmost expansion (Smith *et al.* 2011). Tikaboo Valley is on the northern edge of the current range of Joshua trees, so it is possible that the two tree subspecies have come into contact only within the last 9000–13 000 years. Previously, the two tree subspecies may have been isolated on either side of Death Valley and the Salton Trough. Under this scenario, a very recent origin of a secondary contact zone might mean that there has not been sufficient time for the evolution of strong host fidelity in the moths. Similarly, gene flow across the hybrid zone may have not had sufficient time to homogenize morphological differences between tree subspecies.

It is also unclear whether the patterns of gene flow between subspecies seen in Tikaboo Valley reflect dynamics on a larger scale. The genetic data presented here were sampled from only small fraction of the total range, and so it is unclear whether the differences in host fidelity and gene flow seen within Tikaboo Valley can explain chloroplast introgression on the relatively large geographical scale described previously (Smith *et al.* 2008). However, unpublished data that examine gene flow across the entire range of *Yucca brevifolia* suggest that the patterns seen here are also borne out on a larger scale (J. B. Yoder, unpublished).

The prospect of introgression and gene flow on a broader geographical scale also raises the question of whether the morphological data from allopatric populations might have included admixed individuals and whether this might have impacted our morphological assignments of trees. While it is indeed possible that our sample of trees from allopatric populations may have included admixed individuals, it is unlikely that our finding of more morphologically intermediate trees in sympatry than in allopatry is an artefact. First, trees were sampled from across the entire range, including approximately equal numbers of individuals from many allopatric populations. Most of these are located hundreds of kilometres away from the hybrid zone and outside the region of chloroplast introgression. Thus, if there are admixed individuals within our sample, they are likely to be a small minority. Second, the effect of including admixed individuals would be to make the allopatric populations less distinct, and as a result, most of the trees from allopatric populations would have been classified as ambiguous. Our test, comparing the number of ambiguously assigned trees in sympatry and allopatry is therefore conservative: errors in assignments due to the inclusion of admixed trees in the training data would have resulted in a bias towards accepting the null hypothesis. Finally, by randomly selecting trees from allopatric populations to serve as training data and then classifying the remaining allopatric trees based on these

training data, we can assess how well the logistic regression is able to assign trees to subspecies (and thus the effect of including any putatively admixed trees in our training data). Although there were a large number of ambiguously classified trees from within the allopatric zones based on vegetative data (Table 1), almost none were misclassified (Fig. 5). Similarly, the floral data were able to unambiguously assign trees to subspecies 89% of the time. Thus, if any admixed trees were included in our sample from allopatric populations, it seems to have had little impact on our ability to successfully predict tree subspecies.

The above considerations notwithstanding, our results provide compelling evidence that pollinators play an important role in maintaining species boundaries between incipient species. It seems increasingly clear that pollinator host specificity shapes the direction of gene flow and introgression between subspecies in this system. In addition, the magnitude of gene flow between subspecies suggests that pollinator-mediated selection may also maintain morphological differences between *Y. b. brevifolia* and *Y. b. jaegeriana*. However, direct measures of the covariance between floral morphology and tree fitness are needed to test this idea, as are evaluations of the role that pollinators play in mediating fitness differences. Designing a study to obtain these measures would surely be challenging, but the current results offer some reasons for optimism. Because natural hybrids exist in the zone of sympatry and because these hybrids contain different combinations of genetic and morphological features—including morphologically ‘intermediate’ individuals—it should be possible to statistically partition the contributions of different variables to tree fitness. In addition, the existence of a continuous range of hybrid individuals offers the exciting prospect that an admixture mapping approach (Buerkle & Lexer 2008) could be used to identify genomic regions associated with ecologically important features, such as floral morphology. We are thus optimistic that the Joshua tree system will continue to offer important insights into the mechanisms by which plant pollinator interactions generate diversity.

### Acknowledgements

Funding for this work was provided by grants to CIS from the National Science Foundation (DEB 0918241) and the M. J. Murdock Charitable Trust (2006201:JVZ:11/16/2006) and by a Doctoral Dissertation Improvement Grant to JBY from the National Science Foundation (DEB 091023). TNS and KEG were supported by Willamette University’s Science Collaborative Research Program. Jeff Collins, Theresa Edwards, William Godsoe, Daniel Rowley, Lucille Shaw, Glenn Svensson and Edwin R. Wilmsen assisted with data collection in the field. William Godsoe offered useful discussions about statistical

analyses of the morphological data. The authors gratefully acknowledge this support.

## References

- Aguilera AM, Escabias M, Valderrama MJ (2006) Using principal components for estimating logistic regression with high-dimensional multicollinear data. *Computational Statistics & Data Analysis*, **50**, 1905–1924.
- Armbruster W, Muchhala N (2009) Associations between floral specialization and species diversity: cause, effect, or correlation? *Evolutionary Ecology*, **23**, 159–179.
- Beerli P (2006) Comparison of Bayesian and maximum likelihood estimation of population genetic parameters. *Bioinformatics*, **22**, 341–345.
- Beerli P, Felsenstein J (1999) Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. *Genetics*, **152**, 763–773.
- Beerli P, Felsenstein J (2001) Maximum likelihood estimation of a migration matrix and effective population sizes in *n* subpopulations by using a coalescent approach. *Proceedings of the National Academy of Sciences, USA*, **98**, 4563–4568.
- Bell CD, Soltis DE, Soltis PS (2010) The age and diversification of the angiosperms re-visited. *American Journal of Botany*, **97**, 1296–1303.
- Bradshaw HD, Schemske DW (2003) Allele substitution at a flower colour locus produces a pollinator shift in monkey-flowers. *Nature*, **426**, 176–178.
- Buerkle CA, Lexer C (2008) Admixture as the basis for genetic mapping. *Trends in Ecology & Evolution*, **23**, 686–694.
- Burger W (1981) Why are there so many kinds of flowering plants? *BioScience*, **31**, 572–581.
- Cole K, Ironside K, Eischeid J, Garfin G, Duffy P, Toney C (2011) Past and ongoing shifts in Joshua tree support future modeled range contraction. *Ecological Applications*, **21**, 137–149.
- Darwin C (1876) *The Effects of Cross—and Self-Fertilization in the Vegetable Kingdom*. John Murray, London.
- Dodd ME, Silvertown J, Chase MW (1999) Phylogenetic analysis of trait evolution and species diversity variation among angiosperm families. *Evolution*, **53**, 732–744.
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611–2620.
- Field DL, Ayre DJ, Whelan RJ, Young AG (2011) Patterns of hybridization and asymmetrical gene flow in hybrid zones of the rare *Eucalyptus aggregata* and common *E. rubida*. *Heredity*, **106**, 841–853.
- Flatz R, Yoder JB, Lee-Barnes E, Smith CI (2011) Characterization of microsatellite loci in *Yucca brevifolia* (Agavaceae) and cross-amplification in related species. *American Journal of Botany*, **98**, e67–e69.
- Friedman WE (2009) The meaning of Darwin's "abominable mystery". *American Journal of Botany*, **96**, 5–21.
- Fulton M, Hodges SA (1999) Floral isolation between *Aquilegia formosa* and *Aquilegia pubescens*. *Proceedings of the Royal Society of London Series B*, **266**, 2247–2252.
- Godsoe WKW, Yoder JB, Smith CI, Pellmyr O (2008) Coevolution and divergence in the Joshua tree/yucca moth mutualism. *American Naturalist*, **171**, 816–823.
- Godsoe W, Strand E, Smith CI, Yoder JB, Esque TC, Pellmyr O (2009) Divergence in an obligate mutualism is not explained by divergent climatic factors. *New Phytologist*, **183**, 589–599.
- Grant V (1949) Pollination systems as isolating mechanisms in angiosperms. *Evolution*, **3**, 82–97.
- Grimaldi D (1999) The co-radiations of pollinating insects and angiosperms in the Cretaceous. *Annals of the Missouri Botanical Garden*, **86**, 373–406.
- Hartl DL, Clark A (1989) *A Primer of Population Genetics*, 2nd edn. Sinauer Associates, Sunderland, Massachusetts.
- Hodges S, Arnold ML (1995) Spurring plant diversification: are floral nectar spurs a key innovation? *Proceedings of the Royal Society of London Series B*, **262**, 343–348.
- Janzen D (1979) How to be a fig. *Annual Review of Ecology and Systematics*, **10**, 13–51.
- Kiester AR, Lande R, Schemske D (1984) Models of coevolution and speciation in plants and their pollinators. *American Naturalist*, **124**, 220–243.
- Lenz LW (2007) Reassessment of *Y. brevifolia* and recognition of *Y. jaegeriana* as a distinct species. *Aliso*, **24**, 97–104.
- Machado CA, Robbins N, Gilbert MTP, Herre EA (2005) Critical review of host specificity and its coevolutionary implications in the fig/fig-wasp mutualism. *Proceedings of the National Academy of Sciences*, **102**, 6558–6565.
- Meirmans PG, Van Tienderen PH (2004) genotype and genotype: two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes*, **4**, 792–794.
- Menard S (1995) *Applied Logistic Regression Analysis*. Sage University, Thousand Oaks, California.
- Mora C, Tittensor DP, Adl S, Simpson AGB, Worm B (2011) How many species are there on earth and in the ocean? *PLOS Biology*, **9**, e1001127.
- Nuismer SL, Gomulkiewicz R, Ridenhour BJ (2010) When is correlation coevolution? *The American Naturalist*, **175**, 525–537.
- O'Brien R (2007) A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity*, **41**, 673–690.
- Pellmyr O (1992) Evolution of insect pollination and angiosperm diversification. *Trends in Ecology & Evolution*, **7**, 46–49.
- Pellmyr O (2003) Yuccas, yucca moths and coevolution: a review. *Annals of the Missouri Botanical Garden*, **90**, 35–55.
- Pellmyr O, Segraves KA (2003) Pollinator divergence within an obligate mutualism: two yucca moth species (Lepidoptera; Prodoxidae: *Tegeticula*) on the Joshua Tree (*Yucca brevifolia*; Agavaceae). *Annals of the Entomological Society of America*, **96**, 716–722.
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.
- Sargent RD (2004) Floral symmetry affects speciation rates in angiosperms. *Proceedings of the Royal Society of London Series B*, **271**, 603–608.
- Scascitelli M, Whitney KD, Randell RA, King M, Buerkle CA, Rieseberg LH (2010) Genome scan of hybridizing sunflowers from Texas (*Helianthus annuus* and *H. debilis*) reveals asymmetric patterns of introgression and small islands of genomic differentiation. *Molecular Ecology*, **19**, 521–541.
- Scotland RW, Wortley AH (2003) How many species of seed plants are there? *Taxon*, **52**, 101–104.
- Smith CI, Godsoe WKW, Tank S, Yoder JB, Pellmyr O (2008) Distinguishing coevolution from covariance in an obligate

- pollination mutualism: asynchronous divergence in Joshua tree and its pollinators. *Evolution*, **62**, 2676–2687.
- Smith CI, Drummond CS, Yoder JB, Godsoe W, Pellmyr O (2009) Host specificity and reproductive success of yucca moths (*Tegeticula* spp. Lepidoptera: Prodoxidae) mirror patterns of gene flow between host plant varieties of Joshua tree (*Yucca brevifolia*: Agavaceae). *Molecular Ecology*, **18**, 5218–5229.
- Smith CI, Tank S, Godsoe W *et al.* (2011) Comparative phylogeography of a coevolved community: concerted population expansions in Joshua trees and four yucca moths. *PLoS ONE*, **6**, e25628.
- Stebbins GL (1981) Why are there so many species of flowering plants? *BioScience*, **31**, 573–577.
- Vähä J-P, Primmer CR (2006) Efficiency of model-based Bayesian methods for detecting hybrid individuals under different hybridization scenarios and with different numbers of loci. *Molecular Ecology*, **15**, 63–72.
- Wahlund S (1928) Zusammensetzung von Population und Korrelationserscheinung vom Standpunkt der Vererbungslehre aus betrachtet. *Hereditas* **11**, 65–106.
- Weiblen GD (2002) How to be a fig wasp. *Annual Review of Entomology*, **47**, 299–330.
- Weiblen GD (2004) Correlated evolution in fig pollination. *Systematic Biology*, **53**, 128–139.
- Whittall JB, Hodges SA (2007) Pollinator shifts drive increasingly long nectar spurs in columbine flowers. *Nature*, **447**, 706–709.
- Wright S (1978) *Evolution and the Genetics of Populations Vol. 4: Variability within and among Natural Populations*. University of Chicago Press, Chicago, Illinois.
- Yoder JB, Nuismer SL (2010) When does coevolution promote diversification? *The American Naturalist*, **176**, 802–817.

---

T.N.S. and C.I.S. designed the study. T.N.S., K.E.G., J.B.Y. and R.F. performed the research. C.I.S. wrote the manuscript.

---

### Data accessibility

Morphological and microsatellite data have been deposited in the Dryad repository: <http://dx.doi.org/10.5061/dryad.7tq0s>

### Supporting information

Additional supporting information may be found in the online version of this article.

**Table S1** Tests of Hardy Weinberg equilibrium using Nei's  $G_{IS}$  by locus and by population.

**Table S2** Measures of multicollinearity among predictor variables used to classify trees to subspecies.

**Table S3** Logistic regression results for training data used to assign trees as *Y. b. brevifolia* or *Y. b. jaegeriana*.

**Table S4** Principal component analyses of floral data—component loadings and total variation explained by each component.

**Table S5** Principal component analyses of vegetative data—component loadings and total variation explained by each component.