Molecular Ecology (2009) **18**, 1916–1931 doi: 10.1111/j.1365-294X.2009.04148.x

Not just vicariance: phylogeography of a Sonoran Desert euphorb indicates a major role of range expansion along the Baja peninsula

R. C. GARRICK,* J. D. NASON,† C. A. MEADOWS* and R. J. DYER*
*Department of Biology, Virginia Commonwealth University, Richmond, Virginia 23284, USA, †Department of Ecology, Evolution and Organismal Biology, Iowa State University, Ames, Iowa 50011, USA

Abstract

To examine the generality of population-level impacts of ancient vicariance identified for numerous arid-adapted animal taxa along the Baja peninsula, we tested phylogeographical hypotheses in a similarly distributed desert plant, *Euphorbia lomelii* (Euphorbiaceae). In light of fossil data indicating marked changes in the distributions of Baja floristic assemblages throughout the Holocene and earlier, we also examined evidence for range expansion over more recent temporal scales. Two classes of complementary analytical approaches — hypothesis-testing and hypothesis-generating — were used to exploit phylogeographical signal from chloroplast DNA sequence data and genotypic data from six codominant nuclear intron markers. Sequence data are consistent with a scenario of mid-peninsular vicariance originating *c.* 1 million years ago (Ma). Alternative vicariance scenarios representing earlier splitting events inferred for some animals (e.g. Isthmus of La Paz inundation, *c.* 3 Ma; Sea of Cortez formation, *c.* 5 Ma) were rejected. Nested clade phylogeographical analysis corroborated coalescent simulation-based inferences. Nuclear markers broadened the temporal spectrum over which phylogeographical scenarios could be addressed, and provided strong evidence for recent range expansions along the north–south axis of the Baja peninsula. In contrast to previous plant studies in this region, however, the expansions do not appear to have been in a strictly northward direction. These findings contribute to a growing appreciation of the complexity of organismal responses to past climatic and geological changes — even when taxa have evolved in the same landscape context.

Keywords: Baja California peninsula, *Euphorbia*, landscape history, population structure, statistical phylogeography

Received 29 November 2008; revision received 29 January 2009; accepted 5 February 2009

Introduction

A new era of analytical phylogeography is emerging, driven by advances in the use of coalescent theory to understand how DNA sequences can be used to uncover past divergence events (e.g. vicariance) and long-acting, recurrent processes (e.g. gene flow) that shape present-day distributions of intraspecific diversity (Emerson *et al*. 2001; Hey & Machado 2003). To explicitly account for the inherent 'noise' in genetic data sets, recent approaches to historical inference have employed simulations to model stochastic coalescence of evolutionary lineages. They typically use a scenario-based framework in which a small set of alternative divergence histories are assessed (i.e. hypothesis-testing), with clearly demarcated populations serving as the underlying units of interest (e.g. Knowles & Maddison 2002; Carstens *et al*. 2005; DeChaine & Martin 2006; Garrick *et al*. 2008a). However, a major source of potential error may arise from model misspecification, in which case the best-fit scenario may nonetheless fit the data quite poorly. Although alternative scenario-free (i.e. hypothesis-generating) approaches generally make few assumptions, flexibility comes at the cost of statistical discrimination among alternative explanations for a given phylogeographical pattern (e.g. Knowles 2004). For these reasons, a careful balance

Correspondence: Ryan Garrick, Fax: +1804-828-0503; E-mail: rcgarrick@vcu.edu

of complementary hypothesis-testing and hypothesis-generating approaches is particularly desirable.

In plant phylogeographical studies that focus on chloroplast DNA (cpDNA) sequence variation, historical inferences can be obscured by incomplete lineage sorting due to the typically slow mutation rates of this genome (Schaal *et al*. 1998). Indeed, DNA sequences can fail to faithfully record past events that impacted genetic structuring, and thus, the strongest historical signal may be embedded in haplotype frequency information. Furthermore, this genome reflects only the history of effective seed-mediated gene flow, and therefore, exclusive reliance on cpDNA data ignores the genetic consequences of pollen-mediated dispersal, which is typically the predominant mechanism of effective gene flow in woody plants (El Mousadik & Petit 1996; Hamrick & Nason 2000). The analysis of cpDNA sequence data in combination with nuclear genotype-yielding markers strengthens historical inference. In particular, past events can be examined over a wider temporal spectrum, and hence, different components of a species' evolutionary history can potentially be separated (Crandall *et al*. 2000; Sunnucks 2000; Garrick *et al*. 2008a). For example, there is little chance that Holocene range expansions would be recorded by cpDNA sequence mutations (Templeton 2002), but over the same timescales, nuclear allele and genotype frequency data are likely to be informative (Sunnucks 2000), with the signature of recent (re)colonization evident from a reduction in the number of alleles (Hewitt 1996). In addition, contrasts between maternally inherited cpDNA and biparentally inherited nuclear alleles can be used to infer the relative importance and spatial scales of seed- vs. pollen-mediated gene flow (Ennos 1994; Hamilton & Miller 2002), and how these contribute to phylogeographical structure (e.g. Oddou-Muratorio *et al*. 2001).

The Sonoran Desert of the Baja California peninsula and adjacent continental Mexico ('Baja peninsula' and 'continental Sonora' herein) have featured prominently in animal phylogeographical studies (e.g. Riddle *et al*. 2000; Rodriguez-Robles & De Jesús-Escobar 2000; Hurtado *et al*. 2004; Crews & Hedin 2006; Douglas *et al*. 2006; Recuero *et al*. 2006; Riddle & Hafner 2006; Ross & Markow 2006; Pfeiler *et al*. 2007). These investigations uncovered a prominent and taxonomically pervasive role of vicariance in this landscape setting. Three major events underpin hypotheses for explaining patterns of biotic diversification in the region: (i) permanent separation of Baja peninsula and continental Sonora populations following formation of the Sea of Cortez (*SoC*) *c*. 5 million years ago (Ma; Roberts 1989; although initial rifting may have created a proto-gulf *c.* 12 Ma, Riddle *et al*. 2008), (ii) temporary isolation of southern Baja (Cape Region) from other peninsular populations owing to oceanic inundation of the Isthmus of La Paz (*IoL*) *c.* 3 Ma (Riddle *et al*. 2000), and (iii) a putative mid-peninsular seaway (*MPS*) that temporarily isolated northern Baja populations from the remainder of the peninsula *c.* 1 Ma (Upton & Murphy 1997; Fig. 1).

Although the former existence of the *MPS* remains a topic of considerable debate (Jacobs *et al*. 2004; Lindell *et al*. 2006), geological activity along the Baja peninsula has been cited as the major cause of past vicariance events evident from traditional phylogenetic analyses of animal mitochondrial DNA (Riddle *et al*. 2000). However, changes in local environmental conditions that accompanied Pleistocene and Holocene climatic fluctuations are also likely to have influenced genetic structuring (Grismer 2002a, b). This is especially true for plants, given that species' distributions, phenologies, and physiological tolerances can be strongly tied to precipitation, or the frequency and severity of winter frosts (Van Devender 2002). Indeed, early to mid-Holocene range shifts of Sonoran Desert floral communities are recognizable from plant macrofossils in packrat middens (Van Devender *et al*. 1994), and allozyme data of two columnar cacti provide strong evidence of northward expansion along the peninsula (Nason *et al*. 2002; Clark-Tapia & Molina-Freaner 2003).

To date, few phylogeographical studies of terrestrial plants from Baja peninsula have been conducted, and thus, appraisals of regional patterns of biodiversity and the processes that underpin them are potentially biased due to an overrepresentation of animal (particularly vertebrate) studies (see Riddle *et al*. 2000; Riddle & Hafner 2006). In the present study, we use molecular data from a long-lived arid-adapted plant, *Euphorbia lomelii* V.W. Steinm. (Euphorbiaceae), to explicitly test the set of vicariance scenarios hypothesized for animal taxa. This euphorb is endemic to Sonoran Desert habitats of the Baja peninsula and continental Sonora and may have a long history of occupation in these two regions — packrat midden data indicate the persistence of small semideserts at low elevations during the Early Holocene, at the height of the Last Glacial Maximum, and throughout the Pleistocene (Van Devender *et al*. 1994; Thompson & Anderson 2000). However, the current widespread distribution of desert flora in southwestern North America and Sonora is believed to be very recent, with the Sonoran Desert establishing its present-day northern range limit only *c.* 6000 years ago (Thompson & Anderson 2000). In addition to testing for vicariance using cpDNA sequences coupled with analyses of spatial patterns of differentiation at nuclear loci, we also examined evidence for post-Pleistocene range expansion. In the latter case, we expected that a south-to-north gradient of decreasing genetic diversity would be evident from diploid nuclear markers, and despite slow evolutionary rates, perhaps also for cpDNA. To enable identification of the overlying signals of vicariance and recent range expansion, a flexible set of hypothesis-testing and hypothesis-generating analytical approaches were employed.

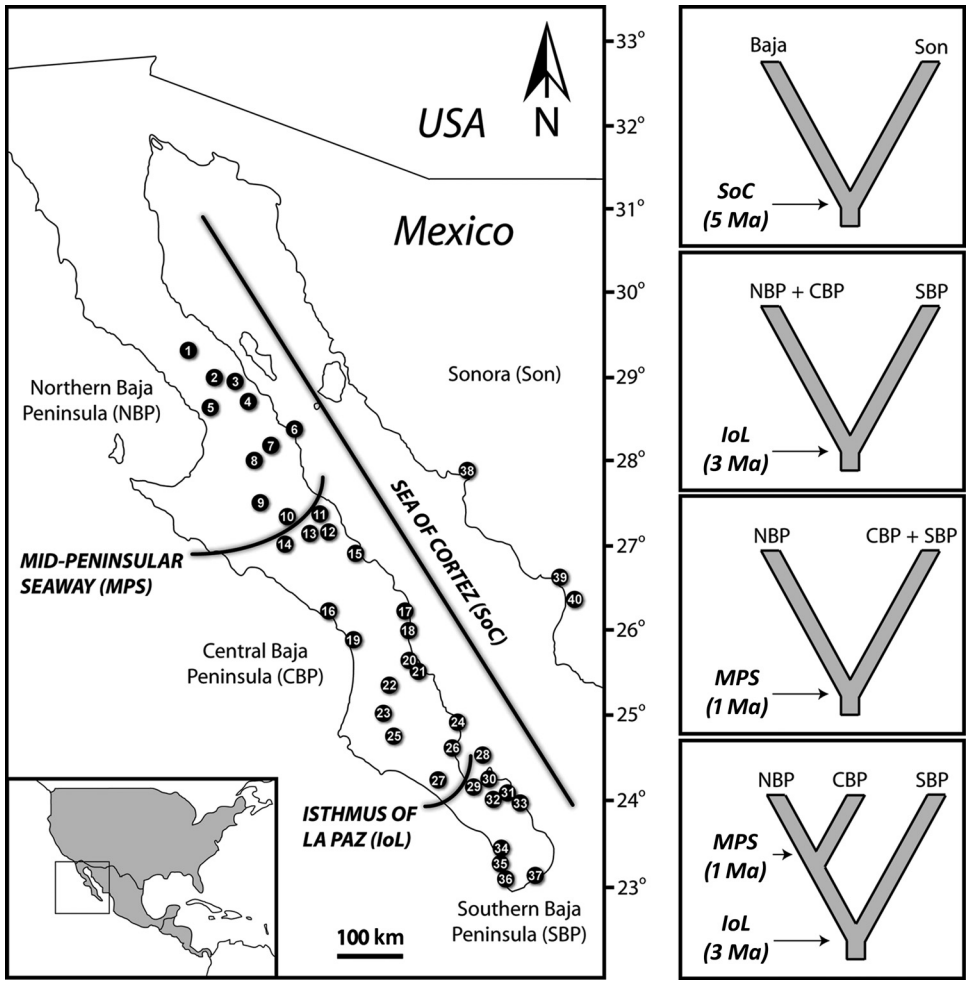

Fig. 1 Left panel: map of the Baja peninsula and continental Sonora showing *Euphorbia lomelii* populations (labelled 1–40) sampled in the present study. Four *a priori* regions (NBP, CBP, SBP and Son) are separated by three hypothesized phylogeographical breaks (*MPS*, *IoL* and *SoC*). Inset: map of Central and North America showing location of the study area. Right panels: population tree representations of alternative vicariance hypotheses, which include single-event scenarios (upper three panels) and a two-tiered vicariance (*TTV*) scenario (lower panel). Timing of splitting events (million years ago, Ma) is based on dates reported in the literature.

Materials and methods

Taxon sampling and data sets

Euphorbia lomelii (synon. *Pedilanthus macrocarpus* Benth.) is a hummingbird-pollinated desert succulent (Dressler 1957). Leaves are ovate (1–2.5 cm), and seeds are ovoid (7–10 mm diameter) and possess no obvious mechanisms for dispersal other than via gravity. The species is common along the Baja peninsula, occurring from just north of Bahia de Los Angeles south to the Cape Region. Smaller populations are also found along coastal regions of southern continental Sonoran Desert in the Mexican state of Sonora.

Euphorbia lomelii was sampled from 40 localities encompassing the species' known range (Fig. 1; Appendix). Genomic DNA was extracted from silica-dried tissue using a DNeasy plant tissue kit (Qiagen). Polymerase chain reactions (PCRs) were performed using a PTC-200 thermocycler (MJ Research) in 10 μL volumes containing 2 μL 5× *GoTaq* Flexi buffer (Promega), 2 mM $MgCl_2$, 200 μM of each dNTP, 5% bovine serum albumin (New England Biolabs), 0.5 μM of each primer, 0.5 U *GoTaq* DNA polymerase (Promega), and 1 μL template DNA. Sequencing was performed on a MegaBACE 1000 (Amersham Biosciences) following the manufacturer's recommendations.

Sequence data were obtained from two chloroplast protein-coding genes, *maturase K* (465-bp) and *NADH dehydrogenase* subunit F (465 or 471-bp), using primers designed from GenBank Accessions for related species (Table 1). Sequences were edited by eye, then aligned and checked for open reading frames in MEGA version 4.0 (Tamura *et al*. 2007). In addition, two intergenic spacers were amplified using

Table 1 Primers used for amplification of *Euphorbia lomelii* chloroplast DNA regions

cpDNA region	Primer sequence (5′–3′)	*T* (°C)	Assay	GenBank Accessions	Source
matK	F: AGTATCTTCTTTAGAAAGGCC R: TAGCATTTGACTCCGTACTACC*	50	DNA sequencing	FJ652116–FJ652145	This study
ndhF	F: TAATAGCTTGGTTGACTGCGG R: TGTAACCTCGATTATAGGAC*	46	DNA sequencing	FJ652148–FJ652177	This study
trn L-F	F: GGTTCAAGTCCCTCTATCCC R: ATTTGAACTGGTGACACGAG	48	RFLP	FJ652180–FJ652181	Taberlet *et al*. (1991)
*trn*T-L	F: CATTACAAATGCGATGCTCT R:TCTACCGATTTCGCCATATC	50	Indel	FJ652182–FJ652184	Taberlet *et al*. (1991)

*primer used for sequencing.
T, annealing temperature; indel, insertion/deletion mutation; RFLP, restriction fragment length polymorphism.

primers from Taberlet *et al*. (1991). The *trn*T^{UGU}–*trn*L$^{UAA.}$ region was assayed for known insertion/deletion polymorphisms (indels; 14-bp and 49-bp) via standard electrophoresis (2% agarose gels, 60 V, 40 mA for 10 h). Similarly, a T/G single nucleotide polymorphism in the *trn*L^{UAA}–*trn*F$^{GAA.}$ spacer was resolved using a *Dra*I restriction-fragment-length polymorphism (RFLP) assay following Garrick *et al*. (2008c). PCR primer sequences, amplification conditions, and GenBank accessions are given in Table 1. Contiguous indels were treated as a single-event and recoded using arbitrary nucleotide characters. Due to the nonrecombining nature of cpDNA, polymorphism data for the four regions were concatenated. These 'combined haplotypes' were determined for 215 individuals (see Appendix). Six codominant nuclear intron markers were assayed in 352 individuals using RFLP analysis following Garrick *et al*. (2008c). Nuclear loci were: *granule-bound starch synthase*, *floral meristem identity protein*, *alcohol dehydrogenase*, *pistillata*, *RNA polymerase II*, and *malate synthase*.

Assessing alternative a priori *population divergence scenarios*

Spatial structuring of cpDNA and nuclear marker data provides insights into historical vicariance. We examined the distribution of genetic variation within *E. lomelii* by quantifying the fit of the data to two complimentary phylogeographical models: a coalescent approach, and one based upon the statistical decomposition of genetic variance. The three major vicariance events hypothesized to have impacted population structures of Sonoran Desert fauna (i.e. *SoC*, *IoL* and *MPS*) were assessed one at a time, and we also tested a scenario representing the two most recent events (*IoL* and *MPS*; 'two-tiered vicariance' herein; Fig. 1). These scenarios are necessarily simple, but should be useful for generating biologically meaningful inferences because they encompass a suite of evolutionary histories that have received some support in other phylogeographical studies.

Coalescent simulations. Coalescent simulations of cpDNA sequence data were performed using divergence models characterized by the following three parameters: population tree topology (where vicariance events are bifurcations), branch lengths (time since splitting), and branch widths (effective population sizes, N_e). To set N_e of extant populations, we first estimated θ ($2N_e\mu$ for cpDNA of hermaphroditic plants, where μ is the per-site per-generation mutation rate) using FLUCTUATE version 1.4 (Kuhner *et al*. 1998). Search settings were: 10 short Monte Carlo chains (10 000 steps), five long chains (100 000 steps) sampling every 20th genealogy, random starting trees, empirical transition/transversion ratio and base frequencies, starting Θ-value from Watterson's estimate, and the growth parameter fixed at zero. The mean Θ-value from five independent runs was accepted (Table S1, Supporting Information). To convert Θ to N_e, we assumed a generation time of 20 years and a mutation rate of 1.57×10^{-9} per year (an average of 30 published cpDNA substitution rates; Fig. S1, Supporting Information). Branch widths of internal nodes (i.e. ancestral populations) were set by summing N_e of the descendant populations. For each divergence model, 200 coalescent trees were simulated in MESQUITE version 2.1 (Maddison & Maddison 2007) using empirical sample sizes. Next, the best-fit substitution model for *E. lomelii* cpDNA was determined by hierarchical likelihood ratio test in MODELTEST version 3.7 (Posada & Crandall 1998). This model (F_{81} + G, gamma shape = 0.0172, base frequencies = A 0.3265, C 0.1555, G 0.1285, T 0.3895) was used to simulate nucleotide characters (933-bp) on the coalescent trees in MESQUITE, using scaling factors that generated levels of diversity (i.e. number of haplotypes and segregating sites) similar to the empirical cpDNA data (Fig. S2, Supporting Information).

Maddison's (1997) Deep Coalescences (DC), which measures the discord between a gene tree and a population tree assuming incomplete lineage sorting, was used to test vicariance hypotheses. The simplifying assumption of no post-divergence long-distance seed dispersal used here is biologically realistic for *E. lomelii* given its seed morphology

and dispersal biology. Gene trees were estimated for the simulated and empirical cpDNA data sets via neighbour-joining, using maximum-likelihood corrected distances (F_{81} + G model and parameters as above), in PAUP* version 4.0b10 (Swofford 2002), and DC was calculated for unrooted gene trees without branch lengths in MESQUITE. Although the speed of neighbour-joining may come at the expense of accuracy for population-level divergences (Kumar & Gadagkar 2000), this inherent error in phylogenetic estimation should not lead to systematically biased phylogeographical hypothesis tests.

We also used a compound genetic diversity-based statistic, Nei & Li (1979) net nucleotide differences between populations (π_{net}). Given two populations, *X* and *Y*, π_{net} is calculated as $\pi_{XY} - ((\pi_X + \pi_Y)/2)$, where π is the average number of nucleotide differences between two randomly chosen sequences. For the two-tiered vicariance scenario, it was modified to accommodate three populations: $\pi_{net} = ((\pi_{XY} + \pi_{XZ} + \pi_{YZ})/3) - ((\pi_X + \pi_Y + \pi_Z)/3)$, where *Z* is the third population. All π-values were calculated in DNASP version 4.10.3 (Rozas *et al*. 2003). Because the interaction between phylogeny- and diversity-based statistics should provide additive power when the specific combinations of these summary statistics per data set are considered, we used DC and π_{net} to generate bivariate null distributions. For all scenarios, a one-tailed test of significance was performed at 0.05 level (π_{net} tends to increase with time since divergence whereas DC decreases, and thus, we assessed the lower and upper tail, respectively).

Variance decomposition. Genetic variance decomposition methods from analysis of molecular variance (AMOVA, Excoffier *et al*. 1992) were used to estimate relative support for the four alternative vicariance scenarios. The distance model of Smouse & Peakall (1999) was used for codominant nuclear markers, whereas cpDNA haplotypes were encoded as haploid makers in an infinite-alleles model. Each scenario (Fig. 1) specifies a hierarchical partitioning of populations within regions, and the nesting scheme that maximized the among-group variance component was considered the best-fit divergence scenario. AMOVA can also be used to evaluate the relative importance of seed- vs. pollen-mediated gene flow in shaping the spatial distribution of genetic variation (Petit *et al*. 2005). For simplicity, we examined the ratio of seed-to-pollen dispersal in *E. lomelii* by comparing Φ_{ST} estimates for each marker set by considering only peninsular populations. All AMOVA analyses were conducted in GENETICSTUDIO (Dyer 2009).

Estimating a population divergence scenario from the data

Although formulating and testing competing historical scenarios using coalescent simulations carries a number of benefits (see Introduction), these approaches cannot determine whether the conditions used in models (i.e. fixed parameters) are correct, or whether the 'true' scenario was included in the set (Hey & Machado 2003; Templeton 2009). Here we used nested clade phylogeographical analysis (NCPA, Templeton *et al*. 1995) as a heuristic approach to estimating a population divergence scenario directly from the cpDNA data set. We recognize that there are limitations associated with the NCPA inference key; accordingly, we employed NCPA as one of several complementary approaches for addressing specific hypotheses generated from information external to the present genetic data (Garrick *et al*. 2008b). A statistical parsimony network was estimated for cpDNA with the 95% confidence criterion enforced in TCS version 1.21 (Clement *et al*. 2000). Ambiguous connections were resolved following Pfenninger & Posada (2002), and the nesting design was constructed by hand. NCPA statistics were calculated in GEODIS version 2.5 (Posada *et al*. 2000), with significance evaluated at the 0.05 level via 10 000 permutations. Although D_c, D_n and *I–T* distances are more sensitive to spatial-genetic structure than the contingency chi-square test (Posada *et al*. 2006), generating multiple statistics per nested clade may lead to high false positives (Panchal & Beaumont 2007). For this reason, we applied the November 2005 inference key (http://darwin.uvigo.es/software/geodis.html) only when the chi-square statistic for a nested clade was significant.

Identifying key components of population history from the data

The spatial distribution of nuclear genetic variation was used to examine three aspects of landscape-level historical events, and on-going, recurrent population processes: (i) the number of natural partitions and their distribution with respect to hypothesized phylogeographical breaks, (ii) signal of post-Pleistocene range expansion evident from latitudinal trends in levels of within-population genetic diversity, and (iii) isolation-by-distance and long-distance dispersal.

To identify natural genotypic clusters, we used GENELAND version 2.0.12 (Guillot *et al*. 2005). This approach is underpinned by the notion that spatial-genetic discontinuities exist between populations and one or more geographically cohesive, multiple-locality clusters capture the major structure in the data (see Chen *et al*. 2007). We examined $K = 1–10$ distinct clusters (five replicates each) with the following search settings: 100 000 MCMC iterations recording every 100th step, spatial Dirichlet model, maximum rate of Poisson process = *n* (where *n* is the number of individuals), and maximum number of nuclei = 3*n*. The modal *K* value from the posterior distribution was accepted, following Guillot *et al*. (2005).

We tested for recent range expansion by examining within-population genetic diversity and among-population

Table 2 Coalescent simulation tests of four alternative vicariance scenarios depicted in Fig. 1. Test statistics are the number of Deep Coalescences (DC), and net nucleotide differences between populations (π_{net}). 'Combined' is a bivariate test statistic that jointly considers DC and π_{net} values obtained from a given sequence data set. '*P* value' is the probability of obtaining test statistic values equal to or more extreme than the observed (Obs.) value at the 0.05 level

			DC		π_{net}		Combined
Vicariance scenario	Timing	Geographical scale	Obs. value	*P* value	Obs. value	*P* value	*P* value
Sea of Cortez	5 Ma	Rangewide	24	0.050	1.262	0.060	0.030
Isthmus of La Paz	3 Ma	Baja Peninsula	65	0.035	0.493	0.025	0.020
Mid-peninsular seaway	1 Ma	Baja Peninsula	17	0.855	1.263	0.350	0.330
Two-tiered vicariance	1–3 Ma	Baja Peninsula	96	0.300	0.934	0.075	0.055

genetic variance, both of which are expected to vary systematically along the axis of expansion (Hewitt 1996). For nuclear diversity, we estimated allelic richness (*A*) using HP-RARE version 1.0 (Kalinowski 2005) and used rarefaction to correct for unequal sample sizes (population 35 had only one sample and was omitted). For cpDNA diversity, we estimated Nei (1987) haplotypic diversity (*H*) using DnaSP. Significance of systematic changes in within-population diversity for both marker types was determined by regression on latitude using the software R (R Development Core Team 2008). Systematic changes in among-population genetic structure were evaluated using a stepwise analysis of molecular variance (STAMOVA; Dyer *et al*. 2004). The STAMOVA framework is a general family of multivariate linear models, of which the AMOVA test is one particular subset. We used the STAMOVA model to treat latitude as a covariate in the analysis of genetic variance. The residual genetic variation left over after regressing the latitude of populations on the genetic structure was then partitioned into within- and among-population components, yielding the statistic $\Phi_{ST\ Baja|Spatial}$. Comparing the noncorrected structure, $\Phi_{ST\ Baja}$, to $\Phi_{ST\ Baja|Spatial}$ provides an estimate of how much of the observed genetic variation is explained by latitude, which we attribute to range expansion.

Lastly, we determined if there was evidence for isolation by distance and/or long-distance dispersal events using Population Graphs — a graph-theoretic approach to analyse how genetic variation is distributed across the landscape. Within a graph, populations are represented as nodes and the genetic covariation among populations determines the topology. The pattern of connections between populations is estimated conditional on the entire data set and can be used to test for isolation-by-graph-distance (IBGD; Dyer & Nason 2004). We prefer graph distance to pairwise F_{ST} because the former is determined by the genetic covariance of all populations simultaneously, rather than only individual pairs of populations. The correlation between interpopulation physical distance and graph distances was assessed using a Mantel test. In addition to overall IBGD, we used graph and physical distances to identify pairs of populations that may indicate long-distance dispersal. Long-distance dispersal would create population pairs that are significantly farther apart (spatially) than predicted by genetic distances, even after accounting for IBGD. When placing the Population Graph onto a map, these population pairs would have edges that are stretched significantly farther than expected (referred to here as *extended edges*). To identify extended edges, the chi-square test was used. All analyses on graph and physical distances were conducted in GENETICSTUDIO.

Results

Assessing alternative a priori *population divergence scenarios*

Coalescent simulations. Coalescent simulations of the cpDNA sequence data showed that the *SoC*, *IoL*, and *TTV* scenarios could be rejected (Table 2). The poor fit of the data to these scenarios was evident from each of the two test statistics (DC and π_{net}) individually, and the bivariate test statistic provided a modest improvement in sensitivity. The *MPS* scenario could not be rejected.

Variance decomposition. The AMOVA was applied to both nuclear and chloroplast data sets under the historical hypotheses presented in Fig. 1, as well as the partitioning suggested by GENELAND analysis (described below). Overall, every partitioning scheme yielded a significant estimate of among-region differentiation (Φ_{RT} $P < 0.001$). Of the two rangewide hypotheses, *SoC* had larger among-region variance components for both nuclear and cpDNA (Fig. 2). For hypotheses focused on peninsular populations, the *MPS* hypothesis had the largest regional effect for cpDNA whereas *IoL* had the largest regional effect for the nuclear data. Chloroplast and nuclear markers revealed markedly different estimates of among-population differentiation when examined for Baja populations ($\Phi_{RT:Pollen}/\Phi_{RT:Seed}$ = 0.1809/0.8589 = 0.2106), suggesting that that pollen-mediated dispersal is much more pervasive.

Table 3 Biological inferences for nested clades with significant phylogeographical structure, as determined via χ^2 contingency tests. Inference abbreviations are: RGF with IBD, restricted gene flow with isolation-by-distance; CRE, contiguous range expansion

Significant nested clades	χ^2 permutation			
	P value	Chain of inference	Inferred process or event	Geographical location of primary inference (population)
Clade 1-2	0.0167	1-2-3-4	RGF with IBD	Within continental Sonora (39)
Clade 1-3	0.0002	1-2-3-4	RGF with IBD	Within Northern Baja (3)
Clade 1-5	< 0.0001	1-2-3-4-9-10	Inadequate sampling	
Clade 1-6	< 0.0001	1-19-20-2-11-12-13-21	Insufficient evidence	
Clade 2-1	< 0.0001	1-19-20	Inadequate sampling	
Clade 2-2	< 0.0001	1-2-3-5-15-21	Insufficient evidence	
Clade 2-6	0.0022	1-19-20-2-11-12	CRE	Central Baja (13) to Sonora (38) and Central Baja (24)
Clade 2-7	< 0.0001	1-2-3-4-9	Allopatric fragmentation	Northern Baja (1-6)/Central Baja (11,14,23)
Clade 3-1	< 0.0001	1-2-11-12-13-21	Insufficient evidence	
Clade 3-2	< 0.0001	1-2-3-5-6-7-8	Inadequate sampling	
Total cladogram	< 0.0001	No interior-tip contrast	Inconclusive outcome	

Estimating a population divergence scenario from the data

All cpDNA haplotypes were connected in a single parsimony network (Fig. 3). Four of the 11 nested clades with significant phylogeographical structure yielded unambiguous NCPA inferences. Polarity of the network was unaltered by the addition of two closely related outgroup species, *E. tithymaloides* and *E. calcarata* (GenBank Accessions FJ652146, FJ652147, FJ652178 and FJ652179; not shown). The *MPS* scenario was corroborated by an inference of allopatric fragmentation (clade 2-7) between a tip clade restricted to northern Baja (clade 1-3) and its more broadly distributed ancestor that is most prominent at the northern extremity of central Baja (clade 1-6; Table 3; Fig. 3). Genetic connectivity across the Sea of Cortez was indicated by an inference of contiguous range expansion (clade 2-6) from central Baja (clade 1-18) to continental Sonora and to a more southerly central Baja locality (clade 1-4; Table 3; Fig. 3). However, the putative geographical origin of this range expansion is not immediately adjacent to the Midriff Islands, a series of 'stepping stone' Gulf islands between northern Baja and the mainland located to the east of populations 3, 4 and 6 (Fig. 1). In this case, a more plausible mode of expansion across the Sea of Cortez is long-distance colonization. NCPA inferences of restricted gene flow with isolation-by-distance in northern Baja (clade 1-3) and continental Sonora (clade 1-2) indicate substructuring over finer geographical scales than those considered in our regional vicariance scenarios.

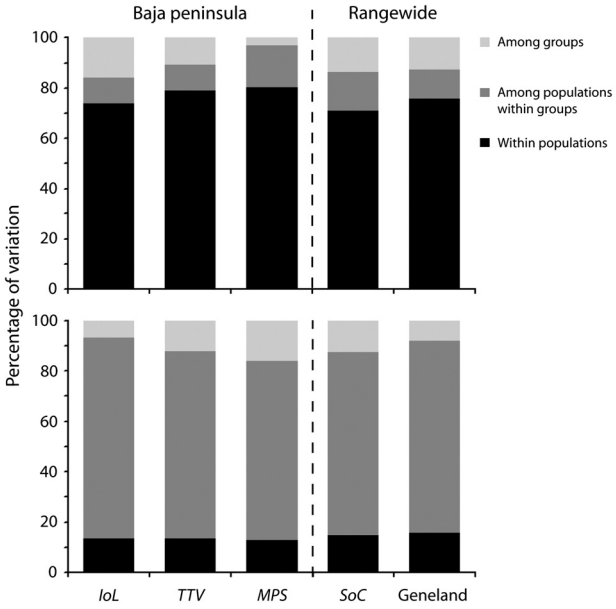

Fig. 2 Top: analysis of molecular variance (AMOVA) comparison of four *a priori* vicariance scenarios (Fig. 1), and an *a posteriori* nesting scheme based on GENELAND (Fig. 4), based on nuclear markers. Sample sizes for Baja peninsula and Rangewide analyses were $N = 327$ diploid individuals and $N = 351$, respectively. Bottom: AMOVA comparisons performed using cpDNA (Baja peninsula $N = 194$ haploid individuals, Rangewide $N = 213$). All AMOVA tests produced significant estimates of regional differentiation.

Identifying key components of population history from the data

The distribution of genetic variation, as partitioned by GENELAND, suggested $K = 2$ groups (but see support for $K = 3$; Fig. 4 inset). Subsequent fixed-K runs were repeated for each of these two values to create a posterior probability map of group membership (Fig. 4). Three aspects of spatial-genetic structuring are particularly noteworthy. First, populations assigned to the same genetic cluster are not always spatially contiguous—both clusters contain member populations that are isolated by large intervening areas (*c.* 170–220 km long). Second, broad-scale genetic structuring is orientated along the north–south axis of the peninsula.

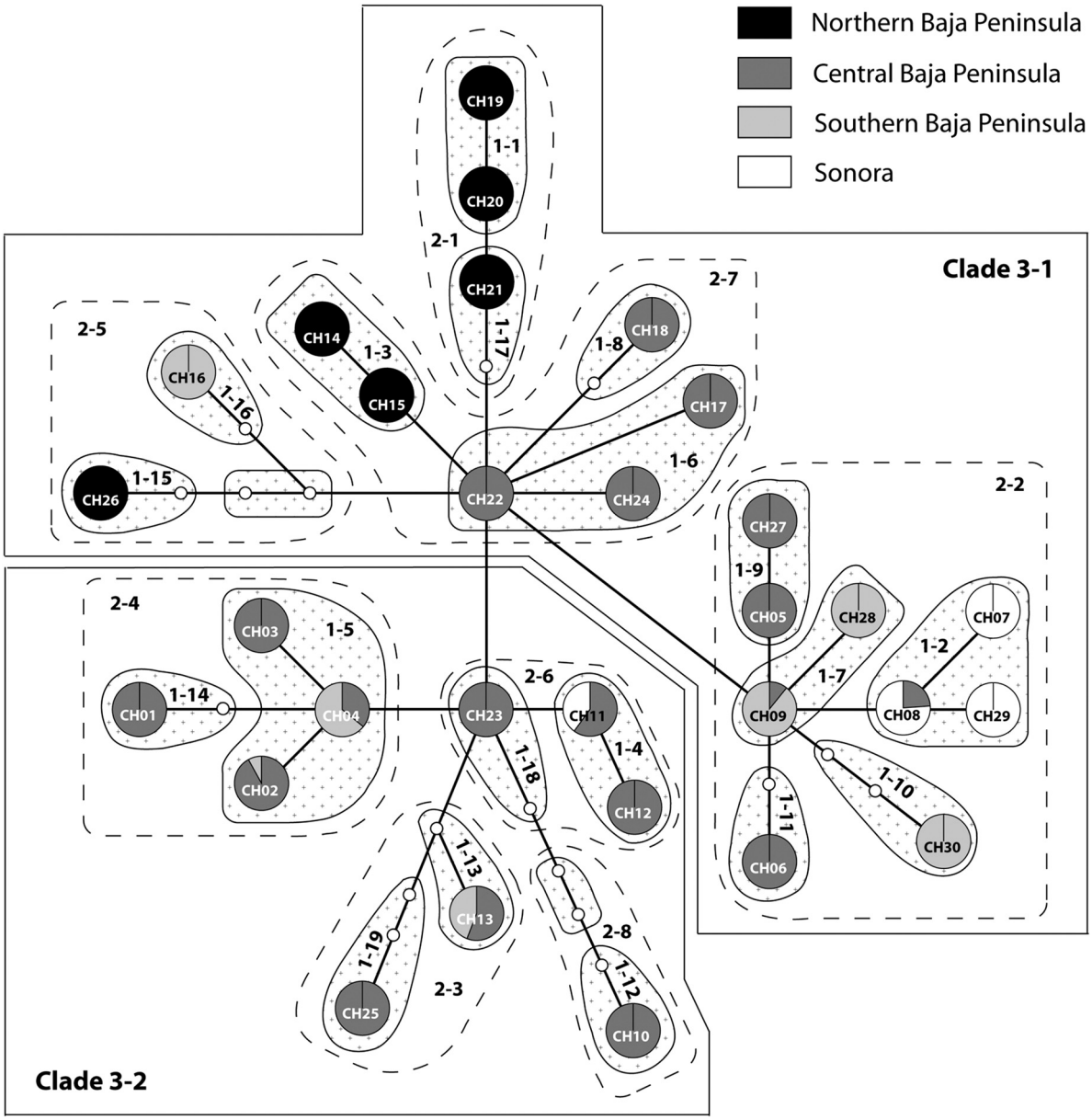

Fig. 3 Statistical parsimony network and nesting design for *Euphorbia lomelii* chloroplast DNA. Pies are unique haplotypes (CH01 to CH30), and pie slices are proportional to relative frequency of a haplotype in each of the four *a priori* regions. Single black lines represent one mutational step, small white circles represent inferred haplotypes that were not sampled or are extinct, and nested clades are labelled with nesting level (prefix) followed by clade number (suffix). The final nesting level, Total Cladogram, is not shown.

Northern and southern Baja regions are relatively homogeneous with respect to population assignments whereas both genetic groups are well represented in central Baja. Finally, despite long-term physiogeographical isolation, continental Sonora populations are not clearly distinct since they group with the population cluster that occupies the southern and central Baja regions (Fig. 4). This $K = 2$ partitioning was used as the additional hypothesis in the AMOVA tests (Fig. 2). In the $K = 3$ analysis, the same three aspects highlighted here were evident; however, the predominantly southern cluster was further subdivided into two *c.* 20 km south of the putative *IoL* break (between populations 29 and 30; Fig. 1).

Nuclear allelic richness showed a complex relationship with latitude (Fig. 5, circles). A simple plot suggested either a higher order polynomial or partitioning of richness values into two separate relationships splitting the Baja peninsula at approximately 25°N (i.e. south of population 23 and coincident with a region of abrupt spatial-genetic discontinuity identified from the GENELAND analyses).

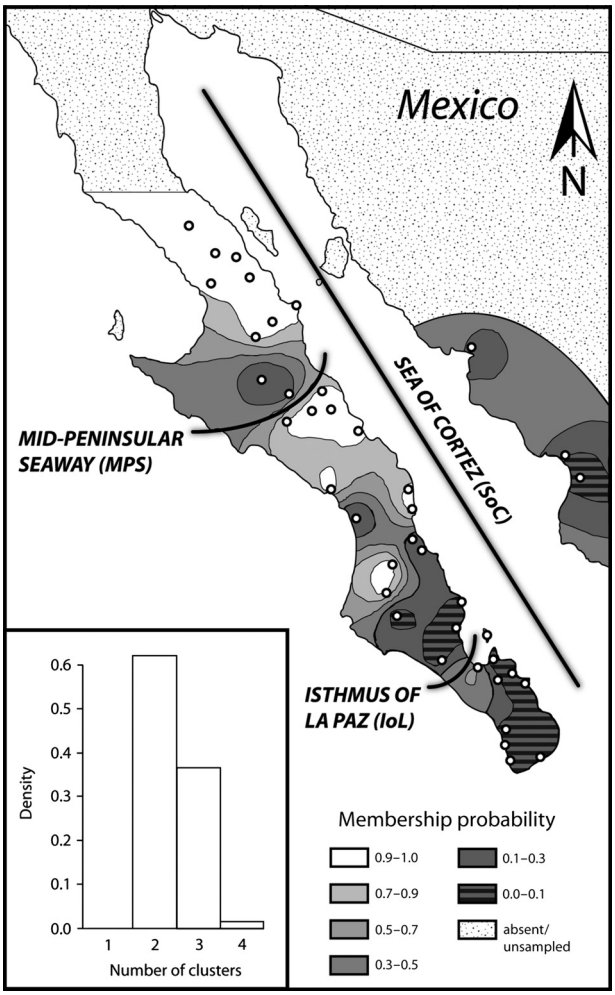

Fig. 4 Map showing the spatial distribution of $K = 2$ clusters, inferred from six nuclear loci analysed with Geneland. Contour lines indicate posterior probability of membership in cluster 1, ranging from white (strongly assigned) to dark grey (weakly assigned; see legend). Inset: Density plot showing modal value of K along the MCMC chains.

Assuming a partition is more biologically reasonable than higher order spatial relationships, both groups of populations showed a significant relationship between latitude and allelic richness (Fig. 5; central/northern group: slope = 0.387, $R^2 = 0.597$, d.f. = 1,20, $F = 32.20$, $P = 1.5e-5$; southern group: slope = 1.334, $R^2 = 0.669$, d.f. = 1,12, $F = 27.20$, $P = 2.16e-4$). However, contrary to our expectations, both of these relationships (as well as the more inappropriate linear model) show a concomitant reduction in diversity in a southward direction suggesting that range expansion proceeded from north to south. Conversely, cpDNA haplotypic diversity (H) showed no significant relationship with latitude (Fig. 5, triangles; pooled linear model: slope = –0.00245, $R^2 = 0.0003$, d.f. = 1,35, $F = 0.010$, $P = 0.922$; populations 11 and 35 were omitted due to small sample sizes).

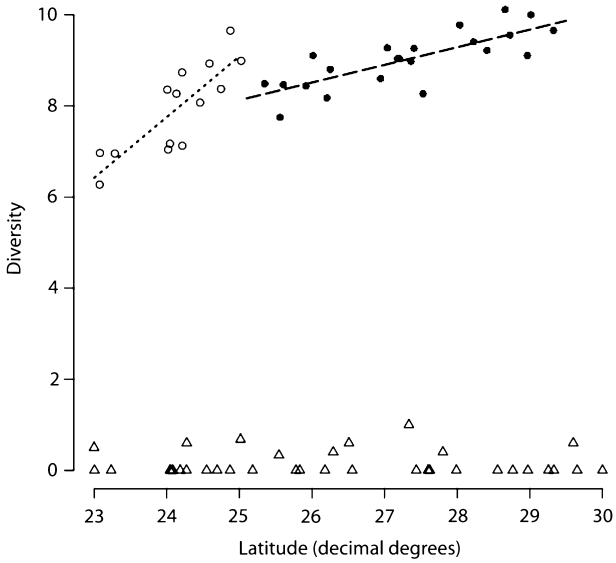

Fig. 5 Latitudinal genetic diversity gradients along the Baja peninsula, plotted by sampling locality from south (left) to north (right). Analysis of the residuals of nuclear allelic richness values suggested the best-fit linear model to be comprised of two separate relationships; one in the southern Cape Region (open circles) and the other comprising the remaining populations to the north (solid circles). Chloroplast DNA haplotypic diversity (triangles) showed no significant relationship with latitude.

Table 4 (a) Analysis of molecular variance (AMOVA), and (b) Stepwise analysis of molecular variance (STAMOVA) calculated for Baja Peninsula populations ($N = 327$ individuals), based on six nuclear loci. 'Subpopulations' are sampling localities (see Appendix). Differences among peninsular subpopulations ($\Phi_{ST\,Baja}$), and among peninsular subpopulations after removing the effect of range expansion by treating the north/south location as a covariate ($\Phi_{ST|Spatial}$), are significant ($P < 0.001$). The impact on genetic structuring of the interaction between latitude and longitude is also significant ($P = 0.023$)

(a) Baja Peninsula

Source of variance	d.f.	MS	$\Phi_{ST\,Baja}$
Among all subpopulations	35	3.437	0.181
Error	291	1.144	

(b) Baja Peninsula

Source of variance	d.f.	MS	$\Phi_{ST\,|\,Spatial}$
Spatial (north/south)	1	25.628	
Among all subpopulations	35	2.696	0.128
Error	288	1.157	

While nuclear genetic differentiation among Baja peninsula populations was relatively large ($\Phi_{ST\,Baja} = 0.181$, $P < 0.001$; Table 4), a significant portion (roughly 29%) of this structure can be attributed to the spatial location of populations

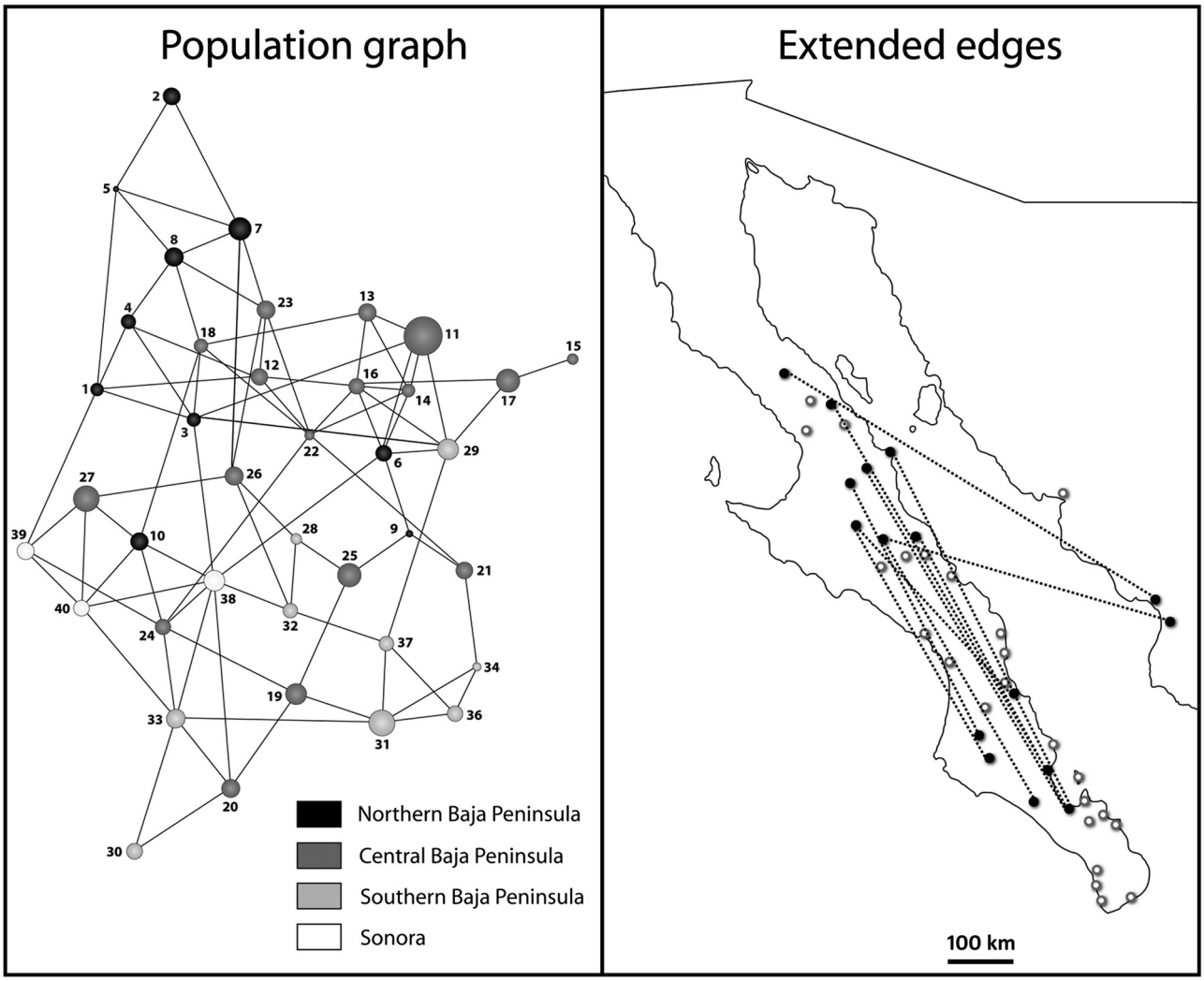

Fig. 6 Left panel: Population Graph for *Euphorbia lomelii* based on six nuclear loci. Spheres represent sampled populations (size is proportional to within-population genetic variance; shading indicates *a priori* region, population numbers follow Fig. 1). Thin black lines are retained edge sets and indicate genetic connectivity. Right panel: map showing only the extended edge sets (broken lines). Contributing localities are represented by solid black circles; open circles indicate noncontributing localities.

along the putative route of range expansion ($\Phi_{ST|Spatial} = 0.128$; Table 4). Both nuclear allelic richness and STAMOVA suggest that range expansion along the long axis of the Baja peninsula has played a major role in shaping spatial-genetic patterns seen in *E. lomelii*.

The Population Graph estimated from the nuclear data (Fig. 6, left panel) consisted of 39 populations (population 35 omitted due to small sample size; Appendix) and 87 edges. The overall topology was completely connected without obvious partitioning due to putative historical vicariance events. Using graph and physical distances, significant IBGD among Baja peninsula populations was detected (Mantel $Z = 16419.17$, $P = 0.001$; $\rho = 0.36$). Moreover, 10 of the 72 edges in the Population Graph were found to connect populations spatially separated by distances significantly farther than expected from the graph distances alone, indicative of long-distance dispersal. These extended edges (Fig. 6, right panel) are orientated along the peninsula's north–south axis and primarily link sampling localities in northern Baja with those from the southern extremity of central Baja. The placement of these extended edges support the hypothesis of recent range expansion (albeit with unspecified directionality) between these two subregions.

Discussion

Geographically concordant phylogeographical breaks in multiple animal taxa from the Baja peninsula may be indicative of landscape-level processes that impacted whole communities. However, a basic appreciation of how plants have responded to past climatic changes and geological events in this dynamic landscape system is only beginning

to emerge. The *Euphorbia lomelii* cpDNA data are consistent with a mid-peninsular vicariance event, and nuclear markers also support the partitioning of populations by both mid-peninsular and Isthmus of La Paz breaks. Moreover, the present work indicates that vicariance is only one of several microevolutionary forces that have impacted genetic structuring. *E. lomelii* nuclear data provide strong evidence for recent range expansion along the Baja peninsula, which has also been inferred for two Baja cactus species (Nason *et al*. 2002; Clark-Tapia & Molina-Freaner 2003). Multiple processes operating over a range of spatial and temporal scales can be difficult to model *a priori*, and the present work illustrates the advantages of a balance between hypothesis-testing and hypothesis-generating approaches to phylogeographical inference.

Chloroplast vs. nuclear genetic structure

Historical levels of seed-mediated gene flow are often an order of magnitude lower than those of pollen-mediated gene flow (e.g. Ennos 1994; El Mousadik & Petit 1996; Hamrick & Nason 2000). Consistent with this observation, spatial-genetic differentiation of *E. lomelii* cpDNA was structured at much finer scales than nuclear markers. Most haplotypes were restricted to (and fixed in) a single population (Fig. 5; Appendix), whereas all known nuclear alleles (excluding those with frequencies < 0.05) were present in the majority of populations, albeit at different frequencies. This is reflected by an estimated pollen-to-seed-mediated gene flow ratio of 0.21 (Fig. 2). Expected levels of phylogeographical congruence across marker types are impacted by these contrasting spatial scales of gene flow. Moreover, there can be considerable differences in the temporal scales over which genotypic, allelic, and phylogenetic data are most informative. The signal of more ancient vicariance events that impacted *E. lomelii* should at least partly be retained by slowly evolving, dispersal-limited cpDNA sequences. This is particularly relevant when large effective population sizes have been maintained (as with this euphorb; Fig. S2) coupled a glacier-free landscape setting, because fixation via genetic drift will proceed more slowly and historical signal should not be overwritten by large-scale extinctions. In contrast, recent range expansions are more likely to be marked by nuclear genotypes and their frequencies, which can change over short, generation-to-generation timescales (Sunnucks 2000). For these reasons, assessment of our phylogeographical hypotheses relied on a combination of DNA sequence and genotypic data.

Genetic connectivity across the Sea of Cortez

Even for organisms with moderate dispersal, the Sea of Cortez represents a major physiogeographical barrier to gene flow. This is evident for numerous taxa, including cactophilic *Drosophila* (Ross & Markow 2006; Pfeiler *et al*. 2007), the senita cactus (Nason *et al*. 2002), two rodent species and a toad (Riddle *et al*. 2000). However, considerable genetic covariance between Baja peninsula and continental Sonora *E. lomelii* populations was evident. Neither the cpDNA sequence data as examined from a coalescent simulation or NCPA perspective, nor the nuclear data as represented by the Population Graph and Geneland analyses, supported the *SoC* vicariance hypothesis. These data were more consistent with colonization or long-distance dispersal from the peninsula to the mainland. Although we did not expect to see evidence for trans-gulf migration, transportation of seeds over water (e.g. rafting or via seed-eating birds) seems biologically plausible. Interestingly, despite low sampling density, Sternburg & Rodriguez (1982) reported that *E. lomelii* hydrocarbon profiles indicate a strong affinity between southern Baja and mainland population samples (cf. southern vs. northern Baja population profiles). This observation indicates that a circum-gulf dispersal route is very unlikely, and lends support to our inference of gene flow across the Sea of Cortez.

Given that long-distance seed dispersal across the Sea of Cortez is likely to be rare for *E. lomelii*, genetic connectivity could be driven by recent colonization with little or no subsequent gene flow (e.g. Nielsen & Wakeley 2001). Three sources of evidence indicate that the directionality of colonization would have been from the peninsula to the mainland. First, all four continental Sonora cpDNA haplotypes occupy tip or recently derived positions on the parsimony network whereas most central and southern Baja haplotypes are interiors (Fig. 3)—a polarity indicative of haplotype age (Castelloe & Templeton 1994). The fact that continental Sonora contains two relatively distantly related groups of haplotypes (i.e. CH11 vs. CH07, CH08 and CH29) suggests that colonization involved several seeds of peninsular origin, although it is unclear whether the arrival of these long-distance dispersers would have been contemporaneous. Second, higher overall cpDNA diversity occurs on the Baja peninsula (Appendix), consistent with bottleneck effects during colonization of the mainland (Comps *et al*. 2001). Interestingly, two haplotypes (CH07 and CH29) are unique to continental populations, indicating that colonization events are not necessarily very recent, or at least occurred on timescales over which new DNA sequence mutations arise. Third, the geographical range of *E. lomelii* on continental Sonora is considerably smaller than the Baja peninsula, indicating a shorter occupancy of extant lineages on the mainland. However, we emphasize that the eastward long-distance dispersal scenario presented here does not preclude local extinction prior to subsequent recolonization of continental Sonora. Other plant species from the region show a similar distribution (e.g. Boojum tree, *Idria columnaris*), and therefore, landscape-level processes such as climate-driven extirpation on the mainland or

contraction into coastal refugia may have contributed to the peninsular-continental range size differences. Indeed, *E. lomelii* is the only one of *c.* 15 species in the *Pedilanthus* clade found on the Baja peninsula, yet all members are present on the Mexican mainland (Roberts 1989; Steinmann 2003) — a pattern inconsistent with a peninsular origin. It is also important to recognize that although rejection of an ancient (*c.* 5 Ma) *SoC* vicariance event by our coalescent approach is compelling, the assumption that vicariance proceeds according to a specific *a priori* model with no post-divergence gene flow may not hold. While NCPA is less restrictive, failure to detect an event may reflect lack of resolution in the sequenced gene regions and so inferences about population history must be viewed as incomplete.

Mid-peninsular vicariance

Our analyses of *E. lomelii* cpDNA lend support to a mid-peninsular vicariance event, and levels of incomplete lineage sorting are consistent with the hypothesized *c.* 1 Ma timescale. Although the bivariate summary statistic approach yielded only a modest improvement in sensitivity in our coalescent-based hypothesis tests (Table 2), the advantages could be more fully realized when attempting to discriminate among scenarios that differ rather subtly, or when post-divergence migration is incorporated into the models. NCPA corroborated a mid-peninsular vicariance event (Table 3) via an inference of past fragmentation — an inference that has low type I and type II error rates (Templeton 2004; Panchal & Beaumont 2007). However, when considering $K = 3$ natural clusters, the GENELAND analysis also supported predictions of the *IoL* scenario with respect to spatial-genetic structure, as did comparison of AMOVA tests conducted using nuclear data from peninsular populations only. Although disagreement between marker types is not uncommon in phylogeographical studies, this does not need to be invoked here. For example, it is possible that the nuclear genetic discontinuity across the Isthmus reflects a relatively recent (e.g. Holocene) origin of restricted pollen dispersal, in which case there is no expectation the cpDNA will be structured in a similar manner. Moreover, some authors speculate that there may have been multiple *IoL* vicariance events throughout the Miocene and Quaternary (e.g. Lindell *et al*. 2008), yet our coalescent simulation analyses of *E. lomelii* cpDNA explicitly tested for a splitting event at least as old as 3 Ma. Within this analytical framework, marine inundations that postdate the chosen timeframe of interest will go unrecognized.

Spatially clustered mid-peninsular phylogeographical breaks have been reported for herpetofauna (Upton & Murphy 1997; Rodríguez-Robles & De Jesús-Escobar 2000; Douglas *et al*. 2006; Recuero *et al*. 2006), birds (Zink *et al*. 2001), several small mammals (Riddle *et al*. 2000) and a spider (Crews & Hedin 2006), and the presence of a seaway *c.* 1 Ma has often been cited as the cause of this suture zone (Riddle *et al*. 2000; Lindell *et al*. 2006). Nonetheless, desert habitats were repeatedly fragmented during cool and wet pluvial periods (Riddle & Hafner 2006), particularly during the mid-Late Pleistocene when the amplitude of climatic oscillations increased (Riddle *et al*. 2008), and Leaché *et al*. (2007) recently provided evidence for multiple temporally distinct mid-peninsular vicariance events affecting several vertebrate species. Of particular relevance to plants, however, is that the location of the hypothesized mid-peninsular seaway is coincident with a marked transition between two major phytogeographical provinces occupying distinct desert subregions (i.e. the Vizcaíno Desert to the north and the Magdalena Desert to the south; Roberts 1989; Grismer 2002a, b). These subregions are separated by volcanic mountain ranges (Sierra San Francisco and the Volcán Las Tres Vírgenes) where extensive lava flows over the past 1.2 million years (Riddle *et al*. 2008), and perhaps as recently as the Late Holocene, effectively wiped out existing vegetation (Roberts 1989). Such abrupt environmental gradients and repeated sources of vicariance have resulted in localized, harsh, low-quality habitats (Jacobs *et al*. 2004), and thus, there may be no need to invoke a mid-peninsular seaway to account for the spatial-genetic patterns seen in *E. lomelii*.

Range expansion along the Baja peninsula

Nuclear data from *E. lomelii* support the notion that range expansion has played a major role in shaping spatial patterns of intraspecific diversity (Table 4; Figs 5 and 6). Phylogeographical studies of low-mobility, arid-adapted arthropods from southwestern North American deserts also provide evidence for range expansions during the Holocene and earlier (e.g. Ayoub & Riechert 2004; Smith & Farrell 2005; Pfeiler *et al*. 2007). Similarly, Fehlberg & Ranker (2009) reported that cpDNA sequence data from brittlebush (*Encelia farinosa*) were consistent with postglacial range expansion in parts of the Sonoran Desert and surrounding regions. However, the northward expansions along the Baja peninsula inferred for two similarly distributed columnar cacti (Nason *et al*. 2002; Clark-Tapia & Molina-Freaner 2003) were not seen in *Euphorbia lomelii*. Allelic richness values, which are well-suited to detecting the impact of bottlenecks associated with range expansion (Comps *et al*. 2001), suggest two expansion events (Fig. 5). The steady increase of nuclear allelic richness with increasing latitude is indicative of southward expansions of *E. lomelii*. Nonetheless, the inferred difference in directionality of range expansion (i.e. this study vs. Nason *et al*. 2002 and Clark-Tapia & Molina-Freaner 2003) is not irreconcilable. For example, in *E. lomelii*, it is possible that the distinct southern Baja relationship between allelic richness and latitude is driven by admixture between historically

isolated gene pools on either side of the Isthmus of La Paz (e.g. Kolbe *et al*. 2004). While this secondary contact scenario is inconsistent with the latitudinal pattern of allelic richness in northern and central Baja populations, the observed northern Baja peak in nuclear genetic diversity could be attributable to higher rates of pollen-mediated gene flow from continental Sonora populations. It is also noteworthy that the cpDNA data showed no evidence of unidirectional, north–south orientated range expansion. Indeed, the central Baja region harbours the most ancestral cpDNA haplotypes (i.e. CH09, CH22 and CH23), and the most phylogenetic diversity (Fig. 3). The fact that this pattern is inconsistent with southward range expansion(s) inferred from nuclear markers does not necessarily reflect conflict in the data — it may be a consequence of the different timescales over which DNA sequence vs. genotypic data tend to be most informative (Crandall *et al*. 2000; Sunnucks 2000; Garrick *et al*. 2008a).

Complementarity of analyses

In the present study, phylogeographical inferences drew on the duality of two classes of analyses, which we generalize as hypothesis-testing and hypothesis-generating, but their caveats warrant consideration. Scenario-based simulation approaches treat the empirical data set as just one of many possible realizations of a given population history. Coalescent analysis of *E. lomelii* cpDNA data exemplifies some benefits of a hypothesis-testing framework: historical signal can be recovered even in the face of incomplete lineage sorting, and the probability of the observed spatial-genetic patterns under a particular scenario can be quantified (Knowles 2004). One particularly limiting feature, however, is the dependence on heavily parameterized demographic or population divergence models. Indeed, any interpretive framework that requires the full phylogeographical model to be specified *a priori* may fail if the true history is not included in the set, or if it is embedded among one or more false events (Templeton 2009).

Hypothesis-generating procedures essentially treat empirical molecular data as being directly informative about the past, and reconstruct population history *a posteriori*. Because this framework does not require the full phylogeographical model to be specified at the outset, few parameters must be fixed (e.g. Templeton *et al*. 1995; Dyer & Nason 2004; Guillot *et al*. 2005), and it also enables consideration of a broad array of histories including unanticipated discoveries. Although the 'evolutionary portraits' generated using these approaches may be painted with broad brush strokes, they usually reveal some major components of an organism's history that warrant further examination using other methods. The major limitation, however, is that it can be difficult to assess the superiority of the estimated historical scenario against alternatives. Taken together, there is considerable scope for integration of phylogeographical analyses that are not subject to the same assumptions and limitations.

Acknowledgements

This work was funded by National Science Foundation grant DEB-0543102 to R.J.D. and J.D.N. The Center for the Study of Biological Complexity at Virginia Commonwealth University provided access to sequencing facilities, Ivalú Cacho supplied outgroup DNA samples, and Greg Plunkett provided access to some software. We also thank three reviewers for comments on the manuscript.

References

Ayoub NA, Riechert SE (2004) Molecular evidence for Pleistocene glacial cycles driving diversification of a North American desert spider, *Agelenopsis aperta*. *Molecular Ecology*, **13**, 3453–3465.

Carstens BC, Brunsfeld SJ, Dembowski JR, Good JM, Sullivan J (2005) Investigating the evolutionary history of the Pacific Northwest mesic forest ecosystem: hypothesis-testing within a comparative phylogeographic framework. *Evolution*, **59**, 1639–1652.

Castelloe J, Templeton AR (1994) Root probabilities for intraspecific gene trees under neutral coalescent theory. *Molecular Phylogenetics and Evolution*, **3**, 102–113.

Chen C, Durand E, Forbes F, François O (2007) Bayesian clustering algorithms ascertaining spatial population structure: a new computer program and comparison study. *Molecular Ecology Notes*, **7**, 747–756.

Clark-Tapia R, Molina-Freaner F (2003) The genetic structure of a columnar cactus with a disjunct distribution: *Stenocereus gummosus* in the Sonoran desert. *Heredity*, **90**, 443–450.

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1657–1659.

Comps B, Gömöry D, Letouzey J, Thiébaut B, Petit RJ (2001) Diverging trends between heterozygosity and allelic richness during postglacial colonization in the European Beech. *Genetics*, **157**, 389–397.

Crandall KA, Bininda-Emonds ORP, Mace G, Wayne RK (2000) Considering evolutionary processes in conservation biology. *Trends in Ecology & Evolution*, **15**, 290–295.

Crews SC, Hedin M (2006) Studies of morphological and molecular phylogenetic divergence in spiders (Araneae: *Homalonychus*) from the American southwest, including divergence along the Baja California Peninsula. *Molecular Phylogenetics and Evolution*, **38**, 470–487.

DeChaine EG, Martin AP (2006) Using coalescent simulations to test the impact of Quaternary climate cycles on divergence in an alpine plant–insect association. *Evolution*, **60**, 1004–1013.

Douglas ME, Douglas MR, Schuett GW, Porras LW (2006) Evolution of rattlesnakes (Viperidae; *Crotalus*) in the warm deserts of western North America shaped by Neogene vicariance and Quaternary climate change. *Molecular Ecology*, **15**, 3353–3374.

Dressler RL (1957) The genus *Pedilanthus* (Euphorbiaceae). *Contributions of the Gray Herbarium*, **182**, 1–188.

Dyer RJ (2009) GeneticStudio: a suite of programs for spatial analysis of genetic-marker data. *Molecular Ecology Resources*, **9**, 110–113.

Dyer RJ, Nason JD (2004) Population graphs: the graph theoretic shape of genetic structure. *Molecular Ecology*, **13**, 1713–1727.

Dyer RJ, Westfall RD, Sork VL, Smouse PE (2004) Two-generation analysis of pollen flow across a landscape. V. A stepwise approach for extracting factors contributing to pollen structure. *Heredity*, **92**, 204–211.

El Mousadik A, Petit RJ (1996) Chloroplast DNA phylogeography of the argan tree of Morocco. *Molecular Ecology*, **5**, 547–555.

Emerson BC, Paradis E, Thébaud C (2001) Revealing the demographic histories of species using DNA sequences. *Trends in Ecology & Evolution*, **16**, 707–716.

Ennos RA (1994) Estimating the relative rates of pollen and seed migration among plant populations. *Heredity*, **72**, 250–259.

Excoffier L, Smouse P, Quattro J (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics*, **131**, 479–491.

Fehlberg SD, Ranker TA (2009) Evolutionary history and phylogeography of *Encelia farinosa* (Asteraceae) from the Sonoran, Mojave, and Peninsular Deserts. *Molecular Phylogenetics and Evolution*, **50**, 326–335.

Garrick RC, Rowell DM, Simmons CS, Hillis DM, Sunnucks P (2008a) Fine-scale phylogeographic congruence despite demographic incongruence in two low-mobility saproxylic springtails. *Evolution*, **62**, 1103–1118.

Garrick RC, Dyer RJ, Beheregaray LB, Sunnucks P (2008b) Babies and bathwater: a comment on the premature obituary for nested clade phylogeographical analysis. *Molecular Ecology*, **17**, 1401–1403.

Garrick RC, Meadows CA, Nicolas AN, Nason JD, Dyer RJ (2008c) A set of polymorphic nuclear intron markers for conservation genetics and phylogeography of *Euphorbia* species (*Pedilanthus* clade). *Conservation Genetics*, **9**, 1673–1676.

Grismer LL (2002a) A re-evaluation of the evidence for a mid-Pleistocene mid-peninsular seaway in Baja California: a reply to Riddle *et al*. *Herpetological Review*, **33**, 15–16.

Grismer LL (2002b) *Amphibians and Reptiles of Baja California Including its Pacific Islands and Islands of in the Sea of Cortés*, pp. 1–39. University of California Press, Berkeley, California.

Guillot G, Mortier F, Estoup A (2005) Geneland: a computer package for landscape genetics. *Molecular Ecology Notes*, **5**, 712–715.

Hamilton MB, Miller JR (2002) Relative rates of pollen and seed gene flow in the island model using nuclear and organelle measures. *Genetics*, **162**, 1897–1909.

Hamrick JL, Nason JD (2000) Gene flow in forest trees. In: *Forest Conservation Genetics: Principles and Practice* (eds Boyle TJB Young A Boshier D), pp. 81–90. CIFOR and CSIRO, Australia.

Hewitt GM (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*, **58**, 247–276.

Hey J, Machado CA (2003) The study of structured populations — new hope for a difficult and divided science. *Nature Reviews Genetics*, **4**, 636–643.

Hurtado LA, Erez T, Castrezana S, Markow TA (2004) Contrasting population genetic patterns and evolutionary histories among sympatric Sonoran Desert cactophilic *Drosophila*. *Molecular Ecology*, **13**, 1365–1375.

Jacobs DK, Haney TA, Louie KD (2004) Genes, diversity, and geologic processes on the Pacific coast. *Annual Review of Earth and Planetary Sciences*, **32**, 601–652.

Kalinowski ST (2005) HP-Rare 1.0: a computer program for performing rarefaction on measures of allelic richness. *Molecular Ecology Notes*, **5**, 187–189.

Knowles LL (2004) The burgeoning field of statistical phylogeography. *Journal of Evolutionary Biology*, **17**, 1–10.

Knowles LL, Maddison WP (2002) Statistical phylogeography. *Molecular Ecology*, **11**, 2623–2635.

Kolbe JJ, Glor RE, Rodriguez-Schettino L *et al*. (2004) Genetic variation increases during biological invasion by a Cuban lizard. *Nature*, **431**, 177–181.

Kuhner MK, Yamato J, Felsenstein J (1998) Maximum likelihood estimation of population growth rates based on the coalescent. *Genetics*, **149**, 429–434.

Kumar S, Gadagkar SR (2000) Efficiency of the neighbor-joining method in reconstructing deep and shallow evolutionary relationships in large phylogenies. *Journal of Molecular Evolution*, **51**, 544–553.

Leaché AD, Crews SC, Hickerson MJ (2007) Two waves of diversification in mammals and reptiles of Baja California revealed by hierarchical Bayesian analysis. *Biology Letters*, **3**, 646–650.

Lindell J, Méndez-de la Cruz FR, Murphy RW (2008) Deep biogeographical history and cytonuclear discordance in the black-tailed brush lizard (*Urosaurus nigricaudus*) of Baja California. *Biological Journal of the Linnean Society*, **94**, 89–104.

Lindell J, Ngo A, Murphy RW (2006) Deep genealogies and the mid-peninsular seaway of Baja California. *Journal of Biogeography*, **33**, 1327–1331.

Maddison W (1997) Gene trees in species trees. *Systematic Biology*, **46**, 523–536.

Maddison WP, Maddison DR (2007) *Mesquite: A Modular System for Evolutionary Analysis*, Version 2.1, http://mesquiteproject.org.

Nason JD, Hamrick JL, Fleming TH (2002) Historical vicariance and postglacial colonization effects on the evolution of genetic structure in *Lophocereus*, a Sonoran Desert columnar cactus. *Evolution*, **56**, 2214–2226.

Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Nei M, Li W-H (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences, USA*, **76**, 5269–5273.

Nielsen R, Wakeley J (2001) Distinguishing migration from isolation: a Markov chain Monte Carlo approach. *Genetics*, **158**, 885–896.

Oddou-Muratorio S, Petit RJ, Le Guerroue B, Guesnet D, Demesure B (2001) Pollen-versus seed-mediated gene flow in a scattered forest tree species. *Evolution*, **55**, 1123–1135.

Panchal M, Beaumont MA (2007) The automation and evaluation of nested clade phylogeographic analysis. *Evolution*, **61**, 1466–1480.

Petit RJ, Duminil J, Fineschi S *et al*. (2005) Comparative organization of chloroplast, mitochondrial and nuclear diversity in plants. *Molecular Ecology*, **14**, 689–701.

Pfeiler E, Erez T, Hurtado LA, Markow TA (2007) Genetic differentiation and demographic history in *Drosophila pachea* from the Sonoran Desert. *Hereditas*, **144**, 63–74.

Pfenninger M, Posada D (2002) Phylogeographic history of the land snail *Candidula unifasciata* (Helicellinae, Stylommatophora): fragmentation, corridor migration, and secondary contact. *Evolution*, **56**, 1776–1788.

Posada D, Crandall KA (1998) ModelTest: testing the model of DNA substitution. *Bioinformatics*, **14**, 817–818.

Posada D, Crandall KA, Templeton AR (2000) GeoDis: a program for the cladistic nested analysis of the geographical distribution of genetic haplotypes. *Molecular Ecology*, **9**, 487–488.

Posada D, Crandall KA, Templeton AR (2006) Nested clade analysis statistics. *Molecular Ecology Notes*, **6**, 590–593.

R Development Core Team (2008) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. http://www.r-project.org.

Recuero E, Martínez-Solano I, Parra-Olea G, García-París M (2006) Phylogeography of *Pseudacris regilla* (Anura: Hylidae) in western North America, with a proposal for a new taxonomic rearrangement. *Molecular Phylogenetics and Evolution*, **39**, 293–304.

Riddle BR, Dawson MN, Hadly EA *et al*. (2008) The role of molecular genetics in sculpting the future of integrative biogeography. *Progress in Physical Geography*, **32**, 173–202.

Riddle BR, Hafner DJ (2006) A step-wise approach to integrating phylogeographic and phylogenetic biogeographic perspectives on the history of the core North American warm deserts biota. *Journal of Arid Environments*, **66**, 435–461.

Riddle BR, Hafner DJ, Alexander LF, Jaeger JR (2000) Cryptic vicariance in the historical assembly of a Baja California Peninsular Desert biota. *Proceedings of the National Academy of Sciences, USA*, **97**, 14438–14443.

Roberts NC (1989) *Baja California Plant Field Guide*, pp. 1–43, Natural History Publishing Co., La Jolla, California.

Rodríguez-Robles JA, De Jesús-Escobar JM (2000) Molecular systematics of New World gopher, bull, and pinesnakes (*Pituophis*: Colubridae), a transcontinental species complex. *Molecular Phylogenetics and Evolution*, **14**, 35–50.

Ross CL, Markow TA (2006) Microsatellite variation among diverging populations of *Drosophila mojavensis*. *Journal of Evolutionary Biology*, **19**, 1691–1700.

Rozas J, Sánchez-De I, Barrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*, **19**, 2496–2497.

Schaal BA, Hayworth DA, Olsen KM, Rauscher JT, Smith WA (1998) Phylogeographic studies of plants: problems and prospects. *Molecular Ecology*, **7**, 465–474.

Smith CI, Farrell BD (2005) Range expansions in the flightless longhorn cactus beetles, *Moneilema gigas* and *Moneilema armatum*, in response to Pleistocene climate changes. *Molecular Ecology*, **14**, 1025–1044.

Smouse PE, Peakall R (1999) Spatial autocorrelation analysis of individual multiallele and multilocus genetic structure. *Heredity*, **82**, 561–573.

Steinmann VW (2003) The submersion of *Pedilanthus* into *Euphorbia* (Euphorbiaceae). *Acta Botanica Mexicana*, **65**, 45–50.

Sternburg C, Rodriguez E (1982) Hydrocarbons from *Pedilanthus macrocarpus* (Euphorbiaceae) of Baja California and Sonora, Mexico. *American Journal of Botany*, **69**, 214–218.

Sunnucks P (2000) Efficient genetic markers for population biology. *Trends in Ecology and Evolution*, **15**, 199–203.

Swofford DL (2002) PAUP*: *Phylogenetic Analysis Using Parsimony (*and Other Methods)*, Version 4. Sinauer Associates, Sunderland, Massachusetts.

Taberlet P, Gielly L, Pautou G, Bouvet J (1991) Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology*, **17**, 1105–1109.

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA 4: Molecular Evolutionary Genetics Analysis (MEGA) Software, Version 4.0. *Molecular Biology and Evolution*, **24**, 1596–1599.

Templeton AR (2002) Out of Africa again and again. *Nature*, **416**, 45–51.

Templeton AR (2004) Statistical phylogeography: methods for evaluating and minimizing inference errors. *Molecular Ecology*, **13**, 789–809.

Templeton AR (2009) Statistical hypothesis testing in intraspecific phylogeography: nested clade phylogeographical analysis vs. approximate Bayesian computation. *Molecular Ecology*, **18**, 319–331.

Templeton AR, Routman E, Phillips CA (1995) Separating population structure from population history: a cladistic analysis of the geographical distribution of mitochondrial DNA haplotypes in the tiger salamander, *Ambystoma tigrinum*. *Genetics*, **140**, 767–782.

Thompson RS, Anderson KH (2000) Biomes of western North America at 18 000, 6000 and 0 ^{14}C yr BP reconstructed from pollen and packrat midden data. *Journal of Biogeography*, **27**, 555–584.

Upton DE, Murphy RW (1997) Phylogeny of the side-blotched lizards (Phrynosomatidae: *Uta*) based on mtDNA sequences: support for a mid-peninsular seaway in Baja California. *Molecular Phylogenetics and Evolution*, **8**, 104–113.

Van Devender TR (2002) Environmental history of the Sonoran Desert. In: *Columnar Cacti and Their Mutualists: Evolution, Ecology, and Conservation* (eds Fleming TH Valiente-Banuet A), pp. 3–24. University of Arizona Press, Tucson.

Van Devender TR, Burgess TL, Piper JC, Turner RM (1994) Paleoclimatic implications of Holocene plant remains from the Sierra Bacha, Sonora, Mexico. *Quaternary Research*, **41**, 99–108.

Zink RM, Kessen AE, Line TV, Blackwell-Rago RC (2001) Comparative phylogeography of some aridland bird species. *Condor*, **103**, 1–10.

Ryan Garrick is a postdoc in the Dyer lab and is interested in phylogeography and population structure of terrestrial invertebrates and their host plants. John Nason is Professor at Iowa State University and has broad interests in plant-insect coevolution and comparative population structure. Crystal Meadows is a Masters student in the Dyer lab, working on parentage analysis of flowering Dogwood. Rodney Dyer is Assistant Professor at Virginia Commonwealth University and his research focuses on understanding how genes are dispersed in space and time.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1 Estimates of Θ for *Euphorbia lomelii*

Fig. S1 Small-scale survey of 30 published chloroplast DNA substitution rates for coding and non-coding gene regions from diverse taxa.

Fig. S2 Model parameters of alternative vicariance scenarios tested in *Euphorbia lomelii*.

Supplementary references

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

Appendix

Euphorbia lomelii collection localities and sample sizes. Sample sizes refer to number of plants per population screened for nuclear markers (N_{nDNA}) and chloroplast DNA (N_{cpDNA}) 'combined haplotypes' (CH; see Table 1)

				Sample sizes		
Region	Locality	Latitude	Longitude	N_nDNA	N_{cp}DNA	Chloroplast haplotypes
NBP	1	29.32541	–114.29353	10	6	CH15
	2	29.01457	–113.94486	3	3	CH15
	3	28.96651	–113.66787	10	6	CH14, CH15, CH26
	4	28.72796	–113.48973	10	6	CH15
	5	28.66056	–113.99141	10	6	CH15
	6	28.40846	–112.86985	10	6	CH15
	7	28.22308	–113.18263	10	6	CH21
	8	28.03661	–113.39991	10	6	CH21
	9	27.52944	–113.31609	10	6	CH20
	10	27.36320	–112.96400	6	5	CH19, CH20
CBP	11	27.40498	–112.52959	10	1	CH24
	12	27.20280	–112.40800	10	6	CH18
	13	27.18232	–112.66552	9	6	CH23
	14	27.03670	–112.98600	10	7	CH22
	15	26.94589	–112.04613	10	2	CH01, CH04
	16	26.24905	–112.40948	9	5	CH10
	17	26.20876	–111.37833	10	6	CH05, CH09, CH27
	18	26.01550	–111.35475	9	5	CH13, CH25
	19	25.91409	–112.08062	10	5	CH13
	20	25.60521	–111.32638	5	5	CH04
	21	25.55757	–111.21563	10	6	CH04
	22	25.34819	–111.60056	10	6	CH02, CH04
	23	25.02470	–111.67500	6	5	CH17
	24	24.87611	–110.69175	11	8	CH08, CH11, CH12
	25	24.74642	–111.54410	10	6	CH02
	26	24.58843	–110.74599	8	5	CH06
	27	24.21150	–110.95100	10	7	CH03
SBP	28	24.45879	–110.36857	10	4	CH04
	29	24.13389	–110.46236	10	7	CH13
	30	24.21441	–110.27252	8	6	CH02, CH04, CH16
	31	24.04380	–109.98900	10	7	CH04
	32	24.01950	–110.09600	10	6	CH04
	33	24.00789	–109.85071	7	5	CH04
	34	23.28550	–110.10429	10	5	CH30
	35	23.08984	–110.10910	1	1	CH09
	36	23.08000	–110.03000	5	4	CH09
	37	23.07570	–109.64869	11	4	CH09, CH28
Son	38	27.90509	–110.57436	9	6	CH08, CH11
	39	26.63783	–109.32700	6	6	CH07, CH08, CH29
	40	26.38014	–109.12633	9	7	CH08