MOLECULAR ECOLOGY

Molecular Ecology (2012) **21**, 4578–4592 doi: 10.1111/j.1365-294X.2012.05715.x

Phylogeography of speciation: allopatric divergence and secondary contact between outcrossing and selfing *Clarkia*

JAMES B. PETTENGILL and DAVID A. MOELLER

Department of Plant Biology, 1445 Gortner Avenue, University of Minnesota, Saint Paul, MN 55108, USA

Abstract

The origins of hybrid zones between parapatric taxa have been of particular interest for understanding the evolution of reproductive isolation and the geographic context of species divergence. One challenge has been to distinguish between allopatric divergence (followed by secondary contact) versus primary intergradation (parapatric speciation) as alternative divergence histories. Here, we use complementary phylogeographic and population genetic analyses to investigate the recent divergence of two subspecies of *Clarkia xantiana* and the formation of a hybrid zone within the narrow region of sympatry. We tested alternative phylogeographic models of divergence using approximate Bayesian computation (ABC) and found strong support for a secondary contact model and little support for a model allowing for gene flow throughout the divergence process (i.e. primary intergradation). Two independent methods for inferring the ancestral geography of each subspecies, one based on probabilistic character state reconstructions and the other on palaeo-distribution modelling, also support a model of divergence in allopatry and range expansion leading to secondary contact. The membership of individuals to genetic clusters suggests geographic substructure within each taxon where allopatric and sympatric samples are primarily found in separate clusters. We also observed coincidence and concordance of genetic clines across three types of molecular markers, which suggests that there is a strong barrier to gene flow. Taken together, our results provide evidence for allopatric divergence followed by range expansion leading to secondary contact. The location of refugial populations and the directionality of range expansion are consistent with expectations based on climate change since the last glacial maximum. Our approach also illustrates the utility of combining phylogeographic hypothesis testing with species distribution modelling and fine-scale population genetic analyses for inferring the geography of the divergence process.

Keywords: approximate Bayesian computation, *Clarkia xantiana*, divergence population genetics, hybrid zone, introgression, parapatry

Received 16 February 2012; revision received 21 May 2012; accepted 25 May 2012

Introduction

Studies of hybrid zones and parapatric distributions have been important for illuminating the ecological and evolutionary processes involved in speciation and the maintenance of species borders (Barton & Hewitt 1985; Bull 1991). Although hybrid zones are quite common in nature and have received extensive study (Barton & Hewitt 1985; Jiggins & Mallet 2000; Hewitt 2001; Brennan *et al*. 2009; Maroja *et al*. 2009; Arnold *et al*. 2011), the history of their formation remains poorly understood for most systems (Harrison 1993; Hewitt 2011). One challenge has been differentiating between secondary contact versus primary intergradation (Durrett *et al*. 2000). Under secondary contact, differentiation occurs primarily in allopatry followed by range expansion to a

Correspondence: David A. Moeller, Fax: (612) 625-1738; Email: moeller@umn.edu

zone of contact (Dobzhansky 1940; Mayr 1942; Coyne & Orr 2004). Under primary intergradation, divergence does not occur in allopatry but rather because of serial local adaptation along a resource gradient or spatially varying selection in a patchy environment (Fisher 1930; Endler 1977; Caisse & Antonovics 1978; Slatkin 1981). Theory has also shown that parapatric speciation can occur because of genetic incompatibilities alone, without disruptive ecological selection (Gavrilets & Li 2000).

Most attempts to distinguish secondary from primary contact have employed clinal analyses of genetic variation within contact zones (e.g. Morgan-Richards *et al.* 2000; Kawakami *et al.* 2008). The utility of such analyses rests on the assumption that the alternative divergence histories will produce different clinal patterns. For example, one prediction is that secondary contact should result in the coincidence (common centre) of clines across hybrid zones especially for neutral markers or characters (Endler 1977; Barton & Hewitt 1985; Durrett *et al.* 2000). Under primary intergradation, non-neutral loci are not expected to form coincident clines and neutral loci may not form clines at any location because of high levels of introgression (Durrett *et al.* 2000). Despite these contrasting theoretical expectations, it can be difficult to clearly distinguish between them using empirical data (Bull 1991; Coyne & Orr 2004). One problem is that when loci are purely neutral (and not linked to loci under selection) clines can degrade quickly following secondary contact resulting in patterns of genetic variation that are indistinguishable from primary intergradation (Durrett *et al.* 2000). The opposite may occur for loci under selection; coincident clines can form under primary intergradation if non-neutral loci respond similarly to an agent of selection (e.g. an environmental gradient; Durrett *et al.* 2000).

Alternatively, approaches that attempt to infer the historical processes that gave rise to extant geographic distributions can be used to differentiate among alternative explanations for parapatric distributions. Such approaches include approximate Bayesian computation (ABC; Beaumont *et al.* 2002) under which the posterior probability for competing phylogeographic hypotheses can be estimated (e.g. Fagundes *et al.* 2007; Ross-Ibarra *et al.* 2009) and methods employing the coalescent to infer the ancestral geography of recently diverged taxa (Lemey *et al.* 2009). A complementary approach is that of palaeodistribution modelling, which identifies potential historic geographic ranges without inferring genealogical relationships. Combining the two methods, phylogeography and niche modelling, is particularly powerful because they provide independent assessments of historical geographic distributions (Carstens & Richards 2007; Kozak *et al.* 2008). By identifying putative ancestral distributions, these historical approaches also provide insight into the relative importance of range expansions or contractions, vicariance events and extirpation in shaping extant genetic and distributional patterns (Carstens & Richards 2007; Swenson 2008; Chan *et al.* 2011).

Here, we examine the evolutionary history of two recently diverged parapatric subspecies of *Clarkia xantiana* A. Gray (Onagraceae). We used phylogeographic and population genetic approaches to discriminate between the alternative hypotheses of secondary contact and primary intergradation for the process of divergence and the formation of the current contact zone between the two subspecies. *Clarkia xantiana* ssp. *xantiana* A. Gray is primarily outcrossing, and *C. xantiana* ssp. *parviflora* (Eastw.) Harlan Lewis and P. H. Raven is primarily selfing, although both taxa are self-compatible (Runions & Geber 2000; Pettengill & Moeller 2012). Previous analyses suggest that they began to diverge within the past ~10 000–65 000 years (Pettengill & Moeller 2012) making it unlikely that the signature of speciation and the dynamics of the contact zone have been overwritten by numerous climatic events or range expansions and contractions. The subspecies are winter annuals endemic to the southern Sierra Nevada foothills and associated mountain ranges of Southern California, USA. Their distributions occupy different sections of a dominantly west-to-east environmental gradient where the western section, occupied by *xantiana,* receives more and less variable precipitation compared to the eastern section, occupied by *parviflora* (Eckhart *et al.* 2010; Eckhart *et al.* 2011). *Xantiana* is also typically found in blue oak, grey pine woodland, whereas *parviflora* occurs more commonly in xeric scrub or pinyon-juniper woodland. They overlap in distribution along a narrow zone (~5–10 km) at the eastern edge of *xantiana*'s and western edge of *parviflora*'s range (Eckhart & Geber 1999). In the zone of sympatry, subspecies often occur within metres to tens of metres of one another, and there is evidence for some introgression in sympatric sites (Pettengill & Moeller 2012).

To elucidate the history of divergence between *xantiana* and *parviflora*, we first tested alternative phylogeographic hypotheses for the divergence process using ABC. Second, we inferred whether allopatric populations of each taxon represent the ancestral geographic location using a coalescent-based Bayesian model of character state evolution. Third, we used palaeo-distribution models to estimate the historical distribution of suitable habitat at the last glacial maximum, which falls between two estimates for the divergence time of the two taxa (10 000–65 000 ybp (years before present); Pettengill & Moeller 2012). We also tested population genetic predictions about primary and secondary contact using clinal analyses based on hybrid indices from

molecular data. Last, we considered our results in the light of historical changes in climate to help identify the causes for the divergence history suggested from our analyses.

Materials and methods

Population sampling, DNA sequencing and microsatellite genotyping

We sampled five individuals from 31 populations (15 of *xantiana*, 16 of *parviflora*) spanning the species' distribution except for the Tehachapi Mountains, which are largely inaccessible (Fig. 1, Table S1, Supporting information). Each individual represented a different maternal family, and the five families were randomly chosen from a larger collection of families from across each population. Tissue was collected from leaves, and DNA was extracted using Qiagen DNeasy plant mini kits (Valencia, CA). We used PCR to amplify eight nuclear genomic regions (nDNA) and three chloroplast regions (cpDNA) from the 154 individuals. The PCR primers used to amplify single-copy nDNA loci were designed from EST sequences isolated from *Clarkia breweri* (see Moeller *et al*. 2011). The three cpDNA regions were *psbA-trnH* (Kress & Erickson 2007), *trnT-trnL* and *trnL-trnF* (Taberlet *et al*. 1991).

Below, we briefly describe the sequencing and phasing of individuals. For a complete description of methods, see Pettengill and Moeller (2012). Following successful amplification, PCR products were sequenced directly using the ABI BigDye v3.1 chemistry and an ABI 3730xl DNA sequencer. We inspected all chromatograms using SEQUENCHER v4.8 (Gene Codes Corp., Ann Arbor, MI), aligned sequences using MUSCLE (Edgar 2004) based on the default settings, and determined haplotype phase using PHASE V2.1 (Stephens *et al*. 2001; Stephens & Donnelly 2003). In total, we acquired 1473 sequences across loci and individuals (4751 bp), 811 were from *parviflora* samples and 662 were from *xantiana* samples (Table S2, Supporting information). Sequence variability across all loci was much greater within *xantiana* than *parviflora* (no. polymorphisms = 332 and 61, respectively; Table S2, Supporting information).

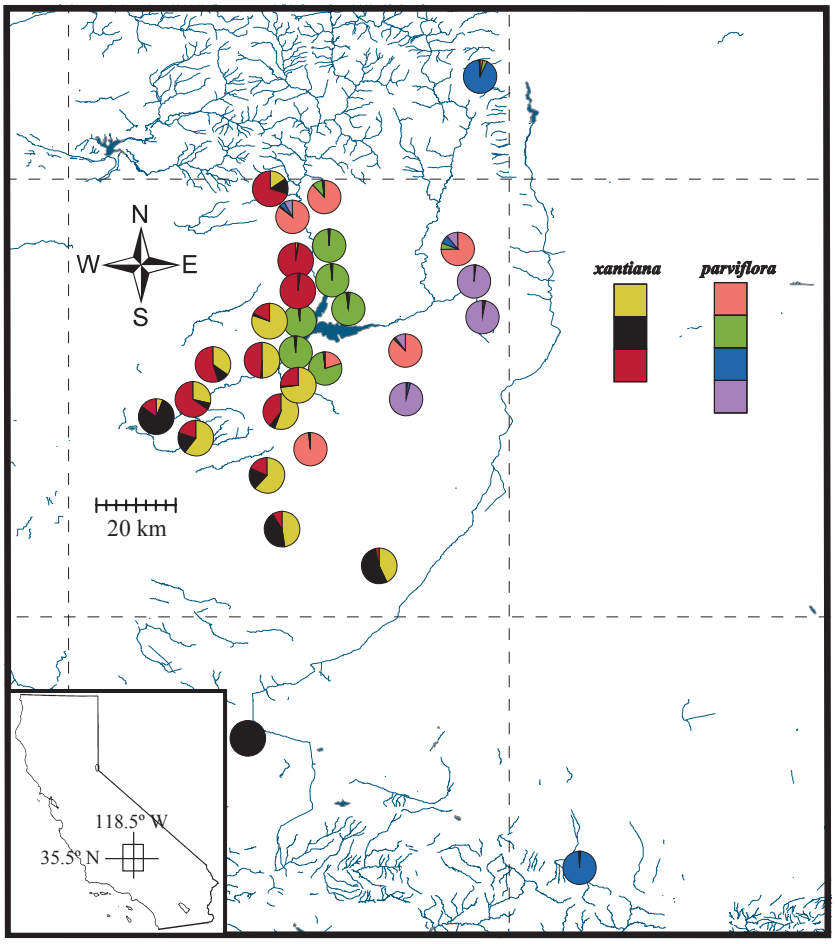

Fig. 1 Map of the distribution of populations sampled from *xantiana* and *parviflora*. Pie charts represent the fractional assignment of individuals within each population to either $k = 3$ or $k = 4$ clusters for *xantiana* and *parviflora* samples, respectively.

We also assayed variation at four dinucleotide-repeat microsatellite loci. Primer sequences and amplification conditions can be found in Moeller *et al*. (2011). We performed PCR separately for each locus using the 6-FAM and NED dyes and combined amplified products before loading onto an ABI 3730xl. Replicate independent PCR reactions were genotyped for a subset of individuals to confirm alleles. We examined electropherograms to determine allelic states. Microsatellite amplification success was nearly 100% with 90 alleles across the four loci (Table S3, Supporting information). Similar to sequence diversity, variability was greater in *xantiana* than *parviflora* (no. alleles = 71 and 37, respectively; Table S3, Supporting information).

Approximate Bayesian computation

We evaluated the statistical support for four alternative phylogeographic models of the divergence history of *xantiana* and *parviflora* (Fig. 2) using an approximate Bayesian computational (ABC) approach. In the first model, no gene flow occurs throughout the divergence process (isolation model). In the second model, gene flow occurs throughout the entire divergence process (island model). The other two models include gene flow but at different times during divergence. In the third, gene flow only occurs in the initial stages of divergence (sympatric model). In the fourth, gene flow occurs following an initial allopatric phase of divergence (secondary contact model). All models consider two diverging lineages: *xantiana* and *parviflora*.

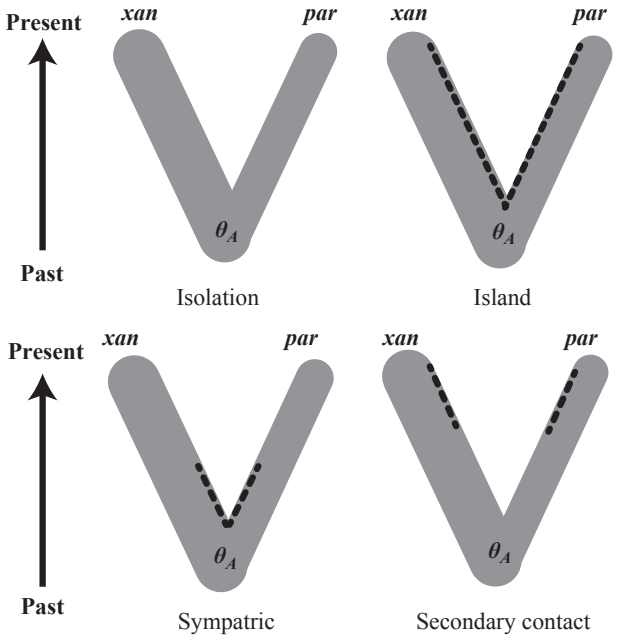

Fig. 2 Alternative phylogeographic models simulated under the neutral coalescent and evaluated using approximate Bayesian computation. Dashed lines indicate the time over which migration may occur under a given model. The two taxa, subspecies *xantiana* and *parviflora*, are indicated by ''*xan*'' and ''*par*'', respectively.

We ran 10^6 simulations under each model based on the neutral multilocus coalescent using the program MSABC (Pavlidis *et al.* 2010), which is a modified version of the program MS (Hudson 2002). We set the number of loci to eight (i.e. the number of nDNA loci we sampled), and sample sizes varied among loci reflecting the number of sequences we obtained for each locus. For the population mutation rate parameter, θ, we used uniform prior distributions that were broader than the empirical minimum and maximum estimate within each taxon across the eight loci (Table S4, Supporting information). We estimated Watterson's θ (Watterson 1975) separately for each population to obtain our minimum and maximum estimates. We also used a uniform prior on the estimate of divergence time, τ_D, which was bounded by twice the 95% credible interval estimate based on previous analyses with IMA2 (see Pettengill & Moeller 2012). We note that our objective was not to estimate divergence time using ABC; rather, we incorporated information on divergence time from more robust methods to test alternative phylogeographic hypotheses. Uniform prior distributions for the time at which migration, τ_M, starts (sympatric model) or stops (secondary contact model) and asymmetric rates of migration, $M_{xan \rightarrow par}$ and $M_{par \rightarrow xan}$, (island, secondary contact, sympatric models) were also used (Table S4, Supporting information). Because of the restriction that divergence must occur before migration can begin to occur, the prior distributions on τ_D and τ_M were not flat (Fig. 3). For the population recombination rate, ρ, we used the average estimate across loci.

For the ABC analysis, we estimated four summary statistics from both our observed nDNA data (using all sequenced sites for each taxon separately) and for the simulations described above: the mean and variance across loci of F_{ST} (Hudson *et al.* 1992) and the mean and variance across loci of the proportion of sites shared between taxa. We conducted model selection by combining the summary statistics and also with each summary statistic separately. We chose these summary statistics because they describe the degree of differentiation among lineages and, therefore, should be informative regarding the divergence process (e.g. Ross-Ibarra *et al*. 2009, Fagundes *et al*. 2007).

The posterior probability for a model was estimated as the proportion of accepted simulations produced under that model (i.e. the rejection method) and using the multinomial logistic regression approach (Beaumont 2008). Analyses performed with a range of tolerances

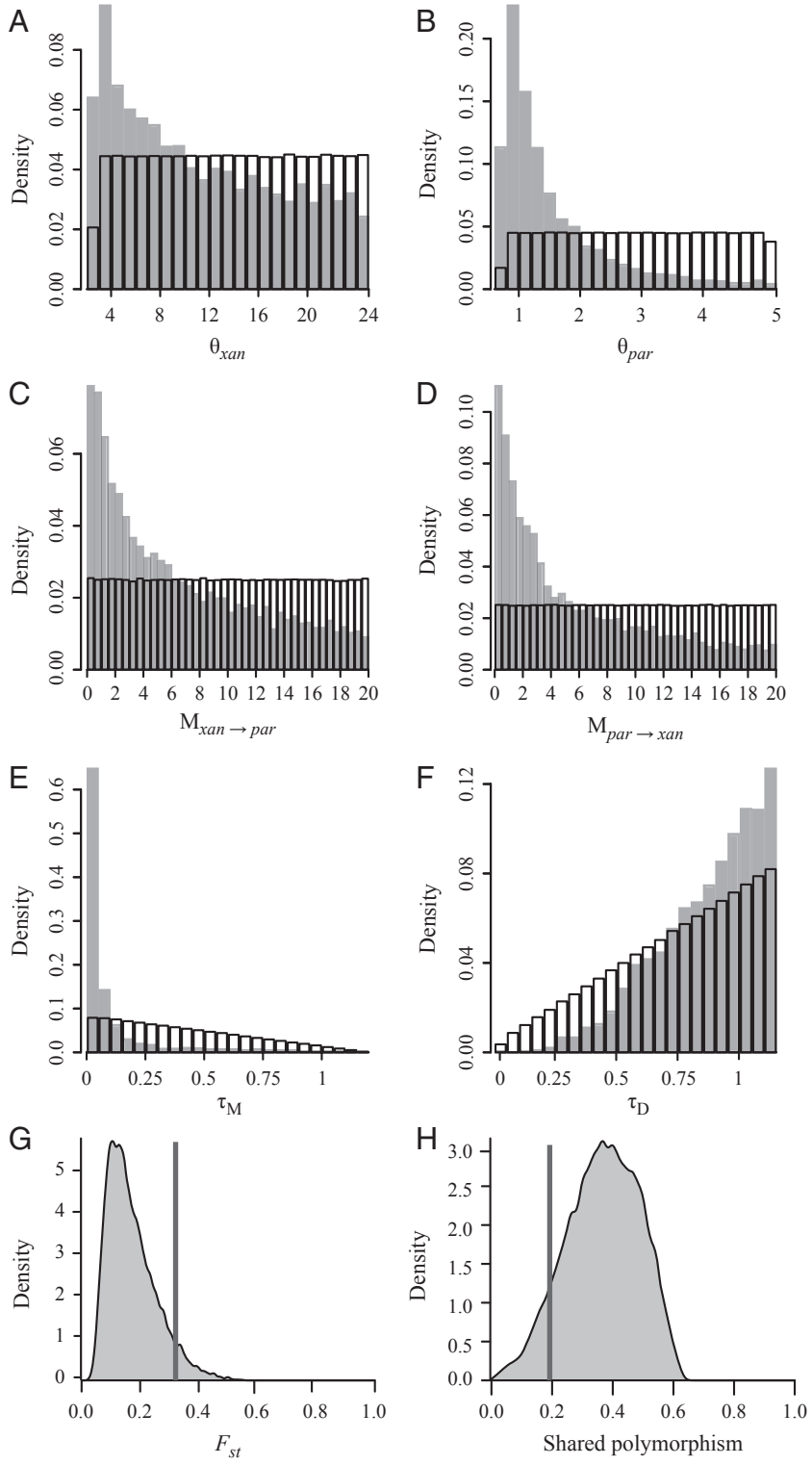

Fig. 3 Prior (open bars) and posterior probability (shaded bars) distributions for effective population sizes, θ_{xan} and θ_{par} (A and B), asymmetric migration rates, $M_{xan \rightarrow par}$ and $M_{par \rightarrow xan}$ (C and D), the timing at which migration stops, τ_M (E), and the time of divergence, τ_D (F), under the secondary contact model of divergence based on the ABC results. Time is units of $4N_e$ looking backward in time. Posterior predictive histograms under the secondary contact model for shared polymorphism (G) and F_{ST} (H); the line represents our observed value for the summary statistics.

(δ = 0.1, 0.05, and 0.02) did not change which model had the highest posterior probability; we present the results based on δ = 0.02. To evaluate the sufficiency of the summary statistics and the robustness of our results, we also conducted posterior predictive checks for the best-fitting model (Csilléry *et al.* 2012). This was accomplished by comparing our observed value for each summary statistic to the distribution produced by

10 000 resimulations using parameters from the posterior (i.e. parameters from 10 000 accepted simulations randomly chosen from all accepted simulations). All analyses were conducted using the R package ABC (Csilléry *et al*. 2012).

Ancestral geography

We reconstructed ancestral geography for each taxon using a Bayesian phylogeographic analysis implemented in BEAST v1.6.1 (Lemey *et al*. 2009). In this analysis, geography is a character taking on as many states as there are population locations and its evolution is modelled to infer the ancestral state. If taxa diverged in allopatry and subsequently came into secondary contact then we expect allopatric populations to have the highest posterior probability of being the ancestral state and sympatric populations to be the derived state as a result of a recent range expansion. By contrast, if the taxa are in primary contact then we expect populations proximal to the contact zone to be ancestral and allopatric populations to be more recently derived. The advantage of this method over heuristic approaches is that discrete geographic states are modelled using a standard continuous-time Markov chain within an analysis where phylogenetic relationships are also estimated and, thus, the analysis integrates over both phylogenetic uncertainty and Markov model parameter uncertainty (Lemey *et al*. 2009).

We ran analyses for each taxon separately. Because there is evidence for introgression in our data set, which may cause problems for ancestral state reconstructions that do not account for interspecific hybridization, we excluded samples from sympatric populations (i.e. 22x/22p, 5x/5p, 6x/6p, 74x/74p, 79x/77p and 9x/9p; Table S1, Supporting information). We have retained populations that lie just outside the narrow contact zone where we have detected no evidence of introgression. Preliminary analyses showed that the potential bias caused by including hybrids was greater for *parviflora* than *xantiana*; in the latter taxon, ancestral-state reconstructions were similar with and without sympatric populations.

All runs of BEAST included both the nDNA and cpDNA loci, which were assigned the appropriate inheritance scalars of 1 and 0.5, respectively. Analyses consisted of 10^8 generations with the first 10% of generations treated as burnin. Posterior probabilities for each geographic population as the ancestral (root) state were estimated from 10 000 trees. We identified the most appropriate nucleotide substitution model for each locus using MODELTEST (Posada & Crandall 1998). Each locus was assigned a separate molecular clock; for those loci that violated the assumptions of a molecular clock, we used a relaxed lognormal clock. The effective-sample size (ESS) values of parameters were monitored to ensure sufficient sampling of parameter space (i.e. >200). We present the average from 10 replicate runs of each data set.

Species distribution modelling

We used species distribution modelling to identify the range of suitable habitat at present and at the last glacial maximum (LGM), ~21 000 ybp, using a maximum entropy approach (MaxEnt; Phillips *et al*. 2006). Predictions of a taxon's geographic distribution were based on presence-only data for a set of georeferenced occurrence locations and layers representing environmental variables. Current and historical (LGM) layers (19 environmental variables and elevation) were obtained at the resolution of 2.5 arc-min from the WorldClim database (Hijmans *et al*. 2005), which is based on data from the PMIP2-project (http://pmip2.lsce.ipsl.fr). Results are based on the Community Climate System Model version 3 (CCSM3; Collins *et al*. 2006) projections. We assumed that topography has not changed since the LGM and used the current projection for elevation as a layer for historical projections. All models used 25% of presence localities to test for model performance, included 100 bootstrap replicates, and used the linear quadratic feature type; all other parameters were set to default values. We used the area under the receiver operator curve (AUC, which ranges from 0.5 to 1.0 with the former indicating predictions no better than random) as a measure of model accuracy.

Phylogeographic structure

We examined phylogeographic structure using INSTRUCT v3.2.09 (Gao *et al*. 2007), which infers clusters (k) based on deviations from Hardy–Weinberg because of population structure and inbreeding. This method is well suited for our system given that *parviflora* is primarily selfing and *xantiana* is mixed mating. We used both haplotype (cpDNA and phased nDNA) and microsatellite data for these analyses. We first used individuals from both subspecies in an analysis with k fixed at 2 to examine the association between taxonomic designation and cluster assignment. Because nearly all individuals were clustered among conspecifics, we conducted separate analyses within each taxon to examine population structure. For all analyses, we ran two independent MCMC chains of 10^5 generations with 10^3 generations as burnin under a model allowing for admixture with the selfing rate estimated at the population level. For the analysis of all individuals with k fixed at 2, we present the results from the MCMC chain with the lowest deviance information criterion (DIC). For intrasubspeci-

fic analyses, we evaluated $k = 1$ through 10 and identified the most likely number of clusters based on the value of k with the lowest DIC value. Results were visualized using DISTRUCT (Rosenberg 2004).

Clinal analyses

We used clinal analyses to examine patterns of genetic variation across the contact zone and to evaluate the strength of the barrier to gene flow between taxa. For our analyses, we estimated the distance to the nearest heterospecific population for each population. The subspecies sort primarily along a west—east axis, with *xantiana* to the west and *parviflora* to the east, and distances to the nearest heterospecific population are negative for *xantiana* and positive for *parviflora* populations. For a hybrid index, we used the results from the INSTRUCT analyses with all individuals for $k = 2$. Therefore, hybrid indices are calculated as the probability that an individual is assigned to one of two populations under which a hybrid index of 1 or 0 indicates 'pure' *xantiana* or 'pure' *parviflora* individuals, respectively. An intermediate value suggests that individuals are of admixed ancestry. We estimated hybrid indices for all loci combined and for each of the three marker types separately: microsatellites, nDNA and cpDNA.

Characteristics of clines were estimated using the maximum-likelihood method implemented in CFIT7 (Gay *et al.* 2008). The AICc was used to determine the best fitting of five clinal models that differ in the degree of introgression that can occur within a hybrid zone. The five models were unimodal, bimodal, bimodal with introgression, trimodal and trimodal with introgression. Briefly, a unimodal model fits all individuals with a single distribution (i.e. a hybrid swarm; Jiggins & Mallet 2000), a bimodal model fits each parental taxon with a separate normal distribution (i.e. no introgression) and a trimodal model fits a third distribution to admixed individuals within the contact zone (i.e. distinct parental taxa plus admixed individuals). After determining the best-fitting model for each type of molecular marker, we tested for coincidence and concordance among clines by analysing five different models that varied in the number of parameters that were forced to be the same among the clines: (i) all parameters including width, centre and those associated with the two exponential tails of the distribution (Gay *et al.* 2008), (ii) only the centre and width, (iii) only the centre, (iv) only the width and (v) none. The significance of differences between the fit of constrained models relative to an unconstrained model was assessed using the AICc. We eliminated the southernmost populations (26x and 27p; Fig. 1) from clinal analyses because we do not know the location of a contact zone in this region. Results were congruent between analyses with and without those two populations.

Results

ABC phylogeographic models

The secondary contact model was the most probable model when all summary statistics were used under both the rejection method [posterior probability (PP) = 0.72] and the multinomial logistic regression method (PP = 0.99; Table 1). The same inference was made when only the proportion of shared polymorphisms was used in the model selection process (using both methods) and when F_{ST} was evaluated with multinomial logistic regression. When F_{ST} was evaluated with the rejection method, we found strongest support for the isolation model.

We also estimated the parameters θ_{xan}, θ_{par}, τ_M (time when migration starts or stops), $M_{xan \rightarrow par}$ and $M_{par \rightarrow xan}$ and τ_D (divergence time) under the secondary contact model of divergence, which was the best-fitting model based on all summary statistics. The results suggest that migration has occurred only recently and for a short time relative to the entire coalescent process [mode $\tau_M = 0.018$ or ~2300 ybp based on a mutation rate of 1.5×10^{-8} (Koch *et al.* 2001) substitutions per site per year; see Pettengill and Moeller 2012 for parameter conversions] and that the magnitude of migration was low ($M_{xan \rightarrow par} = 1.169$; $M_{par \rightarrow xan} = 0.897$; Fig. 3). Posterior predictive checks confirmed that our observed summary statistics fall within distributions created using parameter estimates from 10 000 accepted simulations in our initial analysis (Fig. 3).

Ancestral geography

We found strong support within the derived taxon, *parviflora*, that eastern, allopatric populations are ancestral

Table 1 Phylogeographic model selection results based on posterior probabilities under the rejection (*rej*) and multinomial logistic regression (*mnlog*) methods using all summary statistics, only the proportion of shared polymorphisms and F_{ST}. The model with the highest posterior probability is in bold font

	(All)		(Shared)		(F_{ST})	
Model	*rej*	*mnlog*	*rej*	*mnlog*	*rej*	*mnlog*
Island	0.254	0.002	0.004	0	0.004	0.013
Secondary	**0.723**	**0.998**	**0.694**	**1**	0.073	**0.989**
Sympatric	0.017	0	0.012	0	0.356	0
Isolation	0.008	0	0.004	0	**0.567**	0.001

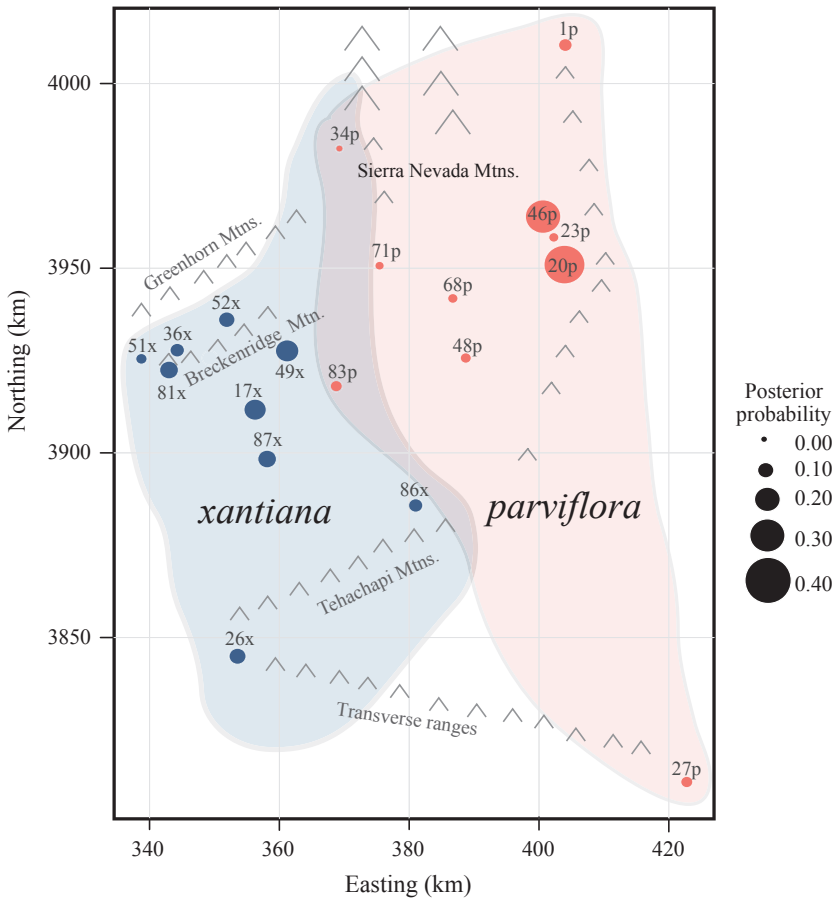

Fig. 4 Ancestral reconstructions for geography based on the analyses from BEAST for two data sets representing *xantiana* and *parviflora*. Sizes of circles indicate the posterior probability of a given location being the root state.

(Fig. 4). Populations 20p and 46p had the highest posterior probabilities of 0.36 and 0.30, respectively. All other *parviflora* populations had posterior probabilities <0.1 (Table S1, Supporting information, Fig. 4). In the progenitor taxon, *xantiana*, we found little conclusive support for an ancestral geographic region. However, populations south of Breckenridge Mountain and the Kern River Valley (especially 49x, 17x, 81x and 87x) have somewhat higher posterior probabilities than more northern populations in the Kern River Valley (36x, 51x and 52x) (Table S1, Supporting information, Fig. 4).

Species distribution modelling

The MAXENT model for both *xantiana* and *parviflora* had high predictive performance with AUC values >0.9 (standard deviations <0.001) for the test data and the training data. Overall, projections based on the current environmental layers suggested that the distribution of the most suitable habitat for both *xantiana* and *parviflora* is generally restricted to the area where they are currently found (Fig. 5). The strongest contributing environmental variables for *xantiana* were as follows: a negative effect of precipitation of the warmest quarter (27.1%), a positive effect of minimum temperature of the coldest month (15.8%), a negative effect of elevation (13.9%) and a unimodal effect of precipitation of the coldest quarter (11.6%). The strongest contributing environmental variables for *parviflora* were similar to those for *xantiana* but minimum temperature in the coldest month contributed the most (30.6%) followed by precipitation in the warmest quarter (20.2%) and precipitation in the coldest quarter (13.4%). The direction of the effects of those variables was the same as those observed for *xantiana*, except that precipitation of the coldest quarter had a positive effect.

The LGM projections for *xantiana* and *parviflora* do not substantially overlap with their current distributions (Fig. 5). For *xantiana*, the most suitable habitat is a relatively small geographic region found at the south-eastern edge and outside its current distribution near the junction of the Sierra Nevada and Tehachapi Mountains; little suitable territory at the LGM exists to the north, where it is primarily found today. For *parviflora*, the most suitable habitat at the LGM is primarily distributed along the eastern slope of the Tehachapi and Sierra Nevada Mountains. The south-eastern section overlaps with the most suitable region for *xantiana* at

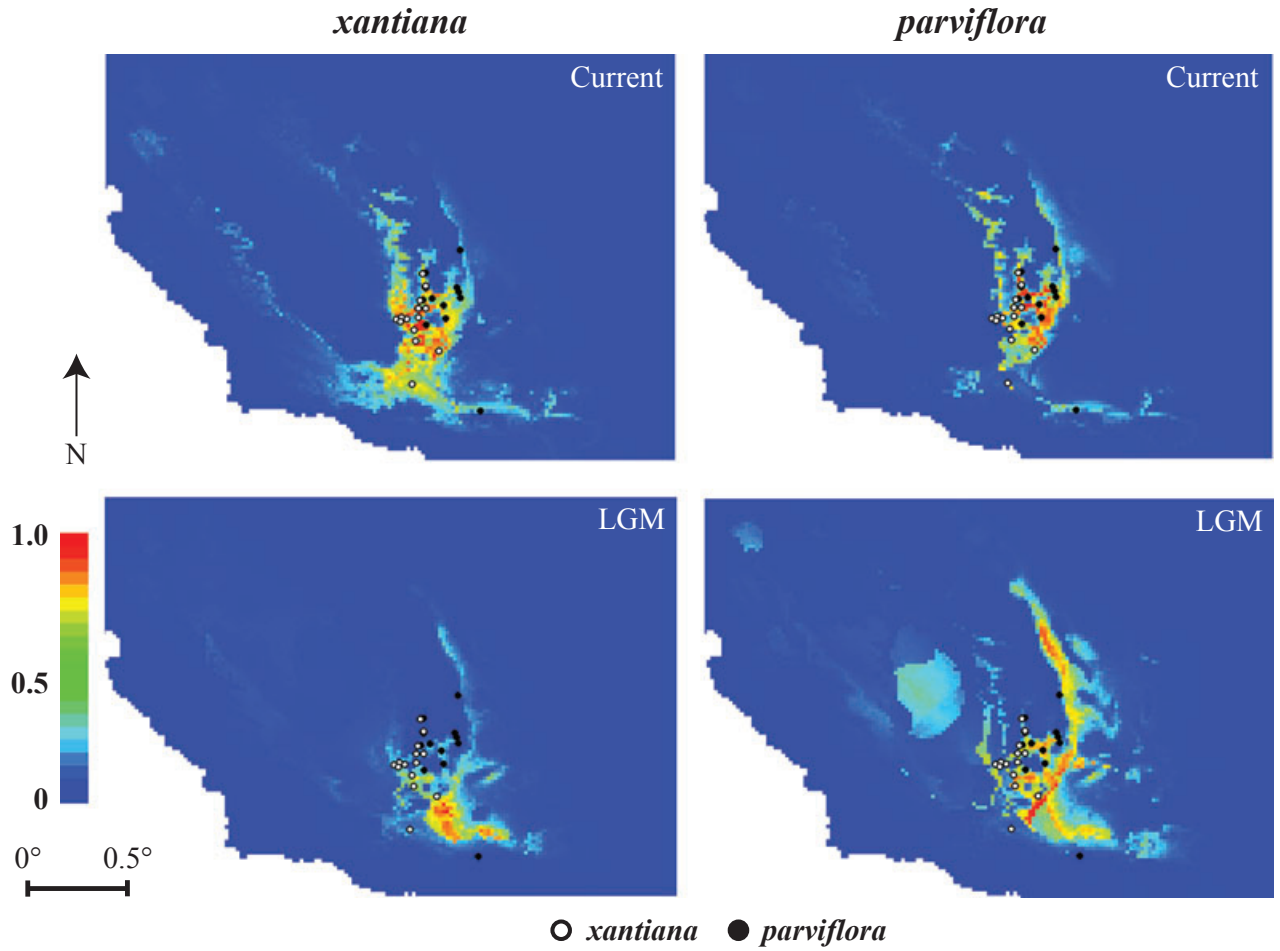

Fig. 5 MaxEnt models projecting the areas of suitable habitat for *xantiana* and *parviflora* based on environmental conditions for the populations included in this study. Upper panels show projections based on current climatic data (i.e. 1950–2000) and lower panels show projections based on environmental conditions at the LGM (21 000 bp)

the LGM. For *parviflora*, there is also a section of suitable territory to the north of its current known distribution on the east side of the Sierra Nevada.

Phylogeographic structure within taxa

We did not find widespread admixture between the two taxa based on the results from InStruct when $k = 2$ (Fig. 6A). The individuals of admixed ancestry were restricted to sympatric sites in the narrow ~5- to 10-km-wide contact zone. The majority of admixed individuals were from the 6x and 6p populations, the northernmost extent of the contact zone.

Within both subspecies, we found differentiation between allopatric populations and those proximal to the zone of sympatry at $k = 2$ (Fig. 6B, C). Within *xantiana*, we detected additional phylogeographic structure ($k = 3$), where southern and low elevation populations differ from northern and high elevation populations (Figs 1 and 6D). Within *parviflora*, we found stronger population structure ($k = 4$), with many populations assigned primarily to a single genetic cluster rather than a mixture of two or more clusters as in most *xantiana* populations (Figs 1 and 6E). Differentiation among *parviflora* populations roughly followed an east—west gradient regardless of latitude. The most eastern populations (1p and 27p) share the same cluster despite having the greatest difference in latitude among populations in our sample. InStruct detected additional clustering within *xantiana* and *parviflora* (lowest DIC for $k = 6$ and 8, respectively, when k was allowed to vary from 1 to 10); however, there was no geographic pattern to these additional clusters.

Clinal variation

Plots of the hybrid indices based on the nDNA, microsatellite and combined (all loci) data sets showed a sharp change at the contact zone and minimal amounts of introgression, which was primarily restricted to some

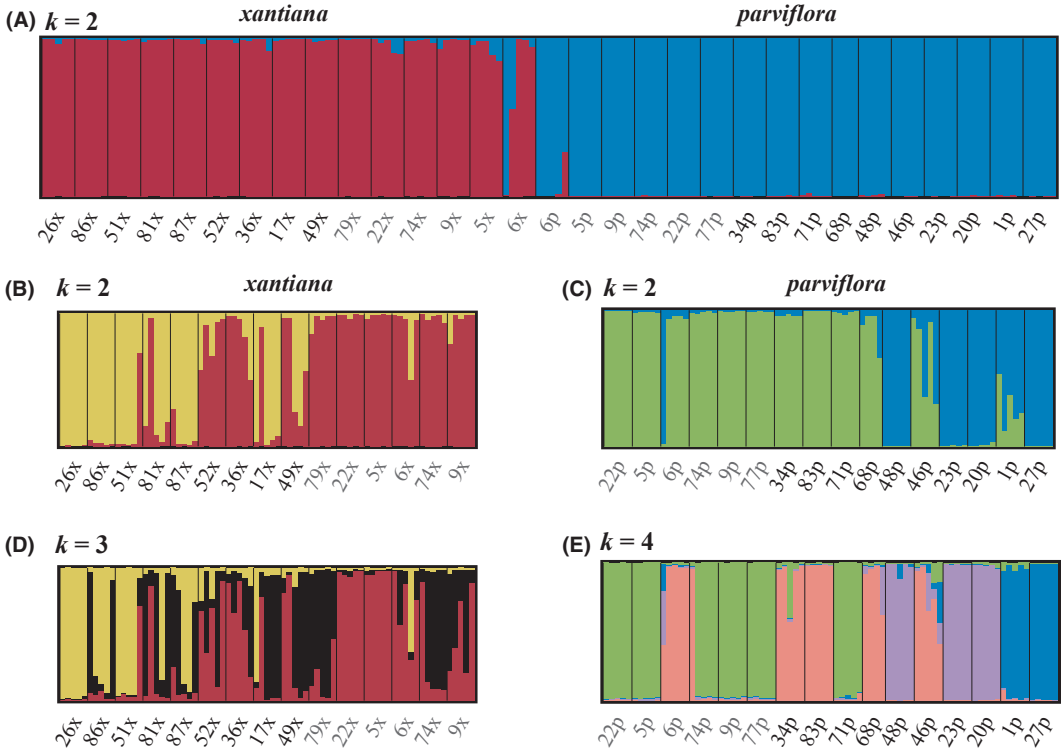

Fig. 6 DISTRUCT diagrams illustrating the composition of clusters across populations and the degree of admixture within individuals based on the results from INSTRUCT runs across three separate data sets: all samples (A), *xantiana* samples (B and D) and *parviflora* samples (C and E). Bars represent individuals, blocks represent populations and different colours indicate different clusters. Populations, shown below the figure, are ordered based on the distance to the nearest heterospecific population and sympatric populations are coloured grey

of the sympatric sites (Fig. 7). We did not detect a cline for cpDNA, most likely due to low nucleotide variation (Fig. 7, Table S2, Supporting information).

Across all four hybrid indices, we found strong statistical support for the trimodal model, under which separate distributions describe each parental taxon plus a third for admixed individuals (ΔAICc to the next best model >20; Table S5, Supporting information). On the basis of the simultaneous analysis of the three independent hybrid indices from the different marker types (i.e. cpDNA, microsatellites and nDNA loci), we found strong support for coincidence and concordance of clines; the model constraining the different indices to have the same cline width and centre was the best-fitting model (Table 2). Under that constrained model, the cline centre was shifted 0.10 km in the direction of *xantiana* from the contact zone and the width was 6.64 km.

Discussion

Although parapatric taxa and hybrid zones have been the subject of intensive research (Harrison 1993; Arnold 2006), tests of alternative phylogeographic hypotheses for their origins have been uncommon. The results of our phylogeographic and population genetic analyses were consistent in supporting the hypothesis that subspecies of *Clarkia xantiana* diverged in allopatry and have recently come into secondary contact. ABC analyses strongly supported a secondary contact model over competing phylogeographic models. Likewise, allopatric populations farthest from the contact zone of the derived taxon, *parviflora*, had the highest posterior probability of being the ancestral geographic state. Palaeodistribution modelling and patterns of population genetic structure suggested that refugial populations of *xantiana* and *parviflora* may have occurred in the extreme southern and eastern portion of their respective distributions and that historical migrations in response to Holocene climate change resulted in secondary contact.

Evidence for secondary contact

Approximate Bayesian computation analyses supported phylogeographic models including an allopatric phase during divergence (Table 1), which discounts the possibility that primary intergradation explains the contact zone between subspecies. The summary statistics illu-

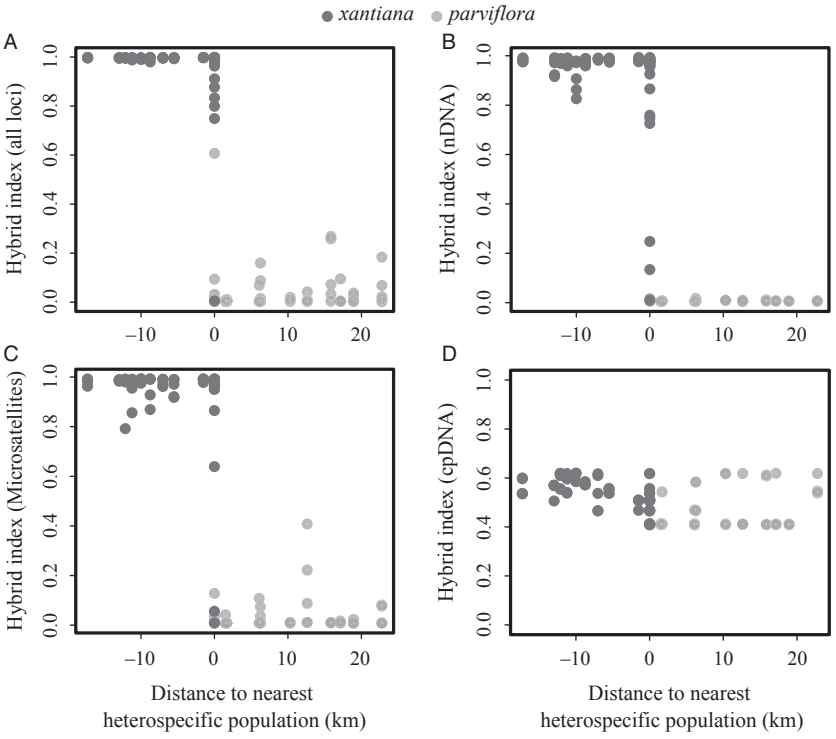

Fig. 7 Mean hybrid index per population for (A) all loci, (B) nDNA loci, (C) microsatellite loci and (D) cpDNA loci plotted against the distance to the nearest heterospecific population.

Table 2 Results of clinal analyses constraining certain parameters to be shared among the hybrid indices estimated separately for the cpDNA, microsatellites and nDNA loci. Negative centres indicate a shift from the centre in the direction of where *xantiana* populations are found. The best-fitting model is in bold; N_P = number of parameters in the model; ln = likelihood; ΔAICc = change in AIC relative to the best-fitting model. An n/a denotes a parameter that was free to vary across marker types and, therefore, a single estimate was not produced

Constraint	N_P	ln	Center (km)	Width (km)	ΔAICc
Center and width	**32**	**1139.81**	**−0.10**	**6.64**	**0.00**
All*	12	880.18	0.80	8.48	62.61
Center	34	1041.56	**−1.51**	n/a	203.29
Width	34	1008.54	n/a	4.44	269.34
None	35	883.91	n/a	n/a	522.11

*Under this model, parameters describing the shape of the distribution rather than just the centre and width are also constrained among the marker types.

minated somewhat different aspects of the divergence process. Model selection based on all summary statistics supported a secondary contact model; the same conclusion was reached when using only the proportion of shared polymorphisms. However, model selection based on F_{ST} and the rejection method most strongly supported an isolation model without any historic or recent introgression. Although it is not surprising that the process of divergence and migration leave different signatures on individual summary statistics, our results confirm that caution must be taken when performing ABC analyses as the choice of summary statistics and model selection criteria might influence the conclusions that one draws (e.g. Joyce & Marjoram 2008; Csilléry *et al*. 2010).

Ancestral geography analyses were consistent with the ABC results in suggesting that divergence occurred through an allopatric phase followed by secondary contact. The most ancestral populations of the derived taxon, *parviflora*, were not those found proximal to the progenitor taxon, *xantiana*, but rather were allopatric populations most distant from the zone of contact (Fig. 4). Furthermore, InStruct analyses detected strong differentiation between those allopatric populations and populations closer to the contact zone. In fact, we found virtually no evidence for individuals found in sympatric and allopatric populations occurring in the same cluster (k = 4, Fig. 6).

We also observed the coincidence and concordance of clines across different molecular markers, which suggests that there is a strong barrier to gene flow. In some circumstances, this pattern may also reflect secondary contact. On the basis of multilocus clines, we found that the cline width was ∼6 km and that the centre was slightly shifted in the direction of *xantiana* by about 0.1 km (Table 2, Fig. 7). Moreover, introgression was almost entirely limited to sites where subspecies

co-occur. These results are consistent with previous observations that introgression appears to be limited to the narrow geographic region where the two taxa co-occur and that gene flow is asymmetric primarily from *parviflora* to *xantiana*.

Divergence and range expansion

In the light of previous phylogenetic work showing that *xantiana* was the progenitor taxon from which *parviflora* diverged ~10 000–65 000 ybp (Pettengill & Moeller 2012), the distribution of suitable habitat at the LGM may provide some insight into the geographic region where divergence occurred and historical patterns of migration. Palaeo-distribution models suggest that suitable habitat at the LGM for *xantiana* was narrowly restricted to a region near the south-eastern edge of its current distribution (Fig. 5). If *parviflora* diverged from *xantiana* around the LGM, then the palaeo-distribution of its progenitor, *xantiana*, suggests that divergence occurred in this southern region. Phylogeographic analyses provide mixed support for this idea. The most southern populations of both taxa did not receive strong support as the ancestral geographic location. Within *xantiana*, allopatric populations that lie just outside the zone of contact (just south of Breckenridge Mountain: 49x, 17x, 81x and 87x) received the highest support. For *parviflora*, some of the most eastern allopatric populations (but not the most southern populations) received the highest support. Thus, it remains unclear as to whether *parviflora* diverged from *xantiana* while the latter was confined to its southern glacial refugium or at a different time when its distribution was more widespread.

For both taxa, our data suggest significant migration from zones of suitable habitat at the LGM to the areas where they are currently most abundant. For *xantiana*, the region of highest suitability at the LGM occurs at the far south-eastern edge of its distribution where populations are currently sparse. Populations are currently most abundant to the north in the Kern River Canyon. InStruct analyses detected differentiation along a roughly south—north axis that may reflect Holocene migration. In *parviflora*, the region of highest suitability at the LGM occurs at the far eastern portion of its current distribution. The most eastern *parviflora* populations also had the highest posterior probability of being the ancestral geographic state (Fig. 4), which suggests that migration occurred westward out of these populations. Our results from the InStruct analyses similarly suggest that differentiation within *parviflora* follows a predominantly east—west axis (Figs 1 and 6). Although our analyses do not provide conclusive evidence, one hypothesis is that *xantiana* migrated primarily north and *parviflora* migrated primarily west from LGM refugia to the current zone of contact between subspecies. These results parallel other historical studies suggesting that the postglacial migration of lineages that diverged in separate refugia resulted in hybrid zones (Swenson and Howard 2005; Rebernig *et al*. 2010; Hewitt 2011).

One noteworthy result regarding divergence is that sympatric populations of *parviflora* occurring in the zone of geographic overlap with *xantiana* (ca. 5 km wide × 50 km long) form a distinct genetic cluster that is virtually unrepresented outside the zone of overlap (Fig. 6E). Common garden studies have suggested that floral traits that determine pollinator attraction have diverged between these sympatric populations and allopatric populations to the east (D.A. Moeller, unpubl. data). Ongoing studies are testing alternative hypotheses for the causes of floral divergence in sympatry.

Holocene climate change

Predictions from our species distribution models for *xantiana* and *parviflora* overlap largely with their current distributions, suggesting that each taxon presently occupies most of the distribution of suitable habitat (see also Eckhart *et al.* 2011). Suitable habitat for both taxa was most strongly associated with summer drought, warmer winter temperatures and winter precipitation. Possible refugia during the LGM existed in areas that are currently warmer and more xeric than the habitat where *xantiana* and *parviflora* are now found. During the Xerothermic (~9000–5000 ybp), the climate in this region became substantially warmer and drier, and the ranges of many taxa shifted to higher elevations and deeper into the foothills of the Sierras (Axelrod 1981). For example, the region identified as suitable habitat at the LGM is currently the western Mojave Desert. As recently as 7800–9000 ybp, this area was predominantly pinyon—juniper—oak woodlands (Leskinen 1975; Van Devender & Spaulding 1979; Betancourt *et al.* 1990; Smith *et al.* 2000), which is similar to habitat where *xantiana* and *parviflora* are currently found.

Phylogeographic patterns across other taxa within the same geographic region also mirror our observations of range shifts for *xantiana* and *parviflora*. A particularly noteworthy comparison is with the grey pine, *Pinus sabiniana*, which is a common tree species in habitats where *xantiana* and *parviflora* are currently found. The extant distribution of *P. sabiniana* is in the foothills surrounding the Great Central Valley of California, but all fossil evidence of it occurs south of its current distribution (Axelrod 1938; Axelrod 1986). Population genetic work by Ledig (2001) suggests that grey pine migrated north into Central California and to higher elevations approximately during the Xerothermic period. Refugial areas for *P. sabiniana* currently have much higher mini-

mum winter temperatures than found in the current range, which may explain its near absence from Southern California at present (reviewed in Ledig 2001). This same environmental factor is one of the two primary factors determining the current distribution of *xantiana* and *parviflora* based on our species distribution models. Taken together, it is possible that shifts in the ranges of *xantiana* and *parviflora* paralleled those of a dominant member of their surrounding plant community, *Pinus sabiniana*. Similar patterns have been found for the dusky footed woodrat (*Neotoma fuscipes*), which also overlaps in distribution with *xantiana* and *parviflora*. Fossil records and a recently discovered relict population (Smith *et al*. 2000) indicate that during the Pleistocene *N. fuscipes* occurred east and farther into the Mojave Desert (Jefferson 1991) and has since receded to higher elevations in the foothills of the Sierras.

Conclusions

The coupling of phylogeographic methods based on the coalescent and palaeo-distribution modelling represents a relatively new approach to elucidating the geographic history of divergence and formation of hybrid zones (Carstens & Richards 2007; Chan *et al*. 2011). Our investigation of *xantiana* and *parviflora* highlight the advantages of such an approach. On the basis of coalescent methods (i.e. ABC and ancestral geography reconstructions), we found consistent support for allopatric divergence followed by range expansion leading to secondary contact. Palaeo-distribution modelling provided independent evidence for range expansion from areas of suitable habitat during the LGM to areas where populations are currently most abundant. It is not until we consider phylogeographic results in the light of historical climate changes since the LGM that we begin to understand the environmental pressures that may have facilitated shifts in geographic ranges and the formation of a hybrid zone. The combined power of these methods represents an exciting advance for studies of phylogeography and the geography of speciation.

Acknowledgements

We thank Erica Beckman, Raven Bier, Rebecca Carter, Justin Iverson, Christy Kelly and Jen Reese for help collecting and processing the molecular data. We also thank Monica Geber and Vince Eckhart for help identifying populations and sharing seeds. Computational resources and assistance were provided by the Minnesota Supercomputing Institute (MSI). Pavlos Pavlidis assisted with the running of the msABC. Funding for this project was provided by the National Science Foundation (DEB-1025004 to D.A. Moeller), the University of Minnesota and the University of Georgia.

References

Arnold ML (2006) *Evolution Through Genetic Exchange*. Oxford University Press, Oxford, UK.

Arnold ML, Ballerini ES, Brothers AN (2011) Hybrid fitness, adaptation and evolutionary diversification: lessons learned from Louisiana Irises. *Heredity*, **108**, 159–166.

Axelrod DI (1938) A Pliocene flora from the Mount Eden beds, Southern California. Contributions to palaeontology. Miocene and Pliocene floras of western North America. *Carnegie Instution*, **476**, 125–183.

Axelrod DI (1981) Holocene climatic changes in relation to vegetation disjunction and speciation. *American Naturalist*, **117**, 847–870.

Axelrod DI (1986) Cenozoic history of some western American pines. *Annals of the Missouri Botanical Garden*, **73**, 565–641.

Barton NH, Hewitt GM (1985) Analysis of hybrid zones. *Annual Review of Ecology and Systematics*, **16**, 113–148.

Beaumont MA, Zhang WY, Balding DJ (2002) Approximate Bayesian computation in population genetics. *Genetics*, **162**, 2025–2035.

Beaumont MA (2008) Joint determination of topology, divergence time, and immigration in population trees. In: *Simulation, Genetics and Human Prehistory* (eds Renfrew C, Matsumura S and Forster P), pp. 134–1541. McDonald Institute Monographs, UK.

Betancourt JL, Van Devender TR, Martin PS (1990) *Packrat Middens: The Last 40,000 Years of Biotic Change*. University of Arizona Press, Tucson.

Brennan AC, Bridle JR, Wang A, Hiscock SJ, Abbott RJ (2009) Adaptation and selection in the *Senecio* (Asteraceae) hybrid zone on Mount Etna, Sicily. *New Phytologist*, **183**, 702–717.

Bull CM (1991) Ecology of parapatric distributions. *Annual Review of Ecology and Systematics*, **22**, 19–36.

Caisse M, Antonovics J (1978) Evolution in closely adjacent plant populations. IX. Evolution of reproductive isolation in clinal populations. *Heredity*, **40**, 371–384.

Carstens BC, Richards CL (2007) Integrating coalescent and ecological niche modeling in comparative phylogeography. *Evolution*, **61**, 1439–1454.

Chan LM, Brown JL, Yoder AD (2011) Integrating statistical genetic and geospatial methods brings new power to phylogeography. *Molecular Phylogenetics and Evolution*, **59**, 523–537.

Collins WD, Bitz CM, Blackmon MI, *et al*. (2006) The Community Climate System Model: CCSM3. *Journal of Climate*, **19**, 2122–2143.

Coyne JA, Orr HA (2004) *Speciation*. Sinauer Associates, Inc., Sunderland, Massachusetts.

Csilléry K, Blum MGB, Gaggiotti O, François O (2010) Approximate Bayesian Computation (ABC) in practice. *Trends in Ecology and Evolution*, **25**, 410–418.

Csilléry K, François O, Blum MGB (2012) abc: an R package for Approximate Bayesian Computation (ABC). *Methods in Ecology and Evolution*, **3**, 475–479.

Dobzhansky T (1940) Speciation as a stage in the evolutionary process. *The American Naturalist*, **74**, 312–321.

Durrett R, Buttel L, Harrison RG (2000) Spatial models for hybrid zones. *Heredity*, **84**, 9–19.

Eckhart VM, Geber MA (1999) Character variation and geographic distribution of *Clarkia xantiana* (Onagraceae):

flowers and phenology distinguish two subspecies. *Madroño*, **46**, 117–125.

Eckhart VM, Geber MA, Morris WF, *et al*.. (2011) The geography of demography: long-term demographic studies and species distribution models reveal a species border limited by adaptation. *American Naturalist*, **178**, S26–S43.

Eckhart VM, Singh I, Louthan AM, *et al*. (2010) Plant-soil water relations and species border of *Clarkia xantiana* ssp. *xantiana* (Onagraceae). *International Journal of Plant Sciences*, **171**, 749–760.

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, **32**, 1792–1797.

Endler JA (1977) *Geographic Variation, Speciation and Clines*. Princeton University Press, Princeton.

Fagundes NJR, Ray N, Beaumont M, Neuenschwander S, *et al*. (2007) Statistical evaluation of alternative models of human evolution. *Proceedings of the National Academy of Sciences, USA*, **104**, 17614–17619.

Fisher RA (1930) *The Genetical Theory of Natural Selection*. Oxford University Press, London.

Gao H, Williamson S, Bustamante CD (2007) A Markov Chain Monte Carlo approach for joint inference of population structure and inbreeding rates from multilocus genotype data. *Genetics*, **176**, 1635–1651.

Gavrilets S, Li H (2000) Patterns of parapatric speciation. *Evolution*, **54**, 1126–1134.

Gay L, Crochet P, Bell D, Lenormand T (2008) Comparing genetic and phenotypic clines in hybrid zones: a window on tension zone models. *Evolution*, **62**, 2789–2806.

Harrison RG (1993) *Hybrid Zones and the Evolutionary Process*. Oxford University Press, New York, New York.

Hewitt GM (2001) Speciation, hybrid zones and phylogeography - or seeing genes in space and time. *Molecular Ecology*, **10**, 537–549.

Hewitt GM (2011) Quaternary phylogeography: the roots of hybrid zones. *Genetica*, **139**, 617–638.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965–1978.

Hudson RR (2002) Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics*, **18**, 337–338.

Hudson RR, Slatkin M, Maddison WP (1992) Estimation of levels of gene flow from DNA sequence data. *Genetics*, **132**, 583–589.

Jefferson GT (1991) A catalogue of late Quaternary vertebrates from California: part two, mammals. *Natural History Museum of Los Angeles County Technical Report*, **7**, 1–129.

Jiggins CD, Mallet J (2000) Bimodal hybrid zones and speciation. *Trends in Ecology and Evolution*, **15**, 250–255.

Joyce P, Marjoram P (2008) Approximately Sufficient Statistics and Bayesian Computation. *Statistical Applications in Genetics and Molecular Biology*, **7**, 26

Kawakami T, Butlin RK, Adams M, Paull DJ, Cooper SJB (2008) Genetic analysis of a chromosomal hybrid zone in the Australian morabine grasshoppers (*Vandiemenella viatica* species group). *Evolution*, **63**, 139–152.

Koch M, Haubold B, Mitchell–Olds T (2001) Molecular systematics of the Brassicaceae: evidence from coding plastidic *matK* and nuclear *Chs* sequences. *American Journal of Botany*, **88**, 534–544.

Kozak KH, Graham CH, Wiens JJ (2008) Integrating GIS-based environmental data into evolutionary biology. *Trends in Ecology and Evolution*, **23**, 141–148.

Kress JW, Erickson DL (2007) A two-locus global DNA barcode for land plants: the coding *rbcL* gene complements the non-coding *trnH-psbA* spacer region. *PLoS ONE*, **2**, e508.

Ledig FT (2001) Genic diversity, genetic structure, and biogeography of *Pinus sabiniana* Dougl. *Diversity and Distributions*, **5**, 77–90.

Lemey P, Rambaut A, Drummond AJ, Sucharc MA (2009) Bayesian phylogeography finds its roots. *Plos Computational Biology*, **5**: e1000520. doi: 10.1371/journal.pcbi.1000520.

Leskinen PH (1975) Occurrence of oaks in late Pleistocene vegetation in the Mojave Desert of Nevada. *Madroño*, **23**, 234–235.

Maroja LS, Andres JA, Harrison RG (2009) Genealogical discordance and patterns of introgression and selection across a cricket hybrid zone. *Evolution*, **63**, 2999–3015.

Mayr E (1942) *Systematics and the Origin of Species*. Columbia University Press, New York, New York.

Moeller DA, Geber MA, Tiffin P (2011) Population genetics and the evolution of geographic range limits in an annual plant. *American Naturalist*, **178**, S44–S61.

Morgan-Richards M, Trewick SA, Wallis GP (2000) Characterization of a hybrid zone between two chromosomal races of the weta *Hemideina thoracica* following a geologically recent volcanic eruption. *Heredity*, **85**, 586–592.

Pavlidis P, Laurent S, Stephan W (2010) msABC: a modification of Hudson's ms to facilitate multi-locus ABC analysis. *Molecular Ecology Resources*, **10**, 723–727.

Pettengill JB, Moeller DA (2012) Tempo and mode of mating system evolution between incipient *Clarkia* species. *Evolution*, **66**, 1210–1225.

Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, **190**, 231–259.

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

Rebernig CA, Schneeweiss GM, Bardy KE, *et al*. (2010) Multiple Pleistocene refugia and Holocene range expansion of an abundant southwestern American desert plant species (*Melampodium leucanthum*, Asteraceae). *Molecular Ecology*, **19**, 3421–3443.

Ross-Ibarra J, Tenaillon M, Gaut BS (2009) Historical divergence and gene flow in the genus *Zea*. *Genetics*, **181**, 1399–1413.

Rosenberg NA (2004) Distruct: a program for the graphical display of population structure. *Molecular Ecology Notes*, **4**, 137–138.

Runions CJ, Geber MA (2000) Evolution of the self-pollinating flower in *Clarkia xantiana* (Onagraceae). I. Size and development of floral organs. *American Journal of Botany*, **67**, 1439–1450.

Slatkin M (1981) Pleiotropy and parapatric speciation. *Evolution*, **36**, 263–270.

Smith FA, Matocq MD, Melendez KF, Ditto AM, Kelly PA (2000) How isolated are Pleistocene refugia? Results from a study on a relict woodrat population from the Mojave Desert, California. *Journal of Biogeography*, **27**, 483–500.

Stephens M, Donnelly P (2003) A comparison of Bayesian methods for haplotype reconstruction from population genotype data. *American Journal of Human Genetics*, **73**, 1162–1169.

Stephens M, Smith N, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics*, **68**, 978–989.

Swenson NG (2008) The past and future of influence of geographic information systems on hybrid zone, phylogeographic and speciation research. *Journal of Evolutionary Biology*, **21**, 421–434.

Swenson NG, Howard DJ (2005) Clustering of contact zones, hybrid zones, and phylogeographic breaks in North America. *American Naturalist*, **166**, 581–591.

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

Van Devender TR, Spaulding WG (1979) Development of vegetation and climate in the southwestern United States. *Science*, **204**, 701–710.

Watterson GA (1975) On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology*, **7**, 256–275.

In his research, J.B.P. employs an integrative approach where multiple methods of inference are used to better understand the evolutionary history and contemporaneous interactions that can explain extant patterns of genetic diversity. D.M. studies the ecology and genetics of speciation, the evolution of mating systems, and the evolution of geographic range limits, especially in the genus Clarkia.

Data accessibility

DNA sequences: GenBank Accessions nos JQ009828–JQ010836. Alignments of phased sequences and microsatellite genotypes have been deposited in DRYAD entry doi: 10.5061/dryad.g2p5g. Location information for all samples can be found in Table S1 (Supporting information).

Supporting information

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

Table S1. Population codes (Site #), number of individuals sampled (*N*), elevation, geographic coordinates, geographic distance to the nearest heterospecific population, and posterior probability (*P*) of being the ancestral location from the BEAST analyses for samples included in this study.

Table S2. Summary statistics for 11 sequenced loci: the number of sequences obtained (*N*), length of fragment (*L*), segregating sites (*S*), polymorphisms unique to *C. xantiana* ssp. *xantiana* (P_x), polymorphisms unique to *C. xantiana* ssp. *parviflora* (P_p), shared polymorphisms (P_s), average posterior probability of phased haplotypes (*Phase*) and nucleotide substitution model (*NST*).

Table S3. Summary statistics for four microsatellite loci: percent successful amplification, number of alleles, allele size ranges, observed heterozygosity (H_o), and inbreeding coefficients (F_{is}).

Table S4. Uniform priors scaled by locus length for Approximate Bayesian analyses. θ_{xan} and θ_{par} describe Watterson's θ for *Clarkia xantiana* ssp. *xantiana* and ssp. *parviflora*, respectively. τ_M is the time that migration starts or stops, τ_D is the divergence time, and the asymmetric migration rates $M_{xan \rightarrow par}$ and $M_{par \rightarrow xan}$ between taxa.

Table S5. Results of cline model selection from CFIT7 for hybrid indices from four different molecular datasets. Negative centers indicate a shift from the location of contact between the two taxa in the direction of where *xantiana* populations are found. Best fitting models are in bold.

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