MOLECULAR ECOLOGY

Molecular Ecology (2012) **21**, 4242–4256 doi: 10.1111/j.1365-294X.2012.05696.x

# A range-wide genetic bottleneck overwhelms contemporary landscape factors and local abundance in shaping genetic patterns of an alpine butterfly (Lepidoptera: Pieridae: *Colias behrii*)

SEAN D. SCHOVILLE, ATHENA W. LAM and GEORGE K. RODERICK

*Department of Environmental Science, Policy and Management, University of California, Berkeley, 130 Mulford Hall #3114, Berkeley, CA 94720-3114, USA*

## Abstract

**Spatial and environmental heterogeneity are major factors in structuring species distributions in alpine landscapes. These landscapes have also been affected by glacial advances and retreats, causing alpine taxa to undergo range shifts and demographic changes. These nonequilibrium population dynamics have the potential to obscure the effects of environmental factors on the distribution of genetic variation. Here, we investigate how demographic change and environmental factors influence genetic variation in the alpine butterfly *Colias behrii*. Data from 14 microsatellite loci provide evidence of bottlenecks in all population samples. We test several alternative models of demography using approximate Bayesian computation (ABC), with the results favouring a model in which a recent bottleneck precedes rapid population growth. Applying independent calibrations to microsatellite loci and a nuclear gene, we estimate that this bottleneck affected both northern and southern populations 531–281 years ago, coinciding with a period of global cooling. Using regression approaches, we attempt to separate the effects of population structure, geographical distance and landscape on patterns of population genetic differentiation. Only 40% of the variation in $F_{ST}$ is explained by these models, with geographical distance and least-cost distance among meadow patches selected as the best predictors. Various measures of genetic diversity within populations are also decoupled from estimates of local abundance and habitat patch characteristics. Our results demonstrate that demographic change can have a disproportionate influence on genetic diversity in alpine species, contrasting with other studies that suggest landscape features control contemporary demographic processes in high-elevation environments.**

*Keywords*: approximate Bayesian computation, climate change, demographic change, landscape genetics, Matthes glaciations, Sierra Nevada

*Received 8 February 2012; revision accepted 30 May 2012*

## Introduction

Environmental heterogeneity plays an important role in regulating population connectivity in natural populations (Turner 1989). The characteristics of suitable habitat (e.g. size, shape and quality of patches) and intervening habitat matrix affect demographic trends by regulating immigration and emigration rates, recruitment and population growth rates (Moilanen & Hanski 1998). By influencing the exchange of breeding migrants, environmental heterogeneity also regulates genetic variability within populations, fitness and the spread of advantageous alleles. Although population connectivity and genetic connectivity are not directly related (Lowe & Allendorf 2010), in situations where

Correspondence: Sean D. Schoville, Université Joseph Fourier Grenoble, Equipe Biologie Computationnelle et Mathématique, Laboratoire TIMC-IMAG, UMR CNRS 5525, 38402 Grenoble, France. Fax: +33 4 56 52 0055; E-mail: sean.schoville@imag.fr

environmental heterogeneity reduces population connectivity, there is typically an increase in genetic differentiation across the landscape (Manel *et al.* 2003). An important objective of landscape genetics research is to understand how environmental factors influence genetic connectivity and to use this information to predict how populations will respond to global change (Segelbacher *et al.* 2010).

Spatial heterogeneity is particularly pronounced in montane environments, where steep environmental gradients and topographical complexity frequently result in patchy species distributions (Mani 1968; Korner 1999). These gradients can have a strong effect on mediating the direction and intensity of gene flow among alpine populations (Stanton & Galen 1997). In a study of 12 alpine plant species in the European Alps, Thiel-Egenter *et al.* (2011) found that the distribution of genetic variation was similar among species and genetic break zones occurred in regions with large changes in elevation. Other studies have also shown that high levels of genetic differentiation in alpine species are related to topographical factors (Murphy *et al.* 2010; Savage *et al.* 2010), as well as climatic gradients (Hirao & Kudo 2004) and the surrounding habitat matrix (Keyghobadi *et al.* 1999).

However, the importance of landscape factors in structuring genetic variation in alpine landscapes should be contrasted to the effect of nonequilibrium population dynamics. It is well known that alpine species have been subject to severe and rapid demographic transitions due to historical climate change (e.g. Schoville & Roderick 2009), as well as geographical shifts during glacial maxima (Elias 1991) and following the retreat of glaciers (Ehrich *et al.* 2007). Demographic change is known to affect the levels of genetic diversity (Tajima 1989), effective population size (Waples 2005) and interpopulation gene flow (Marko & Hart 2011), and it often influences spatial genetic variability to a greater extent than environmental factors (Zellmer & Knowles 2009). Moreover, efforts should be made to separate demographic effects during data analysis, for example, by including factors that account for population structure and population expansion (Gaggiotti *et al.* 2009; Dyer *et al.* 2010).

In this study, we set out to test whether demographic change and landscape factors have had an important role in structuring patterns of genetic variability in the alpine butterfly *Colias behrii* Edwards (1866). Previous research provided evidence of population structure, but suggested that *C. behrii* colonized alpine habitats following the last glacial maximum and lacked substantial genetic variation across its range (Schoville *et al.* 2011). Here, we expand our sampling by examining variation at 14 microsatellite loci and five additional populations across the Sierra Nevada Mountains. We first test whether population structure or demographic change is evident in this species, and then use regression analyses to assess whether environmental factors influence genetic differentiation and genetic diversity.

## Materials and methods

### *Study organism and genetic sampling*

*Colias behrii* is a specialist of subalpine and alpine meadow habitats (>2700 m elevation) and endemic to the Sierra Nevada. Adults have a single annual flight, usually in July–August, when they aggregate in large colonies. The dispersal behaviour of *C. behrii* is likely to be similar to other alpine *Colias*, which do not disperse more than a few kilometres during their adult lifespan and avoid movement through forested habitat (Watt *et al.* 1977; Roland 1982). Samples of 10–31 individual *C. behrii* were collected from 18 sites spanning the species geographical range during 2003–2008 (total $N = 356$, Fig. 1) and subsequently deposited in the

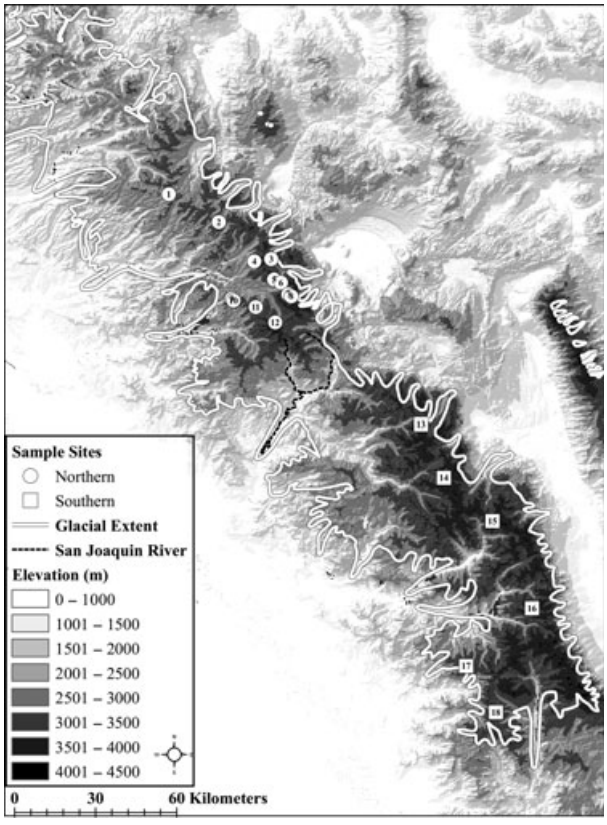


**Fig. 1** Map of the genetic sampling localities and transect sites of *Colias behrii* in the California Sierra Nevada. Northern (circle) and southern (square) population clusters are shown, and site numbers refer to locations listed in Table 1. The extent of glacial ice during the last glacial maximum is taken from Gillespie & Zehfuss (2004).

Essig Museum of Entomology at UC Berkeley. Genotypic data based on sequences (1066 base pairs) of the nuclear gene elongation factor 1 alpha (EF1α) are available for 235 individuals examined in a previous study (Schoville *et al.* 2011). Here, we add data for 14 microsatellite loci from these individuals and an additional five populations (a total of 356 individuals). Microsatellite genotypes were obtained following a previously published protocol (Molecular Ecology Resources Primer Development Consortium *et al.* 2010).

### *Analysis of population genetic diversity and population structure*

Each microsatellite locus was tested for deviation from Hardy–Weinberg equilibrium (HWE) using Fisher's exact tests. Due to the large number of tests ($n = 252$), the nominal level of statistical significance ($\alpha = 0.05$) was adjusted by Dunn–Šidák correction ($1-(1-\alpha)^{1/n}$) to 0.0002035. Pairwise independence of loci within each population was used to assess linkage equilibrium. Statistics measuring the mean number of alleles ($MA_R$), mean observed ($MH_O$) and mean expected heterozygosity ($MH_E$) and fixation index ($F = 1-(H_O/H_E)$) were averaged across loci for each population. These statistics were calculated within GENALEX v6.41 (Peakall & Smouse 2006) and GENEPOP v4.1 (Rousset 2008). We used the program GESTE (Foll & Gaggiotti 2006) to estimate $F_{ST}$ based on the drift of a subpopulation from a common ancestral population. This measure of $F_{ST}$ might be more appropriate in populations with population size fluctuations than an estimator based on migration–drift equilibrium. Genetic diversity at the EF1α locus was reported in the study by Schoville *et al.* (2011), and here, we provide the same measurements of theta based on Watterson's estimator ($\theta_S$) and pairwise diversity ($\theta_\pi$).

To identify distinct populations, we first employed a principal coordinate analysis (PCoA) summarizing microsatellite variation among individuals. The eigenvectors of the PCoA were calculated from a covariance matrix with data standardization in GENALEX and the first two principal coordinates were plotted. Second, Bayesian clustering was used to assign individuals to populations based on inferred population allele frequencies. Analyses of clustering methods have shown that the presence of population admixture can lead to an overestimation of the number of clusters, whereas clustering models that account for admixture remain robust even if admixture is lacking (François & Durand 2010). We therefore analysed the combined microsatellite and EF1α data set allowing for admixture. STRUCTURE v2.3.3 was run with sampling locations as prior information and the F model allowing for correlated allele frequencies across populations (Falush *et al.* 2003). Population cluster values ($K$) varied from 1 to 8 with ten independent runs at each $K$ value, the burn-in period was set to 100 000, and MCMC sampling was run for 1 000 000 repetitions. We determined the optimal $K$ value by examining the difference in log-likelihood among values of $K$ and based on the change in log-likelihood values ($\Delta K$) corrected by the variance among replicate runs (Evanno *et al.* 2005). As an alternative, we analysed our data in the program TESS v2.3.1 (Chen *et al.* 2007) with geographical location as prior information and the admixture model (Durand *et al.* 2009). Population cluster ($K$) values were set from 2 to 8 with 10 independent runs at each $K$ value, the burn-in period was set to 30 000 sweeps, and MCMC sampling was run for 100 000 sweeps. The optimal $K$ value was determined by examining the difference in the deviance information criterion (DIC) value among values of $K$, and the $\Delta K$ of these values was also calculated. For both STRUCTURE and TESS analyses, the 10 runs at the optimal value of $K$ were summarized using the program CLUMPP (Jakobsson & Rosenberg 2007) with the Greedy algorithm, and the program DISTRUCT (Rosenberg 2004) was used to graphically display the output. Finally, we used a Mantel test to assess genetic isolation by distance in the microsatellite data, with significance values assessed by 999 matrix permutations. Genetic distance was measured between pairs of individuals according to the method of Smouse & Peakall (1999), and geographical distance was measured as Euclidean and log-transformed Euclidean distance.

### *Analysis of population size change*

Due to low levels of genetic variability, we tested for evidence of genetic bottlenecks in each population sample. When the effective population size is sharply reduced during a population bottleneck, there is a transient phase where the population exhibits an excess of heterozygosity across loci. The sign-test, standardized difference test and Wilcoxon sign-rank deficiency test were developed to test for departures from equilibrium based on multilocus patterns of heterozygosity (Cornuet & Luikart 1996). These statistics were calculated in the program BOTTLENECK (Piry *et al.* 1999) assuming the two-phase model with variance set to 30 and probability set to 70%, and significance was assessed over 10 000 replicates. An alternative test based on the $M$-ratio statistic (Garza & Williamson 2001) examines the ratio of the number of alleles and the allele range, with the expectation that the number of alleles declines faster than the allele range during a bottleneck. The $M$-ratio was calculated in ARLEQUIN v3.5.1.2 (Excoffier & Lischer 2010) and tested for significance by comparison with data

simulated in the program SIMCOAL v2.1.2 (Laval & Excoffier 2004). Data were simulated under an island model of constant size and migration ($4Nm = 1$) assuming a generalized stepwise mutation model (GSM), with the number of repeats drawn from a geometric distribution ($P = 0.5$). Samples were drawn according to the observed sample size in each of the 18 populations. Critical values ($\alpha = 0.05$) of the *M*-ratio, below which an *M*-ratio value would be likely to indicate a bottleneck, were determined for different levels of theta ($\theta_{\text{per deme}} = 0.004, 0.01, 0.02, 0.04, 0.1, 0.2$) based on 1000 simulated data sets at each level of theta. For the EF1 locus, statistics reporting genetic change were reported in the study by Schoville *et al.* (2011). Here, we again provide the results for Tajima's *D* statistic, Fu's $F_S$ statistic and the mismatch distribution SSD statistic and Raggedness index, with significance calculated using 1000 bootstrap replicates. Significant negative values of Tajima's *D* or Fu's $F_S$ indicate population growth or purifying selection, while significant positive values indicate population bottlenecks or balancing selection. Fu's $F_S$ is considered significant at *P*-values < 0.02. Mismatch distributions examine the frequency distribution of pairwise differences assuming population expansion. Significant SSD or Raggedness index values indicate that an expansion model is rejected by the data.

Two additional approaches were used to evaluate models of past population size change and to estimate associated demographic parameters, including the current population size ($N$), ancestral population size ($N_A$) and time of population size change ($\tau$). First, we employed the full likelihood method of Beaumont (1999), which estimates a linear or exponential change in population size in a single population, and assume the microsatellite loci evolve according to a single-step mutation model. We employed the hierarchical Bayesian model in MSVAR 1.3, which allows for interlocus variability (Storz & Beaumont 2002). Previous work has shown that MSVAR is sensitive to underlying population structure (Chikhi *et al.* 2010). We therefore examined three data sets, including (i) all samples combined, (ii) pooled northern populations and (iii) pooled southern populations. We chose the exponential growth model and designated broad hyperpriors for the model parameters as follows (on a log-scale): means $\alpha_{N0} = \alpha_{N1} = 4$, $\alpha_{\mu} = -4.75$, $\alpha_{xa} = 3$, $\sigma_{N0} = \sigma_{N1} = 3$, $\sigma_{\mu} = 0.75$, $\sigma_{xa} = 2$, $\beta_{N0} = \beta_{N1} = \beta_{\mu} = \beta_{xa} = 0$, and $\tau_{N0} = \tau_{N1} = \tau_{\mu} = \tau_{xa} = 0.5$. The prior range of the microsatellite mutation rate was based on published studies of insects (Lehmann *et al.* 1998; Schug *et al.* 1998). For each data set, three separate MCMC chains were run for $2.5 \times 10^9$ steps with samples recorded every 100 000 steps. After removing a 10% burn-in, the CODA package in the software R v2.13.2 (R Development Core Team 2011) was used to assess convergence by calculating the Gelman–Rubin multivariate scale reduction factor (PRSF) and the effective sample size (ESS) of each parameter across the independent chains. Estimates of the mean and 95% highest posterior density (HPD) intervals were obtained from the marginal posterior distribution of each parameter using the locfit package in R.

As a second approach, we compared the fit of our observed microsatellite and EF1α data to several alternative evolutionary models using approximate Bayesian computation (ABC). This method estimates the posterior distributions of model parameters without calculating the explicit likelihood of these parameters. Instead, data are simulated under each model by randomly drawing values from a prior distribution of parameter values, the observed data are compared to the synthetic data based on the similarity to summary statistic values, and both parameters and models can be estimated based on the fit of the data. We proposed three models of population history (Fig. 2) including population structure and bidirectional gene flow: (i) constant population size, (ii) a historic bottleneck preceding population divergence and (iii) a recent bottleneck following population divergence. We used ABCtoolbox (Wegmann *et al.* 2010) and the coalescent simulator SIMCOAL v2.1.2 (Laval & Excoffier 2004) to generate 500 000 synthetic data sets under each model. Broad uniform priors were set for most model parameters, including population sizes, event times and mutation rates, while truncated normal priors were set for each migration rate (prior values shown in Table S1, Supporting information). Based on a fossil-calibrated EF1α mutation rate for *Colias* (Braby *et al.* 2006), with a mean of 0.001277 per site per million years (confidence intervals of ±0.000024), we specified a broader uniform prior of $7e^{-7}$–$2e^{-6}$. Summary statistics of the data were calculated in ARLEQUIN for northern, southern and both populations combined. The mean and standard deviation of the following statistics were calculated for the microsatellite data: the number of alleles ($K$), the expected heterozygosity ($H$), the Garza–Williamson index, modified Garza–Williamson index, range of alleles ($R$), *F*-statistics ($F_{IS}$, $F_{ST}$ and $F_{IT}$) and the mean number of differences between populations (PI). The mean and standard deviation of the following statistics were calculated for EF1α: the number of segregating sites ($S$), the number of private segregating sites (prS), Tajima's *D*, Fu's $F_S$, the number of pairwise differences ($\pi$), $F_{ST}$, and the mean number of differences between populations (PI). A variety of statistical techniques have been proposed for model fitting in ABC (Csilléry *et al.* 2010). We implemented rejection sampling coupled with feed-forward neural networks (nonlinear regression) because this method reduces the large number of summary

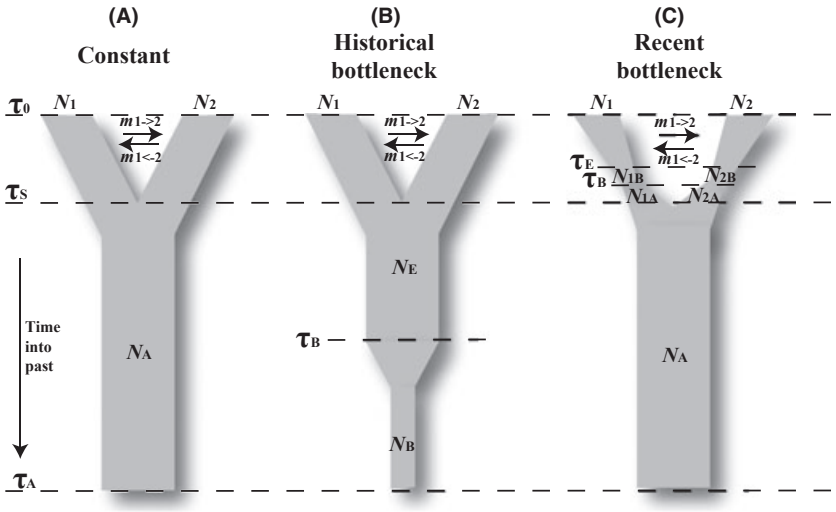


**Fig. 2** Alternative models of population history tested by approximate Bayesian computation (ABC).

statistics into a smaller number of dimensions (Blum & François 2010). We explored several tolerance levels for the rejection step using cross-validation (0.001, 0.005, 0.01, 0.05 and 0.1) and chose to accept 5% of the simulations as more stringent tolerance levels seemed to have little effect on the results. Competing models were compared based on their posterior probability and Bayes Factor scores, and we used cross-validation to ensure that the models could be correctly classified based on our model selection procedure. Finally, we estimated parameters in the best-fitting model and used cross-validation to assess the accuracy of our parameter estimates. These analyses were conducted in R using the abc package (Csilléry *et al.* 2012).

### *Environmental predictors of genetic differentiation*

We used regression-based analyses to examine whether population genetic differentiation ($F_{ST}$) could be explained by environmental predictors. We included population structure in a simple matrix assigning cluster membership and geographical distance calculated as the Euclidean distance between digital longitude and latitude coordinates of each population sample. We chose a small set of predictor variables known to influence dispersal in adult butterflies (Hayes 1981; Keyghobadi *et al.* 1999; Roland & Matter 2007), including elevation (using a digital elevation model from the NASA Shuttle Radar Topography Mission), meadow patch connectivity (CalVeg and National Park Service vegetation databases) and maximum July temperature (PRISM Group 2007). We calculated the climatic envelope for each variable from ~150 occupancy and museum records and then assigned incrementally increasing cost values outside the minimum and maximum of each envelope. Pairwise least-cost distance was measured between populations using the Landscape Genetics ArcToolbox (Etherington 2010), and these values were converted to landscape connectivity measurements using a negative exponential dispersal kernel (Moilanen & Nieminen 2002), as suggested by Foll & Gaggiotti (2006). We first applied the random forest regression method, which iteratively selects the most important variables explaining the variance in $F_{ST}$. This nonparametric method has the advantage of making no a priori assumptions about the relationship between the response and predictor variable, but regression lines are not fit. Second, we applied principal component regression, where environmental variables are first transformed into orthogonal components before being regressed on $F_{ST}$. This method offers the advantage of transforming a set of correlated predictor variables into independent components, although results can be biased if the data exhibit strong spatial autocorrelation (Novembre & Stephens 2008). These methods were implemented in the regressionForest and pls packages of R.

### *Estimation of abundance*

Line-transect surveys were used to estimate local population density as a proxy for abundance data at sample sites, where density is estimated based on the encounter rate of individuals per metre of transect length. Visual range was restricted to 10 m on either side of the transect line, and estimates of perpendicular distance and direction for each individual butterfly were recorded. The frequency of observations was grouped in 1-m bins measured from the transect line. For most sites, density estimates are based on multiple surveys conducted during 2006–2008 (Table S2, Supporting information). These separate surveys were used as replicates to adjust site estimates by survey effort (distance covered). Den-

sity estimates ($D_{LT}$) were adjusted using a detection function estimated from the global data set, which was fit using a half-normal detection curve with a cosine adjustment. This model was selected over a uniform detection function based on the fit of the model. Statistical analyses were implemented in DISTANCE v6.0 (Thomas *et al*. 2009).

### *Environmental predictors of genetic diversity*

To examine whether genetic diversity at sample sites could be explained by local habitat features or abundance, we applied a linear modelling approach to several different measures of intrapopulation genetic diversity, including the mean number of alleles ($MA_R$), mean expected heterozygosity ($MH_E$), Shannon's index and the mean squared allele size difference (MSRD). Habitat features include the area of the sampled meadow patch (in squared metres, $m^2$), the average distance to neighbouring patches (m) and total neighbourhood size ($m^2$). The latter two measurements were calculated in a $5 \times 5$ km grid around the sample site, to reflect the limited dispersal distance measured in *Colias* butterflies. Univariate linear regression was used to assess each predictor variable, and the *F*-test was used to test for significance.

## Results

### *Genetic variability and population structure*

Among 14 microsatellite loci, significant deviation from HWE was observed in three loci (A116, A106 and A6) after Bonferroni correction for multiple testing ($\alpha = 0.0002035$, Table S3, Supporting information). However, significant tests were found in a small number of populations, so we decided to retain them in downstream analyses. There was no evidence of linkage disequilibrium for any combination of loci. No clear geographical pattern was observed in $MA_R$ (ranging from 2.6 to 4.1), $MH_E$ (ranging from 0.280 to 0.401) or $MH_O$ (ranging from 0.234 to 0.334) across the Sierra Nevada (Table 1). Most populations exhibit limited inbreeding with $F > 0.1$, while moderate inbreeding is evident at Cathedral Lakes ($F = 0.355$) and Snow Lake YNP ($F = 0.282$). Genetic differentiation ($F_{ST}$) is low, with populations in the southern Sierra Nevada showing the highest differentiation (ranging from 0.148 to 0.155). Genetic diversity at the EF1α locus ($\theta_S$ and $\theta_\pi$) is very low in all populations.

Principal coordinate analysis indicates north–south differentiation along the first coordinate axis (Fig. 3A), which accounts for ~26% of the microsatellite variability. However, the groups overlap in this first axis and no further structure is evident in the second principal coordinate (12.5% of the variation). Bayesian clustering separates the samples into a northern and southern population cluster (Fig. 3B). In the STRUCTURE analysis, there is a large improvement in the log-likelihood from $K = 1$ to $K = 2$ clusters ($\Delta L = 583.26$) and the rate of change in log-likelihood values diminishes for additional clusters ($\Delta K_2 = 157$, $\Delta K_3 = 0.29$, Fig. S1, Supporting information). At $K = 2$ (Fig. S2, Supporting information), the northern and southern population clusters show evidence of limited admixture in the geographically intermediate Snow Lakes (Sierra NF) and Humphreys' Basin populations. The analysis with TESS (Fig. 3B) has the smallest DIC value at $K = 3$ clusters ($\Delta DIC = -217$), but no individual was completely assigned to a unique third cluster. The results are qualitatively similar at $K = 2$, with limited admixture in geographically intermediate populations. Tests for isolation by distance in the microsatellite data were significant for Euclidean ($r_{xy} = 0.417$, $P = 0.001$) and log-transformed Euclidean distance ($r_{xy} = 0.365$, $P = 0.001$).

### *Genetic evidence of population size change*

Summary statistical tests of microsatellite variation (Table 2) provide evidence for demographic bottlenecks. The sign, standardized difference and Wilcoxon tests indicate that six populations have an excess of heterozygosity indicative of a recent bottleneck, although only two are significant in all three tests. In contrast, *M*-ratio tests suggest that all populations have undergone bottlenecks, with observed values well below critical values. Four populations have significant SSD and Raggedness tests at EF1α (Table 1), rejecting models of population growth.

The MSVAR analysis of the pooled samples indicates weak population expansion (Bayes Factor 6.9 for a model of growth vs. bottleneck). MCMC diagnostics suggest the chains were mixed (PRSF = 1.22), and parameter estimates were drawn from a large sample size (ESS > 250). Estimates of the current population size (mode = 21 302, 95% HPD: 219–878 687), ancestral population size (mode = 12 399, 95% HPD: 51–373 027) and time of population size change (mode = 2150, 95% HPD: 7–845 217) suggest a ~2-fold increase 2000 years ago. A bivariate density plot shows a multimodal distribution in the current and ancestral population effective population size, with a smaller secondary peak favouring a bottleneck scenario (Fig. S3A, Supporting information). Analysis of the northern Sierra Nevada samples results in a signal of population expansion (Bayes Factor 14.6 for a model of growth vs. bottleneck).

**Table 1** Genetic variability of 14 microsatellite loci and the EF1α locus in populations of *Colias behrii*

| | Microsatellite loci | | | | | | | EF1α locus | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *N* | $MA_R$ ±SE | $MH_O$ ±SE | $MH_E$ ±SE | Mean *F* ±SE | $F_{ST}$ | $F_{ST}$ 95% CI | *N* | $\theta_S$ per bp ±SD | $\theta_\pi$ per bp ±SD | Tajima's *D* | Fu's $F_S$ | SSD | Raggedness |
| Emigrant Pass | 20 | 3.143 ±0.573 | 0.246 ±0.067 | 0.322 ±0.073 | 0.264 ±0.091 | 0.0589 | 0.0243–0.0985 | 38 | 4.47E−04 ±3.28E−04 | 0.566 ±0.524 | 0.363 | 0.469 | **0.036** | **0.239** |
| Snow Lake, Yosemite NP | 15 | 2.643 ±0.414 | 0.271 ±0.071 | 0.337 ±0.074 | 0.282 ±0.114 | 0.0495 | 0.0233–0.0791 | n/a | | | | | | |
| Roosevelt Lake | 15 | 3.714 ±0.675 | 0.333 ±0.071 | 0.401 ±0.075 | 0.172 ±0.058 | 0.0695 | 0.0311–0.114 | n/a | | | | | | |
| Saddlebag Lake | 18 | 3.286 ±0.714 | 0.254 ±0.066 | 0.322 ±0.075 | 0.157 ±0.063 | 0.116 | 0.0647–0.171 | 32 | 2.33E−04 ±2.33E−04 | 0.498 ±0.485 | 1.535 | 1.67 | 0.022 | 0.248 |
| Dana Meadows | 17 | 3.214 ±0.576 | 0.290 ±0.076 | 0.305 ±0.074 | 0.030 ±0.058 | 0.0373 | 0.0121–0.0666 | n/a | | | | | | |
| Gaylor Lakes | 24 | 3.857 ±0.818 | 0.315 ±0.082 | 0.357 ±0.077 | 0.166 ±0.083 | 0.0977 | 0.0482–0.153 | 34 | 2.30E−04 ±2.30E−04 | 0.499 ±0.485 | 1.559 | 1.711 | **0.022** | **0.249** |
| Cathedral Lakes | 17 | 3.071 ±0.606 | 0.256 ±0.079 | 0.334 ±0.076 | 0.355 ±0.113 | 0.0513 | 0.0175–0.0906 | n/a | | | | | | |
| Lake Elizabeth | 11 | 2.857 ±0.523 | 0.234 ±0.059 | 0.333 ±0.075 | 0.255 ±0.087 | 0.0985 | 0.0327–0.172 | 20 | 2.65E−04 ±2.65E−04 | 0.521 ±0.509 | 1.531 | 1.467 | 0.027 | 0.273 |
| Mono Pass, Yosemite NP | 23 | 3.571 ±0.600 | 0.289 ±0.071 | 0.341 ±0.071 | 0.193 ±0.090 | 0.0332 | 0.0114–0.0566 | n/a | | | | | | |
| Parker Pass | 10 | 2.786 ±0.482 | 0.300 ±0.090 | 0.315 ±0.079 | 0.086 ±0.086 | 0.0655 | 0.0192–0.119 | 28 | 2.41E−04 ±2.41E−04 | 0.516 ±0.498 | 1.607 | 1.661 | 0.026 | **0.267** |
| Rafferty Creek | 21 | 3.357 ±0.617 | 0.252 ±0.059 | 0.302 ±0.065 | 0.159 ±0.064 | 0.104 | 0.0527–0.159 | 42 | 2.18E−04 ±2.18E−04 | 0.508 ±0.487 | 1.667 | 1.871 | **0.024** | **0.258** |
| Maclure Lake | 21 | 3.143 ±0.563 | 0.286 ±0.074 | 0.337 ±0.075 | 0.157 ±0.098 | 0.046 | 0.0182–0.0773 | 40 | 2.20E−04 ±2.20E−04 | 0.501 ±0.484 | 1.617 | 1.816 | **0.022** | **0.251** |
| Snow Lakes, Sierra NF | 28 | 3.286 ±0.691 | 0.334 ±0.074 | 0.336 ±0.077 | −0.028 ±0.053 | 0.0592 | 0.0263–0.0961 | 28 | 2.41E−04 ±2.41E−04 | 0.071 ±0.164 | −1.151 | −1.155 | 0 | 0.74 |
| Humphreys Basin | 31 | 3.500 ±0.635 | 0.274 ±0.064 | 0.313 ±0.074 | 0.053 ±0.049 | 0.0892 | 0.0534–0.125 | 58 | 4.05E−04 ±2.96E−04 | 0.165 ±0.252 | −1.064 | −1.648 | 0.001 | 0.483 |
| Dusy Basin | 23 | 3.429 ±0.618 | 0.264 ±0.066 | 0.323 ±0.077 | 0.192 ±0.063 | 0.0994 | 0.0563–0.145 | 44 | 8.63E−04 ±4.79E−04 | 0.709 ±0.604 | −0.534 | −1.16 | 0.022 | **0.181** |
| Sixty Lakes Basin | 28 | 4.143 ±0.733 | 0.327 ±0.070 | 0.37 ±0.077 | 0.113 ±0.055 | 0.155 | 0.0951–0.220 | 42 | 6.54E−04 ±4.07E−04 | 0.549 ±0.512 | −0.457 | −0.737 | 0.012 | 0.154 |
| Table Meadows | 13 | 2.857 ±0.645 | 0.291 ±0.082 | 0.306 ±0.081 | 0.028 ±0.073 | 0.148 | 0.0826–0.219 | 24 | 2.51E−4 ±2.51E−4 | 0.464 ±0.469 | 1.232 | 1.362 | 0.015 | 0.22 |
| Lost Canyon | 21 | 2.786 ±0.613 | 0.276 ±0.079 | 0.280 ±0.075 | 0.009 ±0.089 | 0.153 | 0.0986–0.217 | 40 | 4.41E−04 ±3.25E−04 | 1.009 ±0.769 | 2.164 | 1.915 | 0.043 | 0.207 |
| Mean | 19.8 | 3.258 ±0.144 | 0.283 ±0.017 | 0.330 ±0.017 | 0.149 ±0.019 | | | 36.2 | 3.62E−04 ±2.87E−04 | 0.506 ±0.48 | 0.774 | 0.711 | 0.021 | 0.29 |

*N*, number of diploid individuals; $A_R$, mean number of alleles; $MH_O$, mean observed heterozygosity; $MH_E$, mean expected heterozygosity; *F*, fixation index ($(H_E-H_O)/H_E$); $\theta_S$, number of segregating sites per base pair; $\theta_\pi$, number of pairwise differences per base pair; *D*, Tajima's D; $F_S$, Fu's statistic; SSD, sum of squared differences test; Raggedness index. Other abbreviations refer to standard error (SE), standard deviation (SD) or 95% confidence intervals (CI).

Significant departures from the null model of neutrality (*D* and $F_S$) or population growth (SSD and Raggedness) are shown in boldface.

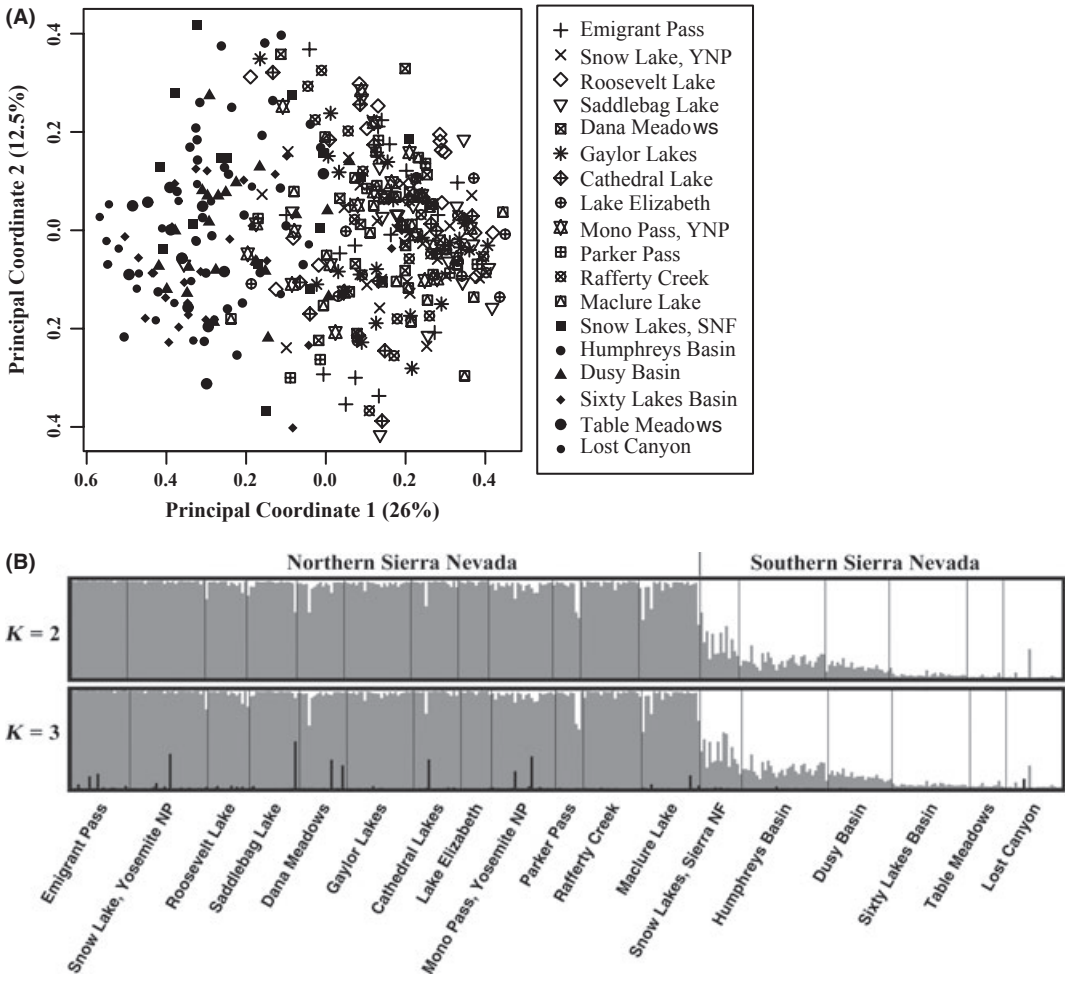


**Fig. 3** Analysis of population structure in *Colias behrii* inferred from a (A) principal coordinate analysis (PCoA), with south Sierra Nevada population labelled with bold symbols, and (B) Bayesian clustering analysis using TESS, with individual admixture proportions for $K = 2$ and $K = 3$ clusters labelled in separate colours.

The PRSF value (3.6) suggests the chains did not converge, but the ESS values were high (>290) and visual inspection of the chains suggested they had stabilized. Estimates of the current population size (mode = 53 935, 95% HPD: 54–47 191 197), ancestral population size (mode = 10 709, 95% HPD: 81–407 115) and time of population size change (mode = 715, 95% HPD: 0–400 676) suggest a >5-fold increase 700 years ago. The bivariate density plot shows a multimodal distribution in the current and ancestral population effective population size, with a smaller secondary peak favouring a bottleneck scenario (Fig. S3B, Supporting information). In the southern Sierra Nevada, a signature of recent population expansion is detected (Bayes Factor 46.23 for a model of growth vs. bottleneck). MCMC diagnostics suggest the chains were well mixed (PRSF = 1.16) and parameter estimates were drawn from a reasonable sample size (ESS > 100). Estimates of the current population size (mode = 204 927, 95% HPD: 2657–266 588 862), ancestral population size (mode = 7208, 95% HPD: 204–221 754) and time of population size change (mode = 1328, 95% HPD: 22–84 040) suggest a >25-fold increase from a small population 1000 years ago. The bivariate density plot shows a distribution favouring a larger current to ancestral effective population size (Fig. S3C, Supporting information).

Approximate Bayesian computation provides evidence of a demographic bottleneck in *Colias behrii*. Cross-validation of the model selection procedure indicates that ABC can assign the three models accurately (Model 1: 77.8%, Model 2: 72.9% and Model 3: 66.3%). Based on posterior model probabilities, the recent bottleneck model is strongly favoured (0.9958) in comparison with the constant size model (0.004) or the historic bottleneck model (0.0001). Similarly, Bayes Factors

**Table 2** Tests for demographic bottleneck based on summary statistic measurements of the microsatellite data

| | Sign-test *P*-value | Standardized difference test *P*-value | Wilcoxon test *P*-value | *M*-ratio Mean | *M*-ratio SD | Critical values for *M*-ratio | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | $\theta$ = 0.004 | $\theta$ = 0.01 | $\theta$ = 0.02 | $\theta$ = 0.04 | $\theta$ = 0.1 | $\theta$ = 0.2 |
| Emigrant Pass | 0.253 | 0.192 | 0.695 | **0.413** | 0.159 | 0.917 | 0.936 | 0.920 | 0.917 | 0.862 | 0.819 |
| Snow Lake, Yosemite NP | **0.001** | **0.005** | **0.034** | **0.380** | 0.165 | 0.917 | 0.933 | 0.912 | 0.912 | 0.850 | 0.798 |
| Roosevelt Lake | 0.166 | 0.192 | 0.465 | **0.338** | 0.110 | 0.932 | 0.935 | 0.917 | 0.911 | 0.852 | 0.804 |
| Saddlebag Lake | **0.037** | **0.026** | **0.040** | **0.412** | 0.124 | 0.933 | 0.933 | 0.921 | 0.912 | 0.860 | 0.817 |
| Dana Meadows | 0.390 | 0.219 | 0.831 | **0.371** | 0.178 | 0.917 | 0.935 | 0.917 | 0.913 | 0.858 | 0.808 |
| Gaylor Lakes | 0.553 | 0.295 | 0.695 | **0.383** | 0.175 | 0.917 | 0.939 | 0.922 | 0.919 | 0.869 | 0.827 |
| Cathedral Lakes | 0.269 | 0.230 | 0.455 | **0.378** | 0.174 | 0.917 | 0.935 | 0.922 | 0.917 | 0.855 | 0.806 |
| Lake Elizabeth | 0.143 | 0.196 | 0.129 | **0.364** | 0.155 | 0.889 | 0.927 | 0.906 | 0.906 | 0.840 | 0.785 |
| Mono Pass, Yosemite NP | **0.032** | **0.012** | 0.110 | **0.388** | 0.167 | 0.933 | 0.939 | 0.921 | 0.920 | 0.867 | 0.825 |
| Parker Pass | 0.227 | 0.265 | 1.000 | **0.356** | 0.157 | 0.889 | 0.929 | 0.908 | 0.907 | 0.831 | 0.778 |
| Rafferty Creek | 0.588 | 0.219 | 0.465 | **0.379** | 0.161 | 0.933 | 0.939 | 0.923 | 0.918 | 0.865 | 0.820 |
| Maclure Lake | 0.164 | **0.044** | 0.206 | **0.380** | 0.188 | 0.933 | 0.939 | 0.921 | 0.917 | 0.868 | 0.824 |
| Snow Lakes, Sierra NF | **0.042** | **0.027** | 0.102 | **0.407** | 0.165 | 0.933 | 0.939 | 0.924 | 0.923 | 0.873 | 0.836 |
| Humphreys Basin | 0.386 | **0.007** | 0.191 | **0.377** | 0.168 | 0.933 | 0.942 | 0.924 | 0.921 | 0.878 | 0.838 |
| Dusy Basin | 0.389 | 0.371 | 0.898 | **0.364** | 0.167 | 0.933 | 0.935 | 0.923 | 0.917 | 0.867 | 0.827 |
| Sixty Lakes Basin | 0.341 | 0.197 | 0.365 | **0.393** | 0.136 | 0.933 | 0.940 | 0.926 | 0.924 | 0.873 | 0.840 |
| Table Meadows | 0.613 | 0.296 | 0.765 | **0.337** | 0.163 | 0.917 | 0.933 | 0.914 | 0.909 | 0.847 | 0.800 |
| Lost Canyon | 0.265 | **0.002** | 0.077 | **0.373** | 0.199 | 0.917 | 0.939 | 0.919 | 0.917 | 0.861 | 0.822 |
| Grand mean | n/a | n/a | n/a | **0.377** | 0.162 | 0.833 | 0.942 | 0.927 | 0.921 | 0.871 | 0.825 |

Significant tests shown in boldface.

**Table 3** Parameter* estimates in the approximate Bayesian computation (ABC) analysis of the recent bottleneck model

| | $N_1$ | $N_{1B}$ | $N_{1A}$ | $N_2$ | $N_{2B}$ | $N_{2A}$ | $\tau_E$ | $\tau_B$ | $\tau_S$ | $M_{12}$ | $M_{21}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Weighted mean | 17 675 | 125 | 24 029 | 13 676 | 124 | 20 194 | 305 | 534 | 5317 | 0.023 | 0.023 |
| Weighted median | 5435 | 105 | 13 691 | 2949 | 103 | 8525 | 281 | 531 | 5372 | 0.022 | 0.022 |
| Weighted 2.5% interval | 350 | 20 | 474 | 343 | 20 | 425 | 20 | 142 | 774 | 0.001 | 0.001 |
| Weighted 97.5% interval | 86 562 | 301 | 89 466 | 82 733 | 300 | 88 303 | 675 | 959 | 9764 | 0.048 | 0.048 |

*Parameters include current effective population size of subpopulation 1 ($N_1$) and subpopulation 2 ($N_2$), bottleneck effective population size of subpopulation 1 ($N_{1B}$) and subpopulation 2 ($N_{2B}$), ancestral effective population size of subpopulation 1 ($N_{1A}$) and subpopulation 2 ($N_{2A}$), time of postbottleneck expansion ($\tau_E$), time of bottleneck ($\tau_B$), time of population splitting ($\tau_S$), migration rate from subpopulation 1–2 ($M_{12}$) and from subpopulation 2–1 ($M_{21}$).

scores favour the recent bottleneck model (3 vs. 1 BF = 241.6, 3 vs. 2 BF = 10427.9). Median parameter estimates based on the recent bottleneck model (Table 3) suggest that population size contracted 531 years ago (95% intervals: 142–959 years), and remained small until populations expanded 281 years ago (95% intervals: 20–675 years). During the bottleneck, the northern and southern populations were reduced to ~1% of their historic population size. Population divergence between the north and south Sierra Nevada occurred at 5372 years (95% intervals: 774–9764 years), and estimated gene flow is symmetrical and quite low ($m$ = 0.022).

### *Environmental predictors of genetic differentiation*

Genetic differentiation ($F_{ST}$) is weakly related to a set of spatial and environmental predictors in both regression approaches (Table S4, Supporting information). Approximately 40% of the overall variation in $F_{ST}$ is explained in the random forest regression, with Euclidean distance (13.9% IncMSE) and distance among mea-

dow patches (15.8% IncMSE) as the most important predictors. Only 35.6% of the variation in $F_{ST}$ is explained in a principal component regression with two components, with Euclidean distance ($b = -0.008$) and meadow patch distance ($b = -0.008$) contributing the largest estimated regression coefficients. In a principal component regression with three components, 54.6% of the variation in $F_{ST}$ is explained, and elevation ($b = -0.044$), Euclidean distance ($b = -0.021$) and meadow patch distance ($b = -0.024$) have the largest regression coefficients.

### *Estimates of abundance and environmental predictors of genetic diversity*

Measurements of local abundance based on line-transect density estimates ($D_{LT}$) varied by ~100-fold (Table 4), with the smallest population at Table Meadows ($1.7 \times 10^{-4}$) and the largest at Mono Pass ($1.6 \times 10^{-2}$). Populations in southeast (Table Meadows and Lost Canyon) are noticeably smaller. However, there is considerable variation at a local scale, even for nearby populations in Yosemite National Park (e.g. Gaylor Lakes $9.9 \times 10^{-3}$ vs. Roosevelt Lake $1.5 \times 10^{-3}$). Measurements of genetic diversity (Tables 1 and 4) do not increase with population abundance. Three statistical measures of genetic diversity, $MH_E$, Shannon's index and MSRD, have no significant relationship to habitat characteristics (Table S5, Supporting information). The average distance to neighbouring patches is a significant predictor of $MA_R$ ($F$ statistic = 5.895, $P = 0.0282$), although this relationship accounts for a small amount of overall variation ($R^2 = 0.282$, residual error = 0.359).

## Discussion

### *Population structure and demographic history of* Colias behrii

In a previous genetic study of *Colias behrii*, variation in sequence data suggested weak population structure, although the overall level of genetic diversity was low (Schoville *et al.* 2011). Here, we provide additional evidence for population divergence in the northern and southern Sierra Nevada, with the split located at a large discontinuity in high-elevation habitat along the San Joaquin River drainage (Fig. 1). Previous work has shown that low-vagility alpine species have deep genetic breaks occurring at the San Joaquin River (Rovito 2010; Schoville & Roderick 2010), but population structure has not been seen in other alpine butterflies (Nice & Shapiro 2001; Schoville & Roderick 2009). In *C. behrii*, isolation does not appear to be complete, as PCoA and Bayesian clustering indicate admixture between the two clusters. We also detected a significant pattern of isolation by distance, suggesting that gene flow might still be occurring. While it is known that Bayesian clustering algorithms sometimes overestimate the number of clusters in the presence of isolation by distance (Frantz

**Table 4** Line-transect estimates of density ($D_{LT}$) with 95% confidence intervals and measures of microsatellite genetic diversity

| Site | $D_{LT}$ mean estimate | $D_{LT}$ harmonic mean | $D_{LT}$ 95% CI | Shannon's index | MSRD |
|---|---|---|---|---|---|
| Emigrant Pass | n/a | n/a | n/a | 0.616 | 4.960 |
| Snow Lake, YNP | 1.74E−03 | 1.74E−03 | 0.00166–0.00182 | 0.604 | 5.303 |
| Roosevelt Lake | 1.46E−03 | 1.46E−03 | 0.00139–0.00152 | 0.763 | 4.967 |
| Saddlebag Lake | 1.42E−03 | 3.49E−04 | 0.00052–0.00386 | 0.617 | 4.213 |
| Dana Meadows | 4.75E−03 | 4.52E−03 | 0.00172–0.01317 | 0.583 | 6.612 |
| Gaylor Lakes | 9.93E−03 | 1.10E−03 | 0.00542–0.01821 | 0.715 | 5.084 |
| Cathedral Lakes | 2.65E−03 | 1.78E−03 | 0.00093–0.00751 | 0.618 | 4.932 |
| Lake Elizabeth | 1.14E−03 | 9.06E−04 | 0.00057–0.00230 | 0.615 | 5.153 |
| Mono Pass, YNP | 1.58E−02 | 1.26E−02 | 0.00930–0.02700 | 0.648 | 4.900 |
| Parker Pass | n/a | n/a | n/a | 0.590 | 5.463 |
| Rafferty Creek | 2.08E−03 | 2.06E−03 | 0.00172–0.00251 | 0.580 | 5.685 |
| Maclure Lake | 6.76E−03 | 6.76E−03 | 0.00648–0.00707 | 0.639 | 4.703 |
| Snow Lake, Sierra NF | n/a | n/a | n/a | 0.631 | 5.986 |
| Humphreys Basin | 1.06E−03 | 1.06E−03 | 0.00101–0.00111 | 0.603 | 3.820 |
| Dusy Basin | 1.09E−03 | 1.09E−03 | 0.00104–0.00114 | 0.623 | 3.954 |
| Sixty Lakes Basin | n/a | n/a | n/a | 0.752 | 5.178 |
| Table Meadows | 1.73E−04 | 1.73E−04 | 0.00017–0.00018 | 0.574 | 5.235 |
| Lost Canyon | 2.71E−04 | 1.03E−04 | 0.00008–0.00094 | 0.521 | 3.777 |

MSRD, mean squared allele size difference.

*et al.* 2009), the concordance of the geographical break in *C. behrii* with other alpine species suggests that this north–south subdivision is not an artefact of sampling.

To investigate the demographic history of *C. behrii*, we examined genetic evidence of population size change based on summary statistics, full likelihood methods and approximate Bayesian computation. Among the summary statistical tests, the *M*-ratio test is expected to be most sensitive for bottleneck detection (Williamson-Natesan 2005), and our results provide strong evidence of bottlenecks in all 18 populations. Tests based on the sign, standardized difference and Wilcoxon statistics were significant in six populations. Counterintuitively, the full likelihood analysis (MSVAR) suggests weak to moderate population growth rather than a recent bottleneck, with dates for expansion occurring around 700–2000 years ago. We then explicitly modelled population structure using approximate Bayesian computation and our model choice procedure favoured a recent bottleneck (531 years ago, 95%: 142–959) followed by rapid growth (281 years ago, 95%: 20–675 years).

Based on the population structure and isolation by distance evident in *C. behrii*, it is not clear whether the single population model of MSVAR can provide meaningful results for our data set. In their simulation study of MSVAR, Chikhi *et al.* (2010) found that population structure can create a false signal of a demographic bottleneck. However, we detected population expansion in the pooled data, as well as subsets of the data where in fact the signal of population expansion was stronger. To explore the possibility that population structure might bias our results, we examined the fit of several structured population models using ABC and included additional data and a calibrated mutation rate for the EF1α locus to help constrain temporal estimates of population size change. Our models were flexible enough to allow for population expansion following a range of bottleneck sizes and bottleneck times, and our results strongly favoured a model with an expansion following a recent bottleneck. We suggest that the weak population expansion detected by MSVAR is not due to population structure, but instead reflects the recent expansion signal in our data with event times that are misleading. In a separate simulation study, Girod *et al.* (2011) found that while MSVAR performs well in detecting demographic change, parameter estimates for demographic expansions were often inaccurate, particularly when converted from scaled parameter values to natural parameter values. Other studies exploring how MSVAR performs under scenarios of rapid population recovery following a bottleneck could shed light on the discrepancy we found in our results.

### *Historical climate change and population responses in the Sierra Nevada*

Climatic oscillations throughout the Quaternary period (2.59 million years to the present) have had profound impacts on the spatial distribution of biodiversity (Sandel *et al.* 2012), including the distribution of genetic variation within species (Hewitt 1996). It is generally accepted that the last glacial maximum (*c.* 31–15 kya) caused major distributional and demographic shifts in alpine species (Varga & Schmitt 2008). While some studies suggest that alpine populations expand after glacial retreats (e.g. Schoville & Roderick 2009; Shafer *et al.* 2010), this is not the case in all alpine species (Galbreath *et al.* 2009). What is notable in the present study is that a bottleneck, which affected populations across the entire species range, is too recent to be consistent with the last glacial maximum.

Holocene climate change is fairly well documented in the Sierra Nevada (Anderson 1990), and with the exception of the Recess Peak glaciation (*c.* 14.1–13.1 kya), there is little evidence of glacial activity in the Sierra Nevada until *c.* 3250 years ago (Bowerman & Clark 2011). Beginning at this time, several advances known as the Matthes glaciations took place, with maximal extents at *c.* 2250, 1650, 750 and 300–220 years ago (Bowerman & Clark 2011). The largest and most recent advance coincides with the 'Little Ace Age' event (1550–1850 AD) recorded in historical meteorological data (Matthes 1939). This time corresponds closely to the median estimate of the population bottleneck in *C. behrii* (531 years ago), although the confidence intervals around this estimate (142–959 years ago) are large. It also seems plausible that glacial activity could lead to a range-wide bottleneck, because mortality is known to increase in alpine butterfly populations following unseasonal snowfall or shortened summer growth periods (Ehrlich *et al.* 1972; Hayes 1981). Bowerman & Clark (2011) estimate that mean summer temperature would have decreased by ~0.2–2.0 °C and winter accumulation of snow would have increased by ~3–26 cm snow water equivalents during this time period. One caveat is that other climatic events are also known to have occurred in the Sierra Nevada in the recent past, including a pronounced drought period (800–1300 AD) that could have affected habitat suitability in alpine meadows (Kleppe *et al.* 2011). Another important caveat is that our estimates of the bottleneck time are dependent on assumptions about the mutation rate at microsatellite loci and the EF1α locus. There is evidence that time dependency of mutation rates affects recent divergence time estimates (Ho *et al.* 2005) and this could bias the EF1α rate for population inference. However, we would expect time dependency to underestimate the mutation

rate at EF1α and therefore to push our temporal estimates forward in time (towards the present). This should have little effect on our conclusions of a recent historical bottleneck in *C. behrii*.

### *Patterns of genetic variation in alpine specie: demographic history vs. environmental factors*

Environmental factors clearly affect the distribution and local population dynamics of alpine species (Bliss 1962), with temperature variation playing a dominant role in regulating the life cycle and daily activity patterns of cold-adapted insects (Downes 1965; Mani 1968). A number of studies have found that contemporary landscape factors explain patterns of genetic differentiation among alpine populations (e.g. Murphy *et al.* 2010; Savage *et al.* 2010), and these landscape factors may in fact influence demographic responses to climate change (Alvarez *et al.* 2009). In alpine butterflies, environmental factors are known to influence landscape connectivity among populations, as in *Parnassius smintheus*, where connectivity is affected by local patch characteristics and the presence of forested habitat (Matter *et al.* 2003; Roland & Matter 2007). In general, connectivity is low for butterflies across alpine landscapes (e.g. Junker *et al.* 2010), with dispersal on the order of several kilometres in *Colias* species and often depending on the shape and size of suitable habitat patches (Watt *et al.* 1977; Hayes 1981). Thus, we would expect to see genetic differentiation and levels of genetic diversity reflect measures of connectivity and habitat quality in *C. behrii*. Instead, we found that environmental variables made small contributions to explaining genetic diversity patterns, with Euclidean geographical distance and distance among meadow patches as the best predictors. Other studies have pointed out that there is often a temporal lag in the response of genetic variation to landscape factors and demographic events in butterfly metapopulations (Orsini *et al.* 2008), suggesting that historical habitat features might be better predictors of contemporary patterns of genetic diversity. However, our study focuses over a broad spatial scale of an entire species range, where local habitat changes are less likely to alter patterns of range-wide population connectivity. Instead, the decoupling of environmental predictors and genetic diversity patterns is likely to be due to the overwhelming signal of a recent population bottleneck. Genetic bottlenecks decrease levels of genetic variability within populations and increase genetic drift among populations (Nei *et al.* 1975), acting to diminish the signal of gene flow patterns among local populations. Are genetic bottlenecks unusual for alpine species? A number of studies (Paun *et al.* 2008; Muster *et al.* 2009; Frei *et al.* 2011; Shama *et al.* 2011) have recently shown that demographic history plays a larger role in shaping current patterns of genetic differentiation in alpine species and that the influence of large and/or recent population size changes may not be uncommon.

## Acknowledgements

We are grateful to the following people for help in the field: Sean Rovito, Tina Cheng, Vance Vredenburg, Rebecca Chong, Dave Daversa and Tate Tunstall. We would like to thank Sarah Stock and Steve Thompson for help in obtaining funding and Olivier François and Flora Jay for discussions on statistical analysis. This research was supported by grants to SD Schoville from the Yosemite Fund (NPS J8C07070002), National Science Foundation (OISE-0965038), White Mountain Research Station, Magy Fellowship and Walker Fund. Specimens were collected under permits granted by Sequoia and Kings National Parks (Study# SEKI-0091), Yosemite National Park (Study# YOSE-00093) and the California Fish and Game Department (#SC-006997).

## References

Alvarez N, Thiel-Egenter C, Tribsch A *et al.* (2009) History or ecology? Substrate type as a major driver of spatial genetic structure in Alpine plants. *Ecology Letters*, **12**, 632–640.

Anderson RS (1990) Holocene forest development and paleoclimates within the central Sierra Nevada, California. *Journal of Ecology*, **78**, 470–489.

Beaumont MA (1999) Detecting population expansion and decline using microsatellites. *Genetics*, **153**, 2013–2029.

Bliss WC (1962) Adaptations of arctic and alpine plants to environmental conditions. *Arctic*, **15**, 117–144.

Blum MGB, François O (2010) Non-linear regression models for approximate Bayesian computation. *Statistics and Computing*, **20**, 63–73.

Bowerman ND, Clark DH (2011) Holocene glaciation of the central Sierra Nevada, California. *Quaternary Science Reviews*, **30**, 1067–1085.

Braby MF, Vila R, Pierce NE (2006) Molecular phylogeny and systematics of the Pieridae (Lepidoptera: Papilionoidea): higher classification and biogeography. *Zoological Journal of the Linnean Society*, **147**, 239–275.

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

Chikhi L, Sousa VC, Luisi P, Goossens B, Beaumont MA (2010) The confounding effects of population structure, genetic diversity and the sampling scheme on the detection and quantification of population size changes. *Genetics*, **186**, 983–995.

Cornuet J, Luikart G (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics*, **144**, 2001.

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

Csilléry K, Blum M, Francois O (2012) abc: an R package for approximate Bayesian computation (ABC). *Methods in Ecology and Evolution*, **3**, 475–479.

Downes JA (1965) Adaptations of insects in the arctic. *Annual Review of Entomology*, **10**, 257–274.

Durand E, Jay F, Gaggiotti OE, François O (2009) Spatial inference of admixture proportions and secondary contact zones. *Molecular Biology and Evolution*, **26**, 1963–1973.

Dyer RJ, Nason JD, Garrick RC (2010) Landscape modelling of gene flow: improved power using conditional genetic distance derived from the topology of population networks. *Molecular Ecology*, **19**, 3746–3759.

Edwards WH (1866) Descriptions of certain species of diurnal Lepidoptera found within the limits of the United States and British America, no. 5. *Proceedings of the Entomological Society of Philadelphia*, **6**, 200, 201, no. 202.

Ehrich D, Gaudeul M, Assefa A *et al.* (2007) Genetic consequences of Pleistocene range shifts: contrast between the Arctic, the Alps and the East African mountains. *Molecular Ecology*, **16**, 2542–2559.

Ehrlich PR, Breedlove DE, Brussard PF, Sharp MA (1972) Weather and the ''regulation'' of subalpine populations. *Ecology*, **53**, 243–247.

Elias SA (1991) Insects and climate change. *BioScience*, **41**, 552–559.

Etherington TR (2010) Python based GIS tools for landscape genetics: visualising genetic relatedness and measuring landscape connectivity. *Methods in Ecology and Evolution*, **2**, 52–55.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564–567.

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure: extensions to linked loci and correlated allele frequencies. *Genetics*, **164**, 1567–1587.

Foll M, Gaggiotti O (2006) Identifying the environmental factors that determine the genetic structure of populations. *Genetics*, **174**, 875–891.

François O, Durand E (2010) THE STATE OF THE FIELD: spatially explicit Bayesian clustering models in population genetics. *Molecular Ecology Resources*, **10**, 773–784.

Frantz A, Cellina S, Krier A, Schley L, Burke T (2009) Using spatial Bayesian methods to determine the genetic structure of a continuously distributed population: clusters or isolation by distance? *Journal of Applied Ecology*, **46**, 493–505.

Frei E, Scheepens J, Stöcklin J (2011) High genetic differentiation in populations of the rare alpine plant species *Campanula thyrsoides* on a small mountain. *Alpine Botany*, **122**, 23–34.

Gaggiotti OE, Bekkevold D, Jørgensen HBH *et al.* (2009) Disentangling the effects of evolutionary, demographic, and environmental factors influencing genetic structure of natural populations: atlantic herring as a case study. *Evolution*, **63**, 2939–2951.

Galbreath KE, Hafner DJ, Zamudio KR (2009) When cold is better: climate-driven elevation shifts yield complex patterns of diversification and demography in an alpine specialist (American Pika, *Ochotona princeps*). *Evolution*, **63**, 2848–2863.

Garza JC, Williamson EG (2001) Detection of reduction in population size using data from microsatellite loci. *Molecular Ecology*, **10**, 305–318.

Gillespie AR, Zehfuss PH (2004) Glaciations of the Sierra Nevada, California, USA. In: *Quaternary Glaciations- Extent and Chronology- Part II: North America* (eds Ehlers J, Gibbard PL), pp. 51–62. Elsevier, Amsterdam.

Girod C, Vitalis R, Leblois R, Fréville H (2011) Inferring population decline and expansion from microsatellite data: a simulation-based evaluation of the Msvar method. *Genetics*, **188**, 165–179.

Hayes JL (1981) The population ecology of a natural population of the pierid butterfly *Colias alexandra*. *Oecologia*, **49**, 188–200.

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.

Hirao AS, Kudo G (2004) Landscape genetics of alpine-snowbed plants: comparisons along geographic and snowmelt gradients. *Heredity*, **93**, 290–298.

Ho SYW, Phillips MJ, Cooper A, Drummond AJ (2005) Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. *Molecular Biology and Evolution*, **22**, 1561–1568.

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, **23**, 1801–1806.

Junker M, Wagner S, Gros P, Schmitt T (2010) Changing demography and dispersal behaviour: ecological adaptations in an alpine butterfly. *Oecologia*, **164**, 971–980.

Keyghobadi N, Roland J, Strobeck C (1999) Influence of landscape on population genetic structure of the alpine butterfly *Parnassius smintheus* (Papilionidae). *Molecular Ecology*, **8**, 1481–1495.

Kleppe JA, Borthers DS, Kent GM *et al.* (2011) Duration and severity of Medieval drought in the Lake Tahoe Basin. *Quaternary Science Reviews*, **30**, 3269–3279.

Korner C (1999) *Alpine Plant Life: Functional Plant Ecology of High Mountain Ecosystems*. Springer, Berlin.

Laval G, Excoffier L (2004) SIMCOAL 2.0: a program to simulate genomic diversity over large recombining regions in a subdivided population with a complex history. *Bioinformatics*, **20**, 2485–2487.

Lehmann T, Hawley WA, Grebert H, Collins FH (1998) The effective population size of *Anopheles gambiae* in Kenya: implications for population structure. *Molecular Biology and Evolution*, **15**, 264–276.

Lowe WH, Allendorf FW (2010) What can genetics tell us about population connectivity? *Molecular Ecology*, **19**, 3038–3051.

Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology & Evolution*, **18**, 189–197.

Mani MS (1968) *Ecology and Biogeography of High Altitude Insects*. Dr. W. Junk N.V. Publishers, The Hague.

Marko PB, Hart MW (2011) The complex analytical landscape of gene flow inference. *Trends in Ecology & Evolution*, **26**, 448–456.

Matter SF, Roland J, Keyghobadi N, Sabourin K (2003) The effects of isolation, habitat area and resources on the abundance, density and movement of the butterfly *Parnassius smintheus*. *American Midland Naturalist*, **150**, 26–36.

Matthes FE (1939) Report of the committee on glaciers. *Transactions of the American Geophysical Union*, **20**, 518–523.

Moilanen A, Hanski I (1998) Metapopulation dynamics: effects of habitat quality and landscape structure. *Ecology*, **79**, 2503–2515.

Moilanen A, Nieminen M (2002) Simple connectivity measures in spatial ecology. *Ecology*, **83**, 1131–1145.

Molecular Ecology Resources Primer Development Consortium, An J, Bechet A *et al.* (2010) Permanent genetic resources added to molecular ecology resources database 1 October 2009–30 November 2009. *Molecular Ecology Resources*, **10**, 404–408.

Murphy MA, Dezzani R, Pilliod DS, Storfer A (2010) Landscape genetics of high mountain frog metapopulations. *Molecular Ecology*, **19**, 3634–3649.

Muster C, Maddison WP, Uhlmann S, Berendonk TU, Vogler AP (2009) Arctic-alpine distributions-metapopulations on a continental scale? *The American Naturalist*, **173**, 313–326.

Nei M, Maruyama T, Chakraborty R (1975) Bottleneck effect and genetic variability in populations. *Evolution*, **29**, 1–10.

Nice CC, Shapiro AM (2001) Patterns of morphological, biochemical, and molecular evolution in the *Oeneis chryxus* complex (Lepidoptera: Satyridae): a test of historical biogeographical hypotheses. *Molecular Phylogenetics and Evolution*, **20**, 111–123.

Novembre J, Stephens M (2008) Interpreting principal component analyses of spatial population genetic variation. *Nature Genetics*, **40**, 646–649.

Orsini L, Corander J, Alasentie A, Hanski I (2008) Genetic spatial structure in a butterfly metapopulation correlates better with past than present demographic structure. *Molecular Ecology*, **17**, 2629–2642.

Paun O, Schonswetter P, Winkler M, Tribsch A; Intrabiodiv Consortium (2008) Historical divergence vs. contemporary gene flow: evolutionary history of the calcicole *Ranunculus alpestris* group (Ranunculaceae) in the European Alps and the Carpathians. *Molecular Ecology*, **17**, 4263–4275.

Peakall R, Smouse PE (2006) GENALEX version 6.1: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, **6**, 288–295.

Piry S, Luikart G, Cornuet JM (1999) Computer note. BOTTLENECK: a computer program for detecting recent reductions in the effective size using allele frequency data. *Journal of Heredity*, **90**, 502.

PRISM Group (2007) Oregon State University. Available from http://www.prismclimate.org.

R Development Core Team (2011) *R Version 2.13.2. The R Foundation for Statistical Computing*. R Development Core Team, Vienna, Austria.

Roland J (1982) Melanism and diel activity of alpine *Colias* (Lepidoptera: Pieridae). *Oecologia*, **53**, 214–221.

Roland J, Matter SF (2007) Encroaching forests decouple alpine butterfly population dynamics. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 13702–13704.

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

Rousset F (2008) Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. *Molecular Ecology Resources*, **8**, 103–106.

Rovito SM (2010) Lineage divergence and speciation in the Web-toed Salamanders (Plethodontidae: *Hydromantes*) of the Sierra Nevada, California. *Molecular Ecology*, **19**, 4554–4571.

Sandel B, Arge L, Dalsgaard B *et al.* (2012) The influence of Late Quaternary climate-change velocity on species endemism. *Science*, **334**, 660–664.

Savage WK, Fremier AK, Bradley Shaffer H (2010) Landscape genetics of alpine Sierra Nevada salamanders reveal extreme population subdivision in space and time. *Molecular Ecology*, **19**, 3301–3314.

Schoville SD, Roderick GK (2009) Alpine biogeography of Parnassian butterflies during Quaternary climate cycles in North America. *Molecular Ecology*, **18**, 3471–3485.

Schoville SD, Roderick GK (2010) Evolutionary diversification of cryophilic *Grylloblatta* species (Grylloblattodea: Grylloblattidae) in alpine habitats of California. *BMC Evolutionary Biology*, **10**, 163.

Schoville SD, Stuckey M, Roderick GK (2011) Pleistocene origin and population history of a neoendemic alpine butterfly. *Molecular Ecology*, **20**, 1233–1247.

Schug MD, Hutter CM, Wetterstrand KA *et al.* (1998) The mutation rates of di-, tri-, and tetranucleotide repeats in *Drosophila melanogaster*. *Molecular Biology and Evolution*, **15**, 1751–1760.

Segelbacher G, Cushman S, Epperson B *et al.* (2010) Applications of landscape genetics in conservation biology: concepts and challenges. *Conservation Genetics*, **11**, 375–385.

Shafer ABA, Côté SD, Coltman DW (2010) Hot spots of genetic diversity descended from multiple Pleistocene refugia in an alpine ungulate. *Evolution*, **65**, 125–138.

Shama L, Kubow K, Jokela J, Robinson C (2011) Bottlenecks drive temporal and spatial genetic changes in alpine caddisfly metapopulations. *BMC Evolutionary Biology*, **11**, 278.

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

Stanton ML, Galen C (1997) Life on the edge: adaptation versus environmentally mediated gene flow in the snow buttercup, *Ranunculus adoneus*. *The American Naturalist*, **150**, 143–178.

Storz JF, Beaumont MA (2002) Testing for genetic evidence of population expansion and contraction: an empirical analysis of microsatellite DNA variation using a hierarchical Bayesian model. *Evolution*, **56**, 154–166.

Tajima F (1989) The effect of change in population size on DNA polymorphism. *Genetics*, **123**, 597–601.

Thiel-Egenter C, Alvarez N, Holderegger R *et al.* (2011) Break zones in the distributions of alleles and species in alpine plants. *Journal of Biogeography*, **38**, 772–782.

Thomas L, Laake JL, Rexstad E *et al.* (2009) *Distance 6.0. Release 2*. Research Unit for Wildlife Population Assessment, University of St. Andrews, UK.

Turner MG (1989) Landscape ecology: the effect of pattern on process. *Annual Review of Ecology and Systematics*, **20**, 171–197.

Varga ZS, Schmitt T (2008) Types of oreal and oreotundral disjunctions in the western Palearctic. *Biological Journal of the Linnean Society*, **93**, 415–430.

Waples RS (2005) Genetic estimates of contemporary effective population size: to what time periods do the estimates apply? *Molecular Ecology*, **14**, 3335–3352.

Watt WB, Chew FS, Snyder LRG, Watt AG, Rothschild DE (1977) Population structure of pierid butterflies. *Oecologia*, **27**, 1–22.

Wegmann D, Leuenberger C, Neuenschwander S, Excoffier L (2010) ABCtoolbox: a versatile toolkit for approximate Bayesian computations. *BMC Bioinformatics*, **11**, 116.

Williamson-Natesan E (2005) Comparison of methods for detecting bottlenecks from microsatellite loci. *Conservation Genetics*, **6**, 551–562.

Zellmer AJ, Knowles LL (2009) Disentangling the effects of historic vs. contemporary landscape structure on population genetic divergence. *Molecular Ecology*, **18**, 3593–3602.

---

S.D.S's research focuses on population history, climate charge and conservation of alpine insects. This work was completed as part of his PhD research on alpine biogeography. A.W.L. is a molecular biologist whose interests include population biology and invasions. She is currently a graduate student at the University of California, Berkeley. G.L.R. is a Professor in the Department of Environmental Science, Policy and Management with research interests in historical population dynamics, invasion biology, and biodiversity, mostly of insects.

---

## Data accessibility

DNA sequences: GenBank accessions FJ851639–FJ851646.

Microsatellite data with sampling locations, and STRUCTURE input files uploaded on DRYAD as entry doi:10.5061/dryad.c7f1f.

## Supporting information

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

**Table S1** Priors and rules used to generate data for approximate Bayesian computation analysis.

**Table S2** Location and dates of line transect surveys.

**Table S3** Test of deviation from Hardy–Weinberg equilibrium (HWE) in microsatellite loci of *Colias behrii*.

**Table S4** Regression models of genetic differentiation with predictors for population structure and population connectivity.

**Table S5** Linear regression models of genetic diversity with predictors for meadow area, average distance to nearby meadow patches, total neighborhood size of patches, and density.

**Fig. S1** Model performance with an increasing number of population clusters ($K$) in (A) STRUCTURE and (B) TESS. Rate of change in model performance ($\Delta K$) based on the number of population clusters ($K$) in (C) STRUCTURE and (D) TESS.

**Fig. S2** Estimation of individual admixture proportions and cluster membership using $K = 2$ clusters in the Bayesian program STRUCTURE.

**Fig. S3** Bivariate density plots of population size parameters based on the MSVAR analysis of (A) all *Colias behrii* samples, (B) northern and (C) southern samples.

Please note: Wiley-Blackwell are 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.