

# INFERRING THE PAST AND PRESENT CONNECTIVITY ACROSS THE RANGE OF A NORTH AMERICAN LEAF BEETLE: COMBINING ECOLOGICAL NICHE MODELING AND A GEOGRAPHICALLY EXPLICIT MODEL OF COALESCENCE

Simon Dellicour,<sup>1</sup> Shannon Fearnley,<sup>2</sup> Anicée Lombal,<sup>1</sup> Sarah Heidl,<sup>2</sup> Elizabeth P. Dahlhoff,<sup>3</sup> Nathan E. Rank,<sup>2</sup> and Patrick Mardulyn<sup>1,4</sup>

<sup>1</sup>*Evolutionary Biology and Ecology, Université Libre de Bruxelles, Avenue FD Roosevelt 50, B-1050 Brussels, Belgium*

<sup>2</sup>*Department of Biology, Sonoma State University, Rohnert Park, California 94928*

<sup>3</sup>*Department of Biology, Santa Clara University, Santa Clara, California 95053*

<sup>4</sup>*E-mail: pmarduly@ulb.ac.be*

Received August 21, 2013

Accepted April 8, 2014

The leaf beetle *Chrysomela aeneicollis* occurs across Western North America, either at high elevation or in small, isolated populations along the coast, and thus has a highly fragmented distribution. DNA sequence data (three loci) were collected from five regions across the species range. Population connectivity was examined using traditional ecological niche modeling, which suggested that gene flow could occur among regions now and in the past. We developed geographically explicit coalescence models of sequence evolution that incorporated a two-dimensional representation of the hypothesized ranges suggested by the niche-modeling estimates. We simulated sequence data according to these models and compared them to observed sequences to identify most probable scenarios regarding the migration history of *C. aeneicollis*. Our results disagreed with initial niche-modeling estimates by clearly rejecting recent connectivity among regions, and were instead most consistent with a long period of range fragmentation, extending well beyond the last glacial maximum. This application of geographically explicit models of coalescence has highlighted some limitations of the use of climatic variables for predicting the present and past range of a species and has explained aspects of the Pleistocene evolutionary history of a cold-adapted organism in Western North America.

**KEY WORDS:** *Chrysomela aeneicollis*, coalescent simulations, leaf beetle, phylogeography, PHYLOGEOSIM, range fragmentation.

Phylogeographic studies offer an opportunity to evaluate the impact of major climatic events on contemporary and historical geographic ranges of species, as these strongly influence current patterns of genetic variation within them (e.g., Hewitt 2003, 2004). Such studies combine data on DNA sequence variation and geographic distribution, and infer the evolutionary history of natural populations over space and time (e.g., Avise 2000; Hewitt 2001).

In view of the results generated by a large body of previous studies (e.g., reviews in Taberlet et al. 1998; Hewitt 1999; Avise 2000), phylogeographic data appear most informative in a time window that spans the second half of the Pleistocene, which is probably a consequence of the tempo of coalescence of lineages through time and the rate of DNA sequence substitutions. During this time frame, the earth experienced a series of major climatic events that



had profound impacts on the distribution of organisms. For example, at least half of Western North America was covered by the Cordilleran ice sheet for ~90,000 years every Milankovitch cycle of ~100,000 years. This extensive ice sheet reached down into the state of Washington at the height of the last ice age (~20,000 years before present). During that same time, relatively mild conditions along the coast allowed north temperate and boreal species to take refuge in coastal habitats, regions that are often occupied today by otherwise Arctic species (e.g., Brunfeldt et al. 2001).

In some cases, prior knowledge about the history of a study organism is sufficient for a priori definition of a few restricted hypotheses to be tested and compared with observed data on genetic variation, often in a Bayesian framework (e.g., Knowles 2001; Fagundes et al. 2007; Laurent et al. 2011). In other cases however, we lack sufficient information about the natural history of the organism of interest to restrict the potential historical hypotheses to a tractable number for statistical modeling of the pattern of observed genetic variation. In these cases, external information can be sought to develop plausible alternative hypotheses and exclude implausible ones. To achieve this goal, several authors have proposed that data on past and present environmental variables within a species range be integrated through ecological niche modeling (Carstens and Richards 2007; Knowles et al. 2007; Hickerson et al. 2010; Knowles and Alvarado-Serrano 2010). Past and present geographic distributions of a species are estimated using these environmental data. These values are used to constrain the space of possible hypotheses to a few alternative historical scenarios, which are then evaluated using molecular genetic data. For this purpose, we have developed a simulation program (PHYLOGEOSIM 1.0; available with a detailed manual at <http://ebe.ulb.ac.be/ebe/Software.html>) that models evolution of sequences in a two-dimensional space represented by a grid. The grid integrates information on the estimated species range to simulate evolution of target DNA sequences in an explicit geographical framework. Such a spatially explicit model of coalescence appears well suited to consider movement of individuals and genes across a species range while attempting to connect current patterns of genetic variation with the evolution of the species range over time. In contrast to a more classic population model, which would consider each continuous portion of the range as a separate panmictic population, this approach may allow us to take into account limited dispersal of individuals within continuous portions of a species range. Therefore, this model can also account for patterns of isolation by distance that are often observed within a range.

In this study, we focus on a North American willow leaf beetle, *Chrysomela aeneicollis*. This beetle is found in isolated populations at high elevation in the Rocky (Montana, Colorado)

and Sierra Nevada (California) mountains, and in a few isolated areas along the west coast of Oregon and California (USA). Beetles are univoltine and overwinter as adults. Adults emerge from winter diapause in May or June to feed, mate, and lay eggs. Larvae hatch and mature through three instars on the same host plant as their parents (Rank 1992a,b). In late summer larvae pupate, and, shortly thereafter, new adults emerge. They feed until they retreat to the leaf litter at the base of their host plant in late fall (Smiley and Rank 1986). Like many leaf beetles, the diet of *C. aeneicollis* is restricted to a few host species (in this case salicylate-rich willows; Rank 1994); furthermore, the high level of population differentiation suggested by molecular markers imply that this species typically has low rates of dispersal (Rank 1992a; Knoll et al. 1996; Mardulyn and Milinkovitch 2005). The highly fragmented distribution and limited dispersal ability of this species provides an opportunity to study the consequences of climate-induced range expansion and contraction on current patterns of genetic variation in a cold-adapted and specialized herbivorous insect.

To date, most phylogeographic studies in North America and Europe have focused on organisms living in temperate climates, and have shown that their ranges were generally more fragmented during glacial episodes (e.g., Hewitt 1996). Indeed, past fragmentation in these species has often led to the divergence of distinct genetic lineages, which sometimes meet again within secondary contact zones, after the distribution of the species re-expands (Hewitt 1999, 2004). Fewer data are available on organisms adapted to a more extreme, cold environment. Given that overall lower temperatures characterize ice ages in the northern hemisphere, we could hypothesize that, contrary to temperate climate organisms, cold-adapted species should display a less fragmented range during glacial periods.

With this in mind, we collected samples across a large portion of the range of *C. aeneicollis* and analyzed its sequence variation at three independent loci. The aims of our study were as follows: (1) to characterize phylogeographic structure within *C. aeneicollis*; (2) to illustrate the process of deriving spatially explicit coalescence models from distribution models to test historical hypotheses; (3) to test whether gene flow currently occurs among the main regions of the contemporary range of *C. aeneicollis*, given its current fragmented distribution and to test whether this gene flow occurred during the last ice age or was a more ancient phenomenon. To avoid the inference of a general pattern from a single case study, we also compared the phylogeographic pattern uncovered for *C. aeneicollis* to that of another cold-adapted leaf beetle, *C. lapponica* (Mardulyn et al. 2011), a closely related species sharing many ecological traits with *C. aeneicollis* (similar life cycle and feeding exclusively on willow and birch) and characterized by a similarly fragmented distribution in Eurasia.

## Material and Methods

### OVERALL EXPERIMENTAL APPROACH

We combined DNA sequence variation data from three putatively unlinked loci with an ecological niche-modeling analysis. Detailed hypotheses about how connectivity has evolved among regions over time were developed by superimposing a two-dimensional grid onto distribution estimates and incorporating cells that corresponded to a minimum probability of occurrence (values provided by the ecological niche modeling) into the species range. Modifying the minimum probability limit suggested different scenarios, which were used as a basis to create alternative phylogeographic hypotheses that differed (1) in level of fragmentation of the species range and (2) over the time at which fragmentation started to occur (e.g., before or after the last glacial maximum [LGM]). These hypotheses were then translated into coalescence models used to simulate evolution of DNA sequences by directly importing distribution grids into the simulation program. We compared simulated and observed data to identify most probable hypotheses to gain insights into the history of migration of *C. aeneicollis*.

### SAMPLING AND SEQUENCING

We collected *C. aeneicollis* individuals from 10 localities, spread over the southern portion (representing >50%) of the species current distribution (Table 1, Fig. 1). Genomic DNA was extracted from 151 individuals using the Qiagen DNeasy<sup>®</sup> Blood & Tissue kit. A half thorax was ground per specimen in the Qiagen ATL buffer and incubated overnight with proteinase K at 56°C. The remaining DNA-extraction steps were conducted as described in the manufacturer's protocol. We sequenced 151 copies of a ~1000 base pair (bp) long fragment of the mitochondrial gene cytochrome oxidase I (COI), 128 copies (from 64 individuals) of a ~500 bp long fragment of the nuclear protein-coding gene 60S acidic ribosomal protein P0 (RpP0), and 136 copies (from 68 individuals) of a ~500 bp long fragment of the nuclear gene Wingless (WgL). All fragments were PCR amplified with the TrueStart Hot Start Taq DNA polymerase, following the guidelines in the manufacturer's protocol (Fermentas International Inc.). The COI fragment was amplified (annealing temperature of 52°C) using primers TL2-N-3014 and C1-J-1751 (Simon et al. 1994), the RpP0 fragment (annealing temperature of 56°C) with primers 5'-ATGGGTAGGGAGGACAAIGCIACITGG-3' and 5'-GCDATIGCICCIACGRGCYGGIG-3' (Gómez-Zurita et al. 2004), and the WgL fragment (annealing temperature of 63°C) with primers 5'-ACTICGCARCACCARTGGAAT-3' and 5'-GARTGYAARTGYCAYGGYATC-3' (Danforth et al. 2004). For nuclear genes, when the algorithm implemented in PHASE (see below) did not manage to recover the haplotype phase for a heterozygote individual (which was the case for 14 PCR products), the corresponding PCR product was cloned. These ampli-

fied products were purified and ligated into a pGEM-3Z vector (Promega), then transferred to *Escherichia coli* JM109 competent cells. At least five clones per PCR product were sequenced, and these sequences were compared to the one initially obtained by direct sequencing. When sequences of all five clones were identical or when evidence of PCR recombination occurred, additional clones were sequenced to ensure an accurate allele sequence reconstruction.

### DATA ANALYSES

Sequences were aligned using the MUSCLE algorithm (Edgar 2004) implemented in the software CODONCODE ALIGNER (version 3.7.1.1, Codon Code Corporation). These alignments were checked manually and pruned at both 5'- and 3'-ends to ensure the absence of a trailing gap in the final dataset. A single gap, 4 bp long, detected in the RpP0 dataset, was recoded as a separate character for all analyses. For heterozygote genotypes, we used the maximum-likelihood method implemented in the software PHASE 2.2.1 (Stephens et al. 2001; Stephens and Donnelly 2003) to attempt reconstructing the haplotype phase of the sequences. We conducted five independent runs of 10,000 iterations per genotype, while thinning at every 100 steps and discarding the first 1000 samples as burn-in. We checked convergence among chains by comparing haplotype reconstructions inferred by each of the five independent runs. The PCR products of phased haplotypes for which a probability > 0.8 could not be obtained were cloned (see above). However, for 10 heterozygote genotypes (four RpP0 and six WgL genotypes), the cloning procedure failed after two attempts. In those cases, we ended using the most probable haplotype phase provided by PHASE although the associated probability was lower than 0.8 (ranging from 0.36 to 0.69).

Median-joining networks (Bandelt et al. 1999) were inferred for each gene fragment using the software NETWORK 4.6.6 (available at <http://www.fluxus-engineering.com>) with  $\epsilon = 0$ . According to their geographic position, sampled populations were pooled into five regions: Oregon, Montana, California, Sierra Nevada, and Colorado (Table 1, Fig. 1). This a priori subdivision of the samples is based on the topography of the Western United States, where geographic distances are large and mountainous regions are separated by dryer lowland habitats, and on the apparent fragmentation of the species range (e.g., gap between California coast and Sierra Nevada populations). Nucleotide diversity (Nei and Li, 1979), relative nucleotide diversity (i.e., nucleotide diversity within one region divided by nucleotide diversity for the rest of the species range; Mardulyn et al. 2009), and the number of different haplotypes within each of five regions were computed with the software SPADS 1.0 (available with a detailed manual, at <http://ebe.ulb.ac.be/ebe/Software.html>; Dellicour and Mardulyn, 2014). We also used SPADS to calculate a series of

**Table 1.** Sampling localities and distribution of haplotypes for the three loci used in this study.

Population locality	Geographical coordinates	<i>n</i> (COI)	Haplotypes (no. of copies) COI	<i>n</i> (RpP0)	Haplotypes (no. of copies) RpP0	<i>n</i> (WgL)	Haplotypes (no. of copies) WgL
(A) Cannon Beach (Oregon)	45°53'N, 123°57'W	12	13, 15(11)	7	8, 14(2), 15, 16(10)	8	1, 2(9), 3(2), 14(2), 20(2)
(B) Flathead Valley (Montana)	48°18'N, 113°21'W	11	19(9), 23, 24	10	3(2), 7(5), 9, 11(5), 12(7)	9	3(6), 4, 5, 6, 7, 8, 9, 10(2), 13(2), 18, 24
(C) Gualala River (California)	38°50'N, 123°30'W	9	20(9)	9	1(6), 2(6), 4(3), 5, 6(2)	9	3(9), 4(9)
(D1) Rock Creek (Sierra Nevada)	37°25'N, 118°24'W	17	4(15), 5, 9	4	7(8)	4	8(4), 11(4)
(D2) Bishop Creek (Sierra Nevada)	37°18'N, 118°56'W	16	2(3), 4(10), 8, 10(2)	4	7(6), 11, 13	4	8(5), 11 (3)
(D3) Big Pine Creek (Sierra Nevada)	37°07'N, 118°29'W	18	2, 4, 8(10), 12(6)	4	7(4), 10(2), 11(2)	4	8(5), 11 (3)
(D4) Taboose Pass (Sierra Nevada)	36°59'N, 118°24'W	37	1, 3, 6(2), 7(4), 8(28), 11	5	7(6), 9(2), 11, 13	5	8(8), 11 (2)
(E1) Twin Lakes (Colorado)	39°02'N, 106°15'W	11	14, 17, 18(3), 21, 22(5)	5	18(6), 19(2), 20(2)	5	12, 16, 17, 19, 21(2), 22, 26, 28(2)
(E2) Lake Windsor Trail (Colorado)	39°14'N, 106°28'W	10	14(6), 16, 18, 21(2)	9	7(4), 17, 19, 18(12)	10	16(4), 17(7), 19, 28, 29(7)
(E3) Timberline Lake Trail (Colorado)	39°17'N, 106°27'W	10	14, 17(2), 18(7)	7	7(2), 18(10), 21(2)	10	16(5), 15(2), 17(7), 20, 21, 23, 25(2), 27

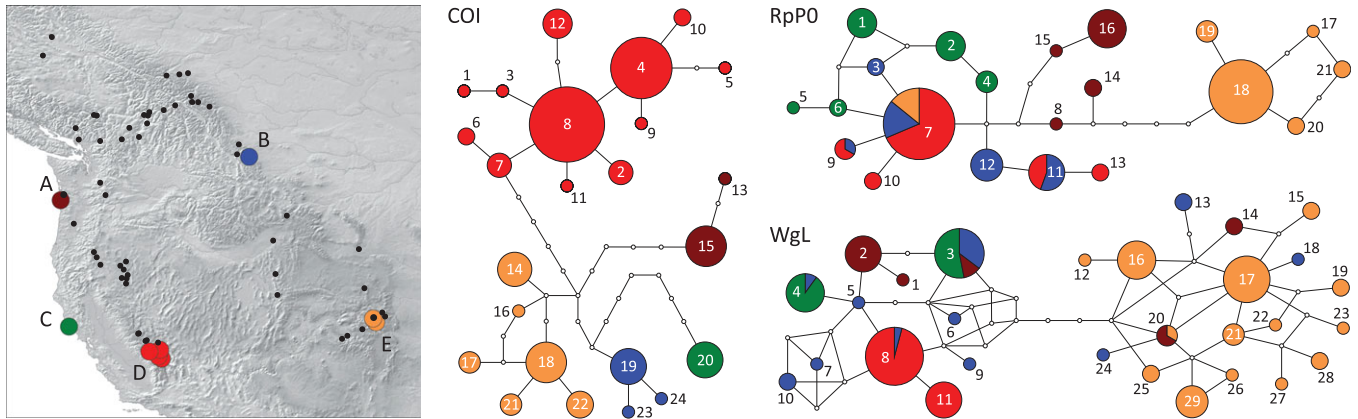
*n* refers to the number of sequenced individuals at each locus. For the two nuclear genes (RpP0 and WgL), there are then *2n* sequences. The population letters A, B, C, D, and E refer to the five geographic regions displayed in Figure 1.

statistics estimating population and/or phylogeographic structure: pairwise  $\Phi_{ST}$ 's (Excoffier et al. 1992) and AMOVA  $\Phi$ -statistics ( $\Phi_{SC}$ ,  $\Phi_{ST}$ , and  $\Phi_{CT}$ , Excoffier et al. 1992; hierarchical analysis) among the same five regions (using *p*-distance as estimate of the number of mutations separating two sequences), the average ratio between the  $\Phi_{ST}$ 's and geographical distances calculated for each pair of regions ( $m\Phi_{STdgeo}$ ), global  $G_{ST}$  (Pons and Petit 1995), and  $N_{ST}$  (Pons and Petit 1996) over all regions. We assessed the statistical significance of all  $G_{ST}$ ,  $N_{ST}$ , and  $\Phi$ -statistic values by recalculating them on 10,000 random permutations of original datasets (permutations of individuals among all populations for  $G_{ST}$ ,  $N_{ST}$ , and  $\Phi_{ST}$ , among populations within groups for  $\Phi_{SC}$  or among groups for  $\Phi_{CT}$ ). The difference between  $N_{ST}$  and  $G_{ST}$ , which highlights the extent of the phylogeographic signal, was tested by recalculating differences on 10,000 random permutations of haplotypes in the original datasets.

### INFERRING THE PAST AND PRESENT SPECIES DISTRIBUTIONS

The current and LGM distributions of *C. aeneicollis* were inferred using the maximum entropy method implemented in MAXENT 3.3.3 (Phillips et al. 2006; Phillips and Dudík 2008). The model

implemented in MAXENT minimizes relative entropy between two probability densities: one estimated from species occurrence data and one estimated from the landscape (Elith et al. 2011). The present distribution was estimated first and the result used to project the species distribution on past climate layers using two distinct LGM models: (i) the CCSM (community climate system model) and (ii) the MIROC (model for interdisciplinary research on climate). These inferences were based on 71 localities and 12 bioclimatic variables. The *C. aeneicollis* localities included 10 sampled populations for this study (Table 1), six additional localities for which we recorded presence of this species within the last 20 years, and 70 historical localities recorded by Brown (1956). Nineteen bioclimatic variables (Bio1–Bio19) at a 2.5 arcmin (~5 km) resolution were initially extracted from the WorldClim database (WorldClim 1.4, Hijmans et al. 2005) for the current time (~1950–2000) and LGM models (CCSM and MIROC). Available bioclimatic variables on WorldClim for the CCSM and MIROC are provided by the PMIP2 database (Paleoclimate Modeling Intercomparison Project Phase II, Braconnot et al. 2007). Relationships among bioclimatic variables were evaluated using Pearson correlation coefficients. To avoid collinear variables (Pearson coefficient > 0.9), seven variables were discarded. The resulting



**Figure 1.** Observed localities for *Chrysomela aeneicollis* and median-joining networks for the three sequenced gene fragments (COI, RpP0, and WgL). Small black dots on the map refer to observations reported by Brown (1956); large colored dots show localities where individuals were collected for sequencing: Oregon (A, brown), Montana (B, blue), coastal California (C, green), Sierra Nevada (D, red), and Colorado Rockies (E, orange). The same colors are used to show the geographic distribution of alleles. Alleles are represented by circles; circle sizes are proportional to allele frequency, and alleles are identified by a unique number (Table 1). Each line in the network represents a single mutational change. Small white circles indicate intermediate alleles not included in our dataset that are necessary to link all observed alleles to the network.

set of 12 variables included annual mean temperature (Bio1), mean diurnal temperature range (Bio2), isothermality (Bio3), maximum temperature of the warmest month (Bio5), minimum temperature of the coldest month (Bio6), annual temperature range (Bio7), mean temperature of the wettest quarter (Bio8), mean temperature of the driest quarter (Bio9), annual precipitation (Bio12), precipitation of the driest month (Bio14), precipitation seasonality (Bio15), and precipitation of the warmest quarter (Bio18). Ten replicates were performed for each analysis, from which we derived an average distribution. We used the default convergence threshold ( $10^{-5}$ ), 5000 iterations and a “random seed” to generate a random partition of our localities into training (90%) and test (10%) localities. These random partitions were used to test the model. The inferences were evaluated using the area under the ROC (receiver operating characteristic) curve. These AUC (area under the curve) values are commonly used to assess MAXENT estimation performance (cf. Marske et al. 2009, 2011).

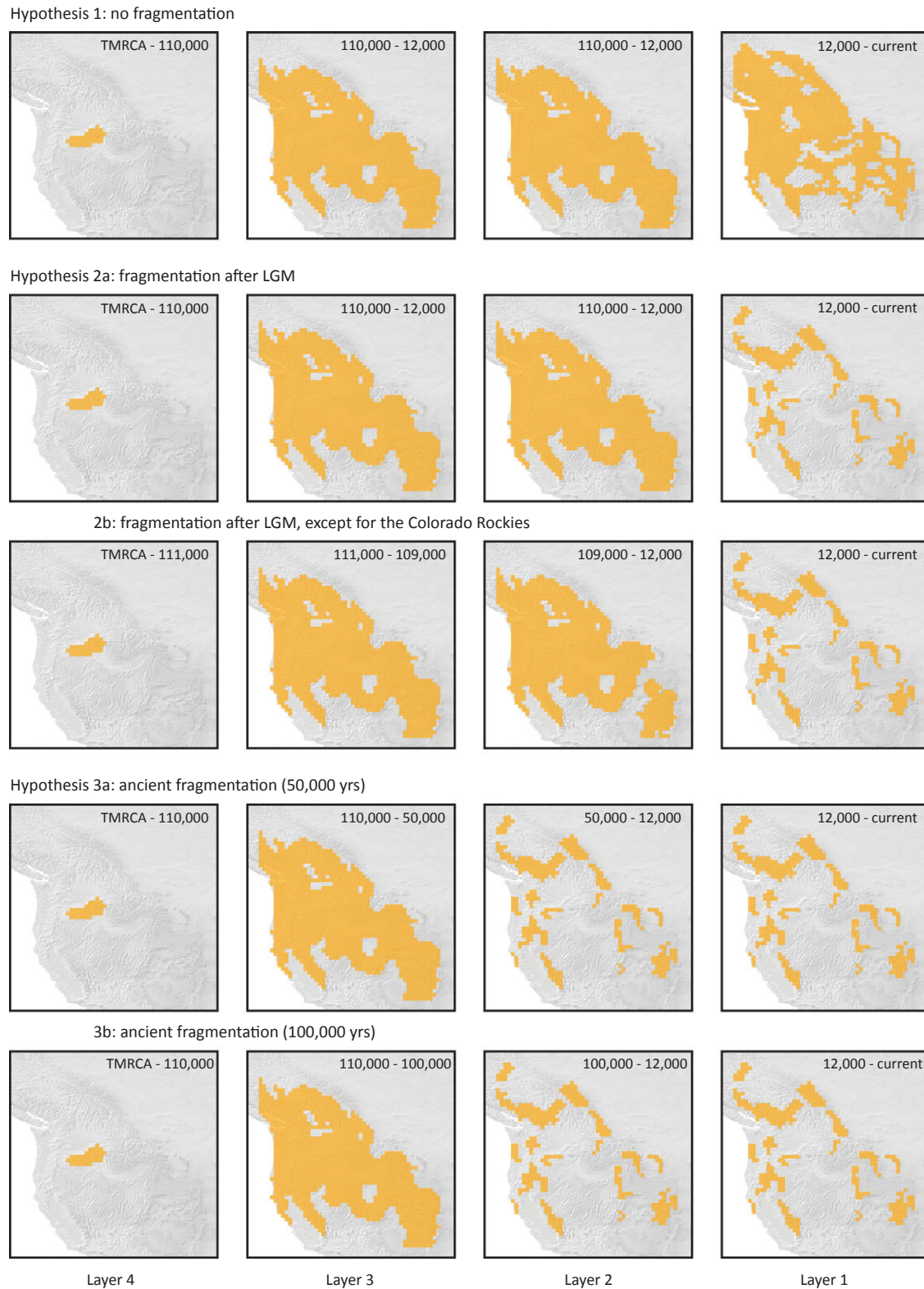
### CHARACTERIZING POPULATION CONNECTIVITY

As a first exploratory analysis of population connectivity, we inferred migration rates among regions from DNA sequence variation using the Bayesian method in LAMARC 2.1.8 (Kuhner 2006) that relies on a Markov chain Monte Carlo (MCMC) algorithm. This method assumes a partially stationary demographic model (i.e., population structure and migration rates are constant, although population size can vary over time), and considers each separate region as a panmictic entity. Although this model is clearly unrealistic for our purpose of inferring history of geographic range modifications in *C. aeneicollis*, we used the results as a first assessment of level of connectivity among regions. For

this analysis, relative effective size was set to 1 for the mitochondrial gene and to 4 for the two nuclear genes. We ran 10 initial chains of 10,000 steps and two final chains of 100,000 steps. Both types of chains were sampled every 100 steps. The following parameters were estimated:  $\theta$  (i.e.,  $4\mu N_e$ , with  $\mu$  the mutation rate per site per generation and  $N_e$  the population effective size), the exponential growth rate  $g$  and the immigration rate  $M$  ( $M = 4mN_e/\theta$  for diploid individuals).

### COMPARING HYPOTHESES OF POPULATION CONNECTIVITY THROUGH COALESCENCE SIMULATIONS

Estimates of present and past distributions of *C. aeneicollis* inferred as described above were partially ambiguous and varied according to LGM model used (CCSM or MIROC). The uncertainty associated to these estimates mainly regarded levels of population connectivity across the range. To take this uncertainty into account, we considered alternative hypotheses about how population connectivity has changed over time and evaluated the possibility that each hypothesis generated observed sequence data. More specifically, we designed three distinct scenarios that vary in development of range fragmentation (Fig. 2): (1) all regions have been connected to each other for a long time (hypothesis of no fragmentation); (2) the current range is highly fragmented since the end of the last glaciation approximately 10,000 years ago (hypothesis of post-LGM fragmentation); (3) the range has been highly fragmented since long before the last glaciation episode (hypothesis of pre-LGM fragmentation). In addition, two possibilities are considered for hypotheses 2 and 3: (2) the south east portion of the range, corresponding to the southern Rockies, were isolated (or not) from the rest of the distribution before the LGM;



**Figure 2.** Three alternative evolutionary hypotheses simulated with PHYLOGEOSIM 1.0. Hypotheses are ordered by length of time since populations became fragmented (1 = no fragmentation; 2 = post-LGM fragmentation, with two variants: southeast portion of the range isolated or not before the LGM; 3 = pre-LGM fragmentation, with two variants: fragmented since 50,000 years or 100,000 years). Each shows chronological layers (the oldest on the left) with potential presence of *C. aeneicollis* for each cell on the grid. Colored cells are thus those accessible to gene copies during the simulation. The time interval (in past years) corresponding to each layer is indicated, from the present, up to time to most recent common ancestor (TMRCA).

(3) two different fragmentation times (50,000 and 100,000 years ago) are considered, both occurring long before the LGM. To compare these hypotheses, we translated them into five models of coalescence integrated in an explicit geographic framework. This was done by superimposing a two-dimensional grid on a map of the study region (Fig. 1), and by identifying grid cells ( $\pm 35$  km wide) in which the species is potentially present.

In the models, a grid cell that is potentially occupied is considered a distinct population, which can exchange migrants with neighbor populations (i.e., cells) and in which coalescence can occur. Four separate grids define each scenario, each representing the potential species' range at a different time period (Fig. 2). Because *C. aeneicollis* is a univoltine species, one year corresponds to one generation. In all models tested, the maximum effective size of cells is held constant over space and time throughout one simulation. This means that global effective size of the species, or of a specific region, is directly proportional to its number of accessible cells. The speed at which all accessible cells are colonized will depend on migration rate among adjacent cells and on reproduction rate. Although this is a strong assumption, allowing maximum cell effective size to vary across time and/or space within a simulation would have unnecessarily increased the number of inferred parameters to an unreasonably large value. As a consequence, an increase in number of accessible cells is automatically associated with a size increase and range expansion in our models (Fig. 2), and a decrease in cell number is associated with a size decrease and range contraction.

Coalescence simulations along the five hypotheses were conducted for the purpose of detecting those that are compatible with our observed sequence data. Prior to a coalescence simulation, the simulation program used (PHYLOGEOSIM) performed a preliminary forward simulation to generate parameter values needed for the backward simulations, that is, backward migration rates and actual (as opposed to maximum) effective population sizes defined for each generation. This strategy is similar to the method developed by Currat et al. (2004) (see also Ray et al. 2010). Details about coalescence simulations (including model design) and the comparison of simulated and observed sequence data through calculation of a series of summary statistics are given in Appendix S1. A global overview of our procedure is also shown in Figure 3

## Results

### CURRENT AND PAST GEOGRAPHIC DISTRIBUTION OF *C. AENEICOLLIS*

The current and LGM (CCSM and MIROC) distributions estimated with MAXENT (Fig. 4) were associated with high AUC values (all values  $>0.918$ , Table S1), indicating that all models

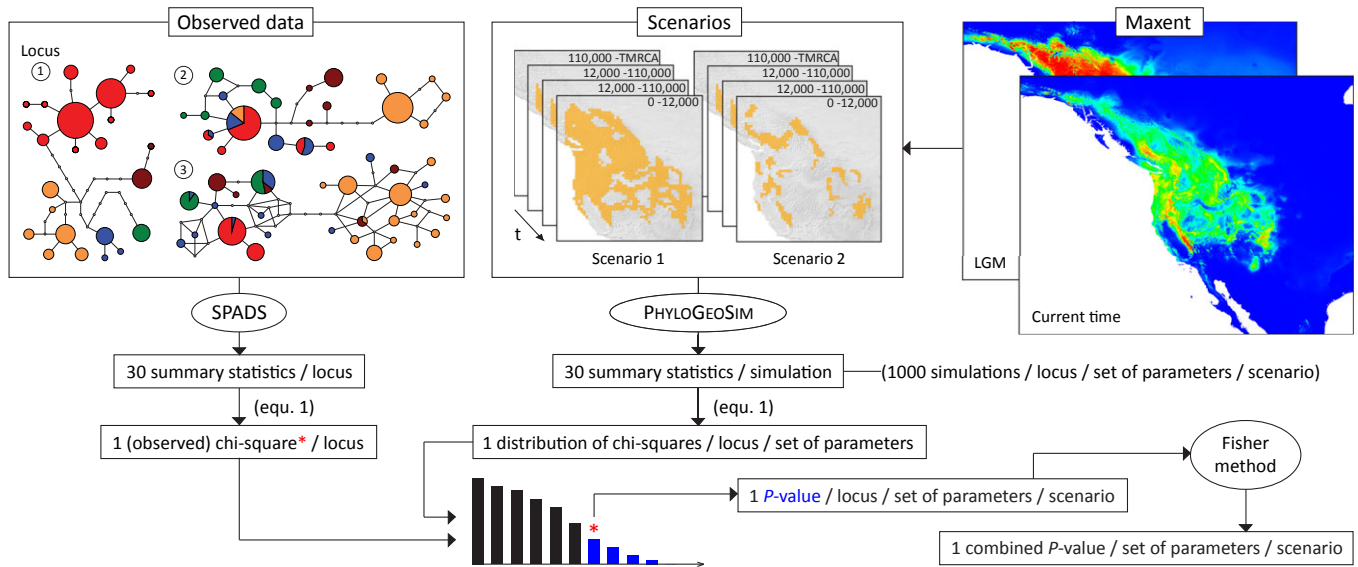
performed significantly better than random predictions (Phillips et al. 2006). It is not surprising that probability of occurrence of the leaf beetle *C. aeneicollis* was greater in many regions of Western North America during the LGM because climate conditions were more favorable at the time (Fig. 4). Yet estimated distributions predict that the species should currently possess a widespread range, as evidenced by the number of cells associated with a probability of occurrence of at least 50% found over a large geographic area. This suggests high levels of gene flow (connectivity) among regions that are characterized by a high probability of occurrence ( $>75\%$ ). Nevertheless, *C. aeneicollis* is not known to occur in many portions of this range, and there is no suitable habitat in many localities. The LGM-estimated distributions suggest that connectivity among regions was even stronger in the past than at present, but they also suggest the presence of an isolated area of suitable habitat in the southeast portion of the range. This area is located further to the southeast in the CCSM reconstruction than the MIROC reconstruction. We have sequenced samples in the southeastern part of Rocky mountains in the vicinity of this southeast habitat shown in the MIROC LGM reconstruction, which were used to test the hypothesis of presence of a historically isolated population in this area (hypothesis 2b in Fig. 2).

### DNA SEQUENCE DATA

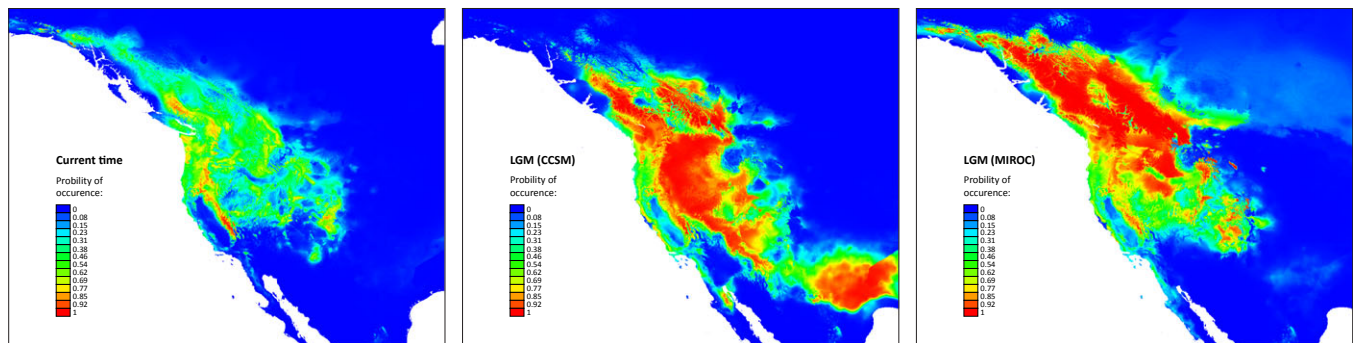
After alignment and pruning, the complete RpP0, WgL, and COI datasets contained 151, 128, and 136 sequences of 1026, 496, and 475 nucleotides, respectively. We found 43 homozygotes and 34 heterozygotes for the RpP0 fragment, and 47 homozygotes against 21 heterozygotes for the WgL locus. In total, there were 24 alleles and 36 polymorphic sites in the COI dataset, 21 alleles and 18 polymorphic sites (plus one 4 bp long gap) in the RpP0 alignment, and 29 alleles and 23 polymorphic sites in the WgL alignment (Table 1). Allele sequences for all loci are available from GenBank under accession numbers KF992624-KF992696.

### GENETIC DIVERSITY AND STRUCTURE ANALYSES

Allele networks are shown in Figure 1, along with the geographic distribution of alleles among the five regions defined a priori. Visual inspection of this figure indicates a strong population structure and robust phylogeographic signal among regions that is most pronounced in the mitochondrial COI dataset. This is confirmed by diversity and structure indices (Table 2). Indeed, AMOVA  $\Phi$ -statistics evaluating partitioning among regions are high ( $>0.6$ ) and statistically significant for all three gene fragments. A strong and statistically significant phylogeographic signal, as measured by the difference between  $N_{ST}$  and  $G_{ST}$  statistics (Pons and Petit 1996) was measured for all three loci. Overall, these results seem to imply strong current fragmentation of the species range.



**Figure 3.** Schematic representation of steps involved in inferring species history from DNA sequence variation and distribution data. “equ. 1” refers to the chi-square computation described in the text, and SPADS, PHYLOGEOSIM, and MAXENT are three computer programs used for analyses (see text and Supporting Information S1 for further details).



**Figure 4.** Inferred distributions of *Chrysomela aeneicollis* in North America obtained with MAXENT and based on current climatic data (current times) or reconstructed last glacial maximum (LGM) palaeoclimatic data (“CCSM” and “MIROC”). Probabilities of occurrence are shown by color. Current datapoint locations on which the MAXENT analysis was based are those shown in Figure 1. The standard deviation associated with inferred distributions is displayed in Figure S1.

### PRELIMINARY GENE FLOW ESTIMATES

Results obtained with the LAMARC software are summarized in Tables S2 and S3. Overall, this analysis calculated low effective numbers of migrants ( $N_e m$ ) among regions per generation, with values well below 1 (i.e., impact of genetic drift on population differentiation is stronger than impact of migration). It reveals strong differentiation among regions, confirming the level of phylogeographic structure suggested by allele networks. Because the model used to infer these numbers of migrants is stationary, these low estimates could be considered as maximum values. Indeed, the same level of population differentiation could also be explained by absence of recent gene flow among regions combined with a past unfragmented species range.

### COMPARISON OF HISTORICAL HYPOTHESES THROUGH STATISTICAL MODELING

In all forward simulations, all cells accessible to individuals during initial range expansion (from 110,000 to 50,000 or 12,000 generations) were colonized before subsequent range fragmentation (Fig. 3). The expansion reached maximum effective size for all accessible cells 60–4000 generations after it began (depending on migration and reproduction rates). The combined  $P$ -values across loci for each combination of parameter values for all hypotheses (assuming a reproduction rate  $t_R = 2$ ) are summarized in Table 3. These  $P$ -values provide a relative estimate of how compatible a model is to observed data. Very similar  $P$ -values were obtained when assuming reproduction rates  $t_R$  of 5 and 10 (Tables S4 and



**Table 2.** Summary statistics computed for each of the three loci (COI, RpP0, and WgL).

Locus	AMOVA										Nucleotide diversity					Relative nucleotide diversity				
	$N_{ST}$		$N_{ST}-G_{ST}$		$m\Phi_{STdgeo}$		$\Phi_{SC}$	$\Phi_{ST}$	$\Phi_{CT}$	Ore.	Mon.	Cal.	S.N.	Col.	Ore.	Mon.	Cal.	S.N.	Col.	
	$G_{ST}$																			
COI	0.613 <sup>1</sup>	0.901 <sup>1</sup>	0.288 <sup>1</sup>	0.076 <sup>1</sup>	0.402 <sup>1</sup>	0.878 <sup>1</sup>	0.795 <sup>1</sup>	0.0003	0.0004	0.0000	0.0013	0.0021	0.06	0.07	0.00	0.22	0.48			
RpP0	0.387 <sup>1</sup>	0.646 <sup>1</sup>	0.260 <sup>1</sup>	0.024 <sup>1</sup>	-0.064	0.690 <sup>1</sup>	0.708 <sup>1</sup>	0.0034	0.0026	0.0039	0.0013	0.0046	0.40	0.28	0.47	0.13	0.99			
WgL	0.294 <sup>1</sup>	0.531 <sup>1</sup>	0.236 <sup>1</sup>	0.010 <sup>1</sup>	-0.094	0.614 <sup>1</sup>	0.647 <sup>1</sup>	0.0090	0.0111	0.0045	0.0010	0.0053	0.77	0.95	0.38	0.08	0.71			

AMOVA  $\Phi$ -statistics are estimated for the partition into the five defined groups of populations: Oregon (Ore.), Montana (Mon.), California (Cal.), Sierra Nevada (S.N.), and Colorado (Col.).

<sup>1</sup>Significant ( $P$ -value < 0.05)  $\Phi$ -statistic value.

S5); hence, varying reproduction rate does not seem to influence our comparison of hypotheses. Across all five hypotheses, one combination of simulation parameter values (migration rate  $fm$  of  $10^{-4}$  and effective population sizes  $N_{e(mt)}$  of 2500 and  $N_{e(nuc)}$  of 10,000) appears the most compatible with observed data, leading to  $P$ -values > 0.05 for pre- (3a + 3b) and post-LGM fragmentation (2a + 2b) hypotheses. Only one other combination ( $fm$  of  $10^{-3}$ ,  $N_{e(mt)} = 2500$  and  $N_{e(nuc)} = 10,000$ ) leads to a  $P$ -value > 0.05 for any single hypothesis (3b, pre-LGM fragmentation). The no fragmentation hypothesis (1) is always rejected ( $P$ -value < 0.05), and most probable scenarios are associated with the pre-LGM fragmentation hypothesis (3a and 3b). Indeed, for the best combination of parameter values, pre-LGM fragmentation hypotheses are four to five times more likely than post-LGM fragmentation hypotheses. Finally, applying a Benjamini–Hochberg correction (Benjamini and Hochberg 1995) for multiple testing slightly increased  $P$ -values associated with a few initially rejected hypotheses to just above 0.05 (Tables 3, S4, and S5), but it did not change the ranking of hypotheses best compatible with observed data. In sum, the hypothesis of no connection among regions since at least 50 thousand years is strongly favored by DNA sequence variation.

## Discussion

In this study, we integrate data on intraspecific DNA sequence variation with the contemporary distribution of *C. aeneicollis* to infer past and present gene flow within this species' geographic range. Five hypotheses derived from past and present estimates of geographic distribution were compared to genetic variation data by simulating DNA sequence evolution in a two-dimensional space approximated by a grid, in which each cell is considered a different population. The explicitly geographic model of coalescence used here appears to offer a more appropriate way to model population evolution for the purpose of estimating past connectivity among regions. Indeed, species (or subsets of species ranges) characterized by a continuous distribution are not naturally subdivided into populations, and delimiting populations in that context may rely on arbitrary decisions regarding the placement of population boundaries.

Describing the species range as a two-dimensional grid on which each cell is reported as accessible or not to individuals allows development of an objective model of population evolution. This was clearly illustrated here by integrating ecological niche-modeling data to design evolutionary hypotheses. The process of overlaying the grid on the map of the species distribution is simple enough that it can be easily applied to other phylogeographic studies. Moreover, this geographically explicit model also offers a natural way to describe movement of individuals across the species range. Although in classic coalescence models, a population is usually considered as a panmictic entity, within which

**Table 3.** *P*-values obtained from the comparison between real and simulated datasets with 1000 simulations per set of parameters and a reproduction rate  $t_R = 2$ .

Scenario	Simulation parameters		Combined <i>P</i> -value	Locus by locus <i>P</i> -values		
	<i>fm</i>	$N_e$ (mitoch./nuclear)		COI	RpP0	WgL
S1	0.00001	250/1000	<0.001	0.003	0.003	0.002
S2a	0.00001	250/1000	<0.001	0.003	<0.001	0.001
S2b	0.00001	250/1000	<0.001	0.005	<0.001	<0.001
S3a	0.00001	250/1000	<0.001	0.005	<0.001	0.001
S3b	0.00001	250/1000	<0.001	0.005	<0.001	<0.001
S1	0.0001	250/1000	0.002	0.017	0.263	0.008
S2a	0.0001	250/1000	<0.001	0.015	0.032	0.002
S2b	0.0001	250/1000	<0.001	0.041	0.031	0.003
S3a	0.0001	250/1000	<0.001	0.048	0.004	0.001
S3b	0.0001	250/1000	<0.001	0.045	0.003	<0.001
S1	0.001	250/1000	<0.001	<0.001	0.001	<0.001
S2a	0.001	250/1000	<0.001	0.009	0.023	0.005
S2b	0.001	250/1000	0.002	0.068	0.062	0.007
S3a	0.001	250/1000	<0.001	0.038	0.006	<0.001
S3b	0.001	250/1000	<0.001	0.011	0.003	0.001
S1	0.000001	2500/10,000	<0.001	0.002	<0.001	<0.001
S2a	0.000001	2500/10,000	<0.001	<0.001	<0.001	<0.001
S2b	0.000001	2500/10,000	<0.001	0.001	0.005	<0.001
S3a	0.000001	2500/10,000	<0.001	0.001	0.001	<0.001
S3b	0.000001	2500/10,000	<0.001	0.002	0.002	<0.001
S1	0.00001	2500/10,000	0.014*	0.164	0.091	0.024
S2a	0.00001	2500/10,000	0.007	0.086	0.078	0.021
S2b	0.00001	2500/10,000	0.020*	0.275	0.090	0.022
S3a	0.00001	2500/10,000	0.015*	0.301	0.068	0.018
S3b	0.00001	2500/10,000	0.010	0.231	0.055	0.018
S1	0.0001	2500/10,000	0.014*	0.005	0.188	0.376
S2a	0.0001	2500/10,000	0.057*	0.017	0.316	0.413
S2b	0.0001	2500/10,000	0.054*	0.013	0.369	0.426
S3a	0.0001	2500/10,000	0.255*	0.062	0.743	0.446
S3b	0.0001	2500/10,000	0.220*	0.078	0.866	0.238
S1	0.001	2500/10,000	<0.001	<0.001	<0.001	<0.001
S2a	0.001	2500/10,000	<0.001	<0.001	<0.001	<0.001
S2b	0.001	2500/10,000	<0.001	<0.001	<0.001	<0.001
S3a	0.001	2500/10,000	<0.001	0.004	0.005	0.018
S3b	0.001	2500/10,000	0.061*	0.019	0.379	0.335
S1	0.000001	25,000/100,000	0.001	0.056	0.027	0.008
S2a	0.000001	25,000/100,000	<0.001	0.031	0.013	0.004
S2b	0.000001	25,000/100,000	<0.001	0.044	0.013	0.004
S3a	0.000001	25,000/100,000	<0.001	0.044	0.003	0.005
S3b	0.000001	25,000/100,000	<0.001	0.019	0.006	0.007
S1	0.00001	25,000/100,000	0.031*	0.093	0.218	0.048
S2a	0.00001	25,000/100,000	0.041*	0.109	0.259	0.050
S2b	0.00001	25,000/100,000	0.036*	0.102	0.238	0.049
S3a	0.00001	25,000/100,000	0.024*	0.097	0.170	0.043
S3b	0.00001	25,000/100,000	0.022*	0.117	0.152	0.034
S1	0.0001	25,000/100,000	<0.001	0.002	0.003	0.019
S2a	0.0001	25,000/100,000	<0.001	0.001	0.011	0.016
S2b	0.0001	25,000/100,000	<0.001	0.001	0.006	0.015

Continued

**Table 4.** Continued.

Scenario	Simulation parameters			Locus by locus <i>P</i> -values		
	<i>fm</i>	<i>N<sub>e</sub></i> (mitoch./nuclear)	Combined <i>P</i> -value	COI	RpP0	WgL
<b>S3a</b>	<b>0.0001</b>	<b>25,000/100,000</b>	<i>&lt;0.001</i>	0.001	0.029	0.021
<b>S3b</b>	<b>0.0001</b>	<b>25,000/100,000</b>	<i>0.007</i>	0.011	0.289	0.045
<b>S1</b>	<b>0.001</b>	<b>25,000/100,000</b>	<i>&lt;0.001</i>	<0.001	<0.001	<0.001
<b>S2a</b>	<b>0.001</b>	<b>25,000/100,000</b>	<i>&lt;0.001</i>	<0.001	<0.001	<0.001
<b>S2b</b>	<b>0.001</b>	<b>25,000/100,000</b>	<i>&lt;0.001</i>	<0.001	<0.001	<0.001
<b>S3a</b>	<b>0.001</b>	<b>25,000/100,000</b>	<i>&lt;0.001</i>	<0.001	<0.001	<0.001
<b>S3b</b>	<b>0.001</b>	<b>25,000/100,000</b>	<i>&lt;0.001</i>	0.001	0.003	0.006

*fm* = forward migration rate and *N<sub>e</sub>* = the maximal effective population size. Combined *P*-values in italic are <0.05, thus corresponding to significantly rejected models, combined *P*-value in bold are ≥0.05, and an asterisk indicates a *P*-value that is ≥0.05 after applying a Benjamini–Hochberg correction for multiple testing. S1 = no fragmentation; S2 = fragmentation after LGM, including (a) or not (b) the Colorado region; S3, ancient fragmentation, that started (a) 50,000 or (b) 100,000 years ago.

each individual has an equal probability to mate with any other, a population defined as a series of contiguous cells on the grid may be characterized by some level of isolation by distance, with the set of possible intrapopulation mating pairs restricted by movement of individuals on the grid. Using an explicitly geographic model of coalescence offers a natural way to implement isolation by distance, which may be particularly important when dealing with populations spanning large areas, or that are characterized by different shapes. For example, a population defined by a line of nine contiguous cells on the grid will be characterized by a different structure than one spanning a square of 3 by 3 cells, although both span an identical surface.

Additionally, this approach allows all possible migration routes among populations to be taken into account for all gene copies, instead of combining migration rates for multiple possible routes between two populations into a single migration rate. On the other hand, one disadvantage of making the model geographically explicit is that it significantly slows down the computation. For the purpose of comparing evolutionary hypotheses or inferring demographic parameters, a large number of simulations are usually required, because it is important to explore as thoroughly as possible the range of all combinations of parameter values. For example, performing an approximate Bayesian computation (ABC) analysis (e.g., Beaumont et al. 2002; Bertorelle et al. 2010) with this model would have been impossible with currently available computing power. An ABC analysis would have done a better job of exploring the space of all combinations of parameter values. However, we are confident that our choice to choose typical combinations of parameter values, homogeneously distributed across their range, was sufficient to identify the portion of the space (and therefore the hypotheses) associated with higher probabilities. This is because the number of parameters used in our model is relatively low (three parameters: cell effective size, migration

rate, reproduction rate), keeping reasonable the size of the space of combination of parameter values to explore.

The estimates of past and present geographic distributions of *C. aeneicollis* show us that climatic conditions were substantially more favorable during the LGM, which suggests its range was more widespread, or at least less fragmented, in the past. However, the model predicts that populations should currently be present in regions of low elevation that separate the mountain ranges in which the beetle is mostly found today. Indeed, the projected current range covers a similar area as the projected past distribution, but associates a lower probability of occurrence with lower elevations. We are left with two possibilities regarding the current distribution of the species: one of strict fragmentation of the range with no individual present in the intermediate cells, which thereby prevents any gene flow among regions, or one that allows some reduced level of connectivity among regions because it assumes that small populations with a few potential migrants are present. Focusing on the estimated LGM distribution (Fig. 4), it appears that an area located to the southeast portion of the range (CCSM model), characterized by a high probability of presence (>0.9) could have been more or less isolated from the main portion of the distribution (also characterized by a probability > 0.9). Once again, however, this southeast area appears connected to the main portion of the species distribution by an area characterized by a lower probability of occurrence (±0.5). This lower probability could also be interpreted by the absence of beetles at the time, which would mean that this area was isolated from the rest of the species range even during the LGM.

To reduce ambiguity associated with estimates of past and present species ranges, we developed alternative models of population evolution using known beetle localities (for the current range) and probabilities of occurrence inferred by niche modeling (for present and past distributions). These models represent

evolutionary hypotheses that differ mainly by population connectivity and precisely predict changes in geographic distribution over time. They were compared with multilocus sequence data by generating simulated sequences with similar characteristics. The analysis identifies hypotheses implementing the longest period of range fragmentation, from today to well beyond the LGM (going backward in time), as most compatible with observed data. The results clearly rejected the scenario that assumes current connectivity among regions (no fragmentation hypothesis), even when assuming extremely limited gene flow. The intermediate hypothesis of current fragmentation that started only after the LGM is not completely rejected ( $P$ -value = 0.057), but is five times less probable than two hypotheses assuming a much more extended period of range fragmentation. Thus, sequence data suggest that sampled regions have been isolated for a long time, which is also corroborated by observed high population structure and strong phylogeographic signal. Although two variants of the pre-LGM fragmentation hypothesis associated with fragmentation times of 50,000 and 100,000 years are equally supported by the data, it is worth noting that these results seem compatible with a range fragmentation scenario that goes back to the end of the penultimate glaciation. This event, the Tahoe glaciation, occurred in Western North America 135–160 kya (Phillips et al. 2009), and was apparently more severe than the latest glaciation. If true, this strong older glacial period could have reduced the range of *C. aeneicollis* to one or a few small refuges, from which the species would have recolonized Western North America as climate warmed. This range contraction would explain the necessity of including a reduced ancestral range in our coalescence model at around that time. The species range would have then become fragmented relatively quickly, and would have remained as such until today. Note that this possibility is somewhat in contradiction with our prior expectation that cold-adapted species should display a less fragmented range during glacial periods (see Introduction). Although it may still hold true for the last glaciation episode, the more severe Tahoe glaciation might have been associated at some point with inhospitable climatic conditions and possibly a range contraction, even for the cold-adapted *C. aeneicollis*.

Interestingly, simulations of DNA sequence evolution contradict historical models based on estimates of past geographic distribution of the species. Indeed, the LGM distribution inferred by MAXENT indicates a high probability of occurrence for a large portion of the range, and appears to show that regions that are presently isolated were connected at that time. However, the sequence data analysis suggests that connectivity among regions occurred much further in the past. Additionally, the estimated time for connectivity to occur did not depend on choice of a substitution rate for sampled DNA fragments, because all simulations were performed by imposing the same number of substitutions as found in observed data. Instead, it is solely based on accessi-

bility of grid cells to individuals over time and the coalescence process.

One possible explanation for the contradiction between sequence data and niche-modeling estimates is the fact that inferred probabilities of occurrence only indicate potential presence of a species in an area, rather than its actual presence. This may be particularly true for leaf beetles, which are characterized by low dispersal and may therefore not have time or opportunity to adjust their ranges in response to a climatic event that increases availability of suitable habitat. Local distribution patterns of willow beetles support this assertion (e.g., Dahlhoff et al. 2008). Also, all environmental parameters important for this species may not have been taken into account for distribution estimates, such as, presence or absence of host plants or soil moisture. In addition, the resolution of bioclimatic data available for estimation of the past and present range may be insufficient for modeling real world conditions. This is especially true for ranges that extend into montane habitats, where steep changes in elevation (and thus suitable habitat) occur over relatively short spatial scales, resulting in microclimates that are challenging to model. Finally, the geographic range of the species that we introduced in hypotheses 1 and 2 (no fragmentation or post-LGM fragmentation) for the last ice age is based on estimated range at the LGM. However, climate conditions during the LGM may not be representative of those from the entire glacial period, and it is possible that a more fragmented distribution, more similar to what is observed today, characterized a large portion of the last ice age. This would explain why pre-LGM fragmentation hypotheses were favored by our analyses. Whatever the cause, our genetic data clearly challenge MAXENT distribution estimates inferred for this study and illustrate in general that estimates of potential range of a species based on ecological modeling cannot be automatically translated into its actual range.

Comparing phylogeographic data from several species that share habitat requirements or distributions may allow for more general inferences about influence of life-history traits or ecological factors on phylogeographic structure (Avice 2000). In the case of *C. aeneicollis*, a useful comparison can be made with *C. lapponica*, which has similar habitat requirements and life-history strategies, but occurs in Eurasia. This species is adapted to cold temperatures and is currently restricted to fragmented populations at high altitudes or latitudes. DNA sequence variation for two of the gene fragments sequenced in the current study were included in a phylogeographic study of *C. lapponica* (Mardulyn et al. 2011). A visual inspection of allele networks inferred in both species highlights common features: strong phylogeographic structure is visible among regions (i.e., mountain ranges), suggesting a long period of isolation among them. At the same time, lineage sorting among regions was not achieved for either species, with the exception of the COI gene fragment in

*C. aeneicollis*. These observations suggest that the two species have gone through a similar history of range fragmentation, with regions that are currently isolated and have been separated since at least the LGM. Our interpretation of this similar pattern is that both species, like many specialized leaf beetles, are characterized by limited dispersal. This suggests that the phylogeographic pattern observed is mainly dependent on long-term history rather than contemporary migration.

Despite a much larger geographic range, *C. lapponica* is characterized by lower values of the parameter  $N_{ST}-G_{ST}$  (which estimates the level of phylogeographic structure) than *C. aeneicollis*. This may suggest a longer history of fragmentation for *C. aeneicollis* than for *C. lapponica*. Most mountain ranges in Western Europe (Pyrenees, Alps, Carpathians) run east–west, whereas those in North America (including the Sierra Nevada, Cascade Range, and the Rockies) run north–south. Thus, while mountain ranges on both continents probably provided opportunity for species to migrate to adjust to climate changes, dispersal among adjacent mountain ranges in Europe was probably easier during glacial episodes than between the Sierra Nevada and Rockies in North America (Hewitt 2003). Further, at the height of the last ice age, large distances and pluvial lakes separated North American ranges. Evidence for strong isolation among mountain ranges in Western North America at some points during the Holocene and Pleistocene has been suggested for other organisms (Knowles 2000; Brunfeld et al. 2001). For example, a recent phylogeographic study of mitochondrial sequence variation of the American pika (*Ochotona princeps*), a small lagomorph whose distribution is restricted to mountain ranges of the west, depicted a similarly strong pattern of genetic differentiation among mountains, and thus could have been subject to a similar fragmentation history (Galbreath et al. 2009).

Overall, this study illustrates how a spatially explicit model of coalescence can be used to easily integrate niche-modeling estimates into a phylogeographic study. It is useful to rigorously compare more realistic historical hypotheses that incorporate a detailed description of the geographic range of a species and its evolution over time. It has improved our understanding of the evolution of the range of cold-adapted organisms in Western North America, under the assumption that it was influenced by climate changes of the Pleistocene, by showing that current fragmentation of the range of *C. aeneicollis* persisted, going backward in time, well beyond the LGM. It follows that the range of other cold-adapted species could have been strongly fragmented in this region during at least a portion of the last glacial period, despite the supposedly more favorable conditions prevailing at the time.

#### ACKNOWLEDGMENTS

We are grateful to M. Quinzin, J. Lundblad, and D. McMillan for help collecting insects, and to J. Smiley, the associate edi-

tor G. Mayer, and three anonymous reviewers for helpful comments and suggestions. Computational resources have been provided by the High Performance Computing Centre co-funded by ULB and VUB (HPC cluster “Hydra”) and by the *Consortium des Équipements de Calcul Intensif* (CÉCI, HPC cluster “HMEM”), funded by the *Fonds de la Recherche Scientifique* (F.R.S.-FNRS) under convention 2.5020.11. R. Leplae from the ULB/VUB Computing Centre, as well as D. François and D. Collignon from the CÉCI, provided invaluable technical assistance for the use of these computer resources during the course of this study. We also wish to thank L. Grumiau for his technical assistance in the laboratory and I. Parmentier for her useful advices on inferring our species distributions. This research project was funded by the Belgian F.R.S.-FNRS through a research grant to PM and a PhD grant to SD, and by an American NSF grant to EPD and NER (DEB-0844404/06). PM is research associate at the F.R.S.-FNRS.

#### DATA ARCHIVING

DNA sequence data available from GenBank (accession numbers KF992624–KF992696).

#### LITERATURE CITED

- Avice, J. C. 2000. *Phylogeography: the history and formation of species*. Harvard Univ. Press, Cambridge, MA.
- Bandelt, H. J., P. Forster, and A. Röhl. 1999. Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* 16:37–48.
- Beaumont, M. A., W. Zhang, and D. J. Balding. 2002. Approximate Bayesian computation in population genetics. *Genetics* 162:2025–2035.
- Benjamini, Y., and Y. Hochberg. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* 57:289–300.
- Bertorelle, G., A. Benazzo, and S. Mona. 2010. ABC as a flexible framework to estimate demography over space and time: some cons, many pros. *Mol. Ecol.* 19:2609–2625.
- Braconnot, P., B. Otto-Bliesner, S. Harrison, S. Joussaume, J. Y. Peterchmitt, A. Abe-Ouchi, M. Crucifix, E. Driesschaert, T. Fichefet, C. D. Hewitt, et al. 2007. Results of PMIP2 coupled simulations of the mid-Holocene and last glacial maximum—part I: experiments and large-scale features. *Clim. Past* 3:261–277.
- Brown, W. J. 1956. The New World species of *Chrysomela* L. (Coleoptera: Chrysomelidae). *Can. Entomol.* 88:1–54.
- Brunfeld, S. J., J. Sullivan, D. E. Soltis, and P. S. Soltis. 2001. Comparative phylogeography of northwestern North America: a synthesis. Pp. 319–339 in J. Silvertown and J. Antonovics, eds. *Integrating ecological and evolutionary processes in a spatial context*. Blackwell Science, Oxford, U.K.
- Carstens, B. C., and C. L. Richards. 2007. Integrating coalescent and ecological niche modeling in comparative phylogeography. *Evolution* 61:1439–1454.
- Curat, M., N. Ray, and L. Excoffier. 2004. SPLATCHE: a program to simulate genetic diversity taking into account environmental heterogeneity. *Mol. Ecol. Notes* 4:139–142.
- Dahlhoff, E. P., S. L. Fearnley, D. A. Bruce, A. G. Gibbs, R. Stoneking, D. M. McMillan, K. Diener, J. T. Smiley, and N. E. Rank. 2008. Effects of temperature on physiology and reproductive success of a montane leaf beetle: implications for persistence of native populations enduring climate change. *Physiol. Biochem. Zool.* 81:718–732.
- Danforth, B. N., S. G. Brady, S. D. Sipes, and A. Pearson. 2004. Single-copy nuclear genes recover Cretaceous-age divergences in bees. *Syst. Biol.* 53:309–326.

- Dellicour, S., and P. Mardulyn. 2014. SPADS 1.0: a toolbox to perform spatial analyses on DNA sequence datasets. *Mol. Ecol. Res.* 14:647–651.
- Edgar, R. C. 2004. MUSCLE: multiple sequence alignment with improved accuracy and speed. *Nucleic Acids Res.* 32:1792–1797.
- Elith, J., S. J. Phillips, T. Hastie, M. Dudík, Y. E. Chee, and C. J. Yates. 2011. A statistical explanation of MaxEnt for ecologists. *Divers. Distrib.* 17:43–57.
- Excoffier, L., P. E. Smouse, and J. M. Quattro. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131:479–491.
- Fagundes, N. J., N. Ray, M. Beaumont, S. Neuenschwander, F. M. Salzano, S. L. Bonatto, and L. Excoffier. 2007. Statistical evaluation of alternative models of human evolution. *Proc. Natl. Acad. Sci. USA* 104:17614–17619.
- Galbreath, K.E., D. J. Hafner, and K. R. Zamudio. 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.
- Gómez-Zurita, J., F. Koplík, K. Theodorides, A. P. Vogler. 2004. Resources for a phylogenomic approach in leaf beetle (Coleoptera) systematics. Pp. 19–35 in P. Jolivet, J. A. Santiago-Blay, and M. Schmitt, eds. *New developments in the biology of Chrysomelidae*. SPB Academic Publishing, The Hague, The Netherlands.
- Hewitt, G. M. 1999. Post-glacial re-colonization of European biota. *Biol. J. Linn. Soc.* 68:87–112.
- . 1996. Some genetic consequences of ice ages, and their role in divergence and speciation. *Biol. J. Linn. Soc.* 58:247–276.
- . 2001. Speciation, hybrid zones and phylogeography—or seeing genes in space and time. *Mol. Ecol.* 10:537–549.
- . 2003. Ice ages, species distributions, and evolution. Pp. 339–361 in L. J. Rothschild and A. M. Lister, eds. *Evolution on planet earth: the impact on the physical environment*. Academic Press, Lond.
- . 2004. Genetic consequences of climatic oscillations in the Quaternary. *Philos. Trans. R. Soc. Lond. B* 359:183–195.
- Hickerson, M. J., B. C. Carstens, J. Cavender-Bares, K. A. Crandall, C. H. Graham, J. B. Johnson, L. Rissler, P. F. Victoriano, and A. D. Yoder. 2010. Phylogeography's past, present, and future: 10 years after Avise, 2000. *Mol. Phylogenet. Evol.* 54:291–301.
- Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* 25:1965–1978.
- Knoll, S., M. Rowell-Rahier, P. Mardulyn, and J. M. Pasteels. 1996. Spatial genetic structure of leaf beetle species with special emphasis on alpine populations. Pp. 379–388 in P. H. A. Jolivet and M. L. Cox, eds. *Chrysomelidae biology, vol. 1: the classification, phylogeny and genetics*. Academic Publishing, Amsterdam, The Netherlands.
- Knowles, L. L. 2000. Tests of Pleistocene speciation in montane grasshoppers (genus *Melanoplus*) from the sky islands of western North America. *Evolution* 54:1337–1348.
- . 2001. Did the pleistocene glaciations promote divergence? Tests of explicit refugial models in montane grasshoppers. *Mol. Ecol.* 10:691–701.
- Knowles, L., and D. F. Alvarado-Serrano. 2010. Exploring the population genetic consequences of the colonization process with spatio-temporally explicit models: insights from coupled ecological, demographic and genetic models in montane grasshoppers. *Mol. Ecol.* 19:3727–3745.
- Knowles, L. L., B. C. Carstens, and M. L. Keat. 2007. Coupling genetic and ecological-niche models to examine how past population distributions contribute to divergence. *Curr. Biol.* 17:940–946.
- Kuhner, M. K. 2006. LAMARC 2.0: maximum likelihood and Bayesian estimation of population parameters. *Bioinformatics* 22:768–770.
- Laurent, S. J., A. Werzner, L. Excoffier, and W. Stephan. 2011. Approximate Bayesian analysis of *Drosophila melanogaster* polymorphism data reveals a recent colonization of Southeast Asia. *Mol. Biol. Evol.* 28:2041–2051.
- Mardulyn, P., and M. C. Milinkovitch. 2005. Inferring contemporary levels of gene flow and demographic history in a local population of the leaf beetle *Goniocenta olivacea* from mitochondrial DNA sequence variation. *Mol. Ecol.* 14:1641–1653.
- Mardulyn, P., Y. E. Mikhailov, and J. M. Pasteels. 2009. Testing phylogeographic hypotheses in a Euro-Siberian cold-adapted leaf beetle with coalescent simulations. *Evolution* 63:2717–2729.
- Mardulyn, P., N. Othmezzouri, Y. E. Mikhailov, and J. M. Pasteels. 2011. Conflicting mitochondrial and nuclear phylogeographic signals and evolution of host-plant shifts in the boreo-montane leaf beetle *Chrysomela lapponica*. *Mol. Phylogenet. Evol.* 61:686–696.
- Marske, K. A., R. A. B. Leschen, G. M. Barker, and T. R. Buckley. 2009. Phylogeography and ecological niche modelling implicate coastal refugia and trans-alpine dispersal of a New Zealand fungus beetle. *Mol. Ecol.* 18:5126–5142.
- Marske, K. A., R. A. B. Leschen, and T. R. Buckley. 2011. Reconciling phylogeography and ecological niche models for New Zealand beetles: looking beyond glacial refugia. *Mol. Phylogenet. Evol.* 59:89–102.
- Nei, M., and W. H. Li. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci. USA* 76:5269–5273.
- Phillips, F. M., M. Zreda, M. A. Plummer, D. Elmore, and D. H. Clark. 2009. Glacial geology and chronology of Bishop Creek and vicinity, eastern Sierra Nevada, California. *Geol. Soc. Am. Bull.* 121:1013–1033.
- Phillips, S. J., and M. Dudík. 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* 31:161–175.
- Phillips, S. J., R. P. Anderson, and R. E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. *Ecol. Modell.* 190:231–259.
- Pons, O., and R. J. Petit. 1995. Estimation, variance and optimal sampling of genetic diversity. I. Haploid locus. *Theor. Appl. Genet.* s91:122–130.
- . 1996. Measuring and testing genetic differentiation with ordered versus unordered alleles. *Genetics* 144:1237–1245.
- Rank, N. E. 1992a. A hierarchical analysis of genetic differentiation in a montane leaf beetle *Chrysomela aeneicollis* (Coleoptera: Chrysomelidae). *Evolution* 46:1097–1111.
- . 1992b. Host plant preference based on salicylate chemistry in a willow leaf beetle (*Chrysomela aeneicollis*). *Oecologia* 90:95–101.
- . 1994. Host plant effects on larval survival in a salicin-using leaf beetle *Chrysomela aeneicollis* Schaeffer (Coleoptera: Chrysomelidae). *Oecologia* 97:342–353.
- Ray, N., M. Currat, M. Foll, and L. Excoffier. 2010. SPLATCHE2: a spatially explicit simulation framework for complex demography, genetic admixture and recombination. *Bioinformatics* 26:2993–2994.
- Simon, C., F. Frati, A. Beckenbach, B. Crespi, H. Liu, and P. Flook. 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene

- sequences and a compilation of conserved polymerase chain reaction primers. *Ann. Entomol. Soc. Am.* 87:651–701.
- Smiley, J. T., and N. E. Rank. 1986. Predator protection versus rapid growth in a montane leaf beetle. *Oecologia* 70:106–112.
- Stephens, M., and P. Donnelly. 2003. A comparison of Bayesian methods for haplotype reconstruction from population genotype data. *Am. J. Hum. Genet.* 73:1162–1169.
- Stephens, M., N. J. Smith, and P. Donnelly. 2001. A new statistical method for haplotype reconstruction from population data. *Am. J. Hum. Genet.* 68:978–989.
- Taberlet, P., L. Fumagalli, A. Wust-Saucy, and J. Cosson. 1998. Comparative phylogeography and postglacial colonization routes in Europe. *Mol. Ecol.* 7:453–464.

Associate Editor: G. Mayer

### Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

**Figure S1.** Standard deviation associated with inferred distributions of *Chrysomela aeneicollis* in North America estimated with MAXENT.

**Table S1.** AUC (area under the curve) scores for each replicate run performed with MAXENT.

**Table S2.** Demographic parameters estimated with LAMARC for the five regions defined a priori.

**Table S3.** Effective number of migrants ( $N_e m$ ) estimated with LAMARC.

**Table S4.**  $P$ -values from comparison between real and simulated datasets ( $t_R = 5$ ).

**Table S5.**  $P$ -values from comparison between real and simulated datasets ( $t_R = 10$ ).

**Appendix S1.** Additional details on coalescence simulations and on comparisons of simulated and observed sequence data.