Report

Coupling Genetic and Ecological-Niche Models to Examine How Past Population Distributions Contribute to Divergence

L. Lacey Knowles,[1,*] Bryan C. Carstens,[1]
and Marcia L. Keat[1]
[1] Department of Ecology and Evolutionary Biology
1109 Geddes Avenue
Museum of Zoology
University of Michigan
Ann Arbor, Michigan 48109-1079

Summary

Understanding the impact of climate-induced distributional shifts on species divergence, like those accompanying the Pleistocene glacial cycles [1, 2], requires tools that explicitly incorporate the geographic configuration of past distributions into analyses of genetic differentiation. Depending on the historical distribution of species, genetic differences may accumulate among ancestral source populations, but there is long-standing debate whether displacements into glacial refugia promoted divergence. Here we integrate coalescent-based genetic models [3, 4] with ecological-niche modeling [5, 6] to generate expectations for patterns of genetic variation based on an inferred past distribution of a species. Reconstruction of the distribution of a montane grasshopper species during the last glacial maximum suggests that *Melanoplus marshalli* populations from the sky islands of Colorado and Utah were likely colonized from multiple ancestral source populations. The genetic analyses provide compelling evidence that the historical distribution of *M. marshalli*—namely, spatial separation of multiple refugia—was conducive to genetic differentiation. The coupling of genetic and ecological-niche modeling provides a new and flexible tool for integrating paleoenvironmental details into species-specific predictions of population structure that can increase our understanding of why the glacial cycles promoted speciation in some taxa and yet inhibited diversification in others [7, 8].

Results

Projections of Species Distribution during Glacial Maximum

Environmental variables that capture biologically meaningful aspects of climate variation were used to reconstruct the past distribution of *Melanoplus marshalli* during the last glacial maximum (21,000 years ago) based on the species' current distribution across the sky islands (i.e., mountain tops) of Colorado and Utah. The model uses 19 bioclimatic parameters (which follow those of ANUCLIM; Australian National University: http://cres.anu.edu.au/outputs/anuclim.php) that include annual trends for temperature and precipitation, aspects

*Correspondence: knowlesl@umich.edu

of seasonality, and extreme or potentially limiting environmental factors. These parameters were derived from both current climate measurements and climate estimates for the last glacial maximum [9] at a spatial resolution of 1 arc-minute. The ecological-niche model generated with the program MAXENT [10] provides a projected distribution that shows *M. marshalli* is restricted to montane habitats with inhospitable intervening habitats limiting contemporary migration in this flightless grasshopper, and the species was displaced from their montane meadows during glacial periods (Figure 1). Furthermore, reconstruction of the past distribution indicates that the intervening region between sky island populations from Colorado and Utah did not contain suitable habitat during the most recent glacial maximum. Instead, projections of the past distribution shows that geographic regions proximate to the contemporary sky island populations may have served as refugia, suggesting that populations were likely recolonized from more than a single ancestral source population as the glaciers retreated (Figure 1). To explore the accuracy of the projected distributions, we performed an additional analysis based on geo-referenced museum collections (i.e., localities not included in the original model); similarity in AUC values (see Experimental Procedures for details) suggest a good fit between the models and the data.

Spatial Arrangement of Genetic Variation

A single maximum-likelihood genealogy was estimated with PAUP* [11] for the mitochondrial and nuclear locus sequenced in 95 individuals sampled from 10 *M. marshalli* populations from Utah and Colorado (Figure 2). The gene trees suggest that the sorting of ancestral polymorphism is not complete—none of the populations are monophyletic for both loci. Haplotypes from the eastern (Colorado) and western (Utah) populations also do not form reciprocally monophyletic clades, and enforcing monophyly on the gene tree topologies significantly decreases the fit of data (p < 0.01; a shift in tree-length from 107 to 153, and from 338 to 447 for SCNPS-2 and COI, respectively).

Comparison of the pair-wise estimates of migration among populations shows that levels of gene flow were similar irrespective of the geographic distance separating populations (Figure 3). Interpretations based solely on the individual migration values from the program MIGRATE-N [12] are potentially problematic because of violations of methodological assumptions (see [13]), which is a general problem with other coalescent-based approaches as well (e.g., a series of analyses among each possible pair of populations with a program like IM [14] does not take into account that the source of migrants, and/or ancestral variation may come from populations not considered in each independent analysis). Nevertheless, the geographic distribution of migration estimates is informative. For example, migration estimates for geographically proximate populations
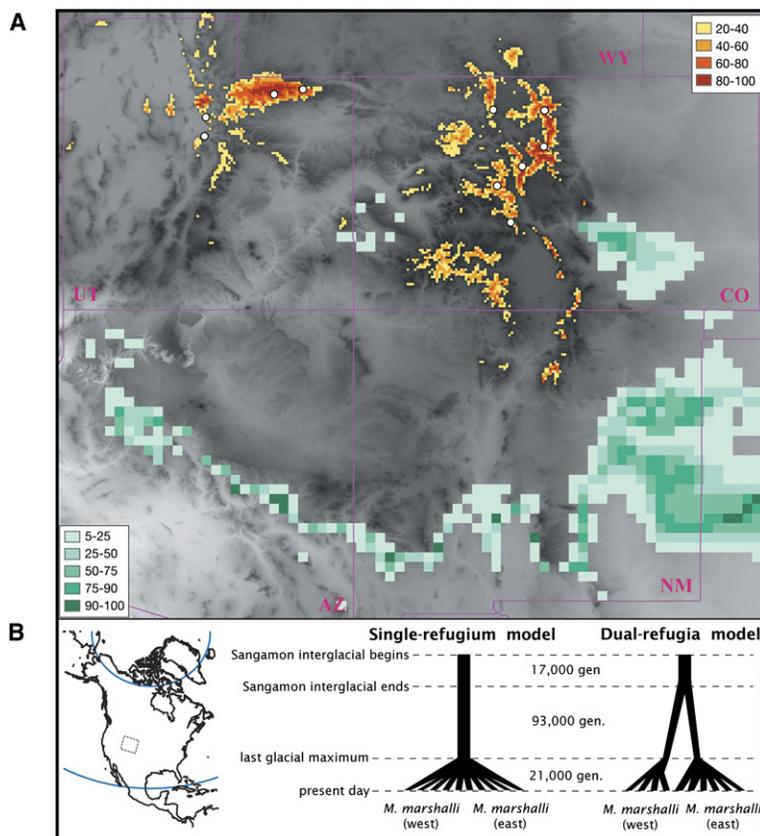
Figure 1. The Coupling of the Ecological-Niche Modeling and Genetic Models

(A) The predicted contemporary distribution of *M. marshalli* and the reconstructed past distribution during the last glacial maximum (21,000 years ago). The projections are based on an integration of contemporary [9] and paleoenvironmental information [37] with the program MAXENT [10]. The darkest shading in the color-scale keys indicate the highest predicted areas for the contemporary ecological niche model (top right) and the predicted past habitat (bottom left); sampled populations are indicated by white circles.

(B) The genetic models representing the geographic configurations of ancestral populations that were used to explore whether the spatial distribution of genetic variation across sky-island populations reflect colonization from a single- versus dual-refugial populations (i.e., the null and alternative hypotheses, respectively).

separated by 50–200 km (i.e., among populations within either the eastern or western regions) were not larger than the migration rate estimates for geographically distant populations separated by as much as 500 km (i.e., between populations from the separate regions). This pattern suggests that genetic drift, rather than gene flow, is the predominant factor underlying the spatial distribution of genetic variation and the observed patterns of incomplete lineage sorting (Figure 2). These analyses also indicate that the genetic model (described below) used to generate expected patterns of genetic variation is justified.

A Genetic Model of the Geography of Population Divergence

Visual inspection of the empirical gene trees (Figure 2) shows how the retention of ancestral polymorphism obscures any obvious correspondence between the spatial distribution of populations and their genetic relatedness. Yet, the pattern of incomplete lineage sorting itself contains information about the demographic past, which can be extracted by placing the observed genetic data in a statistical phylogeographic framework. Two genetic models were used to explore how the geographic configuration of ancestral populations affects the distribution of genetic variation across contemporary populations of *M. marshalli* (Figure 1B). Specifically, we (1) simulated genetic data under the null hypothesis—colonization from a single refugium—and then (2) compared the probability of the observed data (the empirical data) to the probability of the simulated data under the alternative hypothesis—multiple ancestral

source populations [15]. This coalescent framework for hypothesis testing (i.e., rejecting the null hypothesis) takes into account the stochastic effects of genetic drift [4] on patterns of genetic variation, and by simulating nucleotide data under a model of evolution [16] that matches those estimated for the empirical data, the effects of mutational variance are considered as well.

The geographic test of population divergence provides an important complement to inferring how climatically induced distributional shifts contribute to population differentiation based on divergence-time estimates (e.g., distinguishing between glacial versus interglacial differentiation). Inferences based on divergence-time estimates are sensitive to assumptions about the molecular clock. Moreover, the rapidity of distributional shifts exceeds the genetic resolution required for detailed inferences based on divergence-time estimates [17, 18]. For example, estimates of the timing of population divergence among sky-island populations are consistent with their founding from glacial refugia. However, the error surrounding the timing of divergence extends across glacial and interglacials—the 90% highest posterior density interval on estimates of regional divergence between eastern and western populations ranges from 6.89×10^3 to 1.29×10^5 years ago (with the program IM). Because the data includes five anonymous nuclear loci with unknown mutation rates, the geometric mean of the ratios θ_i: $\theta_{i\ COI}$ for each of the five nuclear loci was used to calculate the mutation rate of 1.66×10^{-5} [18] for scaling parameter estimates [14], based on a rate of 2.3% sequence divergence per million years for COI [19]. Despite the increased resolution afforded
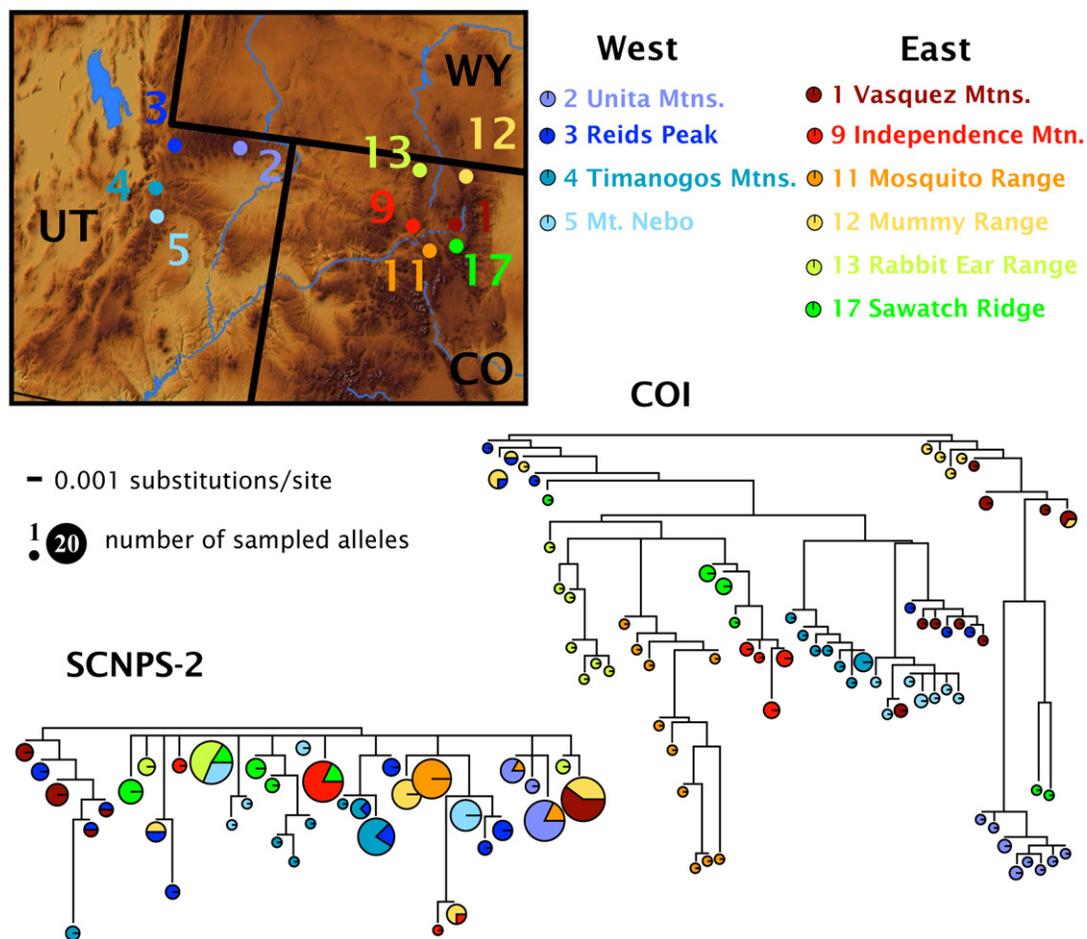
Figure 2. Geographic Distribution of Genetic Variation across the *M. marshalli* Sky-Island Populations from Colorado and Utah

Haplotypes are colored according to the populations from which they were sampled. Genealogies for the mitochondrial and nuclear locus (COI and SCNPS-2, respectively) were estimated by maximum likelihood in PAUP* [11] via a heuristic search with 10 random addition sequence replicates and models of sequence evolution selected with DT-MODSEL [16].

by the coalescent-based approach and analysis of multiple loci [14, 20, 21], which was based on considering six loci (Table S2 in the Supplemental Data available online), the contribution of historical population configurations to divergence cannot be identified based on estimates of the timing of divergence.

Hypothesis Testing within the Statistical Phylogeographic Framework

The single-refugium historical scenario posits that *M. marshalli* populations were founded from a single ancestral population as the Wisconsonian glaciers retreated. This null hypothesis was modeled where each of the contemporary sky-island populations correspond to a population in the population tree that was founded 21,000 years ago from a single ancestral source population (Figure 1B). Alternatively, if the eastern and western *M. marshalli* populations were founded from separate glacial refugia, the contemporary populations would have been colonized from allopatric ancestral-source populations that existed during the Wisconsonian glaciation, approximately for the past 114,000 years based on the geologic dates [22]. This alternative hypothesis was modeled with groups of sky-island populations

sharing a common history if they were founded from the same ancestral refugial-source population as the glaciers retreated (Figure 1B). Parameterization of the model, namely the timing of recolonization and the period that ancestral source populations remained isolated, was based on geologic dates for the most recent and preceding interglacials [22, 23] and corresponds to the paleoenvironmental data [9] used in the ecological-niche models to project the past distribution of *M. marshalli*. The robustness of the conclusions was explored by varying the models parameters; specifically, a range of effective population sizes was explored. The same effect could have been achieved by exploring a range of divergence times (i.e., the number of deep coalescents depends on the population size relative to the divergence time; see e.g., [24]). However, because the times used in the genetic model are based on the independent data from the ecological-niche models, in this case it is preferable to vary the effective population-size parameters.

The theoretical expectation for the amount of discord (i.e., deep coalescence) that would be observed between the dual-refugial model and each locus (i.e., for COI and SCNPS-2 based on the coalescent simulations,
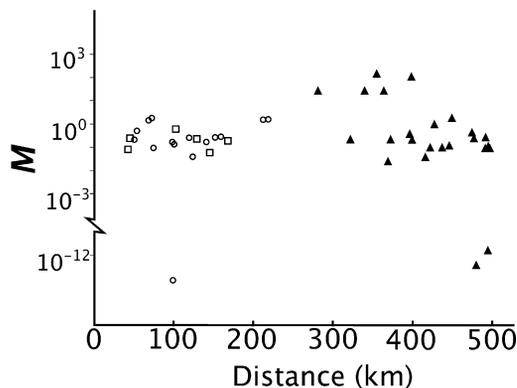
Figure 3. Pair-Wise Comparison of Gene-Flow Estimates and the Geographic Distance Separating *M. marshalli* Populations

Comparisons of populations within the eastern (Colorado) and western (Utah) regions are indicated by the open symbols, whereas comparisons of populations from the separate regions are marked with a closed triangle. There is a break in the y axis in order to accommodate the very low gene-flow estimates, rather than use a log-scale. Estimates of the number of gene copies per generation moving between populations, M, were made with locus-specific transition-transversion rates and base frequencies for COI and SNCPS2-2 (Table S1 for details) and a search strategy that included adaptive heating strategy with 10 short chains (length 5.0×10^5 generations) and 5 long chains (length 5.0×10^7 generations) in MIGRATE-N [12]. The analysis was repeated to verify parameter estimates. The pair-wise population estimates were consistent with migration estimates between the eastern and western regions ($M_{E-W} = 5.5 \times 10^{-6}$; $M_{W-E} = 1.1 \times 10^{-5}$; based on IM analyses).

as explained above) when the data had instead evolved under a history involving a single ancestral source population is quite high (Figure 4). While both loci show some deep coalescence (i.e., 47 and 109 deep coalescence for COI and SCNPS-2, respectively; Figure 4), the amount of discord between the empirical data and the dual-refugial model is significantly lower than expected if the populations had actually been colonized from a single ancestral source population. Additional simulations indicate that this conclusion is robust to the inherent variance associated with parameterization of the genetic model. The results were consistent across repeated analyses with different effective population sizes (N_e) that encompass the 90% posterior density interval on the estimate (an N_e as low as 60,000 or as high as 166,000 individuals) that was derived from a Bayesian analysis of six loci with the program IM [14]; parameter estimates were scaled via the mutation rate of derived from the geometric mean of the ratios θ_i: $\theta_{i\ COI}$ for each nuclear locus as discussed earlier, namely 1.66×10^{-5} [18]. To assure that the population genetic parameters were not influenced by an inappropriate choice for the priors in the IM analysis, the posterior probability distribution of each parameter was investigated over a range of priors to assure that the parameter space was explored fully and effectively. The parameter space was searched by a linear heating scheme and ten metropolis-coupled Markov chains of at least 10 million generations each; the analysis was repeated twice with different random number seeds to confirm convergence.

It is possible that more than two ancestral-source populations (Figure 1) might be involved (see Discussion). However, this scenario would not affect the key
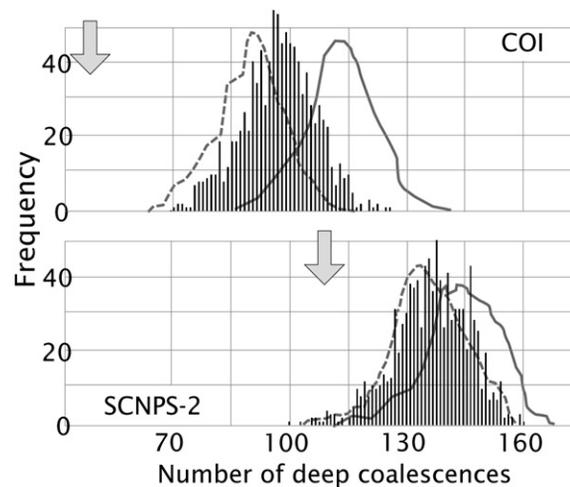

Figure 4. Results of Genetic Analyses

The amount of lineage sorting (number of deep coalescents) observed between the dual-refuge model and COI and SCNPS-2 (marked with the arrow) is compared to the amount of discord (number of deep coalescents) expected between the dual-refugial model and a gene tree if the data had instead evolved under a history involving a single ancestral source. Distributions based on the estimate of N_e (the barred histogram), as well as the upper and lower 90% highest posterior density interval (the dashed and solid lines, respectively) surrounding the estimate from the Bayesian coalescent-based program IM [14] were generated from simulated data with the program MESQUITE 1.0 [42]. The simulations were conducted under conditions that mirrored those of the observed empirical data, specifically: (1) nucleotides matching the length of sequence and model of evolution estimated for COI and SCNPS-2 (see Table S1 for details) were simulated along the branches of gene trees simulated by a neural coalescent under the model of a single ancestral source, (2) the timing of population divergence used in the model was parameterized based on the geologic dates for the most recent and preceding interglacials, (3) estimates of N_e matched the empirical data and encompassed the 90% highest posterior density interval surrounding the estimate, and (4) the number of deep coalescents between each of the 500 simulated gene trees and the dual-refuge model were calculated. The N_e of individual sky-island populations was modeled as being equivalent because the confidence intervals on the per-population estimates broadly overlapped.

question of whether past distributions during the last glacial period contributed to divergence, because this relies on the test of multiple versus single ancestral source populations, not on the determination of how many allopatric glacial refugia there might have been. For the past distribution to be conducive to divergence, species would have to have been displaced into multiple refugia where genetic difference could accumulate among separate ancestral source populations. Otherwise, the effects of the glacial cycles on species divergence would be limited to the drift-induced changes during interglacials when contemporary populations were recolonized from a single refugium [25, 26]. Therefore, while not ideal, the genetic models (Figure 1B) are sufficient for determining whether divergence during glacial periods played a role in promoting divergence.

Discussion

The climatic fluctuations of the Pleistocene induced wide-scale habitat shifts [27] that are hypothesized to

have promoted bursts of speciation in taxa ranging from Neotropical birds to montane insects (e.g., [1, 2]). However, a complexity of species-specific responses to paleoenvironmental conditions are now apparent from numerous palynological studies (e.g., [27]), and methodological limitations of genetic analyses posed by the recency and frequency of the glacial cycles [17, 18] has exacerbated disputes over the role of shifting species distributions in speciation [7, 8].

The coupling of genetic and ecological-niche models offers an opportunity for species-specific, quantitative evaluation of how rapid climatic fluctuations affected divergence. Because ecological-niche models can be projected onto models of past climate, the distribution of suitable habitat can be reconstructed for individual species (e.g., [28]). This provides a framework for generating hypotheses about how the historical configuration of species distributions may have contributed to species divergence and ultimately biodiversity [29]. Coupling these hypotheses with genetic models permits a statistical evaluation that are (1) based on biologically informed models, as opposed to relying on the generic models of computer programs that may not adequately capture the biological phenomena of interest [30, 31], and (2) consider the inherent stochasticity associated with the spatial distribution of genetic variation [3, 4].

The approach used here offers a window into how the glacial cycles promoted divergence in *M. marshalli*—a member of a clade of flightless montane grasshoppers that radiated during the Pleistocene [18, 25, 32]. The projections from the ecological-niche model (Figure 1) show how *M. marshalli* likely tracked the habitat shifts in montane regions [25, 32, 33], and the genetic models indicate that patterns of genetic variation are consistent with the predictions that contemporary sky-island populations were founded from multiple ancestral source populations (Figure 4). This mode of divergence may be key to understanding how some groups were able to speciate during the dynamic Pleistocene. Displacement into multiple glacial refugia provides a geographic setting conducive to divergence, and because interglacials are of relatively short duration compared to glacial periods [22], this is also the distribution that has predominated species histories during the past two million years [33]. With broader application, the coupled models should be useful in identifying paleoecological features influencing past distributions, and consequently, why the glacial cycle effects on species diversification differed among taxa [7, 8].

Future Applications and Methodological Development

This study not only identifies the potential of this integrative approach, but it also illustrates areas for future development. The coupling of genetic and ecological-niche models integrates detailed paleoenvironmental data with genetic predictions [34]. Yet, additional data are needed to improve the models' resolution. For example, to move beyond the coarse-grained analysis applied here (i.e., single versus multiple ancestral source populations), data from landscape features (e.g., [35, 36]) might be used in conjunction with the projected species distribution 21,000 years ago to reconstruct likely routes of displacement and recolonization.

Establishing the connections between ancestral and contemporary populations, as well as how the size of ancestral and contemporary populations may have differed, will increase biological insights by providing more realistic models. Such improvements will no doubt be critical for understanding the spatial and temporal scales of processes contributing to species divergence, such as whether historical or current habitat configurations better explain biodiversity patterns [29].

Experimental Procedures

Projections of Species' Distributions Based on the Ecological-Niche Models

Mean monthly climate estimates at a spatial resolution of 10 arc-minutes were generated for the last glacial maximum of the Pleistocene (LGM) under the Community Climate System Model (CCSM3, [9]) and statistically downscaled [37] by R.J. Hijmans with data provided by the Paleoclimate Modelling Intercomparison Project Phase II (PMIP2). The bioclimatic parameters used in the ecological-niche model of current conditions (1950–2000) were derived from the WorldClim global climate database [37].

Climate models were generated via the maximum entropy machine learning algorithm in MAXENT [10]. A distribution model was generated with the 19 bioclimatic parameters for the current climate and the collection localities. This model was then applied to the set of bioclimatic variables generated from CCSM3 to estimate the geographic distribution and extent of suitable climate for each species during the last glacial maximum. Model performance was evaluated with geo-referenced museum-collection localities as a source to ground-truth the model. The similarity in the AUC (area under the curve) value of 0.996 for the model based on locality data for the 10 populations sampled for this study, and a model based on 18 georeferenced localities from museum collections (AUC = 0.998) suggest a good fit between the model and data.

Genetic Sampling

A balance between sampling individuals versus loci was used for this investigation, while maintaining full coverage of the range of the species (Tables S2 and S3). Two separate sample schemes were used that differed with respect to the number of individuals versus loci studied, which accommodates the shift in information content when making specific historical inferences [38, 39]. To investigate the geography of divergence, the emphasis was on collecting sequence data from multiple individuals because the sampling of populations is critical. To estimate the timing of divergence, sequencing efforts were focused on increasing the number of loci (see [20]), because the correlation that is due to coancestry reduces the information contained in samples from multiple individuals [3, 38].

The anonymous nuclear locus designated SCNPS-2 and COI were sequenced in each specimen, for a total of 2220 bp per individual, to test hypotheses regarding the geography of divergence, namely whether the contemporary sky islands were founded from a single glacial refugium. SCNPS from four additional loci [40] were gathered in a subset of the individuals (n = 5 individuals in each of four populations), for a total of 5667 bp per individual, to estimate the timing of divergence between putative refugial populations. The single-copy nuclear polymorphic sequences (SCNPS) were gathered from anonymous loci identified from a genomic library (see [40] for details) and the mitochondrial gene cytochrome oxidase I (COI) (see [32] for details). SCNPS were initially identified from an interspecific screening set, not levels of variability within *M. marshalli*; this approach avoids an ascertainment bias associated with using an intraspecific screening set (i.e., the lower bound for allele frequencies would depend on the number of individuals used to detect variable loci [41]). A plot of the distribution of pair-wise differences among individuals within the species was not truncated, demonstrating that the interspecific screening set did not introduce an ascertainment bias. Tests also confirmed the lack of recombination in each locus [40], which is an assumption of the coalescent analyses [12, 14]. All loci were

sequenced with an ABI 3730 Automated Sequencer at the University of Michigan DNA Sequencing Core.

Supplemental Data

Three tables are available at http://www.current-biology.com/cgi/content/full/17/10/■ ■ ■/DC1/.

References

1. Hewitt, G.M. (2000). The genetic legacy of the Quaternary ice ages. Nature 405, 907–913.

2. Haffer, J. (1969). Speciation in Amazonian forest birds. Science 165, 131–137.

3. Hudson, R.R. (1990). Gene genealogies and the coalescent process. In Oxford Survey Evolutionary Biology, D. Futuyma and J. Antonovics, eds. (New York: Oxford University Press), pp. 1–44.

4. Knowles, L.L. (2004). The burgeoning field of statistical phylogeography. J. Evol. Biol. 17, 1–10.

5. Austin, M.P. (1985). Continuum concept, ordination methods, and niche theory. Annu. Rev. Ecol. Syst. 16, 39–61.

6. Peterson, A.T. (2001). Predicting species geographic distributions based on ecological niche modeling. Condor 103, 599–605.

7. Hewitt, G.M. (2004). Genetic consequences of climatic oscillations in the Quaternary. Philos. Trans. R. Soc. Lond. B Biol. Sci. 359, 183–195.

8. Bennett, K.D. (2004). Continuing the debate on the role of Quaternary environmental change for macroevolution. Philos. Trans. R. Soc. Lond. B Biol. Sci. 359, 295–303.

9. Collins, W.D., Bitz, C.M., Blackmon, M.L., Bonan, G.B., Bretherton, C.S., Carton, J.A., Chang, P., Doney, S.C., Hack, J.J., Henderson, T.B., et al. (2006). The community climate system model version 3 (CCSM3). J. Clim. 19, 2122–2143.

10. Phillips, S.J., Anderson, R.P., and Schapire, R.E. (2006). Maximum entropy modeling of species geographic distributions. Ecol. Modell. 190, 231–259.

11. Swofford, D.L. (2002). PAUP*. Phylogenetic Analysis Using Parsimony (and other methods), Version 4 (Sunderland, MA: Sinauer Associates).

12. Beerli, P. (2002). MIGRATE: documentation and program, part of LAMARCK. Version. 1.5 (http://evolution.genetics.washington.edu/lamarc/index.html).

13. Slatkin, M. (2005). Seeing ghosts: the effect of unsampled populations on migration rates estimated for sampled populations. Mol. Ecol. 14, 67–73.

14. Hey, J., and Nielsen, R. (2004). Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of Drosophila pseudoobscura and D. persimilis. Genetics 167, 747–760.

15. Knowles, L.L., and Maddison, W.P. (2002). Statistical phylogeography. Mol. Ecol. 11, 2623–2635.

16. Minin, V., Abdo, Z., Joyce, P., and Sullivan, J. (2003). Performance-based selection of likelihood models for phylogeny estimation. Syst. Biol. 52, 674–683.

17. Lascoux, M., Palme, A.E., Cheeadi, R., and Latta, R.G. (2004). Impact of ice ages on the genetic structure of trees and shrubs. Philos. Trans. R. Soc. Lond. B Biol. Sci. 359, 197–207.

18. Carstens, B.C., and Knowles, L.L. (2007). Shifting distributions and speciation: genomic resolution of species divergence during rapid climate change. Mol. Ecol. 16, 619–627.

19. Brower, A.V.Z. (1994). Rapid morphological radiation and convergence among races of the butterfly Heliconius erato inferred from patterns of mitochondrial DNA evolution. Proc. Natl. Acad. Sci. USA 91, 6491–6495.

20. Arbogast, B.S., Edwards, S.V., Wakeley, J., Beerli, P., and Slowinski, J.B. (2002). Estimating divergence times from molecular data on phylogenetic and population genetic timescales. Annu. Rev. Ecol. Syst. 33, 707–740.

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

22. Gibbard, P., and Van Kolfschoten, T. (2004). The pleistocene and holocene epochs. In A Geologic Time Scale, F.M. Gradstein, J.G. Ogg, and A.G. Smith, eds. (London: Cambridge University Press), pp. 441–452.

23. Pierce, K.L. (2004). Pleistocene glaciations of the Rocky Mountains. In The Quaternary Period in the United States, A.R. Gillespie, S.C. Porter, and B.F. Atwater, eds. (New York: Elsevier Press), pp. 63–76.

24. DeChaine, E.G., and Martin, A.P. (2006). Using coalescent simulations to test the impact of Quaternary climate cycles on divergence in an alpine plant-insect association. Evolution Int. J. Org. Evolution 60, 1004–1013.

25. Knowles, L.L. (2001). Genealogical portraits of speciation in montane grasshoppers (genus Melanoplus) from the sky islands of the Rocky Mountains. Proc. R. Soc. Lond. B. Biol. Sci. 268, 319–324.

26. Kadereit, J.W., Griebeler, E.M., and Comes, H.P. (2004). Quaternary diversification in European alpine plants: pattern and process. Philos. Trans. R. Soc. Lond. B Biol. Sci. 359, 265–274.

27. Willis, K.J., Bennett, K.D., and Walker, D. (2004). Introduction to the evolutionary legacy of the Quaternary ice ages. Philos. Trans. R. Soc. Lond. B Biol. Sci. 359, 157–158.

28. Hugall, A., Moritz, C., Moussalli, A., and Stanisic, J. (2002). Reconciling paleodistribution models and comparative phylogeography in the wet tropics rainforest land snail Arosophia bellendenkerensis (Brazier 1875). Proc. Natl. Acad. Sci. USA 99, 6112–6117.

29. Graham, C.H., Moritz, C., and Williams, S.E. (2006). Habitat history improves prediction of biodiversity in rainforest fauna. Proc. Natl. Acad. Sci. USA 103, 632–636.

30. Slatkin, M. (1985). Gene flow in natural populations. Annu. Rev. Ecol. Syst. 16, 393–430.

31. Knowles, L.L., and Carstens, B.C. (2007). Inferring a geographically explicit population divergence models for statistical phylogeographic tests in montane grasshoppers. Evolution Int. J. Org. Evolution 61, 477–493.

32. Knowles, L.L. (2000). Tests of Pleistocene speciation in montane grasshoppers (genus Melanoplus) from the sky islands of western North America. Evolution Int. J. Org. Evolution 54, 1337–1348.

33. Thompson, R.S., Shafer, S.L., Strickland, L.E., Van de Water, P.K., and Anderson, K.H. (2004). Quaternary vegetation and climate change in the western United States. In The Quaternary Period in the United States, A.R. Gillespie, S.C. Porter, and B.F. Atwater, eds. (New York: Elsevier Press), pp. 403–426.

34. Carstens, B.C., and Richards, C.L. (2007). Integrating coalescent and ecological niche modeling in comparative phylogeography. Evolution Int. J. Org. Evolution, in press.

35. Galindo, H.M., Olson, D.B., and Palumbi, S.R. (2006). Seascape genetics: a coupled oceanographic-genetic model predicts population structure of Caribbean corals. Curr. Biol. 16, 1622–1626.

36. Manel, S., Schwartz, M.K., Luikart, G., and Taberlet, P. (2003). Landscape genetics: combining landscape ecology and population genetics. Trends Ecol. Evol. 18, 189–197.

37. Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., and Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. Int. J. Climatol. 25, 1965–1978.

38. Pluzhnikov, A., and Donnelly, P. (1996). Optimal sequencing strategies for surveying molecular genetic diversity. Genetics *144*, 1247–1262.

39. Maddison, W.P., and Knowles, L.L. (2006). Inferring phylogeny despite incomplete lineage sorting. Syst. Biol. *55*, 21–30.

40. Carstens, B.C., and Knowles, L.L. (2006). Variable nuclear markers for *Melanoplus oregonensis* identified from the screening of a genomic library. Mol. Ecol. Notes *6*, 683–685.

41. Wakeley, J., Nielsen, R., Liu-Cordero, S.N., and Ardlie, K. (2001). The discovery of single-nucleotide polymorphisms and inferences about human demographic history. Am. J. Hum. Genet. *69*, 1332–1347.

42. Maddison, W.P., and Maddison, D.R. (2004). MESQUITE: a modular system for evolutionary analysis. Version 1.01 (http://mesquiteproject.org).

Supplemental Data

Coupling Genetic and Ecological-Niche Models to Examine How Past Population Distributions Contribute to Divergence

**L. Lacey Knowles, Bryan C. Carstens,
and Marcia L. Keat**

Table S1. Models of Sequence Evolution Estimated for Each Locus and Used in the Genetic Simulations

Locus	Model	Parameters
COI	TrN +I +G	$p_A = 0.323$, $p_C = 0.155$, $p_G = 0.150$, $p_T = 0.372$; $tv_i = 6.26$, $tv_2 = 12.78$; $p_{INV} = 0.7609$; $a = 0.884$
SCNPS-2	HKY + I	$p_A = 0.355$, $p_C = 0.17$, $p_G = 0.175$, $p_T = 0.3$; $ti/tv = 1.993$; $p_{INV} = 0.7722$
SCNPS-73	F81 + I	$p_A = 0.272$, $p_C = 0.183$, $p_G = 0.27$, $p_T = 0.275$; $p_{INV} = 0.9789$
SCNPS-85	K80 +I	$[p_A = 0.289$, $p_C = 0.232$, $p_G = 0.241$, $p_T = 0.238]$; $ti/tv = 1.204$; $p_{INV} = 0.8846$
SCNPS-89	HKY +I +G	$p_A = 0.293$, $p_C = 0.181$, $p_G = 0.253$, $p_T = 0.273$; $ti/tv = 1.90$; $p_{INV} = 0.8174$; $a = 0.8786$
SCNPS-211	HKY +I +G	$p_A = 0.299$, $p_C = 0.214$, $p_G = 0.231$, $p_T = 0.256$; $ti/tv = 1.03$; $p_{INV} = 0.4899$; $a = 0.7876$

Table S2. Choice of Sample Design and Description of the Variation for Each Locus

Locus	Length	Gene Copies	Indels[a]	s[b]
COI	1264	40	0	47
SCNPS-2	956	40	3 (17, 19, 21)	28
SCNPS-73	853	40	0	8
SCNPS-85	825	40	0	18
SCNPS-89	581	40	2 (3, 2)	26
SCNPS-211	1188	30	0	16
COI	1264	95	0	143
SCNPS-2	956	190	3 (17, 19, 21)	44

Six loci were used in all population genetic analyses to estimate population genetic parameters to maximize the information content available in multilocus data (inheritance scalars were used to take into account the differences in effective population size between mitochondrial and nuclear loci), whereas an emphasis on geographic coverage was used to explore whether the contemporary populations were colonized from single- versus dual-refugial populations (as indicated by the last two rows).
[a] The length of each indel is shown in parentheses.
[b] Number of segregating sites.

Table S3. Sampled Populations and the Number of Individuals Sequenced for Each Locus

Sky Island Populations	Latitude/Longitude	n_{TOTAL}	n_{COI}	n_2	n_{73}	n_{85}	n_{89}	n_{211}
Independence Mtns., CO	39.13 N 106.57 W	9	9	9	5	5	5	.
Mosquito Ridge, CO	39.37 N 106.07 W	10	10	10	.	.	.	.
Mummy Range, CO	40.42 N 105.75 W	10	10	10	.	.	.	.
Rabbit Ear Range, CO	40.40 N 106.63 W	7	7	7	.	.	.	.
Sawatch Range, CO	38.50 N 106.33 W	9	9	9	.	.	.	.
Vasquez Mtns., CO	39.80 N 105.77 W	10	10	10	5	5	5	5
Unita Mtns., UT	40.78 N 109.92 W	10	10	10	.	.	.	.
Reids Peak, UT	40.70 N 110.42 W	10	10	10	5	5	5	5
Timanogos Mtns., UT	40.33 N 111.53 W	10	10	10	5	5	5	5
Mt. Nebo, UT	39.92 N 111.55 W	10	10	10	.	.	.	.
	Total	95	95	95	20	20	20	15