

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

L. LACEY KNOWLES and DIEGO F. ALVARADO-SERRANO

Department of Ecology and Evolutionary Biology, Museum of Zoology, University of Michigan, Ann Arbor, MI 48109-1079, USA

Abstract

Understanding the genetic consequences of shifting species distributions is critical for evaluating the impact of climate-induced distributional changes. However, the demographic expansion associated with the colonization process typically takes place across a heterogeneous environment, with population sizes and migration rates varying across the landscape. Here we describe an approach for coupling ecological-niche models (ENMs) with demographic and genetic models to explore the genetic consequences of distributional shifts across a heterogeneous landscape. Analyses of a flightless grasshopper from the sky islands of the Rocky Mountains of North America are used to show how biologically informed predictions can be generated about the genetic consequences of a colonization process across a spatially and temporally heterogeneous landscape (i.e. the suitability of habitats for the montane species differs across the landscape and is itself not static, with the displacement of contemporary populations into glacial refugia). By using (i) ENMs for current climatic conditions and the last glacial maximum to (ii) parameterize a demographic model of the colonization process, which then (iii) informs coalescent simulations, a set of models can be generated that capture different processes associated with distributional shifts. We discuss how the proposed approach for model generation can be integrated into a statistical framework for estimating key demographic parameters and testing hypotheses about the conditions for which distributional shifts may (or may not) enhance species divergence, including the importance of habitat stability, past gene-flow among currently isolated populations, and maintenance of refugial populations in multiple geographic regions.

Keywords: coalescent simulations, glacial cycles, phylogeography, Pleistocene, refugia, speciation

Received 18 January 2010; revised 5 May 2010; accepted 5 May 2010

Introduction

A lack of stability characterizes the history of many (if not most) species distributions. In some cases distributional shifts have been quite dramatic and punctuated, as in the response to the rapid climate change associated with the Pleistocene glacial cycles (e.g. Taberlet

et al. 1998; Hewitt 2000; Shapiro *et al.* 2004; Tzedakis *et al.* 2004; Knowles & Richards 2005; Smith & Farrell 2005; Carstens & Knowles 2007a). Distributional instability is also not unique to the distant past. Expansions associated with invasive species and pronounced reductions in the ranges of previously more widespread taxa (e.g. Silva *et al.* 2002; Smith *et al.* 2004; Stuart *et al.* 2004; Currat *et al.* 2008) provide many contemporary examples where there have been striking and abrupt shifts in distribution. Understanding the genetic consequences

Correspondence: L. Lacey Knowles, Fax: +1 734 763 4080; E-mail: knowlesl@umich.edu

of such demographic processes is important for understanding under what conditions distributional shifts might threaten species existence, as opposed to playing an active role in promoting species diversity (Hewitt 2000; Knowles 2000, 2009; Steele & Storfer 2006; Currat *et al.* 2008; Gray *et al.* 2008; Moyle *et al.* 2009). Moreover, with the current pace of climate change, there is a poignant urgency for such information.

The genetic consequences of distributional shifts are defined by a species-specific demographic and geographic context (Slatkin 1977; Whitlock & McCauley 1990; Wakeley & Aliacar 2001; Pannell 2003; Currat & Excoffier 2004; Edmonds *et al.* 2004; Wegmann *et al.* 2006; Sork *et al.* 2010). Yet, most genetic studies on the effect of distribution shifts do not incorporate species-specific demographic and geographic parameters, which are the factors that ultimately determine how genetic variation is distributed across a species' range (e.g. rates and directions of movement will differ spatially and temporally depending on the environment). Instead empirical studies have relied heavily on generic and non-spatially explicit models of demographic expansion (e.g. models with large expansions in panmictic populations: Slatkin & Hudson 1991; Rogers & Harpending 1992). We note that there is an extensive literature on the genetics of spatially expanding populations; however, these are almost exclusively simulation studies with hypothetical landscapes (for an excellent review see Excoffier *et al.* 2009).

The lack of an explicit spatial context for studying the demographic properties associated with distributional shifts in empirical systems stands in stark contrast to the work on describing genetic variation across the landscape (reviewed in Storfer *et al.* 2007). This difference cannot be explained by the lack of tools for generating expectations for patterns of genetic diversity from spatially explicit demographic models. For example, it is now possible to generate patterns of genetic diversity for populations expanding across spatially (as well as temporally) heterogeneous landscapes (Currat *et al.* 2004; Wegmann *et al.* 2006; Excoffier *et al.* 2009). What often constrains the construction of more realistic models is insufficient information for generating biologically informed models (Knowles 2009). In the context of studying the genetic consequences of distributional shifts, we generally lack information on how migration rates or population densities vary across the landscape. Here we concentrate on how this gap in knowledge might be bridged.

Our focus is on the utility of coupling distributional information from ecological-niche modelling with coalescent simulations to study the genetic consequences of the colonization process when species experience dramatic shifts in distribution. Specifically, we investigate

the colonization process associated with the climate-induced distributional shifts caused by the Pleistocene glaciations. Ecological-niche models (ENMs) have been used to inform genetic models (e.g. Carstens & Richards 2007; Knowles *et al.* 2007; Richards *et al.* 2007; Carnaval *et al.* 2009; Buckley *et al.* 2010), including identifying areas of habitat stability and/or the location of putative refugia. Unlike these previous treatments, here the predictions from ENMs are used to generate a spatially explicit demographic model, which then forms the basis for genetic simulations across the landscape, resulting in an explicit quantitative link between the spatial environmental information and population history (Fig. 1). Moreover, by transcribing the information on the predicted habitat suitabilities derived from the ENMs into a demographic model, not only can models that accommodate environmental heterogeneity be generated, but these models will also be informed by species-specific habitat requirements. As the focus of this study, we illustrate this approach to model generation with montane flightless grasshoppers. We highlight the caveats with translating the information extracted from ENMs into the relevant demographic parameters for studying the genetic consequences of distributional shifts in response to rapid climate change. These models are used to explore how key population demographic parameters and degrees of environmental heterogeneity can impact patterns of genetic variation. As taxa that show strict habitat requirements—the flightless grasshoppers are restricted to the sky-islands of the Rocky Mountains—other aspects of the colonization process (i.e. the demography associated with expansion from past refugial populations and founding of contemporary sky-island populations) are also explored with the simulations. These include variation in migration routes from glacial refugia to sky islands, as well as the duration of expansion (e.g. the genetic consequences when expansion stops when all the sky-island populations are founded, as opposed to continued expansion after colonization).

This exploration demonstrates how additional realism about climate-induced distributional shifts might be incorporated via ENMs to generate suites of models for studying the genetic consequences of such shifts. However, the generation of models is just the first step. Statistical testing of alternative models, which includes estimation of relevant parameters, is required to distinguish among alternative scenarios. For example, our simulations indicate that significant genetic differentiation among contemporary sky-island grasshopper populations can be generated as a consequence of the demography of the expansion process across a heterogeneous environment without long-term isolation among refugial populations, thereby identifying a

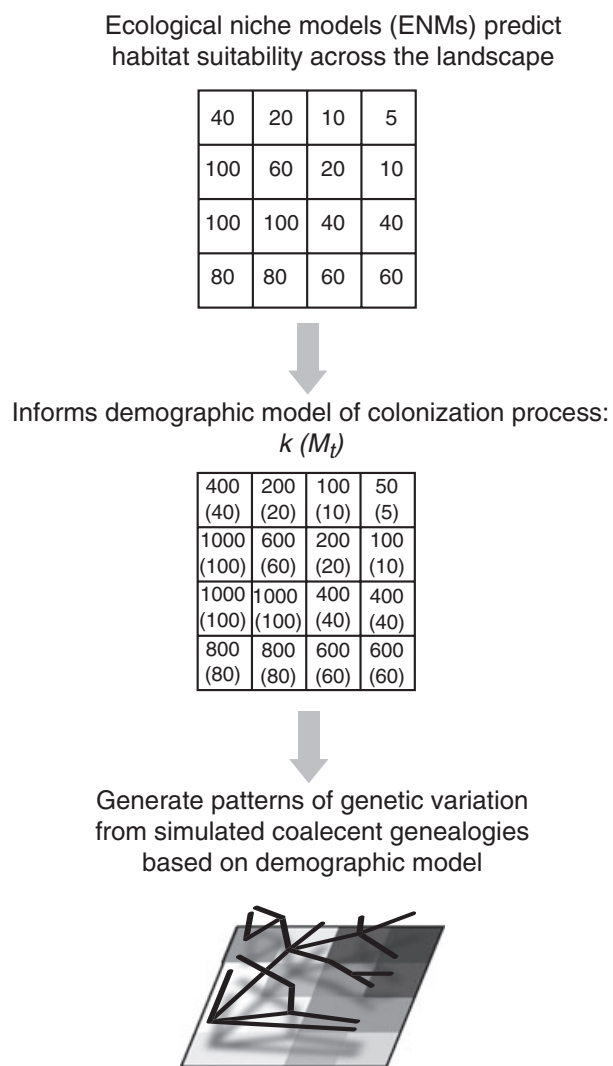


Fig. 1 Flow chart describing how environmental information extracted from ENMs is used to generate a demographic model that captures both spatial and temporal landscape heterogeneity, which then is used to predict patterns of genetic variation. This approach takes into account the impact of landscape heterogeneity on the colonization process (via the influence on population densities and migration rates), and hence generates models for specifying the expected patterns of genetic variation that can be used to study the genetic consequences of distributional shifts. This flow of information establishes direct links between process and pattern; shown across the grid are the predicted habitat suitabilities, the demographic parameters, and the distribution of gene lineage coalescence (darker shadings correspond to concentration of coalescent events). This approach contrasts with those that go from pattern to process, whereby the distribution of genetic variation across the landscape is described and then specific processes that might have been important in generating the patterns of genetic variation are inferred, the limitations of which have been extensively reviewed (see Knowles 2009; Cushman & Landguth 2010; Epperson *et al.* 2010).

potential class of models that should be considered to understand what processes underlie the observed patterns of genetic variation. To distinguish whether the founding of sky-island populations from allopatric glacial refugia vs. the demography of the colonization process itself best explains the high genetic differentiation observed among the grasshopper populations requires a statistical framework for evaluating the probabilities of these scenarios. We discuss how such tests might be conducted and focus specifically on the merits of approximate Bayesian computation (ABC; see Beaumont *et al.* 2002, 2010) as a likely candidate. This approach can accommodate the complexities associated with the biologically informed tempo-spatially explicit models that are generated from the approach we propose, as highlighted by the exploration of possible models generated for the grasshoppers. With further developments (i.e. incorporation of the proposed approach for model generation, Fig. 1, into a framework for parameter estimation and hypothesis testing), the insights that come from taking into account temporal and spatial heterogeneity of the environment should advance our understanding of the genetic consequences of distributional shifts and improve the accuracy of landscape-genetic methods.

Methods

Collection localities for *Melanoplus oregonensis* (a short-wing, flightless grasshopper; Melanoplinae, Acrididae) were compiled from previous fieldwork (summarized in Knowles & Carstens 2007) and from specimens housed at the Insect Division of the University of Michigan Museum of Zoology. All localities were geo-referenced following a point method (Wieczorek *et al.* 2004) with the aid of the USGS' Geographic Name Information System Database (USGS 2009). Only localities with GPS coordinates or detailed locality descriptions were used to reduce potential geo-referencing errors. This selection procedure resulted in a total of 57 unique localities of the species presence across the entire range of the species. These 57 localities were then randomly split into training and testing sets (90% and 10% of the localities, respectively) to generate predictive distribution models under a Maximum Entropy Algorithm in the program MAXENT v. 3.3.1 (Phillips *et al.* 2006); robustness of the model results was confirmed by cross-validation of 50 separate runs with randomly chosen training and testing sets. The choice of MAXENT is based on the results of studies that compared the performance of different algorithms (see Elith *et al.* 2006; Hernandez *et al.* 2006), but the approach we present could be implemented using other approaches for generating ENMs, including mechanistic niche modelling that utilizes

detailed biological information about species (e.g. Kearney & Porter 2009).

*ENMs generated for *M. oregonensis**

The distribution of the species was modelled for two time periods: current conditions (0 years ago) and the last glacial maximum (LGM; 21 000 years ago). The specific parameters used for the models (see Table S1, Supporting information) were chosen after sensitivity analyses of parameter values and overall model performance based on prediction accuracy. Fourteen bioclimatic variables (specific to each time period) at a 30 arc-second resolution (≈ 1 km) from the Worldclim database (Hijmans *et al.* 2005) were used to generate the models. These variables summarize climatic conditions based on patterns of temperature, precipitation, and seasonality (Table S2, Supporting information) and were chosen for consistency between the two time periods, despite correlations among the variables. ENMs generated with the entire set of 14 variables vs. those generated with the exclusion of highly correlated variables resulted in qualitatively similar ENMs for each of the separate time periods; the pattern of correlation among the variables was not the same for current climatic conditions and those of the LGM. The set of variables for current conditions was derived through interpolation of climatic records between 1950 and 2000 from weather stations around the globe (Hijmans *et al.* 2005), whereas the LGM set was derived from the Paleoclimate Modelling Intercomparison Project Phase II (Braconnot *et al.* 2007), and generously provided by R. Hijmans. The modelled distributions of *M. oregonensis* for the time periods were standardized for comparability; cell-specific habitat suitabilities from the two ENMs (i.e. for current conditions and the LGM) were rescaled such that the highest probabilities in each model scaled to 100% probability.

We note that as the basis for informing the demographic simulations, and hence the spatially explicit genetic model, accuracy of the ENM is crucial. Consequently, implementation of our proposed procedure should also consider the importance of algorithm performance, sample size, and geo-referencing imprecision on ENMs (see Guisan *et al.* 2007; Stockwell, Peterson, 2002; Phillips & Dudik 2008; Wisz *et al.* 2008). Likewise, as implemented here, we are assuming that the species niches and dispersal capabilities have not changed over time. Given that closely related taxa (Knowles 2000; Carstens & Knowles 2007b) all show similar habitats (i.e. are restricted to montane habitats) and are flightless, this assumption is reasonable. However, it may or may not be reasonable in other taxa (e.g. Phillips *et al.* 2008; Wiens *et al.* 2009) and should be considered.

Additional information useful for projecting species distributions (e.g. information on vegetation layers) might also be incorporated to improve the accuracy of the ENMs.

Integrating ENM information into the demographic simulations

The predicted distributions for the two time periods from the ENMs were then used to perform spatially-explicit demographic and genetic simulations for generating patterns of molecular variation that incorporate environmental information (using the program SPLATCHE; Currat *et al.* 2004). First, the demographic and migration history was simulated based on the information from the distributional projections from the ENM. Specifically, under this approach a time-forward demographic colonization process was simulated according to migration and density limitations imposed by the environment. Environmental information was represented by two variables across the defined geographic area: the friction to movement of individuals and the carrying capacity of each individual cell across the grid, where the value of individual cells differs (Currat *et al.* 2004). The relative values of the parameters for each cell were based on the predicted habitat suitability derived from the ENM generated with MAXENT. To generate patterns of genetic diversity, coalescent simulations were then conducted based on the simulation of the demographic history (Excoffier *et al.* 2000; Currat *et al.* 2004). The parameter values used for both the demographic and coalescent simulations are summarized in Table S3 (Supporting information) and details about their derivation are described below. A range of parameter values was explored with the simulations to highlight how changes in key demographic parameters and environmental characteristics could result in greater or lesser degrees of genetic differentiation among populations (see discussion for how such parameters might be estimated and hypothesized historical scenarios tested for any given species).

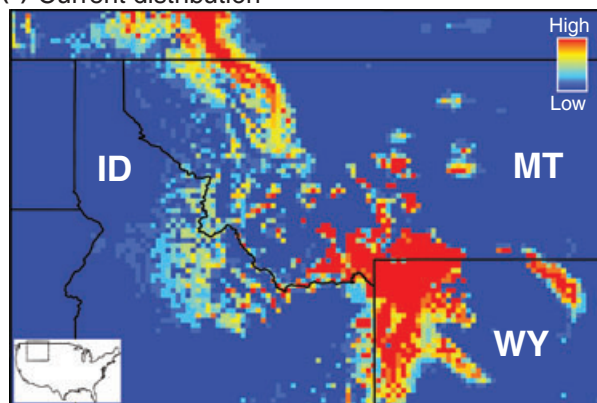
Although the ENMs provide predictions for both past and current species distributions, a combination of the projections (i.e. a composite history) is needed to explore the genetic consequences of the colonization process that integrates this environmental information. Because the ENMs produce discrete time-specific projections, whereas the colonization process is continuous through time, there is some unavoidable arbitrariness involved in combining the information from the current and past ENMs. We recognize this is not ideal. However, integrating ENM information into the demographic and genetic simulations introduces relevant details (e.g. areas differ in their suitability and some

areas have been more or less habitable across time), which is arguably preferable to ignoring biological realities (e.g. if a random diffusion process was used in the demographic and coalescent models that would treat all space as ecologically equivalent). Two different approaches were explored for generating a map with combined cell-specific habitat suitabilities for the two time periods.

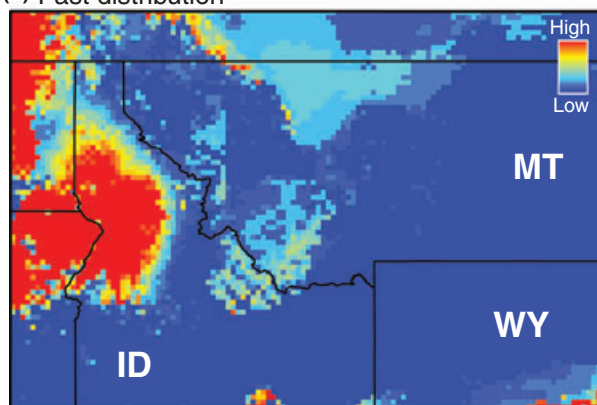
In the first approach (referred to hereafter as the composite method) the combined information from the current and LGM suitability scores from the respective ENMs was calculated as a cell-specific prediction—specifically, the combined prediction = $(0.75 \times \text{current prediction}) + (0.25 \times \text{LGM prediction})$. The Raster Calculator tool in ArcGis 9.3 (ESRI 2008) was used to generate the cell-specific combined predictions. Although other weightings could obviously have been chosen, we used this formula to emphasize the impact of the current distribution of *M. oregonensis* on the colonization process, while incorporating information on the refugial distribution during the LGM. Note that there is little to no overlap in the geographic regions predicted to be highly suitable between the two time periods (i.e. no areas of long-term habitat stability; Fig. 2). Weighting the projections based on the current climatic conditions more heavily increased the likelihood that migration route from refugial to contemporary sky-island populations followed a track in which the area of highest probability corresponded to the contemporary distribution of the grasshoppers across the sky islands. This path of movement is again consistent with the natural history of the species, as well as the ENMs for the current period; *M. oregonensis* is flightless and restricted to the sky islands, or mountaintops of the Northern Rocky Mountains.

For the second approach (referred to hereafter as the partitioned method) the predictions from the two time periods were independently used to inform the demographic simulations. Specifically, the movement of individuals across demes (i.e. the cell-specific friction values) was based on the LGM prediction whereas the carrying capacity was determined by the predictions from the ENM for the current conditions. The primary difference between this approach for combining the predictions from the two time periods and the preceding one is that the migration route is more diffuse initially (as evident from the visualizations of the colonization process; see Fig. 3) because of the low resistance to individual movement in the regions surrounding the refugia (as determined by the ENM for the LGM refugia with the partitioned method). This partitioned model is included here because it offers some insights about the genetic consequences of admixture early in the colonization process

(a) Current distribution



(b) Past distribution



(c) Combined distribution

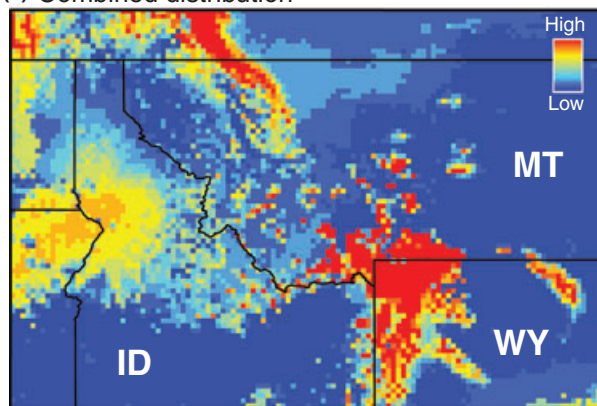


Fig. 2 Projected distributions of *M. oregonensis* for (a) current conditions, (b) past conditions (i.e. LGM), and (c) the combined history (the results from the composite method are shown); all maps were generated from ENMs with the program MAXENT and have been standardized to the same scale (see text for details).

cess (i.e. emigration is more diffuse than directional early in the partitioned model), and not because we think it is necessarily an accurate representation of the postglacial history of this species.

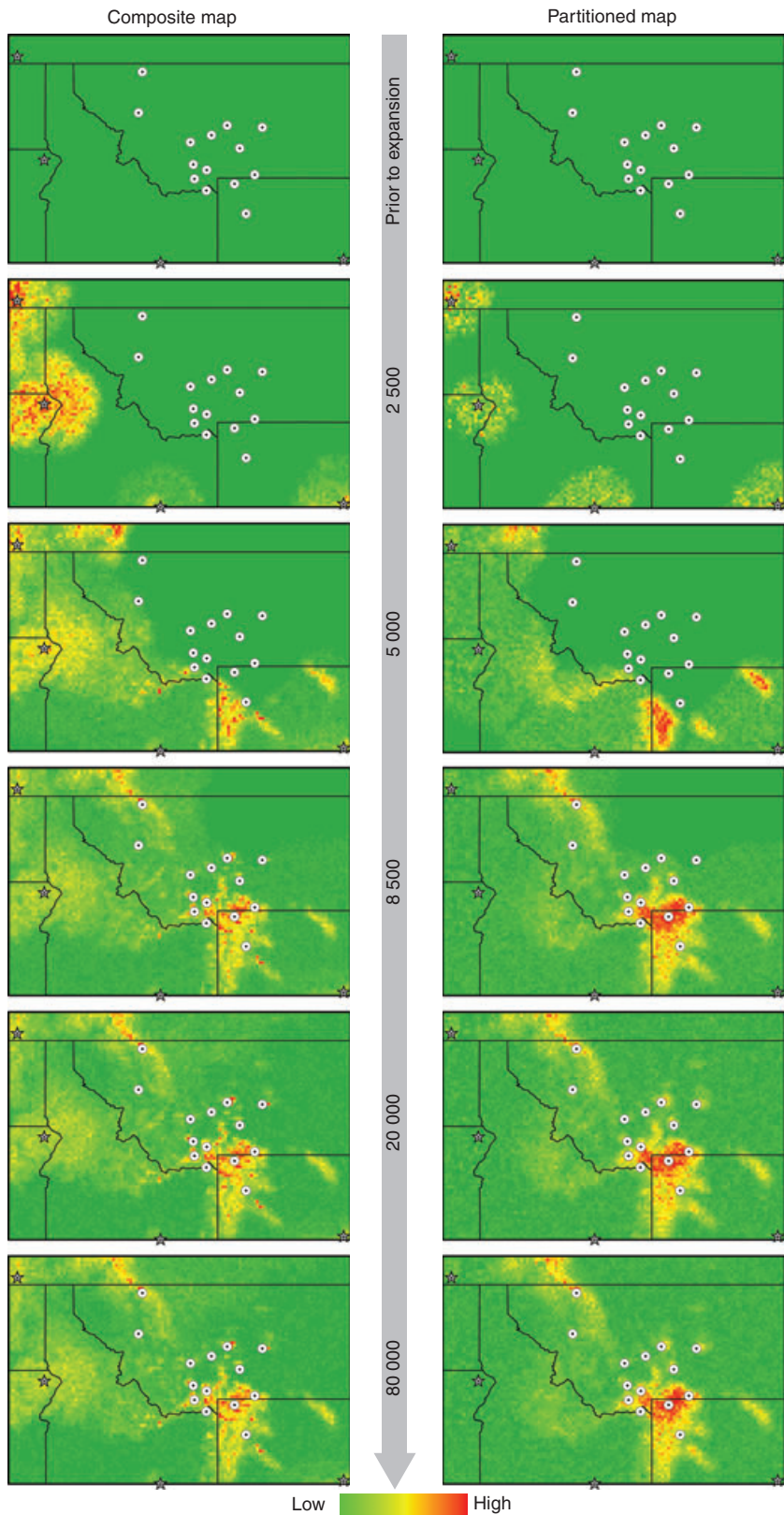


Fig. 3 Schematic of the demographic history (specifically, models in class A; see Table 1 for details) that shows the expansion from glacial refugia that were identified from the ENMs (marked with grey stars) and colonization of the contemporary sky-island populations (marked with white dots). Green colors in the maps signify cells with low numbers of emigrants, whereas cells contributing high numbers of emigrants are shown in red. Results for the two approaches of combining the information from the ENMs for current conditions and the LGM are shown. The figures on the left correspond to the composite model in which the cell-specific habitat suitabilities are added between the time periods. The figures on the right are the results from the partitioned model in which the ENM model for current conditions informs the cell-specific carrying capacities and the ENM of the LGM informs the migration history (see text for details). Different points in the colonization history are shown to trace the migration history across the landscape (i.e. at 2.5, 5.0, 8.5, 20.0 and 80.0 thousand years after the initiation of expansion). Details about the parameters used in the demographic model and the simulations with the program SPLATCHE are given in Tables 1 and S3.

Potential source (i.e. refugial) populations were identified from the ENM for the LGM. For both the composite and partitioned method, a refugial population was identified as any area with 25 contiguous cells (i.e. 25 km²) with predicted habitat suitabilities above the 75th percentile separated from other similarly highly habitable areas by 25 or more cells (regions with 75% predicted suitability separated by less than 25 km were treated as a single refugial population). Using the 75th percentile as the basis for identifying refugia produced four identified refugia (Fig. 2); with higher thresholds there were not contiguous areas (i.e. 25 cells) of high predicted suitability and with thresholds less than 75% there were not discrete regions of high vs. low predicted suitability. Hence, while admittedly arbitrary, to the extent that the grasshopper populations were indeed displaced into multiple allopatric refugia, the criterion recovered probable refugial sources (we discuss the caveats surrounding this criterion for identifying refugial populations in the discussion, as well as note that alternative means for modelling refugial sources could easily be accommodated within the proposed framework). For the demographic model, the central area of a refuge was used as the point of geographic origin from which the expansion process would begin. Other approaches for modelling expansion for the refugia could be accommodated; this approach was chosen such that the effect of differentiation among allopatric refugia could be explored specifically. For example, expansion could be modelled based on the cell-specific predicted habitat suitabilities, as opposed to the approach implemented here.

For both the composite and partitioned methods, the cell-specific predicted habitat suitabilities from the ENMs had to be transformed before they could be used to inform the demographic simulations generated in the program *SPLATCHE* (Curat *et al.* 2004). The grids had to be rescaled to a 0.1° resolution because computational limitations restricts the number of cells for which data can be simulated; this rescaling ensured that the entire study area (i.e. geographic distribution of the species) could be represented in the simulations. Because of this rescaling, the movement of the flightless grasshoppers from one cell to the next across the landscape (i.e. 10 km² per cell) is very unlikely to occur in one generation. Therefore the time until colonization was rescaled assuming that the grasshoppers will move 10 km² within 10 generations. The particular value used for rescaling will differ in a species-specific manner and should be considered when modelling the colonization process (i.e. some species may take fewer or more generations to move 10 km²).

Lastly, the cell-specific scores from the composite and partitioned methods were transformed into categorical

variables for ease of computation. Specifically, the continuous scale of habitat suitabilities (ranging from 0% to 100%) was divided into 20 variables in increments of 5% of predicted suitability, which were then used for the demographic simulations in the program *SPLATCHE*. The predicted habitat suitability from the ENMs was used to inform two key parameters of the demographic model— k and M , the per deme maximum population density (or carrying capacity) and the number of emigrants, respectively (see Table S3, Supporting information) for additional details about model settings). The values of k varied across the landscape as a function of the predicted habitat suitabilities. The values for k ranged from a maximum number of individual for cells (demes) with 100% suitability scores to a minimum number of individuals for those demes with 5% or less predicted suitability, depending on the k_{\max} (either 200, 1000, or 5000) and the 'coarseness' of the environmental heterogeneity (i.e. the number of discrete k -values modelled) in any single simulation (see Table 1 for details). The migration rate, m , was also varied to explore the effect of this parameter on the patterns of genetic differentiation, and ranged from an m of 0.02, 0.1 or 0.5 (see Table 1). For example, with an m set to 0.1, the number of emigrants in a particular generation, M_t , was 0.1% of the population density of the deme at that generation, N_t , and the allocation of emigrants among the surrounding cells differed according to the cell-specific predicted habitat suitabilities from the ENMs in a manner analogous to the friction-values for defining spatial movement (McRae 2006). For example, if the population densities and cell-specific friction values were uniform across the surrounding demes, the same number of emigrants from a specific deme migrated into the surrounding demes. Otherwise, the number of immigrants the surrounding cells received will differ proportionally according to the suitability of the surrounding habitat (i.e. the cell-specific suitabilities). In both cases, the total number of emigrants, M_t , from a cell always equalled $N_t m$ for a deme. In addition to k and m , variation in a third key demographic parameter— r , the logistic growth rate—was explored; specifically, r was varied from 0.02, 0.1, and 0.5.

For each of the two approaches for combining the information from the ENMs on the current and past distribution of *M. oregonensis* (i.e. both the composite and partitioned methods), two scenarios that differed in the initial degree of genetic differentiation among the refugial source populations were explored. To generate differing amounts of standing variation among the refugial populations, the data were simulated with the refugial populations isolated for either 10 000 or 100 000 generations (i.e. $\tau = 10\,000$ and $100\,000$ in *SPLATCHE*; Curat *et al.* 2004) before the demographic expansion.

Table 1 Different demographic models explored with 10 data sets simulated per set of conditions for the combined and partitioned models (i.e. the two schemes used to combine information from the ENMs for current and past conditions; see text for details). Models are grouped according to common parameter values (e.g. all parameters were held at similar values to those models in class A) except for the one parameter of interest (e.g. k_{\max} differed between models in class B compared to those in class A to examine how increasing or decreasing k_{\max} impacted the average degree of genetic differentiation among populations). Differences in the ‘coarseness’ of the environment (see Methods for details) are represented by the number of categories that cells were grouped into based on their respective predicted habitat suitability scores from the ENMs (e.g. grouping cells into 20 categories produces a more coarse environment as opposed to 40 categories that would capture more fine-scale differences in habitat suitability). Note that k was rescaled for each deme according to the deme-specific habitat suitability scores from the ENMs (see text for details), whereas k values were applied equivalently across the landscape under the homogeneous grid parameter setting. Colonization times differed across the models because the time to founding differed depending on the demographic parameter values; the effect of post-founding migration was examined for model A (see Figs 3–5). Average F_{st} values among the populations across the 10 replicate data sets are shown (standard deviations are in parentheses). The different demographic scenarios were also examined for two levels of initial differentiation among the source populations generated with a τ of 10 000 and 100 000 generations, respectively (τ is the number of generations the refugial source populations were isolated before the onset of expansion), which translates into a mean F_{st} of 0.334 vs. 0.883 (standard deviation of 0.120 and 0.026), respectively, among the four refugial populations averaged across 10 genetic simulations

Model class	Demographic parameters				Grid parameters		Colonization time		Mean F_{st} (SD)	
	τ	k_{\max}	m	r	Categories	Grid type	Combined	Partitioned	Combined	Partitioned
A-1	0	1000	0.1	0.1	20	ENM	850	900	0.237 (0.099)	0.265 (0.084)
A-2	10 000	1000	0.1	0.1	20	ENM	850	900	0.375 (0.063)	0.319 (0.074)
A-3	100 000	1000	0.1	0.1	20	ENM	850	900	0.482 (0.031)	0.398 (0.078)
A-4	0	1000	0.1	0.1	20	Homogeneous	650	650	0.184 (0.074)	0.164 (0.045)
A-5	10 000	1000	0.1	0.1	20	Homogeneous	650	650	0.263 (0.061)	0.265 (0.072)
A-6	100 000	1000	0.1	0.1	20	Homogeneous	650	650	0.424 (0.030)	0.453 (0.012)
B-1	10 000	5000	0.1	0.1	20	ENM	700	750	0.244 (0.063)	0.272 (0.075)
B-2	100 000	5000	0.1	0.1	20	ENM	700	750	0.390 (0.041)	0.327 (0.024)
B-3	10 000	5000	0.1	0.1	20	Homogeneous	600	600	0.175 (0.029)	0.162 (0.032)
B-4	100 000	5000	0.1	0.1	20	Homogeneous	600	600	0.345 (0.032)	0.346 (0.036)
C-1	10 000	200	0.1	0.1	20	ENM	1700	2250	0.466 (0.053)	0.393 (0.078)
C-2	100 000	200	0.1	0.1	20	ENM	1750	2450	0.549 (0.030)	0.445 (0.042)
C-3	10 000	200	0.1	0.1	20	Homogeneous	850	850	0.338 (0.062)	0.390 (0.046)
C-4	100 000	200	0.1	0.1	20	Homogeneous	800	850	0.486 (0.025)	0.466 (0.031)
D-1	10 000	1000	0.5	0.1	20	ENM	350	350	0.187 (0.076)	0.174 (0.085)
D-2	100 000	1000	0.5	0.1	20	ENM	350	350	0.396 (0.026)	0.372 (0.019)
D-3	10 000	1000	0.5	0.1	20	Homogeneous	300	300	0.156 (0.035)	0.171 (0.058)
D-4	100 000	1000	0.5	0.1	20	Homogeneous	300	300	0.400 (0.022)	0.370 (0.016)
E-1	10 000	1000	0.02	0.1	20	ENM	2850	3350	0.490 (0.023)	0.459 (0.052)
E-2	100 000	1000	0.02	0.1	20	ENM	2850	3350	0.619 (0.051)	0.488 (0.076)
E-3	10 000	1000	0.02	0.1	20	Homogeneous	1600	1600	0.406 (0.059)	0.395 (0.054)
E-4	100 000	1000	0.02	0.1	20	Homogeneous	1550	1550	0.544 (0.042)	0.505 (0.052)
F-1	10 000	5000	0.02	0.1	20	ENM	1600	1700	0.381 (0.087)	0.422 (0.086)
F-2	100 000	5000	0.02	0.1	20	ENM	1600	1650	0.547 (0.048)	0.510 (0.056)
F-3	10 000	5000	0.02	0.1	20	Homogeneous	1300	1300	0.330 (0.031)	0.277 (0.046)
F-4	100 000	5000	0.02	0.1	20	Homogeneous	1300	1300	0.475 (0.029)	0.440 (0.047)
G-1	10 000	200	0.5	0.1	20	ENM	400	450	0.312 (0.047)	0.237 (0.056)
G-2	100 000	200	0.5	0.1	20	ENM	450	450	0.341 (0.045)	0.340 (0.033)
G-3	10 000	200	0.5	0.1	20	Homogeneous	350	350	0.262 (0.048)	0.257 (0.059)
G-4	100 000	200	0.5	0.1	20	Homogeneous	350	350	0.377 (0.013)	0.438 (0.013)
H-1	10 000	1000	0.1	0.5	20	ENM	400	450	0.430 (0.069)	0.465 (0.068)
H-2	100 000	1000	0.1	0.5	20	ENM	400	450	0.478 (0.049)	0.497 (0.126)
I-1	10 000	1000	0.1	0.2	20	ENM	1950	2000	0.312 (0.044)	0.276 (0.043)
I-2	100 000	1000	0.1	0.2	20	ENM	2000	2050	0.335 (0.039)	0.348 (0.031)
J-1	10 000	1000	0.1	0.1	40	ENM	1000	1050	0.295 (0.085)	0.313 (0.051)
J-2	100 000	1000	0.1	0.1	40	ENM	1000	1100	0.453 (0.039)	0.432 (0.015)
K-1	10 000	1000	0.1	0.1	10	ENM	800	800	0.265 (0.074)	0.314 (0.050)
K-2	100 000	1000	0.1	0.1	10	ENM	800	800	0.451 (0.033)	0.500 (0.023)

These differences translated into a mean F_{st} of 0.334 vs. 0.883 among the four refugial populations averaged across 10 genetic simulations (standard deviation of 0.120 and 0.026, respectively) for a τ of 10 000 and 100 000 generations, respectively.

Genetic simulations and characterization of genetic differentiation

The record of population sizes and migration rates (per deme) from the demographic simulation was used to simulate genetic data under a coalescent framework (for additional details see Excoffier *et al.* 2000; Currat *et al.* 2004). Specifically, the probability of coalescence of genes in a deme at time t is computed according to the population size, N_t , of a given deme at time t from the demographic simulations. Likewise, the probability for a gene in a deme i to migrate to deme j is calculated based on the number of emigrants from deme i to deme j at time t from the demographic simulation. Deme-specific parameters in this case are the cell-specific values from the demographic simulation across the geographic landscape and were explored for a range of parameter values (see discussion for an approach for estimating parameters and testing the probability of alternative models).

DNA sequences were simulated on the gene genealogies from the coalescent simulations (i.e. coalescent and mutational processes are simulated separately). Mutations are randomly assigned to the branches in the gene genealogy according to a Poisson distribution, where the number of mutations is proportional to the branch lengths (measured in t generations) in the coalescent genealogy at a per generation mutation rate of μ (Excoffier *et al.* 2000). Details on the mutational model used to generate patterns of genetic variation are given in Table S3 (Supporting information).

Genetic diversity for each simulation scenario was assessed both at the refugial source populations and at each geographic site corresponding with each of the 15 sky-island populations (i.e. the geographic coordinates for the demes that correspond to the location of actual *M. oregonensis* populations; for locality details see appendix in Knowles & Carstens 2007). F_{st} values were calculated to examine the patterns of differentiation among sky-island populations, as well as among refugial source populations using the program ARLEQUIN v. 3.01 (Excoffier *et al.* 2005). For each of the models, genetic diversities were assayed when each of the 15 sky-island populations had been founded. To explore the effect of post-founding migration, genetic diversities were also assayed at three different times (i.e. at 8.5, 20, and 80 thousand years) after the initiation of expansion in the sky-island populations for one set of models (i.e.

model A: $k_{max} = 1000$, $m = 0.1$, and $r = 0.1$; see Table 1). These different time points correspond to differing migration histories for the 15 *M. oregonensis* populations. At 8500 years the sky-island populations have all just been founded (based on the demographic parameters used in model A to simulate the colonization process). Beyond 8500 years the probability of additional emigration and mixing among the sky-island populations increases. Consequently, with this sampling regime we can evaluate how genetic differentiation is impacted by the colonization event itself (i.e. the founding of sky-islands), as well as post-colonization history (i.e. potential admixture among sky-islands); distinguishing between these scenarios require estimation of model parameters (see discussion for an approach that might be used to estimate parameters and distinguish among hypothesized historical scenarios).

Results and discussion

As with other species distributed in the Rocky Mountains, and high latitudes in general, the projected distributions of *M. oregonensis* from the ENMs confirm other studies on the pronounced impact the Pleistocene glaciations had on species distributions (e.g. Pielou 1991; Weir & Schluter 2004; Provan & Bennett 2008). The ENMs highlight the extreme spatial and temporal heterogeneity in the habitability of areas across the geographic range of *M. oregonensis* (Fig. 2). Consequently, not only does the habitat suitability differ substantially across the landscape, but there is little to no overlap in the projected distribution of contemporary populations compared to past when the species was displaced into glacial refugia.

The lack of habitat stability in *M. oregonensis*, as with other species distributed in temperate regions (e.g. Bennett *et al.* 1991; Comes & Kadereit 1998; Taberlet *et al.* 1998; Petit *et al.* 2002), emphasizes that patterns of genetic variation will reflect both the impact of the current landscape on population structure and past distributional configurations. Moreover, the temporal and spatial heterogeneity in the predicted habitat suitabilities in the montane grasshoppers highlights the importance of considering how the demography of the colonization process is expected to differ across the landscape—that is, different areas are expected to vary in the probability of population densities and migration routes, and consequently need to be taken into account when exploring the genetic consequences of distributional shifts. Hence, this is an empirical setting where there is a critical need for biologically informed genetic models (Knowles 2009) that are not just spatially explicit (reviewed in Excoffier *et al.* 2009), but are also spatio-temporally explicit.

Demography of the colonization process in a heterogeneous environment

Incorporating the potential influence of environmental heterogeneity on the demographic history of *M. oregonensis* produces a demographic expansion that proceeds from refugial populations towards the sky-island populations (Fig. 3), as opposed to a random diffusion process (i.e. a landscape with equal population densities and probabilities of migration across the species distribution). Because of the environmental heterogeneity (Fig. 2), the population densities and migration probabilities differ across the landscape and over time. Consequently, rather than a simple wave of expansion (see Excoffier *et al.* 2009), movement from the refugial populations shows a spatial pattern in which some demes contribute higher numbers of emigrants (shown in red and yellow, Fig. 3), whereas others do not (shown in light green; Fig. 3) creating localized paths of migration (McRae 2006). The geographic concentration of population movements (i.e. specific migration routes) has been inferred for a variety of taxa, and independently verified in some cases with fossil and/or archeological or historical data (e.g. Comes & Kadereit 1998; Taberlet *et al.* 1998; Petit *et al.* 2002; Currat & Excoffier 2004; Zellmer & Knowles 2009; Sork *et al.* 2010).

A general pattern of local concentrations of migration (Fig. 3), as opposed to a continuous expanding wave front, is apparent in both the composite and partitioned methods for combining the information on current and past conditions. Yet, if m as a fraction of population size, or the carrying capacity, k , becomes sufficiently large, genetic variation becomes effectively homogenized at a local geographic scale, and the realized heterogeneity in the environment is lessened, thus impacting the genetic consequences of the colonization process. The reduced impact of environmental heterogeneity is apparent in the lack of concentrated migration routes (in contrast to Fig. 3; see Supporting information S1 for movies contrasting the colonization dynamic), as well as by the patterns of differentiation observed among the founded populations under the differing demographic settings (Table 1). For example, the dampening effect of heterogeneity on genetic divergence for large k or m is apparent if we contrast the average F_{st} values for expansion across a homogeneous vs. heterogeneous landscape. Average F_{st} values increased on average by 0.112 when expansion occurred across a heterogeneous landscape contrasted with expansion across a homogeneous landscape (with $k_{max} = 1000$, $m = 0.1$, and $r = 0.1$; combined model A-2 vs. model A-5). However, when k_{max} is increased to 5000 or m is increased to 0.5, the difference in the

average F_{st} values between expansion across a heterogeneous vs. homogeneous landscape is significantly lessened (i.e. a difference of only 0.069 and 0.031 between the combined model B-1 vs. model B-3, and combined model D-1 vs. model D-3, respectively; Table 1). Theoretical work also highlights how environmental heterogeneity (i.e. variation in km across the landscape), not only the value of km , can exert a significant influence on spatial expansion dynamics (e.g. Wegmann *et al.* 2006). In fact, the impact of a heterogeneous environment on the expansion dynamics can be analogized as an expansion across a homogeneous environment, but with a much smaller km on average (reviewed in Excoffier *et al.* 2009).

The two methods (i.e. the composite and partitioned methods) for combining the information on current and past conditions differ in one, albeit rather subtle way. Emigration is fairly diffuse under the partitioned method, relative to the composite method, early in the history of the expansion (Fig. 3). This difference is because the number of emigrants, M , remains fairly uniformly low given that the ENM for current conditions translates into small carrying capacities, k , for demes in the region of the refugial source populations. With the low habitat suitability in the area immediately surrounding the refugial source populations (Fig. 2), even though there are few migrants, these few migrants diffuse evenly into the surrounding areas (Fig. 3) given the low friction values determined by the high suitability predicted from the ENM for the LGM. However, as the colonization process proceeds (i.e. at the later time points after the initiation of expansion; Fig. 3), the spatial patterns of emigration rates for the composite and partitioned methods become more similar. The emigration patterns resemble the carrying capacity map by 850 generations (i.e. at the founding of the sky-island populations) and stay similar afterwards in both methods. This similarity is because given enough time, all demes become filled with the maximum number of individuals as determined by the habitat suitability, such that the number of emigrants of each deme is also maximized ($M_{max} = km$).

Given the significant impact of spatial heterogeneity on expansion dynamics (Wegmann *et al.* 2006), the information contained in the ENMs for current conditions and the LGM become particularly critical as a means for studying the colonization process in species that have undergone distributional shifts. With the lack of habitat stability in the montane grasshoppers at this timescale, we represented the temporal heterogeneity as spatial heterogeneity. This decision was made because with two snap shots of what is a continuous process, and given there is little to no overlap between the current and past distribution of *M. oregonensis*

(Fig. 2), the integration of spatial information contained in the respective ENMs for current and LGM conditions becomes a way to inform the demographic model. This choice arguably also results in a more biologically informed model, as opposed to incorporating random temporal heterogeneity. Moreover, the general robustness of the pattern of spatial expansion irrespective of whether the composite or partitioned methods are used to combine information from the ENM for current and LGM conditions (Fig. 3) suggests that heterogeneity itself influences the rate and direction of the colonization process, as opposed to a specific route of movement being a primary determinant of the consequences of distributional shifts (at least for the conditions explored here). This suggests that when studying the genetic consequences of climate-induced distributional shifts, inferences are less dependent on specific reconstructions. Hence, assumptions and oversimplifications used to translate information about temporal heterogeneity from projected distributions at discrete time points (i.e. habitat suitability from ENM for current and the LGM) for what is admittedly a continuous process of expansion across time as individuals move across the landscape from glacial refugia to sky-island populations, may not greatly affect conclusions.

Further analyses, including more thorough characterization of the genetic consequences of the expansion process with additional summary statistics (see section on model testing and parameter estimation), are needed to determine how sensitive the results are to inferences about specific migration routes (or as may be the case, mis-specification of the routes). Of course, alternative approaches for accounting for temporal heterogeneity (or other assumptions used to generate the models) could be included among the set of models to be estimated (as described in the last section). Theoretical considerations also suggest the impact of temporal heterogeneity on expansion dynamics is not as great as the effects of spatial heterogeneity, and that there are not any characteristic genetic signatures specific to temporal heterogeneity *per se* (Wegmann *et al.* 2006). Nevertheless, if there were information on the extent to which variation in *km* is correlated (or not) across time, a more realistic demographic model might be implemented that captures temporal heterogeneity as parameter independent of spatial heterogeneity (e.g. Currat & Excoffier 2004). One intriguing aspect for future development of incorporating temporal heterogeneity in the context of our proposed method for coupling ENMs, demographic, and coalescent models, would be the potential to add bioclimatic layers (Table S2, Supporting information) for the intervening time intervals between the present and LGM.

Spatial genetic patterns after expansion from glacial refugia

The colonization process, as modelled here, generated significant differentiation among the *M. oregonensis* sky-island populations (Table 1), which is similar in magnitude with the quantification from empirical assays of genetic variation (see Knowles & Carstens 2007). Much of the differentiation appears to have been generated from expansion across a heterogeneous environment (Fig. 4; Wegmann *et al.* 2006; Excoffier & Ray 2008; Hoffer *et al.* 2009), as opposed to simply the partitioning of genetic variation from initially divergent refugial source populations. For example, expansion across an increasingly heterogeneous landscape resulted in increased average F_{st} values among the colonized populations even if the ancestral source populations were themselves not differentiated (Fig. 4). Moreover, the larger average

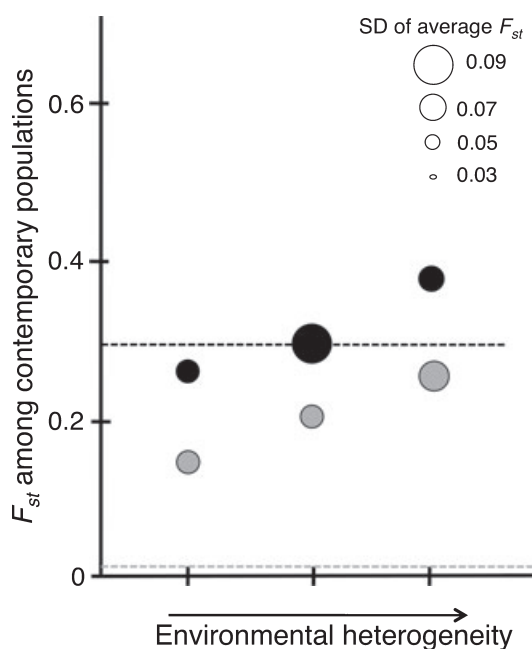


Fig. 4 Effect of increasing environmental heterogeneity on average genetic differentiation among the 15 colonized sky-island populations colonized from either genetically differentiated ancestral refugial populations (shown in black) or undifferentiated ancestral refugial source populations (shown in grey). Average F_{st} values, averaged across the 10 replicate data sets, are presented and the sizes of the circles correspond to the standard deviation of the F_{st} values. The black and grey dashed lines mark the level of genetic differentiation, or lack thereof, among the ancestral refugial populations that were source of founders of sky-island populations colonized from either genetically differentiated ancestral refugial populations (shown in black) or undifferentiated ancestral refugial source populations (shown in grey). See text and Table 1 for details about the categorizations of the ENMs used to generate differing levels of environmental heterogeneity.

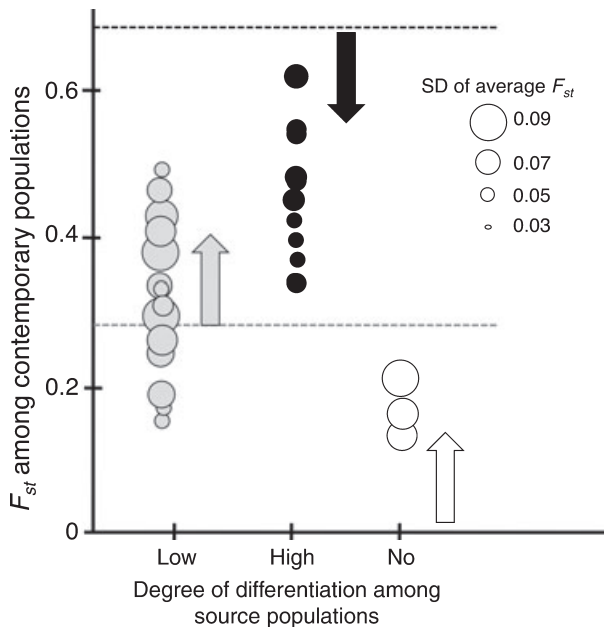


Fig. 5 Effect of the degree of divergence among refugial source populations on average genetic differentiation among the 15 colonized sky-island populations. All the different demographic conditions explored with the simulations (see Table 1 for details) are represented as grey, black, or white circles depending on the degree of differentiation among the source populations; average F_{st} values, averaged across the 10 replicate data sets, are presented and the sizes of the circles correspond to the standard deviation of the F_{st} values. Dashed lines mark the average F_{st} values among the ancestral refugial source populations and the arrows highlight whether founded populations exhibit higher or lower levels of genetic differentiation relative to that of the ancestral source populations.

F_{st} values among highly differentiated refugial source populations were not maintained among the colonized sky-island populations (Fig. 5). Instead, the average F_{st} values among the colonized sky-island populations are quite similar irrespective of the initial level of differentiation among the refugial source populations (i.e. average F_{st} of 0.334 vs. 0.883; Fig. 5). These results show that it is not necessary to have long-term isolation among ancestral source populations, nor do populations have to be founded from genetically distinct refugia to generate substantial differentiation, contrary to what has been intuited for these grasshoppers based on patterns of genetic variation (Knowles 2001; Carstens & Knowles 2007a). However, while these results highlight the potential of expansion across a heterogeneous landscape to generate genetic differentiation, they do not demonstrate whether such a model is more likely than one based on colonization from divergent ancestral refugial source populations. This distinction requires a framework to estimate the respective demographic parameters and probabilities of alternative historical scenarios.

As with the effect of expansion across a heterogeneous landscape, the degree of genetic differentiation will also differ depending on the specific details of the demographic process (Edmonds *et al.* 2004; Klopstein *et al.* 2006; Wegmann *et al.* 2006)—namely, the values of k , m and r . Comparing the effects of each demographic parameter in turn (Fig. 6) not only shows the general impact of specific parameter values, but also that each is capable of generating levels of genetic differentiation among the colonized populations that exceed the initial level of differentiation among the ancestral refugial source populations. However, there can also be a loss of genetic differentiation among the colonized populations relative to the ancestral sources depending on the specific parameter values. For example, if migration, m , becomes sufficiently large such that genetic variation becomes effectively homogenized at a local geographic scale (Fig. 6). Similar decreases in the degree of genetic differentiation are observed with increasing carrying capacities, k , with low F_{st} values among colonized populations when k_{max} is large (Fig. 6).

Depending on the duration of migration associated with colonization process, and specifically the opportunity for post-founding migration among sky-islands, the degree of population differentiation drops significantly (see F_{st} values for model A for time points after 8500 years, Table 2). The effect of post-founding migration is readily apparent in the coalescent genealogies (Fig. 7), which shows that gene lineages from specific populations become much more scattered across the tree with post-founding migration (i.e. when genetic diversities are assayed 80 000 years after the initial expansion). Likewise, the geographic distribution of coalescent events becomes more dispersed as the expansion proceeds over time (Fig. 8). However, even with an extended post-founding migration period, population genetic differentiation persists (Table 2). The slow decay of differentiation appears to be a consequence of the population demography where the values of Nm remain sufficiently small such that most gene lineages coalesce within the sky islands (i.e. there is still a concentration of gene lineage coalescence in the sky-island region; Fig. 8). With larger values of Nm , gene lineages are more likely to migrate out of demes before they coalesce (i.e. coalescent events will be older and concentrate in the region where the lineage originated) (reviewed in Excoffier *et al.* 2009). Because the demography of the expansion process directly impacts the time required for colonization (see Table 1), the period of time associated with the expansion process becomes an important constraint when evaluating the probability of alternative historical scenarios because some parameter values will be very

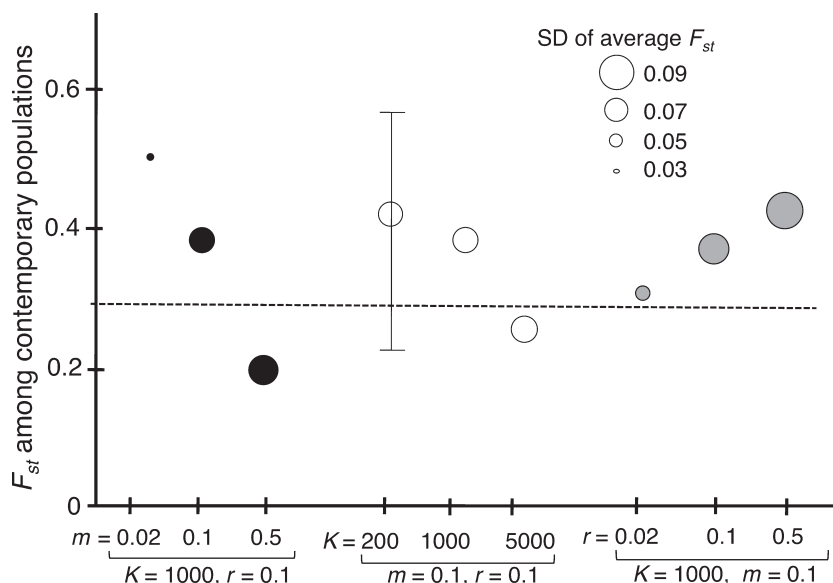


Fig. 6 Examination of the effect of specific demographic parameters by varying either m , k_{\max} or r while holding other parameters constant (shown below the brackets). Only results from the composite model with habitat heterogeneity held constant and a mid-level of differentiation among the ancestral refugial source populations (marked by the dashed line) are shown. Average F_{st} values, averaged across the 10 replicate data sets, are presented and the sizes of the circles correspond to the standard deviation of the F_{st} values, except for the model with $k_{\max} = 200$. For this case, an error bar denotes that there was a very high variance across simulated data sets because of the shift to an extinction-colonization dynamic when k_{\max} was small. Note that k and m was rescaled for each deme according to the deme-specific habitat suitability scores from the ENMs (see text for details).

Table 2 Mean F_{st} values (standard deviations are given in parentheses) among the sky-island populations averaged across the 10 replicates of the simulated genetic data for each of the different times of expansion. These include assaying genetic diversities in the sky islands after they were initially colonized (i.e. at 850 generations), as well as at two time points after their founding, which provided opportunities for inter-island migration (only the results from model class A are shown; see Table 1 for details). Each of these scenarios was also examined for two levels of initial differentiation among the source populations (i.e. a mean F_{st} of 0.334 vs. 0.883 among the four refugial populations averaged across 10 genetic simulations; standard deviation of 0.120 and 0.026, respectively) generated with a τ of 10 000 and 100 000 generations, respectively (τ is the number of generations the refugial source populations were isolated before the onset of expansion)

Time for expansion	Mean F_{st} values			
	$\tau = 10\ 000$		$\tau = 100\ 000$	
	Composite method	Partitioned method	Composite method	Partitioned method
850*	0.375 (0.063)	0.349 (0.074)	0.482 (0.031)	0.450 (0.078)
2000	0.154 (0.028)	0.178 (0.040)	0.304 (0.019)	0.281 (0.027)
8000	0.076 (0.005)	0.127 (0.017)	0.137 (0.024)	0.196 (0.018)

*Note that the time for expansion was 900 rather than 850 generations for colonization of the sky-island populations under the partitioned model (the additional generations were required for founding of all the sky-island populations).

unlikely (see section below on testing hypotheses and estimating parameters).

Lastly, the distribution of coalescences over time (i.e. the branch lengths) in the gene genealogies is also quite revealing about the underlying demographic process (Fig. 8). Most coalescent events are disproportionately distributed towards the tips of the tree when genetic

diversity is assayed right after the sky-island populations have been founded. This pattern is consistent with the signature of a demographic expansion (Excoffier *et al.* 2009; Knowles 2009). In contrast, if the simulations are continued after the initiation of expansion from the refugial source populations (Fig. 3), there is movement of individuals across space, and hence migration among

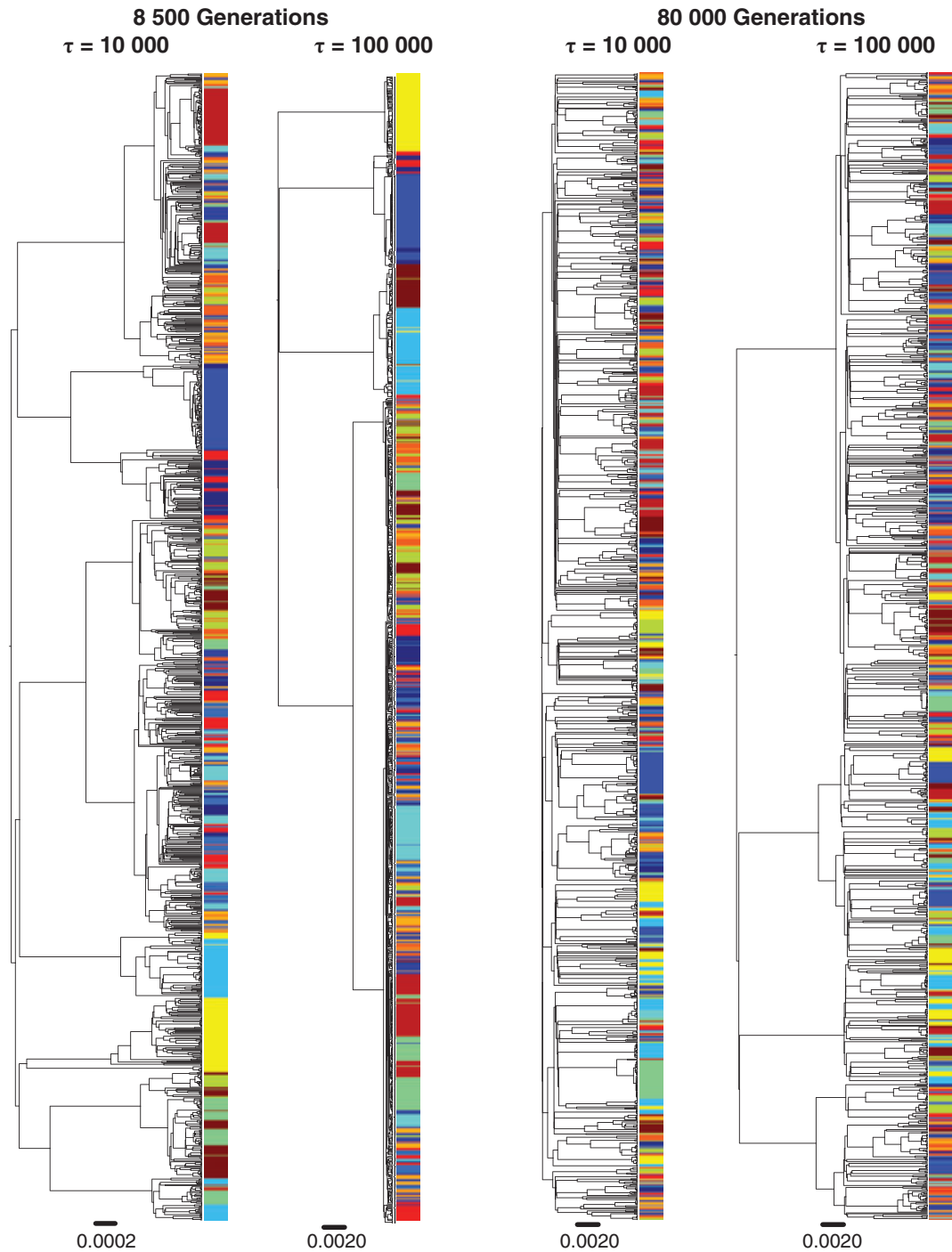


Fig. 7 One realization for the coalescent genealogies for the demographic models (specifically, models in class A; see Table 1) for which genetic variation was assayed at different time points after the initiation of expansion (i.e. 8.5 and 80.0 thousand years), and different degrees of initial differentiation among the refugial source populations (i.e. $\tau = 10\,000$ or $100\,000$). Gene lineages from the 15 different sky-island populations are shown in different colours and the pattern of relationships are denoted by the color bar aside the genealogy (e.g. blocks of similar color denote groups of closely related gene lineages that are also from the same population). The different demographic conditions represented in the figure correspond to a situation in which there is little opportunity for back migration (i.e. the demographic simulation ends shortly after the sky island populations have been colonized), vs. an extended history in which migration continues beyond the initial colonization of the sky island populations. Results are based on the composite method (see text for details); note the scale differs across the gene genealogies.

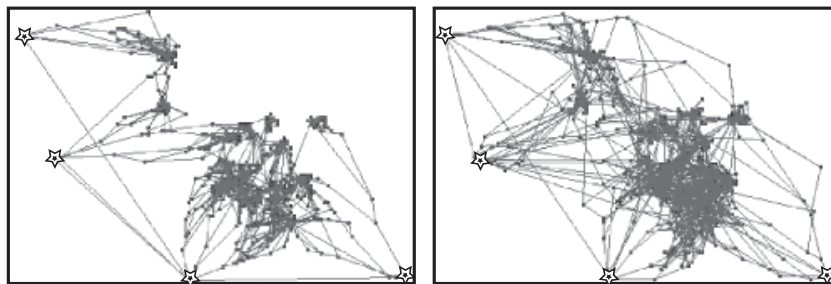


Fig. 8 Schematic of the geographic distribution of gene lineage coalescence at two time points after the initiation of expansion (i.e. 8.5 and 80.0 thousand years) under model A, which shows that most of coalescent events occur in the proximity of the sky-island populations (i.e. areas with greatest concentration of points), relative to the refugial source populations (marked with stars). The spatial dispersion of gene lineage coalescence (i.e. the grey concentrations of branching points) is key in understanding how the demography of the colonization process will impact patterns of genetic variation (i.e. whether coalescent events are old and occur at the geographic sites where the expansion began, as opposed to recent gene lineage coalescence, as shown in these figures).

sky-island populations (commensurate with drops in F_{st} -values; Table 2). However, this spatial movement does not appear to be accompanied by demographic expansion because the distribution of gene lineage coalescence with post-colonization migration become dispersed over time, rather than a concentration of recent gene lineage coalescence events (i.e. gene lineages show a range of branch lengths in the former but not the later case).

Estimating demographic parameters and testing models

Although the proposed approach addresses one critical need in statistical phylogeography—the generation of models (reviewed in Knowles 2009), integration of this procedure with a statistical framework for parameter estimation of testing alternative historical scenarios is needed. Given the complexity of the models being explored to study the genetic consequences of climate-induced distributional shifts, as with other parameter-rich models (see Currat & Excoffier 2004; Fagundes *et al.* 2007; Valdiosera *et al.* 2008; Bramanti *et al.* 2009; Itan *et al.* 2009; Malmström *et al.* 2009), ABC (Beaumont *et al.* 2002, 2010) is ideal for these applications. As a class of coalescent model-based approaches that rely on summary statistics to estimate posterior probabilities of parameters, ABC provides a flexible and powerful approach for model selection in addition to parameter estimation (Beaumont *et al.* 2002). Moreover, uncertainty in parameter values is explicitly modelled, avoiding the problems of inference methods that ignore uncertainty (Knowles 2008).

With respect of the ultimate goal of evaluating the probabilities of the different models, specification of a prior distribution is required for all the parameters of a model. These can be informed from biological information (e.g. empirical estimates of dispersal capabilities) or geological data. For example, with regards to a model

of colonization that proceeds from the LGM to the present, the time of the colonization process (i.e. occurring over 21 000 years) would obviously make some parameter values more or less probable, and hence make some models more or less probable. Clearly the rate of expansion is related to the k , m and r (i.e. contrast the times of colonization in Table 1 across the different models). In this exploration, the simulations were stopped when all the populations were colonized. Yet, under some scenarios it is unlikely that the populations would be founded within 21 000 years, and in other cases, post-colonization migration with the relative rapid movement across the landscape would mean the resulting homogenization of genetic differences would not match the observed data of significant genetic differentiation (see Knowles & Carstens 2007) indicating that these scenarios would also be improbable. In the absence of information to constrain priors for model parameters, broad priors can be chosen (with the caveat that it will dramatically increase the number of simulations needed to estimate the posterior probabilities of parameters), and the impact of mis-specified priors can be quantified (Beaumont *et al.* 2010).

In contrast to the exploration presented here, in which we characterized patterns of genetic differentiation with a single summary statistic— F_{st} —many more summary statistics will be needed to distinguish among historical scenarios; ABC relies on the computation of multiple summary statistics. For example, different models can produce similar F_{st} values (Table 1), which might be compatible with observed empirical data. Yet the genetic data will differ in other characteristic ways depending on the specific details of the history (e.g. Wakeley 2003). Therefore, while a single summary statistic like F_{st} may not provide sufficient power for distinguishing among the scenarios, power for joint parameter estimation and model selection is provided by considering the combination of summary statistics

calculated for any specific set of empirical data (see Beaumont *et al.* 2002). We are in the process of developing a program for implementing ABC analyses that can accommodate the spatio-temporally explicit models generated from the approach described here; without the ability to generate simulations in batch mode (i.e. thousands of runs across a range of values for each demographic parameter) such ABC analyses are beyond the scope of currently available programs.

Conclusions

The study highlights how the genetic consequences of distributional shifts can be studied in a spatio-temporally explicit framework through the coupling of ENMs, demographic, and genetic models, as with other model-based computational approaches (e.g. Knowles 2009; Bruggeman *et al.* 2010; Epperson *et al.* 2010). For species such as *M. oregonensis*, whose current and past distribution is characterized by significant spatial heterogeneity that varies over time (i.e. there is little to no habitat stability because of differences in habitat suitability during different periods of time), environmental heterogeneity may be a primary parameter influencing the genetic consequences of the colonization process (Wegmann *et al.* 2006; Excoffier *et al.* 2009). Consequently, the information contained in the ENMs (i.e. projections of the predicted habitat suitabilities) for current conditions and the past (e.g. 21 000 years ago when many species underwent dramatic changes in distribution) can provide quantitative details about how demographic parameters may vary across the landscape (e.g. population densities and migration rates). We discuss aspects about the predictions of the colonization process that are robust or particularly sensitive to parameter settings in the models, as well as ways in which future work could improve the approach. Nevertheless, the utility and strength of this framework for generating biologically informed models of the colonization process is in the flow of information (Fig. 1) that establishes direct links between process and pattern. For example, analysis of the montane grasshopper showed that when the impact of environmental heterogeneity on the colonization process is taken into account the expansion process itself becomes a likely candidate for generating significant genetic differentiation. Not only are such models a potentially important class of models that should be considered along with a prior focus on the importance of isolation among allopatric glacial refugia (e.g. Knowles 2001; Carstens & Knowles 2007a), but the exploration determined that significant genetic differentiation can be generated when the ancestral refugial source populations were themselves undifferentiated, demonstrating that long-term isolation is not

necessary. Obviously the class of models explored here—demographic expansion across a heterogeneous environment (Wegmann *et al.* 2006)—should be among those that should be tested within a hypothesis testing framework to understand how distributional shifts induced by climate change has contributed to species divergence in this diverse group of montane grasshoppers. The next step in development of this procedure is to integrate the approach described here for model generation into a statistical framework for parameter estimation and hypothesis testing (e.g. using ABC; Beaumont *et al.* 2002, 2010).

Acknowledgements

We would like to thank Victoria Sork and two anonymous reviewers for their helpful suggestions, Qixin He for extensive discussion, helpful comments, and assistance with the graphics, as well as Matthieu Foll for help with the program SPLATCHE and Kevin Gabelman for running some of the models. The National Science Foundation (grants DEB-0715487 and DEB 0918218) and a sabbatical fellowship from the Miller Institute, University of California, Berkeley, to L. Lacey Knowles supported this work.

References

- Beaumont MA, Zhang W, Balding DJ (2002) Approximate Bayesian computation in population genetics. *Genetics*, **162**, 2025–2035.
- Beaumont M, Rasmus N, Robert C *et al.* (2010) In defense of model-based inference in phylogeography. *Molecular Ecology*, **19**, 436–446.
- Bennett KD, Tzedakis PC, Willis KJ (1991) Quaternary refugia of north European trees. *Journal of Biogeography*, **18**, 103–115.
- Braconnot P, Otto-Bliesner B, Harrison S *et al.* (2007) Results of PMIP2 coupled simulations of the Mid-Holocene and Last Glacial Maximum—Part 1: experiments and large-scale features. *Climate of the Past*, **3**, 261–277.
- Bramanti B, Thomas MG, Haak W *et al.* (2009) Genetic discontinuity between local hunter-gatherers and central Europe's first farmers. *Science*, **326**, 137–140.
- Bruggeman DJ, Wiegand T, Fernández N (2010) The relative effects of habitat loss and fragmentation on population genetic variation in the red-cockaded woodpecker (*Picoides borealis*). *Molecular Ecology*, **19**, 3679–3691.
- Buckley TR, Marske K, Attanayake D (2010) Phylogeography and ecological niche modeling of the New Zealand stick insect *Clitarchus hookeri* (White) support survival in multiple coastal refugia. *Journal of Biogeography*, **37**, 682–695.
- Carnaval AC, Hickerson MJ, Haddad CFB, Rodrigues MT, Moritz C (2009) Stability predicts genetic diversity in the Brazilian Atlantic forest hotspot. *Science*, **323**, 785–789.
- Carstens BC, Knowles LL (2007a) Shifting distributions and speciation: species divergence during rapid climate change. *Molecular Ecology*, **16**, 619–628.

- Carstens BC, Knowles LL (2007b) Estimating species phylogeny from gene-tree probabilities despite incomplete lineage sorting: an example from *Melanoplus* grasshoppers. *Systematic Biology*, **56**, 400–411.
- Carstens BC, Richards CL (2007) Integrating coalescent and ecological niche modeling in comparative phylogeography. *Evolution*, **61**, 1439–1454.
- Comes HP, Kadereit JW (1998) The effect of quaternary climatic changes on plant distribution and evolution. *Trends in Plant Science*, **3**, 432–438.
- Currat M, Excoffier L (2004) Modern humans did not admix with Neanderthals during their range expansion into Europe. *PLoS Biology*, **2**, e421.
- Currat M, Ray N, Excoffier L (2004) SPLATCHE: a program to simulate genetic diversity taking into account environmental heterogeneity. *Molecular Ecology Notes*, **4**, 139–142.
- Currat M, Ruedi M, Petit RJ, Excoffier L (2008) The hidden side of invasions: massive introgression by local genes. *Evolution*, **62**, 1908–1920.
- Cushman SA, Landguth EL (2010) Spurious correlations and inference in Landscape Genetics. *Molecular Ecology*, **19**, 3592–3602.
- Edmonds CA, Lillie AS, Cavalli-Sforza LL (2004) Mutations arising in the wave front of an expanding population. *Proceedings of the National Academy of Sciences of the USA*, **101**, 975–979.
- Elith J, Graham CH, Anderson RP *et al.* (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129–151.
- Environmental Systems Research Institute (2008) ArcGIS 9.3. ESRI, Redlands, CA.
- Epperson BK, McRae BH, Scribner K *et al.* (2010) Utility of computer simulations in landscape genetics. *Molecular Ecology*, **19**, 3549–3564.
- Excoffier L, Ray N (2008) Surfing during population expansions promotes genetic revolutions and structuration. *Trends in Ecology and Evolution*, **23**, 347–351.
- Excoffier L, Novembre J, Schneider S (2000) SIMCOAL: a general coalescent program for the simulation of molecular data in interconnected populations with arbitrary demography. *Journal of Heredity*, **91**, 506–509.
- Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, **1**, 47–50.
- Excoffier L, Foll M, Petit RJ (2009) Genetic consequences of range expansions. *Annual Review of Ecology Evolution and Systematics*, **40**, 481–501.
- Fagundes NJR, Ray N, Beaumont M *et al.* (2007) Statistical evaluation of alternative models of human evolution. *Proceedings of the National Academy of Sciences, USA*, **104**, 17614–17619.
- Gray DA, Huang H, Knowles LL (2008) Molecular evidence of a peripatric origin for two sympatric species of field crickets (*Gryllus rubens* and *G. texensis*) revealed from coalescent simulations and population genetic tests. *Molecular Ecology*, **17**, 3836–3855.
- Guisan A, Zimmermann NE, Elith J *et al.* (2007) What matters for predicting the occurrences of trees: techniques, data, or species' characteristics? *Ecological Monographs*, **77**, 615–630.
- Hernandez PA, Graham CH, Master LL, Albert DL (2006) The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography*, **29**, 773–785.
- Hewitt G (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907–913.
- Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965–1978.
- Hofer T, Ray N, Wegmann D, Excoffier L (2009) Large allele frequency differences between human continental groups are more likely to have occurred by drift during range expansions than by selection. *Annals of Human Genetics*, **73**, 95–108.
- Itan Y, Powell A, Beaumont MA, Burger J, Thomas MG (2009) The origins of lactase persistence in Europe. *PLoS Computational Biology*, **5**, e1000491.
- Kearney M, Porter W (2009) Mechanistic niche modelling: combining physiological and spatial data to predict species' ranges. *Ecology Letters*, **12**, 334–350.
- Klopfstein S, Currat M, Excoffier L (2006) The fate of mutations surfing on the wave of a range expansion. *Molecular Biology and Evolution*, **23**, 482–490.
- Knowles LL (2000) Tests of Pleistocene speciation in montane grasshoppers (genus *Melanoplus*) from the sky islands of western North America. *Evolution*, **54**, 1337–1348.
- Knowles LL (2001) Did the Pleistocene glaciations promote divergence? Tests of explicit refugial models in montane grasshoppers. *Molecular Ecology*, **10**, 691–701.
- Knowles LL (2008) Why does a method that fails continued to be used? *Evolution*, **62**, 2713–2717.
- Knowles LL (2009) Statistical phylogeography. *Annual Review of Ecology Evolution and Systematics*, **40**, 593–612.
- Knowles LL, Carstens BC (2007) Estimating a geographically explicit model of population divergence. *Evolution*, **61**, 477–493.
- Knowles LL, Richards CL (2005) Importance of genetic drift during Pleistocene divergence as revealed by analyses of genomic variation. *Molecular Ecology*, **14**, 4023–4032.
- Knowles LL, Carstens BC, Keat ML (2007) Coupled genetic and ecological-niche models to examine how past population distributions contribute to divergence. *Current Biology*, **17**, 1–7.
- Malmström H, Gilbert MTP, Thomas MG *et al.* (2009) Ancient DNA reveals lack of continuity between neolithic hunter-gatherers and contemporary Scandinavians. *Current Biology*, **19**, 1758–1762.
- McRae BH (2006) Isolation by resistance. *Evolution*, **60**, 1551–1561.
- Moyle RG, Filardi CE, Smith CE, Diamond J (2009) Explosive Pleistocene diversification and hemispheric expansion of a 'great speciator'. *Proceedings of the National Academy of Sciences, USA*, **106**, 1863–1868.
- Pannell JR (2003) Coalescence in a metapopulation with recurrent local extinction and recolonization. *Evolution*, **57**, 949–961.
- Petit RJ, Brewer S, Bordacs S *et al.* (2002) Identification of refugia and post-glacial colonisation routes of European white oaks based on chloroplast DNA and fossil pollen evidence. *Forest Ecology and Management*, **156**, 49–74.

- Phillips SJ, Dudik M (2008) Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography*, **31**, 161–175.
- Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, **190**, 231–259.
- Phillips BL, Brown GP, Travis MJJ, Shine R (2008) Reid's paradox revisited: the evolution of dispersal kernels during range expansion. *American Naturalist*, **172**, S34–S48.
- Pielou EC (1991) *After the Ice Age: The Return of Life to Glaciated North America*. University of Chicago Press, Chicago.
- Provan J, Bennett KD (2008) Phylogeographic insights into cryptic glacial refugia. *Trends in Ecology and Evolution*, **10**, 564–571.
- Richards CL, Carstens BC, Knowles LL (2007) Distribution modelling and statistical phylogeography: an integrative framework for generating and testing alternative biogeographical hypotheses. *Journal of Biogeography*, **34**, 1833–1845.
- Rogers AR, Harpending H (1992) Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution*, **9**, 552–569.
- Shapiro B, Drummond AJ, Rambaut A *et al.* (2004) Rise and fall of the Beringian steppe bison. *Science*, **306**, 1561–1565.
- Silva T, Reino LM, Borralho R (2002) A model for range expansion of an introduced species: the common waxbill *Estrilda astrild* in Portugal. *Diversity and Distributions*, **8**, 319–326.
- Slatkin M (1977) Gene flow and genetic drift in a species subject to frequent local extinctions. *Theoretical Population Biology*, **12**, 253–262.
- Slatkin M, Hudson RR (1991) Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics*, **129**, 555–562.
- Smith CI, Farrell BD (2005) Range expansions in the flightless longhorn cactus beetles, *Moneilema gigas* and *Moneilema armatum*, in response to Pleistocene climate changes. *Molecular Ecology*, **14**, 1025–1044.
- Smith MM, Smith HT, Engeman RM (2004) Extensive contiguous north–south range expansion of the original population of an invasive lizard in Florida. *International Biodeterioration and Biodegradation*, **54**, 261–264.
- Sork VL, Davis FW, Westfall R *et al.* (2010) Gene movement and genetic association with regional climate gradients in California valley oak (*Quercus lobata* Née) in the face of climate change. *Molecular Ecology*, **19**, 3806–3823.
- Steele CA, Storfer A (2006) Coalescent-based hypothesis testing supports multiple Pleistocene refugia in the Pacific Northwest for the Pacific giant salamander (*Dicamptodon tenebrosus*). *Molecular Ecology*, **15**, 2477–2487.
- Storfer A, Murphy MA, Evans JS *et al.* (2007) Putting the 'landscape' in landscape genetics. *Heredity*, **98**, 128–142.
- Stuart SN, Chanson JS, Cox NA *et al.* (2004) Status and trends of amphibian declines and extinctions worldwide. *Science*, **306**, 1783–1786.
- Taberlet P, Fumagalli L, Wust-Saucy AG, Cosson JF (1998) Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology*, **7**, 453–464.
- Tzedakis PC, Roucoux KH, De Abreu L, Shackleton NJ (2004) The duration of forest stages in southern Europe and interglacial climate variability. *Science*, **306**, 2231–2235.
- United States Geological Survey (2009) *Geographic Name Information System (GNIS)*. US Department of the Interior, US Geological Survey.
- Valdiosera CE, Garcia-Garitaigoitia JL, Garcia N *et al.* (2008) Surprising migration and population size dynamics in ancient Iberian brown bears (*Ursus arctos*). *Proceedings of the National Academy of Sciences, USA*, **105**, 5123–5128.
- Wakeley J (2003) Inferences about the structure and history of populations: coalescents and intraspecific phylogeography. In: *The Evolution of Population Biology* (eds Singh R, Uyenoyama M). pp. 193–215, Cambridge University Press, Cambridge.
- Wakeley J, Aliacar N (2001) Gene genealogies in a metapopulation. *Genetics*, **159**, 893–905.
- Wegmann D, Currat M, Excoffier L (2006) Molecular diversity after a range expansion in heterogeneous environments. *Genetics*, **174**, 2009–2020.
- Weir JT, Schluter D (2004) Ice sheets promote speciation in boreal birds. *Proceedings Royal Society of London B*, **271**, 1881–1887.
- Whitlock MC, McCauley DE (1990) Some population genetic consequences of colony formation and extinction: genetic correlations within founding groups. *Evolution*, **44**, 1717–1724.
- Wieczorek J, Guo Q, Hijmans R (2004) The point-radius method for georeferencing locality descriptions and calculating associated uncertainty. *International Journal of Geographical Information Science*, **18**, 745–767.
- Wiens JA, Stralberg D, Jongsomjit D, Howell CA, Snyder MA (2009) Niches, models, and climate change: assessing the assumptions and uncertainties. *Proceedings of the National Academy of Sciences, USA*, **106**, 19729–19736.
- Wisz MS, Hijmans RJ, Li J *et al.* (2008) Effects of sample size on the performance of species distribution models. *Diversity and Distributions*, **14**, 763–773.
- Zellmer AJ, Knowles LL (2009) Disentangling the effects of historic vs. contemporary landscape structure on population genetic divergence. *Molecular Ecology*, **18**, 3593–3602.

Dr L. Lacey Knowles is an Associate Professor whose research focuses on the demographic and selective processes driving species divergence. Her research also addresses the methodological and statistical issues involved in studying species diversification, phylogeography, and phylogenetic relationships of recently diverged and rapidly speciating taxa. Diego Alvarado-Serrano is a graduate student whose research focuses on the effects of habitat heterogeneity on species divergence and structures patterns of species diversity, especially in the diverse region of the Ecuadorian Andes.

Supporting information

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

Table S1 Parameters used in the MAXENT models for both the projected distributions of *M. oregonensis* under current

and past conditions (i.e., at the LGM), as well as how the parameters effect the ENMs (for additional details see Phillips *et al.* 2006; Phillips & Dudik 2008). Also listed is the circumscribed geographic area for which the distribution was projected. The predicted distribution was generated for an area that extended beyond the sampled sky-island populations; this area was selected based on the collection localities and the range of the species, the presence of mountain chains, and the southern extent of the glaciers during the last glacial maximum

Table S2 Climatic variables used to generate the predicted distribution of *M. oregonensis* using the program MAXENT (Phillips *et al.* 2006). Codes are given in accordance to the WorldClim Database (Hijmans *et al.* 2005)

Table S3 Fixed parameter settings used in the demographic and genetic models generated with the program SPLATCHE

(Currat *et al.* 2004). See Table 1 for details of other key parameter values that were varied (i.e., k , m , and r)

Movie S1. These visualizations show how demographic parameters impact the pattern of movement across the geographic landscape, and in particular, that the effect of heterogeneity is lessened with large carrying capacities, k , and migration, m . For example, the expansion patterns depicted here, which approximate an expanding wave front, contrast with those with smaller values of k and m (see Fig. 3). Green colors in the maps signify cells with low numbers of emigrants, whereas cells contributing high numbers of emigrants are shown in red.

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