Molecular Ecology (2007) **16**, 619–627 doi: 10.1111/j.1365-294X.2006.03167.x

Shifting distributions and speciation: species divergence during rapid climate change

BRYAN C. CARSTENS and L. LACEY KNOWLES
Department of Ecology & Evolutionary Biology, 1109 Geddes Ave., Museum of Zoology, Room 1089, University of Michigan, Ann Arbor, MI 48109-1079, USA

Abstract

Questions about how shifting distributions contribute to species diversification remain virtually without answer, even though rapid climate change during the Pleistocene clearly impacted genetic variation within many species. One factor that has prevented this question from being adequately addressed is the lack of precision associated with estimates of species divergence made from a single genetic locus and without incorporating processes that are biologically important as populations diverge. Analysis of DNA sequences from multiple variable loci in a coalescent framework that (i) corrects for gene divergence pre-dating speciation, and (ii) derives divergence-time estimates without making a priori assumptions about the processes underlying patterns of incomplete lineage sorting between species (i.e. allows for the possibility of gene flow during speciation), is critical to overcoming the inherent logistical and analytical difficulties of inferring the timing and mode of speciation during the dynamic Pleistocene. Estimates of species divergence that ignore these processes, use single locus data, or do both can dramatically overestimate species divergence. For example, using a coalescent approach with data from six loci, the divergence between two species of montane *Melanoplus* grasshoppers is estimated at between 200 000 and 300 000 years before present, far more recently than divergence estimates made using single-locus data or without the incorporation of population-level processes. *Melanoplus* grasshoppers radiated in the sky islands of the Rocky Mountains, and the analysis of divergence between these species suggests that the isolation of populations in multiple glacial refugia was an important factor in promoting speciation. Furthermore, the low estimates of gene flow between the species indicate that reproductive isolation must have evolved rapidly for the incipient species boundaries to be maintained through the subsequent glacial periods and shifts in species distributions.

Keywords: climate change, divergence time, glacial cycles, Pleistocene, speciation

Received 5 September 2006; revision accepted 15 September 2006

Introduction

The climatic cycles of the Pleistocene had a major impact on patterns of population genetic variation within many species (Hewitt 1996), yet it is unclear how these cycles contributed to species diversification. This uncertainty reflects the extreme difficulty of tracking species divergence during dynamic periods of climate change — not only is there a challenge of obtaining sufficient genetic resolution for differentiating among hypotheses, but there are also analytical difficulties with obtaining accurate divergence-time estimates if speciation was accompanied by gene flow. The periodicity of glacial cycles (Gates 1993) prevents glacial vs. interglacial divergence, or divergence among different glacial periods, from being distinguished without a precise estimate of the timing of speciation. Furthermore, with frequent distributional shifts in response to the glacial cycles (Pielou 1991; Elias 1996; Davis & Shaw 2001; Booth *et al*. 2004; Pierce 2004; Thompson *et al*. 2004; Webb *et al*. 2004), there presumably would have been multiple opportunities for population isolation, but also gene flow given the lack of long-term geographical barriers. When speciation does not occur in strict allopatry (Wu 2001), estimates of divergence

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

will be appreciably underestimated when gene flow is ignored (Nielsen & Wakeley 2001).

The Pleistocene glacial cycles are thought to have played a central role in the formation of the 37 species of montane *Melanoplus* grasshoppers from the sky islands (or mountain tops) of the northern Rocky Mountains (Knowles 2000; Knowles & Otte 2000). Patterns of widespread incomplete lineage sorting, and a largely unresolved polytomy in the species phylogeny, suggest that this diversity hot spot was generated by an explosive and recent radiation (Knowles 2001a). Two models have been proposed to explain how frequent and repeated climate-induced distributional shifts might have promoted diversification in these grasshoppers (Knowles 2001b). Speciation could have been initiated during (i) glacial periods by displacements into multiple refugia, or (ii) interglacials by the founding of the sky islands from ancestral glacial-source populations. Since the timing of species divergence is one of the key components that allows these models to be differentiated, and recognizing the difficulties with obtaining an accurate estimate of when divergence was initiated during periods of rapid climate change (Arbogast *et al*. 2002), we apply an integrative population genetic approach to examine how Pleistocene distributional shifts might have contributed to species diversity.

A multilocus approach was used to estimate the timing of divergence between two species, *Melanoplus oregonensis* and *Melanoplus montanus*; these taxa are members of the diverse clade of montane grasshoppers that radiated in the northern Rocky Mountains (Knowles 2000). DNA sequences from six variable loci (five anonymous nuclear markers identified from a genomic library and one mitochondrial gene) were sequenced in more than 20 individuals for each species. A coalescent model that estimates ancestral genetic diversity and incorporates a migration parameter was used to estimate species divergence (Hey & Nielsen 2004), thereby allowing for the possibility of divergence with gene flow and avoiding a priori assumptions about the processes underlying observed patterns of incomplete lineage sorting between the species. Simulations under conditions matching the empirical data set were then used to determine whether this approach provided an accurate estimate of the timing of grasshopper speciation. These simulations took into account the stochasticity of both mutational and lineage sorting processes, as well as variation in mutation rates. This general approach (Knowles 2004; Knowles & Maddison 2002) provides a framework for asking whether the given data are adequate for estimating the timing of divergence under any specific historical scenario. Verifying that the empirical data can provide an accurate and precise divergence-time estimate is critical. Longstanding debate over the impact of the glacial cycles on species diversity has been fuelled by divergence-time estimates with such broad confidence intervals that various claims about the effects of rapid climate change on species divergence during the dynamic Pleistocene cannot be rejected. As yet, estimates of the timing of speciation have contributed very little to our understanding of how shifting distributions might have promoted speciation because of the logistical and the analytical difficulties discussed below.

Methods

Sampling

Sequence data were collected from 21 individuals from four *Melanoplus oregonensis* populations and 23 individuals from two *Melanoplus montanus* populations. Five anonymous nuclear loci with an average length of 881 bp and 1147 bp of mitochondrial cytochrome oxidase I (COI) were sequenced (see Methods in Knowles 2001a). Across loci (a total of 5.5 kb) there were 409 variable sites, and three of the nuclear loci contained indels (Table 1). The single-copy nuclear polymorphic sequences (SCNPS) were identified from a genomic library using an interspecific screening set that included a single representative from *M. montanus*, *M. oregonensis*, and *Melanoplus marshalli*, a closely related melanopline. The choice of loci was not based on levels

Table 1 Genetic diversity statistics for each locus. From left to right, columns show the length of each locus in base pairs, the number of alleles, the number of insertion-deletions (with the length of the indels in parentheses), the number of polymorphic sites (s) excluding indels, and the nucleotide diversity (π) calculated by SITES, the per-site theta calculated by MIGRATE-N (θ) and its standard deviation, and Tajima's *D*, with the *P* value calculated by simulation using ARLEQUIN

Locus	Length	Alleles	Indels	s	θ	*D*
COI	1147	36	0	125	0.0293 (0.023, 0.038)	−0.320 *P* = 0.401
2	956	25	3 (31, 5, 19)	50	0.0104 (0.008, 0.017)	−0.631 *P* = 0.304
73	853	18	0	10	0.007 (0.005, 0.013)	0.822 *P* = 0.962
85	826	41	0	58	0.0214 (0.009, 0.026)	−1.412 *P* = 0.07
89	581	37	2 (11, 3)	51	0.0336 (0.028, 0.041)	−0.216 *P* = 0.436
211	1188	30	4 (2, 4, 62, 7)	115	0.0146 (0.011, 0.024)	0.897 *P* = 0.765

of variability within *M. oregonensis* or *M. montanus* (which would introduce an ascertainment bias because the lower bound for allele frequencies would depend on the number of individuals used to detect variable loci; Wakeley *et al*. 2001). A plot of the distribution of pairwise differences among individuals within each species was not truncated, demonstrating that the interspecific screening set did not introduce an ascertainment bias.

A genomic library was constructed to identify variable nuclear loci in *Melanoplus* (see Carstens & Knowles 2006) for a detailed protocol). Total genomic DNA was extracted from one *M. oregonensis* using QIAGEN DNeasy kits. The DNA was cut with *Hin*dIII, cloned with the QIAGEN PCR*plus* Cloning kit, and sequenced using an ABI PRISM 3730 Automated Sequencer at the University of Michigan DNA Sequencing Core. *Melanoplus*-specific polymerase chain reaction (PCR) primers were designed using PRIMER 3 1.0 (Rosen & Skaletsky 2000) and OLIGO 4.0 (Molecular Biology Insights, Inc.). PCR subcloning was used to verify that the loci were single copy.

Summary statistics and parameter estimation

For each locus, the number of polymorphic sites (s), genetic diversity (π), and Tajima's *D* (Tajima 1989) were computed with ARLEQUIN 2.1 (Schneider *et al*. 2000). Estimates of $\theta = 4N_e\mu$ for each locus were made using MIGRATE-N (Beerli 2002); estimates of θ from MIGRATE-N can be compared among loci because the estimates are not biased by the number and length of indels. When estimating θ, locus-specific transitions/transversion ratios and base frequencies, and a search strategy that included an adaptive heating scheme with 10 short chains (length 5.0×10^5 generations) and three long chains (length 1.1×10^7 generations) were used; each analysis was repeated with a different random number seed to verify parameter estimates.

Since the shape of each gene genealogy, and specifically the monophyly of *M. oregonensis* and *M. montanus* (or lack thereof), are of interest, a genealogy for each locus was estimated. Genealogies were estimated by maximum likelihood in PAUP* (Swofford 2002), using a heuristic search with 10 random addition sequence replicates and a model of sequence evolution selected using DT-MODSEL (Minin *et al*. 2003). Corrected genetic distances, both the maximum distance within each species and the maximum between species, were calculated for each locus using PAUP*.

An isolation-with-migration model was used to estimate the timing of species divergence ($T = T/\mu$) and five additional parameters with the program IM (Hey & Nielsen 2004): theta of *M. oregonensis* (θ_O) and *M. montanus* (θ_M), the ancestral theta (θ_A), and migration from *M. oregonensis* and *M. montanus* and its reverse ($M_{OM} = M_{OM}/\mu$, $M_{MO} = M_{MO}/\mu$). The priors were truncated at $\theta_1 = \theta_2 = 15$, $\theta_A = 30$, $M_1 = M_2 = 2$, and $T = 7$; the effect of the priors on the posterior probability distribution of parameter estimates was thoroughly investigated. Since the data include five anonymous nuclear loci with unknown mutation rates, the geometric mean of the ratios θ_i: θ_{COI} for each of the five nuclear loci was used to calculate a mutation rate for scaling parameter estimates (Hey & Nielsen 2004), based on a rate of 2.3% sequence divergence per million years for COI (Brower 1994). The parameter space was searched using a linear heating scheme and 10 Metropolis-coupled Markov chains of 5.1 million generations each; the analysis was repeated four times with different random number seeds to confirm convergence.

Since this model assumes that there is no intralocus recombination (Hey & Nielsen 2004), estimates of the per-site recombination rate were calculated for each locus using SITES (Hey & Wakeley 1997), and compared to values obtained from data simulated with no recombination under the estimated model of sequence evolution for each locus. The results from this test indicated that the assumption of nonrecombining loci was justified in all cases, except one locus (SCNPS-85). Recombination rates were also estimated under a coalescent model using the LAMARC package (Kuhner *et al*. 2005), and the results were consistent and suggested that SCNPS-85 might have experienced some intralocus recombination. To determine the potential bias introduced by such recombination, the genetic parameters were estimated under the isolation-with-migration model with and without SCNPS-85.

Coalescent simulations to evaluate the accuracy of divergence-time estimates

A simulation study was conducted to verify the accuracy of the parameters used to differentiate between the two models of species divergence. Divergence of two species from a common ancestor was modelled over a range of divergence equivalent to 8.6×10^4–6.88×10^5 years ($0.5N_e$, $1.0N_e$, $2.0N_e$, and $4.0N_e$ generations); this range encompasses the actual estimated divergence between *M. oregonensis* and *M. montanus* and allows for an assessment of the relationship between divergence time and the confidence intervals of the estimate. Simulated data were intended to closely match the characteristics of the empirical data, both in terms of the model of sequence evolution (Table 2) and the N_e (172 000 individuals; based on the average of estimates from *M. montanus* and *M. oregonensis*) using the program MESQUITE (Maddison & Maddison 2004). Simulated data were analysed with IM using seven Metropolis-coupled chains 1.0 million generations in length, and the following priors: $\theta_1 = \theta_2 = \theta_A = 30$; $M_1 = M_2 = 1$; $T_{0.5N_e} = 4.2$, $T_{1N_e} = 8.4$, $T_{2N_e} = 16.8$, and $T_{4N_e} = 25$.

The analysis was conducted using only the simulated data representing the mitochondrial COI data, as well with an addition of 5 and 10 simulated nuclear loci to evaluate how the estimate was improved by gathering multilocus

Table 2 The maximum genetic distances within and between species, with the average distances in parentheses. The model of sequence evolution and the parameters of the model used to calculate the distances are also shown; all distances are in units of substitutions per site

Locus	Within *montanus*	Within *oregonensis*	Between species	Model	Parameters
COI	0.0460 (0.0205)	0.0414 (0.0156)	0.0602 (0.0230)	TrN + I + Γ	$\pi_A = 0.324$, $\pi_C = 0.158$, $\pi_G = 0.151$, $\pi_G = 0.367$; $tv_1 = 6.35$, $tv_2 = 12.64$; PINV = 0.7313; $\alpha = 0.5278$
2	0.0150 (0.0052)	0.0407 (0.0090)	0.0479 (0.0106)	HKY + I	$\pi_A = 0.352$, $\pi_C = 0.17$, $\pi_G = 0.175$, $\pi_G = 0.303$; ti/tv = 1.188; PINV = 0.6198
73	0.0037 (0.0013)	0.0079 (0.0027)	0.0093 (0.0033)	F81 + I	$\pi_A = 0.273$, $\pi_C = 0.181$, $\pi_G = 0.268$, $\pi_G = 0.178$; PINV = 0.7313
85	0.0313 (0.0097)	0.0400 (0.0050)	0.0522 (0.0082)	K80 + I + Γ	ti/tv = 2.48; PINV = 0.9281; $\alpha = 0.7763$
89	0.0326 (0.0167)	0.0418 (0.0099)	0.0438 (0.0176)	HKY + I + Γ	$\pi_A = 0.295$, $\pi_C = 0.179$, $\pi_G = 0.251$, $\pi_G = 0.274$; ti/tv = 1.985; PINV = 0.814; $\alpha = 0.728$
211	0.0808 (0.0333)	0.0648 (0.0089)	0.0808 (0.0338)	HKY + I + Γ	$\pi_A = 0.294$, $\pi_C = 0.214$, $\pi_G = 0.237$, $\pi_G = 0.255$; ti/tv = 0.963; PINV = 0.4602; $\alpha = 0.8662$

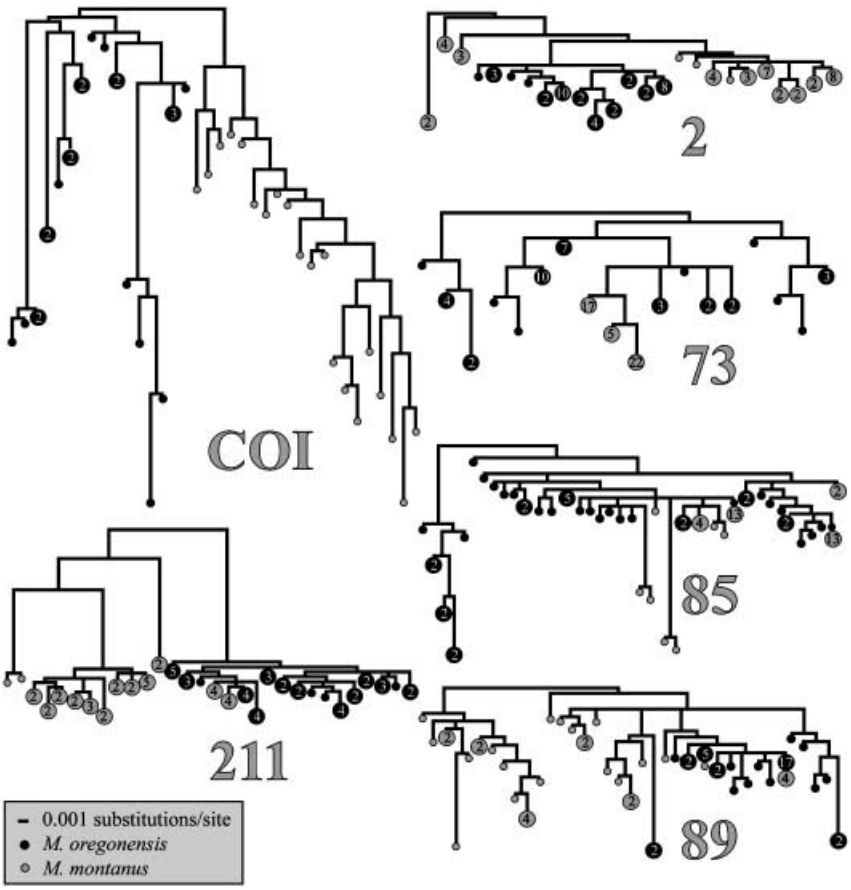

Fig. 1 Variation in the topologies of the gene trees of the six loci. Branch lengths are scaled (in substitutions per site) according to the scale bar shown in the legend. Individual haplotypes for *Melanoplus oregonensis* and *Melanoplus montanus* are represented by black and grey circles, respectively; the numbers within the circles denote haplotypes that were sampled in multiple copies.

data. We recognize that a very small subset of the potential parameter space is being considered here. However, the goal is to investigate the extent to which the data are sufficient to estimate species divergence using a parameter-rich model that allows divergence with gene flow, as well as how the accuracy of the estimated species divergence is impacted by the actual timing of divergence.

Results and discussion

Levels of nucleotide variability differed across loci, but each locus contained substantial variation (Table 1). Maximum-likelihood genealogies show that *Melanoplus oregonensis* and *Melanoplus montanus* were not reciprocally monophyletic at any of the nuclear loci (Fig. 1); genetic distances within

Table 3 Parameterization of the historical split of *Melanoplus oregonensis* and *Melanoplus montanus* from a common ancestor (using the program IM). Shown (from left to right) are the high points of the distribution and the 90% highest posterior density interval; an abbreviation of 'O' and 'M' identifies parameters for *M. oregonensis* and *M. montanus*, respectively

Parameter	High point	$HPD90_{LO}$	$HPD90_{HI}$
θ_O	12.05	9.06	15.72
N_{eO}	1.82×10^5	1.31×10^5	2.42×10^5
θ_M	13.39	9.56	16.55
N_{eM}	2.02×10^5	1.41×10^5	2.47×10^5
θ_A	25.45	16.13	41.74
N_{eA}	3.83×10^5	1.88×10^5	5.99×10^5
M_O	0.21	0.003	0.095
m_o	3.48×10^{-6}	5.23×10^{-7}	2.27×10^{-6}
M_M	0.45	0.017	0.121
m_M	7.47×10^{-7}	4.15×10^{-8}	1.32×10^{-6}
T	4.2	3.23	4.88
Years before present	2.53×10^5	2.08×10^5	3.06×10^5

species are nearly as high as the average distance between species (Table 2). Given the effective population sizes of the species relative to the timing of divergence (Table 3), the high degree of incomplete lineage sorting observed between species, as well as the discordance among loci, should not be taken as evidence against the species status of *M. oregonensis* and *M. montanus* (Hudson & Coyne 2002; Hickerson *et al*. 2006), since it fits with theoretical expectations (Hudson 1990; Maddison 1997; Hudson & Turelli 2003; Maddison & Knowles 2006).

The lack of reciprocal monophyly and the shared polymorphism in the data (Fig. 1) is critical to obtaining a precise estimate of species divergence (Wakeley & Hey 1997). With the stochastic loss of ancestral polymorphism (i.e. genetic information pertinent to identifying the proportion of gene-lineage divergence that pre-dates speciation) the variance on divergence-time estimates necessarily becomes large (Edwards & Beerli 2000; Arbogast *et al*. 2002). As the simulations show, with the loss of ancestral variation over time, the confidence interval on the divergence-time estimates increase (e.g. Fig. 3), causing a loss in precision. Even though incorporating information from multiple loci, as opposed to relying on a single locus, is essential for extracting the historical signal of species splitting from the inherent stochasticity of genetic processes (Edwards & Beerli 2000; Arbogast *et al*. 2002), the genetic resolution is significantly lower for the older divergence times (e.g. $4N$) compared to recent species divergence.

The estimated timing of divergence between *M. oregonensis* and *M. montanus* dates to the pre-Illinoian glacial period, about 253 000 (± 49 000) years before present, assuming one generation per year and a scaled mutation rate of 1.66×10^{-5}. The confidence interval on this estimate (2.08×10^5–3.06×10^5) encompasses this entire glacial period, overlapping somewhat with the proceeding interglacial and just slightly with the following Yarmouth II interglacial (Fig. 4). The precision of this divergence-time estimate based on the multilocus data contrasts with estimate derived from the mitochondrial DNA that extends across four glacial periods (Fig. 4). The estimate is also robust to the inclusion of the locus that might have experienced some intralocus recombination (i.e. SCNPS-85) — the posterior distributions are largely overlapping when T is estimated with or without SCNPS-85.

Despite the precision of the divergence-time estimate, it may not be accurate. The reliability of parameter estimates can be compromised when the number of model parameters exceeds the information content of the data (Hey & Nielsen 2004). The divergence-with-gene-flow model (Hey & Nielsen 2004) has a large number of parameters. Moreover, while the stochasticity of the mutational processes is taken into account with the coalescent model used to estimate the timing of speciation, parameter estimates such as θ or T are converted to biologically relevant values using a mutational scaling factor equal to the geometric mean of the ratios θ_i: θ_{COI} for the five anonymous nuclear loci (Hey & Nielsen 2004). The potential effects of this simplifying assumption for this recently developed approach remains largely unknown, but estimates of species divergence made using this approach are consistent with other estimates in model systems (Hey 2005; Won & Hey 2005). By simulating nucleotide data under models of evolution that match the empirical grasshopper data, we can ask whether this approach can provide an accurate estimate of the timing of speciation given the data (i.e. does the inferred divergence time correspond to the divergence time used to simulate the data). The coalescent simulations clearly demonstrate the accuracy of the divergence-time estimates, as well as the robustness of the approach over a range of differing divergence times (Fig. 3). Estimates of the ancestral effective population size, θ_A, were also inferred accurately across the range of divergence times (data not shown). Comparison of the parameter estimates from the simulated data corresponding to our multilocus data set to those from the simulated data representing the single mitochondrial locus also emphasizes the increased accuracy of the multilocus approach (Fig. 3).

Because the timing of divergence is critical to distinguishing among alternative hypotheses about species diversification (Arbogast *et al*. 2002), estimates of species-divergence times directly influence our conceptual views on the processes that promote speciation. The multilocus and analytical approaches taken here provide not only an accurate, but also a precise estimate of the species-divergence time for distinguishing among divergence at different glacial periods. Nevertheless, longstanding controversy surrounds the role

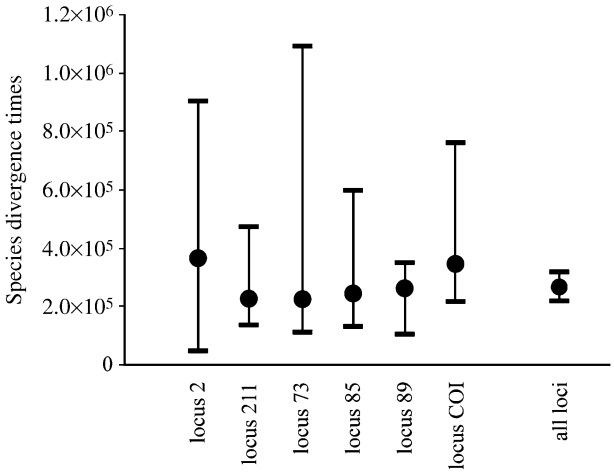

Fig. 2 Estimates of *T* can vary widely across independent estimates based on single loci; the confidence intervals are also much wider when estimates are derived from analyses of single loci compared to multilocus data.

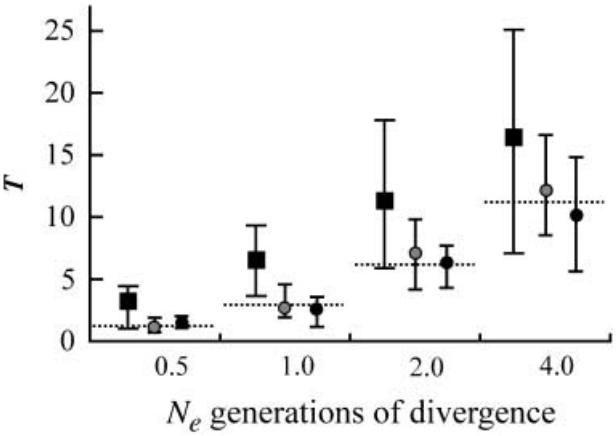

Fig. 3 Accuracy of the estimates of speciation times, *T*, across a range of divergence times (expressed in units of N_e generations) and different sampling regimes; the dashed line indicates the actual time of divergence under which the data were simulated, black squares identify the estimates based on a single locus modelled after the mtDNA, and the grey and black circles identify estimates based on 5 and 10 simulated nuclear loci, respectively. The 90% highest posterior density intervals averaged across replicates are also shown for each set of conditions examined.

of the Pleistocene glacial cycles in speciation (Klicka & Zink 1997, 1998; Arbogast & Slowinski 1998; Weir & Schluter 2004; Zink *et al*. 2004). Much of this debate has been perpetuated by the intractable situation created because the timescale of competing hypotheses about divergence was simply beyond the resolution of the genetic approaches used in the past. For example, divergence-time estimates based on the mitochondrial DNA genetic distances between species, as applied in other taxa (Klicka & Zink 1997; Avise & Walker 1998; Avise *et al*. 1998; DeChaine & Martin 2004;

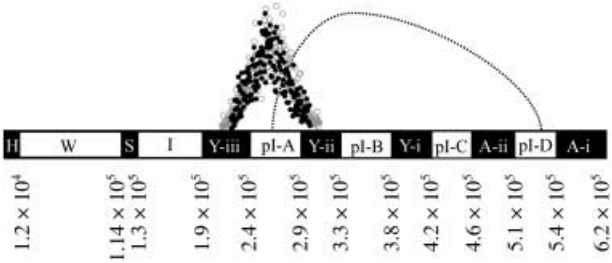

Fig. 4 Species-divergence time estimates set against the timing of glacial-interglacial cycles during the Pleistocene (Gates 1993; Barnola *et al*. 2003; Gibbard & Van Kolfschoten 2004). The 90% highest posterior density (HPD) of species divergence (*T*) estimated from the multilocus data (shown as back and grey circles for the data analysed with and without the nuclear locus SCNPS-85, respectively) compared to that from mitochondrial data alone (shown as a dashed line) are shown. Glacial periods (labelled: W, Wisconsin; I, Illinoian; pI-A, pre-Illinoian A; pI-B, pre-Illinoian B; pI-C, pre-Illinoian C; pI-D, pre-Illinioan D) are shown in white, and interglacials in black (labelled: H, Holocene; S, Sangamon; Y-iii, Yarmouth III; Y-ii, Yarmouth II; Y-i, Yarmouth I; A-ii, Aftonian II; A-i, Aftonian I).

Zink *et al*. 2004) puts the divergence of *M. oregonensis* and *M. montanus* somewhere between 4.9×10^5 and 2.0×10^6 years before present — an estimate as much as four times greater with a confidence interval of 1.5 million compared to 98 000 years for the timing of speciation estimated under the coalescent model (Fig. 2). Note that this has nothing to do with the assumed rate of mutation (the same value was used in each approach), but is caused by ignoring the divergence of gene lineages within the ancestral species (i.e. the divergence that predates speciation). This bias disproportionately affects recent divergence-time estimates compared to older divergences and requires multilocus data to circumvent (Edwards & Beerli 2000).

As with any estimate of species divergence based on a molecular clock there are obvious sources of errors, with difficulties arising from rate heterogeneity among taxa (Yang *et al*. 1994), loci (Swofford *et al*. 1996), and nucleotide sites within a locus (Nei & Li 1979). Although the coalescent-based approach corrects for the discord between the timing of gene divergence and the speciation (Edwards & Beerli 2000), the estimate relies on a molecular-clock calibration (Brower 1994) that, while widely used in arthropods, has not been calibrated in *Melanoplus* grasshoppers specifically. While the conclusion that speciation coincided with the glacial cycles (see also Goropashnaya *et al*. 2004; Johnson & Cicero 2004; Weir & Schluter 2004) is not likely to be influenced, inferences that rely on the finer resolution of the multilocus data, such as distinctions between glacial vs. interglacial divergence, could be effected by calibration errors. Nevertheless, with a parameterized model of speciation, a number of interesting biological conclusions can be drawn about the divergence of *M. oregonensis* and *M. montanus*. Not withstanding the aforementioned

caveats, the bulk of the posterior distribution on the divergence-time estimate occurs during a glacial period (the pre-Illinoian), corroborating analyses of population genetic variation within *M. oregonensis* that suggested speciation was initiated by displacements into multiple, allopatric-glacial refugia (Knowles 2001a). This complementation of studies may be essential for distinguishing between glacial vs. interglacial models of divergence. By itself, the divergence-time estimate does not provide unambiguous evidence for distinguishing between the hypotheses. Moreover, the inferred role of displacements into glacial refugia in promoting species divergence based on patterns of genetic variation within *M. oregonensis* is not contingent upon assumptions about the molecular clock — the conclusion relies upon the geographical configuration of genetic variation (Knowles 2001b; Knowles & Richards 2005).

The multilocus analyses also indicate that species divergence took place without substantial amounts of gene flow (Table 3). Given that speciation predated the most recent Wisconsin glacial period, reproductive isolation must have evolved very rapidly for incipient species divergence to be maintained across subsequent shifts in species distributions. One possible mechanism for the rapid evolution of reproductive isolation in these grasshoppers is sexual selection. Male genitalia in insects are posited to be under sexual selection (Bella *et al*. 1992; Eberhard 1993, 1996; Andersson 1994) and studies have demonstrated that the male genitalia in insects can play an important role in reproductive isolation (Bella *et al*. 1992; Price 1997). Both *M. oregonensis* and *M. montanus*, as well as other montane melanoplines, exhibit pronounced differences in the shape of the male genitalia (Knowles 2000), and are otherwise quite similar morphologically. Moreover, as with analyses of genomic differentiation within species [based on amplified fragment length polymorphism (AFLP) data; Knowles & Richards 2005], this study shows that divergence between species occurred without significant reductions in genetic diversity (Table 1). This pattern of species divergence without purging ancestral variation (Fig. 1), in contrast to speciation associated with severe bottlenecks, raises the intriguing possibility that initial drift-induced differences may be further amplified by selection since the requisite variation for selection to act upon would be readily available. Such an integrative model of drift-induced change and selectively driven divergence may explain how *Melanoplus* grasshoppers were able to radiate during the dynamic Pleistocene.

Conclusions

The integration of multilocus data with coalescent-based models provides resolution for addressing how shifts in species distributions contributed to species divergence during the dynamic Pleistocene. Accuracy of the inferred divergence time estimate was confirmed with a simulation study, thereby providing the first test verifying that the timing of speciation, as inferred from DNA sequences, can be used to distinguish among hypotheses about Pleistocene speciation. As shown here, analysis of data from multiple variable loci in a coalescent framework that corrects for gene divergence pre-dating speciation, and derives divergence-time estimates without making a priori assumptions about the processes underlying patterns of incomplete lineage sorting between species (i.e. allow for the possibility of gene flow during speciation), is critical to overcoming the inherent difficulties that have hampered past inferences about the impact of rapid climate change on species divergence. Two interesting implications about speciation in the montane grasshoppers can be deduced from the parameterized divergence-with-gene-flow model. The estimated timing of divergence between *M. oregonensis* and *M. montanus* corroborates previous studies on population genetic variation within *M. oregonensis* that identified displacements into glacial refugia as promoting divergence (Knowles 2001b; Knowles & Richards 2005). The low estimates of gene flow between the species (Table 3) coupled with a divergence time that pre-dates the most recent Wisconsin glacial period (Fig. 2), indicates that reproductive isolation must have evolved rapidly for the incipient species-boundaries to be maintained through the subsequent glacial periods and shifts in species distributions.

Acknowledgements

We thank members of the Knowles laboratory, particularly M. Keat for assistance in the laboratory. We thank J. Evans for assistance with the simulation study. The research was funded by the following awards to L.L.K.: a National Science Foundation grant (DEB-0447224), the Elizabeth Caroline Crosby Fund, NSF ADVANCE Project, University of Michigan, and a grant from the University of Michigan (Office of the Vice President for Research).

References

Andersson M (1994) *Sexual Selection*. Princeton University Press, Princeton, New Jersey.

Arbogast BS, Slowinski JB (1998) Pleistocene speciation and the mitochondrial DNA clock. *Science*, **282**, 1955a.

Arbogast BS, Edwards SV, Wakeley J, Beerli P, Slowinski JB (2002) Estimating divergence times from molecular data on phylogenetic and population genetic timescales. *Annual Review of Ecology and Systematics*, **33**, 707–740.

Avise JC, Walker D (1998) Pleistocene phylogeographic effects on avian populations and the speciation process. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, **265**, 457–463.

Avise JC, Walker D, Johns GC (1998) Species durations and Pleistocene effects on vertebrate phylogeography. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, **265**, 1707–1712.

Barnola J-M, Raynaud D, Lorius C, Barkov NI (2003) In: *Trends: A Compendium of Data on Global Change*. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tennessee.

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

Bella JL, Butlin RK, Ferris C, Hewitt GM (1992) Asymmetrical homogamy and unequal sex ratio from reciprocal mating-order crosses between *Chorthippus parallelus* subspecies. *Heredity*, **68**, 345–352.

Booth DB, Troost KG, Clague JJ, Waitt RB (2004) The Cordilleran Ice Sheet. In: *The Quaternary Period in the United States* (eds Gillespie AR, Porter SC, Atwater BF). Elsevier Press, New York.

Brower AVZ (1994) Rapid morphological radiation and convergence among races of the butterfly *Heliconius erato* inferred from patterns of mitochondrial DNA evolution. *Proceedings of the National Academy of Sciences, USA*, **91**, 6491–6495.

Carstens BC, Knowles LL (2006) Variable nuclear markers for *Melanoplus oregonensis* identified from the screening of a genomic library. *Molecular Ecology Notes*, **6**, 683–685.

Davis MB, Shaw RG (2001) Range shifts and adaptive responses to Quaternary Climate Change. *Science*, **292**, 673–679.

DeChaine EG, Martin AP (2004) Historic cycles of fragmentation and expansion in *Parnassius smintheus* (Papilionidae) inferred using mitochondrial DNA. *Evolution*, **58**, 113–127.

Eberhard WG (1993) Evaluating models of sexual selection: genitalia as a test case. *American Naturalist*, **142**, 564–571.

Eberhard WG (1996) *Female Control: Sexual Selection by Cryptic Female Choice*. Princeton University Press, Princeton, New Jersey.

Edwards SV, Beerli P (2000) Gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. *Evolution*, **54**, 1839–1854.

Elias SA (1996) *The Ice Age History of National Parks in the Rocky Mountains*. Smithsonian Institution Press, Washington, DC.

Gates DM (1993) *Climate Change and Its Biological Consequences*. Sinauer and Associates, Sunderland, Massachusetts .

Gibbard P, Van Kolfschoten T (2004) The Pleistocene and Holocene Epochs. In: *A Geologic Time Scale* (eds Gradstein FM, Ogg JG, Smith AG). Cambridge University Press, Cambridge, UK.

Goropashnaya AV, Federov VB, Pamilo P (2004) Recent speciation in the *Formica rufa* group ants (Hymenoptera, Formicidae): inference from mitochondrial DNA phylogeny. *Molecular Phylogenetics and Evolution*, **32**, 198–206.

Hewitt GM (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*, **58**, 247–276.

Hey J (2005) On the number of new world founders: a population genetic portrait of the peopling of the Americas. *PloS Biology*, **3**, 965–975.

Hey J, 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.

Hey J, Wakeley J (1997) A coalescent estimator of the population recombination rate. *Genetics*, **145**, 833–846.

Hickerson MJ, Meyer CP, Moritz C (2006) DNA-barcoding will fail to discover animal species. *Systematic Biology*, **5**, 729–739.

Hudson RR (1990) Gene genealogies and the coalescent process. In: *Oxford Survey Evolutionary Biology* (eds Futuyma D, Antonovics J). Oxford University Press, New York.

Hudson RR, Coyne JA (2002) Mathematical consequences of the genealogical species concept. *Evolution*, **56**, 1557–1565.

Hudson RR, Turelli M (2003) Stochasticity overrules the 'three-times' rule: genetic drift, genetic draft, and coalescence times for nuclear loci versus mitochondrial DNA. *Evolution*, **57**, 182–190.

Johnson NK, Cicero C (2004) New mitochondrial DNA data affirm the importance of Pleistocene speciation in North American birds. *Evolution*, **58**, 1122–1130.

Klicka J, Zink RM (1997) The importance of recent ice ages in speciation: a failed paradigm. *Science*, **277**, 1666–1669.

Klicka J, Zink RM (1998) Response: Pleistocene speciation and the mitochondrial DNA clock. *Science*, **282**, 1955a.

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 (2001a) Did the Pleistocene glaciations promote divergence? Tests of explicit refugial models in montane grasshoppers. *Molecular Ecology*, **10**, 691–701.

Knowles LL (2001b) Genealogical portraits of speciation in montane grasshoppers (genus *Melanoplus*) from the sky islands of the Rocky Mountains. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, **268**, 319–324.

Knowles LL (2004) The burgeoning field of statistical phylogeography. *Journal of Evolutionary Biology*, **17**, 1–10.

Knowles LL, Maddison WP (2002) Statisical phylogeography. *Molecular Ecology*, **11**, 2623–2635.

Knowles LL, Otte D (2000) Phylogenetic analysis of montane grasshoppers from western North America (genus *Melanoplus*, Acrididae, Melanoplinae). *Annals of the Entomological Society of America*, **93**, 421–431.

Knowles LL, Richards CL (2005) Importance of genetic drift during Pleistocene divergence as revealed by analyses of genomic variation. *Molecular Ecology*, **14**, 4023–4032.

Kuhner MK, Yamato J, Beerli P *et al.* (2005) LAMARC version 2.0. Available at http://evolution.gs.washington.edu/lamarc.html.

Maddison WP (1997) Gene trees in species trees. *Systematic Biology*, **46**, 523–536.

Maddison WP, Knowles LL (2006) Inferring phylogeny despite incomplete lineage sorting. *Systematic Biology*, **55**, 21–30.

Maddison WP, Maddison DR (2004) *MESQUITE: A modular system for evolutionary analysis. Version 1.01*. Available at http://mesquiteproject.org

Minin V, Abdo Z, Joyce P, Sullivan J (2003) Performance-based selection of likelihood models for phylogeny estimation. *Systematic Biology*, **52**, 674–683.

Nei M, Li W-H (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences, USA*, **76**, 5269–5273.

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

Pielou EC (1991) *After the Ice Age: The Return of Life to Glaciated North America*. University of Chicago Press, Chicago.

Pierce KL (2004) Pleistocene glaciations of the Rocky Mountains. In: *The Quaternary Period in the United States* (eds Gillespie AR, Porter SC, Atwater BF). Elsevier Press, New York.

Price CSC (1997) Conspecific sperm preference in *Drosophila*. *Nature*, **388**, 663–666.

Rosen S, Skaletsky HJ (2000) *PRIMER 3 on the WWW for general users and biologist programmers*. Available at http://frodo.wi.mit.edu/primer3/primer3_code.html.

Schneider S, Roessli D, Excoffier L (2000) *ARLEQUIN: A software for population genetics data analysis*. Genetics and Biometry Laboratory, Department of Anthropology, University of Geneva, Switzerland. Available at http://anthro.unige.ch/arlequin.

Swofford DL (2002) *PAUP*. Phylogenetic Analysis Using Parsimony (and Other Methods)*. Version 4. Sinauer and Associates, Sunderland, Massachusetts.

Swofford D, Olsen G, Waddell P, Hillis D (1996) Phylogenetic inference. In: *Molecular Systematics* (eds Hillis DM, Moritz C, Mable BK), pp. 407–514. Sinauer and Associates, Sunderland, Massachusetts.

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585–595.

Thompson RS, Shafer SL, Strickland LE, Van de Water PK, Anderson KH (2004) Quaternary vegetation and climate change in the western United States. In: *The Quaternary Period in the United States* (eds Gillespie AR, Porter SC, Atwater BF). Elsevier Press, New York.

Wakeley J, Hey J (1997) Estimating ancestral population parameters. *Genetics*, **145**, 847–855.

Wakeley J, Nielsen R, Liu-Cordero SN, Ardlie K (2001) The discovery of single-nucleotide polymorphisms and inferences about human demographic history. *American Journal of Human Genetics*, **69**, 1332–1347.

Webb III, Shuman TB, Williams JW (2004) Climatically forced vegetation dynamics in eastern North America during the late Quaternary Period. In: *The Quaternary Period in the United States* (eds Gillespie AR, Porter SC, Atwater BF). Elsevier Press, New York.

Weir JT, Schluter D (2004) Ice sheets promote speciation in boreal birds. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, **271**, 1881–1887.

Won Y-J, Hey J (2005) Divergence population genetics of chimpanzees. *Molecular Biology and Evolution*, **22**, 297–307.

Wu C-I (2001) The genic view of the process of speciation. *Journal of Evolutionary Biology*, **14**, 851–865.

Yang Z, Goldman N, Friday A (1994) Comparison of models for nucleotide substution used in maximum-likelihood phylogenetic estimation. *Molecular Biology and Evolution*, **11**, 316–324.

Zink RM, Klicka J, Barber BR (2004) The tempo of avian diversification during the Quaternary. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **359**, 215–220.

This study complements earlier work conducted by L. L. Knowles and others on montane grasshoppers of the genus *Melanoplus*. Not only does it provide insights into the processes involved in the radiation of western *Melanoplus* grasshoppers during the dynamic Pleistocene, but it also addresses issues surrounding the estimation of species divergence times. Bryan Carstens conducts phylogeographic research in western North America, investigating the factors that influence the formation of population genetic structure and the divergence of lineages across a broad set of codistributed species. Both authors are interested in quantitative approaches used to test phylogeographic hypotheses, including how to best estimate hierarchical models of population divergence using genomic data.
