Genealogical portraits of speciation in montane grasshoppers (genus *Melanoplus*) from the sky islands of the Rocky Mountains

L. Lacey Knowles†

Department of Ecology and Evolution, State University of New York at Stony Brook, Stony Brook, New York 11794-5245, USA

Grasshoppers in the genus *Melanoplus* have undergone a radiation in the 'sky islands' of western North America, with many species originating during the Pleistocene. Despite their recent origins, phylogenetic analyses indicate that all the species exhibit monophyletic or paraphyletic gene trees. The objectives of this study were to determine whether the monophyletic genealogies are the result of a bottleneck at speciation and to investigate the extent to which the different phylogenetic states of eight species (i.e. monophyletic versus paraphyletic gene trees) can be ascribed to the effects of speciation. A coalescent simulation was used to test for a bottleneck at speciation in each species. The effective population sizes and demographic histories of species were compared across taxa to evaluate the possibility that the paraphyly versus monophyly of the species reflects differential rates of lineage loss rather than speciation mode. While coalescent analyses indicate that the monophyly of *Melanoplus* species might not be indicative of bottlenecks at speciation, the results suggest that the paraphyletic gene trees may reflect the demography of speciation, involving localized divergences in the ancestral species. With respect to different models of Pleistocene divergence, the data do not support a model of founder-effect speciation but are compatible with divergence in allopatric refugia.

**Keywords:** bottlenecks; coalescence; gene genealogy; historical demography; Pleistocene; speciation

1. INTRODUCTION

Gene genealogies are often used to study the geography of species divergence (Avise 1989; Harrison 1999; Crandall & Templeton 1993; Riddle 1996) but they can also provide information about the demography of speciation. For example, using models of evolutionary processes based on coalescent theory (Kingman 1982), gene trees can be used to test different hypotheses about the speciation process, such as whether there was a bottleneck at speciation (e.g. Takahata 1993; Eyre-Walker et al. 1998; Klein et al. 1998; Knowles et al. 1999).

A gene genealogy undergoes a progression from an initial state of polyphyley to paraphyly and eventually monophyly as the time since speciation increases (Neigel & Avise 1986; Ball et al. 1990; Avise 1994). This transition to monophyly requires \(4N_e\) generations, on average, where \(N_e\) is the effective population size (Hudson 1990; Chesser & Baker 1996; for exceptions, see Hoelzer 1997). In addition to the influence of the species' historical population size, the mode of speciation also affects a gene genealogy. A bottleneck at speciation can produce a monophyletic gene tree (Neigel & Avise 1986), whereas a paraphyletic gene tree occurs if a descendant species arises from one of several populations of the ancestral species (i.e. from a subset of the ancestral species' gene tree), such as during peripatric speciation (e.g. Avise et al. 1983; Hey & Kliman 1993). However, although monophyletic and paraphyletic gene trees may be consistent with particular modes of speciation, they do not necessarily reflect the demography of speciation (Knowles et al. 1999). A monophyletic gene tree may arise if there has been sufficient time for coalescence of lineages due to a species' small effective population size. Similarly, a paraphyletic gene tree may reflect the retention of ancestral polymorphism because of a species' large effective population size (Maddison 1997). Therefore, since the speciation process and the demographic history of species affect the rate of transition of gene genealogies, both influences should be considered.

This study addresses the extent to which the mode of speciation in *Melanoplus* grasshoppers has left a signature on the species' genealogies, and whether speciation in this group involved bottlenecks. It has previously been shown that the *Melanoplus* species inhabiting the 'sky islands' of the northern Rocky Mountains originated during the Pleistocene (Knowles 1999, 2000a; Knowles & Otte 1999). Despite their recent origins, many of the species exhibit monophyletic gene trees (table 1; this study) suggesting that they may have experienced a bottleneck at speciation. Moreover, bottlenecks are a common feature of speciation models involving displacement of species into glacial refugia (e.g. Mengel 1964; Hewitt 1996, 2000). The northern Rocky Mountain region was dominated by the Cordilleran ice sheet and a complex of mountain and valley glaciers during the Pleistocene (Hollin & Schilling 1988; Mayweski et al. 1981), which displaced species to ice-free refugia (Ficou 1991; Elias 1996). However, not all the *Melanoplus* species under study have monophyletic gene trees (table 1). Thus, the goal of this study was to determine whether the mode of speciation differs between species. Coalescent simulations were used to evaluate the likelihoods that the monophyletic genealogies reflect bottlenecks at speciation. Estimates of the effective population sizes and consideration of the historical demographies of the species were used to

†Present address: Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721-0088, USA (knowles@u.arizona.edu).
Table 1. Phylogenetic status (from Knowles 2000a) and estimates of genetic diversities (i.e. \( \pi \) and \( \hat{\theta} \), effective population sizes \( (N_e) \), Tajima’s \( D \) and time of origin of each species (estimates were made using the program SITES (Wakeley & Hey 1997) and MEGA (Kumar et al. 1993))

The number of specimens sequenced is given in parentheses. Significant values of Tajima’s \( D \) (i.e., values outside the 95% confidence interval of the neutral expectation based on table 1 in Tajima (1989)) are shown in bold. The range of divergence times encompasses the standard errors calculated for the average pairwise differences between species assuming one generation per year and using Kimura’s two-parameter model to correct for multiple substitutions (Kimura 1980).)

<table>
<thead>
<tr>
<th>species</th>
<th>( \pi ) bp(^a)</th>
<th>( \hat{\theta} ) bp(^b)</th>
<th>Tajima’s ( D )</th>
<th>( N_e )(^c)</th>
<th>divergence time (Myr)</th>
<th>phylogenetic status</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>M. oregonensis</em> (124)</td>
<td>0.1957</td>
<td>0.2977</td>
<td>-1.9159</td>
<td>850 000–1 200 000</td>
<td>1.10–1.64</td>
<td>monophyletic</td>
</tr>
<tr>
<td><em>M. marshalli</em> (5)</td>
<td>0.1926</td>
<td>0.2426</td>
<td>-1.4190</td>
<td>837 400–1 100 000</td>
<td>1.10–1.64</td>
<td>monophyletic</td>
</tr>
<tr>
<td><em>M. indigenus</em> (3)</td>
<td>0.1442</td>
<td>0.1534</td>
<td>-0.7350</td>
<td>627 000–667 000</td>
<td>0.69–1.08</td>
<td>paraplythetic</td>
</tr>
<tr>
<td><em>M. crux</em> (9)</td>
<td>0.1025</td>
<td>0.1377</td>
<td>-1.2903</td>
<td>512 500–688 500</td>
<td>0.69–1.08</td>
<td>paraplythetic</td>
</tr>
<tr>
<td><em>M. mississippi</em> (6)</td>
<td>0.0902</td>
<td>0.1288</td>
<td>-1.9067</td>
<td>392 200–560 000</td>
<td>0.50–0.82</td>
<td>monophyletic</td>
</tr>
<tr>
<td><em>M. triangularis</em> (8)</td>
<td>0.0918</td>
<td>0.0962</td>
<td>-0.4268</td>
<td>399 100–418 300</td>
<td>0.50–0.82</td>
<td>paraplythetic</td>
</tr>
<tr>
<td><em>M. montanus</em> (4)</td>
<td>0.0834</td>
<td>0.0894</td>
<td>-0.6825</td>
<td>362 600–388 700</td>
<td>0.91–1.40</td>
<td>monophyletic</td>
</tr>
<tr>
<td><em>M. moyenii</em> (4)</td>
<td>0.0589</td>
<td>0.0638</td>
<td>-1.6758</td>
<td>256 000–286 000</td>
<td>0.91–1.40</td>
<td>monophyletic</td>
</tr>
</tbody>
</table>

\( \pi \), the average number of pairwise differences between sequences (Tajima 1983).

\( \hat{\theta} \), based on the number of segregating sites (Watterson 1975).

\( N_e \), derived from the expected value of \( 2N_e \mu \) for both \( \pi \) and \( \hat{\theta} \).

Figure 1. The two models used in the coalescent simulation. Under the bottleneck model \((b)\) at time \( t > T \), the \( i \) lineages ancestral to the sampled sequences coalesce to a single ancestral lineage with a probability of 1, whereas under the constant-population-size model \((a)\) the lineages continue to coalesce at the rate of \( ((i − 1)/2N_e) \) (Hudson 1990). Consequently, in the bottleneck model \((b)\) it is assumed that a single individual gave rise to the species, followed by an immediate expansion such that the effective population size, \( N_e \), immediately following the speciation event is the same for both models. However, moderately larger numbers of founders or lower rates of growth will give similar results (B. Rannala, unpublished data).

evaluate whether different rates of lineage loss, rather than the mode of speciation, account for the phylogenetic status of the genealogies.

2. MATERIAL AND METHODS

(a) The study system

The study focused on eight grasshopper species in the genus *Melanoplus* (Stål) (Orthoptera: Acrididae: Melanoplinae: Indigenus and Montanus species groups) distributed in the northern Rocky Mountains (Scudder 1898; Knowles 1999; Otte 2001). The number of specimens sequenced, estimates of divergence times and phylogenetic status for each species are given in table 1. A 1300 base-pair fragment of cytochrome oxidase I mitochondrial DNA was sequenced for each individual (Knowles 2000a,b). The phylogenetic analyses are described by Knowles (2000a,b). The phylogenetic status of each species was consistent across all analyses (Knowles 2000a,b); data were analysed using parsimony, neighbour-joining and maximum-likelihood methods (PAUP*, Swofford 1998). A rate of divergence of 2.3% per million years (Myr) (Brower 1994) was used to calculate species’ divergence times, \( T \), and to estimate the mutation rate per base pair per sequence per generation, \( \mu \). This rate of molecular evolution is similar to other estimates (e.g. Brown et al. 1979).

(b) Demographic and coalescent analyses

To determine whether different rates of lineage loss account for the monophyly versus paraplythy of the species, the effective population size and demographic history of each species were compared. The effective population size of each species was derived from estimates of their respective genetic diversities (i.e. from \( \pi \) and \( \hat{\theta} \); table 1). The frequency distribution of pairwise differences between sequences (i.e. the mismatch distribution) was calculated for each species, and then the shapes of the distributions were compared to detect variation in the species’
Figure 2. Log-likelihood surface (with standard errors) of each of the species’ genealogies under the two coalescent models: a constant population size with no speciation bottleneck (solid line) and a dramatic bottleneck at speciation (dashed line). For each set of parameter values 10000 replicates were simulated.

demographic histories (Slatkin & Hudson 1991; Marjoram & Donnelly 1994). The fit of each species to Wakeley & Hey’s (1997) model of population expansion based on the number of segregating sites was also examined. Tajima’s $D$ (Tajima 1989) was also calculated for each species and compared across species to provide an additional comparison of their demographic histories. While this measure of variation can be used to examine the history of selection, it can also be used to make inferences about population demography. Tajima’s $D$ is expected to be negative under a model of population expansion and positive under population subdivision (e.g. Aris-Brosou & Excoffier 1996).

A likelihood approach was used to evaluate the genetic evidence for a bottleneck associated with speciation in each species. A coalescent process (Kingman 1982) was used to model the species’ genealogies expected under two different demographic scenarios: first, a constant population size with no speciation bottleneck (figure la) and, second, a dramatic bottleneck at speciation (figure lb). The probability of observing $S$ segregating sites in a sample of $n$ sequences (the likelihood, when treating $S$ as the observed data) was calculated under the two models for a range of population sizes, assuming an infinite-sites model of sequence mutation. The probability of $S$ is

$$
\Pr(S|\mu, T, N) = \int \Pr(S|\mu, t) \Pr(t|N, T) dt.
$$

(1)

where $\mu$ is the per-sequence mutation rate (i.e. the per-site mutation rate multiplied by the number of sites in the sequence) and $t = \{t_1, t_{n-1}, \ldots, t_1\}$, is a vector of the coalescence times, where $t_i$ is the waiting time for $n$ sequences to coalesce to $i - 1$ ancestral sequences. Monte Carlo integration and simulation from the coalescent process $\Pr(t|N, T)$ was used to evaluate the above integral, where $T$ is the time of species divergence (program provided by B. Rannala).

3. RESULTS

Estimates of $\pi$ and $\hat{\theta}$, as well as Tajima’s $D$, for each species are presented in table 1. The paraphyletic species (i.e. $M. crux$ and $M. triangularis$) clearly do not have larger effective population sizes than most of the monophyletic species (table 1). This conclusion is independent of the different estimates of $N_e$; in general, estimates of $N_e$ derived from $\pi$ and $\hat{\theta}$ (table 1) correspond to the maximum-likelihood estimates from the coalescent simulations (figure 2), with the exception of $M. oregonensis$.

Demographic histories do not vary in an obvious way across most species. While Tajima’s $D$ is not significant in all species, it is consistently negative (table 1). Moreover, the frequency spectrum of pairwise differences for each species was multimodal, with the exception of a unimodal distribution in $M. oregonensis$ (figure 3). However, a multimodal distribution can result from a variety of demographic scenarios (Slatkin & Hudson 1991; Marjoram & Donnelly 1994), and the small sample sizes (see table 1) in some of the species may have affected the distribution.

$M. oregonensis$ fits a population-expansion model based on the number of segregating sites (Wakeley & Hey 1997), and the unimodal distribution of pairwise differences and significantly negative Tajima’s $D$ are consistent with a model of population expansion in $M. oregonensis$ (Slatkin & Hudson 1991; Rogers & Harpending 1992; Aris-Brosou & Excoffier 1996). Given that the species studied here are
distributed in previously glaciated areas (Knowles 1999), this result is not unexpected. However, it does not indicate that the species is panmictic (Knowles 2000b). Depending on the level of gene flow, even a population with substructure can exhibit a unimodal distribution (Hudson 1990; Bertorelle & Slatkin 1995).

The coalescent simulations indicate that the likelihood of a bottleneck at speciation is not higher than the likelihood of no bottleneck (i.e. the confidence intervals of the highest log-likelihood values overlap between the two models in all species; figure 2). Furthermore, the lack of a significant difference between the two models shows that there has been sufficient time for lineage sorting to occur, i.e. for the species’ gene trees to become monophyletic. While it may be somewhat surprising that there has been sufficient time for coalescence of lineages in species that originated around 1 Myr ago (e.g. M. oregonensis and M. marshalli), as well as in species that originated as recently as 500,000 years ago (e.g. M. indigens and M. missouli), the more recently derived species do have correspondingly smaller effective population sizes compared to the older species (table 1).

4. DISCUSSION

These analyses indicate that the origin of the Melanoplus species studied here did not necessarily involve bottlenecks. The coalescent simulations show that the monophyletic gene trees are consistent with the hypothesis that there has been sufficient time for coalescence of lineages within species (figure 2). Moreover, since there are no obvious differences in the effective population sizes or demographic histories between the species with paraphyletic and monophyletic gene trees (table 1 and figure 3), the paraphyly is not attributable to a slower rate of progression towards monophyly. Consequently, it is more likely that the paraphyletic genealogies reflect a mode of speciation in which descendant species arose from one of several populations of the ancestral species (e.g. Harrison 1991).

Thus, as a complement to studies that have demonstrated that high genetic diversities in recently derived species do not preclude the possibility of a bottleneck at speciation (e.g. Eyre-Walker et al. 1998), the present findings argue that monophyletic genealogies, even in recently derived species, do not necessarily imply a bottleneck at speciation (see also Knowles et al. 1999). Furthermore, these results, in conjunction with phylogeographic analyses of population divergence in Melanoplus (Knowles 2000b), provide not only a clearer portrait of speciation in these grasshoppers but also offer insight into the general process of Pleistocene speciation.

The picture that emerges from both the population- and species-level approaches is one of divergence in allopatric refugia (e.g. Mengel 1964; Hewitt 1996) rather than speciation by bottlenecks, such as postulated by founder-effect models (e.g. Mayr 1954; Templeton 1980; Carson 1992; Gavrilets & Hastings 1996). However, because of the frequency of glacial cycles (Bartlein &
Prentice 1989; Dansgaard et al. 1993), and consequently the short time interval separating shifts in species distributions (Roy et al. 1996), Pleistocene speciation must involve the rapid evolution of reproductive isolation (Knowles 1999, 2000a; Dynesius & Jansson 2000). Indeed, while regional structuring of population variation in Melanoplus supports the hypothesis that divergence occurred in allopatric refugia, evidence of sporadic gene flow indicates that these differences will be lost during interglacial expansion without a mechanism to prevent population mixing (Knowles 2000b).

In the absence of evidence to support drift-induced rapid divergence, some other process, such as selection, and specifically sexual selection, may account for the rapidity of the speciation process in Melanoplus species originating during the Pleistocene. The Melanoplus species studied here occupy similar habitats and are morphologically very similar except in the shape of the male genitalia (Knowles 1999; Otte 2001). Male genitalia in insects are posited to be under sexual selection and characters under sexual selection can evolve rapidly (Eberhard 1985, 1993, 1996; Arnegquist 1998). Differences in male genitalia can also play an important role in reproductive isolation (Eberhard 1996; Arnegquist & Danielsson 1999).

In addition to examples of species divergence in association with long-standing geographic barriers (e.g. Avise 1994), the role of transient allopatry coupled with natural or sexual selection in speciation has recently been recognized in a variety of species (e.g. McCune 1997; Hellberg 1998; Orr & Smith 1998; Schluter 1998). Thus, it is highly likely that speciation in Melanoplus is a product of sexual selection and divergence initiated in allopatric glacial refugia.

I thank Daniel Otte for providing critical information on species distributions and identifications, Bruce Ramnala for providing the program for coalescent simulations and Douglas J. Futuyma (my advisor) and two anonymous reviewers for thoughtful comments that greatly improved the manuscript. This research was supported by a National Science Foundation Dissertation Improvement Grant (DEB-9801284), by a Sigma-Xi Grant-in-aid-of-research and a grant from the Orthopterists’ Society. This is contribution 1078 from the Department of Ecology and Evolution at the State University of New York at Stony Brook.

REFERENCES


