Molecular Ecology (2003) **12**, 711–724

Historical demography and genetic structure of sister species: deermice (*Peromyscus*) in the North American temperate rain forest

XIAOGUANG ZHENG, BRIAN S. ARBOGAST* and G. J. KENAGY
Burke Museum and Department of Zoology, University of Washington, Seattle, Washington 98195, USA

Abstract

To investigate the evolutionary and biogeographical history of *Peromyscus keeni* and *P. maniculatus* within the coastal forest ecosystem of the Pacific Northwest of North America, we sampled 128 individuals from 43 localities from southeastern Alaska through Oregon. We analysed mitochondrial DNA variation using DNA sequence data from the mitochondrial cytochrome-*b* (cyt-*b*) gene and control region, and we found two distinct clades consistent with the morphological designation of the two species. The sequence divergence between the two clades was 0.0484 substitutions per site for cyt-*b* and 0.0396 for the control region, suggesting that divergence of the two clades occurred during the middle to late Pleistocene. We also examined the historical demography of the two clades using stepwise and exponential expansion models, both of which indicated recent rapid population growth. Furthermore, using the program MIGRATE we found evidence of migration from populations north of the Fraser River (British Columbia) to the south in both clades. This study demonstrates the utility of these model-based demographic methods in illuminating the evolutionary and biogegographic history of natural systems.

Keywords: coalescence, migration, mitochondrial DNA, phylogeography, population expansion, speciation

Received 15 August 2002; revision received 15 November 2002; accepted 1 December 2002

Introduction

The importance of identifying patterns of differentiation between closely related species was heralded by Darwin (1859) as a basis for understanding evolution by natural selection. Not until a century later did we have an understanding of the genetic process of the origin of new species (Dobzhansky 1951; Lewontin 1974; Wright 1978). With the further development of molecular biology and genetics in the later 20th century, the study of the speciation process was enhanced through an understanding of the genetic and ecological basis of variation and selection (Schluter 2001; Turelli *et al*. 2001; Wu 2001; Via 2002).

Correspondence: G. J. Kenagy, Burke Museum, Box 353010, University of Washington, Seattle, WA 98195, USA. Fax: 206 685 3039; E-mail: kenagy@u.washington.edu
*Present address: Department of Biological Sciences, Humboldt State University, Arcata, California 95521, USA.

Recent studies have addressed progressive levels of genetic differentiation in the process of speciation, ranging from genetic introgression within species (Bonhomme & Guenet 1996) to hybridization between closely related species (Harrison 1993; Arnold 1997; Salzburger *et al*. 2002) and complete reproductive isolation and genetic differentiation of full species (Palopoli & Wu 1994; Coyne & Orr 1998; Hacia 2001). New multidisciplinary efforts have also been made to understand the genetic basis of variation in functionally important morphological traits (Bradshaw *et al*. 1998). Finally, new research directions are developing in the ecological genetics of speciation, emphasizing ecologically important phenotypic traits in natural populations (Schluter 2001; Via 2002).

An additional important aspect of investigating speciation is to understand the historical demography and geographical movement of taxa. This information provides a context for linking genetic differentiation, ecological adaptation, and environmental change. Efforts leading to these approaches include the recent explosion of studies in phylogeography (reviewed in Avise 2000) and the development

of a variety of model-based programs to estimate historical population parameters (reviewed in Emerson *et al*. 2001), especially with regard to applications of coalescent theory (Hudson 1990), Bayesian and maximum likelihood methods (Kuhner *et al*. 1998; Edwards & Beerli 2000; Beerli & Felsenstein 2001). Some recent examples that utilize these approaches include Miura & Edwards's (2001) study on Hall's babbler, Matocq's (2002) study on the dusky-footed woodrat and Conroy & Cook's (2000) study on the long-tailed vole. In the present study we apply these approaches to a pair of sister species of deermice (*Peromyscus*) in the temperate rain forest of the Pacific Northwest of North America.

As a result of cyclical climatic changes of the Pleistocene and Holocene, the boreal forest habitat of northwestern North America has experienced extensive historical changes in geographical location and degree of continuity (Pielou 1991). These climatically induced fluctuations of habitat have been considered important influences promoting floral and faunal diversification in the region. During these periods populations of many taxa may have repeatedly become isolated and/or recontacted one another, and in some cases speciated as a result (Byun *et al*. 1997; Soltis *et al*. 1997; Barrowclough *et al*. 1999; Arbogast & Kenagy 2001). This system provides an excellent setting for the study of speciation. The widespread and ecologically successful deermice of the genus *Peromyscus* in North America are good subjects for genetic analysis of speciation in natural populations (Osgood 1909; Blair 1950; King 1968; Kirkland & Layne 1989). Within the genus *Peromyscus* the widespread *P. maniculatus* and its sister taxon, the more geographically restricted *P. keeni*, represent an instance of speciation that has developed within the Pacific Coastal rain forest region of North America (Sheppe 1961; Gunn & Greenbaum 1986; Hogan *et al*. 1997).

We selected populations of *P. keeni* and *P. maniculatus* over the Pacific Coastal region from southeast Alaska through Oregon to investigate the historical demography and genetic interaction of these two taxa (Fig. 1a). We obtained gene sequence data from two mitochondrial markers, the cytochrome-*b* gene and the noncoding control region. Using these data we estimated divergence time, effective population size, and migration rate in order to understand population genetic structure of the two species and the postglacial recolonization of this area by both taxa, particularly *P. keeni*.

Materials and methods

Sampling

Tissues of 128 specimens were obtained from 43 localities in the Pacific Northwest, including the entire range of *P. keeni* (Table 1, Fig. 1). Most of the tissue samples were associated with voucher specimens at the Burke Museum (UWBM) and the University of Alaska Museum (UAM). Some tissue samples from the Burke Museum are indicated by collector number (AJS). All Burke Museum specimens were collected in 1994–2001, and tissues were held at –80 °C. All individuals included in this analysis are adults, based on reproductive condition and body measurements. We used total length (head + body plus tail) and tail length to assess species identification according to the morphological criteria of Gunn & Greenbaum (1986).

Laboratory techniques

DNA was extracted using the DNeasy Tissue Kit (Qiagen). We sequenced 479 bp of the cytochrome-*b* (cyt-*b*) gene for all 128 individuals. In alignment with the mouse mitochondrial genome (Bibb *et al*. 1981), this cyt-*b* fragment begins at position 14139 and ends at 14617. The primers used were L14724 (5′-CGAAGCTTGATATGAAAAACCATCGTTG-3′; Irwin *et al*. 1991) and H15906 (5′-CATTTCCGGTTTACAAGACCAGTGTAAT-3′; obtained from D. M. Irwin). Polymerase chain reaction (PCR) reactions were run in a total volume of 50 μL containing 1 × PCR buffer, 2.5 mM $MgCl_2$, 0.2 mM of each dNTP, 1.0 μM of each primer, 1 unit of *Taq* DNA polymerase (Hoffmann-La Roche Inc.) and 1–2 μL DNA. PCR reactions were carried out in a GeneAmp PCR System 9700 thermocycler (Applied Biosystems) as follows: 95 °C for 2 min; 30 cycles of 95 °C for 30 s, 53 °C for 1 min and 72 °C for 1 min; 72 °C for 7 min. Each round of PCR reactions also included one negative control to check for contamination.

We also sequenced 393 bp of the control region for all 128 individuals. In alignment with the mouse mitochondrial genome (Bibb *et al*. 1981), this fragment begins at position 15396 and ends at 15792. The first 22 bp of this sequence are included in the highly reserved $tRNA^{Pro}$ gene at 5′ end of the control region, and about 50–100 bp at 3′ end of the sequence are included in the central conserved block of the control region (Bibb *et al*. 1981; Taberlet 1996). The primers used were CTRL-L (5′-CACYWTYAACWCCCAAAGCT-3′; Bidlack & Cook 2001) and TDKD (5′-CCTGAAGTAGGAACCAGATG-3′; Kocher *et al*. 1993). PCR reactions were run in a total volume of 50 μL containing 1 × PCR buffer, 2.0 mM $MgCl_2$, 0.2 mM of each dNTP, 1.0 μM of each primer, 1 unit of *Taq* DNA polymerase (Hoffmann-La Roche, Inc.), and 1–2 μL DNA. The PCR reactions were carried out in a GeneAmp PCR System 9700 thermocycler (Applied Biosystems) as follows: 94 °C for 45 s; 35 cycles of 94 °C for 10 s, 50 °C for 15 s, and 72 °C for 45 s; 72 °C for 3 min. Each round of PCR reactions also included one negative control to check for contamination.

PCR products were purified with a QIAquick PCR Purification Kit (Qiagen). Sequencing reactions were carried out in a volume of 10 μL with the forward primers using a

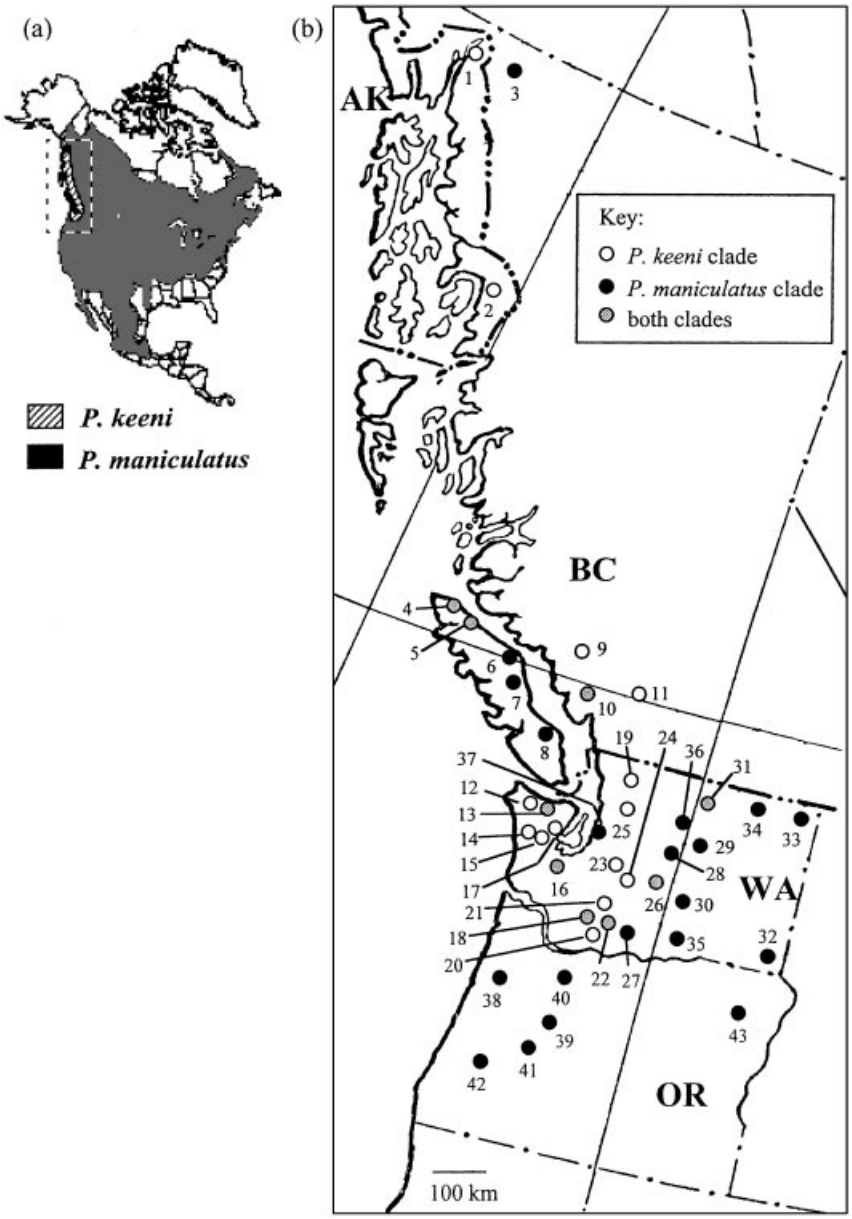

Fig. 1 Overall geographical distribution and sample localities of *Peromyscus keeni* and *P. maniculatus*. (a) Geographic distributions, modified from Hogan *et al*. (1997) and Verts & Carraway (1998). (b) Sample localities. Locality numbers as shown in Table 1. Each locality has individuals belonging to one or both of the two mitochondrial DNA clades, as indicated by symbols in the key. Some localities are displaced slightly for graphical clarity.

Big Dye Terminator Cycle Sequencing Ready Reaction Mix (Applied Biosystems). Samples were run on an ABI 377 automated sequencer (Applied Biosystems). To check the quality of obtained sequence data (in one direction), we sequenced some individuals twice and found identical outcomes. Sequences were aligned using SEQUENCHER 3.0 (Gene Codes Corp.). All sequences were deposited in GenBank (accession nos AY184501–AY184756).

Phylogenetic analysis

We used the computer program PAUP* 4.0b10 (Swofford 2002) to reconstruct phylogenetic relationships among individuals for both cyt-*b* and control region data, using the neighbour-joining method. By applying hierarchical likelihood ratio tests with the program MODELTEST 3.06 (Posada & Crandall 1998), we determined that the HKY model of substitution (Hasegawa *et al*. 1985) plus invariable sites (*I*) and a gamma distribution (Γ) of rate heterogeneity across variable sites provides the best fit to both the cyt-*b* and control region data sets. The estimated parameters under this model were Γ = 0.9191, *I* = 0.7014, Ti/Tv = 9.6873 for cyt-*b*; and Γ = 0.8122, *I* = 0.6677, Ti/Tv = 4.1718 for control region. We used bootstrap analysis with 1000 replicates to evaluate support for relationships, as implemented in PAUP* 4.0b10. One

Table 1 Localities (n = 43) and specimens (n = 128) of *Peromyscus* in the present study

Locality no.	State or Province	Locality: County or Region	Latitude	Longitude	Specimen nos
1	AK	Haines	59°19′ N	135°33′ W	UAM23665, UAM52642, UAM52643, UAM52644, UAM52645
2	AK	Mouth of Unuk River	56°05′ N	131°06′ W	UAM23485, UAM23487, UAM23490, UAM23491, UAM23492
3	BC	Atlin	59°33′ N	133°40′ W	UAM52689, UAM52690, UAM52691, UAM52692, UAM52693
4	BC	Vancouver Island	50°45′ N	127°37′ W	UWBM75398, UWBM75400
5	BC	Vancouver Island	50°31′ N	127°09′ W	UWBM75387, UWBM75390, UWBM75393
6	BC	Vancouver Island	49°59′ N	125°29′ W	UWBM75379, UWBM75380
7	BC	Vancouver Island	50°05′ N	125°27′ W	UWBM75351, UWBM75352, UWBM75353
8	BC	Vancouver Island	49°14′ N	124°06′ W	UWBM75485
9	BC	southwest BC	50°11′ N	123°08′ W	UWBM75463, UWBM75464
10	BC	southwest BC	49°43′ N	123°04′ W	UWBM75480,UWBM75481
11	BC	southwest BC	50°03′ N	121°47′ W	UWBM75449, UWBM75450
12	WA	Clallam	47°60′ N	123°59′ W	UWBM74029, UWBM74030
13	WA	Clallam	48°01′ N	123°35′ W	UWBM74932, UWBM74954, UWBM74957
14	WA	Grays Harbor	47°26′ N	123°33′ W	UWBM73657, UWBM73658, UWBM73659
15	WA	Mason	47°35′ N	123°08′ W	UWBM73644, UWBM73651, UWBM73654
16	WA	Thurston	46°53′ N	123°07′ W	UWBM74331, UWBM74334
17	WA	Jefferson	47°45′ N	123°03′ W	UWBM74036
18	WA	Skamania	46°00′ N	122°03′ W	UWBM74322, UWBM74323
19	WA	Whatcom	48°41′ N	121°55′ W	UWBM73848, UWBM73857
20	WA	Skamania	45°52′ N	121°42′ W	UWBM74226, UWBM74340
21	WA	Skamania	46°18′ N	121°40′ W	UWBM74001, UWBM74002
22	WA	Skamania	45°52′ N	121°38′ W	UWBM73840, UWBM73844
23	WA	Pierce	47°08′ N	121°37′ W	UWBM73999, UWBM74000
24	WA	Pierce	47°05′ N	121°24′ W	UWBM73990, UWBM73996, UWBM73997
25	WA	Snohomish	48°13′ N	121°20′ W	UWBM73879, UWBM73890, UWBM73895
26	WA	Kittitas	47°07′ N	121°03′ W	UWBM73804, UWBM73809, UWBM73822, UWBM73824
27	WA	Klickitat	45°58′ N	120°55′ W	UWBM73835, UWBM73836
28	WA	Chelan	47°46′ N	120°32′ W	UWBM73790, UWBM73791, UWBM73793
29	WA	Okanogan	48°03′ N	120°08′ W	UWBM73611, UWBM73615
30	WA	Grant	47°01′ N	119°59′ W	UWBM72501, UWBM72504, UWBM72505, UWBM72507
31	WA	Okanogan	48°38′ N	119°58′ W	UWBM73638, UWBM73639, UWBM73640
32	WA	Columbia	46°12′ N	117°42′ W	UWBM73914, UWBM73915, UWBM73918
33	WA	Pend Orelle	48°43′ N	117°12′ W	UWBM73945, UWBM73949, UWBM73956
34	WA	Okanogan	48°41′ N	118°52′ W	UWBM74052, UWBM74054
35	WA	Benton	46°02′ N	119°13′ W	UWBM75791, UWBM75795, UWBM75798, UWBM75805
36	WA	Chelan	48°18′ N	120°40′ W	UWBM73774, UWBM73775, UWBM73776
37	WA	King	47°44′ N	122°15′ W	UWBM75844, UWBM75846, UWBM75847, UWBM75848
38	OR	Tillamook	45°10′ N	23°49′ W	UWBM75432, UWBM75433, UWBM75436, UWBM75440, UWBM75447
39	OR	Jefferson	44°27′ N	121°47′ W	UWBM75420, UWBM75425, UWBM75426, UWBM75429
40	OR	Clackamas	45°15′ N	121°46′ W	UWBM75408, UWBM75409, UWBM75415
41	OR	Deschutes	43°51′ N	121°21′ W	UWBM74723, UWBM74724, UWBM74726, UWBM74728, UWBM74729
42	OR	Douglas	43°14′ N	122°26′ W	AJS131, AJS132, AJS134, AJS136, AJS139, AJS140, AJS145
43	OR	Umatilla	45°06′ N	118°36′ W	UWBM75403, UWBM75405, UWBM75407

sequence of *P. leucopus* from GenBank was used as outgroup in the cyt-*b* tree (GenBank no. AF131926; Bradley *et al*. 2000). No adequate outgroup was available for the control region data; we therefore constructed an unrooted tree. As a comparison to the neighbour-joining trees, we also constructed trees using parsimony and maximum likelihood methods.

Molecular diversity and time of divergence between the two clades

Mean nucleotide diversities within each clade and between clades were calculated with the program ARLEQUIN 2.000 (Schneider *et al*. 2000). We used the substitution model of Tamura & Nei (1993) with gamma corrections of 0.9191 for

cyt-*b* and 0.8122 for control region, as estimated with MODELTEST. Mean nucleotide diversity of cyt-*b* data between the two clades was used to estimate divergence time. To account for ancestral polymorphism, we corrected the molecular diversity by $D_{xy} = D - 0.5(D_x + D_y)$ (Edwards 1997), where D_x and D_y are mean molecular diversities of the *keeni* and *maniculatus* clades, respectively; D is the mean nucleotide diversity between two clades, and D_{xy} is corrected molecular diversity between two clades.

With ARLEQUIN 2.000 we can directly calculate D_x and D_y, but not D. We combined the two clades into one single population (t) and calculated mean nucleotide diversity for this population (D_t) using ARLEQUIN 2.000. Then we use D_x, D_y, and D_t to compute D by the following formula:

$$D_t N_t(N_t - 1)/2 = DN_xN_y + D_xN_x(N_x - 1)/2 + D_yN_y(N_y - 1)/2$$

where N_x, N_y and N_t are sample sizes of populations x, y and t, respectively, and where $N_t = N_x + N_y$.

A substitution rate of 0.028 per site per lineage per million years has been suggested for human–chimpanzee mitochondrial DNA cyt-*b*, which is equivalent to 5.6% pairwise sequence divergence per million years (Arbogast & Slowinski 1998). However, given that the relative rate of mitochondrial DNA evolution in rodents is typically faster than that of primates (Britten 1986; Li *et al*. 1987; Avise 2000), and that we have chosen an evolutionary model different from that of Arbogast & Slowinski (1998), we cannot use this substitute rate to calculate a specific divergence date for the *P. maniculatus* and *P. keeni* clades. We can only estimate approximately the general geological time period in which the divergence may have occurred.

Historical population dynamics

We applied distance-based methods (mismatch distribution in ARLEQUIN 2.000, Schneider *et al*. 2000) and coalescence/maximum likelihood methods (FLUCTUATE 1.3, Kuhner *et al*. 1998; MIGRATE 1.3.2, Beerli & Felsenstein 1999, 2001) to estimate effective population size, time and rate of expansion and migration rates among subpopulations.

With the program ARLEQUIN 2.000 we computed frequencies of pairwise differences between haplotypes (the mismatch distribution) to evaluate the hypothesis of recent population growth. A scenario of recent growth is expected to generate a unimodal distribution of pairwise differences, but exact mode of growth (stepwise, exponential or logistic growth) cannot be distinguished (Slatkin & Hudson 1991; Rogers & Harpending 1992; Rogers 1995; Polansk *et al*. 1998; Schneider & Excoffier 1999). ARLEQUIN also uses a nonlinear least squares approach to estimate parameters for a stepwise growth model: $\theta_0 = 2\,\mu N_0$ (before expansion), $\theta_1 = 2\,\mu N_1$ (after expansion) and $\tau = 2ut$ (time of expansion). Notice that $u = m_T\mu$ is the mutation rate for the entire DNA sequence under study, where m_T is the number of nucleotides of the sequence, and μ is the mutation rate per nucleotide (Rogers & Harpending 1992). N_0 and N_1 are effective population size of females before and after expansion, respectively. For the cyt-*b* sequence of this study, m_T is 479; μ is 0.028 per million years (Arbogast & Slowinski 1998), or 2.8×10^{-8} per generation, if we assume a generation time of one year (Kenagy & Barnes 1988). Therefore, u, mutation rate for the cyt-*b* sequence of this study, is 1.3×10^{-5} per generation.

Approximate confidence intervals for these parameters are obtained by a parametric bootstrap approach (1000 bootstrap replicates). The validity of the stepwise expansion model is tested using the same bootstrap parametric approach by a good-of-fitness statistic P = (number of SSD_{sim} larger or equal to SSD_{obs})/B. SSD_{sim} is the sum of square deviation (SSD) between the simulated mismatch distribution and the model expectation, SSD_{obs} is the SSD between observed mismatch distribution and the model expectation and B is the number of simulated samples (ARLEQUIN Version 2.000 Manual; Schneider & Excoffier 1999).

For a model of exponential expansion, we used the program FLUCTUATE 1.3, based on coalescence and maximum likelihood (Kuhner *et al*. 1998) to estimate a present-day value of θ ($\theta = 2\mu N_F$) and exponential growth rate (g). Here μ is mutation rate per nucleotide and N_F is effective population size of females. FLUCTUATE 1.3 uses a Markov chain Monte Carlo approach with Metropolis–Hastings importance sampling to search through genealogies, and it takes the area of the parameter space in which the log likelihood is no more than three units below the maximum as a rough 95% confidence interval (Kuhner *et al*. 1998). Transition/transversion ratio was set to be 9.6873 and 4.1718 for cyt-*b* and control region data, respectively. These values were obtained from MODELTEST 3.06. The initial tree was randomly constructed in the program. Five short chains of 1000 steps each and 2 long chains of 15 000 steps each were run. Sampling increment was set to be 20.

In order to evaluate relationships among populations of each species, we used the program MIGRATE 1.3.2 to estimate effective population sizes ($\theta = 2\mu N_F$) and migration rates ($M = 2mN_F$) among populations (Beerli & Felsenstein 2001). MIGRATE uses a likelihood ratio test and extended coalescent theory to estimate parameters, and it uses a Markov chain Monte Carlo approach with Metropolis–Hastings importance sampling to search through genealogy space. It assumes constant effective population sizes for each population, but allows various effective population sizes for different populations. To use the program on our data, we assigned individuals of each clade into two populations according to geographical ranges (Fig. 4). Note that this designation is somewhat arbitrary. In the

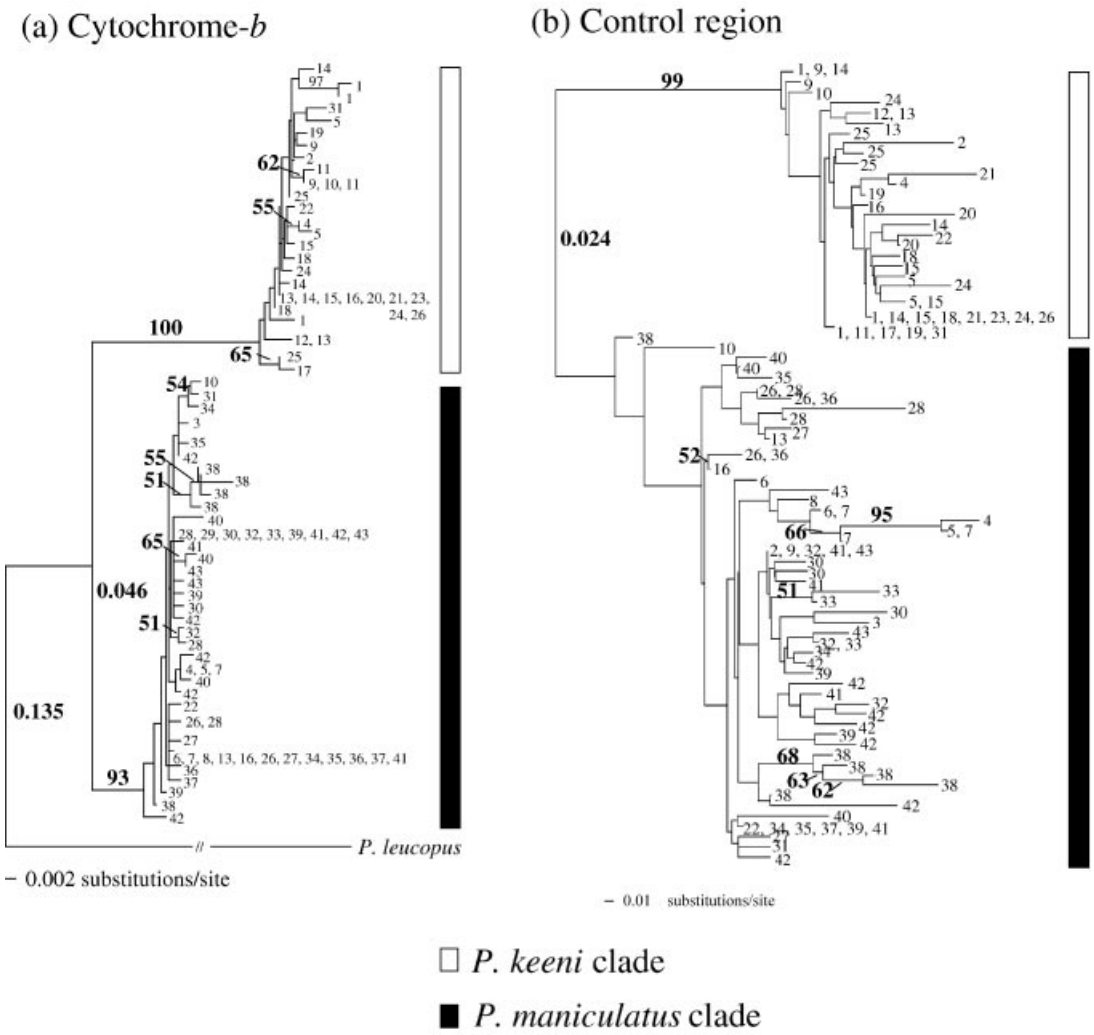

Fig. 2 Neighbour-joining trees for mitochondrial-DNA sequence data based on the HKY + Γ + I model of substitution. (a) Cytochrome-*b* gene (α = 0.9191, p_{inv} = 0.7014, transition/transversion ratio = 9.6873). (b) Control region (α = 0.8122, p_{inv} = 0.6677, transition/transversion ratio = 4. 1718). Bootstrap support of > 50% in 1000 replicates is shown above branches. Branch lengths (substitutions/site) between major clades are shown below branches. Numbers at each branch tip indicate localities where the haplotype occurs. Note that the cytochrome-*b* tree is rooted, while the control region tree is unrooted.

discussion section we address the affects of population designation on parameter estimation. All available samples are included in the analysis in order to provide the best representation of overall genetic variation. F_{ST} estimates of effective population sizes and migration rates were used as initial values. Transition/transversion ratio was estimated from data. Ten short chains with 5000 sampled genealogies each and 3 long chains with 50 000 sampled genealogies each were run. One of every 100 reconstructed genealogies was sampled; therefore the total number of reconstructed genealogies was 500 000 for each short chain and 5000 000 for each long chain. Heating was set to be active, with four temperatures of 1.0, 1.2, 2.5 and 3.0. Three to four runs were repeated for each data set to check consistencies of the results.

Results

To investigate the evolutionary history of *P. keeni* and *P. maniculatus* within the Pacific Northwest of North America we sampled 128 individuals from 43 localities from southeastern Alaska through Oregon (Table 1, Fig. 1b). Evidence from both the cyt-*b* gene and the control region indicates the existence of two clearly distinguishable clades that are sympatric within some areas of southern British Columbia and western Washington (Figs 1 and 2). Of the 128 individuals, 50 belonged to the *keeni* clade and 78 to *maniculatus*. Distances (substitutions per site) between the two clades were 0.046 for cyt-*b* and 0.024 for control region (Fig. 2). Both clades contained numerous haplotypes, but we found no strong support for geographical subdivision within either clade

Table 2 Mean nucleotide diversity (substitutions per site, ± SD) of the two *Peromyscus* clades and estimated time of divergence

Marker	Mean nucleotide diversity			Corrected divergence (D_{xy})	Divergence time of the two clades
	keeni clade (D_x)	*maniculatus* clade (D_y)	between clades (D)		
Cyt-*b*	0.00629 (± 0.00370)	0.00497 (± 0.00303)	0.0484	0.0428	Middle to late Pleistocene
Control region	0.00696 (± 0.00417)	0.01097 (± 0.00608)	0.0396	0.0306	—

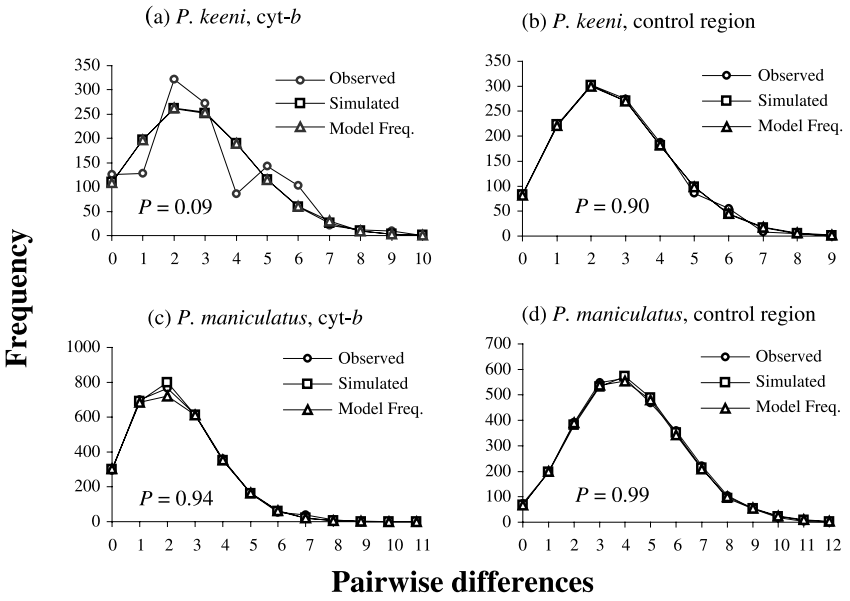

Fig. 3 Mismatch distributions of two *Peromyscus* clades and two mitochondrial-DNA markers. These curves represent the frequency distribution of pairwise differences (a) *P. keeni*, cytochrome-*b* (b) *P. keeni*, control region (c) *P. maniculatus*, cytochrome *b* (d) *P. maniculatus*, control region. *P*-values represent probability that the variance of the simulated data set is equal to or greater than the observed data set (see Materials and methods).

(Fig. 2). Alternative computation of trees using parsimony and maximum likelihood methods resulted in similar tree topologies.

Mean nucleotide diversity for cyt-*b* data was slightly higher in the *keeni* clade than in the *maniculatus* clade, and between-clade diversity was an order of magnitude higher than within-clade diversity (Table 2). The corrected nucleotide diversity between two clades suggests that divergence of the two clades may have occurred in middle to late Pleistocene (Table 2). Mean nucleotide diversity for the control region data was lower in the *keeni* clade than in the *maniculatus* clade. Mean nucleotide diversity between clades was slightly higher for cyt-*b* than that for the control region (Table 2).

Both the tree topology (Fig. 2) and the mismatch distribution (Fig. 3) indicate recent demographic expansion for each clade. From the neighbour-joining trees (Fig. 2) it is apparent that each clade contains one or two widespread haplotypes and numerous small, localized haplotypes. It is also apparent that branches within each clade are small and similar in length, especially in the cyt-*b* tree. These patterns suggest recent expansions in population size and geographical range. Three out of the four mismatch-distribution graphs (Fig. 3) indicate close similarity between observed and expected frequencies of pairwise differences, which indicates recent increase in population size.

In order to recover the details of historical population expansion, we applied two simplified models of expansion to our data. This allowed us to project effective female population size and the time and rate of expansion. In the stepwise expansion model, the expansion for the *keeni* clade, based on cyt-*b*, was estimated at about 120 000 generations ago, or 120 000 years ago, assuming a generation time of one year. This is based on the τ-value of 3.062 (Table 3) and a mutation rate (u) of 1.3×10^{-5} per sequence per generation. For the *maniculatus* clade, expansion occurred about 90 000 years ago based on cyt-*b* (Table 3). Estimated effective population size after expansion (θ_1) is several thousand times higher than before expansion (θ_0) (Table 3).

The exponential expansion model also indicated a rapid increase in effective population size for both clades (Table 3). Assuming a mutation rate (μ) of 2.8×10^{-8} per nucleotide per generation, the effective female population size of the *keeni* clade would be 5.157×10^5 at present and 5.150×10^5 1000 generations ago for cyt-*b* data. For *P. maniculatus*, the effective female population size would be 8.964×10^4 at present and 8.857×10^4 1000 generations ago.

Table 3 Estimated parameters of population expansion for (a) exponential and (b) stepwise expansion models
(a) Exponential expansion model*

Marker	Clade	No. of individuals	No. of haplotypes	Exponential expansion model	
				$\theta = 2\mu N_F$	g
Cyt-*b*	*keeni*	50	25	0.0578 (± 0.0090)	627.41 (± 78.72)
	maniculatus	78	34	0.1004 (± 0.0095)	901.30 (± 58.67)
Control region	*keeni*	50	25	0.1848 (± 0.0405)	900.88 (± 59.41)
	maniculatus	78	51	0.4893 (± 0.0494)	561.01 (± 25.10)

(b) Stepwise expansion model†

		Stepwise expansion model		
		$\tau = 2ut$	$\theta_0 = 2\,\mu N_0$	$\theta_1 = 2\,\mu N_1$
Cyt-*b*	*keeni*	3.062 (1.284, 4.723)	0.008 (0.000, 1.584)	18.518 (5.429, 6776.018)
	maniculatus	2.326 (1.229, 2.888)	0.000 (0.000, 1.333)	237.109 (10.468, 8142.109)
Control region	*keeni*	2.707 (1.344, 3.324)	0.000 (0.000, 1.468)	2725.00 (29.995, 9542.500)
	maniculatus	4.001 (2.545, 5.516)	0.275 (0.000, 1.754)	75.000 (26.760, 7864.875)

*Population parameters under the exponential model are given as maximum likelihood estimates (± SD). †Population parameters under the stepwise model are given as estimates (95% confidence limits).

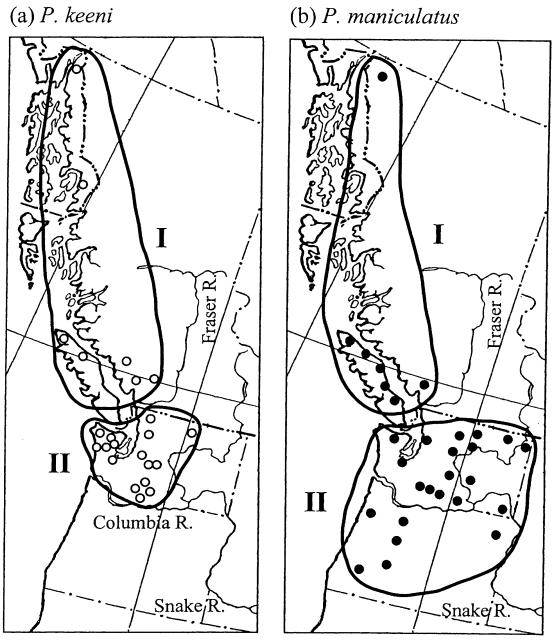

Fig. 4 Maps showing population designations for the MIGRATE analysis, with population I north of the Fraser River and population II south of the Fraser River. Dots indicate localities of specimens of the two mitochondrial-DNA clades (a) *P. keeni* and (b) *P. maniculatus*.

Results of the MIGRATE analysis indicate a strong southward migration for both the *keeni* and *maniculatus* clades (Fig. 4, Table 4a). For the *keeni* clade, the estimated population size of the northern population (area I) was about 10 times larger than that for the south (II). Migration from the north to the south was strong, whereas from south to north was negligible. Estimated population sizes for the *maniculatus* clade were similar in the north and south. Migration rates of *maniculatus* were comparable to those of *keeni*, indicating strong southward migration.

Individuals belonging to the *P. keeni* clade typically have longer tails and a higher tail to head + body ratio than those of the *P. maniculatus* clade (Fig. 5). The absolute tail length distinguishes the *keeni* clade more strongly than the ratio of tail to head + body. Among individuals of the *P. keeni* clade, 94% had tail length ≥ 95 mm; among *P. maniculatus* 91% had tail length < 95 mm (Fig. 5a). Among individuals of the *P. keeni* clade, 88% had a tail to head + body ratio ≥ 1.10; among *P. maniculatus*, 81% had a tail to head + body ratio < 1.10 (Fig. 5b).

Discussion

Peromyscus keeni and *P. maniculatus* form two clearly distinguishable mitochrondrial clades in the Pacific Northwest coastal rain forest ecosystem. However, because the morphological differentiation between the two forms is not strong and because the mitochondrial markers are haploid, we cannot conclude that introgression is occurring between these two sister taxa. This question needs to be addressed with additional markers from the nuclear genome. Analysis of historical demography suggests recent expansion of both clades and asymmetrical migration between northern and southern populations. The strong

Table 4 Estimated effective population sizes and migration rates
(a) Original population designation (Fig. 4): I. North of the Fraser River; II. South of the Fraser River

Clade	Population	*n*	Effective population size, scaled by mutation rate ($\theta = 2\mu N_F$)	Migration rate ($2mN_F$)	
				I to II	II to I
keeni	I	18	0.0217 (0.0146, 0.0307)	—	2.44×10^{-8} (1.83×10^{-8}, 3.1×10^{-5})
	II	32	0.0029 (0.0013, 0.0044)	7.08 (2.75, 12.94)	—
maniculatus	I	14	0.0030 (0.0019, 0.0048)	—	1.63×10^{-8} (1.22×10^{-8}, 4.1×10^{-5})
	II	64	0.0075 (0.0055, 0.0106)	9.49 (5.08, 16.47)	—

(b) Alternative population designation (Fig. 6): I. Southeastern Alaska; II. Washington and southern British Columbia

keeni	I	10	0.0055 (0.0012, 0.0141)	—	2.00 (0.54, 4.63)
	II	40	0.0018 (0.0014, 0.0025)	3.34×10^{-9} (2.50×10^{-9}, 0.8×10^{-5})	—

Parameters are presented as maximum likelihood estimates (95% confidence limits).

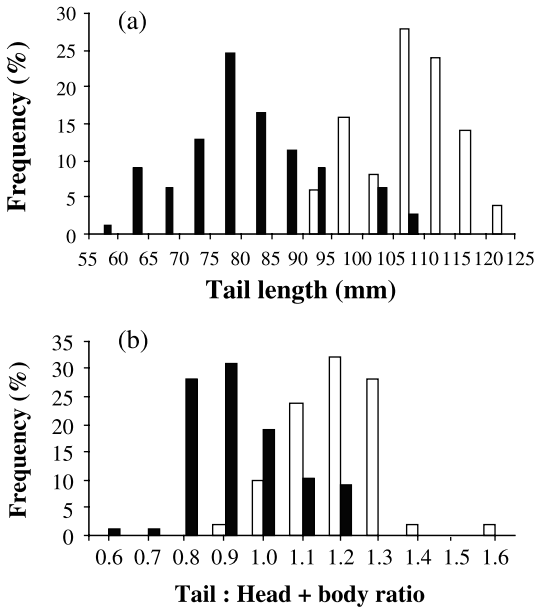

Fig. 5 Frequency distribution of tail length and ratio of tail to head + body for the 128 individuals according to the two mitochondrial-DNA clades of *Peromyscus keeni* (open columns) and *P. maniculatus* (black columns). (a) Tail length. Categories are 55–59 mm, 60–64 mm, etc. (b) Ratio of tail length to head + body length. Categories are 0.55–0.64, 0.65–0.74, etc.

southward migration indicated for the *keeni* clade differs from the expectation of a northward postglacial expansion.

Phylogenetic analysis

Comparison of the cyt-*b* and control region trees indicates a higher rate of evolution in the control region and suggests saturation in the control region. A higher evolutionary rate in the control region is suggested by higher within-clade nucleotide diversity in the control region than in cyt-*b* (Table 2). The saturation in the control region is suggested by a lower transition/transversion ratio in the control region (4.1718) than in cyt-*b* (9.6873) and the observation that a higher proportion of the total genetic diversity is distributed within clades for the control region than for cyt-*b* (Fig. 2, Table 2).

Between-clade sequence divergence is higher in cyt-*b* than in the control region. This result is surprising because the control region contains variable blocks that evolve faster than the entire mtDNA molecule (Taberlet 1996). Our observation may be accounted for by the following two factors. First, about one-fourth of our control region sequence belongs to two conserved blocks flanking one variable block (the $tRNA^{pro}$ gene at the 5′ end and the central conserved block at the 3′ end of the variable block). Data from these conserved regions may have reduced estimates of overall sequence divergence between clades for the control region data. If we had restricted our control region sequencing to only the most variable parts of the control region, it is likely that the between-clade sequence divergence would be substantially higher than we observed. Second, multiple substitutions at one site (saturation) may have occurred between clades in the control region, as discussed above. Since we also observed that within-clade nucleotide diversities were higher in the control region than in cyt-*b* for both clades (Table 2), we believe that saturation between the two clades has played a role in causing the lower between-clade divergence in control region. In the above discussion, we assume that the partial cyt-*b* region that we sequenced is roughly equivalent in terms of average substitution rate to estimates for the whole gene.

Morphological and ecological differentiation

The only widely used criteria for distinguishing *P. keeni* and *P. maniculatus* in the field are tail length and the ratio

of tail to head + body (Sheppe 1961; Allard *et al*. 1987). Allard *et al*. (1987) used the criterion of tail length > 96 mm to distinguish adult *P. keeni* (designated as *P. oreas* in their paper) from *P. maniculatus*. Our data confirmed the general validity of this criterion (Fig. 5). In our samples, all the *maniculatus* individuals with tail length > 95 mm were found either in northwestern Oregon or on Vancouver Island, indicating an uneven distribution of long-tailed *maniculatus* individuals. The significance of an elongated tail for arboreal adaptation (climbing agility and balance) has been suggested for *Peromyscus* (Horner 1954). Ecological differentiation between the two taxa has not been established clearly and probably varies at different places (Sheppe 1961; Dice 1968). One of our general impressions in the field is that where the two forms overlap, *P. keeni* tends to occur in moister and more mature forests. Partial sterility in attempts to cross the two forms has been found in laboratory experiments (Liu 1954), but no decisive conclusion can be drawn since *P. keeni* is usually infertile in the laboratory even when mated with its own kind (Dice 1968). Distinct karyotypic differences were found between the two taxa in coastal regions of Alaska, British Columbia, and Washington, leading previous investigators to conclude that no hybridization occurred between them (Gunn & Greenbaum 1986; Hogan *et al*. 1993). Our mitochondrial data alone are not adequate for concluding whether hybridization occurs between the two taxa, since we do not have other independent criteria (e.g. morphological, ecological, genetic) to distinguish them confidently.

Divergence time and population expansion

Using molecular data to estimate divergence time of lineages has been discussed intensively during the last two decades, partly because of the importance of this issue in evolutionary studies. A number of factors may affect the estimation of divergence times from molecular data. For example, the variability of substitution rate across lineages and among loci, multiple substitution at one site, ancestral polymorphism, change of effective population size over time, evolutionary models selected and the stochastic nature of genetic change (for reviews, see Avise 2000; Arbogast *et al*. 2002). Using multiple independent loci is expected to decrease variance in coalescence time and thus improve the accuracy of estimating divergence time (Arbogast *et al*. 2002). As a further direction of our study, we are currently collecting data on various nuclear loci to understand better the evolutionary history of *P. manicualtus* and *P. keeni*. At the present time, with just mitochondrial data available, we can only estimate approximately that the two clades diverged during middle to late Pleistocene. Fossil records indicate that *Peromyscus* was not a common member of the Pliocene and early Pleistocene faunas, and that the *P. maniculatus* and *P. leucopus* groups are products of late Pleistocene adaptive radiation (Hibbard 1968).

The use of pairwise differences between sampled DNA sequences (the mismatch distribution) to infer modes and parameters of population history has been the focus of numerous recent studies. Several early initial papers (Slatkin & Hudson 1991; Felsenstein 1992; Rogers & Harpending 1992) explored potentials and limitations of pairwise difference-based estimation and triggered much of the later discussion. Theoretical considerations and application to human evolution have demonstrated that the mismatch distribution contains detectable impacts left by population history (Harpending *et al*. 1998; Schneider & Excoffier 1999), especially for populations that are not at demographic equilibrium. However, the efficiency of this method (as compared with phylogenetic estimates) and realistic conclusions that can be inferred from its estimates are of concern (Felsenstein 1992; Bertorelle & Slatkin 1995; Polanski *et al*. 1998; Schneider & Excoffier 1999). To make the infinite-site substitution model more realistic, Schneider & Excoffier (1999) improved the original method of Rogers & Harpending (1992) by considering rate heterogeneity across sites. They demonstrated that the expansion time (τ) and the population size before expansion (θ_0) can be estimated without much bias and with reasonable precision, while the estimation of θ_1 (population size after expansion) is biased upward (Schneider & Excoffier 1999). In this study we applied their methods (ARLEQUIN 2.000) to our cyt-*b* data. The large difference between θ_0 and θ_1 indicates a significant increase in effective population size. The estimated expansion times are 120 000 years ago for *P. keeni* and 90 000 years ago for *maniculatus*, in the late Pleistocene, when global climate was producing dramatic glaciation cycles (Pielou 1991). These results suggest that the major increases in effective population size of *Peromyscus* were associated with a longer period of geological history, not just with the most recent glacial retreat of about 12 000 years ago (Pielou 1991). It should be emphasized that these estimates of expansion time are based on a mutation rate of 1.3×10^{-5} per sequence per generation. A different calibration of mutation rate would certainly affect the estimate of expansion time. Furthermore, the stepwise expansion scenario contains a very simplified reflection of the demographic history. The expansion indicated by the models may actually have consisted of a series of demographic fluctuations over hundreds of thousands of years, but what we observe today at the genetic level is simply the net result of all these fluctuations. Thus, it is not surprising that the estimated expansion times (120 000 and 90 000 years ago) precede the last glacial retreat (about 12 000 years ago). Fossil evidence indicates apparent range expansion during this last glacial retreat (Pielou 1991; FAUNMAP 1994).

Sudden demographic expansion is expected to generate a unimodal distribution of pairwise differences (Rogers &

Harpending 1992). This is because for a population after expansion but before demographic equilibrium, fewer coalescence events occur and thus pairwise differences often fall in an intermediate range. However, Bertorelle & Slatkin (1995) noted that such a distribution might also be the consequences of selection on a particular haplotype (Di Rienzo & Wilson 1991) or of high levels of homoplasy due to high mutation rates (Lundstrom *et al*. 1992). Bottlenecks in population size also generate similar unimodal distributions, but with elevated upper-tail probabilities (Rogers & Harpending 1992). This is perhaps because some older lineages are preserved in a bottlenecked population, as compared to a population with constant small size before expansion. Therefore, an observed unimodal distribution of pairwise differences needs to be interpreted with caution. Among our data sets three of four (*P. keeni* control region, *P. manicualtus* cyt-*b* and *P. maniculatus* control region) showed a good fit with a unimodal distribution without elevated upper-tail probabilities ($P \geq 0.90$, Fig. 3). Because the time scale of our study is relatively small (several hundred thousands of years), a high degree of homoplasy within each clade seems an unlikely explanation of the observed patterns. The effect of selection is also likely to be trivial in our system, as within-species selection at the mitochondrial genome may not be strong enough to promote rapid spreading of favoured alleles in the relatively short time scale of our study. Therefore, rapid expansion seems to be the most plausible explanation for the observed unimodal distribution of pairwise differences.

The *P. keeni* cyt-*b* data set showed a two-peak pattern of pairwise differences instead of a unimodal distribution ($P = 0.09$). This is surprising, as the two mitochondrial markers are linked and should have the same population demographic history. However, for the *maniculatus* clade, the two markers showed similar unimodal distirbutions (Fig. 3).

Felsenstein (1992) noted that estimates incorporating phylogenetic information make more efficient use of the data than estimates based on pairwise differences. However, it is not clear how much of the total information in the data is actually contained in pairwise differences that may allow reasonable reflection of population history. In this study we could not compare the two methods directly because we did not know the true values of the parameters. Estimates obtained by FLUCTUATE indicated that the *maniculatus* clade has larger present-day effective population size than the *keeni* clade. Growth rates (g) indicate rapid expansion. Repetitive runs of FLUCTUATE generated very similar sets of estimates (results not presented). In FLUCTUATE the estimates of g are biased upward, and the estimates for θ are also biased upward slightly due to correlation between the parameters (Kuhner *et al*. 1998). Adding additional loci can rapidly reduce this bias (Kuhner *et al*. 1998).

Since both ARLEQUIN and FLUCTUATE analyse just one model of expansion, we could not estimate which model better describes population history of our samples. In principle, different modes of population demographic history can be directly evaluated against the data by using the maximum likelihood algorithm (Kuhner *et al*. 1998).

Migration

Historical movements of the two clades were investigated by estimating migration rates among current populations. In conventional phylogeography, historical movements are often inferred from clustering of haplotypes (Avise 2000). However, this method is not feasible for our data as the phylogenetic trees do not contain strongly supported major clusters of haplotypes. Under such a circumstance, MIGRATE provides an opportunity to infer historical movements.

The population designations of our samples were arbitrary. To check whether estimated migration patterns fit the conventional hypothesis of postglacial recolonization from southern or eastern refugia (Byun *et al*. 1997; Arbogast 1999), we initially divided the samples into two populations (Fig. 4): north of the Fraser River and south of the Fraser River. Our analysis revealed that migration rates between samples from the Cascade Mountains and samples from Olympic Peninsula and Vancouver Island were substantial in both directions (results not shown), justifying our inclusion of all southern samples in one population.

Results for the *keeni* clade were surprising (Table 4a). If *keeni* individuals in Alaska and British Columbia were recolonizers from the south, i.e. Washington, as suggested by the southern refugium hypothesis (Byun *et al*. 1997; Arbogast 1999), then the direction of migration should be from south to north. However, our estimates of migration rates indicate a strong migration from Alaska and BC southward to Washington. A possible explanation for this discrepancy is that postglacial recolonizers may have actually originated from the coastal refugia of Vancouver Island and southern continental BC, instead of further south from Washington. If we modify the population designations to combine samples from Washington, southern BC and Vancouver Island, then a strong northward migration to Alaska is revealed (Fig. 6, Table 4b). This result indicates both the power of migration analysis for testing the alternative hypothesis of population movement and the influence of population designation on estimates.

For the *maniculatus* clade, a strong southward migration from Alaska and BC to Washington and Oregon was again revealed. Two scenarios can be used to explain such a southward migration: (1) recolonizers originated from Vancouver Island and southern BC, as in the case of the *keeni* clade, or (2) recolonizers originated from an eastern refugium. Because we do not have an adequate number of samples from Alaska and northern BC, we could not

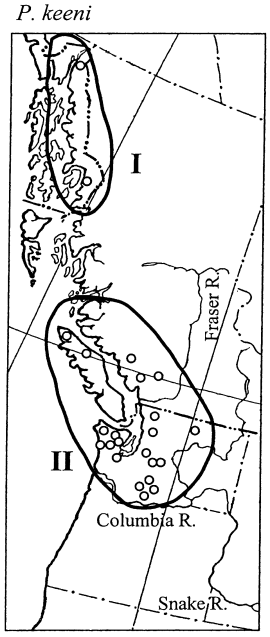

Fig. 6 Map showing alternative population designations of the *keeni* clade for MIGRATE analysis, with population I in southeastern Alaska and population II in Washington and southern British Columbia.

change population designations as we did with the *keeni* clade and thus we were unable to test the first scenario. For the second scenario our results are actually not incompatible with the conventional hypothesis of a southeastern refugium (Arbogast 1999). According to this hypothesis *maniculatus* in Alaska and northern BC may have come from southeastern North America following glacial retreat. If this scenario is true, then the apparent southward coastal migration might have occurred after initial northward recolonization along the cordillera. It could also be the reflection of shared polymorphism between ancestral *maniculatus* populations in eastern and southern refugia. We will need samples from Alaska and northern BC and eastern North America to distinguish these hypotheses further. A recent study on Hall's babbler *Pomatostonus halli* (Miura & Edwards 2001) also reported asymmetric migration between populations. It is not clear whether this asymmetry reflects the real pattern of migration, or whether it is simply a bias caused by using a single mitochondrial locus. One would need to analyse data from nuclear loci to address this question.

The search strategy of MIGRATE can significantly influence accuracy of estimates (MIGRATE documentation). The longer the runs, the more reliable are the estimates. Our experience showed that, for our data, performance of the program stabilized after searching 500 000 genealogies for each short chain and 5000 000 genealogies for each long chain. The use of 'heating' also improved performance. One practical way to check performance of the program is to repeat the analysis several times with the same settings. If repetitive runs generate similar results, then the estimates are more reliable; if not, it is preferred to increase length of runs.

While analysing natural populations, some assumptions of models are usually violated. This will certainly affect results of the analyses, but perhaps not in a significant way. For example, in our analysis, the assumption of constant population size for MIGRATE was apparently violated, since both clades have undergone recent expansion. The resulting effective population sizes are thus averages over a period of time and ought to be smaller than the present-day effective population sizes. However, this bias may not change the major pattern of the migration matrix, i.e. relative magnitude of northward vs. southward migration rates. The assumption of a panmictic population for FLUCTUATE was also violated, because the results of MIGRATE indicated skewed migration between southern and northern populations. This violation could make the coalescence process slower than expected for a panmictic population. As a result, θ might be overestimated. The estimates will be more reliable if effective population size, migration rate and population growth rate can be estimated together. This work is currently being pursued by Peter Beerli, Mary Kuhner and colleagues (pers. comm.).

Model-based estimates of population parameters provide new and exciting ways of revealing the evolutionary history of species. With careful analysis of assumptions, cautious interpretation of results and comparison with other sources of information, these model-base estimates can be extremely helpful for testing alternative hypotheses and pointing out directions for further study.

Acknowledgements

This project was funded by the Mammalogy Program at the Burke Museum, the Graduate School and Royalty Research Funds of the University of Washington, and a Grant in Aid to XGZ from the American Society of Mammalogists. Special thanks to Nicholas Chew, Hopi Hoekstra and Sacha Vignieri for collection of specimens in the field and Joe Cook for tissue samples from the University of Alaska Museum. We thank Peter Beerli, Robb Brumfield, Scott Edwards, Joseph Felsenstein and two anonymous reviewers for advice on the research and manuscript.

References

Allard MW, Gunn SJ, Greenbaum IF (1987) Mensural discrimination of chromosomallycharacterized *Peromyscus oreas* and *P. maniculatus*. *Journal of Mammalogy*, **68**, 402–406.

Arbogast BS (1999) Mitochondrial DNA phylogeography of the New World flying squirrels (*Glaucomys*): implications for Pleistocene biogeography. *Journal of Mammalogy*, **80**, 142–155.

Arbogast BS, Edwards SV, Wakeley J, Beerli P, Slowinski JB (2002) Estimating divergence times from molecular data on phylogenetic

and population genetic time scales. *Annual Review of Ecology and Systematics*, **33**, 707–740.

Arbogast BS, Kenagy GJ (2001) Comparative phylogeography as an integrative approach to historical biogeography. *Journal of Biogeography*, **28**, 819–825.

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

Arnold ML (1997) *Natural Hybridization and Evolution*. Oxford University Press, NewYork.

Avise JC (2000) *Phylogeography: the History and Formation of Species*. HarvardUniversity Press, Cambridge, MA.

Barrowclough GF, Gutierrez RJ, Groth JG (1999) Phylogeography of spotted owl (*Strix occidentalis*) populations based on mitochondrial DNA sequences: gene flow, genetic structure, and a novel biogeographic pattern. *Evolution*, **53**, 919–931.

Beerli P, Felsenstein J (2001) Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. *Proceedings of the National Academy of Sciences USA*, **98**, 4563–4568.

Beerli P, Felsentein J (1999) Maximum likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. *Genetics*, **152**, 763–773.

Bertorelle G, Slatkin M (1995) The number of segregating sites in expanding human populations with implications for estimates of demographic parameters. *Molecular Biology and Evolution*, **12**, 887–892.

Bibb MJ, VanEtten RA, Wright CT, Walberg MW, Clayton DA (1981) Sequences and gene organization of mouse mitochondrial DNA. *Cell*, **26**, 167–180.

Bidlack AL, Cook JA (2001) Reduced genetic variation in insular northern flying squirrels (*Glaucomys sabrinus*) along the North Pacific Coast. *Animal Conservation*, **4**, 283–290.

Blair WF (1950) Ecological factors in speciation of *Peromyscus*. *Evolution*, **4**, 253–275.

Bonhomme F, Guenet J (1996) The laboratory mouse and its wild relatives. In: *Genetics Variants and Strains of the Laboratory Mouse* (eds Lyon MF, Rastan S, Brown SDM), pp. 1577–1596. Oxford University Press, Oxford.

Bradley RD, Tiemann-Boege I, Kilpatrick CW, Schmidly D (2000) Taxonomic status of *Peromyscus boylii sacarensis*: inferences from DNA sequences of the mitochondrial cytochrome *b* gene. *Journal of Mammalogy*, **81**, 875–884.

Bradshaw HD, Otto KG, Frewen BE, McKay JK, Schemske DW (1998) Quantitative train loci affecting differences in floral morphology between two species of Monkeyflower (*Mimulus*). *Genetics*, **149**, 367–382.

Britten RJ (1986) Rates of DNA sequence evolution differ between taxonomic groups. *Science*, **231**, 1393–1398.

Byun SA, Koop BF, Reimchen TE (1997) North American black bear mtDNA phylygeography: implications for morphology and the Haida Gwaii glacial refugium controversy. *Evolution*, **51**, 1647–1653.

Conroy CJ, Cook JA (2000) Phylogeography of a post-glacial colonizer: *Microtus longicaudus* (Rodentia: Muridae). *Molecular Ecology*, **9**, 165–175.

Coyne JA, Orr HA (1998) The evolutionary genetics of speciation. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences*, **353**, 287–305.

Darwin C (1859) *The Origin of Species by Means of Natural Selection*, 6th edn. The Modern Library, New York.

Di Rienzo A, Wilson AC (1991) Branching pattern in the evolutionary tree for human mitochondrial DNA. *Proceedings of the National Academy of Sciences USA*, **88**, 1597–1601.

Dice LR (1968) Speciation. In: *Biology of* Peromyscus *(Rodentia)*. Special Publication no. 2 (ed. King JA), pp. 75–97. American Society of Mammalogists.

Dobzhansky T (1951) *Genetics and the Origin of Species*. Columbia University Press, New York.

Edwards SV (1997) Relevance of microevolutionary processes to higher level molecular systematics. In: *Avian Molecular Evolution and Systematics* (ed. Mindell DP), pp. 251–278. Academic Press, San Diego, CA.

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

Emerson BC, Paradis E, Thebaud C (2001) Revealing the demographic histories of species using DNA sequences. *Trends in Ecology and Evolution*, **16**, 707–716.

FAUNMAP Working Group (1994) *FAUNMAP: a Database Documenting Late Quaternary Distributions of Mammal Species in the United States*. Illinois State Museum, Springfield, IL.

Felsenstein J (1992) Estimating effective population size from samples of sequences: inefficiency of pairwise and segregating sites as compared to phylogenetic estimates. *Genetical Research*, **59**, 139–147.

Gunn SJ, Greenbaum IF (1986) Systematic implications of karyotypic and morphologic variation in mainland *Peromyscus* from the Pacific Northwest. *Journal of Mammalogy*, **67**, 294–304.

Hacia JG (2001) Genome of the apes. *Trends in Genetics*, **17**, 637–645.

Harpending HC, Batzer MA, Gurven M, Jorde LB, Rogers AR, Sherry ST (1998) Genetic traces of ancient demography. *Proceedings of the National Academy of Sciences USA*, **95**, 19671–11967.

Harrison RG (1993) Hybrids and hybrid zone: historical perspective. In: *Hybrid Zones and the Evolutionary Process* (ed. Harrison RG), pp. 3–12. Oxford University Press, Oxford.

Hasegawa M, Kishino K, Yano T (1985) Dating the human–ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution*, **22**, 160–174.

Hibbard CW (1968) Palaeontology. In: *Biology of* Peromyscus *(Rodentia)*. Special Publication no. 2. (ed. King JA), pp. 6–26. American Society of Mammalogists.

Hogan KM, Davis SK, Greenbaum IF (1997) Mitochondrial-DNA analysis of the systematic relationships within the *Peromyscus maniculatus* species group. *Journal of Mammalogy*, **78**, 733–743.

Hogan KM, Hedin MC, Koh HS, Davis SK, Greenbaum IF (1993) Systematic and taxonomic implications of karyotypic, electrophoretic, and mitochondrial-DNA variation in *Peromyscus* from the Pacific Northwest. *Journal of Mammalogy*, **74**, 819–831.

Horner BE (1954) Arboreal adaptations of *Peromyscus* with special reference to use of the tail. *Contributions of the Laboratory of Vertebrate Biology, University of Michigan*, **61**, 1–85.

Hudson RR (1990) Gene genealogies and the coalescent process. *Oxford Surveys in Evolutionary Biology*, **7**, 1–44.

Irwin DM, Kocher TD, Wilson AC (1991) Evolution of the cytochrome *b* gene of mammals. *Journal of Molecular Evolution*, **32**, 128–144.

Kenagy GJ, Barnes BM (1988) Seasonal reproductive patterns in four coexisting rodentspecies from the Cascade Mountains, Washington. *Journal of Mammalogy*, **69**, 274–292.

King JA, ed. (1968) *Biology of* Peromyscus *(Rodentia)*. Special Publication no. 2. American Society of Mammalogists.

Kirkland GL, Layne JN., eds (1989) *Advances In The Study of* Peromyscus *(Rodentia)*. Texas Tech University Press, Lubbock, Texas.

Kocher TD, Conroy JA, McKaye KR, Stauffer JR (1993) Similar morphologies of cichlid fish in Lakes Tanganyika and Malawi are due to convergence. *Molecular Phylogenetics and Evolution*, **2**, 158–165.

Kuhner MK, Yamato J, Felsenstein J (1998) Maximum likelihood estimation of population growth rates based on the coalescent. *Genetics*, **149**, 429–434.

Lewontin RC (1974) *The Genetics Basis of Evolutionary Change*. Columbia University Press, New York.

Li W-H, Tanimura N, Sharp PM (1987) An evaluation of the molecular clock hypothesis using mammalian DNA sequences. *Journal of Molecular Evolution*, **25**, 330–342.

Liu TT (1954) Hybridization between *Peromyscus maniculatus oreas* and *P. m. gracilis*. *Journal of Mammalogy*, **35**, 448–449.

Lundstrom R, Tavaré S, Ward RH (1992) Modelling evolution of the human mitochondrial genome. *Mathematical Biosciences*, **122**, 319–335.

Matocq DM (2002) Phylogeographical structure and regional history of the dusky-footed woodrat, *Neotoma fuscipes*. *Molecular Ecology*, **11**, 229–242.

Miura GI, Edwards SV (2001) Cryptic differentiation and geographic variation in genetic diversity of Hall's babbler *Pomatostomus halli*. *Journal of Avian Biology*, **32**, 102–110.

Osgood WH (1909) Revision of the mice of the American genus *Peromyscus*. *North American Fauna*, **28**, 1–285.

Palopoli M, Wu CI (1994) Genetics of hybrid make sterility between *Drosophila* sibling species: a complex web of epistasis is revealed in interspecific studies. *Genetics*, **138**, 329–341.

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

Polanski A, Kimmel M, Chakraborty R (1998) Application of a time-dependent coalescence process for inferring the history of population size changes from DNA sequence data. *Proceedings of the National Academy of Sciences USA*, **95**, 5456–5461.

Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics*, **14**, 817–818.

Rogers AR (1995) Genetic evidence for a Pleistocene population explosion. *Evolution*, **49**, 608–615.

Rogers AR, Harpending H (1992) Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution*, **9**, 552–569.

Salzburger W, Baric S, Sturmbauer C (2002) Speciation via introgressive hybridization in East Africa cichlids? *Molecular Ecology*, **11**, 619–625.

Schluter D (2001) Ecology and the origin of species. *Trends in Ecology and Evolution*, **16**, 372–380.

Schneider S, Excoffier L (1999) Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: application to human mitochondrial DNA. *Genetics*, **152**, 1079–1089.

Schneider S, Roessli D, Excoffier L (2000) ARLEQUIN, *Version 2.000: a Software for Population Genetic Data Analysis*. Genetics and Biometry Laboratory, Department of Anthropology and Ecology, University of Geneva, Switzerland.

Sheppe W (1961) Systematic and ecological relations of *Peromyscus oreas* and *P. maniculatus*. *Proceedings of the American Philosophical Society*, **105**, 421–446.

Slatkin M, Hudson RR (1991) Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics*, **129**, 555–562.

Soltis DE, Gitzendanner MA, Strenge DD, Soltis PS (1997) Chloroplast DNA intraspecific phylogeography of plants from the Pacific Northwest of North America. *Plant Systematics and Evolution*, **206**, 353–373.

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

Taberlet P (1996) The use of mitochondrial DNA control region sequencing in conservation biology. In: *Molecular Genetic Approaches in Conservation* (eds Smith TB, Wayne RK), pp. 125–142. Oxford University Press, New York.

Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution*, **10**, 512–526.

Turelli M, Barton NH, Coyne JA (2001) Theory and speciation. *Trends in Ecology and Evolution*, **16**, 330–343.

Verts BJ, Carraway LN (1998) *Land Mammals of Oregon*. University of Carlifornia Press, Berkeley.

Via S (2002) The ecological genetics of speciation. *American Naturalist*, **159**, S1–S7.

Wright S (1978) *Evolution and the Genetics of Populations*, Vol. 4. *Variability Within and Among Natural Populations*. The University of Chicago Press, Chicago.

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

This project is part of Xiaoguang Zheng's PhD dissertation in Zoology at the University of Washington and was conducted at the Burke Museum. His research involves genetic aspects of speciation and historical biogeography, and his work is based on mammalian populations in the Pacific Northwest of North America and in Southwestern China. Brian Arbogast worked on phylogeography of North American boreal forest mammals for his PhD at Wake Forest University and expanded on those interests in postdoctoral investigations at the Burke Museum; he recently joined the faculty at Humboldt State University. Jim Kenagy is Curator of Mammals at the Burke Museum, where he has developed a collections-based research programme emphasizing the historical biogeography and evolution of mammals, with particular attention to the Pacific Northwest of North America.
