MOLECULAR ECOLOGY

Molecular Ecology (2009) **18**, 5115–5125 doi: 10.1111/j.1365-294X.2009.04399.x

Nuclear and mitochondrial DNA reveal contrasting evolutionary processes in populations of deer mice (*Peromyscus maniculatus*)

DOU-SHUAN YANG and G. J. KENAGY

Burke Museum and Department of Biology, University of Washington, Seattle, WA 98195, USA

Abstract

We investigated a major geographic break in mitochondrial DNA (mtDNA) haplotypes in deer mice, *Peromyscus maniculatus*, by analysing spatial variation in a 491-bp fragment of the mtDNA control region from 455 samples distributed across a north–south transect of 2000 km in Western North America. To determine whether the mtDNA break was reflected in the nuclear genome, we then compared spatial variation in 13 nuclear microsatellites of 95 individuals surrounding the mtDNA break. Using a canonical correlation analysis we found that nuclear genomic variation was not correlated with mtDNA differentiation. The contrasting patterns of variation in mtDNA and nuclear DNA are consistent with a hypothesis of historic genetic drift that occurred in isolated refugia combined with recent gene flow between the formerly isolated refugial populations. A Mantel test of genetic vs. geographic distance revealed that recent gene flow between deer mouse populations has been high. We conclude that past vicariant events associated with Pleistocene climate changes together with recent gene flow have created the observed intra-specific cytonuclear discordance in Western North America.

Keywords: cytonuclear discordance, environmental history, mtDNA, *Peromyscus maniculatus*, phylogeography, population genetics

Received 19 June 2009; revision received 22 August 2009; accepted 27 August 2009

Introduction

Geographic variation in the genetic diversity of a species reflects the influence of both historic and recent evolutionary processes. Cytoplasmic DNA, such as mitochondrial DNA (mtDNA), and nuclear DNA can differ in their relative responses to historic and recent evolutionary processes due to their different modes of inheritance. The nonrecombinant inheritance of cytological DNA allows shared mtDNA haplotypes to be more easily maintained across species that are recently reproductively isolated. Consequently the mitochondrial phylogeny of recently diverged species pairs may retain shared evolutionary history prior to reproductive isolation, whereas this shared evolutionary history is potentially obscured in the nuclear genome by recombination (Roca *et al.* 2005; Egger *et al.* 2007). Therefore, the presence of discordance between cytoplasmic and nuclear DNA patterns is not surprising in cases of recent speciation (Roca *et al.* 2005; Di Candia & Routman 2007; Egger *et al.* 2007; Leaché & Cole 2007).

In this study, we investigated a previously identified mtDNA genetic break in deer mice, *Peromyscus maniculatus* (Avise *et al.* 1979a; Lansman *et al.* 1983; Dragoo *et al.* 2006) using both mtDNA sequences and nuclear genomic markers. *Peromyscus maniculatus* is continuously distributed throughout most of North America, where it is the most common and widespread terrestrial small mammal (Hall 1981). North America has a Pleistocene history of repeated geoclimatic cycles of glaciation and associated habitat shifts (Pielou 1992; Delcourt & Delcourt 1993). The biota of the Pacific Northwest shows a variety of different kinds of mtDNA discontinuities within this region (reviewed in Soltis *et al.* 1997; Brunsfeld *et al.* 2001; Carstens & Richards 2007). However, the previously studied vertebrate species all have discontinuous geographic ranges in the area of their

Correspondence: Dou-Shuan Yang, Fax: 01-206-616-1274; E-mail: dsyang@u.washington.edu

mtDNA breaks, and thus the breaks may have been caused by genetic drift resulting from recent reproductive isolation. Because of the continuous distribution of *P. maniculatus* and thus expected high gene flow between populations, the mtDNA genetic break in this species would seem unlikely to have originated from reproductive isolation and genetic drift between populations. Moreover, a previous continental-scale study of allozyme data (Avise *et al.* 1979b) found no nuclear differentiation in Western North American *P. maniculatus*, suggesting that a pattern of cytonuclear discordance may exist.

Unlike previous studies of cytonuclear discordance (Roca *et al.* 2005; Di Candia & Routman 2007; Egger *et al.* 2007; Leaché & Cole 2007), the hypothesized pattern of cytonuclear discordance that we propose occurs at the population rather than species level and thus could not be caused by the retention of mtDNA ancestral polymorphism, despite the emergence of reproductive isolation. In contrast, at the population level and thus without reproductive isolation, cytonuclear discordance reflects the maintenance of mtDNA haplotypes in the face of gene flow. One explanation for this may be the uniparental maternal inheritance of mtDNA, which results in an effective population size of mtDNA loci four times smaller than the effective population size of nuclear DNA loci. This smaller effective population size can greatly magnify the effects of genetic drift on mtDNA relative to nuclear DNA, especially in organisms with male-biased dispersal like *P. maniculatus* (King 1968). Thus, population-level patterns of cytonuclear discordance provide an opportunity to observe the contrasting effects of genetic drift and gene flow on populations, via their differing relative impact on mtDNA and nuclear DNA markers.

The goals of this study were to use spatial analysis of mtDNA variation to precisely identify the range of the mtDNA break at a fine scale and to analyse nuclear microsatellites of individuals within the range of the mtDNA break to determine whether the nuclear data are congruent with or conflict with the mtDNA break. We interpret these results in the context of Pleistocene climate history and propose how a pattern of cytonuclear discordance could arise within a single continuously distributed species.

Materials and methods

Sampling

We sampled a total of 455 *P. maniculatus* specimens from 39 counties ranging from British Columbia in the north to California in the south (Fig. 1, Table 1). All British Columbia samples are from Vancouver Island.

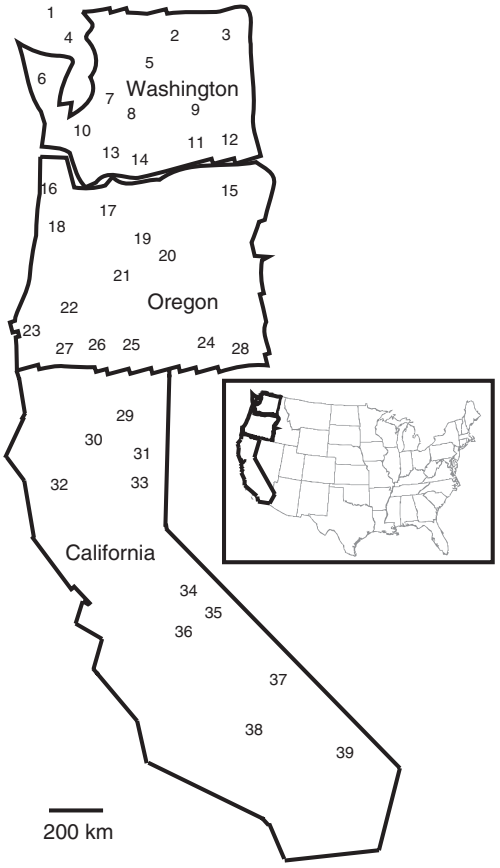

Fig. 1 Map of the 39 localities sampled for this study, with numbers as listed in Table 1. The inset map of the United States highlights the three Pacific States in the range of this study. Locality 1 is in British Columbia, Canada.

All samples are from voucher specimens at the University of Alaska-Fairbanks Museum of the North, the Humboldt State University Natural History Museum, the University of Washington Burke Museum, or the University of California-Berkeley Museum of Vertebrate Zoology.

Genomic DNA extraction and mtDNA sequencing

We extracted whole genomic DNA from all 455 individuals using the standard protocol provided with the Qiagen DNEasy kit. We amplified 491 bp of the mtDNA control region using PCR (Mullis & Faloona 1987) and the primers TDKD (Kocher *et al.* 1993) and THR (Kocher *et al.* 1989) for all 455 individuals. Following amplification we cleaned up the PCR product using the standard protocols of either Qiagen Qiaquick Spin Tubes or EXOSAP-IT (USB Corporation) procedures. Next we performed cycle sequencing in the forward direction using THR as the primer. Then we cleaned up the sequencing reaction products using a standard

Table 1 List of individual specimens sampled for this study, organized by province/state and county, and arranged from north to south by numbered locality

Locality number	Province or State	County	Number of individuals	Museum numbers
1	British Columbia		31	UWBM 75482, 7584–75492, 75379–75384, 75386, 75388–75390, 75392, 75394, 75395, 75397, 75399 75400, UAM 52689–52693
2	Washington	Okanogan	5	UWBM 73611, 73615, 73640, 74052, 74054
3	Washington	Pend Oreille	3	UWBM 73945, 73949, 73956
4	Washington	San Juan	55	UWBM 79937–79971, 79911–79917, 79921–79925, 79931–79954
5	Washington	Chelan	6	UWBM 73774, 73775, 73776, 73790, 73791, 73793
6	Washington	Clallam	1	UWBM 74957
7	Washington	King	3	UWBM 75846–75848
8	Washington	Kittitas	3	UWBM 73804, 73822, 73824
9	Washington	Grant	4	UWBM 72501, 72504, 72505, 72507
10	Washington	Thurston	1	UWBM 74334
11	Washington	Benton	4	UWBM 75791, 75795, 75798, 75805
12	Washington	Columbia	3	UWBM 73914, 73915, 73918
13	Washington	Skamania	1	UWBM 73840
14	Washington	Klickitat	2	UWBM 73835,73836
15	Oregon	Umatilla	3	UWBM 75403, 75405, 75407
16	Oregon	Tillamook	5	UWBM 75432, 75433, 75436, 75440, 75447
17	Oregon	Clackamas	3	UWBM 75408, 75409, 75415
18	Oregon	Lincoln	7	UWBM 79481–79487
19	Oregon	Jefferson	4	UWBM 75420, 75425, 75426, 75429
20	Oregon	Crook	10	UWBM 80087–80090, 80092, 80093, 80095–80098
21	Oregon	Deschutes	11	UWBM 79746–79750, 74723, 74724, 74726, 74728, 74729
22	Oregon	Douglas	7	UWBM 76572, 76577, 76578, 76581, 76583, 76585, 76586,
23	Oregon	Curry	16	UWBM 79973–79988
24	Oregon	Harney	31	UWBM 78504, 78507–78511, 78531, 78532, 78541, 78554, 78555, 80024–80043
25	Oregon	Klamath	13	UWBM 79751–79755, 80009–80012, 80013–80016
26	Oregon	Jackson	20	UWBM 79718–79725, 79756–79760, 80017–80019, 80020–80023,
27	Oregon	Josephine	19	UWBM 79989–80008
28	Oregon	Malheur	44	UWBM 80044–80086, 80099
29	California	Shasta	2	MVZ 200538, 200649
30	California	Tehama	8	MVZ 200598, 200637, 200752, 200754, 200756, 200758, 200760,
31	California	Plumas	3	MVZ 149734, 163033, 163034
32	California	Mendocino	1	HSU 7783
33	California	Nevada	5	MVZ 198293–198295, 199149, 199150
34	California	Tuolumne	20	MVZ 202217, 202224, 202242, 202243, 202246, 202250, 202252, 202258, 202264, 202267, 202273, 202274, 202301, 202313, 202314, 202320, 202321, 202330, 202336
35	California	Mono	9	MVZ 208446–208448, 208461–208463, 208468–208470
36	California	Mariposa	57	MVZ 202087, 202098–202104, 202106, 202109, 202121, 202124, 202128, 202134, 202156, 202163, 202173, 202178, 202179, 202181, 202184, 202185, 202191, 202192, 202193, 202195, 202197, 2202198, 202202, 202207–202210, 202215, 202216, 202340, 202341, 207992–207999, 208003, 208004, 208094–208098, 208109, 209104–209107
37	California	Inyo	2	MVZ 202573, 202574
38	California	Kern	6	MVZ 158882, 158887, 158893, 158895, 158899, 199879
39	California	San Bernadino	4	MVZ 198757–198760

UAM, University of Alaska-Fairbanks Museum; HSU, Humboldt State University Natural History Museum; UWBM, University of Washington Burke Museum; MVZ, Museum of Vertebrate Zoology.

ethanol precipitation protocol. All samples underwent capillary electrophoresis on either ABI 3100 or 3730xl instruments. Bases were scored and aligned manually using Sequencher v4.0 (Gene Codes) software. Three gaps were present within the data set and treated as missing data. There were 106 variable sites, of which 81

were parsimony-informative. All new mtDNA sequences have been deposited with GenBank (Accessions GQ860510–GQ860941).

mtDNA phylogenetic inference

To detect phylogenetic structure and regional population differentiation, we performed maximum likelihood and Bayesian phylogenetic analyses for the entire data set. For both analyses, we first estimated the most suitable model of evolution for the data set and its parameters (GTR+*I*+Γ; Lanave *et al.* 1984) by executing the Mr.ModelTest v2.2 (Nylander 2004) command block in PAUP* v4.0b10 (Swofford 2002). We performed the maximum likelihood analysis with 100 bootstrap pseudoreplicates using the program PHYML v2.2 (Guindon & Gascuel 2003). For this analysis, we applied the model of evolution specified by Mr.ModelTest but allowed PHYML to estimate the parameter values independently. For a Bayesian phylogenetic analysis we used the program MrBayes v3.1.2 (Ronquist & Huelsenbeck 2003) and applied the model and parameters estimated by Mr.ModelTest directly. For this analysis we ran four independent chains with default heating parameters for 2 million generations with a 100 000-generation burn-in period. The analysis sampled log likelihood values every 1000 generations and these values were used by MrBayes to construct a phylogenetic tree with posterior probabilities for nodal support.

Microsatellite analysis

A subset of 95 individuals from nine sampling sites within the mtDNA divergence zone was chosen for additional analysis using 13 nuclear microsatellite loci. We amplified genomic DNA for these 95 individuals using microsatellite primers from Mullen *et al.* (2006). Primer identities, allelic richness, and observed and expected heterozygosities are given in Table 2. We used M13R-sequence tagged fluorescent labels within our PCR reactions to label our amplification products (Mullen *et al.* 2006). Fluorescently labelled amplification products were multiplexed before visualization alongside a ROX400HD ladder on an ABI 3100 capillary electrophoresis instrument. A positive control was included as an allelic ladder for each electrophoresis run. Individual peaks were scored manually using ABI's GeneMapper software and all data were imported into GENALEX 6 (Peakall & Smouse 2006).

We used GENALEX to create a pairwise squared Euclidean genetic distance matrix and imported this matrix into R (R Development Core Team 2008). We then converted this matrix into a Euclidean distance matrix and performed a Principal Co-Ordinate Analysis of variation (PCoA) on the matrix with $k = 13$ eigenvalues (corresponding to the 13 loci genotyped) using the Vegan package (Oksanen *et al.* 2008) for R. Subsequently, we tested the axes of the PCoA for significantly explanatory components of variation using the broken stick distribution (Frontier 1976).

Next we performed a canonical correlation analysis (CCA) of the microsatellite data, constraining variation in the matrix by a vector of mitochondrial haplogroup identity, also using Vegan. This analysis calculates the proportion of variance in the data matrix explained by the constraining variable. Thus, the CCA calculates the proportion of nuclear genetic variance explained by mitochondrial haplogroup identity. We tested

Table 2 List of microsatellite loci used for this study, all of which deviated significantly from Hardy–Weinberg equilibrium except for BW 4-5

Locus	GenBank accession	Allele size range	Alleles	Expected heteryzygosity	Observed heterozygosity
BW 2-110	AF526099	90–190	23	0.922	0.484
BW 4-28	AF526144	257–335	22	0.928	0.742
BW 3-29	AF526123	99–171	24	0.953	0.524
BW 4-129	AF526131	215–383	49	0.978	0.382
BW 4-1	AF526127	137–187	21	0.9	0.686
BW 4-5	AF526155	143–167	7	0.579	0.46
BW 4-54	AF526156	146–248	38	0.967	0.35
BW 4-7	AF526159	288–357	29	0.933	0.522
BW 4-249	AF526147	214–346	41	0.972	0.525
BW 4-276	AF526150	159–258	34	0.953	0.608
BW 4-45	AF526154	218–390	42	0.971	0.35
BW 4-13	AF526132	147–246	27	0.951	0.372
BW 4-74	AF526160	315–388	15	0.653	0.471
Across all loci				0.897	0.498

significance of this proportion of variance using ANOVA, as implemented in **R**.

We also used the Bayesian-inference based program STRUCTURE 2.2.1 to detect clusters of nuclear genotype variation within the microsatellite data (Pritchard *et al.* 2000). To avoid biasing the Structure analysis, we did not use the mtDNA data to assign individuals to *a priori* populations according to mtDNA haplotype. Using the population admixture model with default settings, we performed three runs each for $K = 1$ to $K = 10$, where K is the user-defined number of clusters to test for in the data. Each run consisted of 1 000 000 Markov Chain-Monte Carlo repetitions with a 100 000 repetition burn-in period. We recorded the ln Prob(Data) for each run and averaged the ln Prob(Data) across runs for each value of K. To evaluate the most likely number of clusters within the data for each analysis we used both the mean estimated ln Prob(Data) for each K as well as the ΔK metric suggested by Evanno *et al.* (2005).

Tests to detect historic and current biogeographic processes

To test the hypothesis that populations became isolated in historic refugia and experienced genetic drift and subsequently expanded out of these refugia, we inferred the recent history of each mtDNA haplogroup within the mtDNA divergence zone. Contemporary populations derived from Pleistocene refugial populations should display demographic histories of recent population expansion, corresponding to post-glacial population expansion out of refugia. To statistically test for recent population expansion within mice of each haplogroup, we calculated Tajima's D (Tajima 1989) and Fu's F_S (Fu 1997) statistics for the mtDNA of mice from each haplogroup, each with 10 000 simulated samples to test for significance of the calculated statistic. As a parallel qualitative test for recent population expansion, we performed a mismatch distribution test (Rogers & Harpending 1992) and tested the sum of squared deviation from values derived from a null model of recent exponential population expansion for significance using 1000 bootstrap permutations of the data.

To test for the presence of recent gene flow between populations we used a Mantel test (Mantel 1967; Smouse *et al.* 1986) to calculate the correlation between genetic and geographic distance for the 95 contact-zone individuals. If gene flow does not exist between populations, we expect the correlation between genetic and geographic distance to be strongly positive, whereas if gene flow does occur, then we expect the correlation to be near zero. We calculated confidence intervals for the correlation coefficient and tested the coefficient for significance using random permutation of the data, as implemented in GENALEX.

Results

Variation in mtDNA

Our mtDNA gene tree revealed two major clades of mice in Pacific coastal western North America, distin-

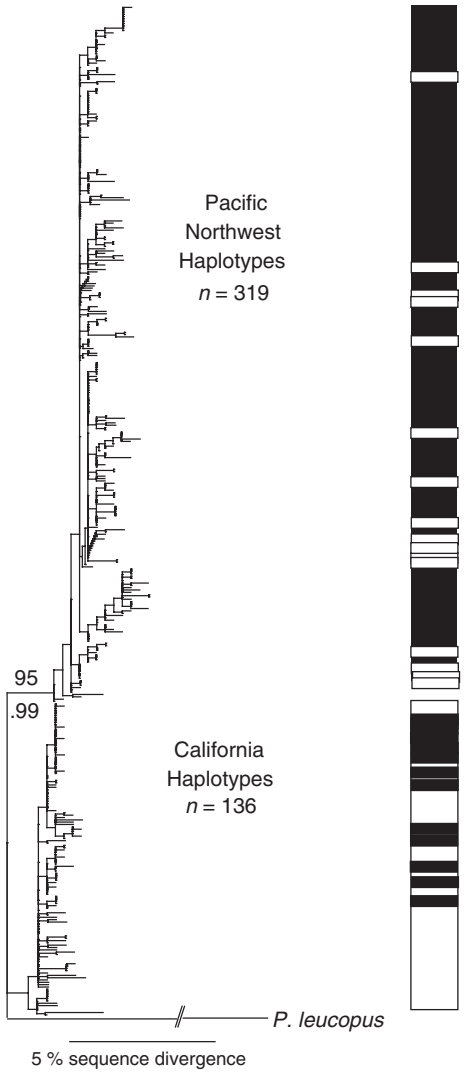

Fig. 2 Maximum likelihood (ML) tree of mtDNA control region data for 455 *P. maniculatus* sampled in the present study. Support for the strong separation of Pacific Northwest and Californian haplotypes is given above the branch for ML bootstrap replicates and below for Bayesian posterior probability. Black in the vertical bar indicates samples from Pacific Northwest localities and white indicates samples from Californian localities. The Bayesian analysis returned an overall identical topology to that of ML. The tree is rooted with a sequence from *Peromyscus leucopus* (Genbank Accession AY540423.1).

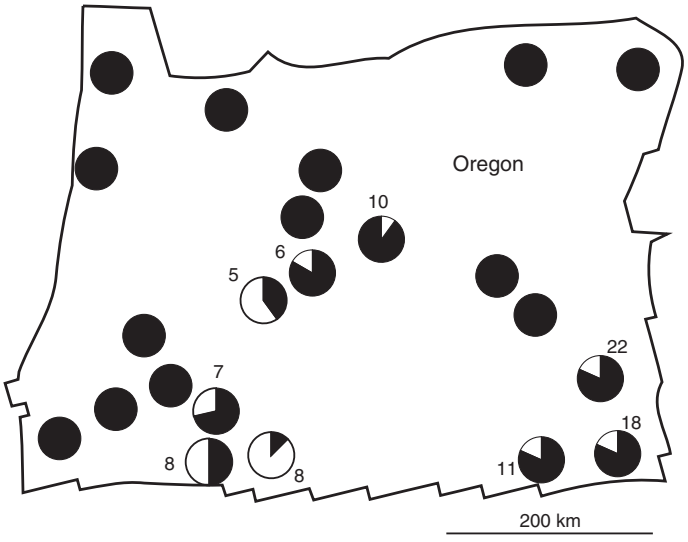

Fig. 3 Pie-chart map showing relative frequency of Pacific Northwest (black) and Californian (white) mtDNA haplotypes at each Oregon locality. Numbers indicate total sample size at the nine sites where both haplotype groups are present. These nine localities represent the mtDNA contact zone.

guished by 12 conserved substitutions (2.7% sequence divergence, Fig. 2). The transitional overlap of these two clades occurs geographically across a mtDNA contact zone in southern and eastern Oregon (Fig. 3; Table S1, Supporting information). Additional transitional overlaps of these clades also occur in northern coastal and eastern California, although we will focus primarily on the Oregon contact zone. No significant regional geographic structure in genetic variation is apparent within the two major haplotype groups (Fig. 2).

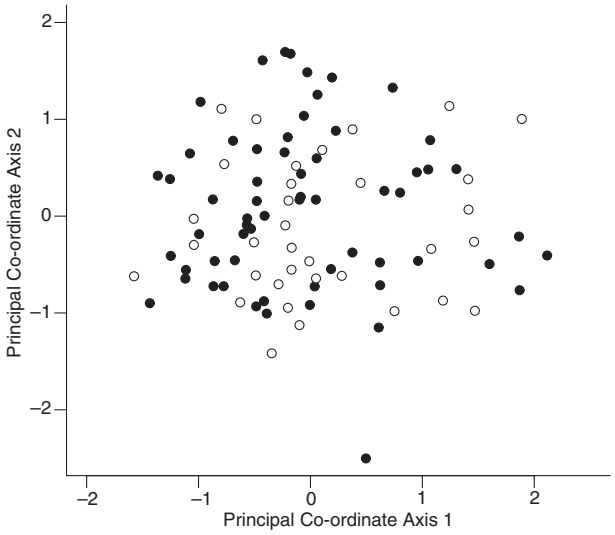

Fig. 4 Principal Co-Ordinate Analysis of nuclear microsatellite genotype variation of individuals within the mtDNA contact zone. Filled circles indicate individuals with Pacific Northwest haplotypes and open circles indicate individuals with Californian haplotypes. Axis 1 contains 11.8% of total variation and Axis 2 contains 10.4% of total variation.

Variation in nuclear microsatellites

The Principal Co-Ordinate Analysis (PCoA) revealed that the nuclear genotypes of mice from Pacific Northwest and Californian haplotype groups within the mtDNA contact zone in southern Oregon formed a single uniform cluster (Fig. 4), in contrast to the mtDNA pattern of a large genetic discontinuity (Fig. 2). None of the first five ordination axes was significantly highly explanatory, although the remaining eight ordination axes were all significantly explanatory (Table 3). Nevertheless, the first two PCoA axes explained the greatest amount of variation, and thus we chose to visualize the microsatellite data in this ordination space. Altogether 5.2% of the genotype data was missing, and the amount of missing data per locus ranged from 0.0% to 14.7%. The allelic richness was very high, with a mean of 34.92 alleles per locus and a range of 2–59

Table 3 Significance of components of variation for canonical correlation analysis, showing observed λ on the top row and expected λ from null distribution on the bottom

	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8	λ_9	λ_{10}	λ_{11}	λ_{12}	λ_{13}
Observed %	11.8	10.4	9.5	8.9	8.4	7.4	6.9	6.9	6.7	6.1	5.9	5.6	5.5
Expected %	24.4	16.8	12.9	10.3	8.4	6.9	5.6	4.5	3.6	2.7	1.9	1.2	0.5

alleles per locus (Table S2, Supporting information). For mice in the Oregon contact zone, all loci except BW 4-5 significantly deviated from Hardy–Weinberg equilibrium (Table 2).

The STRUCTURE analysis without *a priori* designation of populations found that $K = 10$ was the most likely number of clusters of nuclear variation in the data set (mean ln Prob(Data) = −6646.03; Table S3, Supporting information) and that $K = 1$ was the least likely number of clusters (mean ln Prob(Data) = −7567.2, Table S3). However, using the ΔK metric suggested by Evanno *et al.* (2005), which emphasizes the second-order rate of change of ln Prob(Data) as K increases, $K = 5$ was most strongly supported with a Δ ln Prob(Data) of −367 from $K = 4$ (Table S3).

Comparison of mtDNA and nuclear microsatellite variation

Our overall analysis of the microsatellite loci indicates that nuclear genomic variation is not spatially concordant with the major geographic break in southern Oregon. The canonical correlation analysis found that only 1.4% of total nuclear microsatellite variation was explained by mitochondrial haplogroup identity. The 1.4% explained was significant by permutation test (ANOVA: $\chi^2 = 0.0002$, $F = 1.312$, 1 df, 699 permutations, $P = 2.9\%$), although the significant result for this negligible value is likely caused by low sample size. Thus, a significant discordance exists between our observed nuclear and cytogenetic signals, which amount to 'cytonuclear discordance.'

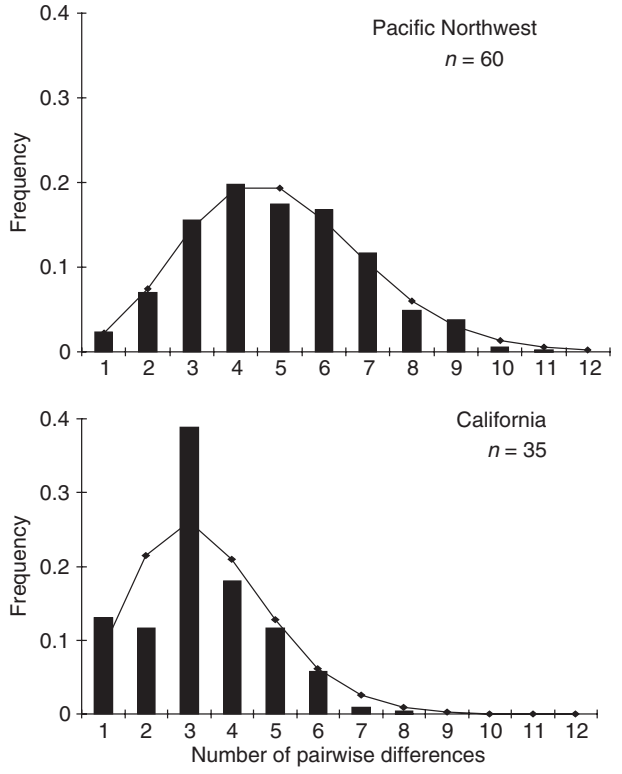

Fig. 5 Mismatch distribution plots for mtDNA data from within the Pacific Northwest and Californian haplogroups for the mtDNA contact zone individuals. The strongly unimodal mismatch distribution reflects historic exponential demographic growth of the Pacific Northwest haplogroup, whereas the complex, multimodal distribution of the Californian group does not support such demographic growth.

Tests to detect Pleistocene and recent biogeographic processes

Pacific Northwest haplotypes within the southern Oregon contact zone possessed a strongly supported signature of recent demographic expansion, whereas Californian haplotypes within the contact zone did not have a similar signature of recent demographic expansion. Contact zone individuals possessing Pacific Northwest haplotypes had significantly negative values of both Tajima's D (−1.78, $P = 0.01$) and Fu's F_S (−25.95, $P < 0.001$). In contrast, contact zone individuals possessing Californian haplotypes had a nonsignificant value of Tajima's D (−0.23, $P = 0.46$) and a marginally significant value of Fu's F_S (−4.25, $P = 0.02$) at the $P = 0.02$ critical α-level. The contrast between the demographic history of Pacific Northwest and Californian haplotypes is also supported by the mismatch distribution (Fig. 5). These analyses demonstrate recent demographic expansion for individuals with Pacific Northwest haplotypes (Σ of Squared Deviation = 0.001, $P = 0.48$; Harpending's Raggedness Index = 0.02, $P = 0.56$) but did not support recent demographic expansion for individuals with Californian haplotypes (Σ of Squared Deviation = 0.029, $P = 0.01$; Harpending's Raggedness Index = 0.128, $P = 0.01$).

The comparison of genetic and geographic distance matrices by Mantel test indicated a lack of isolation by distance in the nuclear genomes of contact-zone mice, suggesting recent gene flow between contact-zone mice. The test found a significant, but minimal relationship between nuclear genetic and geographic distance ($y = 0.0004x + 5.7784$; $r^2 = 0.0363$; $P < 0.001$).

Discussion

Geographic structure in mtDNA

The well supported mtDNA break we observed in western North American *P. maniculatus* is congruent with the mtDNA break detected by Lansman *et al.* (1983) and Dragoo *et al.* (2006). Individuals from Pacific Northwest localities typically belong to one haplotype

group, while individuals from California localities belonged to the other haplotype group. These two haplotype groups are sympatric in southern Oregon, where they co-occurred at nine of our sampling localities in six counties across Oregon (Fig. 1, counties 20, 21, 24 25, 26, and 28; Table S1).

As mentioned above, additional transitional overlaps of mtDNA clades also occurred in northern California, with 16 out of 136 Californian individuals possessing Pacific Northwest haplotypes. These 16 individuals were collected along the Sierra Nevada (Fig. 1, localities 30, 34, 35 and 36; Table S1) and the Pacific coast (locality 32) in Northern California. Lansman *et al.* (1983) previously identified a role of the Sierra Nevada mountain range in partitioning mtDNA diversity between a 'Central States Assemblage' and a 'Californian Assemblage.' Our fine-scale geographic sampling has allowed us to identify the geographic extent of the genetic break in southern Oregon. Unlike the eastern edge of the mtDNA contact zone, no geographic barrier, such as a mountain range, is apparent at the northern and western bounds of the contact zone.

We interpret the presence of a genetic break in the absence of a geographic barrier as the apparent result of historic biogeographic events. Previous research has revealed similar north–south mtDNA divergences in other Pacific Northwest taxa in the absence of structuring geographic barriers (Soltis *et al.* 1997; Miller *et al.* 2006). Regional concordances in phylogeographic patterns across species have been attributed to a shared history of Pleistocene habitat shifts and subsequent vicariance (Hewitt 1996).

Lack of differentiation of nuclear genotypes in the mtDNA contact zone

If the observed mtDNA break reflects reproductive isolation between members of each haplotype group, then the mtDNA pattern should also be reflected geographically in the nuclear genome of contact-zone individuals. However, the nuclear genotypes of contact-zone individuals did not exhibit such a pattern, instead forming a single large cluster in canonical ordination space independent of mtDNA haplotype group, indicating that the nuclear genomes of all individuals are not differentiated according to mtDNA haplotype.

The five clusters supported by ΔK analysis of STRUCTURE results were relatively undifferentiated (mean pairwise F_{ST} between clusters = 0.054) and thus could represent stochastic variation present within the single cluster identified by Principal Co-ordinate Analysis. Unfortunately, we cannot evaluate the hypothesis of a single cluster that was supported by the PCoA ($K = 1$) using the ΔK metric (Evanno *et al.* 2005). Nevertheless, as in the PCoA, differentiation between the five clusters does not correspond to mtDNA haplogroup membership (Fig. 6).

This lack of nuclear differentiation in contrast to mtDNA differentiation could be attributed to a lack of variation at the microsatellite loci. However, we observed high allelic richness in the loci (Table 2). Consequently, we attribute the general deviation from Hardy–Weinberg equilibrium in microsatellite loci to insufficient sample size relative to the high number of alleles.

The lack of correspondence between nuclear DNA variation and mtDNA variation within the mtDNA contact zone represents a case of cytonuclear discordance. The canonical correlation analysis reveals that mtDNA haplotype group identity has almost no influence on nuclear DNA variation. mtDNA data can be a useful tool to detect reproductive isolation between groups when it is combined with other data, e.g. morphological or nuclear genomic data, to more fully reflect evolutionary history (Wilcox *et al.* 1997; Arbogast 1999; Demboski & Cook 2001; Zheng *et al.* 2003; Hebert *et al.* 2004; Haine *et al.* 2006). In this study, interpreting the mtDNA data alone could have led to a conclusion that reproductive isolation existed in a cryptic species complex within *P. maniculatus*. Our results underscore the

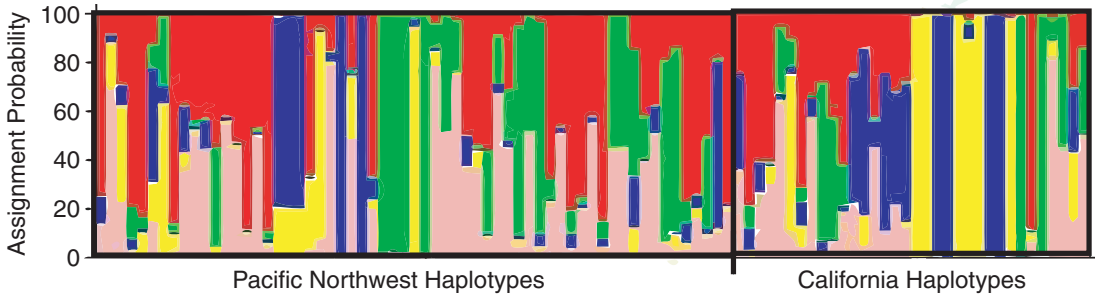

Fig. 6 Cluster assignment test from STRUCTURE based on $K = 5$ clusters. Each color represents an inferred cluster and each individual is represented by a vertical line colored according to its probability of assignment to each cluster. Individuals with Pacific Northwest mtDNA haplotypes belonged to the same clusters as individuals with Californian mtDNA haplotypes.

value of using both cytoplasmic and nuclear DNA markers when attempting to determine the evolutionary history of populations and potential cryptic species relationships.

Biogeographic processes underlying cytonuclear discordance

Our conclusion that the pattern of cytonuclear discordance of *P. maniculatus* in southern Oregon reflects contrasting Pleistocene and recent evolutionary processes is similar to the conclusions of others who have observed cytonuclear discordance in wild populations (Roca *et al.* 2005; Di Candia & Routman 2007; Egger *et al.* 2007; Leaché & Cole 2007). We hypothesize that the mitochondrial divergence was caused by the sundering of a historically continuous distribution of *P. maniculatus* in western North American by Pleistocene glaciation events and the retreat of *P. maniculatus* into northwestern coastal and Californian refugia (Brunsfeld *et al.* 2001; Carstens & Richards 2007). The restriction of *P. maniculatus* into these glacial refugia should have prevented gene flow, and the distinguishing mtDNA substitutions would thus have become fixed within each refugial population by genetic drift. Using the rate of 3.8×10^{-8} mtDNA control region substitutions/year estimated for wild *Mus* species (Goios *et al.* 2007), we can multiply by 250 000 years of late Pleistocene glacial cycles and the 491 bp sequenced for a total of 4.66 expected substitutions per refugial population during this time period. Because each refugial population underwent drift separately, the expected number of substitutions is doubled, for a total of 9.32 expected substitutions between populations. Thus, our observation of 12 fixed substitutions that distinguish the two mtDNA clades is consistent with the hypothesis of mtDNA differentiation within Pleistocene refugia.

The current southern Oregonian mtDNA contact zone is not located at the boundaries of historic refugia, because the populations most likely have shifted due to postglacial population expansion. Under the above hypothesis of mtDNA differentiation in refugia, we postulate that contemporary Pacific Northwest haplotypes originated to the north in a Northwest coastal refugium, while contemporary Californian haplotypes originated in the California refugium. Thus, the presence of the northern haplotypes in the southern Oregon contact zone would indicate that the Pacific Northwest haplotypes have undergone recent spatial and demographic expansion southwards, as suggested by Zheng *et al.* (2003). This postulate is supported by our statistical analyses of recent demographic patterns, which indicate that Pacific Northwest haplotypes in the contact zone have undergone recent demographic expansion.

The generalist life history and ubiquity of contemporary *P. maniculatus* in boreal forest environments allow us to assume that historic *P. maniculatus* were also successful in this environment. Thus, mice in the Californian refugium were likely able to range northward to the maximal glacial leading edge in southern Washington. Therefore, it is puzzling that the contemporary contact zone between Pacific Northwest and Californian haplotypes lies about 320 km to the south of the maximal Pleistocene glacial extent in southern Oregon. Under the hypothesis of differentiation in refugia, it appears that Californian haplotypes have been displaced in the southern portion of their late Pleistocene distribution following the last glacial recession. This postulate is supported by our statistical analyses of demographic patterns, in that Californian haplotypes in the contact zone have not undergone recent demographic expansion. Therefore, the data are consistent overall with a hypothesis of generation of strong mtDNA structure through isolation in interglacial refugia.

In contrast to the mtDNA patterns, no such structure exists in the nuclear DNA of contact-zone individuals. What evolutionary processes could generate this pattern of cytonuclear discordance? One potential explanation is the retention of ancestral polymorphism at the nuclear microsatellite loci throughout the late Pleistocene isolation during which mtDNA differentiation took place. The uniparental inheritance of mtDNA results in an effective population size four times smaller than the effective population size of bi-parentally inherited nuclear DNA, and thus nuclear DNA is more likely to retain ancestral polymorphism. We consider this explanation unlikely, however, because individuals of each haplotype group were present within the same sampling areas at sympatric localities, which indicate that opportunities for interbreeding between haplogroups are readily available. Still, to reject this hypothesis would require dense multilocus nuclear sequence data and a coalescent-based comparison of the relative likelihoods of recent gene flow vs. retention of ancestral polymorphism.

As an alternative, we suggest that strong recent gene flow between mice of the two mtDNA haplotype groups may have generated the observed nuclear homogeneity. *Peromyscus maniculatus* have male-biased dispersal (King 1968), and thus we expect the rate of nuclear gene flow to be higher than the rate of maternally inherited mtDNA gene flow. Our Mantel test suggests that recent gene flow is high and that it was strong enough to mediate the effects of recent genetic drift, thereby homogenizing the nuclear genomes of mice from both haplotype groups. The hypothesis of strong regional gene flow in southern Oregon is also supported by the lack of

geographic structure in mtDNA variation outside the contact zone. Therefore, the observed cytonuclear discordance has likely resulted from the retention of historic genetic drift in mtDNA and the homogenization of nuclear DNA by recent gene flow.

Acknowledgements

This research was supported by the Mammalogy Program at the Burke Museum, by the Theodore Roosevelt Memorial Fund of the American Museum of Natural History, a Grant-in-Aid of the American Society of Mammalogists, and by a fellowship of the Genome Training Grant (University of Washington). We thank the University of California-Berkeley Museum of Vertebrate Zoology, the Humboldt State University Natural History Museum, and the University of Alaska – Fairbanks Museum of the North for access to research specimens. We thank Toby Bradshaw, Joe Felsenstein, Richard Olmstead, Matthew Cronin, and an anonymous reviewer for comments on the research and manuscript.

References

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

Avise JC, Lansman RA, Shade RO (1979a) The use of restriction endonucleases to measure mitochondrial DNA sequence relatedness in natural populations. I. Population structure and evolution in the genus *Peromyscus*. *Genetics*, **92**, 279–295.

Avise JC, Smith MH, Selander RK (1979b) Biochemical polymorphism and systematics in the genus *Peromyscus* VII. Geographic differentiation in members of the Truei and Maniculatus species groups. *Journal of Mammalogy*, **60**, 177–192.

Brunsfeld S, Sullivan J, Soltis DE, Soltis PS (2001) *Comparative Phylogeography of Northwestern North America: a Synthesis*. Blackwell Publishing, Willston, Vermont.

Carstens BC, Richards CL (2007) Integrating coalescent and ecological niche modeling in comparative phylogeography. *Evolution*, **61**, 1439–1454.

Delcourt P, Delcourt HR (1993) Paleoclimates, paleovegetation and paleofloras during the late Quarternary. In: *Flora of North America* (ed. committee FoNAe), pp. 71–94. Oxford University Press, New York.

Demboski JR, Cook JA (2001) Phylogeography of the dusky shrew, Sorex monticolus (Insectivora, Soricidae): insight into deep and shallow history in northwestern North America. *Molecular Ecology*, **10**, 1227–1240.

Di Candia MR, Routman EJ (2007) Cytonuclear discordance across a leopard frog contact zone. *Molecular Phylogenetics and Evolution*, **45**, 564–575.

Dragoo JW, Lackey JA, Moore KE *et al*. (2006) Phylogeography of the deer mouse (*Peromyscus maniculatus*) provides a predictive framework for research on hantaviruses. *Journal of General Virology*, **87**, 1997–2003.

Egger B, Koblmuller S, Sturmbauer C, Sefc K (2007) Nuclear and mitochondrial data reveal different evolutionary processes in the Lake Tanganyika cichlid genus Tropheus. *BMC Evolutionary Biology*, **7**, 137.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

Frontier S (1976) Étude de la decroissance des valeurs propres dans une analyze en composantes principales: comparison avec le modèle de baton brisé. *Journal of Experimental Marine Biology and Ecology*, **25**, 67–75.

Fu YX (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, **147**, 915–925.

Goios A, Pereira Ls, Bogue M, Macaulay V, Amorim AN (2007) MtDNA phylogeny and evolution of laboratory mouse strains. *Genome Research*, **17**, 293–298.

Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*, **52**, 696–704.

Haine E, Martin J, Cook J (2006) Deep mtDNA divergences indicate cryptic species in a fig-pollinating wasp. *BMC Evolutionary Biology*, **6**, 83.

Hall ER (1981) *The Mammals of North America*. John Wiley And Sons, New York.

Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W (2004) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly Astraptes fulgerator. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 14812–14817.

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

King JA (1968) *Biology of Peromyscus (Rodentia)*, 1st edn. The American Society of Mammalogists, Stillwater, Oklahoma.

Kocher TD, Thomas WK, Meyer A *et al*. (1989) Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. *Proceedings of the National Academy of Sciences of the United States of America*, **86**, 6196–6200.

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.

Lanave C, Preparata G, Sacone C, Serio G (1984) A new method for calculating evolutionary substitution rates. *Journal of Molecular Evolution*, **20**, 86–93.

Lansman RA, Avise JC, Aquadro CF, Shapira JF, Daniel SW (1983) Extensive genetic variation in mitochondrial DNA'S among geographic populations of the deer mouse, *Peromyscus maniculatus*. *Evolution*, **37**, 1–16.

Leaché AD, Cole CJ (2007) Hybridization between multiple fence lizard lineages in an ecotone: locally discordant variation in mitochondrial DNA, chromosomes, and morphology. *Molecular Ecology*, **16**, 1035–1054.

Mantel N (1967) The detection of disease clustering and a generalized regression approach. *Cancer Research*, **27**, 209–220.

Miller MP, Bellinger MR, Forsman ED, Haig SM (2006) Effects of historical climate change, habitat connectivity, and

vicariance on genetic structure and diversity across the range of the red tree vole (*Phenacomys longicaudus*) in the Pacific Northwestern United States. *Molecular Ecology*, **15**, 145–159.

Mullen LM, Hirschmann RJ, Prince KL *et al*. (2006) Sixty polymorphic microsatellite markers for the oldfield mouse developed in *Peromyscus polionotus* and *Peromyscus maniculatus*. *Molecular Ecology Notes*, **6**, 36–40.

Mullis K, Faloona F (1987) Specific synthesis of DNA *in vitro* via a polymerase-catalyzed chain reaction. *Methods in Enzymology*, **155**, 335–350.

Nylander J (2004) *MRMODELTEST 2.2*. Evolutionary Biology Centre, Uppsala University, Uppsala, Sweden.

Oksanen J, Kindt R, Legendre P *et al*. (2008) *Vegan: Community Ecology Package. R package version 1.15-1*.

Peakall R, Smouse PE (2006) genalex6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, **6**, 288–295.

Pielou E (1992) *After the Ice Age: The Return of Life to Glaciated North America*. University of Chicago Press, Chicago.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

R Development Core Team (2008) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna.

Roca AL, Georgiadis N, O'Brien SJ (2005) Cytonuclear genomic dissociation in African elephant species. *Nature genetics*, **37**, 96–100.

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

Ronquist F, Huelsenbeck JP (2003) MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19**, 1572–1574.

Smouse PE, Long JC, Sokal RR (1986) Multiple regression and correlation extensions of the Mantel test of matrix correspondence. *Systematic Zoology*, **35**, 627–632.

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

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

Wilcox TP, Hugg L, Zeh JA, Zeh DW (1997) Mitochondrial DNA sequencing reveals extreme genetic differentiation in a cryptic species complex of Neotropical pseudoscorpions. *Molecular Phylogenetics and Evolution*, **7**, 208–216.

Zheng X, Arbogast BS, Kenagy GJ (2003) Historical demography and genetic structure of sister species: deermice (*Peromyscus*) in the North American temperate rain forest. *Molecular Ecology*, **12**, 711–724.

Dou-Shuan Yang is a PhD student in Biology and the Burke Museum, and this work represents a chapter of his dissertation. He is interested in the evolutionary biogeography of small mammals. Jim Kenagy is Curator of Mammals at the Burke Museum, where he works with graduate students interested in collections-based research on the historical biogeography and evolution of contemporary mammal populations.

Supporting information

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

Table S1 Table of localities surveyed including the number of mice from each haplotype group found within the county

Table S2 Frequencies of individual alleles of each microsatellite locus within the six counties surveyed for microsatellite variation

Table S3 Results of Structure analyses

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