Molecular Ecology (2008) **17**, 1313–1327 doi: 10.1111/j.1365-294X.2007.03640.x

Ancient hybridization and mitochondrial capture between two species of chipmunks

JEFFREY M. GOOD,* SARAH HIRD,† NOAH REID,† JOHN R. DEMBOSKI,‡ SCOTT J. STEPPAN,§ TINA R. MARTIN-NIMS§ and JACK SULLIVAN†

**Department of Ecology and Evolutionary Biology, University of Arizona, Biosciences West, Tucson, AZ 85721, USA, †Department of Biological Sciences, Box 443051, University of Idaho, Moscow, ID 83844-3051, USA, ‡Department of Zoology, Denver Museum of Nature & Science, 2001 Colorado Boulevard, Denver, CO 80205-5798, USA, §Department of Biological Sciences, Florida State University, Tallahassee, FL 32306-1100, USA*

Abstract

Models that posit speciation in the face of gene flow are replacing classical views that hybridization is rare between animal species. We use a multilocus approach to examine the history of hybridization and gene flow between two species of chipmunks (*Tamias ruficaudus* and *T. amoenus*). Previous studies have shown that these species occupy different ecological niches and have distinct genital bone morphologies, yet appear to be incompletely isolated reproductively in multiple areas of sympatry. We compared data from four sequenced nuclear loci and from seven microsatellite loci to published cytochrome *b* sequences. Interspecific gene flow was primarily restricted to introgression of the *T. ruficaudus* mitochondrial genome into a sympatric subspecies of *T. amoenus*, *T. a. canicaudus*, with the four sequenced nuclear loci showing little to no interspecific allele sharing. Microsatellite data were consistent with high levels of differentiation between the species and also showed no current gene flow between broadly sympatric populations of *T. a. canicaudus* and *T. ruficaudus*. Coalescent analyses date the mtDNA introgression event from the mid-Pleistocene to late Pliocene. Overall, these data indicate that introgression has had a minimal impact on the nuclear genomes of *T. amoenus* and *T. ruficaudus* despite multiple independent hybridization events. Our findings challenge long-standing assumptions on patterns of reproductive isolation in chipmunks and suggest that there may be other examples of hybridization among the 23 species of *Tamias* that occur in western North America.

Keywords: gene flow, hybridization, multilocus, reproductive isolation, speciation, *Tamias*

Received 14 July 2007; revision accepted 17 November 2007

Introduction

Several recent studies have shown that closely related animal species often share a history of introgressive hybridization (Machado *et al*. 2002; Grant *et al*. 2005; Melo-Ferreira *et al*. 2005; Turner *et al*. 2005; Berthier *et al*. 2006; Putnam *et al*. 2007). These data challenge the notion that recurrent gene flow between distinct animal species is rare (Dobzhansky 1951; Mayr 1963) and suggest that hybridization may be a frequent, potentially important, transient phase before the completion of reproductive isolation in animals. However, because introgressive hybridization is often cryptic, the frequency and genetic impact of hybridization during animal speciation remains an important and unresolved issue in evolutionary biology.

Gene flow between animal species often appears heterogeneous across the genome, leading to the concept of semipermeable species limits (Harrison 1990; Wu 2001). Under this model, genomic regions involved in reproductive isolation are expected to show little or no introgression, whereas regions that are unlinked to hybrid incompatibilities ought to be exchanged more easily between hybridizing species (Rieseberg *et al*. 1999). Thus, genetic data sampled from across the genome can provide important information on both the genetic architecture of reproductive isolation

Correspondence: Jack Sullivan, Fax: + 01 208 885 7905; E-mail: jacks@uidaho. edu

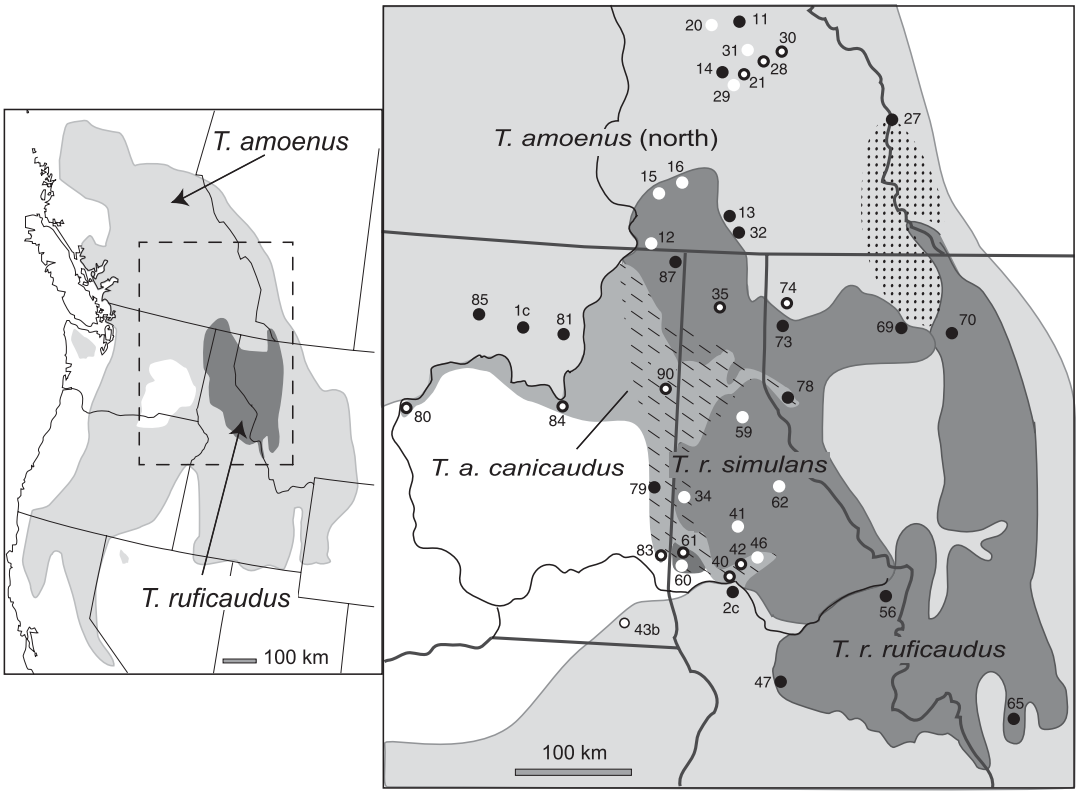

Fig. 1 Distribution of *Tamias ruficaudus* and *T. amoenus* (modified from Hall 1981). The region of sympatry between *T. a. canicaudus* and *T. r. simulans* is indicated with cross-hatching. The zone of current introgression between *T. r. ruficaudus* and *T. amoenus* is stippled and was inferred from incongruence between mtDNA variation and genital morphology (Good *et al*. 2003). Sampling localities are denoted by circles and are shaded based on the kind of genetic sampling represented (black, sequenced mtDNA and/or nDNA loci; white, microsatellites; black and white, sequenced mtDNA and/or nDNA loci and microsatellites). Locality numbers correspond to Good *et al*. (2003), and Table S1.

(Rieseberg *et al*. 1999; Payseur *et al*. 2004) and the history of gene flow among populations or species (Hilton & Hey 1997; Kliman *et al*. 2000; Llopart *et al*. 2005). Recent methodological advances have facilitated the use of multilocus data to reconstruct the evolutionary history of diverging populations (Wakeley & Hey 1997; Wang *et al*. 1997; Pritchard *et al*. 2000; Machado *et al*. 2002; Hey & Nielsen 2007; Putnam *et al*. 2007), which can provide a means for testing demographic models of speciation and quantifying biologically relevant parameters underlying species limits (Broughton & Harrison 2003; Geraldes *et al*. 2006).

Two species of chipmunk (genus *Tamias sensu lato*; but see Piaggio & Spicer 2001) from western North America provide a compelling case study for examining semipermeable species limits (Good & Sullivan 2001; Demboski & Sullivan 2003; Good *et al*. 2003). The yellow-pine chipmunk, *Tamias amoenus*, is widely distributed across xeric forests of western North America (Sutton 1992) whereas the red-tailed chipmunk, *Tamias ruficaudus*, occurs primarily in mesic forests of the northern Rocky Mountains (Fig. 1; Best 1993). These are nonsister species (Levenson *et al*. 1985; Piaggio & Spicer 2001; Demboski & Sullivan 2003) and are sympatric along ecological transitions between xeric and mesic habitats. Furthermore, they show pronounced differentiation in the morphology of the males' os penis or baculum (White 1953; Good *et al*. 2003). Indeed, interspecifc divergence in the size and shape of the baculum has been traditionally viewed as indicative of complete reproductive isolation among species of *Tamias* (Patterson & Thaeler 1982; Sutton 1995).

Nevertheless, at least two independent instances of mitochondrial DNA (mtDNA) introgression have been documented between *T. amoenus* and *T. ruficaudus* (Good *et al*. 2003). Recent, potentially ongoing, hybridization has resulted in the introgression of *T. ruficaudus* mtDNA haplotypes into populations of *T. amoenus* near the northeastern edge of the *T. ruficaudus* distribution (Fig. 1). A second, presumably older hybridization event has been hypothesized to explain the fixation of *T. ruficaudus* mtDNA in *T. a. canicaudus*; based on mtDNA, *T. a. canicaudus* is monophyletic, is nested within *T. ruficaudus*, and is sister to *T. r. simulans*, with which it is partially sympatric. In both cases of mtDNA introgression, *T. amoenus* populations show no evidence of divergence in ecology, external

morphology, or bacular morphology from nearby non-introgressed *T. amoenus* populations (Good *et al*. 2003). In principle, incomplete sorting of ancestral polymorphisms could generate the lack of reciprocal monophyly between these species (Funk & Omland 2003); however, mtDNA variation is well sorted across geography in both species and haplotype sharing is geographically restricted to regions of sympatry (Good & Sullivan 2001; Demboski & Sullivan 2003). This pattern of geography-dependent genetic similarity between species is inconsistent with incomplete lineage sorting and provides strong evidence of interspecific gene flow (Goodman *et al*. 1999; Coyne & Orr 2004).

Here, we use a multilocus approach to elucidate the history and genetic contribution of the introgression in the genomes of *T. amoenus* and *T. ruficaudus*. We focus on two questions. First, what is the extent and pattern of introgression across mitochondrial and nuclear loci in both taxa? Second, is reproductive isolation complete between historically hybridizing populations of *T. a. canicaudus* and *T. r. simulans*? To address these questions, we examined genetic variation at one mitochondrial and 11 nuclear loci. We contrasted patterns of nucleotide sequence variation in the mitochondrion with four protein-coding nuclear loci to characterize levels of intraspecific genetic variation and to test for gene flow since the initial divergence of these species. Finally, we use finer-scale sampling of variation at seven microsatellite loci to characterize contemporary population structure and gene flow between populations of *T. a. canicaudus* and *T. r. simulans*.

Materials and methods

Sampling strategy

Resolution of genetic structure within and between hybridizing species requires sampling that spans the geographical range of both species while sufficiently surveying populations in regions of sympatry. Previously, we examined mitochondrial cytochrome *b* (cyt *b*) variation (701 bp) for 219 *T. ruficaudus* and 197 *T. amoenus* samples from across the range of both species (Demboski & Sullivan 2003; Good *et al*. 2003). We use these published cyt *b* data for all population-level analyses (see below). To build on this work, we expanded our previous genetic and taxonomic sampling to include a total of 1117 bp of cyt *b* and 19 species of *Tamias*. These data were generated to place patterns of mtDNA variation within *T. ruficaudus* and *T. amoenus* in a phylogenetic context with variation observed across the genus. We also sequenced four nuclear loci from 11 *T. ruficaudus* (seven *T. r. simulans* and four *T. r. ruficaudus*) from distinct localities across the species range. For *T. amoenus*, we sampled 34 individuals from 22 locations including eight *T. a. canicaudus*. Of the remaining 26 *T. amoenus*, 19 were sampled from populations directly to the north of the range limit of *T. ruficaudus* and seven were selected from across the species range based on the major mtDNA phylogeographical partitions described by Demboski & Sullivan (2003). The northern *T. amoenus* populations comprise two subspecies (*T. a. affinis* and *T. a. luteiventris*); however, previous genetic analyses have demonstrated that these form a single mtDNA clade (Demboski & Sullivan 2003) and thus will be referred to as *T. amoenus* (north) throughout.

For surveys of microsatellite variation, we expanded our sampling within *T. r. simulans*, *T. a. canicaudus*, and *T. amoenus* (north). Samples of *T. r. simulans* and *T. a. canicaudus* were from populations that were either broadly sympatric or allopatric with respect to each other (Fig. 1). Our sampling of *T. r. simulans* included 20 individuals from five sympatric localities and 24 individuals from six allopatric localities. For *T. a. canicaudus*, we sampled 15 individuals from four sympatric localities and 20 individuals from two allopatric localities. Twenty-six *T. amoenus* (north) individuals were collected in locations that were geographically distant from both *T. r. simulans* and *T. a. canicaudus* as well as from the current introgression zone between *T. r. ruficaudus* and *T. amoenus* (north) (Good *et al*. 2003). Our geographical sampling strategy is depicted in Fig. 1 and Table S1, Supplementary material. All samples were either collected in the field, or obtained through loans of tissues or study skins from various sources (Demboski & Sullivan 2003; Good *et al*. 2003).

Extraction, amplification, sequencing, and fragment analysis

Genomic DNA was extracted from tissues using the PCI/CI 'hot' method (Sambrook & Russell 2001), a modified dodecyl trimethyl ammonium bromide/cetyltrimethyl ammonium bromide (DTAB/CTAB) protocol (Gustincich *et al*. 1991), or using the QIAGEN DNeasy Tissue Kit (QIAGEN). We polymerase chain reaction (PCR) amplified 12 loci: one mitochondrial (cyt *b*), four nuclear protein-coding genes (acrosin, acid phosphatase 5, proto-oncogene c-myc, and recombination activating gene 1), and seven nuclear microsatellite loci (EuAmMS26, EuAmMS35, EuAmMS37, EuAmMS86, EuAmMS108, EuAmMS114 and EuAmMS142; Schulte-Hostedde *et al*. 2000).

For cyt *b*, an 1117-bp fragment of the 5′ end was amplified and sequenced using primers and methodology described in Good *et al*. (2003) and Demboski & Sullivan (2003). This data set also included previously published sequences (Piaggio & Spicer 2001) and four unpublished sequences provided by J. Patton. We also sequenced a portion of acrosin (acr), an acrosomal binding protein present in mammalian sperm and involved in fertilization. Primers for acr were designed by aligning orthologous sequences for *Mus musculus*, *Rattus norvegicus*, *Homo sapiens* and *Spermophilus*

tridecemlineatus retrieved from GenBank and placing primers in conserved regions of exons. The following primer pairs amplified approximately 1600 bp of contiguous DNA, covering the first three exons and two introns: ACR11F (TATGGTRGAGATGCTGCCAACTG) and ACR 15R (GCAGTGAGCAGCTGTAAGCACC), which covers ~1140 bp, from exon 1-exon 2; and ACR13F (CTCACTGGGTGCTTACAGCTGC), ACR14R (CCAGCCAGCCACATAGCAGG) which covers ~400 bp from exon 2-exon 3. Reactions were performed using a touchdown protocol beginning at an annealing temperature of 58 °C cycling twice at each 2° interval (56 °C and 54 °C) down to a final annealing temperature of 53 °C. For acid phosphatase 5 (acp5), a 390-bp region spanning intron 1, including sections of exons 1 and 2, was amplified. For the proto-oncogene, c-myc, we amplified a region spanning the 3′ end of intron 2 (243 bp), all of the translated exon 3 (563 bp), and the 5′ end of the untranslated region (UTR) of exon 3 (150 bp; 956 bp total). For the recombination activating gene 1 (RAG1), sequence data were generated entirely from the divergent 5′ region of the single protein-coding exon (808 bp). Primer sequences and amplification conditions for acp5, c-myc, and RAG1 were published previously (DeBry 1998; Steppan *et al*. 2004a, b). Amplicons were purified via precipitation with polyethylene glycol precipitation or QIAGEN QIAquick spin columns (QIAGEN). Products were sequenced directly for both strands on an ABI PRISM 3100 or 377 automated sequencer using BigDye Terminator chemistry (ABI).

We also amplified seven microsatellite loci using the primers developed by Schulte-Hostedde *et al*. (2000) for *T. amoenus*. PCR conditions involved an initial denaturing step of 94 °C for 3 min, 32 cycles of 45 s at 94 °C, 45 s at the appropriate annealing temperature (see Schulte-Hostedde *et al*. 2000), and 45 s at 72 °C. Amplicons were run on an ABI 3730 DNA sequencer and trace files were analysed with GENEMAPPER version 3.7 software (Applied Biosystems).

Chromatograms were first edited and initially aligned using SEQUENCHER 3.0–4.5 (Gene Codes Corp.). Nuclear sequence alignments were modified using MACCLADE 4.0 (Maddison & Maddison 2003). Levels of heterozygosity were low for all four of the nuclear gene loci and most haplotypes could be resolved unambiguously from diploid PCR sequence. When necessary, we resolved gametic phase of sequences computationally using the computer software PHASE 2.1.1 (Stephens *et al*. 2001). Newly generated DNA sequences are available in GenBank (accessions EU250579–EU250744).

Nucleotide variation

For each sequenced locus, we calculated basic estimates of nucleotide variability within and between *T. amoenus* and *T. ruficaudus* using DNASP 4.10 (Rozas *et al*. 2003). Nucleotide variability was measured as the average pairwise number of nucleotide differences, π (Nei & Li 1979), and the proportion of segregating sites, θ (Watterson 1975). We calculated the normalized difference between π and θ or Tajima's *D* (Tajima 1989) to test for departures in the frequency spectrum from equilibrium conditions.

Genealogical estimation

Genealogical relationships were estimated using two approaches. First, we used a phylogenetic approach based on maximum-likelihood (ML) estimation. Models of sequence evolution were selected using DT-MODSEL (Minin *et al*. 2003) and ML searches were conducted with PAUP* 4.0 (Swofford 2002) using successive approximations (Sullivan *et al*. 2005). Nodal support was estimated using the nonparametric ML bootstraps (Felsenstein 1985) in PAUP*. Second, haplotype networks were estimated using median joining (Bandelt *et al*. 1999) as implemented in NETWORK 4.1.1.2 (Fluxus Technology, Ltd) with insertion/deletion variation treated as a fifth state.

Polymorphism, divergence, and gene flow

We examined patterns of polymorphism to divergence using three related models based on the neutral equilibrium prediction that the ratio of polymorphism to divergence depends on the underlying mutation rate and should be equivalent across site classes and/or loci. First, we used the McDonald–Kreitman test to detect significant departures in the ratio of polymorphism to divergence for synonymous and nonsynonymous sites for protein-coding regions (McDonald & Kreitman 1991). Second, we used the Hudson–Kreitman–Aguade test to contrast patterns of polymorphism to divergence across sequenced loci (i.e. HKA test; Hudson *et al*. 1987). The HKA test was developed to detect heterogeneity among loci due to natural selection; however, heterogeneous gene flow across loci between two species also yields deviations from the neutral model. Thus, the HKA framework also provides a test of an isolation model of speciation (Wang *et al*. 1997). We compared polymorphism to divergence between *T. amoenus* and *T. ruficaudus* as well as divergence to a closely related species, *Tamias senex* (chosen as a geographically distant reference taxon for which some data were available). All five loci were examined using a single HKA test for each of the three possible pairwise species comparisons (*T. amoenus/T. ruficaudus*, *T. amoenus/T. senex*, *T. ruficaudus/T. senex*). In addition, we performed all pairwise comparisons among loci within each group. Estimates of average pairwise divergence for cyt *b* were corrected using the best fit HKY + I model, and significance was evaluated with 10 000 coalescent simulations (Hey & Kliman 1993). Third, we evaluated the fit of all five loci to the general isolation model developed

by Wakeley & Hey (1997) using two statistics as implemented with the WH program. The Wang, Wakeley and Hey (WWH) statistic (Wang *et al*. 1997) was calculated from the sum of the maximum difference across loci between lowest and highest counts of shared polymorphisms and fixed differences, whereas the chi-squared test (Kliman *et al*. 2000) was based on a sum of the discrepancies between observed and expected values for each class of polymorphic site (shared, polymorphic within a single species, or fixed). Statistical significance of the WWH and the chi-squared test statistics were evaluated using 10 000 coalescent simulations.

The two-population isolation model of species divergence (e.g. Wakeley & Hey 1997) has been extended to include gene flow (Nielsen & Wakeley 2001; Hey & Nielsen 2007). We analysed our data using two different implementations of this framework. First, we used the program IMA (Hey & Nielsen 2007) to derive multilocus coalescent estimates of six parameters based on the five-locus sequence data set: three population mutation parameters ($\theta_{amoenus}$, $\theta_{ruficaudus}$, and $\theta_{ancestral}$, where $\theta = 4N_e\mu$), the scaled population splitting time t, and two directional population migration rates (m_1 and m_2). Here, μ is the mutation rate per locus per generation, N_e is the effective population size, and m is the number of migrants per generation. Nuclear loci were analysed under the infinite sites model while the cyt *b* locus was fit to the Hasegawa–Kishino–Yano (HKY) model of sequence evolution. The four-gametic haplotype test (Hudson & Kaplan 1985) was used to identify and remove all recombinant sequences. Second, we used the program MDIV (Nielsen & Wakeley 2001) to fit a simplified version of the model ($\theta_{amoenus} = \theta_{ruficaudus} = \theta_{ancestral}$, $m_1 = m_2$) to derive a coalescent estimate of mtDNA divergence between *T. r. simulans* and *T. a. canicaudus* to approximate the historical timing of introgression.

We used multiple approaches to evaluate contemporary levels of gene flow within and between *T. amoenus* and *T. ruficaudus* based on microsatellite data. Estimates of heterozygosity were calculated using GENEPOP 3.4 (Raymond & Rousset 1995). Exact tests of Hardy–Weinberg equilibrium (HWE) were performed for each locus in each of three groups: *T. a. canicaudus*, *T. r simulans*, and *T. amoenus* (north). To estimate population differentiation, we calculated F_{ST} under an infinite allele model as implemented with GENEPOP and R_{ST} under a stepwise-mutation model as implemented with FSTAT 2.9.3 (Goudet 1995). To examine differentiation at microsatellite loci without imposing a priori population structure on the data, we used a model-based clustering method as implemented in the program STRUCTURE 2.1 (Pritchard *et al*. 2000). We ran five replicates of each of 11 possible population structures, with k ranging from 2 to 12 where k is the number of subdivisions. For each replicate, we ran chains for 10^6 generations and discarded the first 10^5 generations as burn-in.

STRUCTURE does not account for deviations from HWE or provide an explicit test of recent hybridization. We used two approaches to address these issues. First, we analysed the microsatellite data using BAYESASS 1.2 (Wilson & Rannala 2003) and incorporated the inbreeding coefficient as a parameter in estimation of migration among populations. Second, we used NEWHYBRIDS (Anderson & Thompson 2002) to assign explicit probabilities of belonging to one of six hybrid classes (pure population 1, pure population 2, F_1, F_2, backcrossed to population 1 and backcrossed to population 2). We performed two pairwise runs, comparing *T. a. canicaudus* to either *T. r. simulans* or *T. amoenus* (north). The program ran for 2×10^4 iterations per run, the first half of which were discarded as burn-in.

Results

Patterns of nucleotide variation

Sample sizes and patterns of nucleotide variation within each species are summarized in Table 1. For the previously published cyt *b* population data (701 bp), we report nucleotide variability in both the entire sample and in a subsample (51 *Tamias amoenus*, 45 *T. ruficaudus*) restricted to the same geographical sampling regime used for nuclear loci. All multilocus comparisons were conducted using the mtDNA subsample. For the four nuclear loci, we obtained between 54 (acr) and 80 sequences (acp5) for *T. ruficaudus* and *T. amoenus*. Differences in sample sizes were primarily due to the removal of individuals that demonstrated PCR and/or sequencing failure for a given locus.

Levels of variability were over an order of magnitude lower at nuclear loci than was observed for the mitochondrial cyt *b* locus (*T. ruficaudus*, average $\pi = 0.14\%$ vs. 2.02%; *T. amoenus*, average $\pi = 0.31\%$ vs. 4.95%). Variability was also lower in *T. ruficaudus* than in *T. amoenus*, except for c-myc. Within each species, there was considerable variation across nuclear loci. The most remarkable difference was observed in *T. amoenus,* where RAG1 was over an order of magnitude more variable than acr (silent and noncoding sites, $\pi = 2.23\%$ vs. 0.14%). We also observed considerable variation in the frequency spectrum across loci as estimated by Tajima's D. Only cyt *b* in *T. ruficaudus* and RAG1 in *T. amoenus* had a significantly positive skews. A positive Tajima's D was not unexpected given the deep mtDNA phylogeny partitions observed within both species (Good & Sullivan 2001; Demboski & Sullivan 2003).

Genealogical relationships

The phylogeny of *Tamias* from western North America is poorly resolved (Levenson *et al*. 1985; Piaggio & Spicer 2001; Demboski & Sullivan 2003), yet overall phylogenetic patterns are important for evaluating the potential for

Table 1 Nucleotide variation within *Tamias amoenus* and *T. ruficaudus*

					π (percentage)		θ (percentage)			
Locus/species	*N*	Total sites†	Hap.‡	*S*§	All	Silent¶	All	Silent§	Tajima's *D*¶	R_m^{**}
acr										
T. amoenus	38	1558 (1150)	6	7	0.140	0.140	0.107	0.104	0.897	0
T. ruficaudus	16	1557 (1149)	4	4	0.099	0.134	0.077	0.105	0.873	0
acp5										
T. amoenus	62	361 (205)	9	8	0.265	0.435	0.472	0.727	–1.021	1
T. ruficaudus	18	361 (205)	3	2	0.183	0.143	0.161	0.142	0.219	0
c-myc										
T. amoenus	48	899 (341)	11	5	0.098	0.251	0.126	0.270	–0.164	0
T. ruficaudus	14	898 (340)	6	4	0.174	0.157	0.140	0.187	–0.438	0
RAG1										
T. amoenus	52	764 (0)	12	18	0.742	2.226	0.521	1.074	2.577*	2
T. ruficaudus	20	764 (0)	4	4	0.107	—	0.148	—	—	0
cyt *b*										
T. amoenus	197	701 (0)	67	138	4.954	14.011	3.361	8.787	1.779	—
subsample	51	701 (0)	24	114	4.909	14.008	3.615	10.142	1.307	—
T. ruficaudus	219	701 (0)	46	69	2.017	6.974	1.650	4.721	1.333	—
subsample	45	701 (0)	10	41	2.275	7.026	1.338	4.219	2.145*	—

*$P < 0.05$ based on 10 000 coalescent simulations; †total number of nucleotide sites, the number of noncoding sites is in parentheses; ‡number of unambiguously resolved haplotypes, including insertion-deletion variation; §total number of polymorphic sites; ¶based on noncoding regions (acr, acp5, c-myc) or fourfold degenerate synonymous sites (RAG1, cyt *b*); **minimum number of recombination events (Hudson & Kaplan 1985).

incomplete lineage sorting within our study system. For mtDNA data (1117-bp fragment), we estimated a single ML phylogeny (Fig. 2) using the best-fit TrN + I + Γ model. The phylogeny indicates the distant, nonsister relationship between members of the *T. amoenus* clade and the clade containing *T. a. canicaudus* and both subspecies of *T. ruficaudus*, with *T. a. canicaudus* forming a strongly supported clade with *T. r. simulans* (see Good *et al*. 2003; Demboski & Sullivan 2003 for a presentation of variation within each of these clades).

For all four nuclear loci, most PCR products were unambiguously phased into haplotypes. Individuals with more than one polymorphic site were resolved computationally (Stephens *et al*. 2001). Haplotype pairs with a probability < 0.95 were excluded from all haplotype-based analyses. Because phylogenetic relationships among the four loci were poorly resolved, we emphasize a network-based approach (Fig. 3). *T. senex* is included in all networks for reference; however, its outgroup status is unresolved. Since the networks are unrooted, this should not be problematic. For three loci (acr, acp5, c-myc), haplotypes clustered into two groups corresponding to *T. ruficaudus* and *T. amoenus* separated by a minimum of one fixed difference. RAG1 was partitioned into three groups with a deep split within *T. amoenus* and one haplotype shared between multiple *T. ruficaudus* samples and one homozygous *T. a. canicaudus* individual. *Tamias amoenus* appeared paraphyletic at RAG1, even if this shared haplotype was not considered (analysis not shown). The high divergence within *T. amoenus* observed at RAG1 is in general agreement with a deep mtDNA split between *T. amoenus* (north) and other populations of *T. amoenus* (Demboski & Sullivan 2003) whereas the apparent paraphyly may reflect insufficient variation for resolution of deep internal nodes. Interestingly, *T. a. canicaudus* fell out with *T. amoenus* (north) (excluding the single individual grouped within *T. ruficaudus*) consistent with a lack of bacular divergence (see Fig. 2; Good *et al*. 2003) between these geographically adjacent taxa.

Polymorphism, divergence, and gene flow

We observed considerable variation in the ratio of polymorphism to divergence across loci and across species (Table 2). The most extreme deviations involved contrasts of polymorphism within *T. amoenus* to divergence with *T. ruficaudus* or *T. senex*. In these comparisons, both cyt *b* and RAG1 showed high ratios of polymorphism to divergence relative to other loci. None of the sequenced loci showed significant departures in the ratio of polymorphism to divergence for synonymous and nonsynonymous sites based on the McDonald–Kreitman test, suggesting an excess of slightly deleterious nonsynonymous polymorphisms at cyt *b* and RAG1 was not contributing to the overall deviation

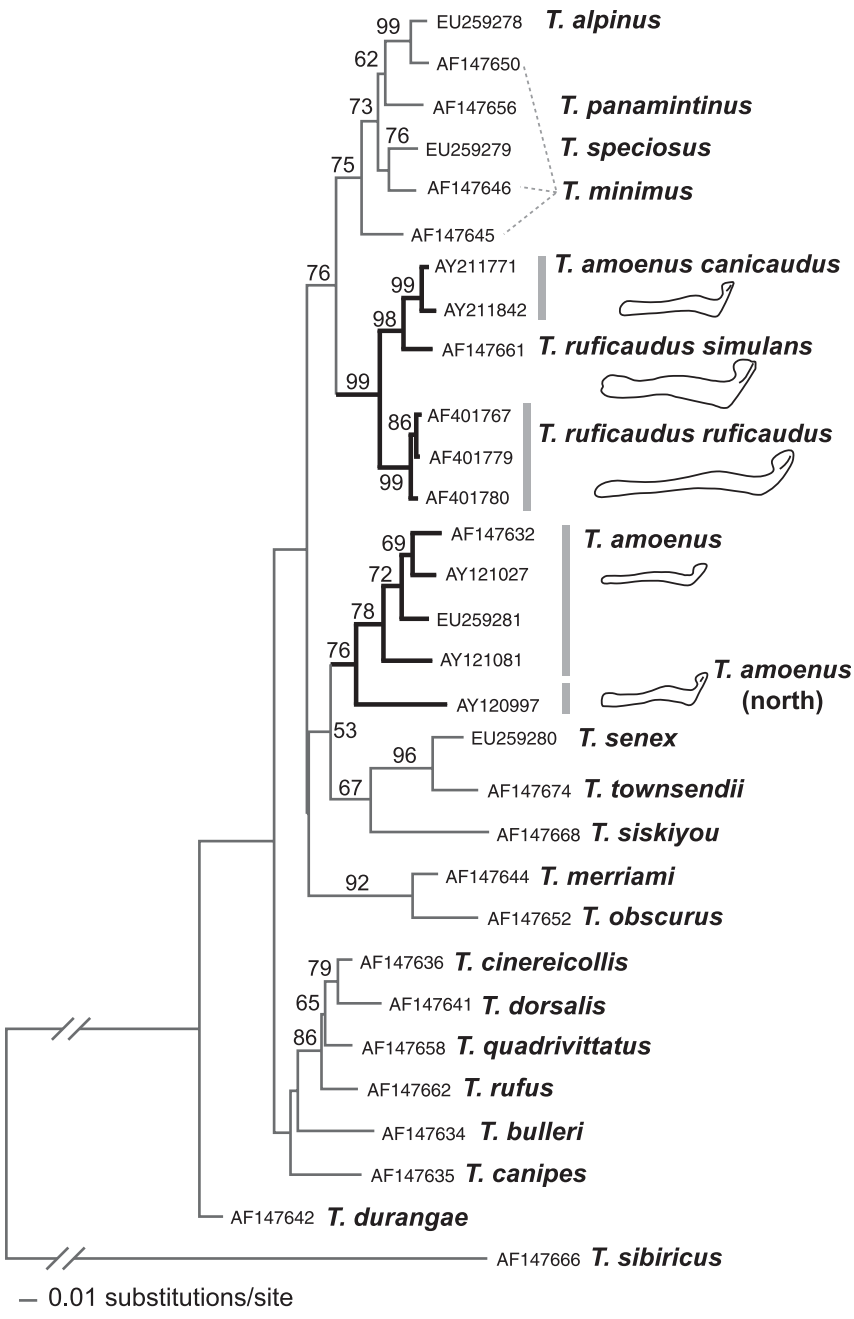

Fig. 2 Maximum-likelihood estimate (TRN + I + Γ) of phylogeny of cyt *b* for 19 species of *Tamias*. Numbers associated with internal branches are ML bootstraps. Taxa studied here have general outlines of their bacular type drawn to the same scale. Good *et al*. (2003) demonstrated low levels of intrataxon variation in bacular morphology for these taxa, consistent with many other studies of bacular variation in chipmunks.

across loci. Alternatively, these patterns are consistent with historical introgression of cyt *b* and possibly RAG1 *T. ruficaudus* alleles into *T. a. canicaudus*.

To evaluate this possibility further, we first tested for heterogeneity among loci in each of three species pairs (*T. amoenus*–*T. ruficaudus*, *T. amoenus*–*T. senex*, *T. ruficaudus* –*T. senex*) using the HKA test. Because of the somewhat low divergence at several loci between *T. amoenus* and *T. ruficaudus*, the inclusion of an additional lineage (*T. senex*) should improve the power of these tests. Multilocus HKA tests across all five loci failed to reject the null model in comparisons between *T. amoenus* and *T. ruficaudus* (P = 0.114) and *T. ruficaudus* and *T. senex* (P = 0.137) but significantly rejected homogeneity across loci when comparing *T. amoenus* and *T. senex* (P = 0.003; Bonferroni corrected α = 0.017). All pairwise HKA tests between *T. ruficaudus* and *T. senex* and all tests excluding cyt *b* were not significant ($P > 0.05$). In contrast, multiple pairwise tests involving cyt *b* indicated a significant excess of polymorphism to divergence in comparisons of *T. amoenus* either to *T. senex* (c-myc, P = 0.006; acr, P = 0.008; Bonferroni-corrected α = 0.01) or *T. ruficaudus* (acr, P = 0.009; Bonferroni-corrected α = 0.01).

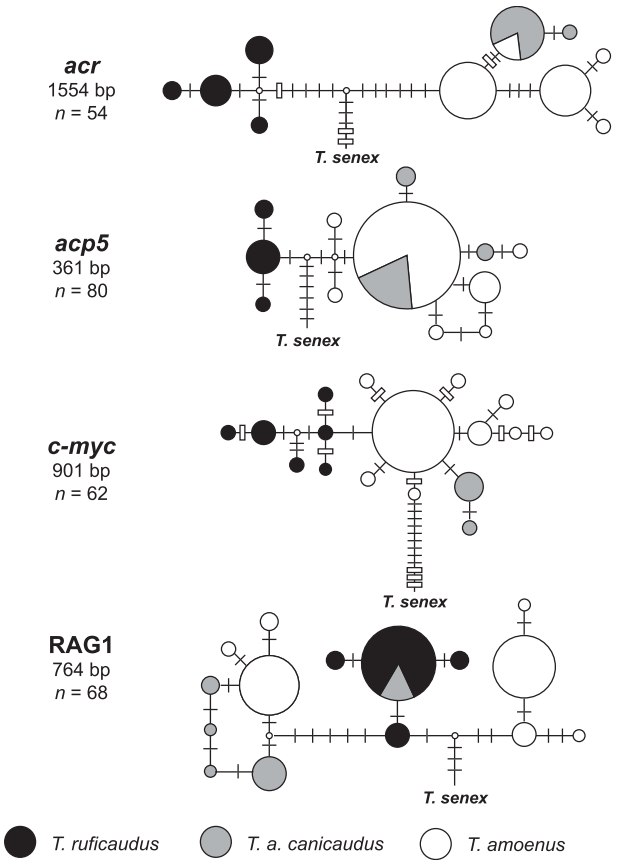

Fig. 3 Median-joining haplotype networks for four sequenced autosomal loci. Haplotypes are indicated as circles and are scaled by frequency. Haplotypes sampled in *T. ruficaudus* are shown in black, *T. a. canicaudus* are in grey, and all other *T. amoenus* samples (*T. amoenus* (north) and other *T. amoenus*) are in white. Substitutions are denoted with closed hashes and insertion/deletion variation is indicated with open hashes.

The HKA tests demonstrate that cyt *b* exhibits an excess of polymorphism to divergence in *T. amoenus* relative to nuclear loci. However, these tests fail to capture the most striking aspect of the data; the overwhelming ratio of shared to fixed differences (25:1) between *T. amoenus* and *T. ruficaudus* despite high levels of pairwise divergence (0.0732 substitutions per site; Table 2) and a relatively small effective population size (~$1/4N_e$ vs. nuclear loci). Therefore, we explicitly tested a model of isolation based on the distribution of fixed, shared, and exclusive polymorphisms between *T. amoenus* and *T. ruficaudus* (Wakeley & Hey 1997; Wang *et al*. 1997). Considering all five loci, we rejected the null model of divergence without gene flow between *T. amoenus* and *T. ruficaudus* based on the chi-squared test ($P = 0.020$) but not the WWH statistic ($P = 0.219$). For the chi-squared test, deviation from the null model was driven by a lower than expected number of fixed differences (13.9 expected, 1 observed) and a higher than expected number of shared polymorphisms (5.7 expected, 25 observed) for cyt *b*.

We used IMA to estimate six parameters ($\theta_{amoenus}$, $\theta_{ruficaudus}$, $\theta_{ancestral}$, m_1, m_2, and t) of the two-population isolation model of species divergence for our five-locus sequence data set. Across independent runs, we explored a broad range of priors, number of chains (1–10), and heating scenarios (linear, two-step, and geometric). Although we were able to achieve adequate mixing of the Markov chains and stable estimates for several of the parameters ($\theta_{amoenus}$, $\theta_{ruficaudus}$, m_1, m_2), we were unable to obtain stable estimates of the full posterior density and the scaled population splitting time t. Thus, we viewed all parameter estimates from our IMA analyses as unreliable. Given the large contrast in genealogical patterns across loci and low levels of variation in our data, it was not surprising that we had little power to estimate all of the parameters of this complex model.

We were, however, able to use the simpler model implemented in MDIV to derive a coalescent estimate of mtDNA divergence between *T. r. simulans* and *T. a. canicaudus* and approximate the historical timing of introgression. The posterior probability distribution of θ for cyt *b* was well resolved (Figure S1a, Supplementary material), with the point estimate of $\theta = 3.06$ (95% credibility interval of 2.05–4.54). There are no available estimates of mutation rate for cyt *b* in *Tamias*. Nevertheless, average mammalian nucleotide substitution rates in mtDNA protein-coding genes have been estimated as 1.8×10^{-9} and 2.7×10^{-8} substitutions/site/year for nonsynonymous and synonymous sites, respectively (Pesole *et al*. 1999). Thus, when weighted across sites, this provides a general estimate of 8.2×10^{-9}. Assuming this general rate per site, μ for the 701-bp segment of cyt *b* would be approximately 5.8×10^{-6} substitutions/locus, yielding a rough estimate of N_e = 265 735. The population migration rate M between *T. a. canicaudus* and *T. r. simulans* was effectively zero (Figure S1b, Supplementary material). The posterior probability of population divergence time was more diffuse, but, using N_e from above and a generation time of 1 year, the estimated divergence time is nevertheless rather old (3.0 million years ago, 95% credibility interval of 1.1–5.0; Figure S1c, Supplementary material).

To resolve further the genetic association of *T. a. canicaudus*, we examined microsatellite variation in expanded samples within *T. r. simulans*, *T. a. canicaudus*, and *T. amoenus* (north). All seven microsatellite loci were polymorphic both within and between these groups (Table 3) and observed heterozygosity ranged from 0.09 to 0.85. Within each of the three groups, 10 of 21 tests exhibited significant departure from HWE expectations ($P < 0.10$). These are likely to be attributable, in part, to population structure not accounted for by these tests. In 92 of the 98 tests of genotype frequencies for locations, we detected no significant departures from HWE expectations (for all exact tests $P > 0.10$).

Table 2 Polymorphism and divergence across five sequenced loci

Locus and species compared	Average differences*	*D*xy† (percentage)	Fixed	Polymorphic species 1‡	Polymorphic species 2	Shared	θ1/*D*xy§	θ2/*D*xy
acr								
amoenus-ruficaudus	12.734	0.818	10	7	4	1	0.131	0.094
amoenus-senex	11.579	0.745	9	7	—	—	0.144	
ruficaudus-senex	9.188	0.591	7	4	—	—	0.130	
acp5								
amoenus-ruficaudus	3.776	1.046	2	8	2	0	0.451	0.154
amoenus-senex	8.887	2.462	7	8	—	—	0.192	
ruficaudus-senex	7.889	2.185	7	2	—	—	0.074	
c-myc								
amoenus-ruficaudus	2.949	0.330	1	5	4	0	0.382	0.424
amoenus-senex	9.521	1.066	9	5	—	—	0.118	
ruficaudus-senex	11.429	1.277	10	4	—	—	0.110	
RAG1								
amoenus-ruficaudus	6.181	0.809	0	17	3	1	0.644	0.183
amoenus-senex	8.423	1.102	3	18	—	—	0.473	
ruficaudus-senex	5.050	0.661	4	4	—	—	0.224	
cyt *b*								
amoenus-ruficaudus	51.342	7.324	1	91	16	25	0.494	0.183
amoenus-senex	73.227	10.446	10	116	—	—	0.346	
ruficaudus-senex	78.873	11.251	40	81	—	—	0.119	

*Average number of differences between groups. Values are for uncorrected nuclear loci and corrected based on the best-fit HKY + I model of sequence evolution for mtDNA. †Average pairwise divergence per site. ‡Number of polymorphic sites exclusive to each species. Species 1 refers to the first species listed in each pairwise comparison. Only a single *Tamias senex* was used, therefore polymorphism was not available for this taxon. §Ratio of polymorphism to divergence calculated as the proportion of segregating sites (θ for all sites, as given in Table 1) divided by the average percent pairwise divergence per site (*D*xy).

Table 3 Microsatellite variation in *Tamias ruficaudus simulans*, *Tamias amoenus canicaudus*, and *T. amoenus* (north)

	T. r. simulans (*n* = 44)				*T. a. canicaudus* (*n* = 35)				*T. amoenus* north (*n* = 26)				Total (*n* = 105)		
Locus	*A**	H_O†	H_E‡	*P*§	*A*	H_O	H_E	*P*	*A*	H_O	H_E	*P*	*A*	H_O	H_E
EuAmMS26	9	0.58	0.81	< 0.01	9	0.54	0.70	< 0.01	7	0.32	0.45	0.040	9	0.50	0.69
EuAmMS35	11	0.73	0.81		10	0.77	0.84		8	0.85	0.74	0.090	15	0.77	0.80
EuAmMS37	10	0.63	0.71	0.046	10	0.64	0.77	< 0.01	2	0.09	0.09		14	0.51	0.59
EuAmMS86	8	0.69	0.80	< 0.01	7	0.71	0.78		7	0.80	0.79		17	0.73	0.79
EuAmMS108	9	0.59	0.72	< 0.01	7	0.67	0.74		3	0.21	0.23		14	0.52	0.61
EuAmMS114	14	0.79	0.88		4	0.29	0.57	< 0.01	8	0.61	0.70		15	0.57	0.73
EuAmMS142	18	0.82	0.91	0.013	10	0.79	0.84	< 0.01	6	0.56	0.65		18	0.75	0.83
Mean	11.3	0.69	0.81	0.093	8.1	0.63	0.75		5.9	0.49	0.52	0.306	14.6	0.62	0.72
Standard error	1.3	0.04	0.03		0.9	0.07	0.04		0.9	0.11	0.10		1.1	0.05	0.04

*Number of alleles; †observed heterozygosity; ‡expected heterozygosity; §*P*-value for exact tests of Hardy–Weinberg equilibrium.

To determine population-level differentiation at microsatellite loci without imposing a priori groups, we used the model-based clustering method implemented in STRUCTURE. Likelihood scores increased from two (k = 2, ln L = –2718) to five populations (k = 5, ln L = –2248), and then became flat. With k = 3, the estimated population clusters coincided with the groupings of *T. r. simulans*, *T. amoenus* (north), and *T. a. canicaudus* (Fig. 4). As k is increased to four or five, additional geographically structured clusters emerged within *T. r. simulans* and *T. a. canicaudus*. Within *T. a. canicaudus*, individuals split into a western group (locality 80) and an eastern group (localities 40, 46, 83, 84, 90) while *T. r. simulans* individuals generally fell out along a north–south partition (north = localities 12, 15, 16, 35; south = localities 34, 41, 42, 59, 60, 61, 62). Pairwise F_{ST} and R_{ST} were quite high among groups when the data were partitioned into three or five clusters (Table 4, k = 3 or 5). Overall, there was little co-ancestry between *T. a. canicaudus* individuals

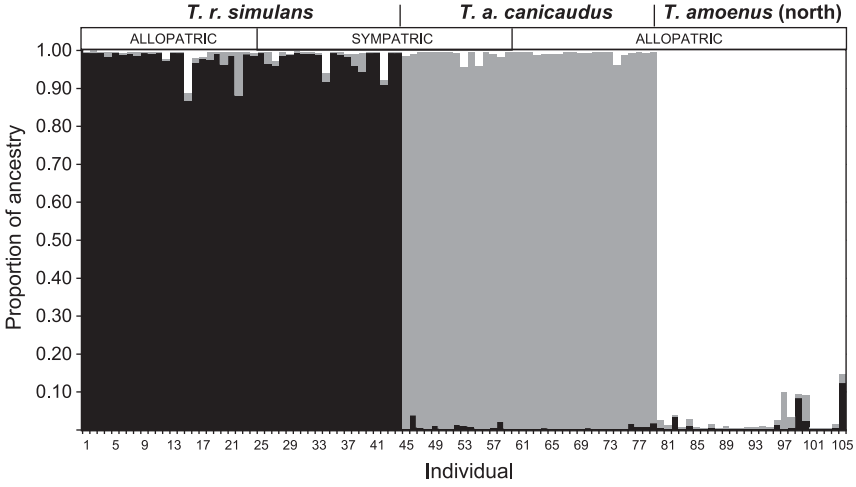

Fig. 4 Assessment of population structure with the program STRUCTURE based on seven microsatellite loci and assuming three populations ($k = 3$). The *y*-axis represents percent of the multilocus genotype in each individual that is attributable to co-ancestry with *T. r. simulans* (black), *T. a. canicaudus* (grey) and *T. amoenus* (north) (white). Individuals are grouped along the *x*-axis according to whether they were sampled from areas of sympatry or allopatry between *T. r. simulans* and *T. a. canicaudus*. All *T. amoenus* (north) sampling localities were allopatric with respect to *T. r. simulans* and *T. a. canicaudus*. Most individuals had unambiguous ancestry.

Table 4 Pairwise estimates of F_{ST} below the diagonal and R_{ST} above the diagonal

A. Pairwise comparisons between groups

	Tamias r. simulans	*Tamias a. canicaudus*	*Tamias amoenus* (north)
T. r. simulans	—	0.578	0.696
T. a. canicaudus	0.176	—	0.554
T. amoenus (north)	0.287	0.260	—

B. Pairwise comparisons between clusters defined by STRUCTURE ($k = 5$)

	T. r. simulans		*T. a. canicaudus*		*T. amoenus* (north)
	norh	south	east	west	
T. r. simulans north	—	0.078	0.751	0.600	0.714
T. r. simulans south	0.131	—	0.747	0.596	0.709
T. a. canicaudus east	0.302	0.205	—	0.637	0.752
T. a. canicaudus west	0.366	0.236	0.238	—	0.676
T. amoenus (north)	0.400	0.293	0.295	0.377	—

and either *T. amoenus* (north) or *T. r. simulans* samples (Fig. 4). Genetic ancestry of *T. r. simulans* and *T. amoenus* (north) was less resolved with nine individuals that showed greater than 5% coancestry. This may be partially due to the influence of gene flow from *T. r. ruficaudus* into some of the *T. r. simulans* populations, possibly leading to spurious assignments (N. Reid and S. Hird, unpublished data). Interestingly, five deviations involved shared ancestry between *T. r. simulans* and *T. amoenus* (north) to the exclusion of *T. a. canicaudus*. Because most of the significant co-ancestry occurs between currently allopatric populations (Fig. 4), it is likely that this was due to historical gene flow.

To test for recent gene flow, we calculated the proportion of the multilocus genotypes that were attributable to recent migration using the program BAYESASS. Estimates of recent migration among *T. r. simulans*, *T. a. canicaudus*, and *T. amoenus* (north) were effectively zero. Likewise, pairwise analyses using NEWHYBRIDS did not strongly assign any individuals to hybrid classes ($P > 0.75$) and most individuals were assigned with confidence to the pure classes ($P > 0.90$). Two *T. r. simulans* individuals had probabilities of greater than 50% of being derived from a backcross; one was from a location sympatric with *T. a. canicaudus* (site 61; Fig. 1) and the other was well outside the range of *T. a. canicaudus* (site 62).

Discussion

Recurrent hybridization in Tamias

Tamias amoenus and *T. ruficaudus* provide a compelling example of introgressive hybridization following initial lineage divergence. Multilocus comparisons revealed striking differences across loci in genealogical patterns (Figs 2 and 3; Table 2). The mtDNA cyt *b* genealogy was not reciprocally monophyletic between *T. ruficaudus* and

T. amoenus (Fig. 2) and had an unexpectedly high ratio of shared to fixed differences between the species (Table 2). By contrast, patterns observed at sequenced nuclear loci suggested two genetically differentiated species. Genealogies for three of the four loci (acr, acp5, c-myc) were characterized by a lack of shared haplotypes, one or more fixed differences, and few shared polymorphisms. The exception to this pattern occurs at RAG1, where one *T. ruficaudus* haplotype was sampled in a single homozygous *T. a. canicaudus* individual (Fig. 3). Finer scale population analyses revealed strong genetic differentiation at seven microsatellite loci with little evidence of co-ancestry (Fig. 4). Overall, these data suggest extensive asymmetric introgression of mtDNA from *T. ruficaudus* (*T. r. simulans*) into *T. amoenus* (*T. a. canicaudus*) with comparatively little introgression at nuclear loci.

Several lines of evidence indicate that mtDNA introgression between *T. a. canicaudus* and *T. r. simulans* was relatively ancient. First, introgression occurred sufficiently long ago for mtDNA genealogies to have developed complete reciprocal monophyly between *T. a. canicaudus* and *T. r. simulans* (Good *et al*. 2003). Second, the introgressed *T. ruficaudus* mtDNA has been present in *T. a. canicaudus* long enough to evolve phylogeographical structure (Good *et al*. 2003). Third, our estimate of the coalescent time between *T. a. canicaudus* and *T. r. simulans* indicates a relatively ancient origin of this hybridization event. Even assuming a rather large 95% credibility interval, the divergence between *T. a. canicaudus* and *T. r. simulans* is not likely to be earlier than mid-Pleistocene (> 1.1 million years ago) and is more likely to be pre-Pleistocene (*c*. 3.0 million years ago). The extent to which this estimate represents the actual timing of introgression depends on the length of time during which hybridization occurred, the pattern of genetic variation segregating in *T. r. simulans* at the time of introgression, and the assumed mutation rate. We chose to present analyses conducted assuming an empirical substitution rate estimated for mammalian mtDNA (Pesole *et al*. 1999), but recent work has suggested that the mtDNA mutation rate may be substantially faster (Ho *et al*. 2005). However, even assuming a mutation rate an order of magnitude faster, the lower bound of the 95% credibility interval still falls within the late Pleistocene.

Interestingly, levels of contemporary reproductive isolation between *T. amoenus* and *T. ruficaudus* vary across populations and/or subspecies. Gene flow appears to have ceased between previously hybridizing populations of *T. r. simulans* and *T. a. canicaudus*. Introgressed mtDNA has obtained reciprocal monophyly between *T. r. simulans* (Good *et al*. 2003) and *T. a. canicaudus* is strongly differentiated from *T. r. simulans* at microsatellite loci, with little residual co-ancestry (Fig. 4). Thus, *T. r. simulans* appears to have been reproductively isolated from *T. a. canicaudus* and *T. amoenus* (north) for an extended period of time. By contrast, recent and potentially ongoing hybridization between *T. r. ruficaudus* and *T. amoenus* (north) populations in the Canadian Rockies suggests that levels of gene flow vary across populations. Thus, reproductive isolation is not yet complete across the entire range of these species. This more recent hybridization event is the subject of an ongoing multilocus study (N. Reid *et al.* unpublished data).

Our data on introgression between *T. amoenus* and *T. ruficaudus* represents the first clear documentation of interspecific hybridization within *Tamias*. Nevertheless, semipermeable species limits may prove to be common in this genus. High levels of bacular divergence proved insufficient to isolate *T. amoenus* and *T. ruficaudus* completely, challenging many of the long-standing assumptions of reproductive isolation within *Tamias* based on genital morphology (White 1953; Patterson & Thaeler 1982; Patterson & Heaney 1987; Sutton 1995; Sutton & Patterson 2000). We do not know the extent of bacular divergence during this ancient introgression event but it is clear that introgression occurred well after the initiation of intraspecific divergence within *T. amoenus* and *T. ruficaudus*. Previous work has shown that *T. amoenus* and *T. ruficaudus* are completely nonoverlapping in bacular form (White 1953; Sutton 1992; Good *et al*. 2003) and it is unlikely that levels of divergence were historically much lower. Moreover, more recent instances of introgressive hybridization between these taxa (see Fig. 1) clearly involve gene flow between taxa with highly divergent bacular morphology (Good *et al*. 2003).

Another likely instance of introgressive hybridization occurs in the southern Rockies between the cliff chipmunk, *Tamias dorsalis*, and the grey-collared chipmunk, *Tamias cinereicollis* (Fitzpatrick & Turelli 2006). These broadly sympatric species are also divergent in bacular morphology yet are paraphyletic with respect to mtDNA (Piaggio & Spicer 2001). The unresolved mtDNA genealogy has been ascribed to incomplete lineage sorting (Piaggio & Spicer 2001). However, an earlier phylogenetic analysis of allozymes suggests the species are divergent nonsister taxa (Levenson *et al*. 1985). Thus, mtDNA paraphyly between this pair, as with *T. amoenus* and *T. ruficaudus*, may reflect extensive mtDNA introgression across species boundaries.

Asymmetric mtDNA introgression and insights into reproductive isolation

Mitochondrial gene flow between *T. amoenus* and *T. ruficaudus* is characterized by complete asymmetric introgression of *T. ruficaudus* mtDNA, with local fixation of the introgressed haplotypes. This pattern holds for ancient introgression into *T. a. canicaudus* and for several more recently introgressed *T. amoenus* (north) populations in the Canadian Rockies (Good *et al*. 2003). By contrast, nuclear introgression between *T. a. canicaudus* and *T. r. simulans*

was fairly limited (Figs 3 and 4) and genital morphology showed no evidence of hybridization in either case (Good *et al*. 2003). Extensive introgression and capture of mtDNA vs. nuclear loci may in principle be driven by selective or neutral processes (Ballard & Whitlock 2004). It is possible that *T. ruficaudus* mtDNA was introduced into *T. amoenus* through hybridization then swept to local fixation by positive natural selection, but this is difficult to evaluate for introgression into *T. a. canicaudus*. Locus-specific selection should generate a shallow, star-like phylogeny and a skew in the frequency spectrum towards rare alleles (Tajima 1989; Simonsen *et al*. 1995), as has been demonstrated in other examples of selective interspecific mtDNA introgression (e.g. Bachtrog *et al*. 2006). This is not the case for *T. a. canicaudus,* where the mtDNA genealogy is fairly structured and the frequency spectrum does not show an excess of rare alleles (Tajima's $D = 0.640$, $n = 46$). However, the power of this test is likely compromised by geographical structure within *T. a. canicaudus* and the ancient timing of the introgression. Indeed, it is likely that the frequency spectrum has had sufficient time to recover (Simonsen *et al*. 1995), even if directional selection drove the initial fixation of introgressed mtDNA.

Alternatively, extensive asymmetric mtDNA introgression may be a consequence of genetic drift and reflect the underlying cause(s) of reproductive isolation. For example, Ferris *et al*. (1983) have argued that extensive introgression of mtDNA between two species of house mice in Scandinavia (*Mus domesticus* introgression into *Mus musculus*), represents a historical founder event near the European hybrid zone between these species (see also Prager *et al*. 1993). Likewise, differential introgression of mtDNA due to unidirectional hybridization has been shown to occur between many animal species pairs (reviewed in Wirtz 1999). A recent theoretical comparison of pre- and post-zygotic models of isolation demonstrated that, under certain conditions, models of prezygotic isolation (e.g. female choice or male–male competition) allow for much more rapid introgression of maternally inherited DNA (Chan & Levin 2005). This result should be strongest when the source species of the mtDNA is relatively rare (Wirtz 1999; Chan & Levin 2005). Secondary contact between *T. r. ruficaudus* and *T. amoenus* in the Canadian Rockies occurs at the far northern extension of the *T. ruficaudus* range where populations densities are generally expected to be low (Brown *et al*. 1996). Likewise, it is possible that ancient hybridization between *T. a. canicaudus* and *T. r. simulans* occurred during climate-induced range shifts that may have contributed to historically low population densities.

Divergent genital morphology may be important for the apparent asymmetry of reproductive isolation in this system (Good *et al*. 2003). The baculum is larger in *T. ruficaudus* (Fig. 2), which may cause asymmetric mechanical isolation in crosses with *T. amoenus*. In general, it has been hypothesized that bacular divergence in *Tamias* and other mammals is driven by direct selection to avoid interspecific hybridization (Patterson & Thaeler 1982). However, several aspects of our data conflict with this hypothesis. First, bacular differentiation is not obviously more exaggerated in areas of sympatry between *T. ruficaudus* and *T. amoenus* (Good *et al*. 2003). Second, the baculum of *T. a. canicaudus* has not diverged from that of *T. amoenus* populations directly to the north despite a history of hybridization with *T. r. simulans* and little/no recent gene flow between *T. a. canicaudus* and *T. amoenus* (north) (Fig. 4). Certainly, the ability of the baculum to respond to this form of selection would depend on the fitness of hybrids. If hybrid offspring have equivalent fitness to parentals, selection for prezygotic isolation through bacular divergence or any other mode of reinforcement would not be expected (Butlin 1989; Coyne & Orr 2004). Regardless, rapid evolution of genital morphology in *Tamias* likely reflects strong sexual selection within species because of cryptic female choice (Eberhard 1996) and thus has the potential to play a role in interspecific isolation. As argued by Coyne and Orr (2004, pp. 229–230), if females choose among male genitalia within a species then discrimination against interspecific genitalia seems plausible if not probable.

Broader implications of mtDNA capture

Mallet (2005) has suggested that as many as 6–10% of animal species hybridize. Interestingly, extensive introgression and capture of mtDNA vs. nuclear loci now appears to be common among hybridizing animal species (Avise 2004). For example, mtDNA capture in *T. amoenus* adds to a growing list of mammalian species with introgressed mitochondrial genomes that includes deer (Cathey *et al*. 1998), goats (Ropiquet & Hassanin 2006), hares (Thulin *et al*. 1997; Thulin & Tegelstrom 2002; Melo-Ferreira *et al*. 2005), house mice (Ferris *et al*. 1983; Prager *et al*. 1993), pocket gophers (Ruedi *et al*. 1997), and voles (Tegelstrom 1987). The practical ramifications of cryptic hybridization in mammals and other taxa echo broadly. For instance, DNA barcoding studies that rely on a single mtDNA marker (e.g. COI; Hebert *et al*. 2003) will be severely compromised by the propensity of mtDNA to introgress. Mitochondrial DNA capture is only detectable using multiple markers, including morphology and/or molecular data from the nuclear genome.

Conclusions

Multilocus analyses of mtDNA and nuclear markers presented here, combined with previous data on phylogeographical and morphological variation, reveal a complex history of population subdivision and gene flow since the initial divergence of *Tamias amoenus* and *T. ruficaudus*

lineages. These data provide an assessment of the genetic impact of historical and recurrent hybridization between these species and offer insight into the current status of reproductive isolation as well as the potential mechanisms underlying speciation. Our work highlights the importance of using information from independent genetic markers when evaluating the evolutionary history of hybridizing species. In particular, resolution of species limits in the face of temporal and spatially varying hybridization may often require a combination of diverse genetic, taxonomic, and population-level sampling.

Acknowledgements

We thank N. Panter and D. Nagorsen (the Royal BC Museum), L. Heaney and B. Patterson (Field Museum), B. McGillivray and B. Weimann (Provincial Museum of Alberta), D. Paulson (Slater Museum of Natural History), A. Sonnedena (Nez Perce Tribe), K. Stone (Southern Oregon University), and University of Idaho Mammalogy classes (1998–2003) for specimens and assistance with field collection. We thank M. Carneiro, M. Dean, C. Drummond, A. Geraldes and W. Maddison for helpful discussions during the course of this project. Pierre Taberlet, John Wiens and several anonymous reviewers provided very helpful comments. We also thank J. Patton (Museum of Vertebrate Zoology) for providing unpublished cyt *b* sequences. This research received support from the UI Initiative for Bioinformatics and the Study of Evolution (IBEST), which has received funding through NIH (NCRR 1P20RRO16448-01) and NSF (EPS-809935).

References

Anderson EC, Thompson EA (2002) A model-based method for identifying species hybrids using multilocus genetic data. *Genetics* **160**, 1217–1229.

Avise JC (2004) *Molecular Markers, Natural History, and Evolution*, 2nd edn. Sinauer & Associates, Sunderland, Massachusetts.

Bachtrog D, Thornton K, Clark A, Andolfatto P (2006) Extensive introgression of mitochondrial DNA relative to nuclear genes in the *Drosophila yakuba* species group. *Evolution*, **60**, 292–302.

Ballard JWO, Whitlock MC (2004) The incomplete natural history of mitochondria. *Molecular Ecology*, **13**, 729–744.

Bandelt HJ, Forster P, Rohl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, **16**, 37–48.

Berthier P, Excoffier L, Ruedi M (2006) Recurrent replacement of mtDNA and cryptic hybridization between two sibling bat species *Myotis myotis* and *Myotis blythii*. *Proceedings of the Royal Society B: Biological Sciences*, **273**, 3101–3109.

Best TL (1993) *Tamias ruficaudus*. *Mammalian Species*, **452**, 1–7.

Broughton RE, Harrison RG (2003) Nuclear gene genealogies reveal historical, demographic and selective factors associated with speciation in field crickets. *Genetics*, **163**, 1389–1401.

Brown JH, Stevens GC, Kaufman DM (1996) The geographic range: size, shape, boundaries, and internal structure. *Annual Review of Ecology and Systematics*, **27**, 597–623.

Butlin RK (1989) Reinforcement of premating isolation. In: *Speciation and its Consequences* (eds Otte D, Endler JA), pp. 158–179. Sinauer & Associates, Sunderland, Massachusetts.

Cathey JC, Bickham JW, Patton JC (1998) Introgressive hybridization and nonconcordant evolutionary history of maternal and paternal lineages in North American deer. *Evolution*, **52**, 1224–1229.

Chan KMA, Levin SA (2005) Leaky prezygotic isolation and porous genomes: rapid introgression of maternally inherited DNA. *Evolution*, **59**, 720–729.

Coyne JA, Orr HA (2004) *Speciation*. Sinauer & Associates, Sunderland, Massachusetts.

DeBry RW (1998) Comparative analysis of evolution in a rodent histone H2a pseudogene. *Journal of Molecular Evolution*, **46**, 355–360.

Demboski JR, Sullivan J (2003) Extensive mtDNA variation within the yellow-pine chipmunk, *Tamias amoenus* (Rodentia: Sciuridae), and phylogeographic inferences for northwest North America. *Molecular Phylogenetics and Evolution*, **26**, 389–408.

Dobzhansky T (1951) *Genetics and the Origin of Species*, 3rd edn. Columbia University Press, New York.

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

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, **39**, 783–791.

Ferris SD, Sage RD, Huang CM, Nielsen JT, Ritte U, Wilson AC (1983) Flow of mitochondrial DNA across a species boundary. *Proceedings of the National Academy of Sciences, USA*, **80**, 2290–2294.

Fitzpatrick BM, Turelli M (2006) The geography of mammalian speciation: mixed signals from phylogenies and range maps. *Evolution*, **60**, 601–615.

Funk DJ, Omland KE (2003) Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annual Review of Ecology Evolution and Systematics*, **34**, 397–423.

Geraldes A, Ferrand N, Nachman MW (2006) Contrasting patterns of introgression at X-linked loci across the hybrid zone between subspecies of the European rabbit (*Oryctolagus cuniculus*). *Genetics*, **173**, 919–933.

Good JM, Sullivan J (2001) Phylogeography of the red-tailed chipmunk (*Tamias ruficaudus*), a northern Rocky Mountain endemic. *Molecular Ecology*, **10**, 2683–2695.

Good JM, Demboski JR, Nagorsen DW, Sullivan J (2003) Phylogeography and introgressive hybridization: chipmunks (genus *Tamias*) in the northern Rocky Mountains. *Evolution*, **57**, 1900–1916.

Goodman SJ, Barton NH, Swanson G, Abernethy K, Pemberton JM (1999) Introgression through rare hybridization: a genetic study of a hybrid zone between red and sika deer (genus *Cervus*) in Argyll, Scotland. *Genetics*, **152**, 355–371.

Goudet J (1995) FSTAT version 1.2: a computer program to calculate F-statistics. *Journal of Heredity* **86**, 485–486.

Grant PR, Grant BR, Petren K (2005) Hybridization in the recent past. *American Naturalist*, **166**, 56–67.

Gustincich S, Manfioletti G, Delsal G, Schneider C, Carninci P (1991) A fast method for high-quality genomic DNA extraction from whole human blood. *BioTechniques*, **11**, 298–302.

Hall ER (1981) *The Mammals of North America. Two volume*. Wiley & Sons, New York.

Harrison RG (1990) Hybrid zones: windows on evolutionary process. *Oxford Surveys in Evolutionary Biology*, **7**, 69–128.

Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society B: Biological Sciences*, **270**, 313–322.

Hey J, Kliman RM (1993) Population genetics and phylogenetics of DNA sequence variation at multiple loci within the *Drosophila melanogaster* species complex. *Molecular Biology and Evolution*, **10**, 804–822.

Hey J, Nielsen R (2007) Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proceedings of the National Academy of Sciences, USA*, **104**, 2785–2790.

Hilton H, Hey J (1997) A multilocus view of speciation in the *Drosophila virilis* species group reveals complex histories and taxonomic conflicts. *Genetical Research*, **70**, 185–194.

Ho SYW, Phillips MJ, Cooper A, Drummond AJ (2005) Time dependency of molecular rate estimates and systematic over-estimation of recent divergence times. *Molecular Biology and Evolution* **22**, 1561–1568.

Hudson RR, Kaplan N (1985) Inferring the number of recombination events in the history of a sample. *Biometrics*, **41**, 572–572.

Hudson RR, Kreitman M, Aguade M (1987) A test of neutral molecular evolution based on nucleotide data. *Genetics*, **116**, 153–159.

Kliman RM, Andolfatto P, Coyne JA *et al*. (2000) The population genetics of the origin and divergence of the *Drosophila simulans* complex species. *Genetics*, **156**, 1913–1931.

Levenson H, Hoffmann RS, Nadler CF, Deutsch L, Freeman SD (1985) Systematics of the holarctic chipmunks (*Tamias*). *Journal of Mammalogy*, **66**, 219–242.

Llopart A, Lachaise D, Coyne JA (2005) Multilocus analysis of introgression between two sympatric sister species of *Drosophila*: *Drosophila yakuba* and *D. santomea*. *Genetics*, **171**, 197–210.

Machado CA, Kliman RM, Markert JA, Hey J (2002) Inferring the history of speciation from multilocus DNA sequence data: the case of *Drosophila pseudoobscura and* close relatives. *Molecular Biology and Evolution*, **19**, 472–488.

Maddison DR, Maddison WP (2003) MACCLADE. Sinauer & Associates, Sunderland, Massachusetts.

Mallet J (2005) Hybridization as an invasion of the genome. *Trends in Ecology & Evolution*, **20**, 229–237.

Mayr E (1963) *Animal Species and Evolution*. Belknap Press, Cambridge, Massachusetts.

McDonald JH, Kreitman M (1991) Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature*, **351**, 652–654.

Melo-Ferreira J, Boursot P, Suchentrunk F, Ferrand N, Alves PC (2005) Invasion from the cold past: extensive introgression of mountain hare (*Lepus timidus*) mitochondrial DNA into three other hare species in northern Iberia. *Molecular Ecology*, **14**, 2459–2464.

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

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

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

Patterson BD, Heaney LR (1987) Preliminary analysis of geographic variation in red-tailed chipmunks (*Eutamias ruficaudus*). *Journal of Mammalogy*, **68**, 782–791.

Patterson BD, Thaeler CSJ (1982) The mammalian baculum: hypotheses on the nature of bacular variability. *Journal of Mammalogy*, **63**, 1–15.

Payseur BA, Krenz JG, Nachman MW (2004) Differential patterns of introgression across the X chromosome in a hybrid zone between two species of house mice. *Evolution*, **58**, 2064–2078.

Pesole G, Gissi C, De Chirico A, Saccone C (1999) Nucleotide substitution rate of mammalian mitochondrial genomes. *Journal of Molecular Evolution*, **48**, 427–434.

Piaggio AJ, Spicer GS (2001) Molecular phylogeny of the chipmunk genus *Tamias* inferred from the mitochondrial cytochrome *b* and cytochrome oxidase II gene sequences. *Molecular Phylogenetics and Evolution*, **20**, 335–350.

Prager EM, Sage RD, Gyllensten U *et al*. (1993) Mitochondrial DNA sequence diversity and the colonisation of Scandinavia by house mice from East Holstein. *Biological Journal of the Linnean Society*, **50**, 85–122.

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

Putnam AS, Scriber JM, Andolfatto P (2007) Discordant divergence times among Z-chromosome regions between two ecologically distinct swallowtail butterfly species. *Evolution*, **61**, 912–927.

Raymond M, Rousset F (1995) GENEPOP (version 1.2) — population genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**, 248–249.

Rieseberg LH, Whitton J, Gardner K (1999) Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. *Genetics*, **152**, 713–727.

Ropiquet A, Hassanin A (2006) Hybrid origin of the Pliocene ancestor of wild goats. *Molecular Phylogenetics and Evolution*, **41**, 395–404.

Rozas J, Sanchez-De I, Barrio JC, Messeguer X, Rozas R (2003) DNASP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*, **19**, 2496–2497.

Ruedi M, Smith MF, Patton JC (1997) Phylogenetic evidence of mitochondrial DNA introgression among pocket gophers in New Mexico (family Geomyidae). *Molecular Ecology*, **6**, 453–462.

Sambrook J, Russell DW (2001) *Molecular Cloning: A Laboratory Manual*, 3rd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Schulte-Hostedde AI, Gibbs HL, Millar JS (2000) Microsatellite DNA loci suitable for parentage analysis in the yellow-pine chipmunk (*Tamias amoenus*). *Molecular Ecology*, **9**, 2180–2181.

Simonsen KL, Churchill GA, Aquadro CF (1995) Properties of statistical tests of neutrality for DNA polymorphism data. *Genetics*, **141**, 413–429.

Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics*, **68**, 978–989.

Steppan SJ, Adkins RM, Anderson J (2004a) Phylogeny and divergence-date estimates of rapid radiations in muroid rodents based on multiple nuclear genes. *Systematic Biology*, **53**, 533–553.

Steppan SJ, Storz BL, Hoffmann RS (2004b) Nuclear DNA phylogeny of the squirrels (Mammalia: Rodentia) and the evolution of arboreality from c-myc and RAG1. *Molecular Phylogenetics and Evolution*, **30**, 703–719.

Sullivan J, Abdo Z, Joyce P, Swofford DL (2005) Evaluating the performance of a successive-approximations approach to parameter optimization in maximum-likelihood phylogeny estimation. *Molecular Biology and Evolution*, **22**, 1386–1392.

Sutton DA (1992) *Tamias amoenus*. *Mammalian Species*, **390**, 1–8.

Sutton DA (1995) Problems of taxonomy and distribution in four species of chipmunks. *Journal of Mammalogy*, **76**, 843–850.

Sutton DA, Patterson BD (2000) Geographic variation of the western chipmunks *Tamias senex* and *T. siskiyou*, with two new subspecies from California. *Journal of Mammalogy*, **81**, 299–316.

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

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

Tegelstrom H (1987) Transfer of mitochondrial DNA from the northern red-backed vole (*Clethrionomys rutilus*) to the bank vole (*Clethrionomys glareolus*). *Journal of Molecular Evolution*, **24**, 218–227.

Thulin CG, Tegelstrom H (2002) Biased geographical distribution of mitochondrial DNA that passed the species barrier from mountain hares to brown hares (genus *Lepus*): an effect of genetic incompatibility and mating behaviour? *Journal of Zoology*, **258**, 299–306.

Thulin CG, Jaarola M, Tegelstrom H (1997) The occurrence of mountain hare mitochondrial DNA in wild brown hares. *Molecular Ecology*, **6**, 463–467.

Turner TL, Hahn MW, Nuzhdin SV (2005) Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biology*, **3**, 1572–1578.

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

Wang RL, Wakeley J, Hey J (1997) Gene flow and natural selection in the origin of *Drosophila pseudoobscura* and close relatives. *Genetics*, **147**, 1091–1106.

Watterson GA (1975) Number of segregating sites in genetic models without recombination. *Theoretical Population Biology*, **7**, 256–276.

White JA (1953) The baculum in the chipmunks of western North America. *University of Kansas Museum Of Natural History Publications*, **5**, 611–631.

Wilson GA, Rannala B (2003) Bayesian inference of recent migration rates using multilocus genotypes. *Genetics*, **163**, 1177–1191.

Wirtz P (1999) Mother species–father species: unidirectional hybridization in animals with female choice. *Animal Behaviour*, **58**, 1–12.

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

Jeffrey Good studies speciation and adaptation in mammals. Sarah Hird and Noah Reid are MS students studying chipmunk hybridization and phylogeny, respectively. John Demboski is Curator of Vertebrate Zoology at the Denver Museum of Nature & Science. Scott Steppan and Jack Sullivan are mammalian systematists.

Supplementary material

The following supplementary material is available for this article:

Fig. S1 Posterior probability distributions of (a) theta (b) migration, and (c) time since divergence calculated using MDIV and the mtDNA (cyt *b*) sequences from *Tamias amoenus canicaudus* and *Tamias ruficaudus simulans*. The scale of the *x*-axis in (c) was set by assuming a per locus mutation rate of 5.8×10^{-6} and a generation length of 1 year. See text for a discussion of the mutation rate.

Table S1 Sampling across localities

This material is available as part of the online article from:
http://www.blackwell-synergy.com/doi/abs/10.1111/j.1365-294X.2007.03640.x
(This link will take you to the article abstract).

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