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 sympathy. 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...
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, interspecific 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 baculiform 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; Coyle & Orr 2004).

Here, we use a multifaceted 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 mitochondrial 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 used 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 (TATGGTRAGATGCTGCAACTG) and ACR 15R (GCA-
TGAGCCAGCTGTAAGCACC), which covers ~1140 bp, from exon 1-exon 2; and ACR13F (CTCAGCTGGCTCCTAC-
AGCTGC), 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 precipita-
tion 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 paur*.
4.0 (Swofford 2002) using successive approximations (Sullivan
et al. 2005). Nodal support was estimated using the
nonparametric ML bootstraps (Felsenstein 1985) in paur*.
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 hetero-
geneity 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
to 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 wwh 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 sima (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 $\theta = 4N(\mu)$, the scaled population splitting time $t$, and two directional population migration rates ($m_1$ and $m_2$). Here, $\mu$ is the mutation rate per locus per generation, $N_r$ 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 number of migrants per generation. Nuclear loci were all recombinant sequences. Second, we used the program of the model ($\theta$ (Hudson & Kaplan 1985) was used to identify and remove sequence evolution. The four-gametic haplotype test (WWH) statistic (Wang et al. 1997) was calculated from the sum of 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.

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.

**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 skew. 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
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 \textit{T. amoenus} clade and the clade containing \textit{T. a. canicaudus} and both subspecies of \textit{T. ruficaudus}, with \textit{T. a. canicaudus} forming a strongly supported clade with \textit{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). \textit{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 \textit{T. ruficaudus} and \textit{T. amoenus} separated by a minimum of one fixed difference. RAG1 was partitioned into three groups with a deep split within \textit{T. amoenus} and one haplotype shared between multiple \textit{T. ruficaudus} samples and one homozygous \textit{T. a. canicaudus} individual. \textit{Tamias amoenus} appeared paraphyletic at RAG1, even if this shared haplotype was not considered (analysis not shown). The high divergence within \textit{T. amoenus} observed at RAG1 is in general agreement with a deep mtDNA split between \textit{T. amoenus} (north) and other populations of \textit{T. amoenus} (Demboski & Sullivan 2003) whereas the apparent paraphyly may reflect insufficient variation for resolution of deep internal nodes. Interestingly, \textit{T. a. canicaudus} fell out with \textit{T. amoenus} (north) (excluding the single individual grouped within \textit{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 \textit{T. amoenus} to divergence with \textit{T. ruficaudus} or \textit{T. senex}. In these comparisons, both cyt \textit{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 \textit{b} and RAG1 was not contributing to the overall deviation.

### Table 1  Nucleotide variation within \textit{Tamias amoenus} and \textit{T. ruficaudus}

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

*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 \textit{b}); **minimum number of recombination events (Hudson & Kaplan 1985).
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).
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 (~1/4Ne 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 (θ*amoenus*, θ*ima*, θ*ima*, m1, m2, and I) 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 (θ*amoenus*, θ*ima*, m1, m2), we were unable to obtain stable estimates of the full posterior density and the scaled population splitting time I. 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 θ = 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 Nμ = 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 Ne 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).
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.

### Table 2: Polymorphism and divergence across five sequenced loci

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

*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)

<table>
<thead>
<tr>
<th>Locus</th>
<th>$T. r. simulans (n = 44)$</th>
<th>$T. a. canicaudus (n = 35)$</th>
<th>$T. amoenus$ north ($n = 26$)</th>
<th>Total ($n = 105$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$A$</td>
<td>$H_o$†</td>
<td>$H_e$‡</td>
<td>$P$§</td>
</tr>
<tr>
<td>EuAmMS26</td>
<td>9</td>
<td>0.58</td>
<td>0.81</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>EuAmMS35</td>
<td>11</td>
<td>0.73</td>
<td>0.81</td>
<td></td>
</tr>
<tr>
<td>EuAmMS37</td>
<td>10</td>
<td>0.63</td>
<td>0.71</td>
<td>0.046</td>
</tr>
<tr>
<td>EuAmMS86</td>
<td>8</td>
<td>0.69</td>
<td>0.80</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>EuAmMS108</td>
<td>9</td>
<td>0.59</td>
<td>0.72</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>EuAmMS114</td>
<td>14</td>
<td>0.79</td>
<td>0.88</td>
<td></td>
</tr>
<tr>
<td>EuAmMS142</td>
<td>18</td>
<td>0.82</td>
<td>0.91</td>
<td>0.013</td>
</tr>
<tr>
<td>Mean</td>
<td>11.3</td>
<td>0.69</td>
<td>0.81</td>
<td>0.093</td>
</tr>
<tr>
<td>Standard error</td>
<td>1.3</td>
<td>0.04</td>
<td>0.03</td>
<td></td>
</tr>
</tbody>
</table>

*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.
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

---

**Table 4** Pairwise estimates of \(F_{ST}\) below the diagonal and \(R_{ST}\) above the diagonal

A. Pairwise comparisons between groups

<table>
<thead>
<tr>
<th></th>
<th><em>Tamias r. simulans</em></th>
<th><em>Tamias a. canicaudus</em></th>
<th><em>Tamias amoenus</em> (north)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>T. r. simulans</em></td>
<td>—</td>
<td>0.578</td>
<td>0.696</td>
</tr>
<tr>
<td><em>T. a. canicaudus</em></td>
<td>0.176</td>
<td>—</td>
<td>0.554</td>
</tr>
<tr>
<td><em>T. amoenus</em> (north)</td>
<td>0.287</td>
<td>0.260</td>
<td>—</td>
</tr>
</tbody>
</table>

B. Pairwise comparisons between clusters defined by *structure* (\(k = 5\))

<table>
<thead>
<tr>
<th></th>
<th><em>T. r. simulans</em></th>
<th><em>T. a. canicaudus</em></th>
<th><em>Tamias amoenus</em> (north)</th>
</tr>
</thead>
<tbody>
<tr>
<td>north</td>
<td>—</td>
<td>0.751</td>
<td>0.600</td>
</tr>
<tr>
<td>south</td>
<td>0.131</td>
<td>—</td>
<td>0.596</td>
</tr>
<tr>
<td><em>T. a. canicaudus</em> east</td>
<td>0.302</td>
<td>—</td>
<td>0.637</td>
</tr>
<tr>
<td>west</td>
<td>0.366</td>
<td>0.205</td>
<td>0.752</td>
</tr>
<tr>
<td><em>T. amoenus</em> (north)</td>
<td>0.400</td>
<td>0.236</td>
<td>0.377</td>
</tr>
</tbody>
</table>

---

© 2008 The Authors

Journal compilation © 2008 Blackwell Publishing Ltd
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 reciprocally 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 paraparous 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 postzygotic 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 sympathy 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 parents, 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 & Tegelström 2002; Melo-Ferreira et al. 2005), house mice (Ferris et al. 1983; Prager et al. 1993), pocket gophers (Ruedi et al. 1997), and voles (Tegelström 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 during the course of this project. Pierre Taberlet, John Wiens and A. Geraldes and W. Maddison for helpful discussions 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


ANCIENT HYBRIDIZATION IN CHIPMUNKS


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 \times 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.