Genetica (2010) 138:869–883
DOI 10.1007/s10709-010-9470-z

Introgression at differentially aged hybrid zones in red-tailed chipmunks

Sarah Hird · Noah Reid · John Demboski · Jack Sullivan

Received: 25 August 2009 / Accepted: 25 January 2010 / Published online: 11 July 2010

Abstract Hybrid zones allow us to investigate the maintenance and the break down of reproductive isolation; they are a window into the speciation process. *Tamias ruficaudus* (red-tailed chipmunk) has a roughly ring-like distribution in the Inland Northwest and includes two morphologically well-differentiated subspecies, *T. r. ruficaudus* (in the eastern portion of its range) and *T. r. simulans* (in the western portion). These taxa meet at two contact zones: the Lochsa River in Idaho and 200 km to the north, near Whitefish, Montana. The Lochsa Zone is encompassed within the Clearwater River Drainage, which has been proposed as a glacial refugium for many taxa throughout the Pleistocene, whereas the Whitefish Zone was under the Cordilleran ice sheet during the most recent glacial maxima approximately 10,000 years ago. Mitochondrial DNA introgression has been documented at both contact zones, yet the subspecies remain significantly distinct with respect to bacular morphology and no intermediate morphologies have ever been reported. Here, we elucidate differentiation and introgression of the nuclear genome using ten microsatellite loci and compare findings to previously described mitochondrial DNA haplotype distribution and introgression. We found significant substructure in the nuclear data; each subspecies is divided into at least two genetically distinct demes. At the Lochsa contact zone, individuals restricted to the mtDNA zone of introgression form a distinct deme at microsatellite loci whereas in the younger, Whitefish contact zone, there is no hybrid-zone specific group. The genetic distances of the demes within these two subspecies indicate recent northward expansion.

Keywords Hybridization · Introgression · Speciation · *Tamias*

S. Hird (✉) · N. Reid
Department of Biological Sciences, Louisiana State University, 202 Life Sciences Building, Baton Rouge, LA 70803, USA
e-mail: shird1@tigers.lsu.edu

J. Demboski
Department of Zoology, Denver Museum of Nature & Science, 2001 Colorado Boulevard, Denver, CO 80205-5798, USA

J. Sullivan
Department of Biological Sciences, University of Idaho, Box 443051, Moscow, ID 83844, USA

Introduction

A central goal of evolutionary biology is to understand the processes that generate biodiversity and, in particular, that promote speciation. Because of the persistence of the Biological Species Concept (BSC) of Mayr (1942), the study of speciation has very frequently been framed in terms of understanding the origin and maintenance of reproductive isolating barriers (e.g., Coyne and Orr 2004). Hybridization, in this view, represents a breakdown of reproductive isolation and has historically been thought of as strictly in opposition to species divergence (e.g., Dobzhansky 1951; Mayr 1963). However, there is now growing support for an opposing view, the speciation-with-gene-flow model (e.g., Wu 2001), which allows for hybridization during the process of speciation and may describe a common mode of divergence (e.g., Rice and Hostert 1993; Nosil 2008). In this alternative view, hybridization may have a variety of consequences and may be an important aspect of speciation.

Hybrid zones have long been used to investigate evolutionary processes such as differentiation and adaptation because they provide a context in which to analyze

migration and gene flow, as well as explore the effects of ecological divergence through the analysis of genetic or morphological clines (Barton and Hewitt 1985). In recent years, studies of hybrid zones have provided evidence in support of the divergence-with-gene-flow model of speciation as they have been shown to act as a filter between incompletely isolated taxa, allowing the introgression of neutral or advantageous alleles while filtering out deleterious ones (Martinsen et al. 2001; Brumfield et al. 2001). A multi-locus genetic approach to the study of hybrid zones allows identification of neutral variation that may indicate patterns of isolation, contact, hybridization or lineage sorting and proves very useful in the study of speciation.

Chipmunks (*Tamias*) provide a useful study system for questions of speciation and hybridization. The 23 chipmunk species in the western North America clade likely represent a rapid radiation; there is only one species in eastern North America and one species in Eurasia. They are classic niche partitioners who competitively exclude congeners from specific habitats but have broadly overlapping fundamental niches (Heller 1971; Heller and Gates 1971). *Tamias* species are frequently difficult to identify using most skeletal and pelage characters; but a generally reliable marker is bacular morphology, a character that tends to exhibit strongly discontinuous variation among taxa (White 1953; Good et al. 2003). Because of the pattern, divergence in bacular morphology has traditionally been thought to represent reproductive isolation, perhaps being a mechanical barrier to gene flow (Patterson and Thaeler 1982). However, there is growing evidence that hybridization (in spite of bacular divergence) is likely an important factor in the diversification of *Tamias*, in western North America. There are at least two documented cases of mitochondrial introgression across non-sister species (Good et al. 2003, 2008; Reid et al. 2010) and phylogenetic evidence for introgressive hybridization among several other species (Reid et al. unpublished data).

Tamias ruficaudus (red-tailed chipmunk) is a northern Rocky Mountain endemic, and it occupies mesic forest habitats in the Inland Northwest. Two subspecies are described: eastern *T. r. ruficaudus* and western *T. r. simulans*. They form a ring-like distribution and meet at two contact zones. The southern contact zone occurs along the Lochsa River in central Idaho and the northern zone is northeast of Whitefish, Montana (Fig. 1). The northern zone (the Whitefish zone) occurs in an area that was covered by the Cordilleran ice sheet during Pleistocene cold periods (Delcourt and Delcourt 1993). The pollen record suggests that suitable forest habitat for *T. ruficaudus* has only been present north of the southern extent of glaciation in the last few thousand years (Mack et al. 1978). Conversely, the southern contact zone (the Lochsa zone) occurs in the Clearwater Drainage, an area that was likely a glacial refugium for mesic forest taxa throughout the climatic changes of the Pleistocene (Daubenmire 1952; Detling 1968; Carstens et al. 2005). Thus, although the precise age of the Lochsa zone is difficult to identify, the Whitefish zone is clearly very recent (post-Pleistocene) whereas the Lochsa zone is more ancient.

The subspecies have long been recognized on the basis of pelage characteristics (Howell 1922). Subsequently, Patterson and Heaney (1987) postulated that these subspecies represent full species based on differentiation of the baculum (*os penis:* Fig. 1), a key taxonomic character in chipmunks (White 1953) and other sciurids. They did not formally recommend the elevation of each subspecies to species status because the location and nature of the contact zones were not established. Good and Sullivan (2001) located and sampled the two contact zones between the subspecies and confirmed a significant difference in subspecific bacular morphology (Canonical Variates Analysis, Good et al. 2003). They documented no morphological intermediates and distinct bacular morphs were segregated on opposite banks of the Lochsa River (Good and Sullivan 2001). Additionally, they documented two strongly supported mitochondrial DNA (mtDNA) clades, assignable to subspecies, but found apparent introgression at both contact zones. An eastern mtDNA haplotype is found in every *T. r. ruficaudus* individual and in many *T. r. simulans* individuals near the contact zones; a western mtDNA clade is restricted to *T. r. simulans* individuals. In the Lochsa zone, the mtDNA contact zone is displaced north and west at least 150 km from the bacular contact zone, whereas in the Whitefish zone it is displaced by around 25 km. Furthermore, Hird and Sullivan (2009) showed that at the Lochsa River, a group of populations within the hybrid zone—the area with *T. r. simulans* bacular morphology and eastern mtDNA haplotypes—has differentiated from both parentals at nuclear loci.

Our primary aim is to characterize and explain the difference in introgression at the two contact zones and secondarily, assess population substructure across the species. Based on the differential, unidirectional, mtDNA introgressions observed by Good and Sullivan (2001), and the observed non-zero levels of gene flow (Hird and Sullivan 2009) we address the following hypotheses:

Hypothesis 1 Similar to the differential introgression of mtDNA, the nuclear loci (ten microsatellite loci) will have differentially introgressed across the contact zones. The older contact zone, at the Lochsa River, will have more introgressed or hybrid genotypes, whereas the purportedly younger contact zone near Whitefish, MT may have more restricted introgression. These introgressions should not be unidirectional, like the mtDNA, since nuclear DNA is not uniparentally inherited.

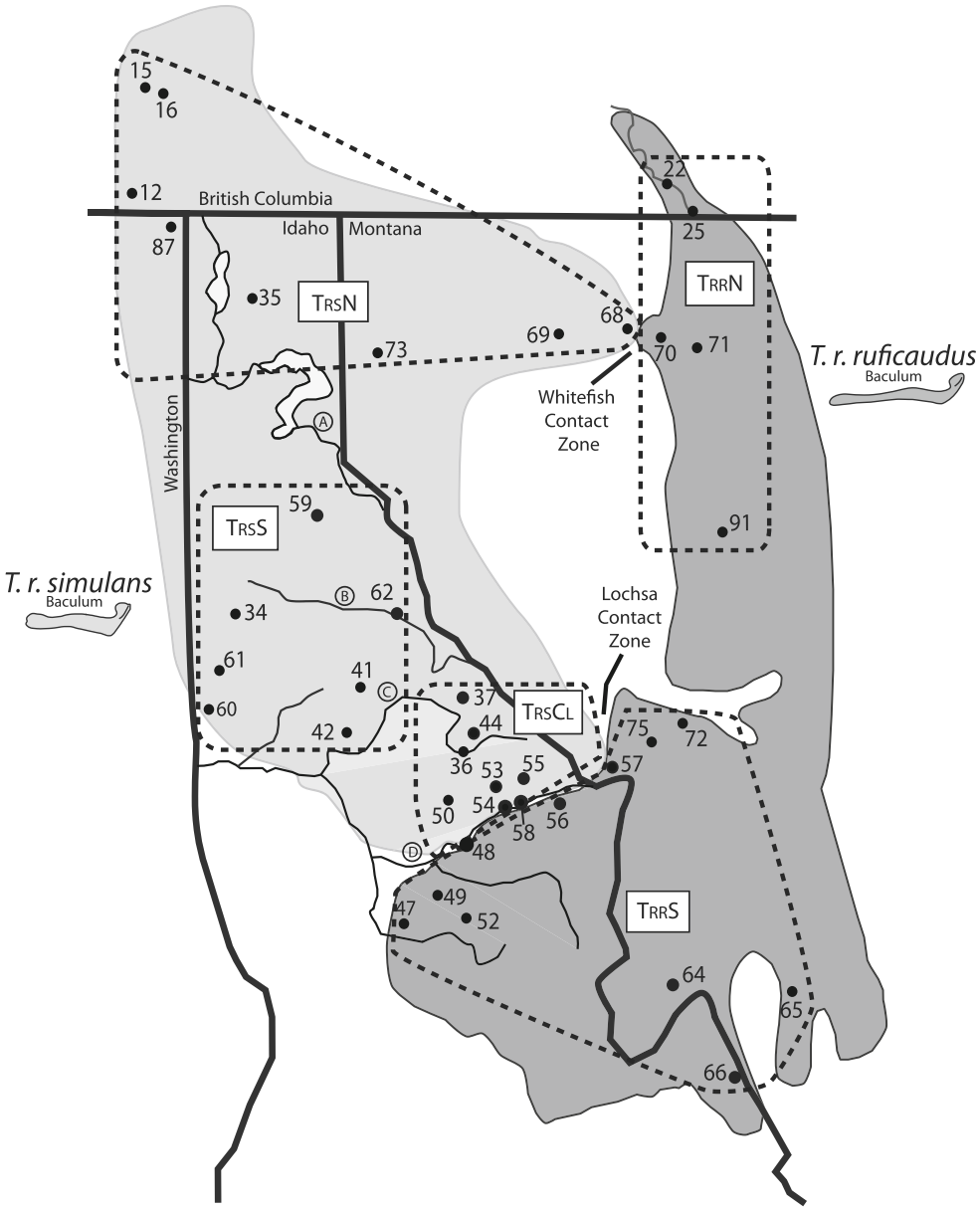

Fig. 1 Distribution of *Tamias ruficaudus*. *Shading* represents the subspecies ranges. *Thick solid lines* indicate state boundaries; thin solid lines indicate rivers *a* is the Clark Fork River, *b* is the St. Joe River, *c* is the North Fork of the Clearwater River, *d* is the Lochsa River. The bacular morphs for each subspecies are shown in the *same shade* as the subspecies distribution and are drawn on the same scale. Numbers refer to collection localities, referenced in "Appendix 1". Dashed lines delimit the five demes assessed using STRUCTURE

Hypothesis 2 *T. ruficaudus* has experienced a recent northward expansion. Based on geologic history, we hypothesize this occurred in the last 10,000 years and this will be analyzed with a coalescent approach, specific population expansion tests and diversity statistics.

Materials and methods

Sampling and DNA extraction

In total, 306 chipmunks were sampled between 1999 and 2007: 175 *T. r. simulans* individuals from 24 localities and 131 *T. r. ruficaudus* individuals from 18 localities (Fig. 1, "Appendix 1"). Populations 48, 54 and 58 had samples collected on both the north bank and south bank of the Lochsa River (thus 48 N and 48 S, etc.). Genomic DNA was extracted from ear clips (stored in 90% ethanol), livers or kidneys, using either the CTAB/DTAB protocol (Gustincich et al. 1991) or the Animal Tissue protocol with a DNeasy Tissue Kit (Qiagen, Valencia, CA). Animal use protocols were approved by the University of Idaho IACUC (protocol number UIACUC-2005-40).

mtDNA sequencing and analysis

An approximately 800 base pair segment of cytochrome *b* was amplified following the protocols in Good et al. (2003).

Additionally, homologous sequences from previous *T. ruficaudus* studies (Good and Sullivan 2001; Good et al. 2003) were downloaded from GenBank, and all sequences were pruned to the maximum length of overlap, 575 base pairs. The primers were designed specifically for chipmunks (Good and Sullivan 2001) and amplify a fragment that exhibits sufficient variation for intraspecific studies (4.7% uncorrected divergence between subspecies, Good and Sullivan 2001). PCR products were sequenced on an ABI 3130 and edited and aligned using SEQUENCHER (Gene Codes Corp., Ann Arbor, MI). The complete mtDNA dataset contained 266 of the 306 *T. ruficaudus* individuals, including at least one individual from each population, and five outgroup individuals. To ease computational load, we condensed redundant sequences using MACCLADE (v. 4.06; Maddison and Maddison, 2003), which resulted in 61 unique haplotypes. DT-MODSEL was used (Minin et al. 2003) to select the simplest model that is expected to perform well in this phylogeny estimation (Sullivan and Joyce 2005), and PAUP* 4.0 (Swofford 2002) was used to conduct an iterative maximum-likelihood (ML) search following Sullivan et al. (2005). Nodal support was evaluated via 500 bootstrap replicates (Felsenstein 1985) using the Fast Step search option, with a single tree was held for each replicate (i.e., MAXTREE = 1). In addition, we assessed nodal support using posterior probabilities generated by MRBAYES (Huelsenbeck and Ronquist 2001; Ronquist and Huelsenbeck 2003). We conducted two independent runs of four Markov chains. Convergence was assumed when the standard deviation of split frequencies decreased to 0.01. Our sample frequency was 1,000 and this resulted in 4,441 trees, the first 25% of which were discarded as burn-in. Independent runs were merged and a majority-rule consensus tree was obtained. Haplotypes were assigned following Good et al. (2003).

Genotyping microsatellites

Ten microsatellite loci were amplified using primer pairs (forward and reverse) and PCR protocols from Schulte-Hostedde et al. (2000): EuAmMS26, EuAmMS35, EuAmMS37, EuAmMS41, EuAmMS86, EuAmMS94, EuAmMS108, EuAmMS114, EuAmMS138 and EuAmMS142 (the loci will be referred to by the numerical portion of their names hereafter). The forward primer of each pair was fluorescently labeled using 6-FAM, HEX, NED, PET, TET or VIC [Applied Biosystems, Inc. (ABI)] on the 5′ end for detection on an ABI 3130.

PCR amplifications of 20 µl were performed using 100 µg of genomic DNA, 5 µM of labeled primer, 10 µM unlabeled primer, 1 mM dNTP, 10× PCR buffer (Invitrogen), 1.5 mM $MgCl_2$ (Invitrogen), 0.2U *Taq* polymerase (Invitrogen). PCR protocols consisted of an initial denaturing step of 94°C for 3 min, followed by 32 cycles of 45 s at 94°C, 45 s at appropriate annealing temperature (Schulte-Hostedde et al. 2000), 45 s at 72°C. One and a half microliters of PCR product were added to 10 uL Hi-Di (ABI) and 0.3 µl GeneScan LIZ500 size standard (ABI) and run on an ABI 3130. Alleles were visualized and called using GENEMAPPER (ABI).

Microsatellite diversity

We used GENEPOP 3.4 (Raymond and Rousset 1995) to test Hardy–Weinberg Equilibrium (HWE), and to estimate observed and expected levels of heterozygosity and population differentiation for subspecies and populations. ARLEQUIN v.3.0 (Excoffier et al. 2005) was used to conduct hierarchical analyses of molecular variation (AMOVA) and calculate pairwise F_{ST} (using the Infinite Alleles Model) and R_{ST} (Stepwise Mutation Model); significance was assessed using the permutation test in the program with 1,000 permutations. *F*STAT (Goudet 1995) was used to calculate allelic richness, F_{ST}, gene diversity, F_{IS} and heterozygosity. We used MICRO-CHECKER (van Oosterhout et al. 2004) to check our dataset for null alleles, allelic dropout and scoring errors.

Population structure

STRUCTURE (Pritchard et al. 2000) was used to estimate individual admixture and population assignment without *a priori* assumptions of population subdivision. Under the admixture model, STRUCTURE estimates coancestry coefficients for individuals assigned to each of *k* populations. We ran six replicates with a burn-in of 5.0×10^4 followed by 1.5×10^5 subsequent generations for each value of *k* ranging from one to 12. The upper bound exceeded the values for *k* that contained discernable clusters. The second order rate of change of the likelihood function (or ΔK, Evanno et al. 2005) was used to detect the value of *k* with the strongest signal. However, it's more appropriate to treat *k* as a random variable in a Bayesian framework because this permits inferences of admixture to integrate across uncertainty in estimates of population structure. Therefore, we used the reversible jump MCMC approach implemented in STRUCTURAMA (Huelsenbeck and Andolfatto 2007) to derive a Bayesian estimate of *k* (STRUCTURAMA uses a Dirichlet process prior on *k*). We ran the MCMC for 100,000 generations with a print frequency of 25 generations, leading to 4,000 samples, the first 200 of which were discarded as burn-in. We assessed the influence of the prior distributions of *k* by varying the expected prior number of populations from one to eight (Table 3). This led to a variety of shapes of the prior distribution. In addition, we used STRUCTURAMA to calculate marginal likelihoods where the number of populations is fixed (instead of being a

random variable); we ran $k = 1$ to $k = 8$ with the same MCMC run parameters as above.

We then used BayesAss (Wilson and Rannala 2003) for assignment because it does not assume HWE within samples. It can identify migrants and F1 hybrids, assess recent migration rates and corroborate assignment of individuals made by the above programs. Data were partitioned by subspecies and run for 3×10^6 iterations. An individual was considered assigned if the probability of assignment was greater than 80%. In order to test for recent hybridization explicitly, we used NewHybrids (Anderson and Thompson 2002) to estimate posterior probabilities for each individual being pure parentals, F1, F2 and backcrossed genotypes, without *a priori* population assignment. We used NewHybrids' default genotype frequency classes that constitute a pairwise test (since there are at most two pure parentals), so we partitioned our data three ways: (1) all individuals together, (2) Whitefish contact zone and (3) Lochsa contact zone. We ran two replicate analyses of each dataset which consisted of a burn-in of 1×10^4 followed by 1×10^4 generations, as recommended by Anderson and Thompson (2002). A single long run with a burn-in of 2.5×10^4 followed by 1×10^5 generations was done to corroborate results. An individual was considered assigned if the probability of a single frequency class exceeded 80%.

We used the program Populations (Langella 2002) to construct neighbor-joining trees based on Cavalli-Sforza and Edwards chord distance (Dc) (Cavalli-Sforza and Edwards 1967) for both the individual populations and demes (hereafter, *demes* refers to the most probable and significant genetic clusters we detected with the Structure analyses; see *Results—Population Structure* below). To assess support, we conducted 1,000 bootstrap replicates over loci.

Coalescent analysis

We used the coalescent-based program IM (Nielsen and Wakeley 2001; Hey and Nielsen 2004) to derive a model-based estimate of gene flow on our mtDNA dataset. The model included the following parameters: θ for *T. r. simulans* (θ_{TRS}), *T. r. ruficaudus* (θ_{TRR}), and the ancestral *T. ruficaudus* population (θ_{TR}), migration from *T. r. simulans* into *T. r. ruficaudus* (m_{TRR}), and from *T. r. ruficaudus* into *T. r. simulans* (m_{TRS}) and time since divergence (t_{div}). In order to help distinguish between hybridization and lineage sorting, we also recorded the distribution of the number (N_{mig}) and timing (t_{mig}) of migration events occurring over the course of the MCMC simulation (Won and Hey 2005). A more recent t_{mig} than t_{div} would support hybridization whereas a t_{mig} older than t_{div} would support lineage sorting.

Upper bounds for priors were estimated through a series of preliminary runs. A burn-in of 6×10^5 was followed by at least 5.0×10^6 additional generations. We partitioned the data by contact zone (161 sequences for the Lochsa contact zone, 90 sequences for the Whitefish contact zone) and performed three independent replicates per dataset. Similarity of posterior distributions and effective sample sizes (ESS) were used to infer convergence (Hey 2005). The parameters θ, m, t_{div} and t_{mig} were translated into effective populations size (N_{eTRS} and N_{eTRR}), migration rate per year (M) and years since divergence and mean migration event (T_{div} and T_{mig}), respectively, using a mitochondrial mutation rate of 0.5%/MYR (Harrison et al. 2003), which has been previously used for mitochondrial studies in ground squirrels (as there is no estimate for mtDNA mutation rate in *Tamias*. This estimate was converted to a per locus mutation rate based on the size of the fragment we amplified; therefore, we used a mutation rate of 3.4×10^{-6} substitutions/locus/year for the Lochsa dataset and 2.8×10^{-6} substitutions/locus/year for the Whitefish dataset.

Population expansion

To test for a population expansion signature in the ten microsatellite loci, we used the Microsoft Excel macro kgtests (Bilgin 2007) to perform a within-locus k-test and an interlocus g-test (Reich and Goldstein 1998). The k-test analyzes distribution of allele length with the assumption that the gene genealogies of a constant sized population are the result of an ancient bifurcation and the allele distribution is bimodal, with few intermediate allele sizes. A population that has recently expanded has gene genealogies that are the result of variously aged bifurcations and the allele distributions are smoother. The g-test assesses the variance of the variance of allele distributions; it assumes that populations of constant size have genes with ancient bifurcations that lead to a large variance in allele frequencies, whereas populations of recent expansion have more tightly clustered bifurcations and therefore have a lower variance.

Results

mtDNA distribution

The mtDNA analysis was done to increase sampling across the species range. Results were very similar to previously published mtDNA studies (Good and Sullivan 2001; Good et al. 2003). There are two major mtDNA clades, an eastern and a western clade (Fig. 2). Only *T. r. simulans* individuals have the western mtDNA and all *T. r. ruficaudus*

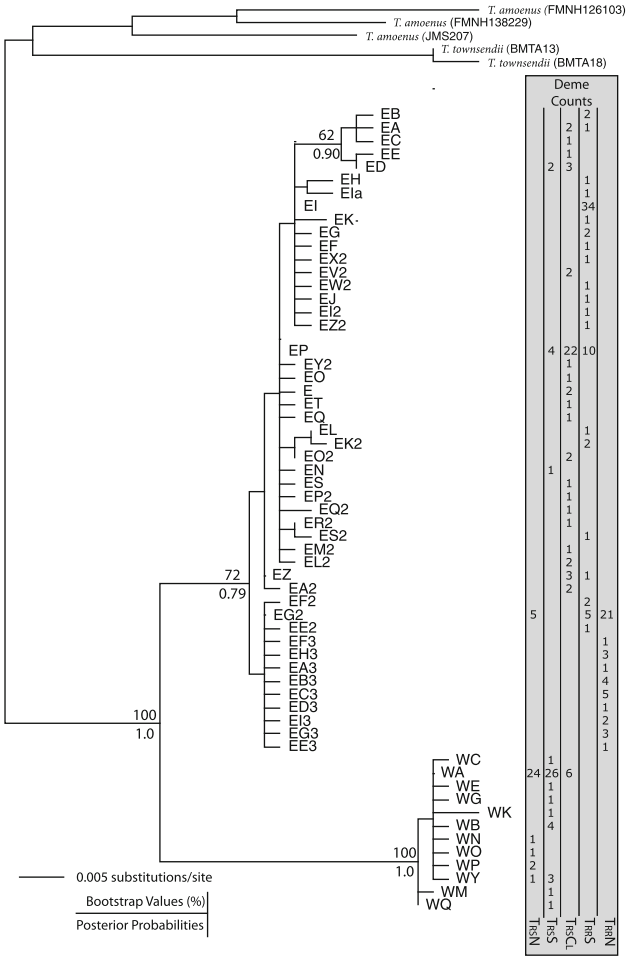

Fig. 2 Maximum likelihood estimate (model: HKY + I) of phylogeny of cytochrome b (575 bp) for 61 unique haplotypes derived from 271 individuals. *Letter/number combinations* correspond to unique haplotypes and the number of individuals belonging to each and their respective demes are shown. *Values above nodes* are bootstrap percentages, *values beneath the node* are posterior probabilities. Names that begin with *E* belong to the eastern clade; *W* is the western clade

individuals have the eastern haplotype. At both contact zones, there are *T. r. simulans* individuals with eastern haplotypes. There appears to be geographically correlated substructure within the eastern clade.

Microsatellite diversity

All ten microsatellite loci were polymorphic within populations and subspecies. Significant departures from HWE occurred in 13 of 20 exact tests (using GENEPOP) when subspecies were analyzed; eight of 40 exact tests were significant within the five demes. When the 42 populations were analyzed separately, nine of 352 tests failed to reject the null hypothesis of HWE. A global test (Fisher's Method) for linkage disequilibrium within subspecies rejected the null hypothesis of genotypes being independent across loci ($P < 0.05$) for seven out of 45 tests. The null hypotheses of identical allelic and genotypic frequencies were rejected ($P < 0.001$) at all ten loci between subspecies and demes.

Table 1 Microsatellite diversity statistics for the five demes (TrsN, TrsS, TrsCl, TrrS, TrrN) and two subspecies (TRS, TRR), including F_{ST}, F_{IS}, observed heterozygosity (H_O), number of alleles corrected for sample size (A*) and the associated *P*-values (*P*)

	F_{ST}	F_{IS}	H_O	A*
TrsN	0.064	−0.019	0.660	1.651
TrsS	0.030	0.016	0.761	1.755
TrsCl	0.007	0.053	0.777	1.821
TrrS	0.056	0.016	0.745	1.766
TrrN	0.085	−0.015	0.606	1.644
P	0.74	0.528	0.004	0.002
TRS	0.092	0.019	0.73	1.745
TRR	0.105	0.005	0.696	1.732
P	0.74	0.62	0.221	0.65

Tests for F_{IS}, heterozygosity, allelic richness and genetic diversity failed to be significant when subspecies were grouped together (using F_{STAT}, Table 1). When data were partitioned by demes, F_{IS} ranged from −0.019 to 0.053 but failed to be significant ($P = 0.52$). Heterozygosity ranged from 0.606 to 0.777 ($P = 0.004$). Total number of alleles ranged from 4 to 24 and allelic richness ranged from 1.644 to 1.821 ($P = 0.002$). Gene diversity (*Hs*) ranged from 0.597 to 0.82 ($P = 0.0009$). Across all measures (except F_{ST}) deme TrsCl had the highest values and demes TrsN and TrrN had the lowest (Table 1).

In the AMOVAs, when analyzing demes as groups of populations, 76.99% of the variation existed within populations, 5.33% existed among populations within demes and 17.69% existed between demes. When we analyzed subspecies as groups of demes, 86.20% of the variation existed within demes, 8.56% existed among demes within subspecies and 5.24% existed between subspecies. F_{ST} and R_{ST} measure genetic divergence among subpopulations using the Infinite Alleles Model and the Stepwise Mutation Model, respectively. F_{ST} was 0.089 and R_{ST} was 0.117 between the subspecies; both of these estimates were significantly different from zero ($P < 0.001$). Assessing the five demes in a pairwise manner, F_{ST} ranged from 0.045 to 0.212 and R_{ST} ranged from 0.033 to 0.331 (Table 2). Again, every pairwise estimate of F_{ST} and R_{ST} was significantly different from zero ($P < 0.001$).

Population structure

Averaged across six independent replicates, the log likelihood values from Structure increased from $k = 1$ to

Table 2 Population differentiation between demes. *FST* is above the diagonal, *RST* is below

	TrsN	TrsS	TrsCl	TrrS	TrrN
TrsN	–	0.099	0.116	0.144	0.212
TrsS	0.097	–	0.045	0.114	0.205
TrsCl	0.314	0.129	–	0.662	0.158
TrrS	0.094	0.132	0.260	–	0.090
TrrN	0.157	0.226	0.331	0.033	–

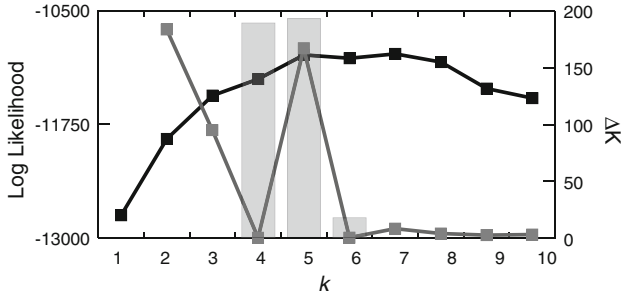

Fig. 3 The three methods used for assessing the most likely number of demes. The *black line* is the average log likelihood averaged over six STRUCTURE runs. The *gray line* is ΔK, a metric for how much better *k* is compared to *k*-1. The *gray bars* represent the posterior probability of *k*, when the expected prior number of populations is two

$k = 5$, decreased slightly at $k = 6$ and increased to the highest value at $k = 7$ (Fig. 3). The variance increased with every increase in *k*. The values for $k = 2$ and $k = 5$ had the two greatest ΔK scores (Fig. 3), indicating the greatest increase in fit from the previous *k*. We used STRUCTURAMA to evaluate the data in two ways. First, we evaluated the marginal likelihood of each value of *k* from one to eight. The likelihood increased substantially from $k = 1$ to $k = 5$, after which the values plateau (P[X|K], Table 3). We also treated the number of populations as a random variable and assessed mean partition of number of populations over a variety of values for α (Table 3). $k = 3$ had the highest posterior probability when expected prior number of populations was one; $k = 4$ was highest when expected prior number of populations was three and four; $k = 5$ was highest when expected prior number of populations was two, five, six, seven and eight.

The coancestry assigned by STRUCTURE at $k = 2$ corresponds broadly to subspecies; 242 of the 306 individuals had a coancestry coefficient of >80%. The patterns within $k = 3$ and $k = 4$ were subsets of the pattern at $k = 5$, which divides the individuals into a northern *T. r. simulans* group (TrsN), a southern *T. r. simulans* (TrsS), a *T. r. simulans* group restricted to the Lochsa River mtDNA introgression zone (TrsCl—for *T. r. simulans* Clearwater),

Table 3 Prior and posterior probabilities for *k* in the program STRUCTURAMA using a Dirichlet process prior and a range of values for α

EPNP	1 (α = 0.0100)		2 (α = 0.1651)		3 (α = 0.3421)		4 (α = 0.5301)		5 (α = 0.7281)		6 (α = 0.9355)		7 (α = 1.1517)		8 (α = 1.3764)	
	Pr[K]	Pr[K|X]	Pr[K]	Pr[K|X]	Pr[K]	Pr[K|X]	Pr[K]	Pr[K|X]	Pr[K]	Pr[K|X]	Pr[K]	Pr[K|X]	Pr[K]	Pr[K|X]	Pr[K]	Pr[K|X]
k																
1	0.9390	0	0.3608	0	0.1259	0	0.0427	0	0.0142	0	0.0046	0	0.0015	0	0.0005	0
2	0.0592	0	0.3753	0	0.2714	0	0.1427	0	0.0650	0	0.0271	0	0.0107	0	0.0040	0
3	0.0018	**0.8001**	0.1871	0	0.2803	0.0692	0.2284	0.0405	0.1429	0.0234	0.0767	0.0158	0.0371	0.0103	0.0167	0.0084
4	0	0.1978	0.0599	0.4728	0.1860	**0.6106**	0.2348	**0.5370**	0.2017	0.4494	0.1391	0.3991	0.0830	0.3473	0.0446	0.2770
5	0	0.0021	0.0139	**0.4830**	0.0896	0.3015	0.1751	0.3894	0.2067	**0.4730**	0.1832	**0.5096**	0.1345	**0.5433**	0.0863	**0.5798**
6	0	0	0.0025	0.0431	0.0335	0.0184	0.1014	0.0324	0.1644	0.0518	0.1872	0.0737	0.1692	0.0926	0.1298	0.1289
7	0	0	0.0004	0.0011	0.0101	0.0003	0.0476	0.0008	0.1060	0.0024	0.1550	0.0018	0.1725	0.0066	0.1581	0.0058
E[K|X]	3.202		4.572		4.269		4.415		4.560		4.646		4.738		4.846	
V[K|X]	0.1659		0.3381		0.3747		0.3944		0.412		0.4195		0.4394		0.4403	
P[X|K]	−13,223		−12,603		−12,327		−12,254		−12,188		−12,180		−12,189		−12,186	

EPNP expected prior number of populations, *k* number of populations. α = shape parameter of Dirichlet process prior, *Pr[K]* prior probability for k, *Pr[K|X]* posterior probability for *k*, *E[K|X]* mean partition for number of populations, *V[K|X]* variance, *P[X|K]* marginal likelihood for data given an integer value for *k* equal to the EPNP of the column (in natural log units)

Highest posterior probability for each EPNP is in bold

a southern *T. r. ruficaudus* (TrrS) and a northern *T. r. ruficaudus* (TrrN, Fig. 1). The results for $k = 7$ are similar to $k = 5$ with the addition of a partition consisting of two *T. r. simulans* individuals from populations 61 and 53 and TrrS splitting into two groups. Although $k = 7$ had the greatest log likelihood, the Structure documentation advises the true value of k is the smallest value for which the likelihood plot "more-or-less plateaus" (Pritchard et al. 2000), which could reasonably be inferred at $k = 5$. Furthermore, $k = 5$ was supported by the ΔK measure, the marginal likelihood measure of Structurama, as well as five of the eight random variable treatments (Table 3, Fig. 3). Therefore, we believe the data indicate five genetic demes within *T. ruficaudus*.

Various methods were used to assess relatedness of individuals and demes and to identify possible hybrid individuals. According to the assignment tests conducted using BayesAss, four individuals had <80% probability of being correctly assigned to their subspecies (as defined by reproductive morphology). One TrsCl individual had 38.7% hybrid classification. Three TrrS individuals had 46.9, 73.6 and 79.6%. The 79.6% individual also had a 16.7% migrant rating, reducing its "pure" percentage to 3.7%. The unrooted neighbor-joining trees from Populations had little support when populations were analyzed (no bootstrap values >75%), but there was moderate support among demes: 65% of the replicates support a TrsN/TrsS bipartition and 98% of the replicates support TrrS/TrrN (Fig. 4). The TrsCl individuals were placed between these two clusters in both population and deme analyses (although this placement does not have a high bootstrap value). Using NewHybrids, at the Whitefish contact zone (demes TrsN and TrrN), most TrsN individuals constituted one of the pure parental classes and TrrN constituted the other. The exceptions were two individuals from population 12 and most individuals from population 68, which were a mix of F2 and backcrossed genotype classes. Across the Lochsa, demes TrsS and TrsCl had all assigned individuals as F2 or backcrossed, however most failed to be assigned. Deme TrrS contained mostly pure individuals with >80% probability. The analysis for all individuals together resulted in TrsN comprising a pure parental class, TrrS and TrrN comprising a second pure parental class and TrsS and TrsCl being predominantly F2 s or backcrossed individuals.

Coalescent estimates

We used IM to estimate θ, m and t_{div} across three independent runs that were allowed to run for at least 5×10^6 iterations prior to assessment of convergence. Posterior distributions were very similar and ESS values exceeded 100 for every parameter. Parameter estimates were averaged across the three runs. All estimates of Ne (for each subspecies and the ancestral *T. ruficaudus* population) varied substantially between the two datasets. The estimates for t_{div} also differed between the two datasets (1.11 for the Lochsa River dataset and 1.62 for the Whitefish dataset). The value of t_{mig} were much more recent than t_{div} in all cases and for the Whitefish dataset, the scaled demographic estimates for mean timing of migration were within the last 10,000 years. Parameter and demographic estimates are given in full in Table 4.

Population expansion

We analyzed our data partitioned by demes for two reasons. First, kgtests assumes that there is no substructure (which is what our demes represent). Second, the program also assumes the data is evolving in a stepwise fashion. We used the allele size permutation test (Hardy et al. 2003) in the program SPAGeDi (Hardy and Vekemans 2002) to test whether our data meet this assumption; it does ($P = 0.008$). When we separated the data by their demes, the only significant result was the k-test of TrsCl ($P = 0.044$). Non-significant results are shown on Table 5.

Discussion

Hybrid zones can directly address the importance of reproductive isolation in speciation. Studies of naturally occurring hybrid zones are particularly relevant between subspecies that are morphologically distinct at a reproductively important character but are ecologically indistinct and have experienced at least some recent gene flow. The mtDNA introgressions within *T. ruficaudus* at both the Whitefish and Lochsa contact zones (Good and Sullivan 2001) were the basis for our hypotheses of gene flow at other loci within the species. The individual markers, pelage (Howell 1922), bacular morphology (Good et al. 2003), mtDNA (Good and Sullivan 2001; Good et al. 2003) and now, nuclear microsatellites, analyzed separately have supported different hypotheses; yet as we sample increasingly larger portions of the genome, a more comprehensive speciation-with-gene-flow model is supported.

Population structure

Several lines of evidence indicate significant substructure within *T. ruficaudus*. The analyses using Structure indicated five discrete populations (Fig. 5). These genetic clusters are generally separated by rivers: the Pend Oreille River and Clark Fork River separate TrsN and TrsS; the North Fork of the Clearwater and St. Joe Rivers delimit

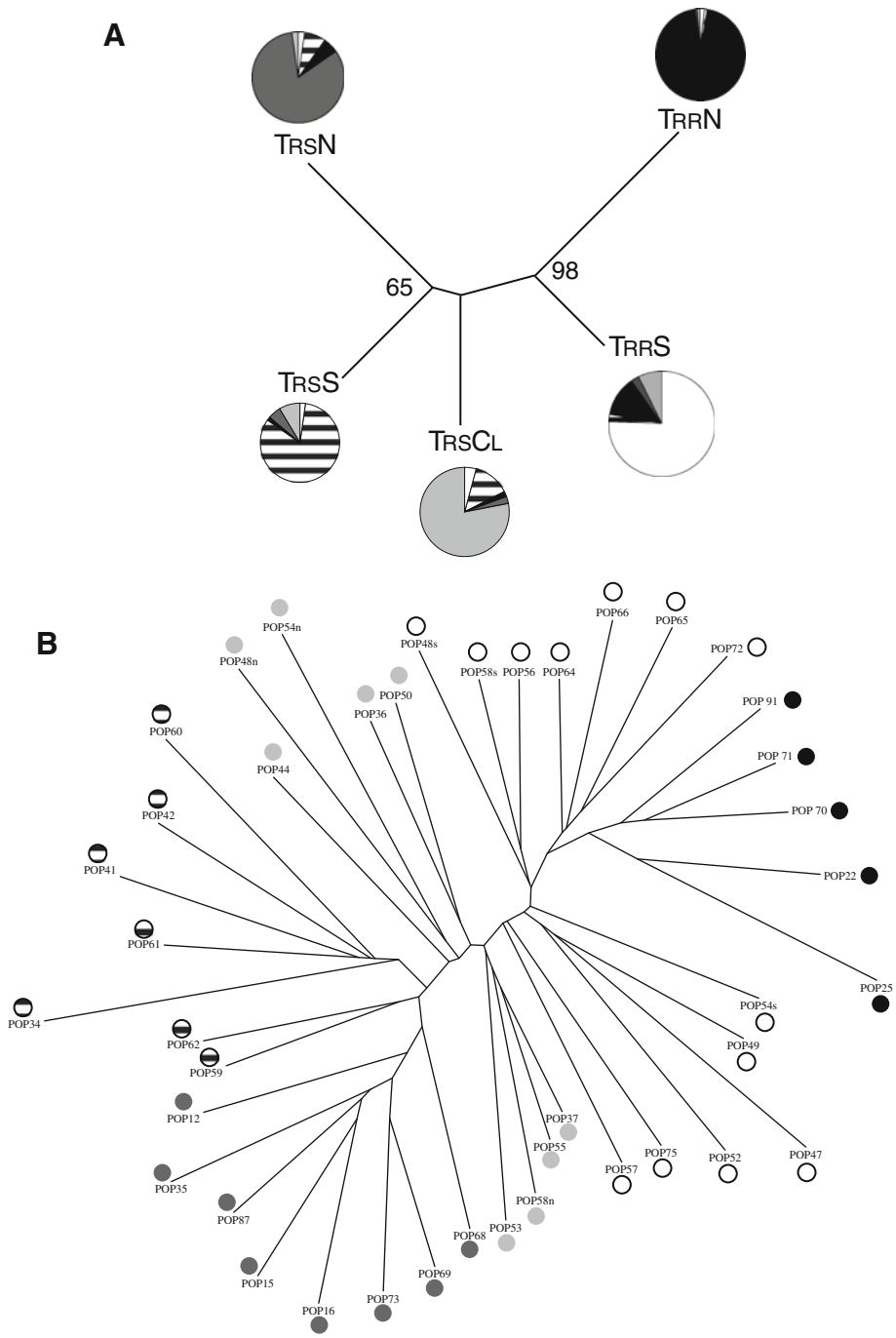

Fig. 4 Unrooted neighbor-joining trees based on genetic distances of the microsatellite data. **a** Operational taxonomic units for the analysis are the demes. Pie charts (representing the demes and scaled to size) show the average coancestry coefficients for all individuals in a deme ($k = 5$, using structure). Bootstrap values greater than 50% shown. **b** Operational taxonomic units for the analysis are the sampling localities; colors represent which deme the sampling locality belongs to [TrsN (*dark gray*), TrsS (*horizontal bars*), TrsCl (*light gray*), TrrS (*white*), TrrN (*black*)]

TrsS and TrsCl; the Lochsa River separates TrsCl and TrrS. Across a range of priors, Structurama placed the majority of the posterior probability as supporting five populations (Table 3). The lack of support for $k = 1$ and $k = 2$, even when the expected prior number of populations is set to one or two, indicate strong signal for substructure. Furthermore, F_{ST} and R_{ST} were significantly different from zero when the demes were grouped, indicating significant population differentiation (Table 2).

Incomplete reproductive isolation

Assessing reproductive isolation with assignment tests may be done in a number of ways. As we did not know the extent of nuclear hybridization in our data, we used Structure and initially assessed $k = 2$ (for the subspecies). This provided the same utility for our data as assigning parentals *a priori*: pure parental populations were detected and these populations were geographically correlated (Fig. 6). To identify a

Table 4 Raw and demographic parameter estimates from IM, with associated 90% highest posterior density intervals (HDP90)

	Lochsa contact zone			Whitefish contact zone		
	HDP90Lo	HiPt	HDP90Hi	HDP90Lo	HiPt	HDP90Hi
θ_{TRS}	27.4564	42.1224	68.5858	1.1782	3.6232	10.1598
θ_{TRR}	9.005	16.6775	30.7325	3.495	8.0796	18.7668
θ_{TR}	21.9821	38.5069	83.8969	11.683	64.1088	145.9639
t_{div}	0.635	1.112	9.735	1.225	1.625	48.625
m_{TRS}	0.0033	0.0036	0.2763	0.0208	0.1575	1.4425
m_{TRR}	0.0464	0.307	0.9996	0.0025	0.0025	0.4542
t_{mig}TRS		0.135			0.025	
t_{mig}TRR		0.165			0.025	
N_{ef} TRS	4,025,865	6,176,311	10,056,575	172,752	531,266	1,489,712
N_{ef}TRR	1,320,384	2,445,387	4,506,232	512,468	1,184,687	2,751,730
N_{ef}TR	3,223,182	5,646,167	12,301,598	1,713,050	9,400,117	21,402,331
T_{div}	186,217	326,100	2,854,839	359,238	476,540	14,259,531
M_{TRS}		0.303			1.141	
M_{TRR}		10.241			0.04	
T_{mig}TRS		39,589			7,331	
T_{mig}TRR		48,387			7,331	

N_{ef} = effective female population size, T_{div} = divergence time in years from present, M = migration rate per year, T_{mig} = mean timing of migration events

Table 5 Within-locus (k) and interlocus (g) tests for population expansion

Demes	k	g
TrsN	0.3406	0.4340
TrsS	0.5856	0.3903
TrsCl	0.0449*	0.5813
TrrS	0.3406	0.6041
TrrN	0.1483	1.5266

Values from k-tests represent P-values; g-tests represent calculated test statistics. Significant values are denoted by asterisks

hybrid zone, we considered any population that contained a hybrid individual to be part of the hybrid zone. An individual was considered a hybrid if the coancestry coefficient met a certain threshold. The hybrid zones were large when a high level of coancestry was required to be considered pure (90% coancestry, Fig. 6a) and decreased to areas closer to the morphological contact zones (which are step clines) when we relaxed the coancestry criterion for individuals to be considered pure (i.e., to 80% then 60%; Fig. 6b, c). In other words, the more stringent our criteria for assigning pure parental individuals, the fewer individuals met that criteria and the hybrid zones increased in area. Additionally, across all three treatments, the hybrid zones were centered on the morphological contact zones; as the hybrid zones grew, they grew on both sides of the contact zones, in most cases. This analysis supports bidirectional gene flow, since the hybrid zones extend into both subspecies ranges at both contact zones and persist even at the lowest stringency for pure status. This finding also supports differential introgression at the two contact zones: the Lochsa hybrid zone is much larger than the Whitefish hybrid zone, which is what one would expect if the Lochsa contact zone is older. An older contact zone has had more time for hybridization events and the spread of new alleles to occur, whereas a younger contact zone has had relatively less time. It should be noted, that age alone may not explain the pattern we see in the size of the two hybrid zones. Available habitat for *T. ruficaudus* to expand may have influenced the current distribution of hybrids; there is less mesic forest habitat near the Whitefish contact zone than in the Lochsa contact zone. However, *T. ruficaudus* has not expanded to fill all available habitat, so that argues for the importance of age of the contact zones for the differential size of the hybrid zones. Second, our sampling in the Whitefish zone is more sparse than the Lochsa zone. The difference in sampling may make it harder to detail the difference in the sizes of the hybrid zones. Our sampling scheme, despite its inconsistencies, provides a reasonable representation of the distribution of haplotypes.

Our analysis using NewHybrids detected numerous individuals of various hybrid classes, although no individuals were assigned as an F1 with high probability. Across the Whitefish contact zone, hybrid individuals occur only in population 68, which is where mtDNA has introgressed. Conversely, across the Lochsa contact zone, both TrsS and TrsCl are largely comprised of F2,

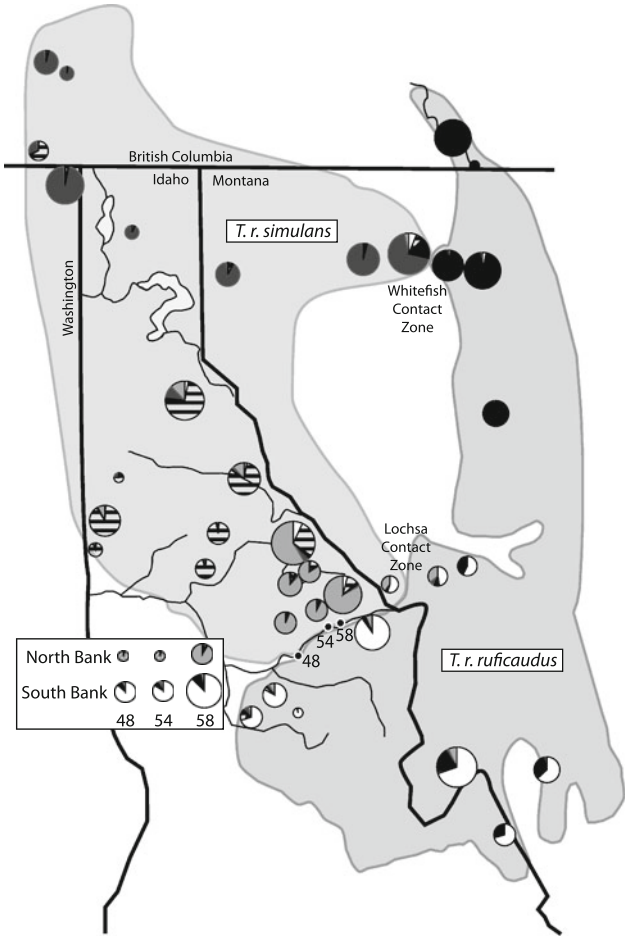

Fig. 5 Mapped coancestry coefficients. *Pie charts* represent the proportion of coancestry to the five demes ($k = 5$, using STRUCTURE) averaged over all individuals in a sampling locality. Pie charts scaled to sample size; colors represent the demes [TRSN (*dark gray*), TRSS (*horizontal bars*), TRSCL (*light gray*), TRRS (*white*), TRRN (*black*)]

backcrossed or unassigned individuals, to the exclusion of a second pure parental class. This would appear to contradict the hypothesis that TRSCL is a hybrid of TRSS and TRRS, but, the unrooted NJ tree based on genetic distance places TRSCL between TRSS and TRRS and places TRSS between TRSN and TRSCL. The complete lack of F1s can be taken as evidence that contemporary hybridizations are rare, but the effects of recent hybridization are still detectable.

Population expansion

The KGTESTS tests detect expansion for only one deme using the within-locus *k*-test and the interlocus *g*-test. The only significant result was the *k*-test for TRSCL where the allele distributions are smoother than expected under a constant sized population (*k*-test), but the variance of the variance in these allele distributions are not sufficiently low enough to indicate expansion (*g*-test). Although this indicates the hybrid zone may have recently undergone or is currently undergoing an expansion, these results are not particularly strong. However, the reduced diversity of mtDNA haplotypes and the reduced levels of coancestry at the microsatellite loci indicate recent expansion. Additionally, the estimates of F_{ST} (0.21) and R_{ST} (0.15) between TRSN and TRRN indicate these two demes are the most differentiated and most recently exposed to secondary contact. Interpreted with the geologic history of the Inland Northwest, wherein a Clearwater refugium south of the Cordilleran ice sheet may have harbored diversity that expanded after glacial recession, the evidence strongly supports a recent northward expansion.

The pattern of mtDNA variation also supports a rapid northward expansion (Good and Sullivan 2001). The TRSN and TRRN demes contain only two mtDNA haplogroups and the majority of individuals belong to a single haplotype. Unique, single mutation haplotypes occurring at low frequency is coincident with a leading edge expansion; populations expand quickly and a single haplotype becomes widespread and abundant and because the expansion was so recent, only single mutations have occurred and these have not had time to spread or differentiate.

Coalescent estimates

At the Whitefish contact zone, the migration rate into *T. r. simulans* is greater than one individual per generation and N_{mig} is two. Into *T. r. ruficaudus*, the migration rate is effectively zero and N_{mig} is zero. This strongly supports unidirectional migration that is further supported by the mtDNA distribution (i.e. two strongly diverged mtDNA haplogroups and a single introgressed population). For both subspecies, T_{mig} is within the last 10,000 years, which is what we expect, given the history of the species range and evidence for a recent northward expansion following ice sheet recession. At the Lochsa contact zone, the migration patterns are the inverse of the Whitefish contact zone. Migration into *T. r. ruficaudus* is greater than ten individuals per generation and N_{mig} is five. Into *T. r. simulans*, migration is almost zero per generation and N_{mig} is zero. This seems counterintuitive, given the direction of the mitochondrial introgression. Based on the mtDNA phylogeny, there must have been at least one migration event into *T. r. simulans*. However, since the Lochsa contact zone has experienced a complex natural history, including episodic vicariance and contact between the two subspecies, the signal for unidirectional migration into *T. r. ruficaudus* may be reasonable. Additionally, IM assumes no substructure within the data, an assumption the Lochsa dataset violates. Thus, given the substructure within the eastern mtDNA haplogroup, the migration of individuals may be

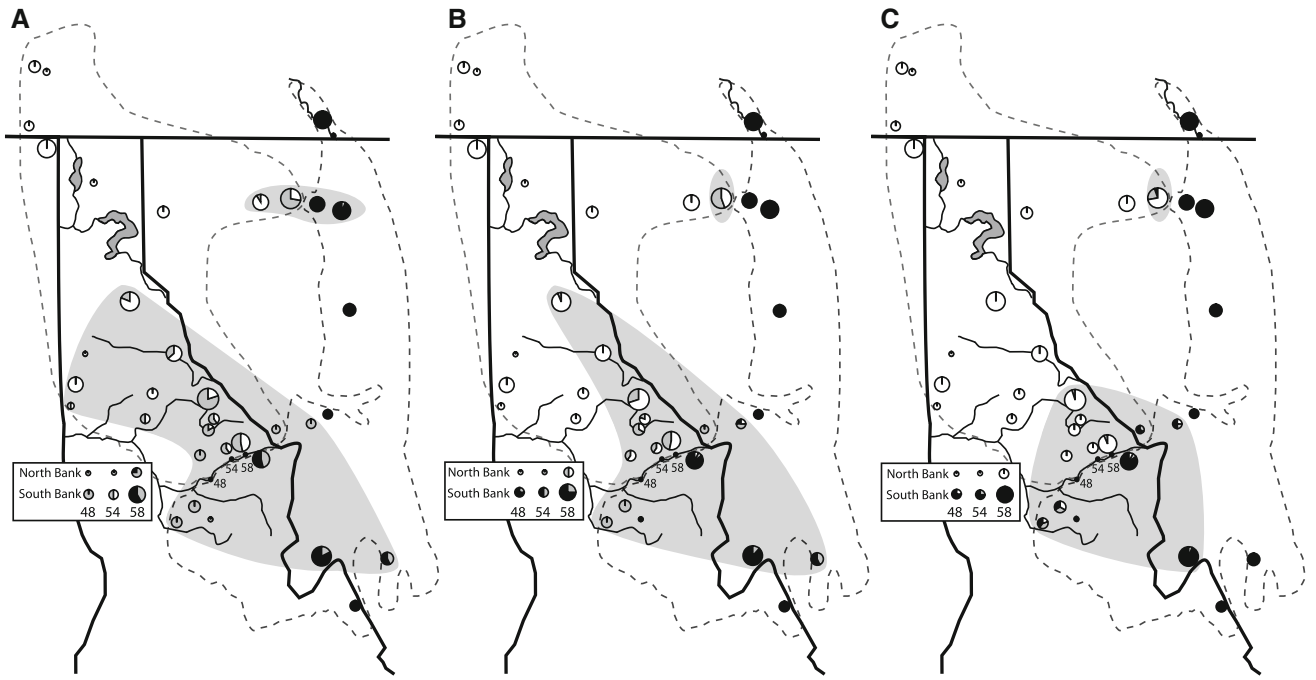

Fig. 6 Analysis of hybrid zones. Each map consists of all individuals within a population shown as a pie chart. The pie charts represent the proportion of individuals that are pure *T. r. ruficaudus* (*black*), pure *T. r. simulans* (*white*) and hybrids (*gray*) according to STRUCTURE analyses at $k = 2$. The levels of coancestry required to be pure is varied across the treatments. **a** Pure individuals exceed 90% coancestry to either subspecies. **b** Pure individuals exceed 80% coancestry. **c** Pure individuals exceed 60% coancestry. Pie charts are scaled to sample size

overwhelmingly into *T. r. ruficaudus*, after divergence in isolation on the *T. r. simulans* side of the Lochsa River. The estimate of t_{mig} is substantially higher at the Lochsa contact zone than the Whitefish zone but still more recent than t_{div}, indicating hybridization.

Finally, our estimate of T_{div} (326,000-476,000 YA, Table 4) is more recent than expected (>1.5 MYA). A study involving a yellow-pine chipmunk subspecies (*T. amoenus canicaudus*) whose mtDNA is nested within *T. ruficaudus* estimated the hybridization event between *T. ruficaudus* and *T. amoenus* to be over 1.5 million YA (Good et al. 2008). This event would need to be more recent than the initial divergence between the *T. ruficaudus* subspecies. There are several possible explanations for the discrepancy in divergence time estimates between these two studies. First, given hybridization, vicariance and contact cycles, the time between initial divergence and final divergence of the *T. ruficaudus* subspecies could have a wide range (the HPDs for all estimates are quite large and extend past the expected time of divergence). Perhaps the time of initial divergence predates the hybridization between *T. ruficaudus* and *T. amoenus*. Second, Wakeley (2000) has shown that substructure within populations has a measurable effect on coalescent estimates of effective population size and divergence times. This could certainly be occurring within our sample, given the amount of nuclear and mitochondrial DNA substructuring. Finally, it is possible that IM is actually estimating a divergence time for a node that is not the split between the eastern and western mtDNA clades, as we assumed. Given the extensive haplotype sharing and large number of private alleles, it may be that IM is estimating a younger node closer to the tips of the tree and not the deepest node in the phylogeny. Regardless of parameter estimates in years, t_{div} substantially precedes t_{mig} at both contact zones, indicating hybridization.

Conclusions

Tamias ruficaudus is comprised of two closely related subspecies each with significant genetic substructure. Given our current sampling and analytical tools it appears that there are three distinct demes in *T. r. simulans* and two in *T. r. ruficaudus* and these five demes are geographically correlated. Previous studies that documented mtDNA introgression at the contact zones were upheld and mtDNA introgression due to hybridization was strongly supported. Additionally, whereas the mtDNA is unidirectionally introgressing, the nuclear genomes are bidirectionally moving across the contact zones. The southern contact zone, at the Lochsa River, has had a longer and more complex history

of isolation and contact and that is shown in the increased genetic diversity and size of the hybrid zone. The northern contact zone, near Whitefish, MT, covered by glaciers until very recently, displays a lack of genetic diversity and a small zone of introgression supporting the hypothesis of a more recently established hybrid zone. So, despite the distinct bacular morphologies, which are thought to indicate reproductive isolation, we have evidence of hybridization at both contact zones. It appears that the supposed morphological barrier to hybridization has been breached repeatedly, yet strictly maintained, an occurrence that may be common across much of *Tamias*. Further study on the genomic scale may elucidate important genetic patterns and mechanisms that underlie these questions and others.

Acknowledgments We thank J. Good and the UI Mammalogy classes (1999–2003) for assistance with field collection. The Field Museum of Natural History, Royal British Columbia Museum, Victoria, Burke Museum of Natural History and Culture and the Connor (Washington State University) provided tissue samples. L. Waits, M. Cantrell, B. Carstens, M. Koopman and T. Pelletier, C. Baers and two anonymous reviewers provided valuable comments on this manuscript. This work was funded by NSF DEB-0717426 (to JS) and DEB-0716200 (to JD). Analyses were run on the bioinformatics core facility supported by the Initiative for Bioinformatics and Evolutionary Studies (IBEST) and funded by NHI (NCRR 1P20RRO16448-01) and NSF (EPS-809935).

Appendix 1

See Table 6.

Table 6 Sampling localities for all individuals; latitude and longitude, number collected (N) and description of localities

Loc	Latitude	Longitude	N	Information
12	49.02 N	−117.43 W	4	British Columbia, Church Creek Road; TRS: RBCM19655, 19666–8 [WI, WA, WI(2)]
15	49.47 N	−117.33 W	6	British Columbia, Giveout Creek FSR KM 9–10; TRS: RBCM19658–62, 20038 [WA(6)]
16	49.439 N	−117.284 W	2	British Columbia, Gold Creek, KM 4.9 Gold Creek RD; TRS: RBCM19654, 20036 [WA(2)]
22	49.25 N	−114.40 W	14	British Columbia, Rainy Ridge Vicinity, Middle Kootenay Pass; TRR: RBCM19875, 19880, 19884–5, 19887, 19906–7, 19914–920 [EC2, EG2, E, EB2(2), EG2(5), EB2, EG2(3), EC2].
25	49.018 N	−114.078 W	1	British Columbia, Source of Akamina Creek, ∼ 30 km W Wall Lake; TRR: RBCM19683 [EG2].
34	47.053 N	−116.87 W	1	Idaho, Benewah Co., 10 mi N Potlatch, Base of Mineral Mt.; TRS: JMG002 [WD]
35	48.513 N	−116.583 W	2	Idaho, Boundary Co., Jeru Creek, 17 mi N. Sandpoint; TRS: JG001, JMS203 [WH, WA].
36	46.642 N	−115.424 W	6	Idaho, Clearwater Co., 0.7 mi E Mouth, Weitas Creek; TRS: JMS151, 154–158 [EP, ER, WA, EP, ET, WA]
37	46.911 N	−115.065 W	20	Idaho, Clearwater Co., 19 mi N Kelly Forks, Rawhide Creek; TRS: JMS144–146, NMR18–26, SMH09–16 [EY2, EO, EV, EQ2, EP(4), WA, EO2, EP, EO2, EP, ED, ET2, EP, EZ, EP (2), WA]
41	46.823 N	−116.177 W	5	Idaho, Clearwater Co., 4 mi N Elk River; TRS: JMS191–195 [WF, WA, EN, WA(2)]
42	46.57 N	−116.270 W	4	Idaho, Clearwater Co., 4.5mi E Teaken, Freeman Creek; TRS: JMS257–260 [WB(4)]
44	46.753 N	−115.369 W	5	Idaho, Clearwater Co., 9mi NW Kelly Forks, Mush Saddle; TRS: JMS238–242 [EP, ES, WA(2), ED]
47	45.840 N	−115.830 W	5	Idaho, Idaho Co., 10.5mi SE Grangeville, FR648A, Cougar Mtn.; TRR:JRD18–22 [EI, E, EI, EX2, EI]
48	46.231 N	−115.412 W	5	Idaho, Idaho Co., 11 mi NE Lowell, Split Creek Pack Bridge; TRS:JMG29 [EA]; TRR: JMS181–2, JMS255–6 [EP, EI, EP, EI]
49	45.968 N	−115.593 W	5	Idaho, Idaho Co., 13 mi S Confluence Lochsa/Selway; TRR: F126113–5, 126123–4 [EH, EI(2), EK, EI]
50	46.439 N	−115.489 W	5	Idaho, Idaho Co., 18 mi E Pierce, Rocky Ridge Lake; TRS: JMS159–163 [EQ, EZ, ER, EP(2)]
52	46.002 N	−115.395 W	1	Idaho, Idaho Co., 25 km N Elk City, FR443,TRR: JMG91 [EJ2]
53	46.510 N	−115.154 W	5	Idaho, Idaho Co., 30 mi E Pierce, Twelve–Mile Saddle; TRS: JMS176–180 [ED, EP(2), EC, EP]
54	46.430 N	−115.129 W	5	Idaho, Idaho Co., 30 mi NE Lowell, Eagle Mt Pack Bridge; TRS:JMS164 [EP]; TRR: JMS165, JMS252–4 [EL, EI(3)]
55	46.540 N	−114.989 W	15	Idaho, Idaho Co., 40 mi E Pierce, Indian Post Office; TRS: JMG153–4, JMS171–175, NMR1–6, SMH1–2 [EM2, EL2, EA2(2), EE, EZ, EA, EP2, EV2, EP(2), ER2, EV2, EP, EU2]
56	46.345 N	−114.640 W	13	Idaho, Idaho Co., Elk Summit/Hoodoo Lake; TRS: JMG74–81, NMR27–9, SMH17–8 [EA, EI, XX, EP, XX(4), EZ, EW2, EP, EI(2), ES2]
57	46.635 N	−114.578 W	3	Idaho, Idaho Co., Lolo Pass/Elk Meadows; TRR: FMNH126126–8 [EF2, EB(2)]
58	46.461 N	−115.018 W	16	Idaho, Idaho Co., Warm Springs Trail Head; TRS: JMG150, 155, 157, JMS298 [EL2, E, W, XX]; TRR: JMG143–44, 46, 48–9, 183–6, 301, 305–6 [EK2, EP, EK2, EI(2), EP(2), EI, EF, XX(3)]

Table 6 continued

Loc	Latitude	Longitude	N	Information
59	47.708 N	−116.377 W	16	Idaho, Kootenai Co., 12.5 mi N Enaville; TRS: JMS200–2, NMR8–17, SMH6–8 [WG, WA(4), WY, WA, WY, WA, WY, EP, WA(2), WM, WN, EP]
60	46.69 N	−116.971 W	2	Idaho, Latah Co., 4 mi SE Moscow, Paradise Ridge; TRS: JMS189–90 [ED(2)]
61	46.808 N	−116.90 W	9	Idaho, Latah Co., Idaho, Latah Co., Moscow Vicinity, 8 mi NE, Moscow Mtn., TRS: JMS122–8, 132–3 [WC, WA(5) WE, XX, WA]
62	47.225 N	−115.606 W	11	Idaho, Shoshone Co., 8.2 mi E Avery; TRS: JMS211–215, 327–8, NMR7, SMH3–5 [EP(3), WA(2), WK, XX, WA(3)]
64	45.730 N	−113.900 W	17	Montana, Beaverhead Co., 8 km E Lost Trail Pass, Trail Creek; TRR: JMG59–64, 66–72, JRD55–8 [EI, XX, EI, XX, EG2(2), EI, E, EF2, EI(2), EJ]
65	45.450 N	−113.080 W	7	Montana, Beaverhead Co., Pioneer Mts., 4mi NW Comet Mtn., FR484; TRR: JMG12–15 [EI(4)], 4 km E Table Mtn.; TRR: JRD145 [EI]; 6 km SE Saddleback Mtn.; TRR: JRD53–54 [EI(2)]
66	45.150 N	−113.500 W	5	Montana, Beaverhead Co., Selway Mtn., Bloody Dick Creek, FR1818; TRR: JRD139–143 [EI, EI2, EI(3)]
68	48.506 N	−114.248 W	18	Montana, Flathead Co., 12 km NNE Whitefish, TRS: JMS225–34, JMG126–33 [WA, EG2, WA(2), EG2, WA, EG2, WA, EG2(2), WA(2), WJ, WK, WJ, WA]
69	48.42 N	−114.825 W	11	Montana, Flathead Co., 35 km W Whitefish, 2 km S Good Cr., TRS: JMG108,13–14,16–19 [WA(7)]
70	48.393 N	−113.945 W	10	Montana, Flathead Co., 8 km E Hungry Horse, Desert Mtn., TRR: JMG134–42,59 [EG2, E02, EG2, E, EG2(5), E]
71	48.28 N	−113.61 W	14	Montana, Flathead, near Essex along Highway 2, TRR: JRD026, NMR40–49, SMH35–7 [ED2]
72	46.586 N	−114.578 W	4	Montana, Granite Co., Brewster Creek, Sliderock Mountain; TRR: FMNH126129–131, 133 [EG2(3), EE2]
73	48.343 N	−115.601 W	6	Montana, Lincoln Co., 4 mi S. Libby, Flower Creek, TRS: JMS199,216–18, SMH19–20 [WA(2), W, WA]
75	46.686 N	−114.077 W	4	Montana, Missoula Co., South Lolo; TRR: FMNH126142–145 [EP, EI, EG(2)]
87	48.85 N	−117.18 W	15	Washington, Pend Oreille Co., Sullivan Creek Drainage, TRS: CMNH83–56;81–574;80–698;80–699;80–701;83–57;98–3910;98–3913;98–3914;98–3915;82–474, JMS205–10 [XX(10), WA(5)]
91	47.404 N	−113.725 W	7	Montana, Missoula Co., Lindbergh Lake; TRR: NMR55–57, SMH39–42 [EH3(3), XX(4)]

Subspecies, sample name and mtDNA haplotypes (in square brackets, correspond to order of individuals listed, XX indicates no mtDNA data collected for corresponding individual) included

References

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

Barton NH, Hewitt GM (1985) Analysis of hybrid zones. Annu Rev Ecol Syst 16:113–148

Bilgin R (2007) KGTESTS: a simple Excel macro program to detect signatures of population expansion using microsatellites. Mol Ecol Notes 7:416–417

Brumfield RT, Jernigan RW, McDonald DB, Braun MJ (2001) Evolutionary implications of divergent clines in an avian (*Manacus*: Aves) hybrid zone. Evolution 55:2070–2087

Carstens BC, Degenhardt JD, Stevenson AL, Sullivan J (2005) Accounting for coalescent stochasticity in testing phylogeographic hypotheses: modeling Pleistocene population structure in the Idaho Giant Salamander *Dicamptodon aterrimus*. Mol Ecol 14:255–265

Cavalli-Sforza LL, Edwards AWF (1967) Phylogenetic analysis: models and estimation procedures. Evolution 21:550–570

Coyne J, Orr HA (2004) Speciation. Sinauer Associates, Sunderland, Massachusetts

Daubenmire R (1952) Plant geography of Idaho. In: Davis RJ (ed) Flora of Idaho. Brigham Young University Press, Provo, Utah, pp 1–17

Delcourt PA, Delcourt HR (1993) Paleoclimates, paleovegetation, and paleoXoras during the late Quaternary. In: Flora of North America Editorial Committee (ed) Flora of North America, vol 1, 1st edn. Oxford University Press, New York, pp 71–94

Detling LE (1968) Historical background of the flora of the Pacific Northwest. Bulletin No. 13. Museum of Natural History, University of Oregon, Eugene

Dobzhansky T (1951) Genetics and the origin of species, 3rd edn. Columbia University Press, New York

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. Mol Ecol 14:2611–2620

Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evol Bioinform 1:47–50

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

Good JM, Sullivan J (2001) Phylogeography of the red-tailed chipmunk (*Tamias ruficaudus*), a northern rocky mountain endemic. Mol Ecol 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

Good J, Hird S, Reid N, Demboski J, Steppan S, Martin-Nims T, Sullivan J (2008) Ancient hybridization and mitochondrial

capture between two distantly related species of chipmunks (*Tamias*: Rodentia). Mol Ecol 17:1313–1327
Goudet J (1995) FSTAT (Version 1.2): a computer program to calculate F-statistics. J Hered 86:485–486
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
Hardy OJ, Vekemans X (2002) SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. Mol Ecol Notes 2:618–620
Hardy OJ, Charbonnel N, Freville H, Heuertz M (2003) Microsatellite allele sizes: a simple test to assess their significance on genetic differentiation. Genetics 163:1467–1482
Harrison R, Bogdanowicz S, Hoffmann R, Yensen E, Sherman P (2003) Phylogeny and evolutionary history of the ground squirrels (Rodentia: Marmotinae). J Mammal Evol 10:249–276
Heller HC (1971) Altitudinal zonation of chipmunks (genus: *Eutamias*): interspecific aggression. Ecology 52:312–319
Heller HC, Gates DM (1971) Altitudinal zonation of chipmunks (genus *Eutamias*): interspecific aggression. Ecology 52:424–433
Hey J (2005) On the number of new world founders: a populations genetic portrait of the peopling of the Americas. PLoS Biol 3(6):e193
Hey J, Nielsen R (2004) Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. Genetics 167:747–760
Hird S, Sullivan J (2009) Assessment of gene flow across a hybrid zone in red-tailed chipmunks (*Tamias ruficaudus*). Mol Ecol 18:3097–3109
Howell AH (1922) Diagnoses of seven new chipmunks of genus *Eutamias*, with a list of the American species. J Mammal 3:178–185
Huelsenbeck J, Andolfatto P (2007) Inference of population structure under a Dirichlet process model. Genetics 175:1787–1802
Huelsenbeck J, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic tress. Bioinformatics 17:754–755
Langella O (2002) Populations 1.2.28: population genetic software (individuals or populations distances, phylogenetic trees). CNRS, France
Mack RN, Rutter NW, Bryant VM Jr, Valastro S (1978) Reexamination of postglacial vegetation history in northern Idaho, Hager Pond, Bonner County. Quat Res 10:241–255
Maddison DR, Maddison WP (2003) MACCLADE. Sinauer & Associates, Sunderland, Massachusetts
Martinsen G, Whitham T, Turek R, Keim P (2001) Hybrid populations selectively filter gene introgression between species. Evolution 55:1325–1335
Mayr E (1942) Systematics and the origin of species. Columbia University Press, New York
Mayr E (1963) Animal species and evolution. Belknap Press, Cambridge
Minin V, Abdo Z, Joyce P, Sullivan J (2003) Performance-based selection of likelihood models for phylogeny estimation. Syst Biol 52:674–683
Nielsen R, Wakeley J (2001) Distinguishing migration from isolation: a Markov chain Monte Carlo approach. Genetics 158:885–896
Nosil P (2008) Speciation with gene flow could be common. Mol Ecol 17:2103–2106
Patterson BD, Heaney LR (1987) Preliminary analysis of geographic variation in red-tailed chipmunks (*Eutamias ruficaudus*). J Mammal 68:782–791
Patterson BD, Thaeler CSJ (1982) The mammalian baculum: hypotheses on the nature of bacular variability. J Mammal 63:1–15
Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959
Raymond M, Rousset F (1995) GENEPOP (version 1.2)—population genetics software for exact tests and ecumenicism. J Hered 86:248–249
Reich DE, Goldstein DB (1998) Genetic evidence for a Paleolithic human population expansion in Africa. Proc Natl Acad Sci U S A 95:8119–8123
Reid N, Hird S, Schulte-Hostedde AI, Sullivan J (2010) Examination of nuclear loci across a zone of mitochondrial introgression between *Tamias ruficaudus* and *Tamias amoenus*. J Mammal (in press)
Rice WR, Hostert EE (1993) Laboratory experiments on speciation: what have we learned in 40 years. Evolution 47:1637–1653
Ronquist F, Huelsenbeck J (2003) MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574
Schulte-Hostedde AI, Gibbs HL, Millar JS (2000) Microsatellite DNA loci suitable for parentage analysis in the yellow-pine chipmunk (*Tamias amoenus*). Mol Ecol 9:2180–2181
Sullivan J, Joyce P (2005) Model selection in phylogenetics. Annu Rev Ecol Evol Syst 36:445–466
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. Mol Biol Evol 22:1386–1392
Swofford DL (2002) PAUP*. phylogenetic analysis using parsimony (*and Other Methods). Sinauer & Associates, Sunderland, Massachusetts
van Oosterhout C, Hutchinson WF, Wills DP, Shipley P (2004) Micro-checker: software for identifying and correcting genotyping errors in microsatellite data. Mol Ecol Notes 4:535–538
Wakeley J (2000) The effects of subdivision on the genetic divergence of populations and species. Evolution 54:1092–1101
White JA (1953) The baculum in the chipmunks of western North America. Univ Kans Mus Nat Hist Publ 5:611–631
Wilson GA, Rannala B (2003) Bayesian inference of recent migration rates using multilocus genotypes. Genetics 163:1177–1191
Won YJ, Hey J (2005) Divergence population genetics of chimpanzees. Mol Biol Evol 22:297–307
Wu C-I (2001) The genic view of the process of speciation. J Evol Biol 14:851–865