Genomic resolution of cryptic species diversity in chipmunks

Nathanael D. Herrera,[1,2] Kayce C. Bell,[3] Colin M. Callahan,[1] Erin Nordquist,[1] Brice A. J. Sarver,[1]
Jack Sullivan,[4,5] John R. Demboski,[6] and Jeffrey M. Good[1,7,8]

[1]Division of Biological Sciences, University of Montana, Missoula, Montana, USA

[2]E-mail: ndh04c@gmail.com

[3]Natural History Museum of Los Angeles County, Los Angeles, California, USA

[4]Department of Biological Sciences, University of Idaho, Moscow, Idaho, USA

[5]Institute for Bioinformatics and Evolutionary Studies (IBEST), University of Idaho, Moscow, Idaho, USA

[6]Department of Zoology, Denver Museum of Nature & Sciences, Denver, Colorado, USA

[7]Wildlife Biology Program, University of Montana, Missoula, Montana, USA

[8]E-mail: Jeffrey.good@umontana.edu

Discovery of cryptic species is essential to understand the process of speciation and assessing the impacts of anthropogenic stressors. Here, we used genomic data to test for cryptic species diversity within an ecologically well-known radiation of North American rodents, western chipmunks (*Tamias*). We assembled a de novo reference genome for a single species (*Tamias minimus*) combined with new and published targeted sequence-capture data for 21,551 autosomal and 493 X-linked loci sampled from 121 individuals spanning 22 species. We identified at least two cryptic lineages corresponding with an isolated subspecies of least chipmunk (*T. minimus grisescens*) and with a restricted subspecies of the yellow-pine chipmunk (*Tamias amoenus cratericus*) known only from around the extensive Craters of the Moon lava flow. Additional population-level sequence data revealed that the so-called Crater chipmunk is a distinct species that is abundant throughout the coniferous forests of southern Idaho. This cryptic lineage does not appear to be most closely related to the ecologically and phenotypically similar yellow-pine chipmunk but does show evidence for recurrent hybridization with this and other species.

KEY WORDS: Hybridization, introgression, *Neotamias*, phylogenomics, speciation, *Tamias*.

Species are fundamental units of biodiversity, yet the basic task of species discovery often remains incomplete, even in well-studied taxonomic groups. Although some of these difficulties stem from philosophical disagreements on species delimitation (de Queiroz 2007), a far more important source of uncertainty involves the misidentification of cryptic lineages that are phenotypically or ecologically similar to known species (Bickford et al. 2007; Struck et al. 2018). Identifying cryptic diversity is fundamental to understanding both the process of speciation and the conservation of species. The emerging threats of accelerated climate change and habitat loss have only increased the urgency of discovering and more fully accounting for patterns of global biodiversity (Bickford et al. 2007; Delic et al. 2017).

Species delimitation historically relied on morphological differences that does not always coincide with genetic patterns of reproductive isolation. Moreover, the evolutionary history of speciation can be further obscured by hybridization and introgression (Lamichhaney et al. 2017; Fišer et al. 2018). Over the past decade, both theoretical and methodological approaches for genetically based species delimitation have seen considerable progress (Yang and Rannala 2010; Fujita et al. 2012; Edwards and Knowles 2014; Mirarab et al. 2014a; Yang 2015; Mirarab et al. 2016; Luo et al. 2018). In conjunction, advances in

high-throughput DNA sequencing provide the ability to reconstruct complex evolutionary histories using sophisticated statistical and phylogenetic approaches that can detect and quantify the extent of gene flow during speciation (Green et al. 2010; Blischak et al. 2018).

Chipmunks of western North America provide a rich system to study speciation across a rapid diversification characterized by ecological adaptation, phenotypic convergence, and overlapping geographic distributions. Of the 26 (see Burgin et al. 2018) currently recognized chipmunk species (Sciuridae: *Tamias*; but see Patterson and Norris 2016), two are widely distributed lineages occupying central and eastern Asia (*Tamias sibiricus*) or eastern North America (*Tamias striatus*; Hall 1981). The remaining 24 species (subgenus *Neotamias*) comprise a recent radiation of habitat specialists and generalists occupying the diverse ecosystems of western North America (Heller 1971; Patterson 1981; Sullivan et al. 2014). Specialization among chipmunks has resulted in strong niche partitioning by habitat and competitive exclusion among co-distributed species (Brown 1971; Heller 1971). Although ecological associations broadly tend to track described species boundaries, these associations can break down in sympatry, with chipmunks that occupy similar or transitory habitats showing phenotypic convergence (e.g., in size and pelage coloration) between nonsister species. As a consequence, internal genital bones often provide the only morphological character that is strongly diagnostic of species boundaries (White 1953). Divergence in genitalia is most pronounced in the male genital bone, the baculum (*os penis*), which evolves rapidly among chipmunk species (Callahan 1977; Sutton and Patterson 2000; Sullivan et al. 2014) likely due to strong sexual selection (Eberhard 1985; Simmons 2014). Bacular divergence was long thought to underlie strong reproductive barriers between chipmunk species (Patterson and Thaeler 1982). However, more recent work has shown that some hybridization occurs between species (see Sullivan et al. 2014 for review). In most cases, levels of nuclear introgression have appeared to be relatively low overall (but see Ji et al. 2021), with the persistence of genetically well-defined species boundaries that largely correspond with established taxonomy based on ecological associations and diagnostic genitalia (Hird and Sullivan 2009; Hird et al. 2010; Good et al. 2015; Bi et al. 2019; Sarver et al. 2021).

Genetic studies have also suggested several potential cases of cryptic diversity in this radiation (Demboski and Sullivan 2003; Reid et al. 2012; Sullivan et al. 2014), and the phylogeny of western chipmunks remains unresolved. One potential source for cryptic diversity lies within the widespread least chipmunk (*Tamias minimus*). For example, *T. m. grisescens* (the Coulee chipmunk, hereafter *grisescens*) is a small, light gray subspecies of *T. minimus* that occupies a restricted
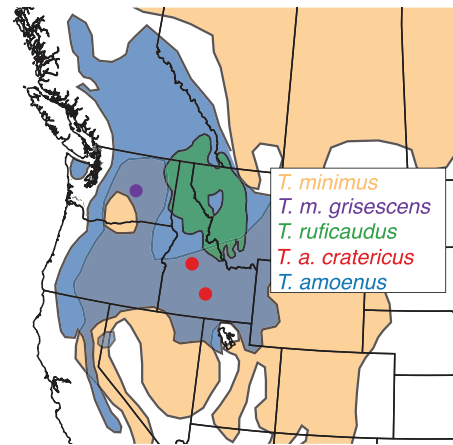

Figure 1. Two potential cryptic species of chipmunk. Ranges of the three described focal species are shown (*T. ruficaudus* [green], *T. amoenus* [blue], and *T. minimus* [orange]). *Tamias minimus grisescens* (purple dot) occupies a restricted range from the Channeled Scablands of central Washington. *Tamias amoenus cratericus* (red dots) is described as a locally adapted form of the yellow-pine chipmunk (*T. amoenus*). See Sullivan et al. (2014) for broader visualization of other western chipmunk species ranges, which show extensive overlap across western North America.

range from the Channeled Scablands of central Washington that was identified by Reid et al. (2012) as a potential cryptic lineage based on four nuclear loci and mitochondrial DNA (mtDNA).

Another potential example of cryptic diversity occurs within the yellow-pine chipmunk (*T. amoenus*), a widespread species associated with xeric forests throughout western North America. Blossom (1937) described a subspecies of *Eutamias amoenus* (i.e., the Crater chipmunk, now *T. amoenus cratericus*, hereafter *cratericus*) as occurring on the periphery of a series of relatively recent lava flows (∼15–2 kya) from Craters of the Moon National Monument and surrounding area of central Idaho. This darker pelage variant of the yellow-pine chipmunk was assumed to reflect local adaptation on the black lava flows of southern Idaho, with rapid transition to more brightly colored *T. amoenus* morphs found in adjacent xeric forest habitats of the region (Fig. 1). However, comparison of bacula by White (1953) and Sutton (1982, including the baubellum, the female genital bone or *os clitoris*) of multiple *Tamias* species concluded that *cratericus* may be more closely related to other chipmunk species (*Tamias umbrinus* or *Tamias ruficaudus*) based on similar genital morphology and nearby geographic proximity. Finally, the mitochondrial genome sequenced from *cratericus* appears to be fairly divergent from other species and has been detected at an additional locality to the north of Craters of the Moon (Demboski and Sullivan 2003), suggesting that *cratericus* may be more widespread (Reid et al. 2012).

Here, we use linked-read sequencing to generate a reference genome for the least chipmunk (*T. minimus*). We then use new and published targeted capture sequencing data from 21,551 autosomal and 493 X-linked loci, and complete mitochondrial genomes to infer the phylogenetic relationships among 22 described chipmunk species and the enigmatic *cratericus* and *griseecens* chipmunks. We then use additional population-level sequencing to characterize the geographic range and evolutionary history of *cratericus* relative to three other potentially co-distributed chipmunk species.

Materials and Methods

SAMPLING AND ETHICS STATEMENT

To assess genus-wide diversity, we combined extensive fieldwork with collections from several natural history museums to obtain samples from 14 western chipmunk species and the eastern chipmunk, *T. striatus*, as an outgroup (Table S1). We combined data from these samples with published data (Bi et al. 2019; Sarver et al. 2021) from seven additional species of western chipmunks resulting in broad sampling covering 22 of 26 described species of chipmunk (57 individuals). The samples that were not included were of *T. solivagus*, *T. durangae*, and *T. bulleri* (all western chipmunk species endemic to Mexico), and the Eurasian *T. sibiricus*. To better characterize the geographic and ecological extent of *cratericus*, we collected an additional 64 individuals from 10 populations throughout central Idaho covering a broad range of habitats from the lava flows of Craters of the Moon up to the subalpine zone. At each sampling locality, we assigned species designations based on the qualitative assessment of bacular morphology when possible and then genetically assigned lineages post hoc based on phylogenetic results. All animals sampled for this project were collected following procedures approved by the University of Montana Animal Care and Use Committee (protocol numbers AUP 042-16JGDBS & 022-10JGDBS). We used acceptable field collecting methods, performed according to the recommendations of the American Society of Mammalogists Animal Care and Use Committee (Sikes et al. 2011).

GENOME SEQUENCING AND ASSEMBLY

Liver tissue from a female *T. minimus* was frozen in liquid nitrogen immediately after euthanasia and sent to the DNA Technologies and Expression Analysis Core at the UC Davis Genome Center for DNA extraction and library preparation. High molecular weight genomic DNA was isolated following Jain et al. (2018). Briefly, 40 mg of flash frozen liver tissue was homogenized by grinding in liquid nitrogen and lysed with 2 ml of lysis buffer containing 10 mM Tris-HCl pH 8.0, 25 mM EDTA, 0.5% (w/v) SDS, and 100 μg/ml proteinase K. The lysate was cleaned with equal volumes of phenol/chloroform using phase lock gels (Quantabio Cat # 2302830). The DNA was precipitated by adding 0.4× volume of 5 M ammonium acetate and 3× volume of ice-cold ethanol, washed twice with 70% ethanol, and resuspended in buffer (10 mM Tris, pH 8.0). The integrity of the high molecular weight genomic DNA was verified on a Pippin Pulse gel electrophoresis system (Sage Sciences, Beverly, MA). Following DNA isolation, a 10× Genomics Chromium library was prepared following the manufacturer's protocol and sequenced on the Illumina HiSeq X platform by Novogene Corporation. We then used Supernova V2.0.0 (Weisenfeld et al. 2017) to generate a de novo genome assembly from the 10× Genomics Chromium library sequence data using default settings with –maxreads = 1,13,50,00,000 (the number of reads estimated to yield 56× coverage for an estimated genome of ∼2.3 Gigabases; Table S2).

EXON CAPTURE ENRICHMENT AND SEQUENCING

We generated new genome-wide resequencing data from 112 individuals and 15 chipmunk species (including *T. striatus* as an outgroup; Table S1) using a custom in-solution targeted capture experiment designed by Bi et al. (2019) to enrich and sequence exons (including flanking introns and intergenic regions) from 10,107 nuclear protein-coding genes (9.4 Megabases [Mb] total and the complete ∼16.5 kilobase [kb] mitochondrial genome). This capture strategy was iteratively developed from a series of genomic experiments. First, RNA-seq was used to sequence and assemble transcriptomes from multiple tissue types sampled from a single chipmunk species (*T. alpinus*), which served as the reference for exome capture probe design (Bi et al. 2012). These transcriptome data were used to develop an array-based capture targeting ∼8000 nuclear genes as well as the complete mitochondrial genome, resulting in ∼6.9 Mb of assembled exonic regions (including flanking introns and intergenic regions; Bi et al. 2012, 2013; Good et al. 2015). Finally, an additional ∼2.4 Mb of genic regions were identified from the original transcriptome through AmiGO and NCBI protein databases and were added to the capture design. Published data from Sarver et al. (2021; *Tamias canipes*, *Tamias cinereicollis*, *Tamias dorsalis*, *Tamias quadrivitattus*, *Tamias rufus*, and *Tamias umbrinus*) were based on the earlier array-based design and thus represent a subset of the newly targeted genomic regions.

To localize the targeted genomic regions within the de novo *T. minimus* genome, we first used BLAT to identify best hits in the reference. We then sorted and filtered the contigs to include only the best match hits, discarding contigs with equally likely hits across the genome. To identify X-linked contigs in the genome assembly, we calculated the normalized female versus male ratio of sequencing coverage for targeted regions mapping to each contig, which should be approximately 2:1 for genes on the X chromosome.

Total genomic DNA was isolated from heart or liver tissue using Qiagen DNEasy DNA blood and tissue kits. Illumina sequencing libraries were prepared for each sample following the protocol of Meyer and Kircher (2010). Sequencing libraries were then enriched in solution using probes designed as described above and manufactured by NimbleGen (SeqCap EZ Developer kits). Hybridization reactions were performed in four separate equimolar pools of indexed libraries. To assess capture efficiency, we first sequenced a subset of pooled libraries on an Illumina MiSeq sequencer (150 bp paired-end reads) at the University of Montana genomics core (Missoula, MT). Final pooled libraries were sequenced on an Illumina HiSeq 2500 (150 bp paired-end reads) at the University of Oregon (Eugene, OR).

ILLUMINA DATA PROCESSING AND GENOTYPING

Raw read data were cleaned using the *HTStream* pipeline (available from https://github.com/s4hts/HTStream; accessed July 21, 2018) to trim adapters and low-quality bases, merge overlapping reads, and remove putative PCR duplicates. Cleaned reads were mapped to the reference genome using the BWA-MEM alignment algorithm as implemented in BWA version 0.7.15 (Li and Durbin 2009; Li 2013) with default options and then sorted with *samtools* version 1.4 (Li et al. 2009). Mapped reads were assigned read groups and duplicate reads were identified post-mapping using Picard version 2.5.0 (http://broadinstitute.github.io/picard/). Initial capture efficiency statistics were calculated using QC3 version 1.33 (Guo et al. 2014). Regions with insertions or deletions (indels) were identified and realigned, and single nucleotide variants were called using UnifiedGenotyper within the Genome Analysis Toolkit (GATK) version 3.6 (McKenna et al. 2010; DePristo et al. 2011; Van der Auwera et al. 2013) with the EMIT_ALL_SITES argument set. This resulted in a final VCF with a call at every position. We then soft filtered based on minimum mapping quality (MQ $<$ 20), a minimum quality (QUAL $<$ 20), a minimum quality by depth (QD $<$ 2), and a minimum sequencing depth (DP $<$ 5), resulting in a final filtered single sample VCF file used for all downstream analyses.

PHYLOGENETIC INFERENCE AND SPECIES DELIMITATION

To obtain sequence alignments for phylogenetic analyses, we constructed consensus sequences of the targeted regions using the filtered VCF for each individual by injecting all confidently called sites back into the original reference using FastaAlternateReferenceMaker within the GATK. IUPAC ambiguity codes were inserted at putative heterozygous positions using the -IUPAC flag. Remaining ambiguous positions (MQ $<$ 20, QUAL $<$ 20, QD $<$ 2, DP $<$ 5) were hard masked (i.e., replaced with an "N") using GNU awk and bedtools version 2.25 (Quinlan and Hall 2010).

We used a two-tiered approach to infer phylogenetic relationships within *Tamias*. First, we used a concatenated alignment of the targeted regions and 100 base pairs (bp) of flanking sequence to estimate a single bifurcating phylogeny using a maximum likelihood (ML) search as implemented in IQ-Tree version 1.5.5 (Nguyen et al. 2015). For all phylogenetic analyses, we ran separate analyses for autosomal and X-linked regions. We use ModelFinder Plus (-m MFP) (Kalyaanamoorthy et al. 2017) and Bayesian information criterion (BIC) to select the best nucleotide substitution model for the concatenated data. To assess nodal support, we performed 1000 ultrafast bootstrap replicates (UFBoot) with the "-bb" command (Minh et al. 2013; Hoang et al. 2017). Concatenation of multiple loci ignores coalescent stochasticity and can result in erroneously high bootstrap support for nodes that may show substantial incongruence when considered on a per locus basis (Kumar et al. 2012). Therefore, we also calculated gene concordance factors (gCF) and site concordance factors (sCF) as implemented in IQ-Tree version 2.0.6 (Minh et al. 2020a, 2020b). Gene concordance factors are the proportion of inferred single locus trees (gene trees) that support a particular branch in the reference tree (ML concat tree), whereas site concordance factors are the proportion of sites within the concatenated alignment that support a branch in the reference tree.

To obtain gene trees, we extracted alignments from the targeted regions and 100 bp of flanking sequence by combining all targets found within stepped 50 kb intervals using bedtools version 2.2.5 and combined regions by scaffold using AMAS (Borowiec 2016). This approach was chosen to capture the spatial distribution of individual targets while ensuring we include enough informative sites to infer independent phylogenetic histories. Although interval size is somewhat arbitrary here, we considered a range of interval sizes and found smaller intervals ($\leq$10 kb) resulted in many intervals without targets and few informative sites. Likewise, much larger interval sizes ($\geq$100 kb) were more likely to conflate independent phylogenetic histories. For each interval, we filtered positions with missing data for $>$85% of individuals using AMAS and excluded intervals containing less than 1 kb of capture data. We then used IQ-Tree version 1.5.5 to estimate local maximum likelihood trees with 1000 ultra-fast bootstraps (using -m MFP, -bb 1000), with *T. striatus* as the outgroup. Gene trees, when necessary, were unrooted using the R package *ape* (Paradis et al. 2004). We then used unrooted gene trees to estimate a consensus species tree using ASTRAL-III version 5.6.3 (Zhang et al. 2018). Assuming sets of independent and accurately estimated gene trees, ASTRAL estimates an unrooted species tree by finding the species tree that has the maximum number of shared induced quartet trees with the given set of gene trees (Mirarab et al. 2014a, 2014b).

Next, we estimated coalescent-based species trees with SVDquartets (Chifman and Kubatko 2014) using 100,000 random quartets and 100 bootstrap replicates as implemented in PAUP* version 4.0a166 (Swofford 2003). SVDquartets assesses support for quartets of taxa directly from site-pattern frequencies of variable sites only. This approach differs from summary methods, such as ASTRAL, because it does not independently estimate gene trees, avoiding the issue of gene-tree estimation error (Chifman and Kubatko 2015; Chou et al. 2015). For SVDquartets, we extracted single nucleotide variants distanced at least 10 kb along the genome and excluded sites with missing information for >85% of the individuals. For both SVDquartets and ASTRAL analyses, species trees were estimated with and without assigning species identities and using sites or intervals for autosomal and X-linked loci separately as previously described. When applicable, we included the eastern chipmunk, *T. striatus*, as the outgroup.

Finally, we estimated a species tree using the Bayesian program BPP version 4.4.1 (Rannala and Yang 2017; Flouri et al. 2018). BPP uses a full likelihood-based implementation of the multispecies coalescent and accommodates uncertainty due to gene tree heterogeneity (both the topology and branch lengths) and incomplete lineage sorting. We first performed preliminary runs using a diffuse prior (shape parameter $\alpha = 3$) to assess the fit of the prior means with our data. For our model parameters, we assigned inverse-gamma priors to the population size parameter θ [$\sim$IG (3, 0.002) with mean 0.00], and to the divergence times τ [$\sim$ IG (3, 0.01) with mean 0.005 for the age of the root]. We assessed convergence based on consistency across five replicate MCMC runs, using a different starting tree and seed for each. For each run, we generated 200,000 samples with a sample frequency of two iterations after a burn-in of 16,000 iterations. Although BPP can potentially analyze thousands of genetic loci, we found that runtimes were excessive even with our relatively modest genome-wide capture dataset. Therefore, we explored several pruning options and settled on selecting the largest contig with a minimum interval distance of 1 Mb. This resulted in a sequence alignment consisting of 88 loci (ranging in size from $\sim$1000 to 4600 bp; median 2343 bp) for a total alignment length of 209,725 bp.

DIVERGENCE TIME ESTIMATES AND GENETIC DIVERGENCE

We performed least squares-based fast dating (LSD2) on the inferred species tree as implemented in IQ-Tree (To et al. 2016) to estimate divergence times. We used a calibration date of 7 Myr based on the Late Miocene date of Dalquest et al. (1996), which is the earliest known fossil occurrence for the root split between *T. striatus* and the western chipmunk radiation. Next, we estimated interspecific absolute genetic divergence (d_{XY}) over 100 kb nonoverlapping windows, considering both polymorphic and monomorphic sites using the program *pixy* (Korunes and Samuk 2021).

INFERENCE OF POPULATION STRUCTURE

We plotted PC1 and PC2 from a principal components analysis (PCA) using the R package SNPRelate (Zheng et al. 2012). To estimate individual co-ancestry, we performed ML estimation of individual admixture proportions using the program ADMIXTURE version 1.30 (Alexander et al. 2009). We tested values of K (the number of population clusters) ranging from 1 to 6 under the default settings in ADMIXTURE and selected the model with the lowest cross-validation error. For both PCA and population cluster analyses, we used a filtered SNP dataset where we applied LD-based pruning to filter our SNP dataset using PLINK version 1.9 (Chang et al. 2015) to minimize nonindependence due to linkage. We used a 1 kb sliding window with a step size of 100 bp and pairwise r^2 of 0.8 as a cut-off for removing highly linked SNPs. This resulted in a final set of 93,247 SNPs.

INTROGRESSION AND PHYLOGENETIC NETWORKS

An advantage of SVDquartets is that it detects deviations in site-pattern frequencies expected under the multispecies coalescent by evaluating support for all three resolutions for each quartet, similar to the ABBA-BABA test (Green et al. 2010; Durand et al. 2011). Therefore, we used the software package HyDe (Blischak et al. 2018) to test for hybridization using the invariants framework of SVDquartets. Here, the quartet with the majority of the support should correspond to the species tree. With incomplete lineage sorting (ILS) and in the absence of introgression, the two minor quartet topologies should show similarly low support. Alternatively, introgression will lead to an imbalance of support toward the topology with the two taxa exchanging alleles as sister taxa (Pease and Hahn 2015; Blischak et al. 2018; Kubatko and Chifman 2019). HyDe also estimates γ, which is the parental contribution in a putatively hybrid genome, where a value of 0.5 would indicate a 50:50 genomic contribution from each parent.

We also modeled hybridization and ILS under the coalescent network framework as implemented in PhyloNet 3.6.6 (Yu and Nakhleh 2015). Networks were computed under maximum pseudo-likelihood (*InferNetwork_MPL*; Yu and Nakhleh 2015) using rooted autosomal gene trees. We modeled 0–5 migration events, associated individuals to species (option -a), and optimized branch lengths and inheritance probabilities to compute likelihoods for each proposed network (option -o). We calculated Akaike information criteria corrected for small sample sizes (AICc) and Bayesian Information Criteria (BIC) using the highest likelihoods per run to compare the resulting networks (Yu et al. 2012, 2014). Networks were visualized with IcyTree (https://icytree.org; accessed December 2020).
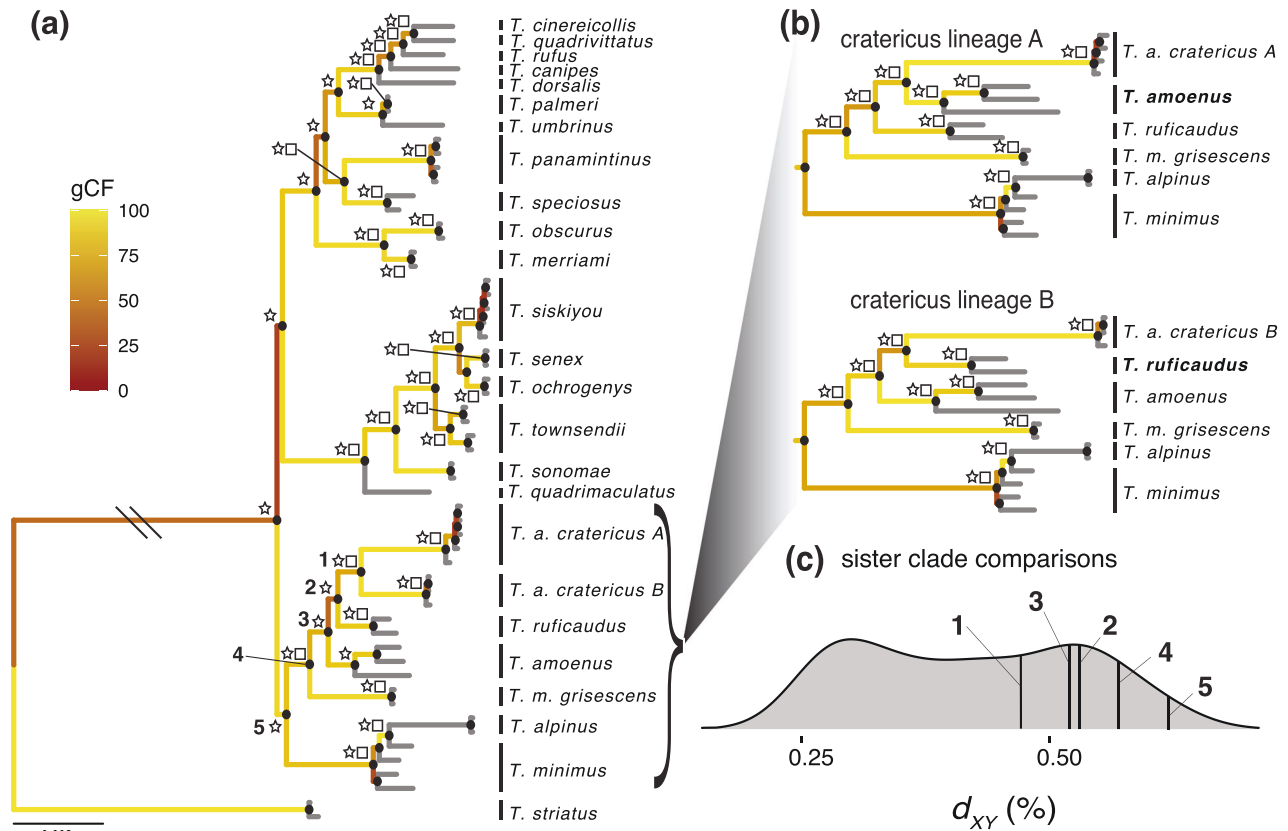
Figure 2. Genome-wide phylogeny of Tamias. (a) Concatenated ML tree estimated with IQ-Tree; node labels represent ASTRAL posterior probability equal to one (star), SVDquartets bootstrap greater than 90 (square), and bootstrap proportions greater than 90 (black circles). Branches are colored according to gene concordance factor (gCF). (b) ML reconstruction of the amoenus-ruficaudus-cratericus clade with the exclusion of cratericus lineage B and cratericus lineage A, respectively. (c) Distribution plot of d_{XY} for comparisons between sister clades. Numbers on plot correspond to nodes defining sister clade comparisons labeled on the tree in panel A.

Results

DRAFT GENOME ASSEMBLY AND SEQUENCE CAPTURE

We generated ∼217 Gb of *T. minimus* genomic sequence data using 10X Genomics technology (Pleasanton, CA, USA) and assembled a 2.48 Gb draft genome with a contig N50 of 196.09 kb and as scaffold N50 of 58.28 Mb, respectively (Table S2). We then analyzed up to 9.4 Mb of sequence capture data from 121 individuals, combining newly generated (112 individuals from 15 species) and published data (nine individuals from seven species; Bi et al. 2019; Sarver et al. 2021; Table S1) for 21 western chipmunk species and the outgroup *T. striatus*. Sequencing efforts produced ∼4 million reads per sample with an average of 0.5% of targets showing no coverage. Approximately 75% of raw reads were unique, resulting in an average target coverage of 33× across samples (range: 5−91×) with ∼73% of targeted bases sequenced to at least 10× coverage (Table S3). Comparison of male versus female coverage resulted in the identification of three

contigs (79.7 Mb, 35.6 Mb, 5.9 Mb) in the genome assembly that are likely on the X chromosome.

THE *TAMIAS* PHYLOGENY

We first estimated a genus-wide ML phylogeny from a concatenated set of 21,551 autosomal (5,365,556 bp) and 493 X-linked (105,671 bp) loci from the combined capture datasets, as well as an ML phylogeny based on the full mtDNA genome (Fig. 2a; Figs. S1 and S2). The concatenated autosomal ML analysis recovered a fully resolved phylogeny with high support for all branches (UFBoot > 90%) and is concordant with previous studies (Reid et al. 2012; Sullivan et al. 2014; Sarver et al. 2021). *Tamias m. grisescens* appeared as a distinct lineage in these analyses and was not most closely related to other *T. minimus* samples. Likewise, *cratericus* was a distinct, monophyletic lineage that was most closely related to *T. ruficaudus* and not other *T. amoenus*. Within *cratericus*, we also recovered a deep split between the two sampled localities (hereafter *cratericus* lineages

A and B; Fig. 2a). ML reconstruction of the mitochondrial genome resulted in a different topology from the nuclear dataset (Fig. S2) and was largely concordant with previous studies based on the cytochrome *b* gene (Demboski and Sullivan 2003; Good et al. 2003, 2008, 2015; Reid et al. 2012; Sarver et al. 2017) with *cratericus* most closely related to *T. quadrimaculatus*, and then to *T. minimus*. Our mitochondrial genome reconstruction also recovered an individual from the putative *cratericus* lineage B (CRCM06-160; Fig. S2) as having a *T. amoenus* mtDNA genome, consistent with introgression. The multispecies-coalescent species trees estimated from ASTRAL (158 genes trees estimated from 50 kb genomic intervals) and SVDquartets (13,482 unlinked SNPs) was largely concordant with the results of the concatenated ML analyses. Strong support was recovered for the branching relationships among species groups within *Tamias* (ASTRAL posterior probabilities > 0.9 and SVDquartets bootstrap support > 90). All analyses recovered both *grisescens* and *cratericus* as distinct, monophyletic lineages. However, one notable difference across analyses was the relationship of *cratericus* with respect to *T. ruficaudus* and other *T. amoenus*. Both concatenated ML analyses and ASTRAL recovered *cratericus* as the sister lineage to *T. ruficaudus*, whereas SVDquartets recovered *cratericus* as more closely related to other *T. amoenus* but with weak support (Figs. S3–S6). Overall, the splits between *T. amoenus-cratericus-T. ruficaudus* were not well supported and remained unresolved (Fig. 2a; Figs. S1–S6).

Given the deep split within the *cratericus* clade and low support for the branching structure, we hypothesized that there might be unequal ancestry between the two *cratericus* lineages, *T. amoenus* and *T. ruficaudus*. To explore this, we re-analyzed these data while excluding either *cratericus* lineage A or lineage B. When *cratericus* lineage B was excluded, we consistently recovered *cratericus* as the sister lineage to *T. amoenus* with strong support. In contrast, analyses excluding *cratericus* lineage A recovered *cratericus* as sister to *T. ruficaudus* with strong support (Fig. 2b).

POPULATION GENOMICS OF *CRATERICUS*

To elucidate the evolutionary history and geographic extent of the cryptic *cratericus* lineages, we sequenced an additional 64 chipmunks from an additional 12 localities (seven *cratericus/T. amoenus* localities; five *T. minimus* localities; Fig. 3) ranging from relatively low desert sage scrub to higher elevation temperate coniferous forests throughout south central Idaho. Initial qualitative assessment of gross bacular morphologies from these samples suggested that *cratericus* may be more widespread in central Idaho, ranging north from the temperate coniferous forest/sage scrub transitional zone of the Snake River Plain to the temperate/subalpine zones of the Salmon River. Following the procedures above, we generated an alignment of 7,813,766 bp

for autosomal loci and 175,884 bp for X-linked loci with >80% of individuals genotyped at a minimum coverage of 5×.

The ML tree from the concatenated set of autosomal and X-linked loci were largely congruent with the genus-wide analysis. *Tamias minimus* (minus *grisescens*) was paraphyletic, with *T. alpinus* nested within the main *T. minimus* clade. *T. m. grisescens* is a distinct lineage that was more closely related to the *T. amoenus-cratericus-T. ruficaudus* complex than to other *T. minimus*. Our expanded sampling substantially increased the diversity detected within *cratericus* lineage A, which was found at six additional localities to the near exclusion of *T. amoenus*. We again recovered a deep split within *cratericus*, with lineage B still represented by three samples from a single locality (Fig. 3; Fig. S7).

However, the most notable difference between this expanded subset and our genus-wide analysis was that *cratericus* now appeared sister to other *T. amoenus* (Fig. 3b; Figs. S7–S12). One potential source of phylogenetic discordance between the datasets may be driven by samples from one locality, known as Iron Creek (IC; yellow star; Fig. 3c). At this locality, both *cratericus* and standard *T. amoenus* bacular types were found to co-occur without other obvious phenotypic or habitat differences. This was the only site where both forms were found to be sympatric. Collectively, these results suggest that *cratericus* is distributed across the coniferous forest of south central Idaho, north of the Snake River Plain to the near exclusion of *T. amoenus*, save for one sympatric locality. At this site, we observed the maintenance of genetically defined species boundaries despite some evidence for admixture between *cratericus* and *T. amoenus* (see below).

Next, we evaluated if the change in the relationship between *T. ruficaudus- cratericus-T. amoenus* was influenced by possible gene flow at the sympatric IC locality. Consistent with this, *cratericus* was again sister to *T. ruficaudus* in a concatenated analysis when excluding all IC samples. However, the species trees estimated from ASTRAL (158 genes trees estimated from 50 kb genomic intervals) and SVDquartets (17,594 unlinked SNPs) for the expanded subset resulted in *cratericus* being sister to *T. amoenus*, regardless of whether IC individuals were included. Branch support increased when IC samples were excluded but we still found a large degree of incongruence based on gCF and sCF. Finally, the results of our BPP analyses were largely consistent across the five replicate runs using different starting species trees, with four of the five runs converging on the same maximum *a posteriori* probability (MAP) tree, with a posterior probability of ~100%. BPP recovered *cratericus* as being sister to *T. ruficaudus* (Fig. S12).

ML reconstruction of mitochondrial genomes from the expanded subset also showed evidence for mtDNA introgression between *cratericus* and *T. amoenus* (Fig. S13). As with the genus-wide analysis, *cratericus* and *T. amoenus* were paraphyletic, but
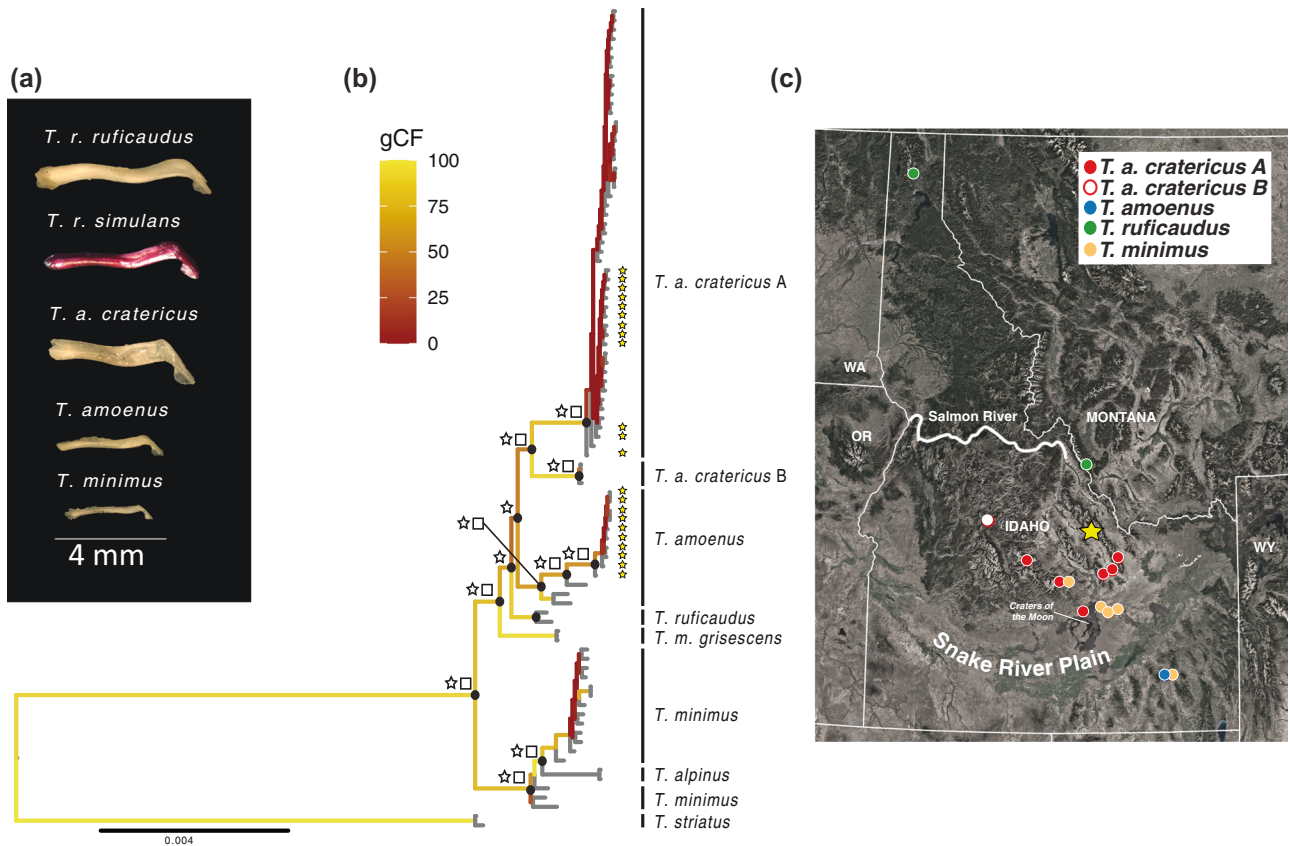
Figure 3. ML concatenated phylogeny for cratericus-amoenus-ruficaudus. (a) Representative bacular morphologies for the five taxa. (b) Concatenated ML tree estimated with IQ-Tree for the full population level sampling of cratericus; node labels represent ASTRAL posterior probability equal to one (star), SVDquartets bootstrap greater than 90 (square), and bootstrap proportions greater than 90 (black circles). Branches are colored according to gene concordance factor (gCF). Yellow stars at the tips indicate cratericus and *T. amoenus* samples from the sympatric Iron Creek locality (yellow star on map). (c) Sampling localities for population sampling. Note that our sampling also includes non-Idaho populations for *T. minimus*, *T. amoenus*, and *grisescens* that are not shown.

the paraphyly in the mtDNA tree was not solely driven by discordance within the IC locality. One *cratericus* mtDNA clade was composed of individuals from the majority of *cratericus* lineages A and B but excluding IC and was sister to a monophyletic *T. minimus* clade. The second major group was composed of *T. amoenus*, including IC *T. amoenus* and *cratericus* from IC as well as a single individual from lineage B (CRCM06-160 as in the genus-wide analysis; Fig. S13).

Principal components of genetic variance (PCA) further supported a history of hybridization and mixed ancestry between *T. amoenus-cratericus-T. ruficaudus*. PC1 largely partitioned the most divergent lineage, *T. minimus*, from a cluster consisting of *cratericus-T. amoenus-T. ruficaudus*. PC2 splits *T. amoenus* and *cratericus*, with *T. ruficaudus* falling intermediate. Cross-validation with ADMIXTURE suggested five population clusters that were largely concordant with the PCA (Fig. 4B; Figs. S14 and S15), although we noted

that this analysis is not well suited to partition genetic clusters at this scale of interspecific divergence (Lawson et al.

2018) or when sample sizes are strongly uneven between putative clusters. Interestingly, both methods showed a tendency to cluster *cratericus* lineage B with *T. ruficaudus* and suggested a closer relationship between *cratericus* and *T. ruficaudus* than to *T. amoenus*. However, we also detected a moderate proportion of shared ancestry between sympatric (IC) *cratericus* and *T. amoenus*, likely driven by gene flow in sympatry.

PATTERNS OF INTROGRESSION AND NETWORK ANALYSES

We tested for further evidence of introgression using multiple analyses based on the multispecies network coalescent (PhyloNet), quartet-based analyses under the invariants framework (HyDe), and estimates of admixture proportions based on the D-statistic (Green et al. 2010; Durand et al. 2011; Yu and Nakhleh 2015; Blischak et al. 2018). Collectively, these analyses supported varying levels of recurrent gene flow between both *cratericus-T. ruficaudus*, and *cratericus-T. amoenus*. PhyloNet supported a model of more ancient introgression, which has likely
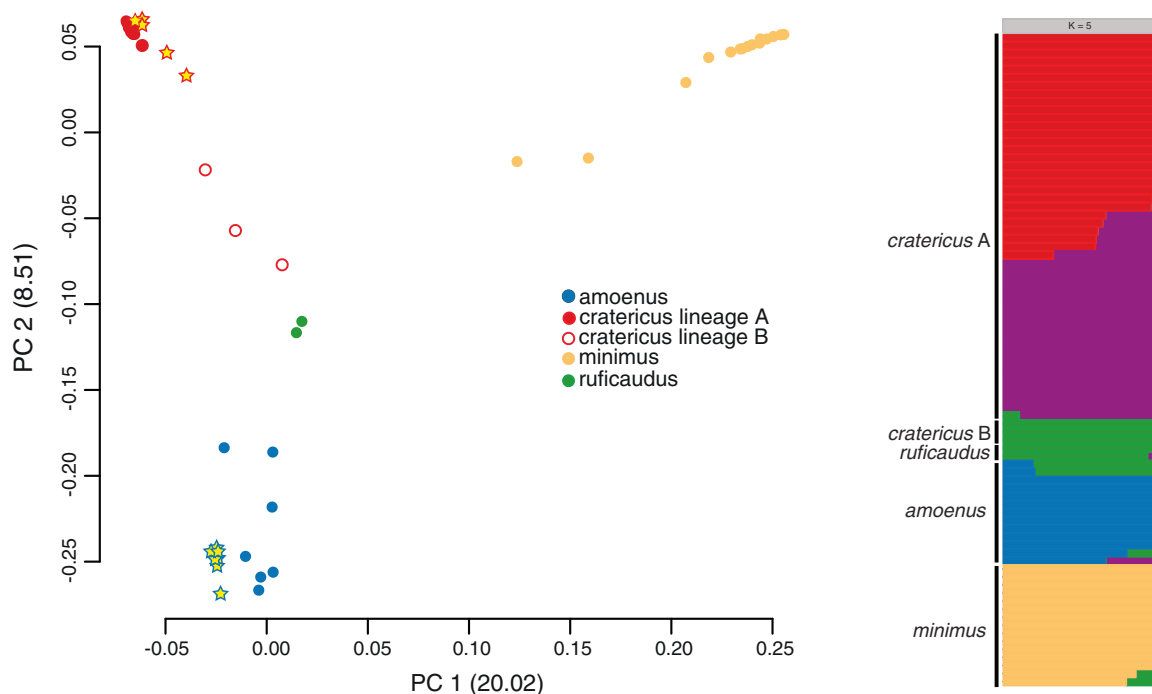
Figure 4. Population structure between cratericus-amoenus-ruficaudus sampled. (a) PCA results from SNPRelate. (b) Admixture output for highest marginal-likelihood run at $k = 5$. Stars indicate samples from the sympatric locality Iron Creek.

affected the inferred split between *T. amoenus-cratericus-T. ruficaudus*. The best supported network analysis (five reticulations) suggested that *T. ruficaudus* and *cratericus* were sister lineages, with reticulation between ancestral populations of *T. amoenus* and the *cratericus* lineage. Other less well-supported reconstructions placed *T. amoenus* as more closely related to *cratericus* with a complex pattern of recurrent gene flow between *cratericus* lineage A and B with both *T. ruficaudus* and *T. amoenus* (Fig. S16). Quartet-based analyses of introgression resulted in three significant triplet comparisons supporting hybridization (Table S4). First, *cratericus* lineage A showed shared ancestry between *T. amoenus* and *cratericus* lineage B, consistent with a close relationship between *cratericus* lineages A and B and recent hybridization between both lineages and *T. amoenus*. We also detected a pattern of shared ancestry between both *cratericus* lineages (A and B) with *T. ruficaudus* and *T. amoenus*. However, the proportion of shared ancestry was asymmetrical with respect to the amount of introgression inferred from *T. amoenus* into either *cratericus* lineage (*T. amoenus-cratericus* lineage A $\gamma = 0.344$; *T. amoenus-cratericus* lineage B $\gamma = 0.541$). Under the ABBA-BABA framework, we evaluated phylogenetic patterns of shared single nucleotide variants between both *T. amoenus-cratericus* and *T. ruficaudus-cratericus*. We found that most sites supported a pattern of differential introgression between *cratericus* lineage A and both *T. amoenus* and to a lesser extent *T. ruficaudus* relative to *cratericus* lineage B (Fig. S17; Tables S4–S6).

Discussion

Discovery of cryptic species is fundamental to understand the process of speciation and cataloging global diversity, which is facing unprecedented pressures due to the emerging threats of climate change and habitat loss. We generated a de novo assembled reference genome for a western chipmunk species, *Tamias minimus*, and used new and published genome-wide targeted capture sequencing data and whole mitochondrial genomes to test for cryptic speciation and hybridization in western chipmunks. Our genus-wide analysis represents the most comprehensive phylogenetic assessment of the western chipmunks to date and, we suggest, reveals at least two new chipmunk species. Below we discuss the evolutionary implications of our findings, focusing first on western chipmunk diversity and then the complex evolutionary history of the Crater chipmunk and other co-distributed species.

SYSTEMATICS OF THE WESTERN CHIPMUNK RADIATION

Due to their abundance and conspicuous nature, chipmunks have been the focus of a broad range of biological studies including physiology (e.g., Levesque and Tattersall 2009), niche partitioning and competitive exclusion (e.g., Grinnell and Storer 1924; Heller 1971), behavior (e.g., Broadbrooks 1970), and responses to climate change (e.g., Moritz et al. 2008; Bi et al. 2019). Early phylogenetic studies using morphology and allozymes

(Levenson et al. 1985), host/ectoparasite (Jameson 1999), chromosomal (Nadler and Block 1962), and molecular data (Piaggio and Spicer 2000, 2001) recognized three distinct clades within *Tamias*: *T. sibiricus*, *T. striatus*, and the western North America species arbitrarily classified as either subgenera or genera (e.g., Jameson 1999; Patterson and Norris 2016). Relationships among the 24 currently recognized species within the western chipmunk clade (subgenus *Neotamias*) have remained somewhat obscure due to a lack of resolution among internal nodes and apparently rampant mitochondrial introgression (Piaggio and Spicer 2000; Reid et al. 2012; Sullivan et al. 2014). Our analysis of western chipmunks represents the most comprehensive genome-wide assessment of this group to date and included 22 of the 26 recognized species of *Tamias*.

We recovered strongly supported, monophyletic species groups within the western chipmunks that were in general agreement with some other studies (e.g., Reid et al. 2012). Most of our tree-based approaches converged on a similar topology where we found strong support and congruence among the western chipmunk species groups. However, one notable difference was the relationship between *T. ruficaudus-T. amoenus-T. minimus*. In contrast to Reid et al. (2012), we recovered *T. ruficaudus* as being sister to either *cratericus* or to a clade composed of *T. amoenus-cratericus* (see discussion below), whereas Reid et al. (2012) found *T. ruficaudus* sister to the enigmatic, and similarly distinct lineage *grisescens*. In addition, Reid et al. (2012) also recovered *T. amoenus* as sister to *T. minimus*, whereas we found *T. minimus* as the sister lineage to the main *T. amoenus-T. ruficaudus-cratericus-grisescens* clade. Given a lack of resolution among internal nodes in the previous study, this difference likely reflects the power afforded by much more extensive genetic sampling.

A major conclusion of our genomic study was the resolution of two cryptic lineages of chipmunk. Although some of the relationships among described species have been unclear until recently, the identity of species as the fundamental units of diversity have remained fairly stable over the last century. The vast majority of chipmunk species were described to some degree by the late 19th and early 20th centuries by the early naturalists Allen, Merriam, and others based on phenotypic and ecological characteristics (for a review, see Thorington et al. 2012). Subsequent discussions have focused on determining if subspecific variation within some species warrant species-level recognition (Patterson 1984; Patterson and Heaney 1987; Good et al. 2003) or if some current species are actually geographically isolated populations of more widespread forms (Piaggio and Spicer 2000, 2001; Rubidge et al. 2014). The putative *grisescens* lineage reflects these dynamics. Howell (1925) described the distinctive *T. m. grisescens* based on its pelage as a least chipmunk subspecies restricted to the Channeled Scablands of central Washington. Reid et al. (2012) then identified this taxon as a potentially cryptic

lineage and found it consistently nested with *T. amoenus* based on limited nuclear and mtDNA data. Although a formal description awaits additional geographic sampling, our data clearly show that *grisescens* does not fall within the considerable genetic diversity of the least chipmunk (*T. minimus*) or other sampled species, and we propose that *grisescens* is in fact a novel, cryptic chipmunk species (i.e., *Tamias grisescens* following nomenclature of Howell 1925). Of note, *grisescens* is much more genetically distinct from other chipmunk species than the Alpine chipmunk (*T. alpinus*); a lineage that is paraphyletic with respect to *T. minimus* but has long been recognized as a distinct species based on phenotypic divergence, ecological differentiation, and reproductive isolation (Grinnell and Storer 1924; Heller 1971; Rubidge et al. 2014; Bi et al. 2019). Absolute genetic divergence (d_{XY}) between *grisescens* and its sister clade (*cratericus-T. ruficaudus-T. amoenus*) was among some of the highest sister-pair estimates observed (Fig. 2c; Fig. S18; $d_{XY} = 0.57\%$) across the western chipmunks. We also found *grisescens* nested within *T. amoenus* in our mtDNA analyses, suggesting a history of mtDNA introgression or incomplete lineage sorting (Figs. S2 and S13). Our study lacks the sampling needed to fully investigate the evolutionary history of *grisescens* (i.e., two individuals from the same locality). However, its range is presumed to be limited to central Washington (Howell 1925), which could contribute to vulnerability due to habitat loss, wildfire, and drought due to rapid climate change.

We also show that *cratericus* is likely a distinct chipmunk species that is relatively divergent from its most closely related sister taxon, *T. ruficaudus* (Fig. 2). Blossom (1937) first described *cratericus* as a duller, dark variant of the yellow-pine chipmunk associated with the recent volcanic lava flows of Craters of the Moon. The color morphology of *cratericus* rapidly transitions to a more brightly colored pelage morph more like standard *T. amoenus* phenotypes and, we show, is found throughout the adjacent xeric forest habitats of the region. Thus, the localized melanic pelage presumably reflects very recent adaptation for crypsis on the black lava flows, a pattern that is also observed in local populations of the Great Basin pocket mouse (*Perognathus parvus*) and pika (*Ochotona princeps)* from Craters of the Moon (Blossom 1937). More surprisingly, *cratericus* appears to be the locally dominant form throughout central Idaho (see below). Our divergence time estimate puts the split between *cratericus* and *T. ruficaudus* at ∼0.73–1.22 Mya (Fig. S19). This massively predates the estimated age of the quite recent Craters lava flows (∼15–2 kya), suggesting local adaptation to this novel habitat well after the origin of this lineage.

Overall, we identified two cryptic species that show relatively deep divergences versus putative sister species when compared to estimates of d_{XY} between many other described western chipmunk species (Fig. 2c; Fig. S18). Neither of these cryptic

lineages were sister taxa to the species they are currently described as belonging to. Furthermore, there are likely to be additional instances of cryptic diversity within this radiation given the propensity for phenotypic convergence of standard taxonomic characters (e.g., pelage and size) and the broad ecological and geographic distribution of generalist species like *T. amoenus* and *T. minimus*. We also found that most western chipmunk groups displayed a high degree of discordance between the mtDNA phylogeny and the nuclear genome, consistent with previous studies documenting recurrent mtDNA introgression layered across the history of this group (e.g., Demboski and Sullivan 2003; Sullivan et al. 2014; Sarver et al. 2017). Indeed, mtDNA tends to do a relatively poor job of tracking species boundaries within western chipmunks (Sullivan et al. 2014; Sarver et al. 2021). Interestingly, previous works have suggested distinct species boundaries with comparably low levels of nuclear introgression relative to more rampant mtDNA introgression (Good et al. 2015; Sarver et al. 2021; however, see Ji et al. 2021). From this perspective, the complex evolutionary history and more extensive nuclear gene flow observed between *cratericus* and both *T. amoenus* and *T. ruficaudus* is noteworthy relative to other chipmunk studies to date.

THE COMPLEX EVOLUTIONARY HISTORY OF *TAMIAS CRATERICUS*

Our data and analyses provide compelling evidence to support *cratericus* as at least one distinct species. However, the full evolutionary history of this taxon appears to be obscured by a history of rapid speciation and likely recurrent instances of hybridization with populations of *T. amoenus* and, possibly, *T. ruficaudus*. To reconstruct this complex evolutionary history, we considered how incomplete lineage sorting and gene flow structured through space and time may lead to contrasting patterns of shared genetic variation among populations. When we assessed the placement of *cratericus* in a genus-wide context, we found that it was most closely related to *T. ruficaudus* albeit with low support (Fig. 2). Interestingly, SVDquartets converged on a different topology than both ASTRAL and BPP. ASTRAL and BPP agreed with the concatenated analysis and found *cratericus* sister to *T. ruficaudus*, whereas SVDquartets consistently grouped *cratericus* with *T. amoenus*. Both ASTRAL and SVDquartets appeared sensitive to sampling and the branches in question received moderate to weak support across all analyses (Fig. 2; Figs. S3–S6, S8–S11). In contrast, although we found strong support with BPP, these analyses were limited to a relatively small subset of data that could produce misleading results given the potential for differential introgression.

Low branch support could be due to either a lack of information across all loci used to resolve relationships (e.g., due to very short internal branches and/or homoplasy) or because gene trees have independent phylogenetic histories that differ from the species tree due to incomplete lineage sorting and/or hybridization. These two sources of discordance likely compromise our ability to resolve this portion of the *Tamias* phylogeny; both gene concordance factors (gCF) and site concordance factors (sCF) for the *cratericus-T. ruficaudus* split were comparably low (gCF: 37.3; sCF: 38.6; Fig. 2, Fig. S1).

One major criticism of coalescent-based summary methods, such as ASTRAL, is that they may incorrectly infer a species tree because they are sensitive to stochastic variation in phylogenetic signal between loci (Mirarab et al. 2016; Morales-Briones et al. 2018). However, the discordance observed between our approaches likely reflect biological causes of discordance such as ILS and introgression. For example, gene tree estimation error often results in a skewed ratio of gCF:sCF values (Minh et al. 2020a), a pattern we do not see, even across the weakest supported branches.

Although individual tree-based approaches may be limited when local genealogies are poorly resolved due to rapid divergence or limited data, our data also strongly supported the conclusion that much of the phylogenetic discordance likely reflects introgressive hybridization. We consistently detected genealogical asymmetry between *cratericus* lineage A and *cratericus* lineage B with respect to both *T. ruficaudus* and *T. amoenus*. When we assessed the relationship of either *cratericus* lineage in the absence of the other, we found strong support for two different topologies. When we excluded *cratericus* lineage B, *cratericus* lineage A was consistently sister to *T. amoenus*. Conversely, when we excluded *cratericus* lineage A, we recovered *cratericus* lineage B as the sister group to *T. ruficaudus* (Fig. 2b). Further, we inferred multiple reticulations across the *amoenus-cratericus-ruficaudus* splits, suggesting there has been both recent and ancient hybridization between these lineages during their diversification. Both contemporary and past hybridization events have resulted in extensive shared polymorphism among the three lineages (Figs. S16 and S17) and highlight how rapid diversification and introgressive hybridization can confound our ability to infer a complex evolutionary history among species. For example, most analyses decisively group cratericus lineage A and lineage B as monophyletic. However, our ADMIXTURE analysis clustered *cratericus* lineage B with *T. ruficaudus*, likely due to a high degree of shared ancestral polymorphism between *cratericus* lineage B and *T. ruficaudus*. As noted above, this analytical approach is expected to be sensitive to higher levels of genetic divergences and uneven sampling across putative clusters.

Integrating our phylogenetic and population genomic data, we propose a working model for the evolutionary history of *cratericus* (Fig. 5). There was an initial (likely geographic) split between *cratericus* and other *T. ruficaudus*, followed by intermittent hybridization between ancestral populations of *cratericus*, *T. ruficaudus*, and the slightly more distantly related *T. amoenus*.
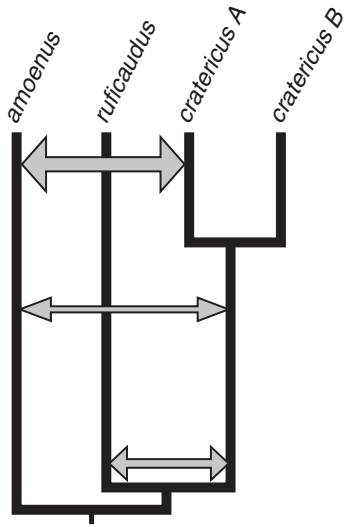
Figure 5. Proposed species tree and evolutionary history of *T. cratericus*. *Tamias cratericus* has a complex evolutionary history punctuated by layered ancient and contemporary gene flow between co-distributed species. Arrow thickness indicates inferred amount of introgression.

Determining if the apparent genetic split within *cratericus* is associated with reproductive isolation between these lineages or divergence in other morphological or ecological traits awaits further sampling. Lineage A appears to be the more widespread and predominant form, but the rarity of lineage B may be due to a lack of sampling in the presumed northern range of *cratericus*. We also detected gene flow between *T. amoenus* and the more widely distributed *cratericus* lineage A (Figs. 3 and 4). Although a history of recurrent hybridization likely skews overall phylogenetic relationships, the sympatric population (IC) allows us to test key predictions about the extent of reproductive isolation between these ecologically and phenotypically similar species. In this area of sympatry, individuals clearly clustered within *cratericus* lineage A or with other *T. amoenus*. Thus, species boundaries appear to be largely maintained, despite some level of gene flow between *cratericus* and *T. amoenus* (Figs. 3 and 4; Fig. S14), consistent with a semipermeable species boundary (Harrison and Larson 2014). Our results also indicate that *cratericus* is the predominant chipmunk species associated with xeric forests of south-central Idaho north of the Snake River Plain and south of the Snake River. This previously unknown lineage appears to exist largely to the exclusion *T. amoenus*. Since the foundational work of Grinnell a century ago (Grinnell and Storer 1924), western chipmunks have been considered exemplars of ecological niche partitioning. Given the lack of obvious ecological and phenotypic differentiation between *cratericus* and *T. amoenus*, sympatric populations between these species may provide a compelling eco-evolutionary context to understand both the local maintenance of species boundaries and niche partitioning through competitive exclusion or other mechanisms.

Finally, our current sampling was not circumscribed by clear geographic barriers to the north and west, indicating the full range of *cratericus* may be considerably larger and likely extends into the remote wilderness tacts of central Idaho. Although these portions of the range are more difficult to access, this diurnal cryptic species is also highly abundant and readily visible around the visitor center of a National Monument that hosts ~200,000 annual visitors—underscoring that cryptic biodiversity may persist even in highly conspicuous systems.

Conclusions

The need to detect and preserve biodiversity has never been more urgent given the emerging anthropogenic threats of global climate change and extensive habitat loss. Western chipmunks have received considerable attention over the last century from biologist and yet cryptic diversity persists even within this well-studied group of mammals. Overall, our study demonstrates the importance of continued specimen collection, incorporating multiple sources of data (genomic, morphological, and ecological) for detecting and describing cryptic diversity and highlights the complexity in defining species boundaries. Our findings further underscore the importance of studies that incorporate more extensive geographic sampling when characterizing genetic diversity across broadly distributed groups. The maintenance and expansion of natural history museum archives, including both historic and contemporary specimen collections, will be essential to this task (Bi et al. 2013; Holmes et al. 2016).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA ARCHIVING

Read sequence data are available for download at the SRA under BioProject accession number PRJNA833404. The de novo genome assembly is available for download from figshare repo: https://doi.org/10.6084/m9.figshare.19853902.v1

LITERATURE CITED

Alexander, D.H., Novembre, J. & Lange, K. (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Resour.*, 19, 1655–1664.

Bi, K., Vanderpool, D., Singhal, S., Linderoth, T., Moritz, C. & Good, J.M. (2012) Transcriptome-based exon capture enables highly cost-effective comparative genomic data collection at moderate evolutionary scales. *BMC Genom.*, 13, 1–14.

Bi, K., Linderoth, T., Vanderpool, D., Good, J.M., Nielsen, R. & Moritz, C. (2013) Unlocking the vault: next-generation museum population genomics. *Molecular Ecology*, 22, 6018–6032.

Bi, K., Linderoth, T., Singhal, S., Vanderpool, D., Patton, J.L., Nielsen, R., Moritz, C. & Good, J.M. (2019) Temporal genomic contrasts reveal rapid evolutionary responses in an alpine mammal during recent climate change. *Plos Genetics*, 15, e1008119.

Bickford, D., Lohman, D.J., Sodhi, N.S., Ng, P.K., Meier, R., Winker, K., Ingram, K.K. & Das, I. (2007) Cryptic species as a window on diversity and conservation. *Trends in Ecology & Evolution*, 22, 148–155.

Blischak, P.D., Chifman, J., Wolfe, A.D. & Kubatko, L.S. (2018) HyDe: a python package for genome-scale hybridization detection. *Systematic Biology*, 67, 821–829.

Blossom, P.M. (1937) Description of a race of chipmunk from south central Idaho. *Occasional Papers Museum Zool.*, 1–3.

Borowiec, M.L. (2016) AMAS: a fast tool for alignment manipulation and computing of summary statistics. *PeerJ*, 4, e1660.

Broadbrooks, H.E. (1970) Home ranges and territorial behavior of the yellow-pine chipmunk (*Eutamias amoenus*). *Journal of Mammalogy*, 51, 310–356.

Brown, J.H. (1971) Mechanisms of competitive exclusion between two species of chipmunk. *Ecology*, 52, 305–311.

Burgin, C.J., Colella, J.P., Kahn, P.L. & Upham, N.S. (2018) How many species of mammals are there? *Journal of Mammalogy*, 99, 1–14.

Callahan, J.R. (1977) Diagnosis of *Eutamias obscurus* (Rodentia: Sciuridae). *Journal of Mammalogy*, 58, 188-201.

Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M. & Lee, J.J. (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*, 4, 7.

Chifman, J. & Kubatko, L. (2014) Quartet inference from SNP data under the coalescent model. *Bioinformatics*, 30, 3317–3324.

Chifman, J. & Kubatko, L. (2015) Identifiability of the unrooted species tree topology under the coalescent model with time-reversible substitution processes, site-specific rate variation, and invariable sites. *Journal of Theoretical Biology*, 374, 35–47.

Chou, J., Gupta, A., Yaduvanshi, S., Davidson, R., Nute, M., Mirarab, S. & Warnow, T. (2015) A comparative study of SVDquartets and other coalescent-based species tree estimation methods. *BMC Genom.*, 16. https://doi.org/10.1186/1471-2164-16-S10-S2.

Dalquest, W.W., Basklin, J.A. & Schultz, G.E. (1996) Fossil mammals from a Late Miocene (Clarendonian) site in Beaver county, Oklahoma. In: Genoways H. H., Baker R. J. (eds). Contribuion in Mammalogy. A memmorial Volume Honoring Dr. J. Knox Jones, JR. Museum of Texas Tech University: Lubbock, TX, USA, pp. 107–137.

de Queiroz, K. (2007) Species concepts and species delimitation. *Systematic Biology*, 56, 879–886.

Delic, T., Trontelj, P., Rendos, M. & Fiser, C. (2017) The importance of naming cryptic species and the conservation of endemic subterranean amphipods. *Science Reports*, 7, 3391.

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

DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., delAngel, G., Rivas, M.A., Hanna, M., et al (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, 43, 491–498.

Durand, E.Y., Patterson, N., Reich, D. & Slatkin, M. (2011) Testing for ancient admixture between closely related populations. *Molecular biology and evolution*, 28, 2239–2252.

Eberhard, W.G. (1985) Sexual selection and animal genitalia. Harvard Univ. Press, Cambridge, MA and London, England.

Edwards, D.L. & Knowles, L.L. (2014) Species detection and individual assignment in species delimitation: can integrative data increase efficacy? *Proc. Biol. Sci.*, 281, 20132765.

Fišer, C., Robinson, C.T. & Malard, F. (2018) Cryptic species as a window into the paradigm shift of the species concept. *Molecular Ecology*, 27, 613–635.

Flouri, T., Jiao, X., Rannala, B. & Yang, Z. (2018) Species tree inference with BPP using genomic sequences and the multispecies coalescent. *Molecular biology and evolution*, 35, 2585–2593.

Fujita, M.K., Leache, A.D., Burbrink, F.T., McGuire, J.A. & Moritz, C. (2012) Coalescent-based species delimitation in an integrative taxonomy. *Trends in Ecology & Evolution*, 27, 480–488.

Good, J.M., Demboski, J.R., Nagorsen, D.W. & Sullivan, J. (2003) Phylogeography and introgressive hybridization: chipmunks (genus *Tamias*) in the northern rocky mountains. *Evolution; Internation Journal of Organic Evolution*, 57, 1900–1916.

Good, J.M., Hird, S., Reid, N., Demboski, J.R., Steppan, S.J., Martin-Nims, T.R. & Sullivan, J. (2008) Ancient hybridization and mitochondrial capture between two species of chipmunks. *Molecular Ecology*, 17, 1313–1327.

Good, J.M., Vanderpool, D., Keeble, S. & Bi, K. (2015) Negligible nuclear introgression despite complete mitochondrial capture between two species of chipmunks. *Evolution; Internation Journal of Organic Evolution*, 69, 1961–1972.

Green, R.E., Krause, J., Briggs, A.W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M.H., et al (2010) A draft sequence of the Neandertal genome. *Science.*, 328, 710–722.

Grinnell, J. & Storer, T.I. (1924) Animal life in the Yosemite. An account of the mammals, birds, reptiles, and amphibians in a cross-section of the Sierra Nevada. Univ. of California Press, Berkeley, CA.

Guo, Y., Zhao, S., Sheng, Q., Ye, F., Li, J., Lehmann, B., Pietenpol, J., Samuels, D.C. & Shyr, Y. (2014) Multi-perspective quality control of Illumina exome sequencing data using QC3. *Genomics*, 103, 323–328.

Hall, E.R. (1981) The mammals of North America. John Wiley & Sons, Inc, New York, NY.

Harrison, R.G. & Larson, E.L. (2014) Hybridization, introgression, and the nature of species boundaries. *J. Heredity*, 105(Suppl 1), 795–809.

Heller, H.C. (1971) Altitudinal zonation of chipmunks (*Eutamias*): interspecific aggression. *Ecology*, 52, 312–319.

Hird, S. & Sullivan, J. (2009) Assessment of gene flow across a hybrid zone in red-tailed chipmunks (*Tamias ruficaudus*). *Molecular Ecology*, 18, 3097–3109.

Hird, S., Reid, N., Demboski, J. & Sullivan, J. (2010) Introgression at differentially aged hybrid zones in red-tailed chipmunks. *Genetica*, 138, 869–883.

Hoang, D.T., Chernomor, O., Haeseler, A.V., Minh, B.Q. & Vinh, L.S. (2017) UFBoot2: improving the ultrafast bootstrap approximation. *Molecular biology and evolution*, 35, 518–522.

Holmes, M.W., Hammond, T.T., Wogan, G.O., Walsh, R.E., LaBarbera, K., Wommack, E.A., Martins, F.M., Crawford, J.C., Mack, K.L., Bloch, L.M., et al (2016) Natural history collections as windows on evolutionary processes. *Molecular Ecology*, 25, 864–881.

Howell, A.H. (1925) Preliminary descriptions of five new chipmunks from North America. *J. Mammal*, 6, 51–54.

Jain, M., Koren, S., Miga, K.H., Quick, J., Rand, A.C., Sasani, T.A., Tyson, J.R., Beggs, A.D., Dilthey, A.T., Fiddes, I.T., et al (2018) Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nature Biotechnology*, 36, 338–345.

Jameson, E.W. (1999) Host-ectoparasite relationships among North American chipmunks. *Acta Theriol.*, 44, 225–231.

Ji, J., Jackson, D.J., Leaché, A.D. & Yang, Z. (2021) Significant cross-species gene flow detected in the *Tamias quadrivittatus* group of North American chipmunks. *bioRxiv.*,. https://doi.org/10.1101/2021.12.07.471567

Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A. & Jermiin, L.S. (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods*, 14, 587–589.

Korunes, K.L. & Samuk, K. (2021) pixy: unbiased estimation of nucleotide diversity and divergence in the presence of missing data. *Molecular Ecology Resources*, 21, 1359–1368.

Kubatko, L.S. & Chifman, J. (2019) An invariants-based method for efficient identification of hybrid species from large-scale genomic data. *Bmc Evolutionary Biology*, 19, 112.

Kumar, S., Filipski, A.J., Battistuzzi, F.U., Kosakovsky Pond, S.L. & Tamura, K. (2012) Statistics and truth in phylogenomics. *Molecular biology and evolution*, 29, 457–472.

Lamichhaney, S., Han, F., Webster, M.T., Andersson, L., Grant, B.R. & Grant, P.R. (2017) Rapid hybrid speciation in Darwin's finches. *Science*, 359, 224–228.

Lawson, D.J., van Dorp, L. & Falush, D. (2018) A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. *Nature Communication*, 9, 3258.

Levenson, H., Hoffmann, R.S., Nadler, C.F., Deutsch, L. & Freeman, S.D. (1985) Systematics of the Holarctic chipmunks (*Tamias*). *Journal of Mammalogy*, 66, 219–224.

Levesque, D.L. & Tattersall, G.J. (2009) Seasonal changes in thermoregulatory responses to hypoxia in the Eastern chipmunk (*Tamias striatus*). *Journal of Experimental Biology*, 212, 1801–1810.

Li, H. (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*,. https://doi.org/10.48550/arXiv.1303.3997

Li, H. & Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25, 1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and S. Genome Project Data Processing (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, 25, 2078–2079.

Luo, A., Ling, C., Ho, S.Y.W. & Zhu, C.D. (2018) Comparison of methods for molecular species delimitation across a range of speciation scenarios. *Systematic Biology*, 67, 830–846.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., et al (2010) The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20, 1297–1303.

Meyer, M. & Kircher, M. (2010) Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols*, 2010:pdb prot5448.

Minh, B.Q., Nguyen, M.A. & von Haeseler, A. (2013) Ultrafast approximation for phylogenetic bootstrap. *Molecular biology and evolution*, 30, 1188–1195.

Minh, B.Q., Hahn, M.W. & Lanfear, R. (2020a) New methods to calculate concordance factors for phylogenomic datasets. *Molecular biology and evolution*, 37, 2727–2733.

Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A. & Lanfear, R. (2020b) IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular biology and evolution*, 37, 1530–1534.

Mirarab, S., Bayzid, M.S., Boussau, B. & Warnow, T. (2014a) Statistical binning enables an accurate coalescent-based estimation of the avian tree. *Science*, 346, 1250463.

Mirarab, S., Reaz, R., Bayzid, M.S., Zimmermann, T., Swenson, M.S. & Warnow, T. (2014b) ASTRAL: genome-scale coalescent-based species tree estimation. *Bioinformatics*, 30, i541–i548.

Mirarab, S., Bayzid, M.S. & Warnow, T. (2016) Evaluating summary methods for multilocus species tree estimation in the presence of incomplete lineage sorting. *Systematic Biology*, 65, 366–380.

Morales-Briones, D.F., Liston, A. & Tank, D.C. (2018) Phylogenomic analyses reveal a deep history of hybridization and polyploidy in the neotropical genus *Lachemilla* (Rosaceae). *New Phytol.*, 218, 1668–1684.

Moritz, C., Patton, J.L., Conroy, C.J., Parra, J.L., White, G.C. & Beissinger, S.R. (2008) Impact of a century of climate change on small-mammal communities in Yosemite National Park, USA. *Science*, 322, 261–264.

Nadler, C.F. & Block, M.H. (1962) The chromosomes of some North American chipmunks (Sciuridae) belonging to the genera *Tamias* and *Eutamias*. *Chromosoma*, 13, 1–15.

Nguyen, L.T., Schmidt, H.A., von Haeseler, A. & Minh, B.Q. (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular biology and evolution*, 32, 268–274.

Paradis, E., Claude, J. & Strimmer, K. (2004) APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, 20, 289–290.

Patterson, B.D. (1981) Morphological shifts of some isolated populations of *Eutamias* (Rodentia: Sciuridae) in different congeneric assemblages. *Evolution; Internation Journal of Organic Evolution*, 35, 53–66.

Patterson, B.D. (1984) Geographic variation and taxonomy of colorado and hopi chipmunks (genus *Eutamias*). *Journal of Mammalogy*, 65, 442–456.

Patterson, B.D. & Heaney, L.R. (1987) Analysis of geographic variation in red-tailed chipmunks (*Eutamias ruficaudus*). *Journal of Mammalogy*, 68, 782–791.

Patterson, B.D. & Norris, R.W. (2016) Towards a uniform nomenclature for ground squirrels: the status of the Holarctic chipmunks. *Mammalia*, 80. https://doi.org/10.1515/mammalia-2015-0004

Patterson, B.D. & Thaeler, C.S. (1982) The mammalian baculum: hypotheses on the nature of bacular variability. *Journal of Mammalogy*, 63, 1–15.

Pease, J.B. & Hahn, M.W. (2015) Detection and polarization of introgression in a five-taxon phylogeny. *Systematic Biology*, 64, 651–662.

Piaggio, A.J. & Spicer, G.S. (2000) Molecular phylogeny of the chipmunk genus *Tamias* based on the mitochondrial cytochrome oxidase subunit II gene. *J. Mammal. Evol.*, 7, 147–166.

Piaggio, A.J. & Spicer, G.S. (2001) Molecular phylogeny of the chipmunks inferred from mitochondrial cytochrome b and cytochrome oxidase II gene sequences. *Molecular Phylogenetics and Evolution*, 20, 335–350.

Quinlan, A.R. & Hall, I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26, 841–842.

Rannala, B. & Yang, Z. (2017) Efficient Bayesian species tree inference under the multispecies coalescent. *Systematic Biology*, 66, 823–842.

Reid, N., Demboski, J.R. & Sullivan, J. (2012) Phylogeny estimation of the radiation of western North American chipmunks (*Tamias*) in the face of introgression using reproductive protein genes. *Systematic Biology*, 61, 44–62.

Rubidge, E.M., Patton, J.L. & Moritz, C. (2014) Diversification of the alpine chipmunk, *Tamias alpinus*, an alpine endemic of the Sierra Nevada, California. *BMC Evol. Biol.*, 14, 1–15.

Sarver, B.A., Demboski, J.R., Good, J.M., Forshee, N., Hunter, S.S. & Sullivan, J. (2017) Comparative phylogenomic assessment of mitochondrial introgression among several species of chipmunks (*Tamias*). *Genome Biol. Evol.*, 9, 7–19.

Sarver, B.A.J., Herrera, N.D., Sneddon, D., Hunter, S.S., Settles, M.L., Kronenberg, Z., Demboski, J.R., Good, J.M. & Sullivan, J. (2021) Diversification, introgression, and rampant cytonuclear discordance in rocky mountains chipmunks (Sciuridae: *Tamias*). *Syst. Biol,*. 70, 908–921.

Sikes, R.S., Gannon, W.L., Care, T.A. & Mammalogists, U.C.O.TAS.O. (2011) Guidelines of the American Society of Mammalogists for the use of wild mammals in research. *Journal of Mammalogy*, 92, 235–253.

Simmons, L.W. (2014) Sexual selection and genital evolution. *Austral Entomol.*, 53, 1–17.

Struck, T.H., Feder, J.L., Bendiksby, M., Birkeland, S., Cerca, J., Gusarov, V.I., Kistenich, S., Larsson, K.H., Liow, L.H., Nowak, M.D., et al (2018) Finding evolutionary processes hidden in cryptic species. *Trends in Ecology & Evolution*, 33, 153–163.

Sullivan, J., Demboski, J.R., Bell, K.C., Hird, S., Sarver, B., Reid, N. & Good, J.M. (2014) Divergence with gene flow within the recent chipmunk radiation (*Tamias*). *Heredity*, 113, 185–194.

Sutton, D.A. (1982) The female genital bone of chipmunks. *Southwestern Nat.*, 27, 393–402.

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

Swofford, D.L. (2003) PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4.0b10. Sinauer Associates.

Thorington, R.W. Jr., Koprowski, J.L., Steele, M.A. & Wahatton, J.F. (2012) Squirrels of the world. The Johns Hopkins Univ. Press, Baltimore, MD.

To, T.H., Jung, M., Lycett, S. & Gascuel, O. (2016) Fast dating using least-squares criteria and algorithms. *Systematic Biology*, 65, 82–97.

Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., et al (2013) From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinform.*, 43, 11 10 11–11 10 33.

Weisenfeld, N.I., Kumar, V., Shah, P., Church, D.M. & Jaffe, D.B. (2017) Direct determination of diploid genome sequences. *Genome Research*, 27, 757–767.

White, J.A. (1953) The baculum in the chipmunks of western North America. *Univ. Kansas Publ. Museum Nat. Hist.*, 5, 611–631.

Yang, Z. (2015) The BPP program for species tree estimation and species delimitation. *Curr. Zool.*, 61, 854–865.

Yang, Z. & Rannala, B. (2010) Bayesian species delimitation using multilocus sequence data. *Proc. Natl. Acad. Sci. U S A*, 107, 9264–9269.

Yu, Y. & Nakhleh, L. (2015) A maximum pseudo-likelihood approach for phylogenetic networks. *BMC Genom.*, 16, 1–10.

Yu, Y., Degnan, J.H. & Nakhleh, L. (2012) The probability of a gene tree topology within a phylogenetic network with applications to hybridization detection. *Plos Genetics*, 8, e1002660.

Yu, Y., Dong, J., Liu, K.J. & Nakhleh, L. (2014) Maximum likelihood inference of reticulate evolutionary histories. *Proc. Natl. Acad. Sci. U S A*, 111, 16448–16453.

Zhang, C., Rabiee, M., Sayyari, E. & Mirarab, S. (2018) ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *Bmc Bioinformatics [Electronic Resource]*, 19, 153.

Zheng, X., Levine, D., Shen, J., Gogarten, S.M., Laurie, C. & Weir, B.S. (2012) A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, 28, 3326–3328.

Associate Editor: M. Kronforst
Handling Editor: A. McAdam

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Figure S1 – Maximum likelihood phylogram of concatenated dataset (5,365,556 bp) for the total Tamias tree (54 Ind.).

Figure S2 – Maximum likelihood phylogram of mitochondrial genomes (16,697 bp).

Figure S3 – ASTRAL species tree obtained from 158 gene trees without assignment of individuals to species.

Figure S4 – ASTRAL species tree obtained from 158 gene trees with assignment of individuals to species.

Figure S5 – *SVDquartets* species tree generated with 13,482 unlinked SNPs without assignment of individuals to a species.

Figure S6 – *SVDquartets* species tree generated with 13,482 unlinked SNPs with assignment of individuals to species.

Figure S7 – Maximum likelihood phylogram of concatenated dataset (7,813,766 bp) for the expanded Tamias subset tree (84 Ind.).

Figure S08 – ASTRAL species tree obtained from 161 gene trees without assignment of individuals to species.

Figure S09 – ASTRAL species tree obtained from 161 gene trees with assignment of individuals to species.

Figure S10 – *SVDquartets* species tree generated with 17,594 unlinked SNPs without assignment of individuals to a species.

Figure S11– *SVDquartets* species tree generated with 17,594 unlinked SNPs with assignment of individuals to species.

Figure S12 – BPP species tree obtained from 88 loci (∼300 kbp) with assignment of individuals to species.

Figure S13 – Maximum likelihood phylogram of mitochondrial genomes (16,697 bp).

Figure S14 – Clustering of individuals in ADMIXTURE analyses with varying number of clusters for K = 2 to K = 8.

Figure S16 –Rooted pseudo-maximum likelihood networks for our expanded subset for *cratericus*, with 0 to 5 reticulations inferred from 161 gene trees with PhyloNet.

Figure S17 – Proportion of sites supporting the three possible rooted phylogenies between the two *cratericus* lineages and *T. amoenus* (top) and *T. ruficaudus* (bottom).

Figure S18 – Concatenated ML tree from IQ-Tree with nodes labeled with species-clade comparisons for *dxy*Table shows dxy for each comparison.

Figure S19 – Chronogram of *Tamias* produced from the least square-based fast dating (LSD2) method.

Supplement Material