Genomic resolution of cryptic species diversity in chipmunks

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

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

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

[3]*Department of Biological Sciences, University of Idaho, Moscow, Idaho, USA*

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

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

[6]*Wildlife Biology Program, University of Montana, Missoula, Montana, USA.*

Correspondence: Nathanael D. Herrera ndh04c@gmail.com; Jeffrey M. Good, jeffrey.good@umontana.edu

ABSTRACT

Discovery of cryptic species is essential to understanding 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 (*T. 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.

KEYWORDS

speciation; hybridization; introgression; phylogenomics; *Tamias*; *Neotamias*

INTRODUCTION

Species are fundamental units of biodiversity, yet the basic task of species discovery often remains incomplete, even in well-studied taxonomic groups. Whereas 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 has historically relied on morphological differences. However, molecular data provide evidence that morphological divergence does not always coincide with 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 and species delimitation across a rapid diversification characterized by ecological adaptation, phenotypic

58 convergence, and overlapping geographic distributions. Of the 25 or 26 (see Burgin et al. 2018)
59 currently recognized chipmunk species (Sciuridae: *Tamias*; but see Patterson and Norris 2016),
60 two are widely distributed lineages occupying central and eastern Asia (*T. sibiricus*) or eastern
61 North America (*T. striatus*; Hall 1981). The remaining 23 species (subgenus *Neotamias*)
62 comprise a recent radiation of habitat specialists and generalists occupying the diverse
63 ecosystems of western North America (Heller 1971; Patterson 1981; Sullivan et al. 2014).
64 Specialization among chipmunks has resulted in strong niche partitioning by habitat and
65 competitive exclusion among co-distributed species (Brown 1971; Heller 1971). Although
66 ecological associations broadly tend to track species boundaries, these associations can break
67 down in sympatry, with chipmunks that occupy similar or transitory habitats showing phenotypic
68 convergence (e.g., in size and pelage coloration) between non-sister species. As a consequence,
69 internal genital bones often provide the only morphological character that is strongly diagnostic
70 of species boundaries (White 1953). Divergence in genitalia is most pronounced in the male
71 genital bone, the baculum (*os penis*), which evolves rapidly among chipmunk species (Callahan
72 1977; Sutton and Patterson 2000; Sullivan et al. 2014) likely due to strong sexual selection
73 (Eberhard 1985; Simmons 2014). Bacular divergence was long thought to underlie strong
74 reproductive barriers between chipmunk species (Patterson and Thaeler 1982). However, more
75 recent work has shown that some hybridization occurs between species (see Sullivan et al. 2014
76 for review). In most cases, levels of nuclear introgression have appeared to be relatively low
77 overall (but see Ji et al. 2021), with the persistence of genetically well-defined species
78 boundaries that largely correspond with established taxonomy based on ecological associations
79 and diagnostic genitalia (Hird and Sullivan 2009; Hird et al. 2010; Good et al. 2015; Bi et al.
80 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 (*T. minimus*). For example, *T. m. grisescens* (the Coulee chipmunk, hereafter *grisescens*) is a small, light gray subspecies of *T. minimus* that occupies a restricted range from the Channeled Scablands of central Washington that was identified by Reid and colleagues (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, *T. amoenus cratericus*, hereafter *cratericus*) as occurring on the periphery of a series of relatively recent lava flows (~10 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 (Figure 1). However, comparison of bacula by White (1953) and Sutton (1982, including the baubellum, the female genital bone; *os clitoris*), of multiple *Tamias* species concluded that *cratericus* may be more closely related to other chipmunk species (*T. umbrinus* or *T. 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

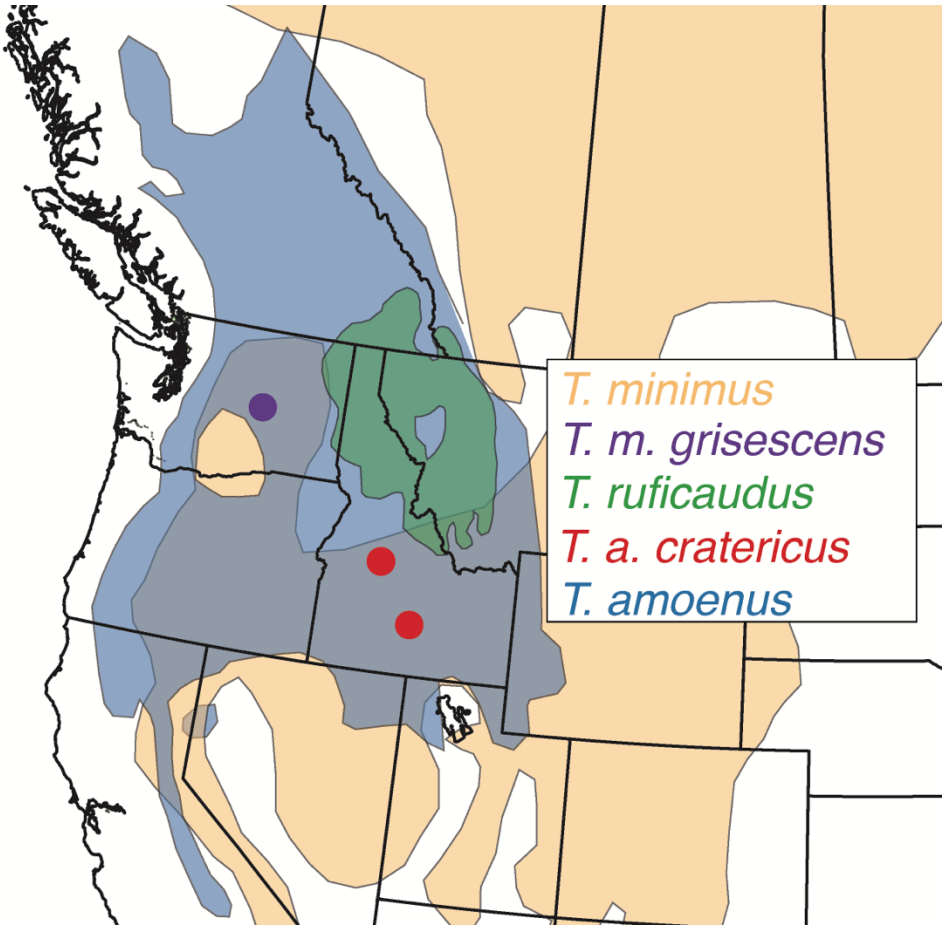

Figure 1 – Two potential cryptic species of chipmunk. *T. m. grisescens* (Purple) occupies a restricted range from the Channeled Scablands of central Washington. *T. a. cratericus* (Red) is described as a locally adapted form of the yellow-pine chipmunk (*T. amoenus*).

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, including the enigmatic *cratericus* and *grisescens* 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, *Tamias striatus*, as an outgroup (Supplementary 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 the 26 described species of chipmunk (57 individuals). Not included were samples 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. 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, conducted according to the recommendations of the American Society of Mammalogists Animal Care and Use Committee (Sikes and Mammalogists 2016).

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 100µg/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 0.4X volume of 5M ammonium acetate and 3X volume of ice-cold ethanol, washed twice with 70% ethanol, and resuspended in buffer (10mM 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 10X 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 10X Genomics Chromium library sequence data using default settings with --maxreads=1135000000 (the number of reads estimated to yield 56× coverage for an estimated genome of ~2.3 Gigabases; Supplementary 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; Supplemental 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

158 et al. 2012). These transcriptome data were used to develop an array-based capture targeting
159 ~8,000 nuclear genes as well as the complete mitochondrial genome, resulting in ~6.9 Mb of
160 assembled exonic regions (including flanking introns and intergenic regions; Bi et al. 2012; Bi et
161 al. 2013; Good et al. 2015). Finally, an additional ~2.4 Mb of genic regions were identified from
162 the original transcriptome through AmiGO and NCBI protein databases and were added to the
163 capture design. Published data from Sarver et al. (2021; *T. canipes*, *T. cinereicollis*, *T. dorsalis*,
164 *T. quadrivitattus*, *T. rufus*, and *T. umbrinus*) were based on the earlier array-based design and
165 thus represent an overlapping subset of the genomic regions targeted here.
166 To localize the targeted genomic regions within the *de novo T. minimus* genome, we first
167 used BLAT to identify the best hit in the reference. We then sorted and filtered the contigs to
168 include only the best match hits, discarding contigs with equally likely hits across the genome.
169 To identify X-linked contigs in the genome assembly, we calculated the normalized female
170 versus male ratio of sequencing coverage for targeted regions mapping to each contig, which
171 should be approximately 2:1 for genes on the X chromosome.
172 Total genomic DNA was isolated from heart or liver tissue using Qiagen DNEasy DNA
173 blood and tissue kits. Illumina sequencing libraries were prepared for each sample following the
174 protocol of Meyer and Kircher (2010). Sequencing libraries were then enriched in solution using
175 probes designed as described above and manufactured by NimbleGen (SeqCap EZ Developer
176 kits). Hybridization reactions were performed in four separate equimolar pools of indexed
177 libraries. To assess capture efficiency, we first sequenced a subset of pooled libraries on an
178 Illumina MiSeq sequencer (150 bp paired-end reads) at the University of Montana genomics core
179 (Missoula, MT). Final pooled libraries were sequenced on an Illumina HiSeq 2500 (150 bp
180 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; last 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 v0.7.15 (Li and Durbin 2009; Li 2013) with default options and then sorted with *samtools* v1.4 (Li et al. 2009). Mapped reads were assigned read groups and duplicate reads were identified post-mapping using Picard v2.5.0 (http://broadinstitute.github.io/picard/). Initial capture efficiency statistics were calculated using QC3 v1.33 (Guo et al. 2014). Regions with insertions or deletions (indels) were identified and realigned, and single nucleotide variants (SNVs) were called using UnifiedGenotyper within the Genome Analysis Toolkit (GATK) v3.6 (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 v2.25 (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 v.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 v. 2.0.6 (Minh et al. 2020a; Minh et al. 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 v2.2.5 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

227 informative sites to infer independent phylogenetic histories. Although interval size is somewhat
228 arbitrary here, we considered a range of interval sizes and found smaller intervals (≤ 10 kb)
229 resulted in many intervals without targets and few informative sites. Likewise, much larger
230 interval sizes (≥ 100 kb) were more likely to conflate independent phylogenetic histories. For
231 each interval, we filtered positions with missing data for >85% of individuals using AMAS and
232 excluded intervals containing less than 1 kb of capture data. We then used IQ-Tree v.1.5.5 to
233 estimate local maximum likelihood trees with 1000 ultra-fast bootstraps (using -m MFP, -bb
234 1000), with *T. striatus* as the outgroup. Gene trees, when necessary, were unrooted using the R
235 package *ape* (Paradis et al. 2004). We then used unrooted gene trees to estimate a consensus
236 species tree using ASTRAL-III v5.6.3 (Zhang et al. 2018). Assuming sets of independent and
237 accurately estimated gene trees, ASTRAL estimates an unrooted species tree by finding the
238 species tree that has the maximum number of shared induced quartet trees with the given set of
239 gene trees (Mirarab et al. 2014a; Mirarab et al. 2014b).
240 Next, we estimated coalescent-based species trees with SVDquartets (Chifman and
241 Kubatko 2014) using 100,000 random quartets and 100 bootstrap replicates as implemented in
242 PAUP* v4.0a166 (Swofford 2003). SVDquartets assesses support for quartets of taxa directly
243 from site-pattern frequencies of variable sites only. This approach differs from summary
244 methods, such as ASTRAL, because it does not independently estimate gene trees, avoiding the
245 issue of gene-tree estimation error (Chifman and Kubatko 2015; Chou et al. 2015). For
246 SVDquartets, we extracted single nucleotide variants (SNVs) distanced at least 10 kb along the
247 genome and excluded sites with missing information for >85% of the individuals. For both
248 SVDquartets and ASTRAL analyses, species trees were estimated with and without assigning
249 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 α=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 [~IG (3, 0.002) with mean 0.00], and to the divergence times τ [~ IG (3, 0.01) with mean 0.005 for the age of the root]. We assessed convergence based on consistency across 5 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. While BPP can potentially analyze 1000's 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 ~1,000-4,600 bp; median 2,343 bp) for a total alignment length of 209,725 bp.

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 v.1.30 (Alexander et al. 2009). We tested values of K (the number of population clusters) ranging from 1-6 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 v1.9 (Chang et al. 2015) to minimize non-independence 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 multi-species 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 towards 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 to 5 migration events, associated individuals to species

296 (option -a), and optimized branch lengths and inheritance probabilities to compute likelihoods
297 for each proposed network (option -o). We calculated Akaike information criteria corrected for
298 small sample sizes (AICc) and Bayesian Information Criteria (BIC) using the highest likelihoods
299 per run to compare the resulting networks (Yu et al. 2012; Yu et al. 2014). Networks were
300 visualized with IcyTree (https://icytree.org; last accessed Dec 2020).
301

302 # RESULTS

303 ## DRAFT GENOME ASSEMBLY AND SEQUENCE CAPTURE

304 We generated ~217 Gb of *T. minimus* genomic sequence data using 10X Genomics technology
305 (Pleasanton, CA, USA) and assembled a 2.48 Gb draft genome with a contig N50 of 196.09 kb
306 and as scaffold N50 of 58.28 Mb, respectively (Supplementary Table S2). We then analyzed up
307 to 9.4 Mb of sequence capture data from 121 individuals, combining newly generated (112
308 individuals from 15 species) and published data (nine individuals from seven species; Bi et al.
309 2019; Sarver et al. 2021; Supplementary Table S1) for 21 western chipmunk species and the
310 outgroup *T. striatus.* Sequencing efforts produced ~4 million reads per sample with an average of
311 0.5% of targets showing no coverage. Approximately 75% of raw reads were unique, resulting in
312 an average target coverage of 33× across samples (range: 5–91×) with ~73% of targeted bases
313 sequenced to at least 10× coverage (Supplementary Table S3). Comparison of male versus
314 female coverage resulted in the identification of three contigs (79.7 Mb, 35.6 Mb, 5.9 Mb) in the
315 genome assembly that are likely on the X chromosome.
316

317 ## 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 (Figure 2A; Supplemental Figure S1-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; Figure 2A). ML reconstruction of the mitochondrial genome resulted in a different topology from the nuclear dataset (Supplemental Figure S2) and was largely concordant with previous studies based on the cytochrome *b* gene (Demboski and Sullivan 2003; Good et al. 2003; Good et al. 2008; Reid et al. 2012; Good et al. 2015; 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; Supplemental Figure 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

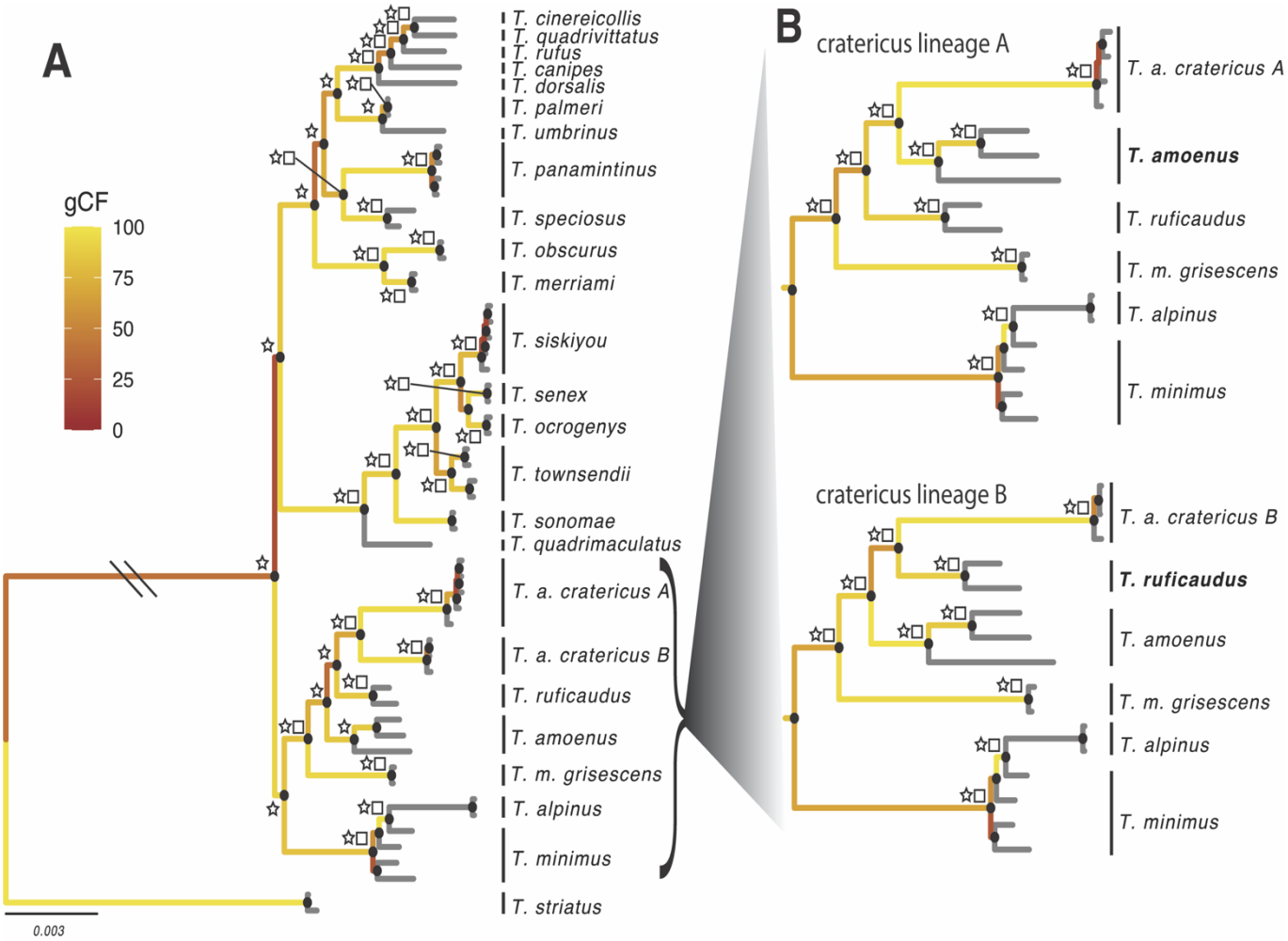

Figure 2 – Genome-wide phylogeny of *Tamias*. A) Concatenated ML tree estimated with IQ-Tree; Node labels represent: ASTRAL posterior probability equal to one (star), bootstrap greater than 90 (square), and bootstrap proportions greater than 90 (grey circles). Branches are colored according to gene concordance factor (gCF). **B)** ML reconstruction of the *amoenus-ruficaudus-cratericus* clade with the exclusion of *cratericus* lineage and *cratericus* lineage respectively.

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 (Supplemental Figures S3-S6). Overall, the splits between *T. amoenus-cratericus-T. ruficaudus* were not well-supported and remained unresolved (Figure 2A; Supplemental Figure 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 (Figure 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; Figure 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. *T. minimus* (minus *grisescens*) was monophyletic, with *T. m. grisescens* as 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

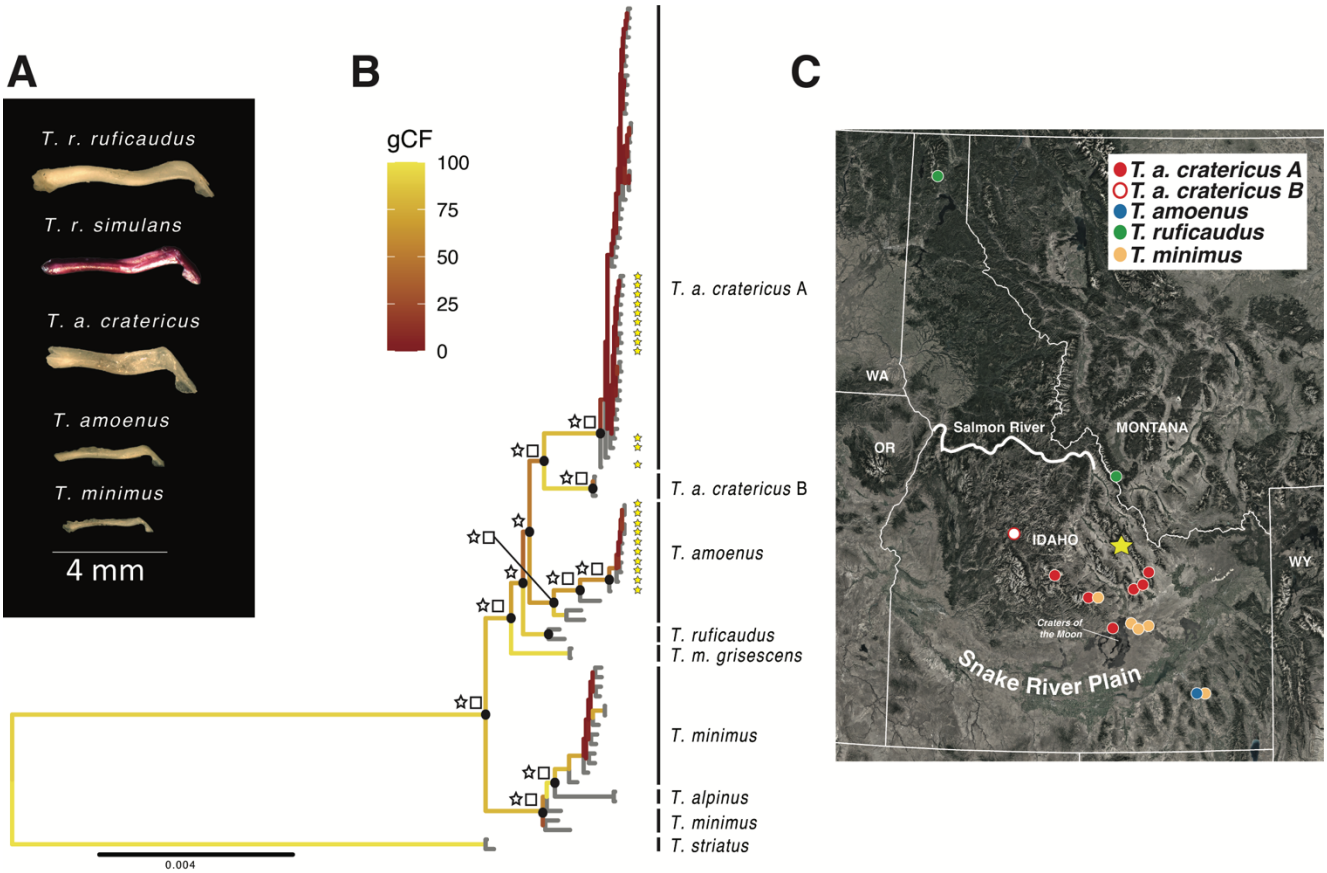

Figure 3 – ML concatenated phylogeny for *cratericus-amoenus-ruficaudus*. A) Representative bacular morphologies for the five taxa. **B)** Concatenated ML tree estimated with IQ-Tree for the full population level sampling of *cratericus*; Node labels represent: ASTRAL posterior probability equal to one (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 diversity detected within *cratericus* lineage A, which was found at 6 additional localities to the near exclusion of *T. amoenus*. We again recovered a deep split within *cratericus*, with lineage B still represented by only a single locality (Figure 3; Supplemental Figure 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* (Figure 3B; Supplemental Figures S10-13). One potential source of phylogenetic discordance between the datasets may be driven by samples from one locality, known as Iron Creek (IC; yellow star; Figure 3C). At this

377 locality, both *cratericus* and standard *T. amoenus* bacular types were found to co-occur without
378 other obvious phenotypic or habitat differences. This was the only site where both forms were
379 found to be sympatric. Collectively, these results suggest that *cratericus* is distributed across the
380 coniferous forest of south central Idaho, north of the Snake River Plain to the near exclusion of
381 *T. amoenus*, save for one sympatric locality. At this site, we observed the maintenance of
382 genetically defined species boundaries despite some evidence for admixture between *cratericus*
383 and *T. amoenus* (see below).

384 Next, we evaluated if the change in the relationship between *T. ruficaudus- cratericus-T.*
385 *amoenus* was influenced by possible gene flow at the sympatric IC locality. Consistent with this,
386 *cratericus* was again sister to *T. ruficaudus* in a concatenated analysis when excluding all IC
387 samples. However, the species trees estimated from ASTRAL (158 genes trees estimated from
388 50 kb genomic intervals) and SVDquartets (17,594 unlinked SNPs) for the expanded subset
389 resulted in *cratericus* being sister to *T. amoenus*, regardless of whether IC individuals were
390 included (Supplemental Figures S10-13). Branch support increased when IC samples were
391 excluded but we still found a large degree of incongruence based on gCF and sCF. Finally, the
392 results of our BPP analyses were largely consistent across the 5 replicate runs using different
393 starting species trees, with 4 of the 5 runs converging on the same maximum *a posteriori*
394 probability (MAP) tree, with a posterior probability of ~100%. BPP recovered *cratericus* as
395 being sister to *T. ruficaudus* (Supplemental Figure S9).

396 ML reconstruction of mitochondrial genomes from the expanded subset also showed
397 evidence for mtDNA introgression between *cratericus* and *T. amoenus* (Supplemental Figure
398 S8). As with the genus-wide analysis, *cratericus* and *T. amoenus* were paraphyletic, but the
399 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; Supplemental Figure S8).

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 split *T. amoenus* and *cratericus*, with *T. ruficaudus* falling intermediate. Cross-validation with ADMIXTURE suggested five population clusters that were largely concordant with the PCA (Figure 4B; Supplemental Figures S14-15), although we note that this analysis is not well-suited to partition genetic clusters at this scale of interspecific divergence (Lawson et al. 2018). 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 affected the inferred split between *T. amoenus*-*cratericus*-*T. ruficaudus*. The best supported network analysis (5 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* (Supplemental Figure S16). Quartet-based analyses of introgression resulted in three significant triplet comparisons supporting hybridization (Supplemental 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 γ =0.344; *T. amoenus*-*cratericus* lineage B γ = 0.541). Under the ABBA-BABA framework, we evaluated phylogenetic patterns of shared SNVs 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

442 extent *T. ruficaudus* relative to *cratericus* lineage B (Supplemental Figure S17; Supplemental

443 Table S4-S5).

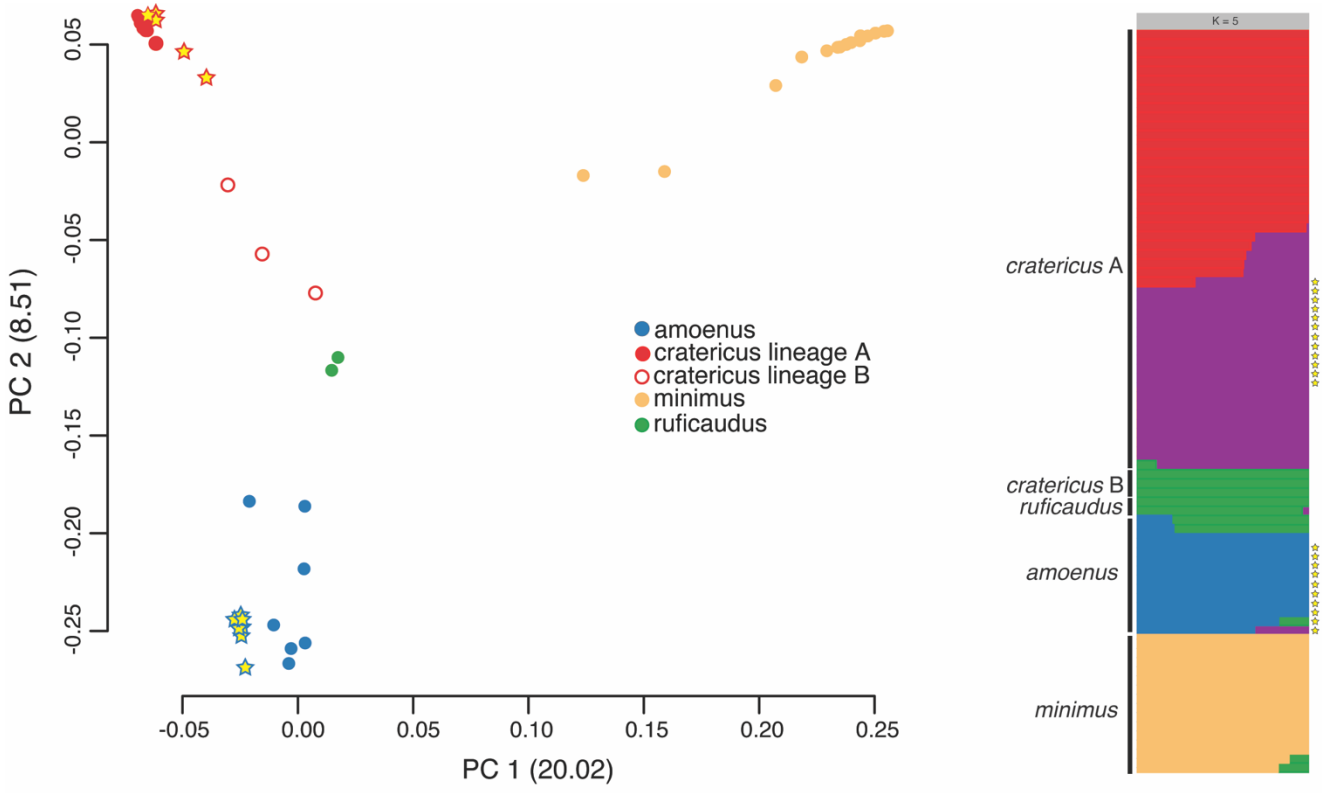

Figure 4 –Population structure between *cratericus-amoenus-ruficaudus* sampled. A) PCA results from SNPRelate. **B)** Admixture output for highest marginal-likelihood run at k=5. Stars indicate samples from the sympatric locality Iron Creek.

444 ## DISCUSSION

445 Discovery of cryptic species is fundamental to understanding the process of speciation and

446 cataloging global diversity, which is facing unprecedented pressures due to the emerging threats

447 of climate change and habitat loss. We generated a *de novo* assembled reference genome for a

448 western chipmunk species, *Tamias minimus*, and used new and published genome-wide targeted

449 capture sequencing data and whole mitochondrial genomes to test for cryptic speciation and

450 hybridization in western chipmunks. Our genus-wide analysis represents the most

451 comprehensive phylogenetic assessment of the western chipmunks to date and, we suggest,

452 reveals at least two new undescribed chipmunk species. Below we discuss the evolutionary

453 implications of our findings, focusing on western chipmunk diversity and the complex speciation
454 dynamics of the Crater chipmunk and other co-distributed species.
455
456 **SYSTEMATICS OF THE WESTERN CHIPMUNK RADIATION**
457 Due to their abundance and conspicuous nature, chipmunks have been the focus of a broad range
458 of biological studies including physiology (e.g., Levesque and Tattersall 2009), niche
459 partitioning and competitive exclusion (e.g., Grinnell and Storer 1924; Heller 1971), behavior
460 (e.g. Broadbrooks 1970), and responses to climate change (e.g., Moritz et al. 2008; Bi et al.
461 2019). Early phylogenetic studies using morphology and allozymes (Levenson et al. 1985),
462 host/ectoparasite (Jameson 1999), chromosomal (Nadler and Block 1962) and molecular data
463 (Piaggio and Spicer 2000; Piaggio and Spicer 2001) recognized three distinct clades within
464 *Tamias*: *T. sibiricus*, *T. striatus*, and the western North America species (arbitrarily classified as
465 either subgenera or genera; (e.g., Jameson 1999; Patterson and Norris 2016). Relationships
466 among the 23 currently recognized species within the western chipmunk clade (subgenus
467 *Neotamias*) have remained somewhat obscure due to a lack of resolution among internal nodes
468 and apparently rampant mitochondrial introgression (Piaggio and Spicer 2000; Reid et al. 2012;
469 Sullivan et al. 2014). Our analysis of western chipmunks represents the most comprehensive
470 genome-wide assessment of the group to date and included 22 of the 25 recognized species of
471 *Tamias*.
472 We recovered strongly supported, monophyletic species groups within the western
473 chipmunks that were in general agreement with some other studies (e.g., Reid et al. 2012).
474 However, one notable difference was the relationship between *T. ruficaudus*-*T. amoenus*-*T.*
475 *minimus*. In contrast to Reid et al. (2012), we recovered *T. ruficaudus* as being sister to either

476 *cratericus*, or to a clade composed of *T. amoenus-cratericus* (see discussion below) whereas
477 Reid et al. (2012) found *T. ruficaudus* sister to the enigmatic, and similarly distinct lineage
478 *grisescens.* In addition, Reid et al. (2012) also recovered *T. amoenus* as sister to *T. minimus*,
479 whereas we found *T. minimus* as the sister lineage to the main *T. amoenus-T. ruficaudus-*
480 *cratericus-grisescens* clade. Given a lack of resolution among internal nodes in previous studies,
481 this difference likely reflects the power afforded by much more extensive genetic sampling.
482 A major conclusion of our genomic study was the resolution of two cryptic lineages of
483 chipmunk. While some of the relationships among described species have been unclear until
484 recently, the identity of species as the fundamental units of diversity have remained fairly stable
485 over the last century. The vast majority of chipmunk species were described to some degree by
486 the late 19th and early 20th Centuries by the early naturalists Allen, Merriam, and others based
487 on phenotypic and ecological characteristics (for a review see Thorington Jr. et al. 2012).
488 Subsequent discussions have focused on determining if subspecific variation within some species
489 warrant species-level recognition (Patterson 1984; Patterson and Heaney 1987; Good et al. 2003)
490 or if some current species are actually geographically isolated populations of more widespread
491 forms (Piaggio and Spicer 2000; Piaggio and Spicer 2001; Rubidge et al. 2014). The putative
492 *grisescens* lineage reflects these dynamics. Howell (1925) described the distinctive *T. m.*
493 *grisescens* based on its pelage as a least chipmunk subspecies restricted to the Channeled
494 Scablands of central Washington. Reid et al. (2012) then identified this taxon as a potentially
495 cryptic lineage and found it consistently nested with *T. amoenus* based on limited nuclear and
496 mtDNA data. Although a formal description awaits additional geographic sampling, our data
497 clearly show that *grisescens* does not fall within the considerable genetic diversity of the least
498 chipmunk (*T. minimus*) or other sampled species, and we propose that *grisescens* is in fact a

499 novel, cryptic chipmunk species (i.e., *Tamias grisescens* following Howell 1925). Of note,
500 *grisescens* is much more genetically distinct from other chipmunk species than the Alpine
501 chipmunk (*T. alpinus*); a lineage that is paraphyletic with respect to *T. minimus* but has long been
502 recognized as a distinct species based on phenotypic divergence, ecological differentiation, and
503 reproductive isolation (Grinnell & Storer; Heller; Rubidge 2014; Bi et al. 2019). We also found
504 *grisescens* nested within *T. amoenus* in our mtDNA analyses, suggesting a history of mtDNA
505 introgression or incomplete lineage sorting (Supplemental Figure S3). While our study lacks the
506 sampling needed to fully investigate the evolutionary history of *grisescens* (two individuals from
507 the same locality), its range in central Washington could contribute to vulnerability due to habitat
508 loss, wildfire, and drought due to rapid climate change.

509 We also show that *cratericus* is likely a distinct chipmunk species that appears to be most
510 closely related to *T. ruficaudus*. Blossom (1937) first described *cratericus* as a duller, dark
511 variant of the yellow-pine chipmunk that is associated with the recent volcanic lava flows of
512 Craters of the Moon. The color morphology of *cratericus* rapidly transitions to a more brightly
513 colored pelage morph more like standard *T. amoenus* phenotypes and, we show, is found
514 throughout the adjacent xeric forest habitats of the region. This localized pelage presumably
515 reflects recent adaptation for crypsis on the black lava flows, a pattern that is also observed in
516 local populations of the Great Basin pocket mouse (*Perognathus parvus*) and pika (*Ochotona*
517 *princeps)* from Craters of the Moon (Blossom, 1937). More surprisingly, *cratericus* appears to be
518 the locally dominant form throughout central Idaho (see below).

519 Overall, we also found that most western chipmunk groups displayed a high degree of
520 discordance between the mtDNA phylogeny and the nuclear genome, consistent with previous
521 studies documenting recurrent mtDNA introgression layered across the history of this group

(e.g., Demboski & Sullivan, 2003; Sullivan et al. 2014; Sarver et al. 2017). 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 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, our results suggest *cratericus* is most closely related to *T. ruficaudus* albeit with low support (Figure 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 (Figure 2; Supplemental Figures S3-S6). In contrast, while we found strong support with BPP, these analyses were limited to a relatively small subset of data which 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, Supplemental Figure 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.

While 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* (Figure 2; Supplemental Figures S8-S9). Further,

568 we inferred multiple reticulations across the *amoenus-cratericus-ruficaudus* splits, suggesting
569 there has been both recent and ancient hybridization between these lineages during their
570 diversification. Both contemporary and past hybridization events have resulted in extensive
571 shared polymorphism among the three lineages (Supplemental Figures. S16-S17) and highlight
572 how rapid diversification and introgressive hybridization can confound our ability to infer a
573 complex evolutionary history among species.

574 Integrating our phylogenetic and population genomic data, we propose a working model
575 for the evolutionary history of *cratericus* (Figure 5). There was an initial (likely geographic) split
576 between *cratericus* and other *T. ruficaudus*, followed by intermittent hybridization between
577 ancestral populations of *cratericus, T. ruficaudus,* and the slightly more distantly related *T.*
578 *amoenus.* Determining if the apparent genetic split within *cratericus* is associated with
579 reproductive isolation between these lineages or divergence in other morphological or ecological
580 traits awaits further sampling. While lineage A appears to be the widespread and predominant
581 form, the rarity of lineage B may be due to a lack of sampling in the presumed northern range of
582 *cratericus.* We also detected gene flow between *T. amoenus* and the more widely distributed
583 *cratericus* lineage A (Figures 3-4). While a history of recurrent hybridization likely skews
584 overall phylogenetic relationships, the sympatric population (IC) allows us to test key predictions
585 about the extent of reproductive isolation between these ecologically and phenotypically similar
586 species. In this area of sympatry, individuals clearly clustered within *cratericus* lineage A or with
587 other *T. amoenus*. Thus, species boundaries appear to be largely maintained, despite some level
588 of gene flow between *cratericus* and *T. amoenus* (Figures 3-4; Supplemental Figure S14),
589 consistent with a semipermeable species boundary (Harrison and Larson 2014). Our results also
590 indicate that *cratericus* is the predominant chipmunk species associated with xeric forests of

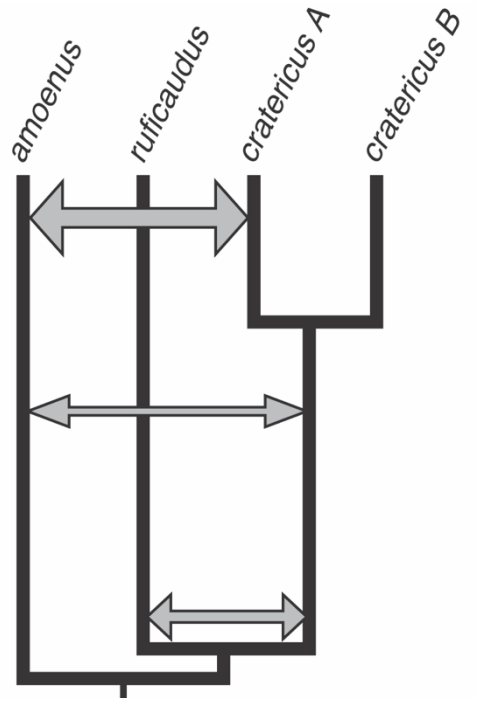

Figure 5 – Proposed species tree and evolutionary history of *T. cratericus*. *T. cratericus* has a complex evolutionary history punctuated by layered ancient and contemporary gene flow between co-distributed species. Arrow thickness indicates inferred amount of introgression.

591 south-central Idaho north of the Snake River Plain and south of the Snake River. This previously
592 unknown lineage appears to exist largely to the exclusion *T. amoenus.* Since the foundational
593 work of Grinnell a century ago (Grinnell and Storer 1924), western chipmunks have been
594 considered exemplars of ecological niche partitioning. Given the lack of obvious ecological and
595 phenotypic differentiation between *cratericus* and *T. amoenus*, sympatric populations between
596 these species may provide a compelling eco-evolutionary context to understand both the local
597 maintenance of species boundaries and niche partitioning through competitive exclusion or other
598 mechanisms.

599 Finally, our current sampling is not circumscribed by clear geographic barriers to the
600 north and west, indicating the full range of *cratericus* may be considerably larger and likely
601 extends into the remote wilderness tracks of central Idaho. Although these portions of the range

602 are more difficult to access, this diurnal cryptic species is also highly abundant and readily
603 visible around the visitor center of a National Monument that hosts ~200,000 annual visitors –
604 underscoring that cryptic biodiversity may persist even in highly conspicuous systems.
605
606 **Conflict of Interest:** The authors declare no conflict of interest.
607
608 **Data Archive:** Read sequence data are available for download at SRA under the BioProject
609 accession number XXXX.
610
611 **Author contributions:** NDH, JRD, JS, KCB, and JMG conceived and designed the study. JMG,
612 JRD, and JS acquired funding. NDH, KCB, JRD, and JMG conducted fieldwork. CMC, EN, and
613 NDH generated sequence data. NDH conducted data analyses, with guidance from BAJS and
614 JMG. BAJS contributed analytical tools. All authors discussed the results. NDH and JMG wrote
615 the manuscript with feedback from all authors.
616
617 **Acknowledgements:** Our research would not be possible without the irreplaceable support of
618 natural history museums. We are grateful to the collections staff of The Denver Museum of
619 Nature & Science and Joseph A. Cook and the University of New Mexico Museum of Southwest
620 Biology for providing tissue loans. We also thank Michael Fazekas, Roger Rodriguez, Patricia
621 McDonald, Randle McCain, Bryan McLean, Schuyler Liphardt, and Lois Alexander for help
622 with fieldwork. We thank David Xing, Jessi Kopperdahl, Mickael Fazekas, and Sara Keeble for
623 assisting with molecular work. We thank members of the Good lab, the University of Montana
624 UNVEIL network, Ke Bi, and Craig Moritz for helpful discussions. Funding support for this

research was provided by a grant from the National Science Foundation (NSF) EPSCoR (OIA-1736249 to JMG), NSF (DEB-0716200 to JRD), a travel grant from the Drollinger-Dial Foundation, the Gordon and Betty Moore Foundation (GBMF2983), the Rose Community Foundation, and research funds from the University of Montana and Denver Museum of Nature & Science. This study included research conducted in the University of Montana Genomics Core, supported by a grant from the M. J. Murdock Charitable Trust (to JMG). Computational resources and support from the University of Montana's Griz Shared Computing Cluster (GSCC), supported by grants from the Nation Science Foundation (CC-2018112 and OAC-1925267), contributed to this research. The DNA isolation, library preparation, and sequencing of the draft *Tamias minimus* genome was carried out at the DNA Technologies and Expression Analysis Core at the UC Davis Genome Center, supported by NIH Shared Instrument Grant 1S10OD010786-01.

LITERATURE CITED

Alexander, D. H., J. Novembre, and K. Lange. 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Resources 19:1655-1664.

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

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

Bickford, D., D. J. Lohman, N. S. Sodhi, P. K. Ng, R. Meier, K. Winker, K. K. Ingram, and I. Das. 2007. Cryptic species as a window on diversity and conservation. Trends Ecol Evol 22:148-155.

Blischak, P. D., J. Chifman, A. D. Wolfe, and L. S. Kubatko. 2018. HyDe: A Python Package for Genome-Scale Hybridization Detection. Syst Biol 67:821-829.

Blossom, P. M. 1937. Description of a race of chipmunk from south central Idaho. Occasional Papers of the Museum of Zoology: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 Chipmun k (*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., J. P. Colella, P. L. Kahn, and N. S. Upham. 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., C. C. Chow, L. C. Tellier, S. Vattikuti, S. M. Purcell, and J. J. Lee. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 4:7.

Chifman, J. and L. Kubatko. 2014. Quartet inference from SNP data under the coalescent model. Bioinformatics 30:3317-3324.

Chifman, J. and L. Kubatko. 2015. Identifiability of the unrooted species tree topology under the coalescent model with time-reversible substitution processes, site-specific rate variation, and invariable sites. J Theor Biol 374:35-47.

Chou, J., A. Gupta, S. Yaduvanshi, R. Davidson, M. Nute, S. Mirarab, and T. Warnow. 2015. A comparative study of SVDquartets and other coalescent-based species tree estimation methods. BMC Genomics:1-11.

de Queiroz, K. 2007. Species concepts and species delimitation. Syst Biol 56:879-886.

Delic, T., P. Trontelj, M. Rendos, and C. Fiser. 2017. The importance of naming cryptic species and the conservation of endemic subterranean amphipods. Sci Rep 7:3391.

Demboski, J. R., and J. Sullivan. 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., E. Banks, R. Poplin, K. V. Garimella, J. R. Maguire, C. Hartl, A. A. Philippakis, G. del Angel, M. A. Rivas, M. Hanna, A. McKenna, T. J. Fennell, A. M. Kernytsky, A. Y. Sivachenko, K. Cibulskis, S. B. Gabriel, D. Altshuler, and M. J. Daly.

2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 43:491-498.

Durand, E. Y., N. Patterson, D. Reich, and M. Slatkin. 2011. Testing for ancient admixture between closely related populations. Mol Biol Evol 28:2239-2252.

Eberhard, W. G. 1985. Sexual Selection and Animal Genitalia. Harvard University Press, Cambridge, MA and London, England.

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

Fišer, C., C. T. Robinson, and F. Malard. 2018. Cryptic species as a window into the paradigm shift of the species concept. Mol Ecol 27:613-635.

Flouri, T., X. Jiao, B. Rannala, and Z. Yang. 2018. Species Tree Inference with BPP Using Genomic Sequences and the Multispecies Coalescent. Mol Biol Evol 35:2585-2593.

Fujita, M. K., A. D. Leache, F. T. Burbrink, J. A. McGuire, and C. Moritz. 2012. Coalescent-based species delimitation in an integrative taxonomy. Trends Ecol Evol 27:480-488.

Good, J. M., J. R. Demboski, D. W. Nagorsen, and J. Sullivan. 2003. Phylogeography and Introgressive Hybridization: Chipmunks (Genus *Tamias*) in the Northern Rocky Mountains. Evolution 57:1900-1916.

Good, J. M., S. Hird, N. Reid, J. R. Demboski, S. J. Steppan, T. R. Martin-Nims, and J. Sullivan. 2008. Ancient hybridization and mitochondrial capture between two species of chipmunks. Mol Ecol 17:1313-1327.

Good, J. M., D. Vanderpool, S. Keeble, and K. Bi. 2015. Negligible nuclear introgression despite complete mitochondrial capture between two species of chipmunks. Evolution 69:1961-1972.

Green, R. E., J. Krause, A. W. Briggs, T. Maricic, U. Stenzel, M. Kircher, N. Patterson, H. Li, W. Zhai, M. H. Fritz, N. F. Hansen, E. Y. Durand, A. S. Malaspinas, J. D. Jensen, T. Marques-Bonet, C. Alkan, K. Prufer, M. Meyer, H. A. Burbano, J. M. Good, R. Schultz, A. Aximu-Petri, A. Butthof, B. Hober, B. Hoffner, M. Siegemund, A. Weihmann, C. Nusbaum, E. S. Lander, C. Russ, N. Novod, J. Affourtit, M. Egholm, C. Verna, P. Rudan, D. Brajkovic, Z. Kucan, I. Gusic, V. B. Doronichev, L. V. Golovanova, C. Lalueza-Fox, M. de la Rasilla, J. Fortea, A. Rosas, R. W. Schmitz, P. L. F. Johnson, E. E. Eichler, D. Falush, E. Birney, J. C. Mullikin, M. Slatkin, R. Nielsen, J. Kelso, M. Lachmann, D. Reich, and S. Paabo. 2010. A draft sequence of the Neandertal genome. Science 328:710-722.

Grinnell, J., and T. I. Storer. 1924. Animal Life in the Yosemite. An Account of the Mammals, Birds, Reptiles, and Amphibians in a Cross-section of the Sierra Nevada. University of California Press, Berkeley, CA. USA.

Guo, Y., S. Zhao, Q. Sheng, F. Ye, J. Li, B. Lehmann, J. Pietenpol, D. C. Samuels, and Y. Shyr. 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, USA.

Harrison, R. G., and E. L. Larson. 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., N. Reid, J. Demboski, and J. Sullivan. 2010. Introgression at differentially aged hybrid zones in red-tailed chipmunks. Genetica 138:869-883.

Hird, S. and J. Sullivan. 2009. Assessment of gene flow across a hybrid zone in red-tailed chipmunks (*Tamias ruficaudus*). Mol Ecol 18:3097-3109.

Hoang, D. T., O. Chernomor, A. v. Haeseler, B. Q. Minh, and L. S. Vinh. 2017. UFBoot2: Improving the Ultrafast Bootstrap Approximation. Molecular Biology and Evolution 35:518-522.

Howell, A. H. 1925. Preliminary Descriptions of Five New Chipmunks from North America. Journal of Mammalogy 6:51-54.

Jain, M., S. Koren, K. H. Miga, J. Quick, A. C. Rand, T. A. Sasani, J. R. Tyson, A. D. Beggs, A. T. Dilthey, I. T. Fiddes, S. Malla, H. Marriott, T. Nieto, J. O'Grady, H. E. Olsen, B. S. Pedersen, A. Rhie, H. Richardson, A. R. Quinlan, T. P. Snutch, L. Tee, B. Paten, A. M. Phillippy, J. T. Simpson, N. J. Loman, and M. Loose. 2018. Nanopore sequencing and assembly of a human genome with ultra-long reads. Nat Biotechnol 36:338-345.

Jameson, E. W. 1999. Host-ectoparasite Relationships among North American Chipmunks. Acta Theriologica 44.

Ji, J., D. J. Jackson, A. D. Leaché, and Z. Yang. 2021. Significant cross-species gene flow detected in the *Tamias quadrivittatus* group of North American chipmunks. BioRxiv.

Kalyaanamoorthy, S., B. Q. Minh, T. K. F. Wong, A. von Haeseler, and L. S. Jermiin. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods 14:587-589.

Kubatko, L. S. and J. Chifman. 2019. An invariants-based method for efficient identification of hybrid species from large-scale genomic data. BMC Evol Biol 19:112.

Kumar, S., A. J. Filipski, F. U. Battistuzzi, S. L. Kosakovsky Pond, and K. Tamura. 2012. Statistics and truth in phylogenomics. Mol Biol Evol 29:457-472.

Lamichhaney, S., F. Han, M. T. Webster, L. Andersson, B. R. Grant, and P. R. Grant. 2017. Rapid Hybrid Speciation in Darwin's Finches. Science 359:224-228.

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

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

Levesque, D. L., and G. J. Tattersall. 2009. Seasonal changes in thermoregulatory responses to hypoxia in the Eastern chipmunk (*Tamias striatus*). J Exp Biol 212:1801-1810.

Li, H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv.

Li, H., and R. Durbin. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754-1760.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and S. Genome Project Data Processing. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078-2079.

Luo, A., C. Ling, S. Y. W. Ho, and C. D. Zhu. 2018. Comparison of Methods for Molecular Species Delimitation Across a Range of Speciation Scenarios. Syst Biol 67:830-846.

McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, and M. A. DePristo. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20:1297-1303.

773 Meyer, M., and M. Kircher. 2010. Illumina sequencing library preparation for highly multiplexed
774 target capture and sequencing. Cold Spring Harb Protoc 2010:pdb prot5448.
775 Minh, B. Q., M. W. Hahn, and R. Lanfear. 2020a. New Methods to Calculate Concordance
776 Factors for Phylogenomic Datasets. Mol Biol Evol 37:2727-2733.
777 Minh, B. Q., M. A. Nguyen, and A. von Haeseler. 2013. Ultrafast approximation for
778 phylogenetic bootstrap. Mol Biol Evol 30:1188-1195.
779 Minh, B. Q., H. A. Schmidt, O. Chernomor, D. Schrempf, M. D. Woodhams, A. von Haeseler,
780 and R. Lanfear. 2020b. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic
781 Inference in the Genomic Era. Mol Biol Evol 37:1530-1534.
782 Mirarab, S., M. S. Bayzid, B. Boussau, and T. Warnow. 2014a. Statistical binning enables an
783 accurate coalescent-based estimation of the avian tree. Science 346:1250463.
784 Mirarab, S., M. S. Bayzid, and T. Warnow. 2016. Evaluating Summary Methods for Multilocus
785 Species Tree Estimation in the Presence of Incomplete Lineage Sorting. Syst Biol
786 65:366-380.
787 Mirarab, S., R. Reaz, M. S. Bayzid, T. Zimmermann, M. S. Swenson, and T. Warnow. 2014b.
788 ASTRAL: genome-scale coalescent-based species tree estimation. Bioinformatics 30:
789 i541-548.
790 Morales-Briones, D. F., A. Liston, and D. C. Tank. 2018. Phylogenomic analyses reveal a deep
791 history of hybridization and polyploidy in the Neotropical genus *Lachemilla* (Rosaceae).
792 New Phytol 218:1668-1684.
793 Moritz, C., J. L. Patton, C. J. Conroy, J. L. Parra, G. C. White, and S. R. Beissinger. 2008.
794 Impact of a century of climate change on small-mammal communities in Yosemite
795 National Park, USA. Science 322:261-264.
796 Nadler, C. F., and M. H. Block. 1962. The Chromosomes of Some North American Chipmunks
797 (Sciuridae) Belonging to the Genera *Tamias* and *Eutamias**. Chromosoma 13:1-15.
798 Nagorsen, D. W., N. Panter, and D. Copley. 2020. Phenotypes and distribution of yellow-pine
799 chipmunk (*Neotamias amoenus*) of hybrid ancestry from the Rocky Mountains of
800 Canada. Western North American Naturalist 81:328-343.
801 Nguyen, L. T., H. A. Schmidt, A. von Haeseler, and B. Q. Minh. 2015. IQ-TREE: a fast and
802 effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol
803 Evol 32:268-274.
804 Paradis, E., J. Claude, and K. Strimmer. 2004. APE: Analyses of Phylogenetics and Evolution in
805 R language. Bioinformatics 20:289-290.
806 Patterson, B. D. 1981. Morphological Shifts of Some Isolated Populations of *Eutamias*
807 (Rodentia: Sciuridae) in Different Congeneric Assemblages. Evolution 35:53-66.
808 Patterson, B. D. 1984. Geographic Variation and Taxonomy of Colorado and Hopi Chipmunks
809 (Genus *Eutamias*). Journal of Mammalogy 65:442-456.
810 Patterson, B. D., and L. R. Heaney. 1987. Analysis of Geographic Variation in Red-Tailed
811 Chipmunks (*Eutamias ruficaudus*). Journal of Mammalogy 68:782-791.
812 Patterson, B. D., and R. W. Norris. 2016. Towards a uniform nomenclature for ground squirrels:
813 the status of the Holarctic chipmunks. Mammalia 80.
814 Patterson, B. D., and C. S. Thaeler. 1982. The Mammalian Baculum: Hypotheses on the Nature
815 of Bacular Variability. Journal of Mammalogy 63:1-15.
816 Pease, J. B., and M. W. Hahn. 2015. Detection and Polarization of Introgression in a Five-Taxon
817 Phylogeny. Syst Biol 64:651-662.

Piaggio, A. J. and G. S. Spicer. 2000. Molecular Phylogeny of the Chipmunk Genus *Tamias* Based on the Mitochondrial Cytochrome Oxidase Subunit II Gene. Journal of Mammalian Evolution 7:147-166.

Piaggio, A. J. and G. S. Spicer. 2001. Molecular phylogeny of the chipmunks inferred from Mitochondrial cytochrome b and cytochrome oxidase II gene sequences. Mol Phylogenet Evol 20:335-350.

Quinlan, A. R., and I. M. Hall. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26:841-842.

Rannala, B. and Z. Yang. 2017. Efficient Bayesian Species Tree Inference under the Multispecies Coalescent. Systematic Biology 66:823-842.

Reid, N., J. R. Demboski, and J. Sullivan. 2012. Phylogeny estimation of the radiation of western North American chipmunks (*Tamias*) in the face of introgression using reproductive protein genes. Syst Biol 61:44-62.

Rubidge, E. M., J. L. Patton, and C. Moritz. 2014. Diversification of the Alpine Chipmunk, *Tamias alpinus*, an alpine endemic of the Sierra Nevada, California. BMC Evolutionary Biology 14:1-15.

Sarver, B. A., J. R. Demboski, J. M. Good, N. Forshee, S. S. Hunter, and J. Sullivan. 2017. Comparative Phylogenomic Assessment of Mitochondrial Introgression among Several Species of Chipmunks (*Tamias*). Genome Biol Evol 9:7-19.

Sarver, B. A. J., N. D. Herrera, D. Sneddon, S. S. Hunter, M. L. Settles, Z. Kronenberg, J. R. Demboski, J. M. Good, and J. Sullivan. 2021. Diversification, Introgression, and Rampant Cytonuclear Discordance in Rocky Mountains Chipmunks (Sciuridae: *Tamias*). Syst Biol.

Schultz, N. G., J. Ingels, A. Hillhouse, K. Wardwell, P. L. Chang, J. M. Cheverud, C. Lutz, L. Lu, R. W. Williams, and M. D. Dean. 2016. The Genetic Basis of Baculum Size and Shape Variation in Mice. G3 (Bethesda) 6:1141-1151.

Sikes, R. S., & Mammalogists, T. A. C. and U. C. of the American Society of Mammalogists. 2016. Guidelines of the American Society of Mammalogists for the use of wild mammals in research and education. Journal of Mammalogy 97:663-688.

Simmons, L. W. 2014. Sexual selection and genital evolution. Austral Entomology 53:1-17.

Simmons, L. W., and R. C. Firman. 2013. Experimental evidence for the evolution of the Mammalian baculum by sexual selection. Evolution 68:276-283.

Struck, T. H., J. L. Feder, M. Bendiksby, S. Birkeland, J. Cerca, V. I. Gusarov, S. Kistenich, K. H. Larsson, L. H. Liow, M. D. Nowak, B. Stedje, L. Bachmann, and D. Dimitrov. 2018. Finding Evolutionary Processes Hidden in Cryptic Species. Trends Ecol Evol 33:153-163.

Sullivan, J., J. R. Demboski, K. C. Bell, S. Hird, B. Sarver, N. Reid, and J. M. Good. 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. The Southwestern Naturalist 27:393-402.

Sutton, D. A., and B. D. Patterson. 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 Jr., R. W., J. L. Koprowski, M. A. Steele, and J. F. Wahatton. 2012. Squirrels of the world. The Johns Hopkins University Press, Baltimore, Maryland.

Van der Auwera, G. A., M. O. Carneiro, C. Hartl, R. Poplin, G. Del Angel, A. Levy-Moonshine, T. Jordan, K. Shakir, D. Roazen, J. Thibault, E. Banks, K. V. Garimella, D. Altshuler, S. Gabriel, and M. A. DePristo. 2013. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics 43:11 10 11-11 10 33.

White, J. A. 1953. The baculum in the chipmunks of western North America. University of Kansas Publications Museum of Natural History 5:611-631.

Yang, Z. 2015. The BPP program for species tree estimation and species delimitation. Current Zoology 61:854–865.

Yang, Z., and B. Rannala. 2010. Bayesian species delimitation using multilocus sequence data. Proc Natl Acad Sci U S A 107:9264-9269.

Yu, Y., J. H. Degnan, and L. Nakhleh. 2012. The probability of a gene tree topology within a phylogenetic network with applications to hybridization detection. PLoS Genet 8: e1002660.

Yu, Y., J. Dong, K. J. Liu, and L. Nakhleh. 2014. Maximum likelihood inference of reticulate evolutionary histories. Proc Natl Acad Sci U S A 111:16448-16453.

Yu, Y., and L. Nakhleh. 2015. A maximum pseudo-likelihood approach for phylogenetic networks. BMC Genomics 16:1-10.

Zhang, C., M. Rabiee, E. Sayyari, and S. Mirarab. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinformatics 19:153.

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