bioRxiv preprint doi: https://doi.org/10.1101/112581; this version posted March 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

**Manuscript type: Article - Discovery**

# Conflicting evolutionary histories of the mitochondrial and nuclear genomes in New World *Myotis*

Roy N. Platt II[1], Brant C. Faircloth[2], Kevin A.M. Sullivan[1], Troy Kieran[3], Travis C. Glenn[3], Michael W. Vandewege[1], Thomas E. Lee[4], Robert J. Baker[1], Richard D. Stevens[5], David A. Ray[1*]

[1] Department of Biological Sciences, Texas Tech University, Lubbock, USA

[2] Department of Biological Sciences and Museum of Natural Science, Louisiana State University, Baton Rouge, USA

[3] Environmental Health Science, University of Georgia, Athens, USA

[4] Department of Biology, Abilene Christian University, Abilene, USA

[5] Natural Resource Management, Texas Tech University, Lubbock, USA

*Corresponding author: E-mail: david.4.ray@gmail.com

## Abstract

The diversification of *Myotis* into more than 100 species in just a few million years is one of the most extensive mammalian radiations available for study. Efforts to understand relationships within *Myotis* have primarily utilized mitochondrial markers, and trees inferred from nuclear markers lacked resolution. Our current understanding of relationships within *Myotis* is therefore biased towards a set of phylogenetic markers that may not reflect the phylogenetic history of the nuclear genome. To resolve this, we sequenced the full mitochondrial genomes of 37 representative *Myotis*, primarily from the New World, in conjunction with targeted sequencing of 3,648 ultraconserved elements (UCEs). We inferred the phylogeny of *Myotis* and explored the effects of concatenation and summary phylogenetic methods, as well as combinations of markers based on informativeness or levels of missing data, on our phylogenetic results. Of the 295 phylogenies generated from the nuclear UCE data, all are significantly different from phylogenies inferred using mitochondrial genomes. Even within the nuclear genome quartet frequencies indicate that around half of all UCE loci conflict with the estimated species tree. Several factors can drive such conflict, including incomplete lineage sorting, introgressive hybridization, or even phylogenetic error. Despite the degree of discordance between nuclear UCE loci and the mitochondrial genome and among UCE loci themselves, the most common nuclear topology is recovered in one quarter of all analyses with strong nodal support. Based on these results, we re-

bioRxiv preprint doi: https://doi.org/10.1101/112581; this version posted March 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

examine the evolutionary history of *Myotis* to better understand the phenomena driving their unique nuclear, mitochondrial, and biogeographic histories.

**Keywords**

incomplete lineage sorting, summary tree methods, concatenation, Vespertilionidae, phylogenomics, UCE, ultraconserved elements, reticulation

bioRxiv preprint doi: https://doi.org/10.1101/112581; this version posted March 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

## Introduction

The genus *Myotis* (Order Chiroptera, Family Vespertilionidae) contains more than 100 species that originated during the last 10-15 million years (Stadelmann, et al. 2007), making it one of the most successful, extant, mammalian species radiations. Members of *Myotis* are distributed worldwide, excluding polar regions, and generally share a similar ecological niche: aerial insectivory. *Myotis* species often exhibit little morphological differentiation and, as a result, the rate of cryptic speciation within the genus is thought to be high. For example, specimens identified as *M. nigricans* and *M. albescens* form multiple paraphyletic lineages distributed throughout the phylogeny of Neotropical *Myotis* (Larsen, et al. 2012).

Confounding matters, the morphological variation that exists is often a poor indicator of species-level relationships. Early classifications of *Myotis* identified three major morphotypes (Findley 1972); each were assumed to be monophyletic and were recognized at the subgeneric level (Simmons 2005). Subsequent phylogenetic analyses of the mitochondrial cytochrome-b (*cytb*) gene recovered paraphyletic origins of the morphologically defined subgenera, suggesting convergent evolution in *Myotis* (Ruedi and Mayer 2001). These same analyses demonstrated that geography was a better predictor of phylogenetic relationship than morphology (Ruedi and Mayer 2001; Stadelmann, et al. 2007). Generally, *Myotis* phylogenies from mitochondrial data contain a single bifurcation at the base of the tree that splits Old World from New World species. An additional bifurcation within New World species separates Nearctic (NA) from Neotropical (NT) species. The NA/NT bifurcation is not absolute, with at least five NA species located in the Neotropics and *vice versa*. The Old World/New World bifurcation is stricter, with only two Old World species, *M. brandtii* and *M. gracilis,* present in the New World clade.

The ability of mitochondrial markers to resolve a well-supported topology does not guarantee that the mitochondrial tree represents the species tree (for examples see Willis, et al. 2014; Li, et al. 2016; Leavitt, et al. 2017). Despite containing 37 genes, the lack of recombination and uniparental inheritance of the mitochondrion means that it is transmitted as a single genetic unit. This makes mitochondria susceptible to evolutionary processes that may cause its history to diverge from the history of the species (Edwards and Bensch 2009). The most widely accepted phylogenies of *Myotis* rely heavily on mitochondrial data and even phylogenies containing nuclear data demonstrate an over reliance on mitochondrial markers for resolution. For example alignments of the nuclear *RAG*2 and mitochondrial *cytb* contained 162 and 560 variable characters respectively (Stadelmann, et al. 2007). Phylogenetic analyses of *RAG*2 in *Myotis* results in a tree primarily composed of polytomies

bioRxiv preprint doi: https://doi.org/10.1101/112581; this version posted March 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

(Stadelmann, et al. 2007). Combining these two markers increases phylogenetic resolution, but the results are heavily influenced by larger numbers of mitochondrial characters, potentially masking signal form the nuclear marker (Stadelmann, et al. 2007; Larsen, et al. 2012; Ruedi, et al. 2013; Haynie, et al. 2016).

It is difficult to draw major conclusions from studies limited in the number of characters (Ruedi and Mayer 2001; Stadelmann, et al. 2007; Larsen, et al. 2012; Ruedi, et al. 2013; Haynie, et al. 2016) or taxa (Platt, et al. 2015). Current data seem to indicate that nuclear and mitochondrial markers recover similar topologies. Platt et al. (2015) generated a phylogeny based entirely on nuclear data using 85,028 shared transposable element insertions. Their results generally confirmed the mitochondrial phylogenies of *Myotis*, but only included seven taxa in their analysis. In order to fully resolve relationships and understand the *Myotis* radiation it is necessary to increase character and taxon sampling.

Recently, targeted sequencing methods have been developed that utilize baits to enrich and sequence ultraconserved elements (UCEs; Faircloth, et al. 2012), and this method has resolved a number of difficult phylogenetic problems (for examples see Crawford, et al. 2012; McCormack, et al. 2013; Green, et al. 2014; Faircloth, et al. 2015; McGee, et al. 2016). Generally speaking, the conserved "core" of UCE regions allows thousands of homologous loci to be enriched from divergent organismal genomes while the sequence that flanks the core UCE region contains a majority of phylogenetically informative sites – allowing researchers to collect a large number of phylogenetically informative, homologous loci from throughout the genome in a cost-effective and efficient manner. Broad sampling of the nuclear genome should help to resolve a phylogeny without an over reliance on mitochondrial loci. In addition, increasing the number of nuclear loci sampled from a handful of genes to thousands can recover accurate trees despite high levels of incomplete lineage sorting (Maddison and Knowles 2006)

Here, we used targeted sequencing of UCEs to collect ~1.4 Mbp from ≥3,600 nuclear loci in 37 taxa, primarily representing New World *Myotis*. Combinations of the UCE data were analyzed using concatenation and tree summary methods to estimate the *Myotis* phylogeny. Analysis of the nuclear UCE data recovered 295 trees representing 175 distinct topologies. The nuclear topologies were compared to trees generated from full mitochondrial genomes to test for conflict between the two types of makers. Our results show that, despite the range of trees recovered from the nuclear data, nuclear and mitochondrial markers always depict conflicting phylogenies. Given that the nuclear and mitochondrial trees are distinct from one another it is necessary to reinvestigate conclusions made based solely on the mitochondrial phylogeny.

bioRxiv preprint doi: https://doi.org/10.1101/112581; this version posted March 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

**Results**

We used targeted sequencing of UCEs to collect sequence data from 3,648 nuclear loci which we assembled into concatenated alignments as large as 1.37 Mb. In addition, we assembled mitochondrial genomes for most taxa. We then used the data to infer the phylogenetic history of New World *Myotis* in three phases: UCE and mitogenome assembly, initial phylogenetic analysis, extended phylogenetic analyses.

UCE and mitochondrial assembly and alignment – We averaged 3.29 million reads per sample after demultiplexing. These reads were assembled into an average of 5,778 contigs per sample (min = 1,562 *M. martiniquensis*, max = 11,784 *M. nigricans* 3). Recovery of UCE loci varied across taxa. Of the 5,500 loci in the Amniote probe set, we successfully recovered 3,898 UCE loci, 3,648 loci from five or more samples, 212 loci in all 37 samples (Table 2). On average, 3,332 UCE loci were recovered per sample, ranging from 1,106 (*M. martiniquensis*) 4,008 (*M. keaysi*). Repetitive sequences, identified via RepeatMasker searches, were minimal occupying less than 0.02% of sites across all UCE alignments. Sequence coverage of the mitochondrial genomes averaged 58x (range >1x - 297x). Mitochondrial genome assemblies varied in quality. Some were almost entirely complete while others were missing sections. We found three premature stop codons in mtDNA protein coding genes. Subsequent manual alignment and validation suggested that these regions were miscalled by MitoBim, and we corrected the errors prior to analysis.

Initial phylogenetic analyses – Initial analysis of the nuclear data used loci that were present in 20 or more taxa. This resulted in an alignment of 1,144,471 bp from 2,890 nuclear loci containing 35,284 parsimony informative characters. The 2,890 loci were split into 27 partitions as recommended by PartitionFinder (Lanfear, et al. 2012). Maximum likelihood and Bayesian analyses recovered the same topology and found similar support for most nodes (Figure 1A). Maximum likelihood analysis recovered 100% support for 31 of 35 bipartitions, and 33 bipartitions were present in ≥98% bootstrap replicates. Nodes with the least support were still present in 86% - 88% of bootstrap replicates. After 50 bootstrap replicates the average weighted Robinson-Foulds distance between replicate sets was less than 0.23% (Pattengale, et al. 2009). Bayesian analysis recovered an identical topology, with the only difference being that all bipartitions were supported with a clade probability value of >0.99. Visual inspection of the parameter files in Tracer v1.6 showed good sampling with a likelihood score of $-2.419 \times 10^{-6}$ and an effective sample size (ESS) of 637 for the likelihood parameter. All other parameters had effective

bioRxiv preprint doi: https://doi.org/10.1101/112581; this version posted March 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

sample sizes greater than 500. The average standard deviation of split frequencies (ASDF) across all runs was less than 1% after 5,000 generations.

Thirty-seven mitochondrial protein coding, rRNA and tRNA genes were concatenated into single alignment of 15,520 bp containing 5,007 informative characters. Alignments for 30 samples were ≥ 90% complete, and alignments for five samples were 68-84% complete. Only 21% and 50% of nucleotide positions were present in the *M. albescens*[3] (TK 61766) and *Myotis levis* alignments. Maximum likelihood and Bayesian analyses of the mitochondrial data recovered similar topologies (Figure 1B), varying in the *M. thysanodes*, *M. evotis*, and *M. keeni* relationships. Neither method recovered significant support for these relationships. Bootstrap replicates of the maximum likelihood analysis meet the stopping criterion after the first 50 of 10,000 replicates (average weighted Robinson-Foulds value = 2.28%). The RAxML mitochondrial phylogeny was well supported with 29 of 35 nodes present in ≥96% of bootstrap replicates. The remaining six nodes were present in 47% to 70% of bootstrap replicates. Bayesian analysis of the mitochondrial data reached convergence, defined as an ASDF of <1%, after the first 424,500 of one million generations. The final ASDF, after discarding 25% of samples, was 0.49%. The trace files from all four independent runs shows proper mixing of samples and the effective sample size for all parameters was greater than 200. The log likelihood score for the Bayesian mitochondrial tree was $-1.205 \times 10^{-5}$ with an ESS of 1,446. In all, posterior probabilities were lower than the bound established as significant (≥ 0.95) for five nodes.

Mitochondrial and nuclear analyses recovered different topologies (Figure 1). We stripped branch lengths from all trees and compared the topologies using an approximately unbiased test to determine whether differences in the tree represented conflicting signals in the marker sets. When the nuclear data is constrained to the mitochondrial tree (*p*-value = $1 \times 10^{-66}$) or *vice versa* (*p*-value = $2 \times 10^{-5}$), likelihood scores are significantly worse than expected given similar evolutionary histories. These results reject the hypothesis that the mitochondrial and nuclear UCE phylogenies reflect similar evolutionary histories.

<u>Extended phylogenetic analyses</u> - Many factors can bias phylogenetic analyses resulting in inaccurate trees (Sanderson and Shaffer 2002). Rather than assuming our initial nuclear UCE tree was an accurate estimate of the phylogenetic relationships of *Myotis*, we wanted to build a range of plausible topologies from the nuclear UCE data. To do this we reanalyzed the nuclear UCE data set with minor deviations in locus sampling, partitioning, inference method, *etc*. In all this effort resulted in 291 unique phylogenetic analyses. Individual results or topologies are not the focus of these analyses. Rather, the goal was to recover as many, reasonable, nuclear UCE topologies as possible in an effort to account for phylogenetic

bioRxiv preprint doi: https://doi.org/10.1101/112581; this version posted March 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

uncertainty not present in the initial analysis and to compare the range of nuclear UCE trees to the mitochondrial genome tree.

We investigated the effects of matrix composition (or completeness) on our phylogenetic inference by generating 10 different alignments having levels of matrix completeness spanning 15-95% at 10% intervals and including a final matrix of 100% completeness. Loci in these alignments were partitioned using three separate schemes: all loci were partitioned individually, loci were unpartitioned, or loci were combined into optimum partitions using PartitionFinder. The result was 10 different alignments with three partitioning schemes each. These were analyzed using Bayesian and maximum likelihood methods. Due to computational limits we abandoned the fully partitioned, Bayesian analyses. The length, number of loci, and optimum number of partitions per alignment is shown in Table 2. Bootstrap topologies stabilized in 9 of 10 alignments after 50 replicates and all Bayesian runs converged in less than ten thousand generations. In general, the same alignment produced the same topology regardless of inference method or partitioning scheme with the only exception being the terminal relationships *M. levis*/*M. albescens* clade in the optimum vs. unpartitioned Bayesian analysis of the 100% complete data matrix.

Trees were generated from data matrices incorporating loci of differing lengths (Hosner, et al. 2016). All 3,648 loci were ordered based on their length and split into nine bins of 365 loci and 1 bin of 363 loci, so that the first bin contained the 365 shortest loci, the second bin contained the 366$^{th}$ to the 731$^{st}$, and so on. The number of informative characters per bin ranged from 1,115 to 6,995 and the number of informative characters was correlated with average locus length (Supplemental Figure 2). On average, only 2.6% of characters in each bin were parsimony-informative. Each of the ten length-based alignments recovered slightly different topologies. Terminal relationships were generally stable across analyses with the majority of differences between topologies found in the early bifurcations of the ingroup (*Myotis*).

From the above analyses combining different matrix composition, inference method, partitioning, and locus-length variants we observed that, in general, larger alignments produced well resolved topologies with significant nodal support regardless of the phylogenetic method or partitioning scheme used. On the other hand, re-analyses of smaller portions of the data were more likely to recover unique topologies. Given that the overall goal of the extended analyses was to generate as many reasonable nuclear UCE based topologies for comparison with the mitochondrial tree, we decided to randomly sample small portions of the nuclear UCE loci to create alignments that are more likely to result in unique topologies. We randomly subsampled the 3,648 enriched loci to create 100 unique data

bioRxiv preprint doi: https://doi.org/10.1101/112581; this version posted March 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

sets. Loci were concatenated in each replicate data set and analyzed using maximum likelihood in RAxML. Of the 100 alignments analyzed, 80 unique topologies were generated (mean Robinson-Foulds distance = 4.3).

In addition to concatenated analyses, three summary-based species tree programs were used on datasets of varying matrix-completeness (ASTRAL-II, ASTRID, SVDquartets). Normalized quartet scores from ASTRAL-II (Mirarab and Warnow 2015) analyses were quite consistent with scores ranging from 0.540 to 0.553, and between 7,745,739 (100% complete 212 gene trees) and 63,042,410 (15% 3648 loci) induced quartet gene trees. SVDquartets (Chifman and Kubatko 2014) sampled all 66,045 quartets. On average, the total weight of incompatible quartets was 2.84%. Similar to the concatenated analysis, we inferred coalescent-based species from the same 100 subsamples of 365 loci described above. Despite being generated from the same underlying data, summary and concatenation methods only recovered the same tree in one of 100 attempts.

Finally, we used weighted and unweighted statistical binning to combine individual gene trees into supergenes, estimate the supergene phylogeny, and then infer the species tree from the supergene trees. The 3,648 loci were combined into 528 binned loci with 480 bins containing seven loci each and 48 bins containing six loci each. Normalized quartet scores were 0.672 for the binned-unweighted and 0.673 for the binned-weighted ASTRAL-II analysis. Given the relative even distribution of loci into bins the negligible difference in quartet/species tree discordance is expected. Both binning methods recovered the same topology which was the same tree recovered in the initial nuclear UCE analyses and was the most common topology observed across all analyses.

Topology comparisons – After rejecting topological congruence between the initial nuclear UCE and mitochondrial phylogenies we used various methods to re-analyze the nuclear UCE data in an effort to identify alternative nuclear topologies. Topological congruence between the mitochondrial sequence data and nuclear topologies resulting from the extended analyses were tested to see if any were statistically congruent with the mitochondrial phylogeny of *Myotis*. Site-log likelihood scores for the mitochondrial alignments when constrained to all 175 unique nuclear UCE topologies were generated in RAxML and analyzed in Consel using the Shimodaira-Hasegawa (Shimodaira and Hasegawa 1999) and approximately unbiased tests (Shimodaira 2002). In each case, the mitochondrial data produced significantly worse likelihood scores, rejecting congruence between the nuclear UCE and mitochondrial phylogenies (Supplemental Table 1).

When visualizing all topologies in tree space, nuclear trees co-localized and were distinct from mitochondrial topologies (Figure 2A). Pairwise comparisons of Robinson-Foulds symmetrical differences

bioRxiv preprint doi: https://doi.org/10.1101/112581; this version posted March 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

show that 98.75% of nuclear UCE vs. nuclear UCE (Figure 2B) trees are more similar to each other than the mitochondrial trees are to even the most similar nuclear UCE tree (Figure 2C). The most frequently observed topology was recovered in 45 of the 294 nuclear analyses and was identical to the tree recovered in the initial nuclear UCE analysis (Figure 1a). Of the 45 analyses that recovered the most frequently observed topology, 38 were Bayesian and RAxML searches that varied by matrix completeness and partitioning scheme. The fact that these analyses recovered the same topology is expected given that they are not independent. For example, a RAxML analysis of the 15% complete data set uses 1.26 Mb of the 1.38 Mb of data from the 25% complete dataset. These two alignments are 91% identical. Analyses that directly varied the alignments and/or sampled less data (e.g. randomly sampling loci) were more likely to generate unique topologies than the nested analyses described above. Of the 200 analyses that randomly sampled UCE loci 164 unique topologies were observed. This implies that when analyses of large data sets produce well-resolved trees with significant nodal support, sampling smaller portions of the data, may provide a mechanism for creating phylogenetic uncertainty not represented by typical tree scoring metrics. To account for the phylogenetic uncertainty present in our dataset, we generated a consensus tree from all nuclear topologies using an 85% threshold to resolve bipartitions (Figure 3).

**Discussion**

We generated phylogenies from 3,648 UCE loci and mitochondrial genomes of 35 *Myotis* bats. Initial analyses of the mitochondrial and nuclear UCE phylogenies recovered distinct topologies (Figure 1). Rather than rejecting concordance between the two data types from a single analysis we took steps to re-analyze the nuclear UCE data in an effort to generate as many viable nuclear topologies as possible. We recovered 175 unique nuclear topologies using multiple methodologies, sampling strategies, and parameters. None of these nuclear topologies were similar to the topology produced from the mitochondrial data suggesting that nuclear UCE loci and the mitochondrial genomes of *Myotis* have distinct evolutionary histories. The conflict between the mitochondrial and nuclear data may be driven by error in phylogenetic estimation or may reflect genuine conflict between the two marker types (Degnan and Rosenberg 2009; Huang, et al. 2010). We relied on multiple tree-inference methods (e.g. summary vs. concatenation), manipulated phylogenetic parameters (e.g. partitioning strategy), and sampling criteria (e.g. loci sampled in all taxa) to minimize the impacts of phylogenetic error on the data set. In most cases, varying parameter or methodologies generated unique topologies, often due to rearrangements of a few terminal taxa. *M. volans* and *M. brandtii* were often placed as either sister to the remaining NW *Myotis* or as an early bifurcation between the NA and NT clades. *M. vivesi* was often

bioRxiv preprint doi: https://doi.org/10.1101/112581; this version posted March 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

found as sister to the clade containing *M. lucifugus*, *occultus* and *fortidens* or as sister to the clade containing the NT *Myotis*.

Around 98.3% of all nuclear tree vs. nuclear tree comparisons contain fewer than 30 symmetric differences (Figure 2b) but there are no mitochondrial vs. nuclear tree comparisons with less than 30 symmetric differences (Figure 2c). Interestingly, the most common nuclear topology most often recovered by concatenation analyses (Figure 1A). Summary methods failed to recover the most common nuclear topology except when loci were binned together prior to gene tree estimation. Summary methods also tended to recover more unique topologies than concatenation methods when analyzing data from the same gene(s). For example, the random sample analyses recovered 80 unique topologies using concatenation (RAxML) and 89 unique topologies with summary methods (ASTRAL-II). This likely has to do with the limited number of informative characters per locus and by extension limited phylogenetic signal per gene tree (Supplemental Figure 2). In these instances, limited phylogenetic signal per gene would likely lead to increased opportunity for phylogenetic error in gene tree estimation. Further supporting this idea, binning of compatible UCE loci may have indirectly increased phylogenetic signal resulting in the same topology that many of the concatenation analyses recovered. No other summary/coalescent method recovered this topology.

Previous work with UCE loci demonstrated that support for deep divergences varied based on the number of loci examined (McCormack, et al. 2013). Further, bootstrap replicates and clade probability values can be inaccurate metrics of nodal support (Douady, et al. 2003; Hedtke, et al. 2006). Varying the input data and phylogenetic parameters can produce a range of reasonable nuclear topologies that may be more useful than overreliance on a tree resulting from a one or two analyses. Here, by considering the different topologies that result from various analyses (e.g. partitioning strategies, inference methods, etc.), we can account for phylogenetic uncertainty better than considering a single nuclear or mitochondrial topology alone.

The mitochondrial alignment constrained to any of the 175 nuclear topologies generated significantly worse likelihood scores than expected by chance (Supplemental Table 2) and a comparison of topologies in tree space shows that the mitochondrial topologies are unique from all nuclear topologies (Figure 2A). Pairwise tree distances demonstrate that all but the most divergent nuclear topologies are more similar to each other (Figure 2B) than any nuclear vs. mitochondrial comparison (Figure 2C). Despite the number of different analyses, the nuclear data never recover a topology that is similar, much less identical, to the mitochondrial topology.

bioRxiv preprint doi: https://doi.org/10.1101/112581; this version posted March 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Multiple studies have recovered effectively the same relationships among *Myotis* using mitochondrial markers to the one presented here (Ruedi and Mayer 2001; Stadelmann, et al. 2007; Roehrs, et al. 2010; Larsen, et al. 2012; Ruedi, et al. 2013; Haynie, et al. 2016). Thus, we are confident that the mitochondrial phylogeny we recovered here, and by others, reflects the true mitochondrial tree. However, the mitochondrial topology may not adequately reflect the species history, particularly when considering the factors that cause incongruence between nuclear and mitochondrial gene trees. Possible causes of conflicting gene trees are horizontal transfer, gene duplication, introgressive hybridization, and incomplete lineage sorting. Some of these phenomena are more likely to have influenced the *Myotis* radiation than others.

Horizontal transfer of genes is thought to be rare in eukaryotes, but, vespertilionids in general (Thomas, et al. 2011; Platt, et al. 2014), and *Myotis* (Pritham and Feschotte 2007; Ray, et al. 2007; Ray, et al. 2008) in particular, have experienced horizontal transfer of DNA transposons. These events would not be reflected in our phylogeny since repetitive sequences were removed prior to phylogenetic analyses. More generally, gene duplications could create conflicting signal among individual UCE markers (ex. comparing non-orthologous UCE loci), but the number of gene duplication events would have to be very high to impact enough of the 3,648 UCE loci to confound the mitochondrial and nuclear phylogenies. Further ruling out gene duplication events as the dominant cause of conflicting phylogenetic signal is the fact that such events are likely depressed in *Myotis* as evidenced by their smaller genome size (~2.2 Gb) and trend towards DNA loss (Kapusta, et al. 2017) combined with low rates of paralogy in UCEs general (Derti, et al. 2006).

Introgressive hybridization and reticulation could significantly influence the phylogenies of *Myotis* in a way that leads to conflicting signal between the nuclear and mitochondrial genomes (Sota 2002; Good, et al. 2015). Hybridization in bats may be relatively common given their propensity to swarm at cave entrances for breeding purposes. In European *Myotis*, swarming has allowed for high degrees of hybridization between *M. brandtii, M. mystacinus, and M. alcathoe* (Bogdanowicz, et al. 2012). Further, *M. evotis*, *thysanodes*, and *keeni* all experienced historical gene flow during their divergence (Carstens and Dewey 2010; Morales, et al. 2016). It is also possible to explain the differences between the mitochondrial and nuclear UCE phylogenies if *Myotis* experienced extensive incomplete lineage sorting during their radiation. Two factors can influence the rate of lineage sorting, the fixation rate and the speciation rate (Hudson, et al. 2002). Increasing the time to fixation and/or decreasing the amount of time between cladogenic events will increase the likelihood of incomplete lineage sorting. *Myotis* are generally long lived species (Dzeverin 2008) and underwent a rapid radiation between 5-10

bioRxiv preprint doi: https://doi.org/10.1101/112581; this version posted March 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

MYA (Lack, et al. 2010), suggesting that *Myotis* species are likely to experience higher levels of lineage sorting. The importance of these events -introgressive hybridization and incomplete lineage sorting- in driving the differences between the mitochondrial and nuclear phylogenies cannot be determined with the current data.

**Evolutionary history of *Myotis* –** Our previous understanding of relationships within *Myotis* is heavily biased with mitochondrial data because nuclear markers were harder to collect and produced fewer informative sites (Ruedi and Mayer 2001; Stadelmann, et al. 2007; Lack, et al. 2010; Roehrs, et al. 2010; Larsen, et al. 2012; Ruedi, et al. 2013; Haynie, et al. 2016). Our UCE-based results indicate that nuclear trees vary substantially from the mitochondrial tree. Given that the nuclear and mitochondrial trees are different, we find it necessary to re-evaluate *Myotis* in the context of the nuclear data.

Paraphyly of *M. nigricans* and *M. albescens* was inferred from previous mitochondrial phylogenies and confirmed in the UCE tree (Larsen, et al. 2012). Larsen et al (Larsen, et al. 2012) identified a minimum of four and potentially twelve lineages in *M. albescens* and *M. nigricans*. Our sampling included four *M. albescens* and three *M. nigricans*, compared to Larsen's 17 and 29 samples. Despite different mitochondrial and nuclear topologies overall, our mitochondrial and nuclear phylogeny recovered the same paraphyletic clade of three *M. albescens* samples and *M. levis*. Close relationships between these taxa was found in previous work and expected. More importantly we did not find that M. albescens was paraphyletic across much of NT Myotis. We also found that *M. nigricans* is monphlyletic in the nuclear tree, but paraphyletic in the mitochondrial tree. These results from *M. nigricans* and *M. albescens* are interesting but further inference is limited due to low sample sizes for these taxa.

The original subgeneric taxonomy of *Myotis* was based on three morphotypes that were later shown to be the result of convergent evolution (Ruedi and Mayer 2001). If lineage-sorting affected the mitochondrial phylogeny, it is possible that the morphotypes truly are monophyletic. However, superimposing the previous subgeneric/morphological classification onto the species tree shows interspersed distribution of morphotypes throughout even the most conservative nuclear tree (Figure 3). Many strongly supported terminal relationships link species with different morphotypes. Based on these results, it appears that the three major morphotypes in *Myotis* are indeed a result of convergent evolution, as suggested by previous work (Ruedi and Mayer 2001; Stadelmann, et al. 2007).

Among the more dramatic differences between the nuclear and mitochondrial topologies is the placement of *M. volans* and *M. brandtii* as sister to all New World taxa by the nuclear data. Our mitochondrial analyses place *M. volans* within a Nearctic clade and *M. brandtii* directly in-between the

bioRxiv preprint doi: https://doi.org/10.1101/112581; this version posted March 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Nearctic and Neotropical bifurcations as has been previously reported (Stadelmann, et al. 2007). Clade probability values and bootstrap frequencies support these placements in trees from both data types. Our placement of *M. brandtii* as sister to all other New World *Myotis* more closely affiliates it with Old World taxa. This make sense given that the *M. brandtii* distribution is also Old World. On the other hand, a placement of *M. volans* sister to all New World taxa (and *M. brandtii*) in the nuclear tree is a significant departure from previous work and, at first glance, does not make as much sense in a biogeographic framework. *M. volans* is distributed across western and northwestern North America as far as far north as Alaska. *M. brandtii* is distributed across much of Northern Europe. The key may lie in understanding a third species, *M. gracili*s.

*M. gracilis*, along with *M. brandtii*, are the only two *Myotis* geographically distributed in the Old World, but phylogenetically affiliated with the New World (Stadelmann, et al. 2007). If the sister relationship between *M. brandtii* and *M. gracilis* (not sampled here) holds when nuclear data are examined, then we can envision a scenario where *M. gracilis, M. brandtii,* and *M. volans* represent speciation events that occurred during the transition of *Myotis* from the Old World to the New World. It is important to remember that this interpretation relies on a fairly dramatic departure from the currently accepted mitochondrial relationships of *M. volans* (represented here by a single sample) to other *Myotis* species, and this hypothesis should be viewed as highly speculative. Increasing the number of *Myotis* lineages sampled will shed additional light on this hypothesis.

Other taxa with conflicting positions between datasets include *M. lucifugus* + *M. occultus*, *M. fortidens*, and *M. vivesi*. In general, these relationships are characterized by very short branches and are the most likely to be affected by incomplete lineage sorting or limited phylogenetic information. This could explain the strong support with the mitochondrial tree compared to the nuclear species tree, while allowing for a number of nuclear loci to disagree with the species tree, as well.

There are a number of monophyletic groups identified with nuclear data (Fig. 1A) that exhibit distinct biological characteristics. For example all of the long eared bats (*septentrionalis*, *auriculus*, *evotis*, *thysanodes* and *keenii*) represent a monophyletic group of higher elevation, forest-dwelling species that glean insects off of surfaces (Fitch and Shump 1979; O'Farrell and Studier 1980; Warner 1982; Manning and Jones 1989; Caceres and Barclay 2000). The group represented by *fortidens*, *lucifugus* and *occultus* represent a relatively long-haired form of *Myotis*. While having a distinct dental formula, *fortidens* was historically described as a subspecies of M. *lucifugus* (Miller Jr and Allen 1928) and *occultus* has alternately represented its own species or been considered a subspecies of *lucifugus* (Hollister 1909; Valdez, et al. 1999; Piaggio, et al. 2002). The clade consisting of *keaysi*, *oxyotus*, *ruber*,

bioRxiv preprint doi: https://doi.org/10.1101/112581; this version posted March 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

*simus*, *riparius*, *albescens* and *diminutus* represents a NT group of primarily woolly-haired bats (LaVal 1973). If the mitochondrial genome has been subjected to phenomena that obscure the true species tree then these species groups, along with their synapomorphic morphological features, can be reevaluated.

**Conclusion -** Relationships within *Myotis*, which until now have relied heavily on mitochondrial data, have served as the basis for species identification (Puechmaille, et al. 2012), evolutionary hypotheses (Simões, et al. 2007), and even conservation recommendations (Boyles and Storm 2007). Previous studies using nuclear data have largely been uninformative or utilized too few samples to draw definitive conclusions. Trees estimated from ~3,650 nuclear loci and 295 different phylogenetic analyses recovered 175 topologies, none of which are congruent with the mitochondrial phylogeny of *Myotis*. Conflict between the mitochondrial and nuclear trees as well as among individual nuclear loci suggest that the *Myotis* radiation may have been accompanied by high levels of incomplete lineage sorting and possible hybridization. Rather than placing emphasis on the mitochondrial tree, it may be more appropriate to consider it for what it really is: a single gene on par with a single UCE locus, albeit one with many more phylogenetically informative characters. If true, then the mitochondrial genome is as likely to reflect the true species tree as any UCE locus chosen at random. Large amounts of lineage sorting make phylogenetic inference difficult and potentially impossible. Other phenomena such as reticulation, hybridization, and introgression have likely influenced the genomes of *Myotis* and should be accounted for in subsequent work. It is possible that the *Myotis* radiation is more accurately reflected as a hard polytomy or a phylogenetic network rather than a strictly bifurcating phylogeny.

## Materials and Methods:

Taxon Selection - Taxa were selected to span the major phylogenetic break points with emphasis on the Nearctic and Neotropical bifurcation as recovered in previous mitochondrial phylogenies (Stadelmann, et al. 2007; Ruedi, et al. 2013) (Table 1). In addition, multiple individuals morphologically identified as *M. nigricans* and *M. albescens* were included to test paraphyly as demonstrated by Larsen et al. (2012). Three Old World species of *Myotis* and the outgroup, *E. fuscus,* were included to root phylogenetic analyses. All field identifications were confirmed from voucher specimens. Information for all specimens examined is available in Table 1.

UCE preparation, sequencing, and processing - Genomic DNA was extracted from 33 samples using either a Qiagen DNEasy extraction kit or a phenol-chloroform/ethanol precipitation. DNA was fragmented using the Bioruptor UCD-300 sonication device (Diagenode, Denville, NJ, USA). Libraries

bioRxiv preprint doi: https://doi.org/10.1101/112581; this version posted March 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

were prepared using the Kapa Library Preparation Kit KR0453-v2.13 (Kapa Biosystems, Wilmington, MA, USA) following the manufacturer's instructions with five minor modifications. First, we used half volume reactions. Second, subsequent to end repair, we added Sera-Mag Speedbeads (Thermo-Scientific, Waltham, MA, USA; prepared according to (Glenn, et al. 2016)) at a ratio of 2.86:1 for end repair cleanup. Third, we ligated universal iTru y-yoke adapters (Glenn, et al. 2016) onto the genomic DNA. Fourth, following adapter ligation, we performed one post-ligation cleanup followed by Dual-SPRI size selection using 55 µL of speedbead buffer (22.5mM PEG, 1M NaCl) and 25 µL of Speedbeads. Finally, we performed a PCR at 95 °C for 45 sec, then 14 cycles of 98 °C for 30 sec, 60 °C for 30 sec, 72 °C for 30sec, then 72 °C for a 5 minute final extension and a 12 °C hold using iTru5 and iTru7 primers to produce Illumina TruSeqHT compatible libraries (Glenn, et al. 2016).

Libraries were quantified on a Qubit 2.0 (Life Technologies) and 83 ng from each library was added to create 5 pools of 6 or 7 libraries each. We then split the pools in two. One subsample was enriched for UCE loci, the other was not. UCE loci in the enriched library pools were captured using Tetrapods 5K version 1 baits from MYcroarray (Ann Arbor, MI, USA) following their MYbaits protocol v. 2.3.1 with overnight incubations (Faircloth, et al. 2012). Enriched libraries were quantified with a Qubit and pooled with other unrelated samples prior to sequencing on an Illumina HiSeq 3000 to produce paired-end reads of ≤ 151 bases. The unenriched samples were sequenced on a separate run using a single lane of Illumina HiSeq 2500. All samples were demultiplexed with Illumina software fastq2bcl. Reads were quality filtered by removing any potential adapter sequence and trimming read ends once the average Phred quality over a four base window score dropped below 20 using the Fastx toolkit (Gordon and Hannon 2010).

Quality filtered raw sequence reads were assembled into contigs using the Trinity assembler (Grabherr, et al. 2011) and a minimum kmer coverage of 2, and we used Phyluce to identify those assembled contigs that were UCE loci. We also harvested UCE loci from *Eptesicus fuscus* (GCA_000308155.1), *Myotis brandtii* (GCA_000412655.1), *M. davidii* (GCA_000327345.1), and *M. lucifugus* (GCF_000147115.1) genome assemblies using the Phyluce package (Faircloth 2016). Once extracted from Trinity and genome assemblies, we aligned all UCE loci MAFFT (Katoh, et al. 2002), trimmed the aligned data with gBlocks (Castresana 2000). Repetitive sequences (i. e. transposable elements) in each alignment were identified with RepeatMasker and trimmed where found.

<u>Mitochondrial genome assembly and annotation</u> – Raw reads from the unenriched libraries were used to generate mitochondrial genomes via MitoBim (Hahn, et al. 2013) in most cases. This program used MIRA (B, et al. 1999) to map reads to a *M. brandtii* reference genome (Genbank accession

bioRxiv preprint doi: https://doi.org/10.1101/112581; this version posted March 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

number KT210199.1). Alternative methods of mitochondrial genome assembly were used when MitoBim assembly failed. These taxa include *M*. *albescens* (TK61766), *M*. *albescens* (TK 101723), *M*. *albescens* (RDS 7889), *M*. *fortidens*, M. *keeni*, *M*. *melanorhinus*, *M*. *nigricans* (QCAZ 9601), *M*. *septentrionalis*, *M*. *simus*, *M*. *velifer*, and *M*. *volans*. For these samples, we first identified reads that were mitochondrial in origin using BLAST searches against the *M. brandtii* mitochondrial genome (KT210199.1). Those reads were assembled using Trinity v2.2.0 with the –single option. For taxa where we either could not assemble useable mitochondrial genomes, we retrieved proxy data from GenBank as follows: *M. brandti*i (KT210199.1), *E. fuscus* (KF111725.1), *M. lucifugus* (KP273591.1), and *M. davidii* (KM233172.1).

Once assembled, each mitogenome was annotated via MITOS (Bernt, et al. 2013). Annotated genes were manually validated via BLAST to confirm sequence identity and length. Protein coding genes were checked for stop codons using EMBOSS's transeq program (Rice, et al. 2000). When a stop codon was found, we used the raw reads to verify the sequence. We used BWA v0.7.12 (Li and Durbin 2009) to align the reads to the Mitobim assembled mitogenome to verify base calls from Mitobim. The protein coding rRNA and tRNA genes from each assembly were aligned using MUSCLE and concatenated into a single alignment for phylogenetic analyses, as described below.

Initial phylogenetic analyses **-** Initial phylogenies derived from UCE loci and mitochondrial coding regions were generated using maximum likelihood and Bayesian methodologies. For the first round of phylogenetic analyses all UCE loci present in 20 or more taxa were concatenated into a single alignment. PartitionFinder v1.1.1 (Lanfear, et al. 2012) was used to identify and combine loci into an optimal partitioning scheme using the hcluster heuristic algorithm. We assumed a GTR+Γ model for all loci (Darriba and Posada 2015). Initial trees were generated using RAxML v7.4.1 (Stamatakis 2006) with linked branch lengths RaxML (v8.1.3) was used to estimate and score the maximum likelihood phylogeny with the rapid bootstrapping option and 1,000 bootstrap replicates. We define strongly supported bipartitions as those present in 95-100% of bootstrap replicates and moderately supported bipartitions are present in 85-95% of bipartitions (Wiens, et al. 2008). A Bayesian phylogeny was generated with the MPI version of ExaBayes (v1.4.1) using two independent runs of 4 chains each. ExaBayes runs were terminated after 1 million generations only if the average standard deviation of split frequencies was less than 0.01. The first 25% of samples were discarded after which every 100$^{th}$ generation was sampled. The "-M 3" option was used to reduce the memory footprint of all ExaBayes runs. Proper sampling, post burn-in was inspected via Tracer v1.6. (Rambaut, et al. 2014). Effective sample sizes greater than 200

bioRxiv preprint doi: https://doi.org/10.1101/112581; this version posted March 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

were considered acceptable. Posterior probability values greater than 95% were considered to be significant.

The mitochondrial phylogeny was generated using methods similar to those described above with the following exceptions. All 37 coding regions, including protein coding genes, tRNA and rRNAs, were concatenated into a single alignment. Genes were partitioned individually except in the instances where two genes overlapped. These regions were partitioned separately from the individual genes resulting in three partitions for the two genes: a partition for gene A, a partition for gene B, and a partition for the overlapping nucleotides of gene A and B. The fast bootstrapping search in RAxML was run for 10K replicates and Bayesian analyses were performed across four independent runs with four chains of ten million generations.

The mitochondrial and nuclear phylogenies were compared to each other to identify discordance between marker types. Site-log likelihood scores were calculated for the mitochondrial alignments when constrained to the nuclear tree while the nuclear alignment was scored against the mitochondrial tree using RAxML. Model parameters were then recalculated for each constrained alignment (-f G). Site-log likelihood scores were analyzed using CONSEL (Shimodaira and Hasegawa 2001) and compared using the approximately unbiased test (Shimodaira 2002). P-values less than 0.05 were used to indicate that the trees produced by mitochondrial or nuclear data differed.

*Extended phylogenetic analyses* - Many factors can influence phylogenetic inference. To reduce the likelihood of any single factor influencing the results, we re-examined the nuclear UCE dataset using the strategies described below. A flow chart of analyses is available in Supplemental Figure 1.

Aligned UCE loci were binned based on the number of taxa represented in the alignment (phyluce_align_get_only_loci_with_min_taxa; Faircloth 2016), or degree of completeness. Groups included loci present in 100% (number of specimens (n) = 37), 95% (n = 35), 85% (n = 31), 75% (n = 27), 65% (n = 24), 55% (n = 20), 45% (n = 16), 35% (n = 12), 25% (n = 9), and 15% (n = 5) of specimens examined. These 10 groups were non-exclusive, so a locus that was assembled in all specimens (100% complete) would also be included with loci present in only 55% of specimens. On the other hand, a locus found in only 55% of specimens would not be included in the 100% complete data set. Each set of UCE alignments was concatenated using phyluce_align_format_nexus_files_for_raxml and a nexus character block was created using the phyluce_align_format_nexus_files_for_raxml –charsets option. These datasets then served as the basis for downstream phylogenetic analyses. For example, when a partitioning methodology (discussed below) was tested, it was performed for each of the 100%, 95%,

bioRxiv preprint doi: https://doi.org/10.1101/112581; this version posted March 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

85%, etc. alignments. In addition to partitioning schemes, the effect of missing data was examined using Bayesian and maximum likelihood methods.

Concatenated alignments were analyzed using three different partitioning schemes. Unpartitioned alignments were simply concatenated UCE loci treated as a single genetic unit (No Partitions). Fully partitioned alignments were concatenated alignments of UCE loci that were partitioned by locus (All Partitions). Finally, PartitionFinder v1.1.1 (Lanfear, et al. 2012) was used to combine individual loci into an optimal partitioning scheme (Optimal Partitions). Rather than searching for best-fit substitution models for each UCE locus or partition, the GTR+Γ model of sequence evolution was assigned to all loci (Darriba and Posada 2015). Initial trees for PartitionFinder were generated using RAxML v7.4.1 (Stamatakis 2006) with linked branch lengths. Partitioning schemes were heuristically searched using the hcluster algorithm.

Maximum likelihood trees were inferred for the concatenated alignments using RAxML v8.1.3 (Stamatakis 2014). The three partitioning schemes (described above) were applied to each analysis. The best scoring (lowest -lnL) tree from each dataset was identified from 100 random starting trees and bootstrapped 100 times using the GTR+Γ in both cases. The autoMRE function in RAxML v8.1.3 was used to determine the need for additional bootstrap replicates beyond the initial 100 (Pattengale, et al. 2009). A stopping criterion was set *a priori* if the weighted Robinson-Foulds distance was less than 5% in 95% of random permutations of computed bootstrap replicates (Pattengale, et al. 2009). If necessary, an additional 100 bootstrap replicates were computed until the convergence stopping criteria were met. Finally, bipartition frequencies of bootstrap replicates were drawn onto the best scoring tree from the initial RAxML searches for each of the respective data sets.

Bayesian analyses were conducted using ExaBayes v1.4.1 (Aberer, et al. 2014). For all Bayesian analyses four independent runs of four chains each were run in parallel for a minimum of one hundred thousand generations sampling every thousandth generation and applying a GTR+Γ substitution model for each partition. Two of the partitioning schemes (described above) were used for each analysis: No Partitions and Optimal Partitions. After one hundred thousand generations, analyses continued until the standard deviation of the split frequency between chains was less than 0.01. An extended majority rule consensus tree was created from all trees after the first 25% of trees were discarded using TreeAnnotator v1.7.0 (Rambaut and Drummond 2013) and parameter estimates across all runs were calculated with Tracer v1.6 (Rambaut, et al. 2014).

*Sampling loci by number of informative characters* – Previous coalescent analyses of UCE data have shown that sub-sampling the most informative loci can result in different topologies (Meiklejohn,

bioRxiv preprint doi: https://doi.org/10.1101/112581; this version posted March 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

et al. 2016). Under these assumptions, UCE loci were sorted into ten groups based on their length and the predicted correlation between length and number of informative characters was confirmed (Supplemental Figure 2). UCE loci in the same size cohort were combined into a single alignment. Rather than using coalescent based analyses we used concatenation of UCE loci to identify different topologies based on length. UCE loci were individually partitioned and the maximum likelihood tree was estimated with the rapid bootstrapping option in RaxML (bootstrap replicates = 100) using the GTR+Γ substitution model.

*Random sampling of loci* – In large phylogenetic analyses, systematic error can result in highly supported, but incorrect topologies as a result of compounding non-phylogenetic signal (Rodríguez-Ezpeleta, et al. 2007). By randomly reducing the dataset and replicating the ML analyses, we can reduce the potential effects of compounding error. Roughly 10% of the dataset, 365 loci, were randomly sampled and concatenated to create 100 new alignments. ML methods were similar to those used when sampling loci by the number of informative characters.

*Summary methods* – Gene trees for individual UCE loci recovered in five or more taxa were inferred using the GTR+Γ substitution model and fast bootstrapping (-f a) option in RAxML (replicates = 1,000). In general, gene trees were classified based on the degree of completeness (i.e. number of taxa represented) similar to the way we treated individuals as described above.

Species trees were estimated and bootstrapped using three different programs. ASTRAL-II v4.10 (Mirarab and Warnow 2015) was used to build a summary tree. Support values for bipartitions in the tree were generated from 100 bootstrap replicates using site as well as site and locus resampling (Seo 2008). Species trees were estimated from ASTRID v1.4 (Vachaspati and Warnow 2015) using bionj and bootstrapped for 100 replicates. SVDquartets (Chifman and Kubatko 2014), as implemented in PAUP v4.0a150 (Swofford 2003), was used to estimate a species trees from a random subset of 200,000 quartets and 1,000 bootstrap replicates.

Errors in gene tree estimation may reduce the accuracy of summary methods (Liu, et al. 2009; Leaché and Rannala 2011; DeGiorgio and Degnan 2014; Mirarab, et al. 2016). We used weighted (Bayzid, et al. 2015) and unweighted (Mirarab, et al. 2014) statistical binning to combine gene trees into compatible supergenes using the pipeline available in Bayzid et al. (Bayzid, et al. 2015). The gene trees used for the summary tree methods described above were used rather than re-estimating trees. Bifurcations supported by more than 50% of the bootstrap replicates were retained for each gene tree. Alignments from compatible trees were concatenated into a single supergene alignment. Trees for supergenes were estimated using RAxML. The best trees for each supergene, as defined by log

bioRxiv preprint doi: https://doi.org/10.1101/112581; this version posted March 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

likelihood score, were retained from 500 searches. Bipartition support was estimated from 500 bootstrap replicates. For all analyses, the GTR+Γ model of substitution was used and each gene in the supergene alignment was partitioned separately. The resulting supertrees were then used for species tree estimation using ASTRAL-II. For the unweighted analysis, all supertrees were included in the pool of trees. For the weighted analysis, supertrees were weighted according to the number of genes combined in the supergene alignment. For example, if a supergene was a composite of six genes, the supertree was present 6 times compared to a composite of five genes which would be represented only five times. Support for the weighted and unweighted species trees was estimated by site and site and locus re-sampling (Seo 2008) for 100 bootstrap replicates in ASTRAL-II.

*Meta-analyses* - Trees recovered from all analyses were compared to each other in tree space. Unweighted Robinson-Foulds distances were calculated among all trees. This distance matrix was transformed into two dimensions using the stochastic CCA algorithm for nonlinear dimension reduction in TreeScaper v1.09 (Huang, et al. 2016). Coordinates were then visualized in R using hexagonal binning in the hexbin library v1.27.1 (Lewin-Koh 2011).

We compared the mitochondrial data to all nuclear trees. Branch lengths have different meanings based on the type of analysis. For example, ASTRAL-II branch lengths are representative of coalescent units. ASTRID doesn't even calculate branch lengths. For accurate tree comparisons, branch lengths were stripped from all trees using regular expressions and Sed. Site log-likelihood scores were calculated for each nuclear tree, without branch lengths, and using the mitochondrial alignment. Model parameters were re-estimated for each tree. Site-log likelihood scores were compared with the approximately unbiased (Shimodaira 2002) and Shimodaira-Hasegawa (Shimodaira and Hasegawa 1999) test in CONSEL and values less than 0.05 were indicative of differences. Finally, we used RAxML to generate a 85% and majority rule consensus trees from all nuclear trees.

**Acknowledgments:**

We would like to thank the following museums, collection managers, and collaborators for the tissue loans necessary to complete this work: Joseph Cook (Museum of Southwestern Biology), Museum of Vertebrate Zoology, Manuel Ruedi (Natural History Museum of Geneva), Santiago Burneo (Pontificia Universidad Catolica del Ecuador Museo de Zoologia), Heath Garner, Robert Bradley and Caleb Phillips (Texas Tech University Natural Science Research Laboratory), Link Olson (University of Alaska Museum of the North), Cody Thompson and Priscilla Tucker (University of Michigan Museum of Zoology). In addition, we would like to thank the Texas Tech HPCC (www.hpcc.ttu.edu) for providing the

bioRxiv preprint doi: https://doi.org/10.1101/112581; this version posted March 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

computational resources necessary to complete this project. This work was supported by the National Science Foundation, DEB-1355176. Additional support was provided by College of Arts and Sciences at Texas Tech University. All sequence data is available at NCBI's Short Read Archive (SRP095250). UCE contig assemblies are available at (PENDING ACCESSION NUMBER). A supplemental file contains all estimated species trees in nexus format (supMaterial-AllTrees-nexus.txt). Individual UCE alignments in fasta format and an archived file containing all individual UCE gene trees in Newick format is available through Dryad (PENDING ACCESSION NUMBER).

bioRxiv preprint doi: https://doi.org/10.1101/112581; this version posted March 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

## Figure and Table Captions

**Figure 1.** Comparison of nuclear and mitochondrial phylogenetic trees in *Myotis*. Bayesian trees generated from (A) 2,890 nuclear UCE loci and (B) 37 mitochondrial protein, tRNA, and rRNA genes (B) Posterior probability values greater than 0.95 are shown as a "*". Values below branches are percentages of maximum likelihood bootstrap replicates supporting that clade. Conflicting tips between data types (nuclear vs. mitochondrial) are indicated with lines between the topologies. The mitochondrial Bayesian and maximum likelihood trees resolved different relationships among *M. thysanodes*, *evotis*, and *keeni* as indicated by the dotted lines. Species with more than one sample are designated with a superscript that is referenced in Table 1. Specimens derived from whole genome alignments are designated with a superscript "G".

**Figure 2.** Differences between mitochondrial and nuclear topologies. (A) All trees recovered from mitochondrial genes and the extended analysis of the nuclear data were visualized in tree space using multidimensional scaling. Nuclear trees (green) were distinct from mitochondrial trees forming a large cluster. Most nuclear trees were found within a limited region of tree space. Mitochondrial trees (blue) were distinct from any of the 294 nuclear trees recovered. Tree vs tree comparisons show that most (B) nuclear trees are more similar to each other than they are to (C) mitochondrial trees. Symmetric differences are equal to twice the Robinson-Foulds distance between two trees.

**Figure 3.** Consensus tree. A consensus tree of New World Myotis was generated from 294 nuclear topologies with a threshold cutoff of 85%. Values shown above the branches represent the percentage of nuclear analyses that support a given bipartion. Previous subgeneric classifications based on morphology are listed at each tip (*Myotis* "M", *Selysius* "S", *Leuconoe* "L". Biogeographic regions are color coded. Species with more than one sample are designated with a superscript that is referenced in Table 1. Specimens derived from whole genome alignments are designated with a superscript "G". This consensus tree represents a very conservative estimate of the *Myotis* radiation.

**Table 1.** Specimens examined.

Collection abbreviations: Museum of Southwestern Biology (MSB), Museum of Vertebrate Zoology (MVZ), Natural History Museum of Geneva (MHNG), Pontificia Universidad Catolica del Ecuador Museo de Zoologia (QCAZ), Texas Tech University Natural Science Research Laboratory (TK), University of Alaska Museum of the North (UAM), University of Michigan Museum of Zoology (UMMZ)

bioRxiv preprint doi: https://doi.org/10.1101/112581; this version posted March 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

**Table 2.** General alignment information. For a subset of analyses a series of alignments were generated based on the number of taxa per locus. Thirty-seven taxa were examined so an alignment with all 37 taxa was considered 100% complete. Parsimony-informative characters make up a small portion of the total alignment. The optimum partitioning scheme was calculated with PartitionFinder.

**Sup Fig 1.** Analytical flow chart. In the initial analysis (A) mitochondrial and nuclear data were analyzed using Bayesian and maximum likelihood methods. Trees were compared using the approximately unbiased and the Shimodaira-Hasegawa tests and determined to be conflicting. Extended analysis (B) of the data used multiple methods and sampling strategies to generate 292 different phylogenetic inferences. (C) All nuclear and mitochondrial trees were compared in tree space and with topological tests. An 85% meta-consensus tree from all analyses was used to represent a conservative estimate of the *Myotis* radiation. Trees from the extended analysis were compared to the mitochondrial trees using the approximately unbiased and the Shimodaira-Hasegawa tests and determined to be conflicting.

**Sup Fig 2.** UCE loci sorted by length. The length of a UCE locus is correlated with the number of phylogenetically informative characters. UCE loci were sorted by length and ten bins of alignments were created so that the shortest loci were combined into one alignment, the next shortest set of loci were combined … etc. Parsimony informative characters made up a minor part of each alignment.

**Sup Table 1.** Tree topology tests. Trees were compared to each other using the approximately unbiased and the Shimodaira-Hasegawa tests. Likelihood scores were calculated for the mitochondrial alignment when constrained to all topologies (mitochondrial and nuclear) recovered herein. Alignment and tree topology incompatibility were identified as p-values < 0.01.

**Sup File** -AllTrees-nexus.txt – Trees generated from all analyses in nexus format.

bioRxiv preprint doi: https://doi.org/10.1101/112581; this version posted March 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

## Literature Cited


Aberer AJ, Kobert K, Stamatakis A. 2014. ExaBayes: Massively parallel Bayesian tree inference for the whole-genome era. Molecular Biology and Evolution 31:2553-2556.

Genome sequence assembly using trace signals and additional sequence information [Internet]. 1999 1 August 2015]. Available from: http://www.bioinfo.de/isb/gcb99/talks/chevreux/main.html

Bayzid MS, Mirarab S, Boussau B, Warnow T. 2015. Weighted statistical binning: Enabling statistically consistent genome-scale phylogenetic analyses. PLoS ONE 10:e0129183.

Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, Fritzsch G, Pütz J, Middendorf M, Stadler PF. 2013. MITOS: Improved de novo metazoan mitochondrial genome annotation. Molecular Phylogenetics and Evolution 69:313-319.

Bogdanowicz W, Piksa K, Tereba A. 2012. Hybridization hotspots at bat swarming sites. PLoS ONE 7:e53334.

Boyles JG, Storm JJ. 2007. The perils of picky eating: Dietary breadth is related to extinction risk in insectivorous bats. PLoS ONE 2:e672.

Caceres MC, Barclay RMR. 2000. *Myotis septentrionalis*. Mammalian Species:1-4.

Carstens BC, Dewey TA. 2010. Species delimitation using a combined coalescent and information-theoretic approach: An example from North American *Myotis* bats. Systematic Biology 59:400-414.

Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Molecular Biology and Evolution 17:540-552.

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

Crawford NG, Faircloth BC, McCormack JE, Brumfield RT, Winker K, Glenn TC. 2012. More than 1000 ultraconserved elements provide evidence that turtles are the sister group of archosaurs. Biology Letters 8:783-786.

Darriba D, Posada D. 2015. The impact of partitioning on phylogenomic accuracy.

DeGiorgio M, Degnan JH. 2014. Robustness to divergence time underestimation when inferring species trees from estimated gene trees. Systematic Biology 63:66-82.

Degnan JH, Rosenberg NA. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends in Ecology & Evolution 24:332-340.

Derti A, Roth FP, Church GM, Wu Ct. 2006. Mammalian ultraconserved elements are strongly depleted among segmental duplications and copy number variants. Nat Genet 38:1216-1220.

Douady CJ, Delsuc F, Boucher Y, Doolittle WF, Douzery EJP. 2003. Comparison of Bayesian and maximum likelihood bootstrap measures of phylogenetic reliability. Molecular Biology and Evolution 20:248-254.

Dzeverin I. 2008. The stasis and possible patterns of selection in evolution of a group of related species from the bat genus *Myotis* (Chiroptera, Vespertilionidae). Journal of Mammalian Evolution 15:123-142.

Edwards S, Bensch S. 2009. Looking forwards or looking backwards in avian phylogeography? A comment on Zink and Carrowclough 2008. Molecular Ecology 18:2930-2933.

Faircloth BC. 2016. PHYLUCE is a software package for the analysis of conserved genomic loci. Bioinformatics 32:786-788.

Faircloth BC, Branstetter MG, White ND, Brady SG. 2015. Target enrichment of ultraconserved elements from arthropods provides a genomic perspective on relationships among Hymenoptera. Molecular Ecology Resources 15:489-501.

Faircloth BC, McCormack JE, Crawford NG, Harvey MG, Brumfield RT, Glenn TC. 2012. Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. Systematic Biology:sys004.

Findley JS. 1972. Phenetic relationships among bats of the genus *Myotis*. Systematic Biology 21:31-52.

bioRxiv preprint doi: https://doi.org/10.1101/112581; this version posted March 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Fitch JH, Shump KA. 1979. *Myotis keenii*. Mammalian Species Archive 121:1-3.

Glenn TC, Nilsen R, Kieran TJ, Finger JW, Pierson TW, Bentley KE, Hoffberg S, Louha S, Garcia-De-Leon FJ, Angel del Rio Portilla M, et al. 2016. Adapterama I: Universal stubs and primers for thousands of dual-indexed Illumina libraries (iTru & iNext). bioRxiv.

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

Gordon A, Hannon G. 2010. Fastx-toolkit. http://hannonlab.cshl.edu/fastx_toolkit.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotech 29:644-652.

Green RE, Braun EL, Armstrong J, Earl D, Nguyen N, Hickey G, Vandewege MW, St. John JA, Capella-Gutiérrez S, Castoe TA, et al. 2014. Three crocodilian genomes reveal ancestral patterns of evolution among archosaurs. Science 346.

Hahn C, Bachmann L, Chevreux B. 2013. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. Nucleic Acids Research 41:e129.

Haynie ML, Tsuchiya MTN, Ospina-Garcés SM, Arroyo-Cabrales J, Medellín RA, Polaco OJ, Maldonado JE. 2016. Placement of the rediscovered *Myotis planiceps* (Chiroptera: Vespertilionidae) within the *Myotis* phylogeny. Journal of Mammalogy 97:701-712.

Hedtke SM, Townsend TM, Hillis DM. 2006. Resolution of phylogenetic conflict in large data sets by increased taxon sampling. Systematic Biology 55:522-529.

Hollister N. 1909. Two new bats from the southwestern United States: publisher not identified.

Hosner PA, Faircloth BC, Glenn TC, Braun EL, Kimball RT. 2016. Avoiding missing data biases in phylogenomic inference: An empirical study in the landfowl (Aves: Galliformes). Molecular Biology and Evolution 33:1110-1125.

Huang H, He Q, Kubatko LS, Knowles LL. 2010. Sources of error inherent in species-tree estimation: Impact of mutational and coalescent effects on accuracy and implications for choosing among different methods. Systematic Biology 59:573-583.

Huang W, Zhou G, Marchand M, Ash JR, Morris D, Van Dooren P, Brown JM, Gallivan KA, Wilgenbusch JC. 2016. TreeScaper: Visualizing and extracting phylogenetic signal from sets of trees. Molecular Biology and Evolution.

Hudson RR, Coyne JA, Huelsenbeck J. 2002. Mathematical consequences of the genealogical species concept. Evolution 56:1557-1565.

Kapusta A, Suh A, Feschotte C. 2017. Dynamics of genome size evolution in birds and mammals. Proceedings of the National Academy of Sciences 114:E1460-E1469.

Katoh K, Misawa K, Kuma Ki, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Research 30:3059-3066.

Lack JB, Roehrs ZP, Stanley CE, Ruedi M, Van Den Bussche RA. 2010. Molecular phylogenetics of *Myotis* indicate familial-level divergence for the genus *Cistugo* (Chiroptera). Journal of Mammalogy 91:976-992.

Lanfear R, Calcott B, Ho SYW, Guindon S. 2012. PartitionFinder: Combined selection of partitioning schemes and substitution models for phylogenetic analyses. Molecular Biology and Evolution 29:1695-1701.

Larsen RJ, Knapp MC, Genoways HH, Khan FAA, Larsen PA, Wilson DE, Baker RJ. 2012. Genetic diversity of neotropical *Myotis* (Chiroptera: Vespertilionidae) with an emphasis on South American species. PLoS ONE 7:e46578.

LaVal RK. 1973. A revision of the Neotropical bats of the genus *Myotis*: Natural History Museum, Los Angeles County.

bioRxiv preprint doi: https://doi.org/10.1101/112581; this version posted March 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Leaché AD, Rannala B. 2011. The accuracy of species tree estimation under simulation: A comparison of methods. Systematic Biology 60:126-137.

Leavitt DH, Marion AB, Hollingsworth BD, Reeder TW. 2017. Multilocus phylogeny of alligator lizards (Elgaria, Anguidae): Testing mtDNA introgression as the source of discordant molecular phylogenetic hypotheses. Molecular Phylogenetics and Evolution.

Lewin-Koh N. 2011. Hexagon binning: an overview. ftp://ftp.naist.jp/pub/lang/R/CRAN/web/packages/hexbin/vignettes/hexagon_binning.pdf.

Li G, Davis BW, Eizirik E, Murphy WJ. 2016. Phylogenomic evidence for ancient hybridization in the genomes of living cats (Felidae). Genome Research 26:1-11.

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

Liu L, Yu L, Kubatko L, Pearl DK, Edwards SV. 2009. Coalescent methods for estimating phylogenetic trees. Molecular Phylogenetics and Evolution 53:320-328.

Maddison WP, Knowles LL. 2006. Inferring phylogeny despite incomplete lineage sorting. Systematic Biology 55:21-30.

Manning RW, Jones JK. 1989. *Myotis evotis*. Mammalian Species Archive 329:1-5.

McCormack JE, Harvey MG, Faircloth BC, Crawford NG, Glenn TC, Brumfield RT. 2013. A Phylogeny of birds based on over 1,500 loci collected by target enrichment and high-throughput sequencing. PLoS ONE 8:e54848.

McGee MD, Faircloth BC, Borstein SR, Zheng J, Darrin Hulsey C, Wainwright PC, Alfaro ME. 2016. Replicated divergence in cichlid radiations mirrors a major vertebrate innovation. Proceedings of the Royal Society B: Biological Sciences 283.

Meiklejohn KA, Faircloth BC, Glenn TC, Kimball RT, Braun EL. 2016. Analysis of a rapid evolutionary radiation using Ultraconserved Elements: Evidence for a bias in some multispecies coalescent methods. Systematic Biology 65:612-627.

Miller Jr GS, Allen G. 1928. The American bats of the genus *Myotis* and *Pixonyx*. Bulletin of the United States National Museum 144:1-218.

Mirarab S, Bayzid MS, Boussau B, Warnow T. 2014. Statistical binning enables an accurate coalescent-based estimation of the avian tree. Science 346.

Mirarab S, Bayzid MS, Warnow T. 2016. Evaluating summary methods for multilocus species tree estimation in the presence of incomplete lineage sorting. Systematic Biology 65:366-380.

Mirarab S, Warnow T. 2015. ASTRAL-II: Coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. Bioinformatics 31:i44-i52.

Morales AE, Jackson ND, Dewey TA, O'Meara BC, Carstens BC. 2016. Speciation with gene flow in North American *Myotis* bats. Systematic Biology.

O'Farrell MJ, Studier EH. 1980. *Myotis thysanodes*. Mammalian Species 137:1-5.

Pattengale ND, Alipour M, Bininda-Emonds ORP, Moret BME, Stamatakis A. 2009. How many bootstrap replicates are necessary? In: Batzoglou S, editor. Research in Computational Molecular Biology: 13th Annual International Conference, RECOMB 2009, Tucson, AZ, USA, May 18-21, 2009. Proceedings. Berlin, Heidelberg: Springer Berlin Heidelberg. p. 184-200.

Piaggio AJ, Valdez EW, Bogan MA, Spicer GS. 2002. Systematics of *Myotis occultus* (Chiroptera: Vespertilionidae) inferred from sequences of two mitochondrial genes. Journal of Mammalogy 83:386-395.

Platt RN, Vandewege MW, Kern C, Schmidt CJ, Hoffmann FG, Ray DA. 2014. Large numbers of novel miRNAs originate from DNA transposons and are coincident with a large species radiation in bats. Molecular Biology and Evolution 31:1536-1545.

bioRxiv preprint doi: https://doi.org/10.1101/112581; this version posted March 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Platt RN, Zhang Y, Witherspoon DJ, Xing J, Suh A, Keith MS, Jorde LB, Stevens RD, Ray DA. 2015. Targeted capture of phylogenetically informative Ves SINE Insertions in genus *Myotis*. Genome Biology and Evolution 7:1664-1675.

Pritham EJ, Feschotte C. 2007. Massive amplification of rolling-circle transposons in the lineage of the bat *Myotis* lucifugus. Proceedings of the National Academy of Sciences 104:1895-1900.

Puechmaille SJ, Allegrini B, Boston ESM, Dubourg-Savage M-J, Evin A, Knochel A, Le Bris Y, Lecoq V, Lemaire M, Rist D, et al. 2012. Genetic analyses reveal further cryptic lineages within the *Myotis nattereri* species complex. Mammalian Biology 77:224-228.

Rambaut A, Drummond A. 2013. TreeAnnotator v1. 7.0.

Rambaut A, Suchard M, Xie D, Drummond A. 2014. Tracer v1.6.

Ray DA, Feschotte C, Pagan HJT, Smith JD, Pritham EJ, Arensburger P, Atkinson PW, Craig NL. 2008. Multiple waves of recent DNA transposon activity in the bat, *Myotis lucifugus*. Genome Research 18:717-728.

Ray DA, Pagan HJT, Thompson ML, Stevens RD. 2007. Bats with hATs: Evidence for recent DNA transposon activity in genus *Myotis*. Molecular Biology and Evolution 24:632-639.

Rice P, Longden I, Bleasby A. 2000. EMBOSS: the European molecular biology open software suite. In: Elsevier Current Trends.

Rodríguez-Ezpeleta N, Brinkmann H, Roure B, Lartillot N, Lang BF, Philippe H. 2007. Detecting and overcoming systematic errors in genome-scale phylogenies. Systematic Biology 56:389-399.

Roehrs ZP, Lack JB, Van Den Bussche RA. 2010. Tribal phylogenetic relationships within Vespertilioninae (Chiroptera: Vespertilionidae) based on mitochondrial and nuclear sequence data. Journal of Mammalogy 91:1073-1092.

Ruedi M, Mayer F. 2001. Molecular systematics of bats of the genus *Myotis* (Vespertilionidae) suggests deterministic ecomorphological convergences. Molecular Phylogenetics and Evolution 21:436-448.

Ruedi M, Stadelmann B, Gager Y, Douzery EJP, Francis CM, Lin L-K, Guillén-Servent A, Cibois A. 2013. Molecular phylogenetic reconstructions identify East Asia as the cradle for the evolution of the cosmopolitan genus *Myotis* (Mammalia, Chiroptera). Molecular Phylogenetics and Evolution 69:437-449.

Sanderson MJ, Shaffer HB. 2002. Troubleshooting molecular phylogenetic analyses. Annual Review of Ecology and Systematics 33:49-72.

Seo T-K. 2008. Calculating bootstrap probabilities of phylogeny using multilocus sequence data. Molecular Biology and Evolution 25:960-971.

Shimodaira H. 2002. An Approximately Unbiased test of phylogenetic tree selection. Systematic Biology 51:492-508.

Shimodaira H, Hasegawa M. 2001. CONSEL: For assessing the confidence of phylogenetic tree selection. Bioinformatics 17:1246-1247.

Shimodaira H, Hasegawa M. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic iInference. Molecular Biology and Evolution 16:1114.

Simmons NB. 2005. Order Chiroptera. Mammal species of the world: a taxonomic and geographic reference 1:312-529.

Simões BF, Rebelo H, Lopes RJ, Alves PC, Harris DJ. 2007. Patterns of genetic diversity within and between *Myotis d. daubentonii* and *M. d. nathalinae* derived from cytochrome *b* mtDNA sequence data. Acta Chiropterologica 9:379-389.

Sota T. (Sota2002 co-authors). 2002. Radiation and reticulation: extensive introgressive hybridization in the carabid beetles *Ohomopterus* inferred from mitochondrial gene genealogy. Population Ecology 44:0145-0156.

bioRxiv preprint doi: https://doi.org/10.1101/112581; this version posted March 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Stadelmann B, Lin LK, Kunz TH, Ruedi M. 2007. Molecular phylogeny of New World *Myotis* (Chiroptera, Vespertilionidae) inferred from mitochondrial and nuclear DNA genes. Molecular Phylogenetics and Evolution 43:32-48.

Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688-2690.

Stamatakis A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312-1313.

Swofford DL. 2003. PAUP*. Phylogenetic analysis using parsimony (* and other methods). Version 4.

Thomas J, Sorourian M, Ray D, Baker RJ, Pritham EJ. 2011. The limited distribution of Helitrons to vesper bats supports horizontal transfer. Gene 474:52-58.

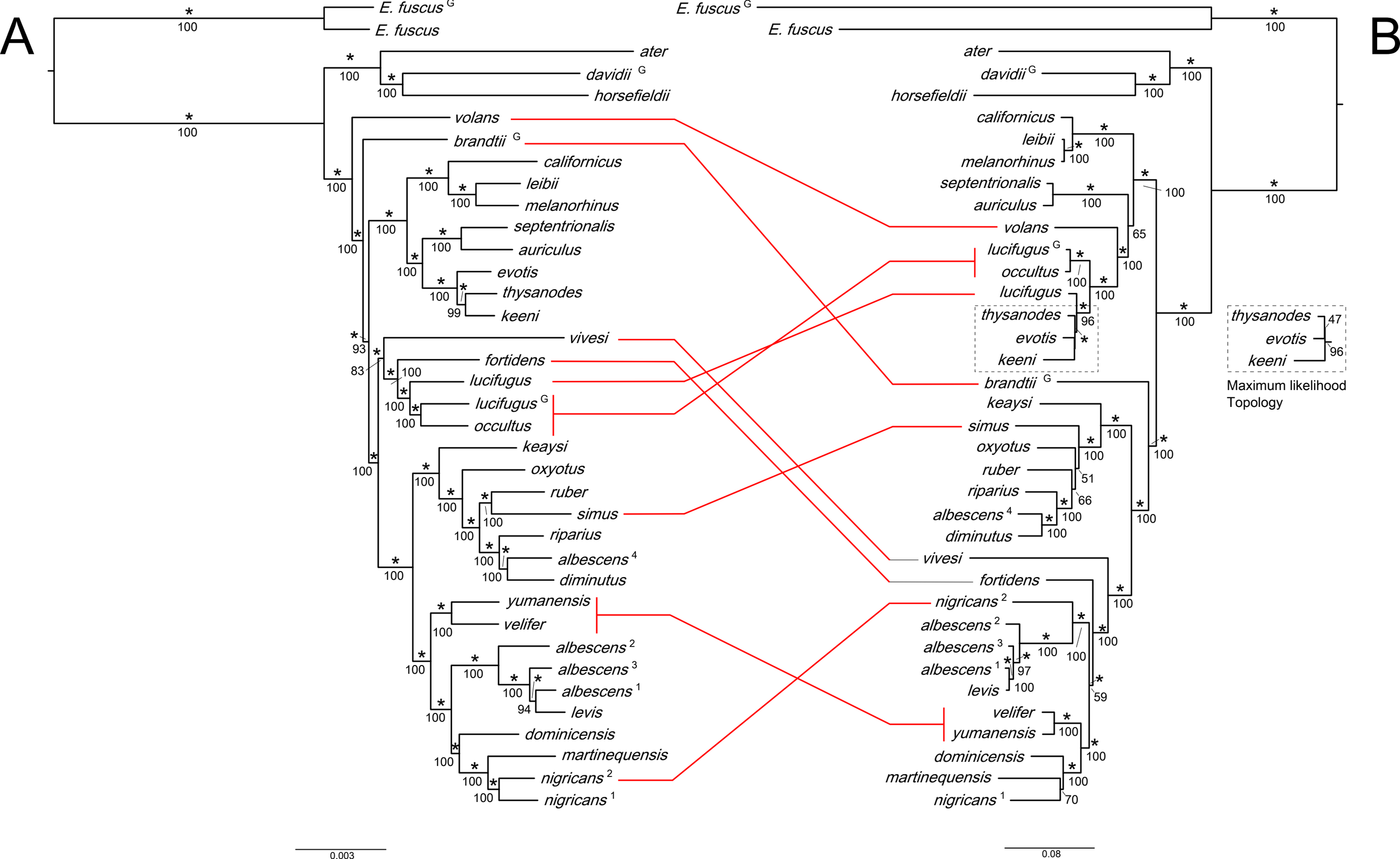
Vachaspati P, Warnow T. (Vachaspati2015 co-authors). 2015. ASTRID: Accurate Species TRees from Internode Distances. BMC Genomics 16:S3.

Valdez EW, Choate JR, Bogan MA, Yates TL. 1999. Taxonomic status of *Myotis occultus*. Journal of Mammalogy 80:545-552.

Warner RM. 1982. *Myotis auriculus*. Mammalian Species 191:1-3.

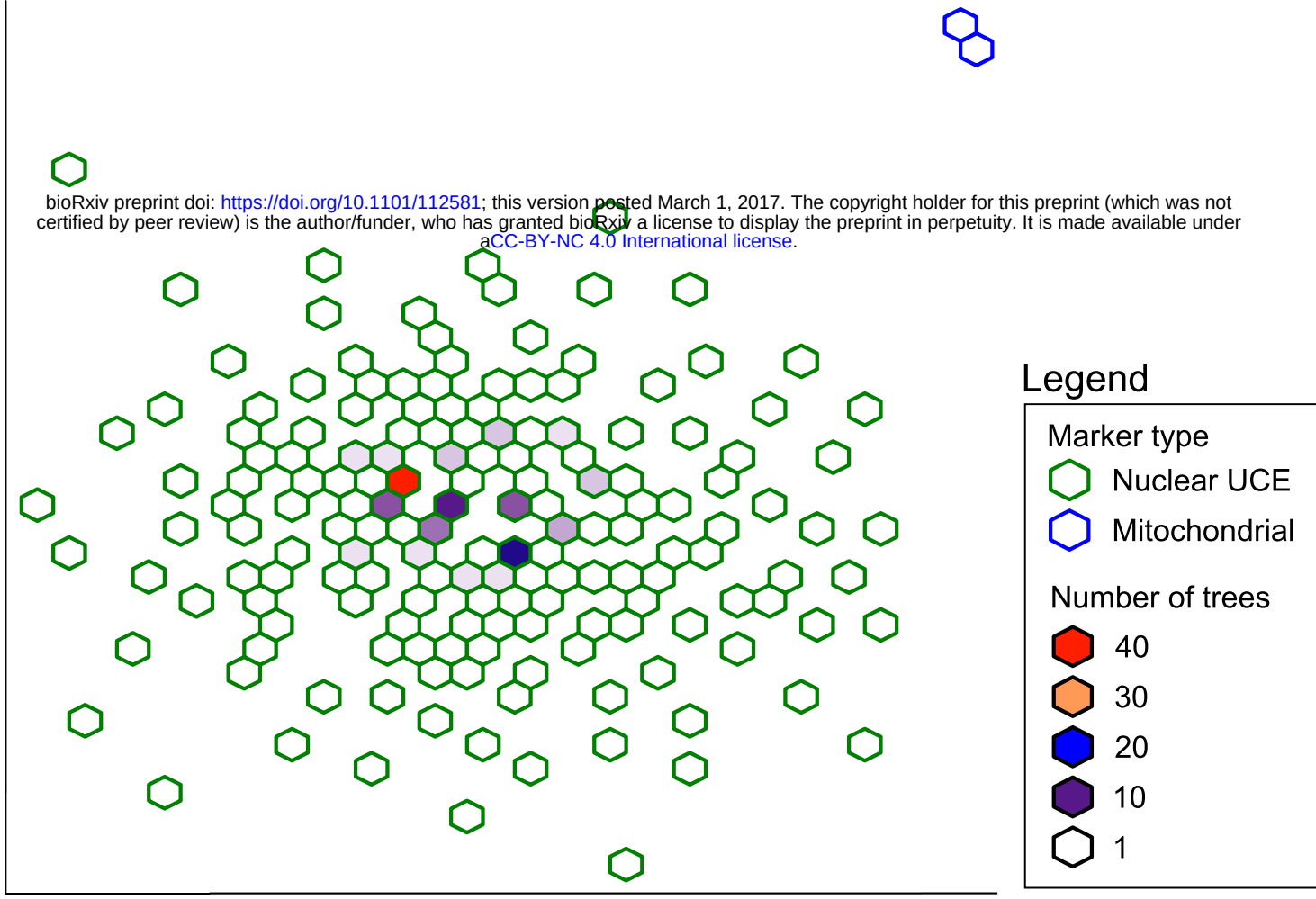
Wiens JJ, Kuczynski CA, Smith SA, Mulcahy DG, Sites JJW, Townsend TM, Reeder TW. 2008. Branch lengths, support, and congruence: Testing the phylogenomic approach with 20 nuclear loci in snakes. Systematic Biology 57:420-431.

Willis SC, Farias IP, Ortí G. 2014. Testing mitochondrial capture and deep coalescence in Amazonian cichlid fishes (Cichlidae: *Cichla*). Evolution 68:256-268.

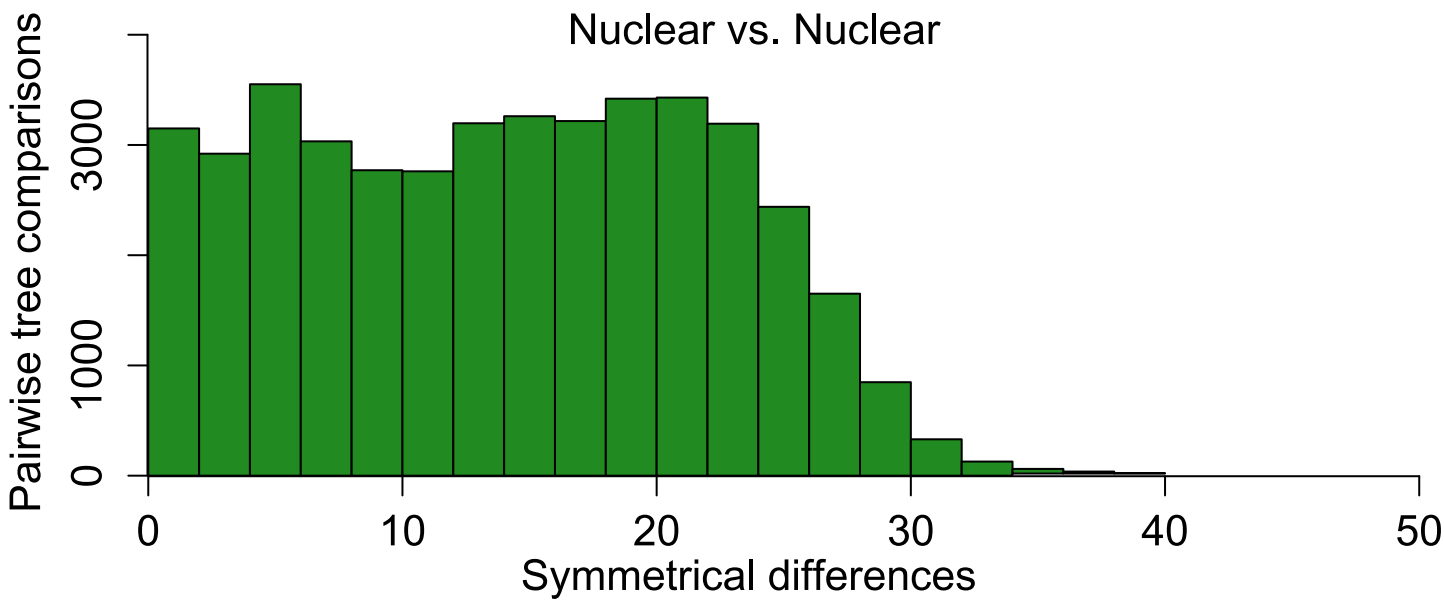

A
B
E. fuscus G
E. fuscus
ater
davidii G
horsefieldii
volans
brandtii G
californicus
leibii
melanorhinus
septentrionalis
auriculus
evotis
thysanodes
keeni
vivesi
fortidens
lucifugus
lucifugus G
occultus
keaysi
oxyotus
ruber
simus
riparius
albescens 4
diminutus
yumanensis
velifer
albescens 2
albescens 3
albescens 1
levis
dominicensis
martinequensis
nigricans 2
nigricans 1
0.003
0.08
Maximum likelihood
Topology

bioRxiv preprint doi: https://doi.org/10.1101/112581; this version posted March 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

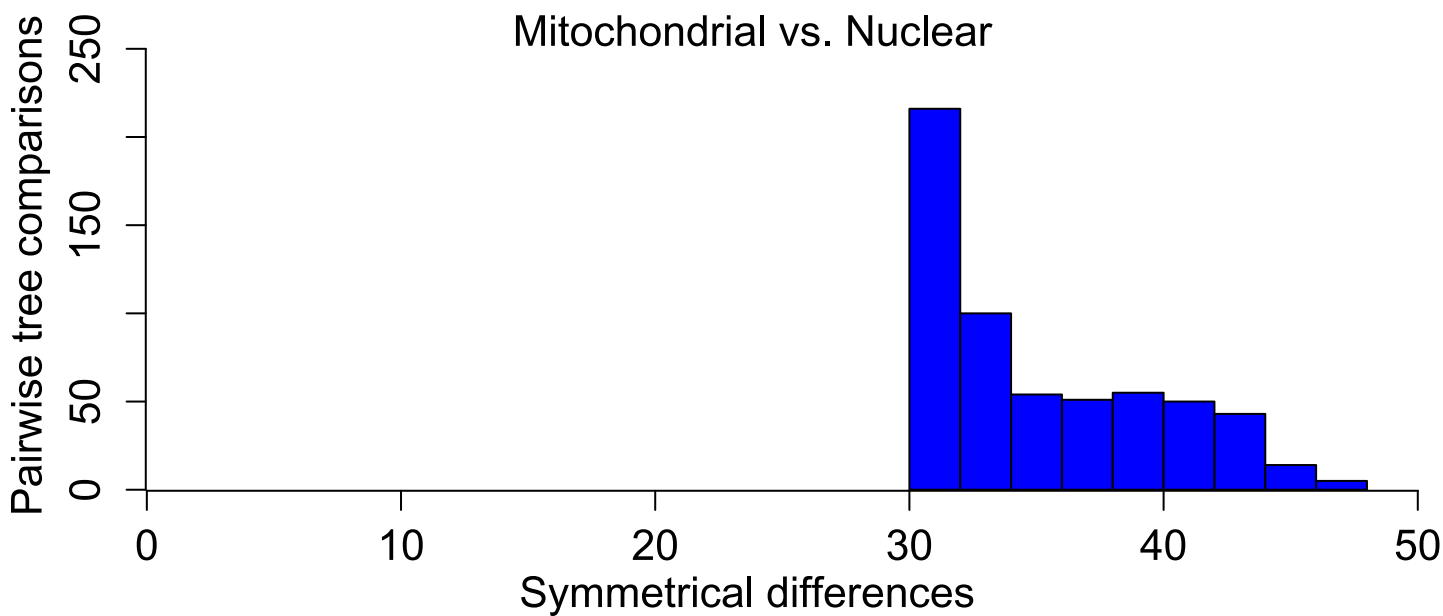
A

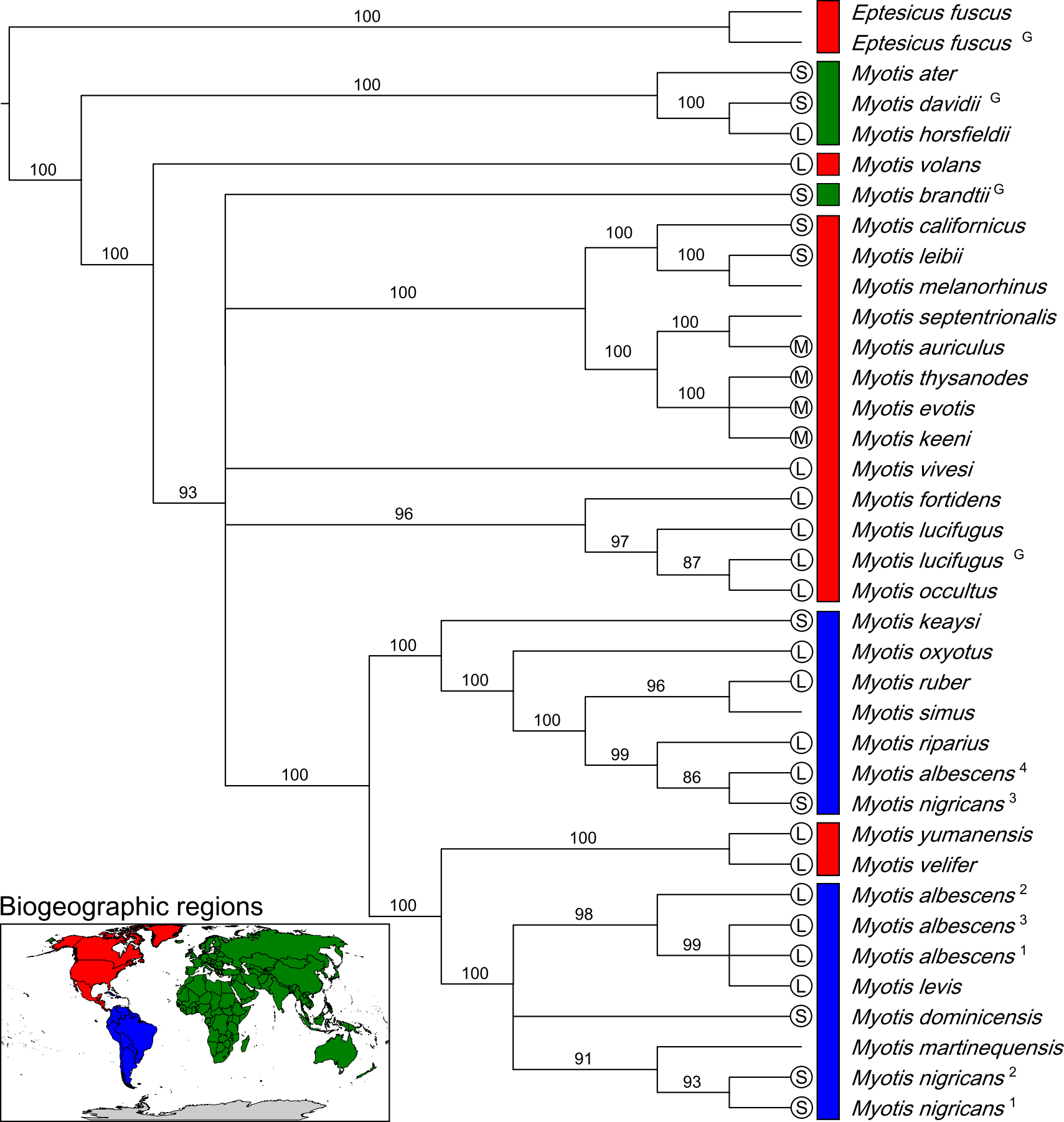



B




C




Eptesicus fuscus
Eptesicus fuscus G
Myotis ater
Myotis davidii G
Myotis horsfieldii
Myotis volans
Myotis brandtii G
Myotis californicus
Myotis leibii
Myotis melanorhinus
Myotis septentrionalis
Myotis auriculus
Myotis thysanodes
Myotis evotis
Myotis keeni
Myotis vivesi
Myotis fortidens
Myotis lucifugus
Myotis lucifugus G
Myotis occultus
Myotis keaysi
Myotis oxyotus
Myotis ruber
Myotis simus
Myotis riparius
Myotis albescens 4
Myotis nigricans 3
Myotis yumanensis
Myotis velifer
Myotis albescens 2
Myotis albescens 3
Myotis albescens 1
Myotis levis
Myotis dominicensis
Myotis martinequensis
Myotis nigricans 2
Myotis nigricans 1
Biogeographic regions

bioRxiv preprint doi: https://doi.org/10.1101/112581; this version posted March 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

**Table 1 - Specimens Examined**

| Genus | Specific epithet | Name herein | Museum identification num. | Num. UCE Loci | UCE contig accession | Mitochondrial genome accession |
|---|---|---|---|---|---|---|
| *Eptesicus* | *fuscus* | *fuscus* | TK 178736 | 2,849 | Pending | Pending |
| *Eptesicus* | *fuscus* | *fuscus* [G] | GCA_000308155.1 | 2,467 | Pending | Pending |
| *Myotis* | *albescens* | *albescens* [1] | RDS 7889 | 2,764 | Pending | Pending |
| *Myotis* | *albescens* | *albescens* [2] | QCAZ 9157 | 1,185 | Pending | Pending |
| *Myotis* | *albescens* | *albescens* [3] | TK 61766 | 2,872 | Pending | Pending |
| *Myotis* | *albescens* | *albescens* [4] | TK 101723 | 2,990 | Pending | Pending |
| *Myotis* | *atacamensis* | *atacamensis* | M4430 | 2,774 | Pending | Pending |
| *Myotis* | *auriculus* | *auriculus* | MSB 40883 | 2,229 | Pending | Pending |
| *Myotis* | *brandtii* | *brandtii* [G] | GCA_000412655.1 | 2,446 | Pending | Pending |
| *Myotis* | *septentrionalis* | *septentrionalis* | RDS 7705 | 2,916 | Pending | Pending |
| *Myotis* | *californicus* | *californicus* | UMMZ 175828 | 2,948 | Pending | Pending |
| *Myotis* | *davidii* | *davidii* [G] | GCA_000327345.1 | 2,450 | Pending | Pending |
| *Myotis* | *dominicensis* | *dominicensis* | TK 15624 | 2,576 | Pending | Pending |
| *Myotis* | *evotis* | *evotis* | MSB 47323 | 2,586 | Pending | Pending |
| *Myotis* | *fortidens* | *fortidens* | MSB 54941 | 2,791 | Pending | Pending |
| *Myotis* | *horsfieldii* | *horsefeldi* | MHNG 1926.039 | 3,017 | Pending | Pending |
| *Myotis* | *keaysi* | *keaysi* | TK 13525 | 3,195 | Pending | Pending |
| *Myotis* | *keenii* | *keenii* | UAM 113849 | 2,723 | Pending | Pending |
| *Myotis* | *leibii* | *leibii* | TK 24872 | 3,119 | Pending | Pending |
| *Myotis* | *levis* | *levis* | RDS 7781 | 2,538 | Pending | Pending |
| *Myotis* | *lucifugus* | *lucifugus* | MSB 46679 | 2,736 | Pending | Pending |
| *Myotis* | *lucifugus* | *lucifugus* [G] | GCA_000147115.1 | 2,429 | Pending | Pending |
| *Myotis* | *martiniquensis* | *martiniquensis* | TK 151413 | 856 | Pending | Pending |
| *Myotis* | *melanorhinus* | *melanorhinus* | M8944 | 3,177 | Pending | Pending |
| *Myotis* | *nigricans* | *nigricans* [1] | QCAZ 9601 | 2,854 | Pending | Pending |
| *Myotis* | *nigricans* | *nigricans* [2] | RDS 7791 | 3,159 | Pending | Pending |
| *Myotis* | *nigricans* or *diminutus* | *nigricans* [3] | QCAZ 9168 | 3,078 | Pending | Pending |
| *Myotis* | *occultus* | *occultus* | MSB 121995 | 2,957 | Pending | Pending |
| *Myotis* | *oxyotus* | *oxyotus* | UMMZ RCO1013 | 3,106 | Pending | Pending |
| *Myotis* | *riparius* | *riparius* | TK 145199 | 2,890 | Pending | Pending |
| *Myotis* | *ruber* | *ruber* | MVZ 185692 | 2,757 | Pending | Pending |
| *Myotis* | *simus* | *simus* | TK 22688 | 2,924 | Pending | Pending |
| *Myotis* | *thysanodes* | *thysanodes* | 07LEP | 2,821 | Pending | Pending |
| *Myotis* | *velifer* | *velifer* | MSB 70877 | 2,704 | Pending | Pending |
| *Myotis* | *vivesi* | *vivesi* | MSB 42658 | 2,469 | Pending | Pending |
| *Myotis* | *volans* | *volans* | MSB 40886 | 2,819 | Pending | Pending |
| *Myotis* | *yumanensis* | *yumanensis* | RDS 7734 | 2,589 | Pending | Pending |

Collection abbreviations: Museum of Southwestern Biology (MSB), Museum of Vertebrate Zoology (MVZ), Natural History Museum of Geneva (MHNG), Pontificia Universidad Catolica del Ecuador Museo de Zoologia (QCAZ), Texas Tech University Natural Science Research Laboratory (TK), University of Alaska Museum of the North (UAM), University of Michigan Museum of Zoology (UMMZ). Samples beginning with "GCA_" represent genome assemblies available through NCBI.

**Table 2. Character information**

| Min. num. of taxa | Percent of taxa per locus | Number of Loci | Parsimony-informative characters | Variable uninformative | Alignment length | Optimum Partitions |
|---|---|---|---|---|---|---|
| 37 | 100 | 212 | 3,480 | 4,778 | 112,125 | 6 |
| 35 | 95 | 1,193 | 18,288 | 24,189 | 575,321 | 15 |
| 31 | 85 | 2,034 | 29,179 | 41,711 | 903,903 | 16 |
| 27 | 75 | 2,481 | 33,031 | 48,732 | 1,041,099 | 20 |
| 24 | 65 | 2,668 | 34,373 | 51,148 | 1,091,620 | 27 |
| 20 | 55 | 2,890 | 35,284 | 53,200 | 1,144,471 | 27 |
| 16 | 45 | 3,064 | 36,539 | 54,782 | 1,187,492 | 31 |
| 12 | 35 | 3,232 | 37,259 | 56,453 | 1,227,093 | 34 |
| 9 | 25 | 3,379 | 37,894 | 57,672 | 1,260,248 | 37 |
| 5 | 15 | 3,648 | 38,718 | 62,588 | 1,377,262 | 31 |

General alignment information. For a subset of analyses a series of alignments were generated based on the number of taxa per locus. Thirty-seven taxa were examined so an alignment with all 37 taxa was considered 100% complete. Parsimony-informative characters make up a small portion of the total alignment. The optimum partitioning scheme was calculated with PartitionFinder.