

## Evidence that *Myotis lucifugus* “Subspecies” are Five Nonsister Species, Despite Gene Flow

ARIADNA E. MORALES\* AND BRYAN C. CARSTENS

Department of Evolution, Ecology and Organismal Biology, Ohio State University, 318 W. 12th Avenue, Columbus, OH 43210, USA

\*Correspondence to be sent to: Department of Evolution, Ecology and Organismal Biology, Ohio State University,  
318 W. 12th Avenue, Columbus, OH 43210, USA;  
E-mail: moralesgarcia.1@osu.edu.

Received 13 September 2017; reviews returned 7 December 2017; accepted 10 February 2018

Associate Editor: Frank Burbrink

**Abstract.**—While genetic exchange between nonsister species was traditionally considered to be rare in mammals, analyses of molecular data in multiple systems suggest that it may be common. Interspecific gene flow, if present, is problematic for phylogenetic inference, particularly for analyses near the species level. Here, we explore how to detect and account for gene flow during phylogeny estimation using data from a clade of North American *Myotis* bats where previous results have led researchers to suspect that gene flow among lineages is present. Initial estimates of phylogenetic networks and species trees indicate that subspecies described within *Myotis lucifugus* are paraphyletic. In order to explore the extent to which gene flow is likely to interfere with phylogeny estimation, we use posterior predictive simulation and a novel Approximate Bayesian Computation approach based on gene tree distances. The former indicates that the species tree model is a poor fit to the data, and the latter provides evidence that a species tree with gene flow is a better fit. Taken together, we present evidence that the currently recognized *M. lucifugus* subspecies are paraphyletic, exchange alleles with other *Myotis* species in regions of secondary contact, and should be considered independent evolutionary lineages despite their morphological similarity. [*Myotis lucifugus*; phylogenomics; reticulations; species delimitation with gene flow; ultraconserved elements.]

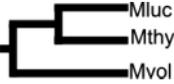
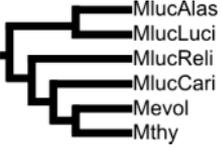
Interspecific gene flow, hybridization, and horizontal gene transfer are processes that occur throughout the evolutionary histories of many taxa (e.g., Melo-Ferreira et al. 2014; Sullivan et al. 2014; Kumar et al. 2017; Morales et al. 2017) but are difficult to incorporate into phylogenetic and species delimitation analyses (but see Camargo et al. 2012 and Jackson et al. 2017a). Therefore, genetic exchange among species has been largely ignored in phylogenetic reconstructions, despite simulation studies that demonstrate that gene flow can interfere with phylogenetic signal (Eckert and Carstens 2008; Leache et al. 2014) and parameter estimation (Leache et al. 2014). Recent methodological advances based on phylogenetic networks and explicit incorporation of gene flow may provide an opportunity to address this shortcoming (e.g., Solís-Lemus and Ané 2016; Jackson et al. 2017b).

Several strategies have been developed in recent years to account for both gene flow and incomplete lineage sorting (ILS) in a phylogenetic context. One is to infer phylogeny and secondarily use an isolation with migration approach to estimate migration rates on branches of interest (e.g., Hey 2010). The implicit assumption of this strategy is that gene flow will not interfere with phylogenetic signal. A second approach is the framework developed by Joly et al. (2009) to test for hybrid speciation. It is based on the assumption that under a scenario of only ILS, gene divergence times must predate speciation times, whereas in a case of hybridization the gene divergence events may occur after speciation. A third strategy utilizes phylogenetic networks, which offer a complement to traditional species tree methods but are generally constrained by lack of scalability

because there are limits to the number of migration (or hybridization) events that can be modeled and limits on the number of loci and individuals per species that can be analyzed (Than et al. 2008; Meng and Kubatko 2009; Solís-Lemus and Ané 2016). Finally, other researchers have assessed the extent to which gene flow should be taken into account in a particular system using simulations such as Approximate Bayesian Computation (ABC; Camargo et al. 2012) or posterior predictive simulation (PPS) (Gruenstaedl et al. 2016). The diversity of approaches is a clear sign that researchers have recognized that gene flow is a process that should be considered in lower-level phylogenetic investigations, even if it is not yet clear exactly which strategies are the most appropriate in a given empirical system.

Here, we evaluate the extent to which gene flow could be problematic for phylogeny estimation and species delimitation in North American *Myotis* bats, a clade in which paraphyletic patterns have been attributed to gene flow among species (Dewey 2006; Carstens and Dewey 2010). We specifically focus on described subspecies within *Myotis lucifugus*, a widely distributed nominal species suspected to include 4–6 independent evolutionary lineages that potentially exchange alleles with a variety of other *Myotis* (Carstens and Dewey 2010). After collecting genomic data and conducting delimitation analyses that support these suspicions, we develop a novel ABC framework that uses gene tree distance metrics to summarize the data and to calculate the relative influence of gene flow on a locus-by-locus basis. Finally, we infer phylogenetic relationships using a network approach that accommodates genetic exchange among species.

TABLE 1. Phylogenetic position of *Myotis lucifugus* in several studies

Topology	Molecular markers	Sampling (ind/sp)	Subspecies included	Observed paraphyly	References
	ND1 + Cyt-b	1	<i>Myotis lucifugus carissima</i>	No	Ruedi and Mayer (2001)
	Rag2 + Cyt-b	1	<i>M. l. carissima</i>	No	Stadelmann et al. (2007)
	Cyt-b	17–51	<i>Myotis lucifugus alascensis</i> <i>M. l. carissima</i> <i>Myotis lucifugus lucifugus</i> <i>Myotis lucifugus relictus</i>	Yes	Dewey (2006)
	6 autosomal + Cyt-b	4–8	<i>M. l. alascensis</i> <i>M. l. carissima</i> <i>M. l. lucifugus</i> <i>M. l. relictus</i>	Yes	Carstens and Dewey (2010)
	Cyt-b Rag2 + Cyt-b	1–3 1	<i>M. l. carissima</i> <i>M. l. carissima</i>	No No	Larsen et al. (2012) Ruedi et al. (2013)

Mvol = *Myotis volans*; Mthy = *M. thysanodes*; Mocc = *M. occultus*; MlucAlas = *M. lucifugus alascensis*; MlucCari = *M. l. carissima*; MlucLuci = *M. l. lucifugus*; MlucReli = *M. l. relictus*; WLE = clade of the western long-eared bats including *M. evotis*, *M. thysanodes*, and *M. keenii*.

### The *Myotis lucifugus* Complex

Inferring the species boundaries and phylogenetic relationships within *Myotis* has challenged evolutionary biologists because ecomorphs associated with phenotypic characteristics in this genus appear to have evolved independently in different subclades (Ruedi and Mayer 2001; Stadelmann et al. 2007; Ghazali et al. 2016) to the extent that several *Myotis* species are phenotypically similar without necessarily being closely related. This phenotypic convergence has the potential to influence phylogeny inference, even in well-studied species, because it complicates the recognition of species boundaries and taxon sampling. At least compared to other bat genera, *Myotis* is well-studied. For example, *M. lucifugus* a member of the North American Nearctic subclade (Stadelmann et al. 2007; Ruedi et al. 2013), has been investigated from physiological, embryological, and ecological perspectives (e.g. Kwiecinski et al. 1987; Schutt et al. 1994; Norquay and Willis 2014; Czenze and Willis 2015) and has a full genome available (Ensembl release 59; Flicek et al. 2014). Despite these efforts, there is uncertainty regarding the species and subspecies status within *M. lucifugus* (Table 1). While five subspecies are currently recognized (*Myotis lucifugus alascensis*, *Myotis lucifugus carissima*, *Myotis lucifugus lucifugus*, *Myotis lucifugus pernox*, and *Myotis lucifugus relictus*; Simmons 2005), some of these descriptions are based on geographic ranges and subtle qualitative differences in

morphological characters. For example, *M. l. pernox* was described as “...externally resembling *M. l. lucifugus*, but with larger foot ...” (Hollister 1911). Furthermore, while Simmons (2005) treats *Myotis occultus* as a distinct species, earlier authors (Barbour and Davis 1969; Valdez et al. 1999) included *occultus* as a subspecies within *M. lucifugus* on the basis of genetic and morphological similarities.

The phylogenetic position of *M. lucifugus* remains uncertain (Table 1). For example, Ruedi and Mayer (2001) used two mitochondrial markers, cytochrome-B and ND-1, to estimate phylogenetic relationships among *Myotis* bats and their findings showed *M. lucifugus* and *Myotis thysanodes* as sister species. Stadelmann et al. (2007) used cytochrome-B and a nuclear gene (Rag2) and augmented the number of taxa sampled from North America. Their results described *M. lucifugus* as the sister taxon of *M. occultus*, this group being sister to the western long-eared bats (*Myotis evotis*, *M. thysanodes*, and *Myotis keenii*). Larsen et al. (2012) studied the phylogenetic relationships of *Myotis* bats in the Americas using cytochrome-B and an extensive sampling of South American species, many of which were lacking from previous studies. Their results showed *M. lucifugus* as a sister species to the western long-eared species, and *M. occultus* as the most divergent lineage within the clade, a result also supported by Ruedi et al. (2013). The aforementioned phylogenetic studies only included

a few samples of one subspecies (i.e., *M. l. carissima*). While most authors assume that the *Myotis lucifugus* subspecies are monophyletic, genetic evidence from mitochondrial data in Dewey (2006) suggested that they may be paraphyletic.

## MATERIAL AND METHODS

### Sampling, Capture Probe Design, Library Preparation, and Sequencing

Data were collected from 131 individuals ( $n$  per species/subspecies ranged from 1 to 35) from *Myotis* species throughout North America (Fig. 1 and Table 1). All specimens included in this study were carefully identified in the field first and subsequently confirmed by curators of museum collections (Supplementary Table S1 available on Dryad at <http://dx.doi.org/10.5061/dryad9d6b2>).

A set of capture probes for ultraconserved elements (UCEs) was previously designed to infer species relationships among the western long-eared *Myotis* bats (Morales et al. 2017), which used the *Myotis lucifugus* genome as a reference (Ensembl release 59; Flicek et al. 2014), and were derived from the set of probes designed for placental mammals (McCormack et al. 2012). These loci have been informative for phylogenetic and population level inferences going from deep to shallow timescales (e.g., Smith et al. 2014; Meiklejohn et al. 2016; Morales et al. 2017). The *Myotis* probes were purchased from MYcroarray, Inc. (Ann Arbor, MI, USA) and used to enrich all samples included in this study.

Genomic DNA was isolated using a DNeasy blood and tissue kit (QIAGEN). To increase the amount of DNA in some samples that were donated by museum collections, whole genome amplifications were performed using REPLI-g kits (QIAGEN). DNA fragmentation and TruSeq library preparation followed the methodology described by Morales et al. (2017). Enrichment of capture probes was performed following the protocol provided by MYcroarray, which is based on the workflow described by Gnirke et al. (2009). The final library was sequenced at the Georgia Genomics Facility using 1.5 lanes of an Illumina NextSeq 150 bp paired-end high-output flow cell run.

### Contigs Assembly, Mapping, and Haplotype Reconstruction

Raw reads were demultiplexed using Casava (Illumina, Inc.). Quality of the reads was assessed and adapters were trimmed using illumiprocessor (Faircloth 2013) and Trimmomatic v.0.32 (Bolger et al. 2014). Consensus contigs were generated *de novo* using VelvetOptimizer v.2.2.5 (<http://www.vicbioinformatics.com/software/velvetoptimizer.shtml> last accessed 26 July 2016) and Velvet assembler v.1.2.09 (Zerbino and Birney 2008). Consensus contigs were aligned to target loci using LASTZ (Harris 2007). Any contigs that did not match the target loci or that matched more than one locus were removed. Retained contigs were mapped against an index of target loci, using BWA-MEM v. 0.7.8-r455 (Li and Durbin 2009). Individual consensus sequences were generated using SAMTools v.0.1.19 (Li et al. 2009). Files were imported from SAM to BAM format, sorted, and indexed. Files in

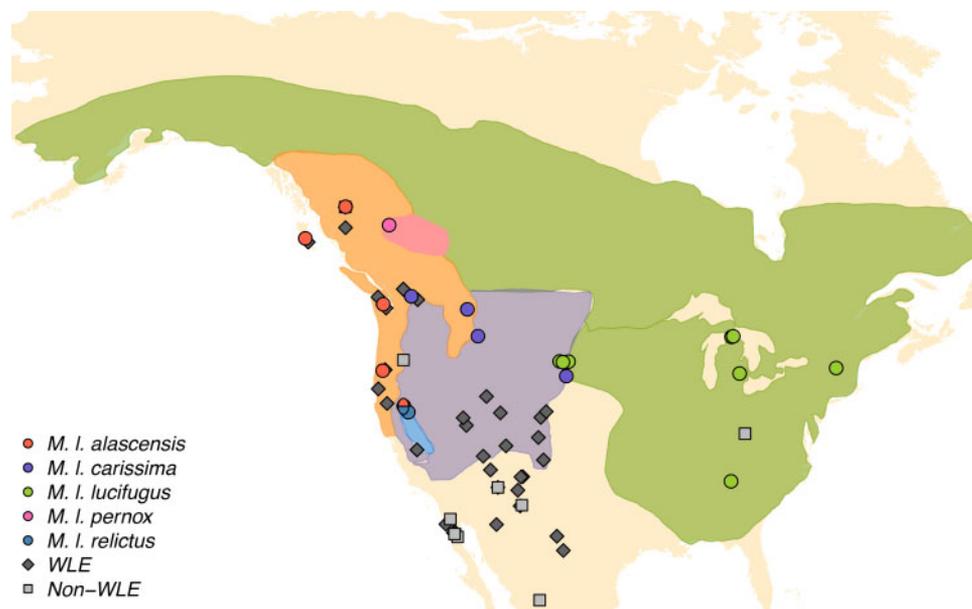


FIGURE 1. Map showing sampling localities of *Myotis* species in the Nearctic subclade. Shaded areas represent the distribution of *M. lucifugus* recognized subspecies. Circles show sampling for *M. lucifugus* lineages. Diamonds show sampling for the western long-eared (WLE) species *M. evotis*, *M. thysanodes*, and *M. keenii*. Squares show sampling for non-WLE species *M. occultus*, *M. sodalis*, and *M. volans*.

VCF and FASTQ format (with hard-masked low quality bases <Q20) were then generated and seqtk was used to covert output-masked FASTQ files to FASTA format (<https://github.com/lh3/seqtk> last accessed 26 July 2016). Each locus was aligned using MAFFT v7.123b (Kato and Standley 2013). To resolve gametic phase in multiple heterozygous sites, PHASE v.2.1.1 was used (Stephens et al. 2001; Stephens and Donnelly 2003). All bioinformatic analyses were performed through the Ohio Supercomputer Center, last accessed November 2017.

### Summary Statistics

To assess the level of polymorphism of phased loci, summary statistics were calculated for each locus across all lineages using PopGenome v.2.2.3 (Pfeifer et al. 2014). Overall, 1370 UCEs were analyzed (unless otherwise noted) to estimate the total number of segregating sites (S), number of haplotypes (Hap), haplotype diversity (Hd), nucleotide diversity ( $\pi$ ), Watterson's  $\theta$ , and Tajima's *D*.

In order to select a model of sequence evolution, maximum likelihood scores under each model were calculated using PHYML (Guindon and Gascuel 2003), and MrAIC (Nylander 2004; last accessed January 2017.) was used to calculate the Akaike's Information Criterion (AIC) and conduct model selection.

### Species Delimitation and Species Tree Reconstruction of *M. lucifugus* Lineages

Previous work suggests both that *M. lucifugus* subspecies may be evolving independently and that gene flow may be occurring among closely related *Myotis* (e.g., Carstens and Dewey 2010; Morales et al. 2017). To evaluate if *M. lucifugus* subspecies are lineages evolving independently, we performed a Bayesian analysis to simultaneously delimit lineages and infer their relationships as implemented in BPP v.3.3 (Yang and Rannala 2014). Putative lineages (specified *a priori*) are collapsed sequentially to determine both the best delimitation scenario and a species tree. BPP uses a reversible jump Markov chain Monte Carlo approach (Yang and Rannala 2010) to calculate posterior probabilities for models of several numbers of potential lineages using gene trees and incorporating uncertainty of phylogenetic inference in a Bayesian framework (Yang and Rannala 2014). For this analysis, individuals were assigned to groups following the subspecies classification. Alignments of 125 UCE loci were used as input; this is the number of loci recovered for all lineages in at least one individual. The most compelling justification for this reduced data set was that only one sample (two alleles) from *M. l. pernox*, which has a restricted distribution (Fig. 1), was available and sequencing results were suboptimal from this sample. Starting priors for theta ( $\theta$ ) were set as gamma(2,1000), and for divergence ( $\tau$ ) were set as gamma(2,1000). Two independent analyses were

performed, with 500,000 generations sampled every five steps, a burnin equivalent to 10%.

To compare the species relationships coestimated in BPP during the species delimitation analysis, a species tree was estimated in \*BEAST v.1.8.1 (Heled and Drummond 2010; Drummond et al. 2012). This analysis was performed using the same set of loci used in the delimitation analysis with the following settings: base frequencies were set to empirically estimated values, a relaxed molecular clock, a Yule process on the species tree prior with a random starting tree, 100 million generations sampled every 10,000 steps, and a burnin of 10%. The same model of sequence evolution (HKY) was used for all loci as this model was found to be representative for all of them. Convergence was assessed in Tracer v.1.6 (Rambaut et al. 2014; last accessed January 2017.) based on effective sample size (ESS) values equal or greater than 200, and maximum Clade Credibility trees were built using Tree Annotator v.1.8.1 (Heled and Drummond 2010).

### Posterior Predictive Evaluation of Species Tree Model

To assess the fit of the multispecies coalescent model (MSCM) used in \*BEAST to estimate gene trees and the species tree, PPS analyses were performed as implemented in P2C2M v.0.7.6 (Gruenstaedl et al. 2016). P2C2M compares the posterior and posterior predictive distributions of each gene tree and the species tree using PPS and uses summary statistics to calculate the differences between distributions. These summary statistics account for the number of deep coalescences (ndc), the coalescent likelihood (coal), and the likelihood of coalescent waiting times (lcwt). The expectation is no difference between distributions when the MSCM has a good fit to the data. This analysis was performed with 100 replicates.

### Estimating the Phylogenetic Placement of *M. lucifugus* Lineages

Species trees were estimated using two coalescent-based approaches. The first was SVDquartets (Chifman and Kubatko 2014, 2015), as implemented in PAUP v.4.0a149 (Swofford 2002). SVDquartets is a single-site method that computes a quartet score based on singular value decomposition of a matrix of site pattern frequencies that correspond to a split on a phylogenetic tree. Quartet scores are used to select the best-supported topology for quartets of taxa, and these scores are then used to infer the species tree. Considering that our data set has 262 tips (alleles), there are nearly 1 billion of possible quartets, conducting an exhaustive analysis of all possible quartets may have consumed thousands of computational hours. We thus used a subsampling approach, where one allele per lineage was sampled at random with replacement such that 1000 independent data sets were built and analyzed. Each analysis was performed with 100 bootstrap replicates to calculate support values. In order to summarize these trees, a

TABLE 2. Mean values and standard deviation ( $\pm$ SD) of summary statistics for 1370 loci from North American *Myotis* species

	Individuals [loci]	Sequence length in bp	Segregating sites	Haplotypes	Haplotype diversity	Nucleotide diversity	Watterson's theta	Tajima's D
All lineages	131 [1370]	132 (43)	38 (26)	12 (7)	0.33 (0.16)	0.02 (0.02)	9.02 (5.95)	-2.21 (0.37)
Mluc_alas	9 [1362]	130 (47)	21 (22)	1.5 (0.7)	0.63 (0.11)	11.95 (13.05)	10 (10.71)	12 (13.05)
Mluc_cari	5 [1307]	130 (47)	3 (4)	1.5 (0.6)	0.64 (0.09)	1.79 (2.53)	1 (2.08)	2 (2.53)
Mluc_luci	22 [1370]	128 (47)	28 (21)	4.0 (2.4)	0.38 (0.15)	4.91 (4.23)	7 (5.73)	5 (4.23)
Mluc_ern	1 [320]	135 (48)	na	1.0 na	na	na	na	na
Mluc_reli	2 [1179]	130 (48)	22 (21)	1.2 (0.4)	0.67 (0.02)	14.85 (14.39)	12 (11.79)	15 (14.39)
Mocc	3 [812]	132 (49)	9 (16)	1.2 (0.4)	0.64 (0.08)	0.77 (4.35)	5 (8.78)	6 (10.73)
Mevo	34 [1369]	128 (47)	10 (9)	4.8 (3.1)	0.45 (0.16)	1.21 (1.32)	2 (2.16)	1 (1.32)
Mkee	12 [1346]	129 (47)	5 (5.35)	2.2 (1.2)	0.37 (0.31)	1.68 (1.99)	2 (1.91)	2 (1.99)
Mthy	20 [1362]	128 (47)	7 (7)	3.4 (2.2)	0.59 (0.17)	1.36 (1.89)	2 (2.13)	1 (1.89)
Msod	11 [968]	127 (49)	22 (20)	1.6 (1.1)	0.58 (0.15)	4.45 (9.68)	10 (10.34)	12 (12.74)
Mvol	9 [1368]	128 (47)	10 (14)	2.0 (1.0)	0.43 (0.15)	1.86 (4.03)	3 (4.93)	3 (4.83)
Mrip	3 [1342]	129 (47)	16 (20)	1.5 (0.7)	0.64 (0.09)	4.19 (9.30)	8 (9.89)	9 (12.06)

Mluc\_alas = *Myotis lucifugus alascensis*; Mluc\_cari = *M. l. carissima*; Mluc\_luci = *M. l. lucifugus*; Mluc\_ern = *M. l. pernox*; Mluc\_reli = *M. l. relictus*; Mocc = *M. occultus*; Mevo = *M. evotis*; Mkee = *M. keenii*; Mthy = *M. thysanodes*; Msod = *M. sodalis*; Mvol = *Myotis volans*; Mrip = *M. riparius*.

majority rule consensus tree was subsequently inferred in PAUP.

Species trees were also estimated using ASTRAL v. 5.0.3 (Mirarab et al. 2014). ASTRAL takes as input rooted or unrooted gene trees and maximizes the number of quartet trees shared between the gene trees and the species tree. Input gene trees were estimated in RAxML v. 8.1.16 (Stamatakis 2014). Each locus was analyzed using 10 independent runs, a GTR+GAMMA model, a rapid-hill climbing algorithm, and 100 bootstrap replicates. For each locus, the best tree was chosen based on the log-likelihood (lnL). To infer branch support on the main species tree, two metrics were used: bootstrap and ASTRAL branch support values. Bootstrap values represent the proportion of times that a node was recovered after resampling the entire data set with replacement. To infer bootstrap values in the main species tree, a species tree was estimated on each of the 100 bootstrap replicates, and then those species trees were used to calculate support values of each node. Alternatively, ASTRAL branch support values represent the proportion of quartet trees (estimated from gene trees) that are also present in the species tree (Sayyari and Mirarab 2016). This method uses quartet support values to compute the local posterior probability (PP1) that a given branch is in the species tree, and also calculates the length of the branch in coalescent units. ASTRAL values were inferred from gene trees to score the existing species tree.

These analyses were performed using the entire data set, including 1370 loci and 12 lineages, considering each lineage of *M. lucifugus* a species and including the closest related species to this group, *M. evotis*, *M. keenii*, *M. occultus*, *Myotis sodalis*, *M. thysanodes*, and *Myotis volans*. *Myotis riparius* was defined as an outgroup (Table 2).

#### Comparing Phylogenetic Models With and Without Gene Flow

Given the unexpected findings showing paraphyly in the *M. lucifugus* subspecies, we explored whether a topology that includes gene flow might be a better option to describe the species relationships of this group. Previous model-based frameworks have been developed to estimate species relationship while accounting for gene flow (e.g., PHRAPL; Jackson et al. 2017a,b). However, because the number of potential models increases faster than factorial, exhaustive model testing with scenarios that include more than four lineages are not computationally feasible with approaches like PHRAPL or any other tool developed so far. Therefore, we implemented a novel ABC approach that was designed to evaluate support for phylogenetic models that include and exclude intraspecific gene flow (Fig. 2), where underlying topologies are "known" (can be prespecified), therefore the number of models tested

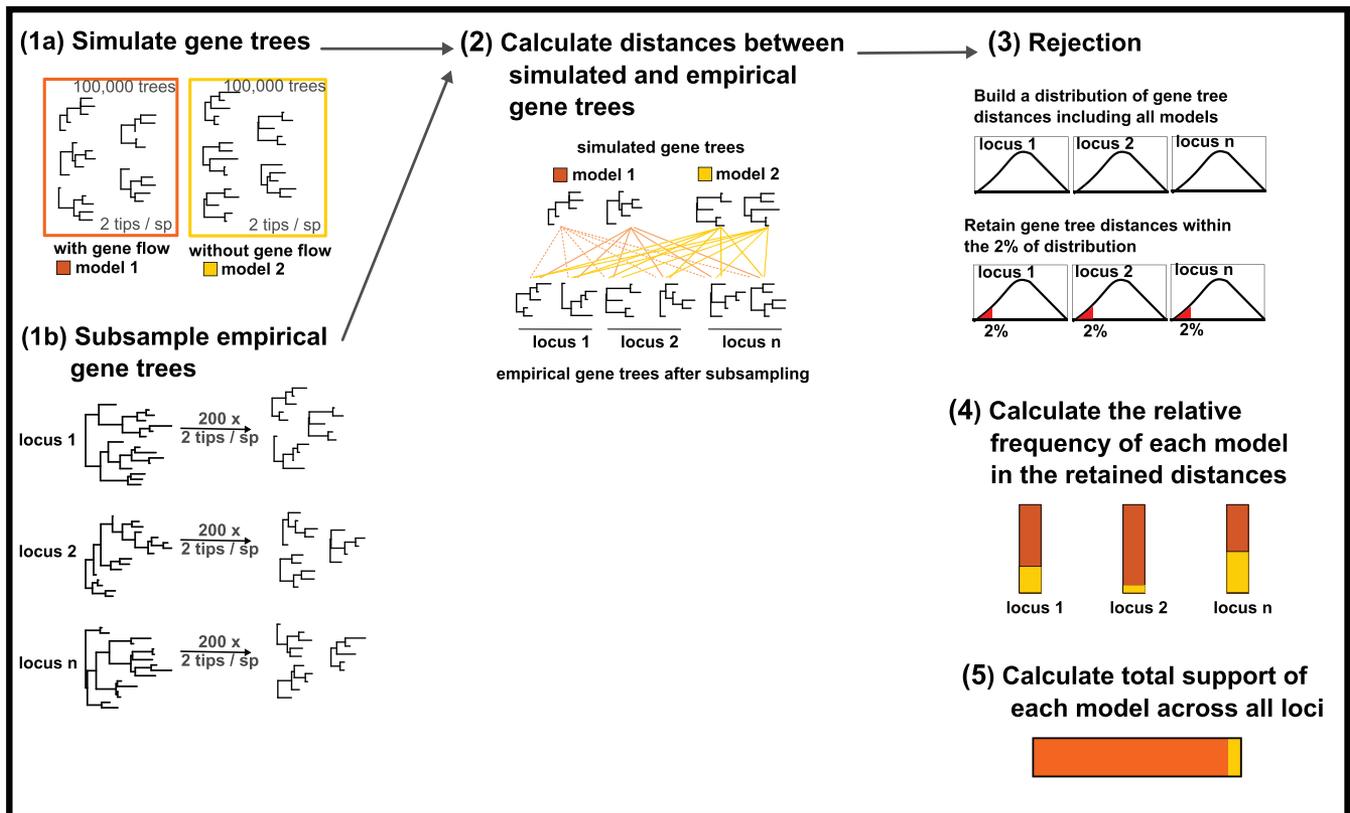


FIGURE 2. Workflow of the methodology followed to compare phylogenetic trees with and without gene flow among taxa and calculate the support of each model.

is considerably smaller. Because this framework relies on topologies, we used two gene tree metrics as summary statistics, the Robinson–Foulds (RF; Robinson and Foulds 1981) and Kuhner–Felsenstein (KF; Kuhner and Felsenstein 1994) distances. The main difference between these metrics is that RF measures differences in topology only while KF measures differences in both topology and branch lengths. RF measures the number of internal branches that exist in one tree but not in other. On the other hand, KF uses the branch tree information to calculate the sum of squares of the differences between each branch length in the two trees compared. For both metrics, a value of 0 means identical topologies and it increases as the match between trees worsens.

In summary, our framework (1a) simulates gene trees under specified models (phylogenies with and without gene flow) using *ms* (Hudson 2002). The underlying topology of each model can be estimated with traditional species tree methods that do not account for gene flow (e.g., ASTRAL, SVDQuartets), and prior distributions of parameters such as divergence times, population sizes, and migration rates are prespecified. As in any model-testing framework, the hypotheses tested and parameters used to simulate data will be based on previous knowledge of the system, and model misspecification could be a problem. (1b) Given that empirical gene trees may include multiple alleles per

species and, for computational feasibility, simulated trees only include two alleles per taxa, a subsampling with replacement approach was thus used to account for disparities across loci in the number of alleles and simplify computations of distances between simulated and observed trees. In this way, all compared trees have the same number of tips, and because each gene tree is subsampled at random hundreds of times, all alleles have the same probability of being included in the analysis. (2) Then, the RF and KF distances are calculated between simulated and empirical trees (after subsampling), and (3) a posterior distribution of gene tree distances is built by retaining only those trees that are among the closest 0.02 trees from the empirical data. The rejection threshold can be adjusted as desired; here, we retained 2% of the trees as we consider this is a conservative value. (4–5) Finally, the posterior probability of phylogenies that do and do not include gene flow can be approximated by calculating their proportional representation in the posterior distribution. Our approach scales easily to the genomic data collected here because posterior probabilities are calculated in parallel on a locus-by-locus basis, then the results are summarized across loci. A pipeline including a set of custom R functions was written to conduct Steps 1–5 ([https://github.com/ariadnamorales/ABC\\_gene\\_trees](https://github.com/ariadnamorales/ABC_gene_trees)).

*Simulated data.*—The performance of this approach to accurately differentiate between models with and without gene flow was assessed via simulation testing, emulating our *Myotis* data set with 11 lineages and models of interspecific gene flow. First, a species tree was simulated and used as underlying topology to design three models as follows: not allowing gene flow between lineages, allowing gene flow among sister species, and allowing gene flow among nonsister species. To simulate 100,000 trees under each model (Step 1a), prior distributions for divergence time ( $\tau$ ) were drawn from a normal distribution ranging from  $\tau=0.25$  to  $\tau=2.5$ . For models including gene flow, matrices of migration between sister and nonsister species were generated at random. Prior distributions for migration ( $m$ ) and theta ( $\theta$ ) were drawn from uniform distributions, where the upper and lower bounds ranged from  $m_{lb}=0.1$  to  $m_{ub}=5$  and  $\theta_{lb}=0.1$  to  $\theta_{ub}=10$ . In addition, 100 trees were simulated under each model, used as pseudo-observed data and analyzed following our framework. Each tree was simulated with four alleles per taxa and then, two tips were subsampled 200 times (as implemented in Step 1b). Finally, the accuracy of our approach was measured as the proportion of times that gene trees supported the true model.

*Empirical data.*—Data collected from 1370 UCE in *Myotis* bats were analyzed as described above. To simulate trees under each model (Step 1a), prior distributions for divergence time ( $\tau$ ) were drawn from a normal distribution, where the mean and variance were derived from previous estimates of divergence times for *Myotis* bats (Ruedi et al. 2013; Morales et al. 2017) ranging from  $\tau=0.25$  to  $\tau=2.5$ , enclosing 10 intervals (because we are assuming 10 divergence events across 11 lineages). Prior distributions for migration ( $m$ ) and theta ( $\theta$ ) were drawn from uniform distributions, where the upper and lower bounds ranged from  $m_{lb}=0.1$  to  $m_{ub}=5$  and  $\theta_{lb}=0.1$  to  $\theta_{ub}=10$ , according to previous estimates (Ruedi et al. 2013; Morales et al. 2017). Because gene flow among *Myotis* species has been reported when they occur in sympatry (e.g., Morales et al. 2017), two models of interspecific gene flow were designed. The first model only allows migration between *M. lucifugus* lineages with continuous distributions (Supplementary Table S3 available on Dryad). The second model allows migration between all *Myotis* species from North America included in this study, which show sympatric distribution in at least some part of their range (Supplementary Table S3 available on Dryad). While this design relies on sympatry in the present and thus may not correspond directly to historical sympatry, our knowledge of historical species range is limited by a lack of recent *Myotis* fossils from North America. Finally, two alleles per species were subsampled at random in each gene tree 200 times (Step 1b). In this way, all the subsampled empirical trees compared against the simulated trees (100,000 trees for each model) yielded 20,000,000 comparisons per locus per model.

The following steps (2–5) in this analysis were performed as described above.

### Phylogenetic Networks

A maximum pseudolikelihood (pL) approach as implemented in SNaQ was used to estimate phylogenetic networks (Solís-Lemus and Ané 2016). Pseudolikelihood frameworks are useful to analyze genomic data sets because computationally they are easier to estimate and they can approximate faster the likelihood function of a set of observed data. The SNaQ framework uses gene trees and a species tree as a starting point to perform a search that accounts for ILS under the coalescent model, and horizontal transfer of genes under reticulations events in a network. Gene trees are summarized as quartet concordance factors, including all possible combinations of four taxa. However, given that our data set has 262 tips (i.e., two alleles per individual), there are nearly 1 billion possible quartets, thus preventing an exhaustive analysis, we used a subsampling approach, where one allele per lineage was sampled at random with replacement such that 100 subsampled trees per locus were analyzed. This analysis was performed using gene trees from 1370 loci. Therefore, after subsampling, 137,000 trees with one tip per lineage were used as input to calculate a table of quartet concordance factors. Gene trees were estimated in RAxML, and the starting species tree was estimated in ASTRAL as described above, where *M. riparius* was set as outgroup. Gene trees were subsampled and relabeled with a custom script. To determine whether a tree with ILS or a network fits the observed data, 11 estimations, each with 10 replicates, were performed allowing a maximum number of reticulation nodes or genetic exchange events ( $h$ ) ranging from 0 to 10; where  $h=0$  will yield a tree,  $h=1$  will yield a network with one reticulation, and so forth. Then, two methods were used to estimate whether a tree or a network with a given  $h$  provides the best fit to the data. First, a goodness-of-fit test, based on quartet concordance factors, was used. This test compares the score of each network based on the negative log-pL, where the network with the lowest value has the best fit. A model selection approach based on slope heuristics methods was also performed as implemented in capushe (Baudry et al. 2012). Slope heuristics is a data-driven method used to calibrate penalized criteria for model selection, where penalties are based on a multiplicative factor. Finally, after identifying a network with the optimum  $h$  value, that value was used to perform 100 bootstrap replicates to calculate support values.

## RESULTS

### Summary Statistics

Tissue samples were acquired from 131 individuals including *Myotis lucifugus* recognized subspecies

(*alascensis*, *carissima*, *lucifugus*, *pernox*, and *relictus*; Simmons 2005) and closely related members of the Nearctic subclade (Fig. 1; *M. evotis*, *M. keenii*, *M. occultus*, *M. sodalis*, *M. thysanodes*, and *M. volans*; Ruedi et al. 2013). Data were also collected from *M. riparius* and used as an outgroup. Using the custom sequence capture probe set developed by Morales et al. (2017), sequence data were collected from 1.5 Illumina NextSeq lanes which produced more than 800 million paired-end reads. Contigs were assembled and only those with coverage greater than  $10 \times$  that aligned to UCEs were retained. After filtering, 1370 UCE loci were analyzed. These loci averaged 132 bp (range 83–387), 38 segregating sites (range 1–129), and 12 haplotypes (range 2–42; Table 2). The average value of haplotype diversity was 0.328 (range 0.025–0.690), whereas the average value of nucleotide diversity was 0.02 (range 0.0002–0.1143). *Myotis lucifugus relictus* contained the highest haplotype and nucleotide diversity values whereas *M. l. lucifugus* showed the lowest values. Mean values of Tajima's *D* were not significantly different between nominal entities.

#### Species Delimitation and Species Tree Reconstruction of *M. lucifugus* Lineages

Genetic evidence has suggested that several lineages within *M. lucifugus* are evolving independently in the presence of gene flow (Carstens and Dewey 2010). To evaluate whether the five subspecies described should be considered independent lineages, we conducted species delimitation and estimated species relationships simultaneously using BPP (Yang and Rannala 2010). Results from this analysis indicate that there are five species ( $pp = 0.872$ ), and the species tree estimated shows that *M. l. alascensis* and *M. l. carissima* are sister species, *M. l. pernox* is sister to the common ancestor of this clade, whereas *M. l. lucifugus* and *M. l. relictus* are the most divergent lineages. However, the species tree estimated using \*BEAST (Heled and Drummond 2010; Drummond et al. 2012), suggests that *M. l. relictus* and *M. l. carissima* are sister species, and *M. l. alascensis* is sister to the common ancestor of the former, whereas *M. l. pernox* and *M. l. lucifugus* are the most divergent lineages, respectively. The \*BEAST results were questionable given ESS <200 in most cases, even when the number of generations was increased by millions several times.

#### Posterior Predictive Evaluation of Species Tree Model

While the \*BEAST results were questionable (ESS <200 in most cases), PPS using P2C2M indicate that the species tree model utilized in \*BEAST does not have a good fit to the data, as fewer than 20% of the gene trees show a good fit to the MSCM (loci with good fit according to  $coal = 10\%$ ,  $lcwt = 0\%$ ,  $ndc = 21\%$ ). Further, these results are consistent with previous analyses conducted

in a closely related clade, the western long-eared bats, where we recently demonstrated that estimates of gene flow should be incorporated into species tree inference (Morales et al. 2017).

#### Estimating the Phylogenetic Placement of *M. lucifugus* Lineages

To infer the phylogenetic placement of *M. lucifugus* lineages with respect to other *Myotis* from the Nearctic clade, a species tree was estimated using two coalescent-based approaches. These results demonstrate that *M. lucifugus* subspecific lineages, suspected to be independent by Carstens and Dewey (2010) and supported by our BPP analysis, are paraphyletic. However, the phylogenetic placement these lineages in a species tree changes depending on the method used (Fig. 3). In one hand, the phylogeny estimated by SVDQuartets recovered two main clades. In the first clade, *M. l. carissima* is sister to *M. occultus* and *M. thysanodes*, *M. evotis* and *M. keenii* are sister species, and *M. l. pernox* and *M. l. alascensis* are the most divergent species within the group. In the second clade, *M. l. lucifugus* and *M. sodalis* are sister species, *M. l. relictus* is sister to the former, and *M. volans* is the most divergent species within this group. Bootstrap support values at each node were higher than 98 (Fig. 3a). On the other hand, in the phylogeny estimated by ASTRAL, *M. l. pernox* is sister to the western long-eared group (*M. evotis*, *M. thysanodes*, and *M. keenii*), *M. l. carissima* is sister to this clade with *M. occultus* as the most divergent species of this clade. *Myotis l. alascensis* and *M. l. relictus* are members of a different clade, where *M. l. relictus* is sister to *M. volans*, and *M. l. alascensis* is sister to them. Finally, *M. l. lucifugus* and *M. sodalis* are the most divergent lineages within this group. Bootstrap support values at each node were higher than 98 whereas ASTRAL support values ranged from 0.27 to 0.49, with effective numbers of gene trees per branch ranging from 311 to 1361 (Fig. 3b). For five lineages that were only described as subspecies on the basis of subtle morphological differences and have been studied by several decades as the same entity, these results are unexpected.

#### Comparing Phylogenetic Models With and Without Gene Flow

Given the paraphyly inherent to *M. lucifugus* subspecies, we explored whether a species tree that includes gene flow might be a better option to describe the evolutionary history of this group. We implemented a framework based on ABC using gene tree distances as summary statistics to compare phylogenetic models in the presence and absence of gene flow. This approach allowed us to test models of isolation and gene flow between species in sympatric and allopatric distributions. Simulation testing of the proposed ABC method demonstrates that it is likely to identify

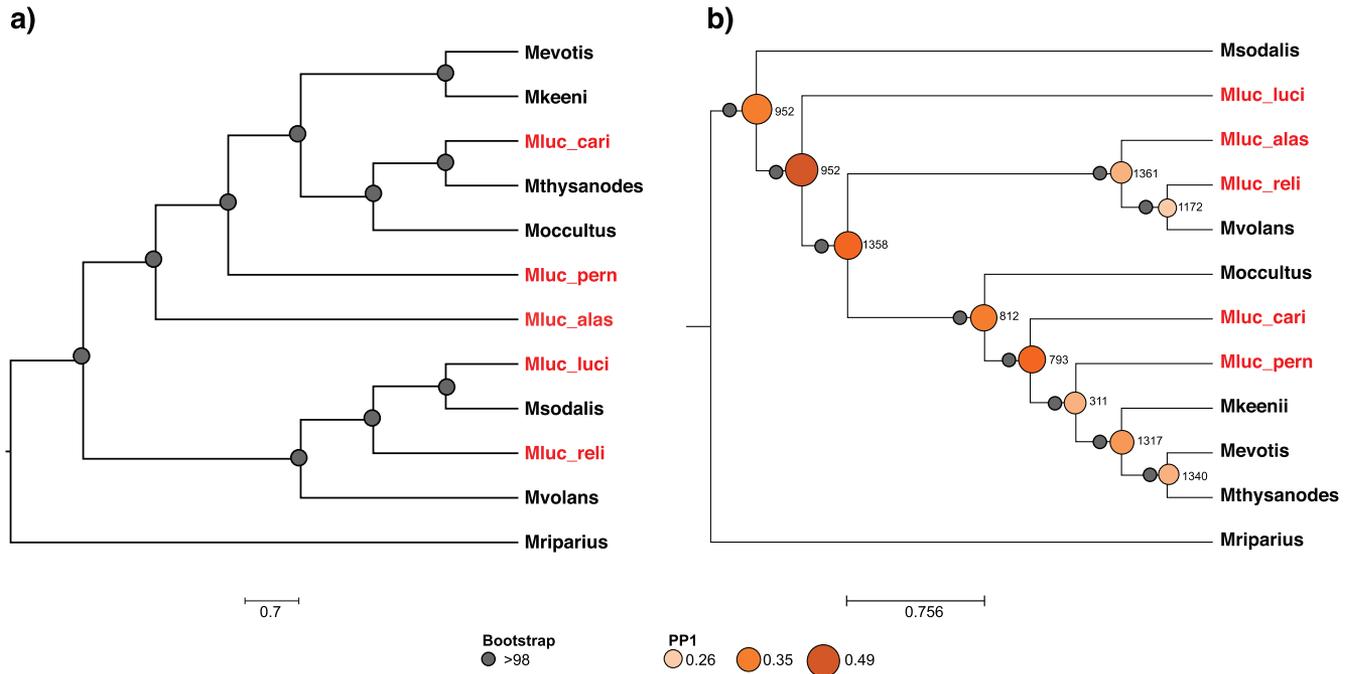


FIGURE 3. Species trees estimated for *Myotis* species in the Nearctic subclade using a) SVDQuartets and b) ASTRAL. Gene trees from 1370 UCE loci were analyzed and *M. riparius* was set as outgroup. Lineages described as *M. lucifugus* subspecies are labeled to be the lighter color, and other North American *Myotis* species are labeled in the darkest. Left circles between nodes represent support values higher than 98 calculated based on 100 multilocus bootstrap replicates for SVDQuartets and ASTRAL analyses. Right circles represent the local posterior probability (PP1) estimated in ASTRAL, the size and color increase from low (light and small circles) to high support (dark and big circles). In the ASTRAL tree, numbers between nodes show the effective number of gene trees per branch.

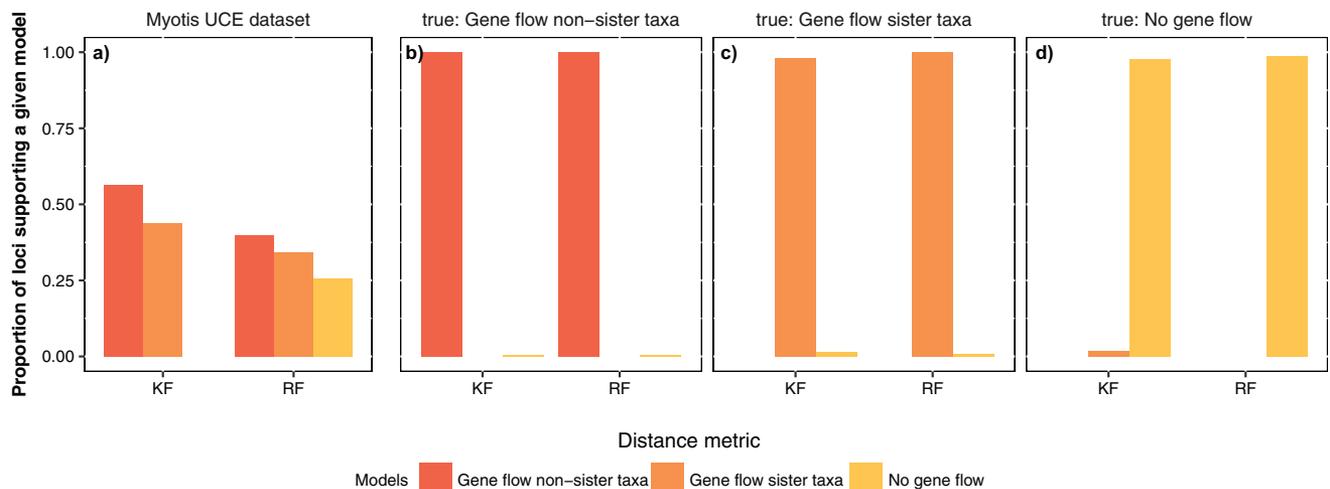


FIGURE 4. Proportion of loci supporting phylogenetic models with and without gene flow. Results from data collected in North American *Myotis* species (a) and results for simulated gene trees under a model of gene flow among nonsister species (b), gene flow between sister species (c), and no gene flow (d). Kuhner-Felsenstein (KF) and Robinson-Foulds (RF) distances.

phylogenies biased by gene flow, as results indicate that the true model is identified in >98% of replicates using either KF or RF distances (Fig. 4).

When data from *Myotis* bats were analyzed with a fixed topology and migration matrices designed based on sympatric distributions, models with gene flow achieved the highest support regardless of the distance metric used (Fig. 4). However, the posterior probability

of the model given the data is dependent on the metric used. The model with gene flow between all the sampled North American *Myotis* showed the highest posterior probability (KF = 0.57; RF = 0.4; Fig. 4 and Supplementary Table S2 available on Dryad), while the model that included only gene flow between *M. lucifugus* lineages with contiguous distributions was lower (KF = 0.43; RF = 0.35; Fig. 4 and Supplementary Table S1

available on Dryad) and the model without any gene flow was lower still (KF  $\sim$ 0.0; RF = 0.25). These results support the suggestion that gene flow in North American *Myotis* bats occurs at interspecific levels whenever species come into sympatry, regardless of whether they are sister species or not.

### Phylogenetic Networks

Given our results confirming gene flow between nonsister taxa in North American *Myotis*, we conducted network analyses to infer species relationships while accounting for genetic exchange (or reticulations). Networks estimated by SNaQ also demonstrate that *M. lucifugus* subspecies are not part of a monophyletic clade (Fig. 5). The species tree analysis ( $h=0$ ) showed the lowest support while a network with  $h=4$  showed the highest support (pL = 8821.4419) and best fitting according to the goodness-of-fit test and the slope heuristics. However, pL values when  $h$  equal to 3, 6, and 7 were quantitatively similar (pL = 8821.4422, 8821.4426, and 8821.4557, respectively; Fig. 5a). The underlying topologies in these networks were identical, each showing two reticulation events with the same  $\gamma$  values for minor hybridization events (Fig. 5b). The  $\gamma$  value is the probability of a certain lineage having descended from the hybridization of a sister edge while  $1-\gamma$  is the probability of having descended from the original tree branch. These phylogenetic networks showed two major splits, the first group includes *M. volans*, *M. thysanodes*, and *M. sodalis*, and the second group includes *M. lucifugus* lineages, *M. evotis*, *M. keenii*, and *M. occultus*. The most recent hybrid event recovered suggest genetic exchange from *M. sodalis* to *M. thysanodes* with a  $\gamma=0.452$ , while the most ancient hybridization event shows a deep reticulation in the clade with a  $\gamma=0.469$ .

### DISCUSSION

Gene flow among mammals was once thought to be rare, however, recent findings have suggested that it may be a common process that can complicate species relationship inferences (e.g., Melo-Ferreira et al. 2014; Sullivan et al. 2014; Kumar et al. 2017). We document an example of interspecific gene flow among nonsister species that occur in sympatry in a genus of bats, where genetic exchange may be hindering phylogenetic signal. Species delimitation results presented here confirmed previous suspicions (i.e., Carstens and Dewey 2010) of multiple species within the currently recognized *M. lucifugus*, but may be influenced by the presence of gene flow in this system. As evident in the results from the P2C2M analysis, the MSCM implemented in BPP is not a good fit to our data, likely because alleles are shared among nonsister lineages due to gene flow. While the phylogenetic analyses demonstrate paraphyletic patterns among these lineages, in systems

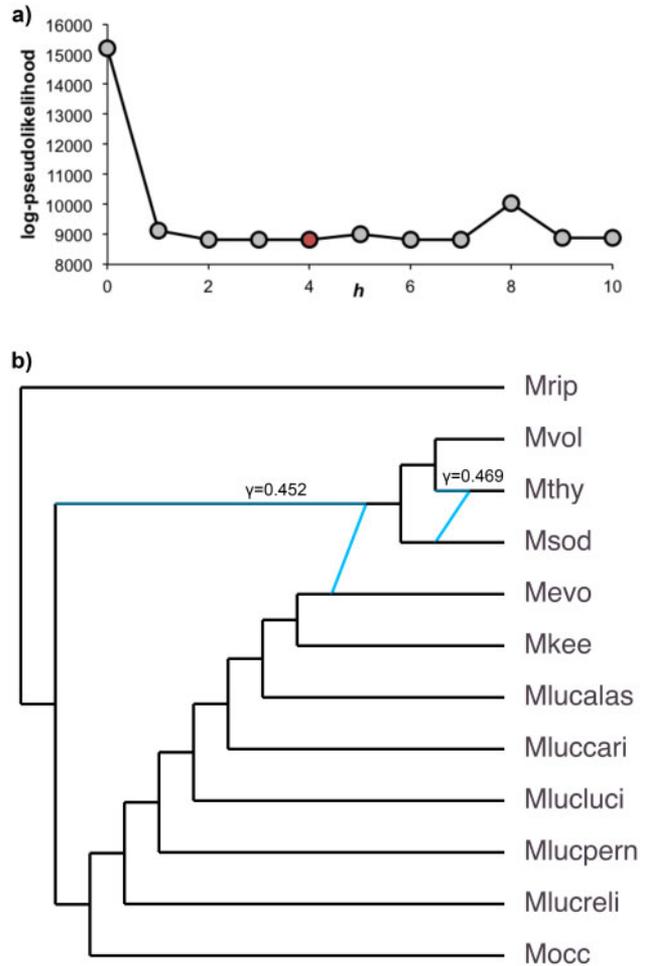


FIGURE 5. a) Plot showing log-pseudolikelihood values obtained from the SNaQ analysis when given values were allowed as the maximum number of reticulation nodes ( $h$ ), where  $h=4$  showed the lowest (best) value with the best fit. b) Reticulate evolution estimated for *Myotis* species in North America using SNaQ when  $h=2, 3, 4, 6,$  and  $7$ . *Myotis riparius* was set as outgroup. Gene trees from 1370 UCE loci were subsampled 100 times, for a total of 137,000 gene trees that were used as input. Dark edges represent the major tree and major hybrid edges. Light solid arrows represent minor hybrid edges ( $\gamma_{vol-thy-sod}=0.452$ ,  $\gamma_{thy}=0.469$ ).

like *Myotis*, phylogenetic inferences are challenging. Our strategy involved using multiple methods with differing assumptions, with the hope that we could identify congruence across phylogeny estimates. However, while species tree methods like ASTRAL and SVDQuartets show the *M. lucifugus* lineages as paraphyletic groups, the recovered topologies are conspicuously different. We question the tree inferred using SVDQuartets because the western long-eared species (*M. evotis*, *M. thysanodes*, and *M. keenii*) do not form a clade as they do in the ASTRAL analysis, which matches previous phylogeny estimates using everything from mitochondrial (Dewey 2006) to UCE data (Morales et al. 2017; Platt et al. 2018). However, given that neither method incorporates gene flow, the pertinent question for interpretation is the extent to which gene flow leads to inaccuracy in the estimation of either the topology or

the divergence times. Simulation studies have suggested that species relationships and divergence time inferences are possible under certain conditions that include gene flow (Eckert and Carstens 2008; Leache et al. 2014). The challenge for empiricists is to understand how estimates of phylogeny may have been biased in their system. For example, could gene flow between nonsister lineages in our system lead to the apparent paraphyly and unclear lineage relationships of the *M. lucifugus* subspecies? Our results indicate that geographic proximity appears to be predictive of clade relationships in the species tree. For instance, *M. l. alascensis* and *M. l. relictus* exhibit contiguous distributions that are partly sympatric, these are also in sympatry with *M. volans*, and these three lineages form a clade. Similarly, *M. l. pernox* and *M. l. carissima* form a clade with the western long-eared species, to which they are broadly sympatric. Our skepticism about the topology estimated in the ASTRAL analysis is largely based on these patterns.

It is clear that methods development is needed if we are to fully take advantage of the recent advances in sequencing technology that have enabled researchers to collect genomic data sets in nonmodel systems such as *Myotis* bats. Here, we applied two analyses designed to evaluate if a species tree is a reasonable model for North American *Myotis*. First, we developed an ABC approach based on gene tree metrics to compare phylogenetic hypotheses that can incorporate gene flow. Results from these estimations support the suggestion that gene flow in North American *Myotis* occurs at interspecific levels whenever species come into sympatry, regardless of whether they are sister species or not. However, it remains to be explored if genetic interchange occurred in the past during early stages of divergence (i.e., speciation with gene flow), or if it occurred after lineages diverged in isolation (i.e., secondary contact), or if it has occurred constantly (i.e., genomic regions are subject to strong divergent selection to maintain species barriers despite gene flow). Based on previous investigations including the western long-eared species showing that gene flow occurred only during early stages of divergence (Morales et al. 2017), we suspect that this could be the case for the North American clade. Secondarily, results from phylogenetic network estimations support that at least four events of genetic exchange have occurred and at least one these occurred deeper in the topology. Even though the species relationships are different from those inferred by species tree methods, these findings suggest that genetic exchange occurred at early stages of divergence among some ancestral taxa and current lineages might be at different stages of the speciation spectrum. Simulation studies using genomic data and large trees have suggested that when genetic exchange has occurred between ancestral nodes, hybrid detection methods are good at identifying taxa in which gene flow has occurred, but these methods are not as effective to unambiguously picking out parental or sister species (Kubatko and Chifman 2015). Given that ABC results suggest that reticulations have occurred more than four times and some of these represent ancestral events,

species relationships might be challenging to infer with current tools, especially among lineages that recently diverged.

Our results emphasize the need for new phylogenetic tools for species tree inference that can accommodate genetic exchange among lineages. While most species tree methods assume that incongruence between gene trees and species trees results from ILS, our investigation (among others) demonstrates that such assumptions might not suffice. Phylogenetic networks provide one alternative and current implementations represent a great advance to infer whether reticulation events occur between ancestral or more derived nodes, but have limitations, including the amount of data, number of individuals per taxon and number of taxa. We have addressed some of these issues by implementing a subsampling approach, where one or a few alleles per species were sampled hundreds of times for each locus and analyzed independently. Even when we drastically reduced our data set and used only a single individual per species, computational time was substantial, and it became a constraint as we attempted to analyze multiple individuals per lineage. Tools like ASTRAL have addressed this problem allowing multiple individuals per lineage, missing taxa from gene trees and still being able to include them in the species tree estimation at the cost of ignoring genetic exchange. Alternatively, methods such as PHRAPL (Jackson et al. 2017b) can handle big data sets, allow missing data and account for parameters such as divergence and gene flow but are limited by the number of lineages that can be analyzed due to model space conflicts.

#### *Taxonomic and Conservation Implications on North American Myotis*

*Myotis* bats are one of the most taxonomically complicated groups within the order Chiroptera. The apparent morphological resemblance between species that occur in sympatry and potential interspecific gene flow have led to contradicting species relationships in studies based on generic morphological features (i.e., length of the forearms, length of the skull) and a few loci. It is not surprising that novel phylogenetic analyses with genomic data are contradicting previously described species relationships and species limits. Here, we have shown that the *M. lucifugus* subspecies recognized in Simmons (2005) are lineages evolving independently that seem to be paraphyletic, we thus suggest they should be considered independent species, under the phylogenetic species concept, as follows: *M. alascensis*, *M. carissima*, *M. lucifugus*, *M. relictus*, and *M. pernox*. For five lineages that were only described as subspecies on the basis of morphological similarities and have been studied by several decades as the same entity, this result is unexpected. Even though these results might represent a big change in the current taxonomy of North American *Myotis*, similar changes have been derived from studies using molecular data. For example,

when DNA analyses of two mitochondrial loci were used for the first time to infer a phylogeny of this genus, species thought to be closely related, given their phenotypic resemblance, were found to be paraphyletic (Ruedi and Mayer 2001). Moreover, genomic studies using UCEs for phylogenetic inferences of *Myotis* in the Americas have also found paraphyletic patterns in *M. lucifugus* lineages and estimated species relations that contradict previous findings based on mitochondrial loci (Platt et al. 2018). However, we do not argue that current topologies, estimated here or in other studies, represent conclusive evidence given the lack of methods to incorporate gene flow into phylogenomic estimations. The morphological resemblance among nonsister species leads to suspect that other evolutionary forces might have a great effect on the phenotypic similarities within this group. For example, strong environment-driven selection might be occurring in some genomic regions, favoring similar phenotypes among sympatric species that exchanged genes at some point in their evolutionary history. To address current confusions due to apparent morphological similarity among these lineages, exhaustive taxonomic revisions based on geometric morphometric and genomic analyses including coding regions coupled with novel analytical approaches are required to identify features that could help to reclassify specimens within this genus. Finally, our findings also shed light upon conservation priorities. For instance, *M. relictus* and *M. pernox* have small distributions and biological surveys are urgent to evaluate the status of their populations, which could be vulnerable or under threat. *Myotis pernox* represents a priority due to their range is less than 1000 km from localities in Washington where white-nose syndrome occurrences were reported in March 2016 (<https://www.whitenosesyndrome.org>).

In summary, we document an example of interspecific gene flow among nonsister species in a genus of mammals, where genetic exchange may be hindering phylogenetic signal. Our findings demonstrate that there are at least five independent lineages within the *M. lucifugus* group, which are paraphyletic and should be considered different species. In addition, results of simulation analyses and model testing demonstrated that gene flow should be considered while species relationships are inferred because it is a common process that has occurred throughout the evolutionary history of *Myotis* bats in North America. We implemented an approach for explicit modeling of genetic exchange among sympatric and allopatric lineages that can be extended to other taxonomic groups to explore if gene flow should be accounted for during phylogenetic estimations.

#### SUPPLEMENTARY MATERIAL

Data and Supplementary Material available from the Dryad Digital Repository: <http://dx.doi.org/10.5061/dryad.9d6b2>.

#### FUNDING

The National Science Foundation funded this research (DEB-1257784) and additional funding was provided through a Doctoral Dissertation Improvement Grant (DEB-1701810) awarded to A.E.M. and B.C.C. The Ohio Supercomputer Center allocated resources to support part of this study (PAS1184-1). Support for A.E.M. was provided in part by a graduate fellowship at The Ohio State University funded by Consejo Nacional de Ciencia y Tecnología (Reg. 217900 CVU 324588).

#### ACKNOWLEDGMENTS

We thank museum curators who lent us tissue samples of specimens under their care: AMNH, American Museum of Natural History, New York, USA; ASNHC, Angelo State Natural History Collection, Texas, USA; CIIDIR-Durango, Colección de Mamíferos, Centro Interdisciplinario de Investigación para el Desarrollo Integral Regional, Durango, México; UABC, Colección de Mamíferos, Universidad Autónoma de Baja California, México; UMMZ, University of Michigan Museum Zoology; Tanya Dewey donated DNA extractions. We thank Brian O'Meara, Nathan Jackson, and members of the Carstens lab for conversations related to this work.

#### REFERENCES

- Barbour R., Davis W. 1969. Bats of America. Lexington (KY): The University Press of Kentucky.
- Baudry J., Maugis C., Michel B. 2012. Slope heuristics: overview and implementation. *Stat. Comput.* 22:455–470
- Bolger A.M., Lohse M., Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120.
- Carstens B.C., Dewey T.A. 2010. Species delimitation using a combined coalescent and information-theoretic approach: an example from North American *Myotis* bats. *Syst. Biol.* 59:400–414.
- Chifman J., Kubatko L. 2014. Quartet inference from SNP data under the coalescent. *Bioinformatics.* 30:3317–3324.
- Chifman J., Kubatko L. 2015. Identifiability of the unrooted species tree topology under the coalescent model with time-reversible substitution processes, site-specific rate variation, and invariable sites. *J. Theor. Biol.* 374:35–47.
- Czenze Z.J., Willis C.K.R. 2015. Warming up and shipping out: arousal and emergence timing in hibernating little brown bats (*Myotis lucifugus*). *J. Comp. Physiol. B* 185:575–586.
- Dewey T.A. 2006. Systematics and phylogeography of North American *Myotis* (Chiroptera: Vespertilionidae) [PhD dissertation]. University of Michigan. Ann Arbor, MI.
- Drummond A.J., Suchard M.A., Xie D., Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* 29:1969–1973.
- Eckert A.J., Carstens B.C. 2008. Does gene flow destroy phylogenetic signal? The performance of three methods for estimating species phylogenies in the presence of gene flow. *Mol. Phylogenet. Evol.* 49:832–842.
- Faircloth B.C. 2013. Illumiprocessor: a trimmomatic wrapper for parallel adapter and quality trimming. Available from: <http://dx.doi.org/10.6079/J9ILL>.
- Flicke P., Amode M.R., Barrell D., Beal K., Billis K., Brent S., Carvalho-Silva D., Clapham P., Coates G., Fitzgerald S., Gil L., Girón C.G., Gordon L., Hourlier T., Hunt S., Johnson N., Juettemann T., Kähäri A.K., Keenan S., Kulesha E., Martin F.J., Maurel T., McLaren W.M.,

- Murphy D.N., Nag R., Overduin B., Pignatelli M., Pritchard B., Pritchard E., Riat H.S., Ruffier M., Sheppard D., Taylor K., Thormann A., Trevanion S.J., Vullo A., Wilder S.P., Wilson M., Zadissa A., Aken B.L., Birney E., Cunningham F., Harrow J., Herrero J., Hubbard T.J., Kinsella R., Muffato M., Parker A., Spudich G., Yates A., Zerbino D.R., Searle S.M. 2014. Ensembl 2014. *Nucleic Acids Res.* 42:D749–D755.
- Ghazali M., Moratelli R., Dzeverin I. 2016. Ecomorph evolution in *Myotis* (Vespertilionidae, Chiroptera). *J. Mamm. Evol.* 1–10.
- Gnrke A., Melnikov A., Maguire J., Rogov P., LeProust E.M., Brockman W., Fennell T., Giannoukos G., Fisher S., Russ C. et al. 2009. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat. Biotechnol.* 27:182–189.
- Gruenstaedl M., Reid N.M., Wheeler G.L., Carstens B.C. 2016. Posterior predictive checks of coalescent models: P2C2M, an R package. *Mol. Ecol. Resour.* 16:193–205.
- Guindon S., Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* 52:696–704.
- Harris R.S. 2007. Improved pairwise alignment of genomic DNA. Ph.D. Thesis, The Pennsylvania State University. State College (PA).
- Heled J., Drummond A.J. 2010. Bayesian inference of species trees from multilocus data. *Mol. Biol. Evol.* 27:570–580.
- Hey J. 2010. The divergence of chimpanzee species and subspecies as revealed in multipopulation isolation-with-migration analyses. *Mol. Biol. Evol.* 27:921–933.
- Hollister N. 1911. Four new mammals from the Canadian Rockies. *Smithson. Misc. Collect.* 56:1–4.
- Hudson R.R. 2002. Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics* 18:337–338.
- Jackson N., Carstens B.C., Morales A.E., O'Meara B.C. 2017a. Species delimitation with gene flow. *Syst. Biol.* 66:799–812.
- Jackson N., Morales A.E., Carstens B.C., O'Meara B.C. 2017b. PHRAPL: Phylogeographic inference using approximate likelihoods. *Syst. Biol.* 66:1045–1053.
- Joly S., McLenachan P.A., Lockhart P.J. 2009. A statistical approach for distinguishing hybridization and incomplete lineage sorting. *Am. Nat.* 174:E54–E70.
- Katoh K., Standley D.M. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30:772–780.
- Kubatko L., Chifman J. 2015. An invariants-based method for efficient identification of hybrid speciation from large-scale genomic data. [bioRxiv doi.org/10.1101/034348](https://doi.org/10.1101/034348)
- Kuhner M.K., Felsenstein J. 1994. A simulation comparison of phylogeny algorithms under equal and unequal evolutionary rates. *Mol. Biol. Evol.* 11:459–468.
- Kumar V., Lammers F., Bidon T., Pfenninger M., Kolter L., Nilsson M.A., Janke A. 2017. The evolutionary history of bears is characterized by gene flow across species. *Sci. Rep.* 7:46487.
- Kwiecinski G.G., Wimsatt W.A., Krook L. 1987. Morphology of thyroid C-cells and parathyroid glands in summer-active little brown bats, *Myotis lucifugus lucifugus*, with particular reference to pregnancy and lactation. *Am. J. Anat.* 178:421–427.
- Larsen R.J., Knapp M.C., Genoways H.H., Khan F.A.A., Larsen P.A., Wilson D.E., Baker R.J. 2012. Genetic diversity of neotropical *Myotis* (Chiroptera: Vespertilionidae) with an emphasis on South American species. *PLoS One* 7:1–9.
- Leache A.D., Harris R.B., Rannala B., Yang Z. 2014. The influence of gene flow on species tree estimation: a simulation study. *Syst. Biol.* 63:17–30.
- Li H., Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760.
- Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25:2078–2079.
- McCormack J.E., Faircloth B.C., Crawford N.G., Gowaty P.A., Brumfield R.T., Glenn T.C. 2012. Ultraconserved elements are novel phylogenomic markers that resolve placental mammal phylogeny when combined with species-tree analysis. *Genome Res.* 22:746–754.
- Meiklejohn K.A., Faircloth B.C., Glenn T.C., Kimball R.T., Braun E.L. (2016) Analysis of a rapid evolutionary radiation using ultraconserved elements: evidence for a bias in some multispecies coalescent methods. *Syst. Biol.* 65:612–627.
- Melo-Ferreira J., Farello L., Freitas H., Suchentrunk F., Boursot P., Alves P.C. 2014. Home-loving boreal hare mitochondria survived several invasions in Iberia: the relative roles of recurrent hybridisation and allele surfing. *Heredity* (Edinb). 112:265–273.
- Meng C., Kubatko L.S. 2009. Detecting hybrid speciation in the presence of incomplete lineage sorting using gene tree incongruence: a model. *Theor. Popul. Biol.* 75:35–45.
- Mirarab S., Reaz R., Bayzid M.S., Zimmermann T., Swenson M.S., Warnow T. 2014. ASTRAL: genome-scale coalescent-based species tree estimation. *Bioinformatics* 30:541–548.
- Morales A.E., Jackson N.D., Dewey T.A., O'Meara B.C., Carstens B.C. 2017. Speciation with gene flow in North American *Myotis* bats. *Syst. Biol.* 66:440–452.
- Norquay K.J.O., Willis C.K.R. 2014. Hibernation phenology of *Myotis lucifugus*. *J. Zool.* 294:85–92.
- Nylander J.A.A. 2004. MrAIC.pl. <https://github.com/nylander/MrAIC>.
- Ohio Supercomputer Center. 2016. Owens Supercomputer Center. Columbus, OH: Ohio Supercomputer Center. <http://osc.edu/ark:19495/hpc6h5b1>
- Pfeifer B., Wittelsbürger U., Ramos-Onsins S.E., Lercher M.J. 2014. PopGenome: An efficient Swiss army knife for population genomic analyses in R. *Mol. Biol. Evol.* 31:1929–1936.
- Platt R.N., Faircloth B.C., Sullivan K.A.M., Kieran T., Glenn T.C., Vandeweghe M.W., Lee T.E., Baker R.J., Stevens R.D., Ray D.A. 2018. Conflicting evolutionary histories of the mitochondrial and nuclear genomes in New World *Myotis*. *Syst. Biol.* 67:236–249.
- Rambaut A., Suchard M.A., Xie D., Drummond A.J. 2014. Tracer v1.6. <http://beast.bio.ed.ac.uk/Tracer>.
- Robinson D.F., Foulds L.R. 1981. Comparison of phylogenetic trees. *Math. Biosci.* 53:131–147.
- Ruedi M., Mayer F. 2001. Molecular systematics of bats of the genus *Myotis* (Vespertilionidae) suggests deterministic ecomorphological convergences. *Mol. Phylogenet. Evol.* 21:436–448.
- Ruedi M., Stadelmann B., Gager Y., Douzery E.J.P., Francis C.M., 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). *Mol. Phylogenet. Evol.* 69:437–449.
- Sayyari E., Mirarab S. 2016. Fast coalescent-based computation of local branch support from quartet frequencies. *Mol. Biol. Evol.* 33:1654–1668.
- Schutt W.A., Cobb M.A., Petrie J.L., Hermanson J.W. 1994. Ontogeny of the pectoralis muscle in the little brown bat, *Myotis lucifugus*. *J. Morphol.* 220:295–305.
- Simmons N.B. 2005. Order Chiroptera: mammal species of the world: a taxonomic and geographic reference. In: Wilson D.E., Reeder D.M., editors. *Mammal species of the world*. Baltimore: Johns Hopkins University Press. p. 312–529.
- Solis-Lemus C., Ané C. 2016. Inferring phylogenetic networks with maximum pseudolikelihood under incomplete lineage sorting. *PLoS Genet.* 12:1–21.
- Smith B.T., Harvey M.G., Faircloth B.C., Glenn T.C., Brumfield R.T. (2014) Target capture and massively parallel sequencing of ultraconserved elements for comparative studies at shallow evolutionary time scales. *Syst. Biol.* 63:83–95.
- Stadelmann B., Lin L.K., Kunz T.H., Ruedi M. 2007. Molecular phylogeny of New World *Myotis* (Chiroptera, Vespertilionidae) inferred from mitochondrial and nuclear DNA genes. *Mol. Phylogenet. Evol.* 43:32–48.
- Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313.
- Stephens M., Donnelly P. 2003. A comparison of Bayesian methods for haplotype reconstruction from population genotype data. *Am. J. Hum. Genet.* 73:1162–1169.
- Stephens M., Smith N.J., Donnelly P. 2001. A new statistical method for haplotype reconstruction from population data. *Am. J. Hum. Genet.* 68:978–989.
- Sullivan J., Demboski J.R., Bell K.C., Hird S., Sarver B., Reid N., Good J.M. 2014. Divergence with gene flow within the recent chipmunk radiation (*Tamias*). *Heredity* (Edinb). 113:185–194.

- Swofford D. 2002. Phylogenetic analysis using parsimony (\* and other methods). version 4. Sunderland, MA: Sinauer Associates.
- Than C., Ruths D., Nakhleh L. 2008. PhyloNet: a software package for analyzing and reconstructing reticulate evolutionary relationships. *BMC Bioinformatics* 9:322.
- Valdez E., Choate J., Bogan M., Yates T. 1999. Taxonomic status of *Myotis occultus*. *J. Mammal.* 80:545–552.
- Yang Z., Rannala B. 2010. Bayesian species delimitation using multilocus sequence data. *Natl. Acad. Sci.* 107:1–6.
- Yang Z., Rannala B. 2014. Unguided species delimitation using DNA sequence data from multiple loci. *Mol. Biol. Evol.* 31: 3125–3135.
- Zerbino D.R., Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* 18:821–829.