Fauskee, B. D., Crowl, A. A., Piatkowski, B., Yoder, A. D., & Tiley, G. P. (2024). Ancient introgression in mouse lemurs (*Microcebus*: Cheirogaleidae) explains 20 years of phylogenetic uncertainty. *Bulletin of the Society of Systematic Biologists*. https://doi.org/10.18061/bssb.v3i1.9319

Investigations

Ancient Introgression in Mouse Lemurs (*Microcebus*: Cheirogaleidae) Explains 20 Years of Phylogenetic Uncertainty

Blake D. Fauskee[1], Andrew A. Crowl[1], Bryan Piatkowski[2,3], Anne D. Yoder[1], George P. Tiley[4]

[1] Duke University, [2] Biosciences Division, Oak Ridge National Laboratory, [3] Division of Computational Biology, Mayo Clinic, [4] Royal Botanic Gardens, Kew

Keywords: Phylogenetic Networks, Gene Tree Discordance, Introgression, Phylogenomics, Mouse Lemurs, Madagascar

https://doi.org/10.18061/bssb.v3i1.9319

Bulletin of the Society of Systematic Biologists

Abstract

Mouse lemurs (genus *Microcebus*) are a clade of approximately 26 named species of small, nocturnal primates endemic to Madagascar. The genus radiated one to ten million years ago and is morphologically cryptic, with most species having been named within the past 20 years largely based on phylogenetic analysis of short fragments of mitochondrial data. More recent work has been focused on revisiting species designations with autosomal nuclear data using more sophisticated statistical approaches. The order of speciation events in *Microcebus* remains contentious, particularly with regard to the placement of the *M. ravelobensis* clade. We investigated support for previous phylogenetic hypotheses based on available whole-genome assemblies from six species and an outgroup. We recovered over 4,000 one-to-one orthologs from these assemblies and used concatenation and coalescent species tree methods to evaluate if differences between previous studies were due to methodological differences or to limitations from too few loci. Observed gene tree discordance was high with patterns inconsistent with incomplete lineage sorting alone. Therefore, we estimated phylogenetic networks to investigate ancient introgression events that may explain observed gene tree distributions and previous phylogenetic conflicts. A network model, invoking some role for introgressive hybridization in the early evolution of *Microcebus*, better characterizes phylogenetic relationships than does any binary species tree. Our results provide insights into the biogeographic history of a threatened and diverse group of primates while also highlighting an important role for phylogenetic network methods in resolving cases of phylogenetic uncertainty.

Introduction

The mouse lemurs (*Microcebus* Cheirogaleidae) are a recent radiation of primates with arguably 26 species (Hotaling et al., 2016; Poelstra et al., 2020). The genetic diversity within the genus *Microcebus* is remarkable among primates and is estimated to have been generated over a one to 10 Myr period, with the estimated ancestral age varying with analytic method (Tiley et al., 2020). Mouse lemurs occupy a wide variety of ecological niches despite maintaining largely cryptic morphological variation (Zimmermann & Radespiel, 2014). Indeed, it is this notable lack of morphological variation that has led to some degree of controversy and suspicion as to whether the explosion in named species is valid (e.g., Tattersall, 2007).

The majority of mouse lemur species are microendemics with restricted ranges across Madagascar's ecologically heterogeneous landscape (Fig. 1), though there are a few generalist species with large ranges across the dry deciduous forests in the west and in a variety of habitats in the northeast (Tiley et al., 2022). Among them is a clade of largely dry-adapted taxa typically referred to as the gray clade, which contains two well-differentiated species, *Microcebus murinus* and *M. griseorufus*. These two species have recently been shown to be completely reproductively isolated, even in contact zones where they co-occur (Poelstra et al., 2022). The majority of mouse lemur species are restricted to Madagascar's wet eastern forests and form a group known as the red clade due to their primarily rufus coloration (Rasoloarison et al., 2013); although, some dispersal events between forest types have been reconstructed with phylogenetic analysis (Yoder et al., 2016). A third clade of mouse lemurs restricted to the dry deciduous forests of northwestern Madagascar is comprised of three named species, *M. ravelobensis, M. bongalovensis,* and *M. danfossi* (Olivieri et al., 2007), and will be henceforth referred to as the *M. ravelobensis* clade. This northwestern region is an ecological transition zone, in the rain shadow of Madagascar's high-

est mountain, where on the eastern side lie the Diana and Sava regions recognized for their diverse biota (J. L. Brown et al., 2014; Callmander et al., 2011; Vences et al., 2009). Remaining populations of the *M. ravelobensis* clade exist in a highly fragmented landscape, likely the product from a combination of Late Pleistocene and Anthropogenic vegetative change (e.g., Matsumotot & Burney, 1994; Railsback et al., 2020), with rivers (Olivieri et al., 2007) and roads (Radespiel et al., 2008) imposing present-day barriers to gene flow between species and poplations.

Within the red clade, *Microcebus mittermeieri* and *Microcebus lehilahytsara* are often recognized as separate species in the general literature. However, they were proposed as being synonymous by Poelstra et al. (2020), a taxonomy that now seems to be generally accepted by the International Union for the Conservation of Nature (IUCN). Using RADseq data, Poelstra et al. (2020) demonstrated that the two taxa followed a single isolation-by-distance pattern, with high levels of gene flow from *M. mittermeieri* to *M. lehilahytsara*. These results thus strongly suggested that *M. mittermeieri* and *M. lehilahytsara* are best considered a single species, with *M. lehilahytsara* having taxonomic priority. We henceforth use this updated taxonomy.

Within the gray clade, *M. murinus* is the most widespread species, occupying a large continuous range in western Madagascar as well as disjunct patches further to the north and southeast (Fig. 1). In the northwest, *M. murinus* exists in sympatry with *M. ravelobensis*. However, *M. murinus* may have only recently expanded into this northwestern region rather than having had a long-term presence (Schneider et al., 2010). Most population genetic and demographic inferences of mouse lemurs have been based on limited mitochondrial sequence data or microsatellites, so it remains an open question as to whether *M. murinus* or a common ancestor of the gray clade may have been historically present in northwestern Madagascar rather than expanding its range from southwestern Madagascar, as implied by Blair et al. (2014).

The phylogenetic placement of the *M. ravelobensis* clade within the larger phylogeny has been a persistent conundrum (Fig. 1). Three well-supported topologies have been recovered in past studies (Fig. 1) (dos Reis et al., 2018; Everson et al., 2023; Heckman et al., 2007; Herrera & Dávalos, 2016; Hotaling et al., 2016; Louis et al., 2006; Olivieri et al., 2007; Springer et al., 2012; Weisrock et al., 2010, 2012; Yoder et al., 2000). It has been difficult to discern the underlying causes of disagreements among studies given that most analyses have been limited to a few mitochondrial or nuclear loci. For example, the data type and amount of data have varied from study to study, ranging from four mitochondrial loci used by Louis et al. (2006) to 334 nuclear loci from a target-capture bait set (Everson et al., 2023 [Preprint]). Additionally, conflicting topologies in the published literature may be due to methodological differences among multilocus studies, such as concatenated maximum likelihood (ML; Felsenstein, 1981) versus the multispecies coalescent (MSC; Rannala & Yang, 2003) analyses. There are well-known biases for ML estimators when the time between speciation events is very short, such that the species tree may not be reflected by a majority of gene trees (Degnan & Rosenberg, 2006; Kubatko & Degnan, 2007). Multiple mouse lemur genomes have recently become available (Hunnicutt et al., 2020; Larsen et al., 2017; Poelstra et al., 2020), allowing for the first phylogenomic investigation of the earliest divergences in *Microcebus*.

To further investigate these phylogenetic conflicts, we take advantage of the recent development of phylogenetic network methods that consider incomplete lineage sorting (ILS) and account for introgressive hybridization (introgression), modeled as episodic cross-species gene flow (Flouri et al., 2020; Solís-Lemus & Ané, 2016; Yu et al., 2012; Zhang et al., 2018). While the importance of introgression has long been recognized in plants and animals (Mallet, 2005; Rieseberg, 1997), especially groups where life histories facilitate gene flow such as wind-pollinated plants (e.g., Dunning et al., 2019; Zhou et al., 2022) or sympatry with synchronous reproduction in vertebrates (e.g., Meier et al., 2017), there is increasing recognition that introgression is an important evolutionary process in mammals (e.g., Jones et al., 2018). Within primates, whole-genome sequence data has recently revealed evidence of cross-species gene flow in the early diversification of Old-World monkeys (Vanderpool et al., 2020). Here, we leverage available genomic data to investigate phylogenetic incongruence in mouse lemurs and the potential role for introgression in the early evolutionary history of a species-rich primate genus.

Methods

Genome sequences and orthology inference

Whole-genome sequences were previously assembled for six mouse lemur species and an outgroup from the sister genus *Mirza* (Supplementary Table S1). Notably, the genome for *Microcebus murinus* is a chromosome-level assembly with RNA-based annotations while, the other genomes are highly fragmented and annotated from *in silico* and homology evidence alone. To avoid potential biases due to heterogenous data quality or reference-based alignment with respect to an ingroup taxon, only the haplotype consensus protein-coding gene models were used for analysis. Orthologous groups were inferred with the pipeline described in Yang and Smith (2014). Only one-to-one orthologs that were present in all seven species were used for downstream phylogenetic analyses. These orthologs were selected using the "1to1" approach outlined in Yang and Smith (2014) which ignores putative orthologs that have duplicated taxa present.

Alignment and gene tree inference

Orthoclusters were aligned with MAFFT v7.310 (Katoh & Standley, 2013) with E-INS-I iterative refinement method. Due to the incompleteness of some genome annotations, the initial alignments contained many gaps. Gaps in these alignments were trimmed using trimAL v1.2 (Capella-Gutiérrez et al., 2009) with the '–gappyout' setting. Individual gene trees were then estimated from these trimmed

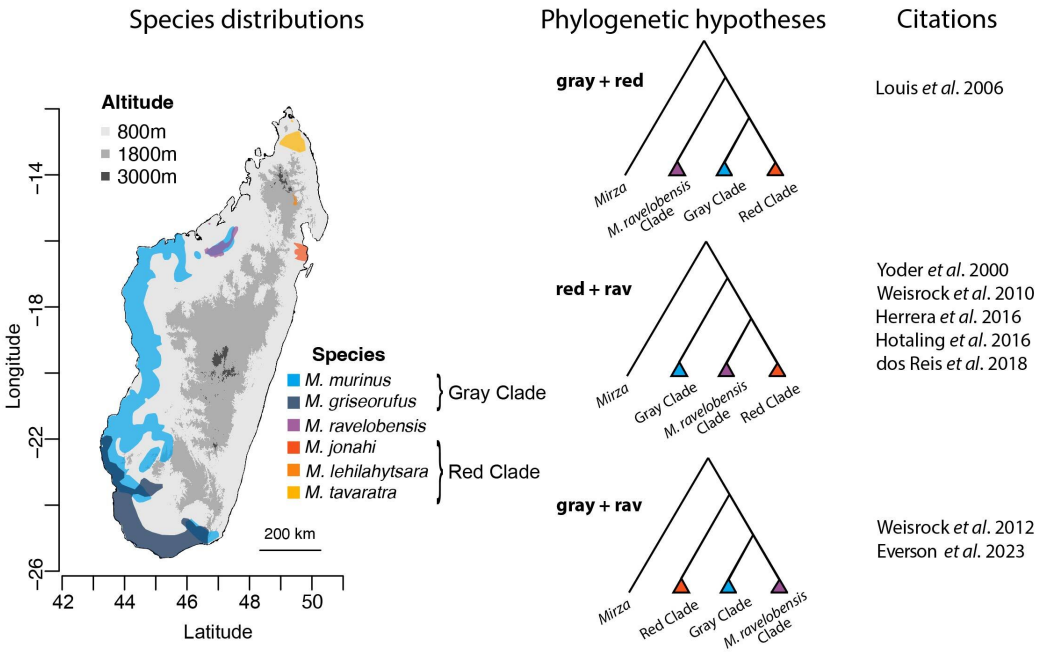

Figure 1. Species distributions of focal lineages and their phylogenetic hypotheses.

Microcebus ravelobensis is found in northwestern Madagascar between *M. murinus* and eastern-distributed members of the red clade. Mouse lemurs are typically restricted to lowland forests and are not found in the Central Highlands (altiudes greater than 800 meters) with some exceptions. There has been disagreement among studies regarding the phylogenetic placement of the *M. ravelobensis* clade; though, the majority of analyses considering concatenate ML and coalescent methods have found *M. ravelobensis* sister to the red clade.

alignments using IQ-TREE 2 v2.1.3 (Minh et al., 2020) with 100 standard nonparametric bootstrap replicates and 10 independent runs. The substitution models for each locus were selected using ModelFinder (Kalyaanamoorthy et al., 2017) implemented in IQ-TREE 2 (-m MFP option). All resulting bootstrap consensus gene trees were then rooted with *Mirza zaza* using the pxrr program in the Phyx phylogenetic toolkit v1.3 (J. W. Brown et al., 2017).

Spurious long branches were then pruned from gene trees using TreeShrink v1.3.9 (Mai & Mirarab, 2018) using the per-species method. The false error rate was set to 0.05 (α=0.05) and the -i option was added to output pruned alignments. Gene trees were then inferred again from the pruned alignments using the same methods described above.

Species tree inference

Three methodologies were used to infer species trees: a concatenated ML approach, a summary statistic MSC approach using ASTRAL-III (Zhang et al. 2018), and an invariant approach consistent under the MSC using SVDQuartets (Chifman & Kubatko, 2014). For concatenation, individual pruned and gap-trimmed alignments were concatenated using AMAS (Borowiec, 2016). The resulting supermatrix was then partitioned by gene, and the MFP+MERGE option in IQ-TREE 2 was used to simultaneously find the best partitioning scheme by merging partitions and select the best substitution model for each final partition. To speed up the partition merging process, only the top 10% of partition schemes was considered. Finally, a concatenated phylogeny was generated in IQ-TREE 2 version 2.2.2.6 with 1000 ultrafast bootstrap (UFBoot) replicates and 10 independent runs. The concatenated phylogeny was then rooted with *Mirza zaza* using the pxrr program in the Phyx phylogenetic toolkit v1.3 (J. W. Brown et al., 2017).

A species tree was also inferred using ASTRAL-III v5.7.8 (Zhang et al. 2018) To perform the initial ASTRAL analysis, ML bootstrap consensus trees for each of the 4,147 orthologs were given as input, and one ASTRAL run was conducted on the consensus trees. An additional bootstrapping analysis was conducted within ASTRAL-III (-b option) using the 100 bootstrap replicates from each locus (414,700 trees in total). Branch lengths were then optimized in coalescent units using qAge (Peng et al., 2022) implemented in PAUP* v4.0a (Swofford, 2002).

The SVDQuartets (Chifman & Kubatko, 2014) analysis was carried out within PAUP* v4.0a (Swofford, 2002) using all 4,147 ortholog alignments as input. Each ortholog alignment was combined into one nexus file and partitioned by gene. SVDQuartets (Chifman & Kubatko, 2014) was then run using exhaustive quartet-sampling, and node support was assessed with 100 nonparametric bootstrap replicates. Branch lengths were then optimized in coalescent units using qAge (Peng et al., 2022) implemented in PAUP* v4.0a (Swofford, 2002). Branch length estimation was carried out under the GTR model using estimated model paramaters (patProb=expBL).

Gene tree conflict analysis

Multiple analyses were conducted to assess topological discordance amongst gene trees. First, a gene tree discordance analysis was conducted in DiscoVista (Sayyari et al., 2018,

mode 2). For this analysis, users select specific species relationships to analyze. Five separate species relationships were selected for this analysis: 1. the Gray Clade (*M. murinus* + *M. griseorufus*), 2. the Red Clade (*M. tavaratra* + *M. jonahi* + *M. lehilahytsara*), 3. The Gray Clade + Ravelobensis (*M. murinus* + *M. griseorufus* + *M. ravelobensis*), 4. the Red Clade + Ravelobensis (*M. tavaratra* + *M. jonahi* + *M. lehilahytsara* + *M. ravelobensis*), and 5. Ravelobensis Sister (*M. murinus* + *M. griseorufus* + *M. tavaratra* + *M. jonahi* + *M. lehilahytsara*). For this analysis, users also specify a cutoff for gene trees that "strongly" support or reject the relationship in question. Here, a cutoff of 70% bootstrap support was used. Next, a relative frequency analysis was conducted also in DiscoVista (Sayyari et al., 2018, mode 6). This analysis calculated the relative frequency of gene trees supporting certain backbone phylogenetic relationships. The backbone relationships assessed here were: 1. Gray Clade + Ravelobensis, 2. Red Clade + Ravelobensis, and 3. Gray Clade + Red Clade. For both DiscoVista (Sayyari et al., 2018) analyses, rooted gene trees were used as input.

Phylogenetic concordance factors were also calculated as an additional metric of gene tree discordance. Both gene concordance factors (gCFs) (Minh et al., 2020) and site concordance factors (sCFls) (Mo et al., 2023) were calculated. For site concordance factors, the maximum likelihood-based approach was used (–scfl in iq. IQ-TREE 2 versions ⩾ 2.2.2.6). gCFs are the percentage of decisive ortholog trees supporting each node in the species tree. Minh et al. (2020) defines a gene tree as "decisive" if it contains at least one of taxa present in the taxon subset under question in the species tree. sCFls are the percentage of alignment sites supporting each node in the species trees. With the maximum likelihood-based concordance factor approach, a substitution model is applied, selected with ModelFinder (Kalyaanamoorthy et al., 2017), and a maximum likelihood tree search is carried out. Both gCFs and sCFls were calculated for each of the three species-tree topologies. To calculate sCFls, 1,000 quartets were sampled (–scfl 1000 in IQ-TREE 2).

Tree topology tests

Using IQ-TREE 2 version 2.2.2.6, several RELL approximation (Kishino et al., 1990) tree topology tests were conducted to compare species tree topologies from different species tree inference methods. These tests include the Kishino-Hasegawa (KH) test (Kishino & Hasegawa, 1989), a bootstrap proportion (BP) test, the Shimodaira-Hasegawa (SH) test (Shimodaira & Hasegawa, 1999), and an expected likelihood weights test (ELW) (Strimmer & Rambaut, 2002). For each of these tests, model parameters were estimated from an initial parsimony tree (-n 0 option in IQ-TREE 2), and 10,000 RELL replicates were used. Additionally, an approximately unbiased test (AU) test (Shimodaira, 2002) was conducted in IQ-TREE 2 version 2.2.2.6.

Phylogenetic network analysis

The Species Networks applying Quartets (SNaQ; Solís-Lemus & Ané, 2016) function in the PhyloNetworks package (Solís-Lemus & Ané, 2016) was applied to our ortholog dataset to investigate potential network topologies. SNaQ estimates phylogenetic networks based on expectations of quartet concordance factor (CF) distributions under the MSC model to differentiate gene tree discordance due to ILS from gene tree discordance arising through horizontal inheritance. When coupled with BUCKy (Larget et al., 2010), which can estimate quartet CF distributions from Bayesian posterior samples, SNaQ can additionally account for gene tree estimation error.

To perform this network analysis, ortholog trees were estimated with MrBayes v3.2.7a (Ronquist et al., 2012) using the GTR+I+G model (nst=6, rates=invgamma) and a variable rate prior (prset ratepr=variable). Four chains were used for each MCMC run and were carried out for 2.5 million generations, sampling every 100 generations. Several individual trace plots were visually inspected in Tracer v1.7.2 (Rambaut et al., 2018) to assess parameter convergence. Posterior samples for each ortholog were then used to conduct a Bayesian concordance analysis using BUCKy v1.4.4, discarding the first 1,000 trees from each posterior sample as burnin. The table of quartet CFs generated with BUCKy were then used as input to SNaQ.

SNaQ allows users to specify the maximum number of reticulations (h_{max}) allowed in a network search. We determined the best value of h_{max} by performing independent analyses from 0 to 3 using slope heuristics (Baudry et al., 2012). The rooted species tree output from ASTRAL-III was used as the starting tree for h_{max} values of 0 and 1. For h_{max} = 2, the network output from the h_{max} = 1 analysis was used as the starting tree. Likewise, the output from h_{max} = 2 was used as the starting tree for the h_{max} = 3 analysis. For each value of h_{max}, we performed 10 independent runs to ensure reasonable exploration of the search space and that the best network and pseudolikelihood score was recovered. We also defined *Mirza zaza* as the outgroup during network estimation.

To assess variability in the inheritance probability (γ) estimation, we calculated one standard deviation around the inheritance probability by using the 100 inheritance probability estimates from the bootstrap replicates. The best network across all h_{max} values analyzed was selected, and branch support on that network was assessed using a bootstrapping analysis. Bootstrapped networks were obtained from the *bootsnaq* function with 100 replicates and ten independent searches per replicate, using the BUCKy concordance factor table as input. Lastly, we conducted the same analysis using the SVDQuartets topology as the starting tree for h_{max} values of 0 and 1.

Quantifying introgression using allele frequencies

We sought to identify genomic patterns of admixture by estimating Patterson's *D* and f_4-ratio statistics for these mouse lemur species. The *D* statistic, or ABBA-BABA test, aims to identify excess allele sharing between populations of a rooted trio through analysis of biallelic SNPs (Hibbins & Hahn, 2022; Patterson et al., 2012). The site pattern encountered most frequently in the tree (((P1,P2),P3),O) with

ancestral (A) and derived (B) alleles is expected to be BBAA in which both P1 and P2 have the derived allele. The frequencies of the two discordant genealogical relationships, ABBA and BABA, are expected to be equal under a scenario of incomplete lineage sorting, and deviation from this equality suggests gene flow has occurred between either P1 or P2 and the more distantly related P3. In a similar fashion, one can calculate the f_4-ratio which approximates the fraction of admixed ancestry (Patterson et al., 2012).

We used the program Dsuite v0.5 r50 (Malinsky et al., 2021) to estimate the *D* and f_4-ratio statistics. SNP-sites v2.5.1 (Page et al., 2016) was used to identify SNPs from the concatenated ML supermatrix. We tested both the ML tree and the major topology of the best network. For each analysis, we report statistics for those rooted trios consistent with the tree topology and assessed significance using jack-knifing with 20 blocks of the data. Raw P-values were adjusted using the Benjamini-Hochberg procedure (Benjamini & Hochberg, 1995). Furthermore, we calculated the *f*-branch statistic (Malinsky et al., 2018) to determine whether potentially correlated *D* statistics for extant species could reflect ancestral gene flow that involved internal branches of the phylogeny.

Comparing network and tree models

We evaluated the strength of evidence for our network hypothesis over possible binary trees or more complex introgression scenarios using Bayes factors. BPP v4.6.2 (Flouri et al., 2018) was used to estimate marginal likelihoods (Rannala & Yang, 2017) using stepping-stone sampling (Xie et al., 2010) with 24 steps. Marginal likelihoods were estimated for the three binary trees using the MSC model and the best network recovered with SNaQ using the multispecies coalescent with introgression (MSci) model (Flouri et al., 2020), also referred to as the multispecies network coalescent (Yu et al., 2012) based on the 4,147 alignments after trimming low-confidence alignment columns and spurious tips. We also evaluated the strength of evidence for two-rate introgression models where episodic gene flow occurred from *M. ravelobensis* into the common ancestor of *M. murinus* and *M. griseorufus* and the extant *M. murinus* lineage (two-rate model A) or the common ancestor and the extant *M. griseorufus* lineage (two-rate model B). Posteriors collected 10,000 samples with an interval of 50 after a burnin of 50,000 samples. Small pilot runs were used to determine appropriate MCMC settings. Four independent runs of all models were used for parameter estimation and to check convergence and that mixing was sufficient for models with and without introgression. Control files and data are available on Dryad (https://doi.org/10.5061/dryad.xgxd254nj). Marginal likelihoods and their standard errors were calculated from the power posteriors using the bppr R package (https://github.com/dosreislab/bppr).

Results

Orthology and alignment

In total, the orthology inference pipeline implemented in this study recovered 4,147 one-to-one orthologs that were present in all seven sampled species. After alignment and gap trimming, the supermatrix totaled 6,731,343 bases. Individual ortholog alignments, after gap-trimming, ranged in length from 114 bases to 15,006 bases. Missing data ranged from 5.04% missing sequences in *Microcebus ravelobensis* to 3.11% in *M. tavaratra* (Supplementary Table S2).

Species tree estimation

The ASTRAL and concatenated analyses produced congruent, well-supported trees (bootstrap and UFBoot = 100, local posterior probability = 1) (Fig. 2A,B). The ASTRAL and concatenated species trees both placed *M. ravelobensis* sister to the gray clade (*M. griseorufus* and *M. murinus*; Figs. 2A and 2B). The best tree from SVDQuartets recovered *M. ravelobensis* as sister to all other *Microcebus*, with 81.6% bootstrap support and 100% bootstrap support across all other nodes (Fig. 2C). However, optimizing branch lengths with qAge (Peng et al., 2022) revealed the branch connecting *M. ravelobensis* to the rest of *Microcebus* to be extremely short (~0).

Gene tree discordance

To investigate the discordance observed between the ASTRAL and concatenated analyses versus the SVDQuartets analysis, gene- and site-concordance factors (henceforth referred to as gCF and sCFl, respectively) were calculated for each node on each tree. The *M. griseorufus* + *M. murinus* relationship, which is recovered in all three species tree analyses, had the highest gCF and sCFl values: 62.7 and 58.7 (or 58.5), respectively, for the ASTRAL and concatenated tree and 62.8 and 64.3, respectively, for the SVDQuartets tree (Fig. 2).

The node with the lowest gCF and sCFl was the *M. ravelobensis* sister to the red and gray clades, a relationship recovered only by SVDQuartets, which has a gCF of 11.9 and sCFl of 24.3 (Fig. 2C). The next lowest gCF and sCFl values across the three species analyses are 33.6 and 42.5, respectively, which both occur in the gray clade containing *M. ravelobensis* which was recovered by the ASTRAL and concatenated analyses but not SVDQuartets (Fig. 2A,B).

Our gene tree discordance analyses in DiscoVista reveal high levels of gene tree discordance but generally support the ASTRAL/concatenated topology (Fig. 3A, B). Among the five relationships analyzed, the gray clade has the highest amount of gene tree support (Fig. 3A). Concerning *M. ravelobensis*, *M. ravelobensis* sister to the gray clade has the highest support among the gene trees, echoing the ASTRAL/concatenated topology (Fig. 3A). The SVDQuartets topology (Ravelobensis Sister in Fig. 3A) has the lowest support among gene trees (Fig. 3A). The relative frequency analysis also identifies the ASTRAL/concatenated topology

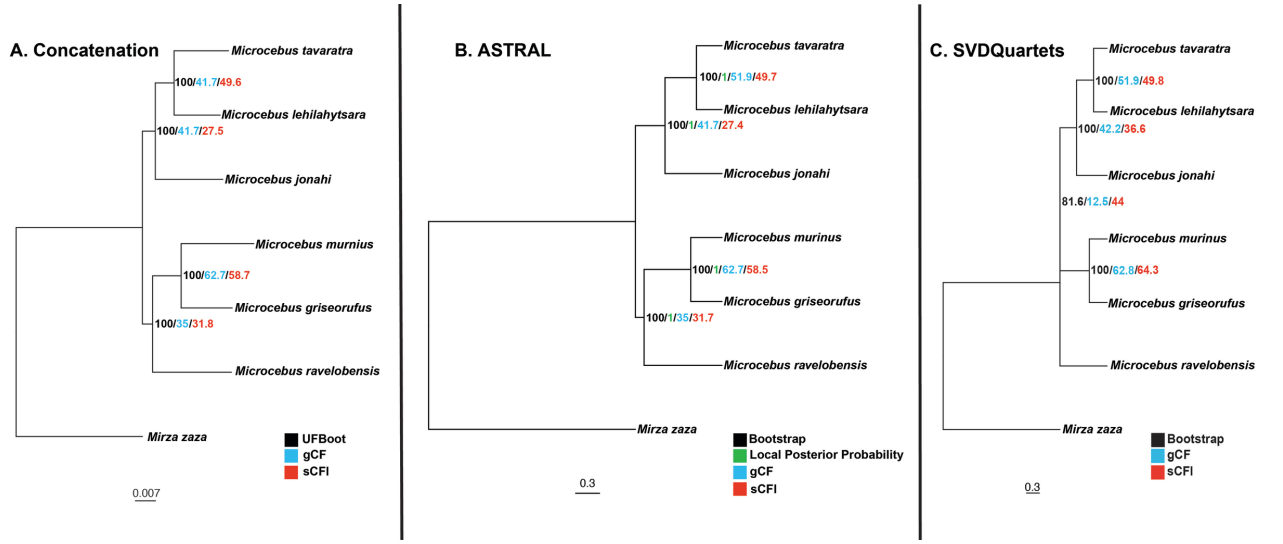

Figure 2. Phylogenetic relationships of Microcebus inferred from three species tree analyses.

A. Concatenated maximum likelihood phylogeny from IQ-TREE 2. Node support values in black represent ultrafast bootstrap (UFBoot) percentages from 1000 UFBoot replicates. Branch lengths represent substitutions/site. B. ASTRAL species tree. Node support values in black represent bootstrap percentages from 100 bootstrap replicates, and values in green represent local posterior probabilities. Branch lengths are in coalescent units. Terminal branches were arbitrarily set to a value of 0.1. C. SVDQuartets species tree. Node support values in black represent bootstrap percentages from 100 bootstrap replicates. Branch lengths are in coalescent units. Blue and red support values on all three trees represent gCF and sCFl values, respectively.

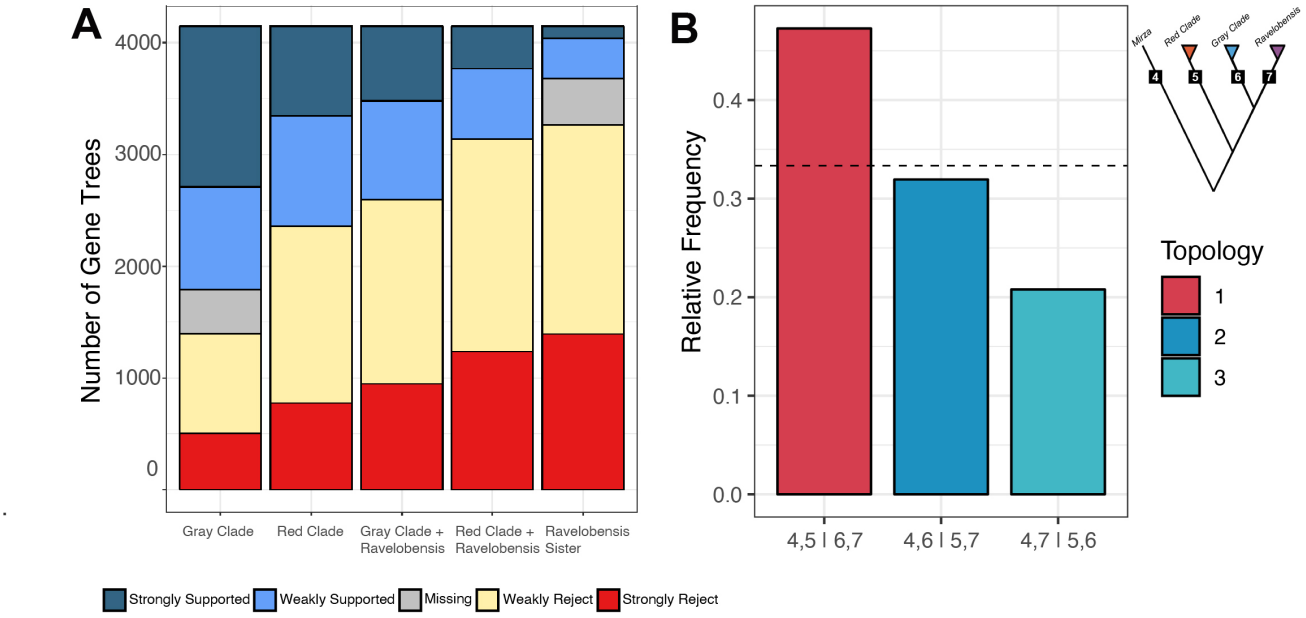

Figure 3. DiscoVista analyses of gene tree discordance.

A. Number of gene trees supporting or rejecting certain species relationships in Microcebus. Dark blue represents gene trees that support a given relationship with ⩾ 70% bootstrap support, light blue represents gene trees that support a given relationship with < 70% bootstrap support. Red represents gene trees that support a conflicting relationship with ⩾ 70% bootstrap support, and yellow represents gene trees that support a conflicting relationship with < 70% bootstrap support. Gray represents missing data. B. Proportion of gene trees supporting a specific species tree arrangement. The dashed black line represents a 0.33 proportion of gene trees.

as occurring at the highest frequency among the gene trees (topology 1 in Fig. 3B). Both Red Clade + Ravelobensis and Ravelobensis Sister are supported in less than one-third of the gene trees (Fig. 3B).

Tree topology tests

Several tree topology tests were conducted in IQTREE 2 (Minh et al., 2020) to compare the ASTRAL/concatenation and SVDQuartets topologies (Table 1). These tree topology tests revealed that the ASTRAL/concatenation topology had a higher likelihood. Furthermore, all tree topology tests

Table 1. Tree topology test results.

Tree Topology	lnL	Δ lnL	bp-RELL	p-KH	p-SH	ELW	p-AU
(Mrav,(Grey,Red))	-11326641.7	3518.8	0*	0*	0*	0*	3.29×10^{-60}*
(Red,(Grey,Mrav))	-11323122.9	-	1	1	1	1	1

lnL = log-likelihood, ΔlnL = change in log likelihood, bp-RELL = bootstrap proportion using the RELL method, p-HK = p-value from one sided Kishino-Hasegawa test, p-SH = p-value from Shimodaira-Hasegawa test, ELW = expected likelihood weight, p-AU = p-value from approximately unbiased test. Asterisks indicate a significant rejection of a tree topology.

conducted here significantly rejected the SVDQuartets topology.

Phylogenetic network analyses

A plot of pseudolikelihood network scores revealed the best network had one hybridization event (h_{max} = 1, Supplemental Fig. S1) as the pseudolikelihood scores for h_{max} = 2 and h_{max} = 3 increased very little. The best network's major topology placed *M. ravelobensis* sister to the red clade with a reticulation between *M. ravelobensis* and the Gray Clade (Fig. 4). The hybrid edge received relatively high bootstrap support of 95%. The inheritance probability (γ) estimated for this reticulation, which in our case is the proportion of the genome contributed from the *M ravelobensis* lineage into the common ancestor of the Gray Clade, revealed that a large proportion of the genome was exchanged in this reticulation event (γ = 0.48) (Fig. 4). This inheritance probability of 0.48 has a standard deviation of 0.12. Marginal likelihoods strongly favored the network recovered by SNaQ over any binary tree or the more complex two-rate networks (Table 2). Multiple runs for the network model show all parameters converged with good mixing despite the complexity of introgression models and autocorrelation of population sizes and divergence times under the MSC (Supplementary Figs. S2-S4). The best binary tree model placed *M. ravelobensis* sister to the gray clade. In the case of using the SVDQuartets topology as the starting tree, SNaQ rejects the starting topology and recapitulates the same network relationship as in Fig. 4 (Supplementary Fig. S5).

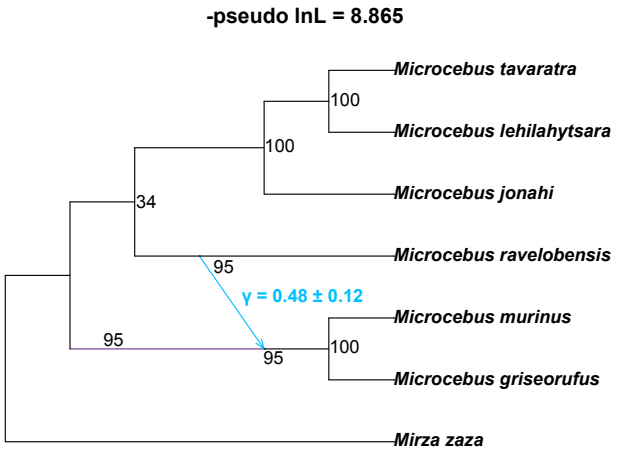

Figure 4. Microcebus phylogenetic network.

Node support values in black represent bootstrap support percentages from 100 bootstrap replicates. These node support values correspond to hybrid edge and hybrid node support in addition to bifurcating node support. The minor hybrid edge is colored blue, while the major hybrid edge is colored purple. The inheritance probability (γ) for the minor hybrid edge is depicted in blue with one standard deviation.

Gene tree frequency tests for introgression

Analysis of introgression using *D* and f4-ratio statistics provide strong support for interspecific introgression. After controlling the false-discovery rate, 13 of 20 rooted trios showed significant (adjusted P < 0.05) introgression when using either the ML tree topology or the major topology of the best network (Table 3; Supplementary Table S3). The f_4-ratio statistics suggest that roughly 4% (ML topology) to 7% (network topology) of the genome is introgressed on average across these significant tests, with the highest amounts (ca. 11-13%) shared between *M. ravelobensis* and both *M. griseorufus* and *M. murinus* when using the network topology (Fig. 5). Indeed, *f*-branch statistics also show such excess gene flow between *M. ravelobensis* and extant members of the Gray Clade (Fig. 5) but were unable to test if this signal could be attributed to introgression involving an ancestor of the gray clade due to topological incompatibilities. Statistics estimated from sequences that were trimmed of spurious branches that may arise from alignment or orthology errors (Fig. 5; Supplementary Table S3) increased the signal for introgression from site patterns between *M. ravelobensis* and species in the grey clade compared to sequences without trimming (Supplementary Fig. 7; Supplementary Table S3), suggesting our results reflect biological processes and not technical artifacts.

Discussion

Gene-tree variation and the disagreement between gene trees and species trees is now well-recognized as an outcome of evolutionary processes rather than as a problem to overcome (Bravo et al., 2019; Maddison, 1997). We can derive expectations for distributions of gene trees due to ILS (Hudson, 1983; Pamilo & Nei, 1988; Rannala & Yang, 2003; Rosenberg, 2002), which provides the foundation for observing deviations from ILS that are perhaps better explained by introgression (e.g., Cai & Ané, 2021; Meng & Kubatko, 2009). Introgression has long been recognized in plants (e.g., Riesberg 1997) and animals (e.g., Mallet, 2005), but genomic data and phylogenomic methods are now highlighting a role for introgression in explaining contentious phylogenetic relationships (e.g., Stull et al., 2020). Our analyses provide additional evidence for ancient introgression underlying topological incongruence near the ancestral radiation of primate groups. For example, potential

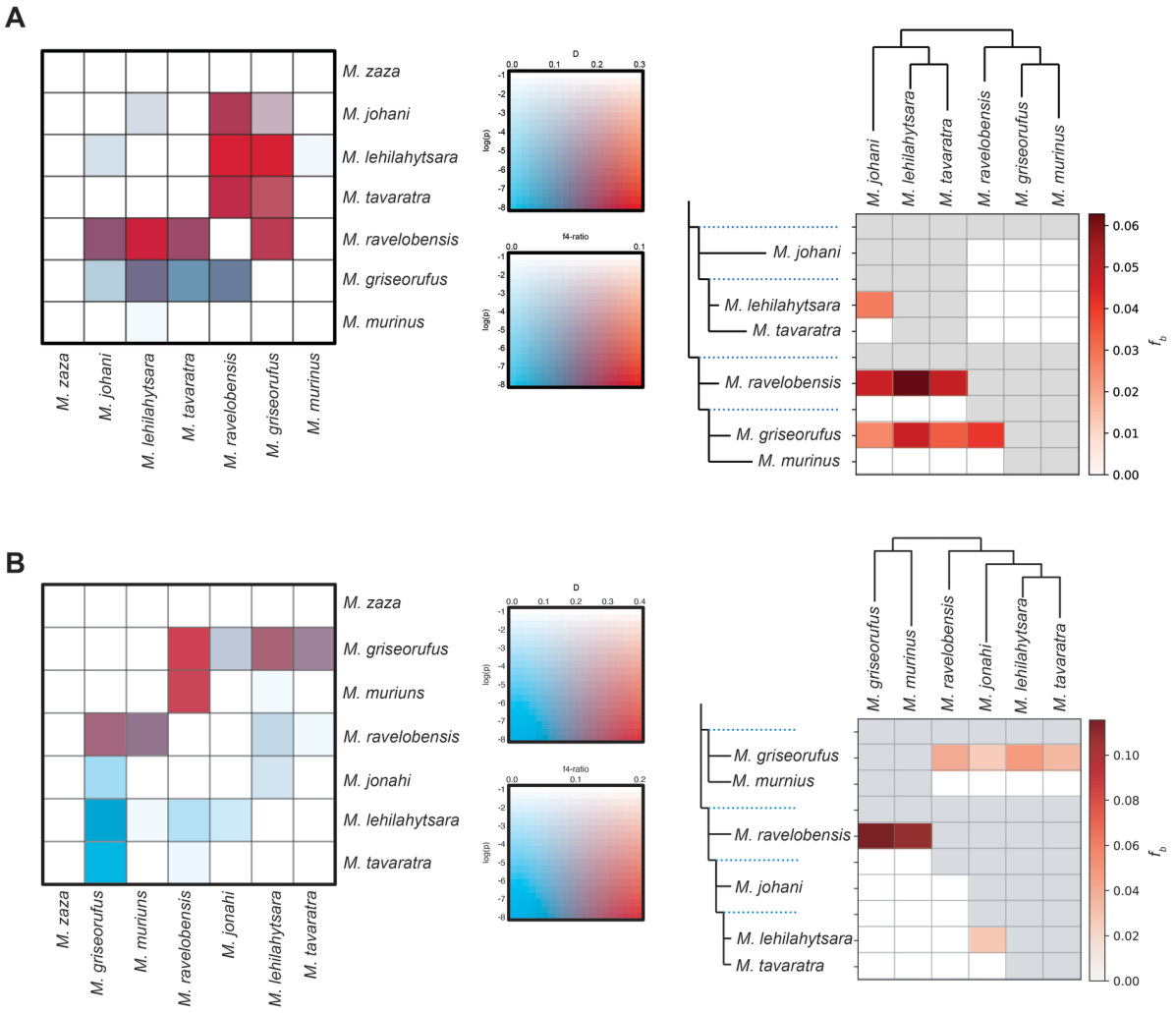

Figure 5. Signatures of introgression among species of Microcebus.

Gene tree frequency tests were performed using (A) the maximum likelihood species tree topology and (B) the major phylogenetic network topology. Left: D (above diagonal) and f4-ratio (below diagonal) statistics depicting estimated gene flow between species pairs Right: f-branch statistic with compatible tests shown in grey.

Table 2. Bayes Factors for competing evolutionary hypothesis

Topology	marginal lnL	SE	2lnBF†
(Mrav,(Grey,Red))	-11255520	4.164485	-1478
(Red,(Grey,Mrav))	-11255150	4.190757	-738
(Grey,(Mrav,Red))	-11255557	4.088052	-1552
Network Hypothesis	-11254781	4.148756	0
Two-rate Model A	-11255439	4.377716	-658
Two-rate Model B	-11255465	4.22489	-684

†log Bayes Factors calculated with respect to the best network model

widespread ancient introgression was recently characterized among Old-World monkeys (Vanderpool et al., 2020). Similarly, introgression is emerging as a common observation in mouse lemur genomic data (Poelstra et al., 2020), a finding that will perhaps be prevalent within Lemuriformes as increasing genomic data are generated (e.g., Everson et al., 2023 [Preprint]).

A phylogenomic perspective of mouse lemur evolution

Our species tree analyses recovered two distinct topologies. Both concatenation and ASTRAL-III placed the *M. ravelobensis* clade sister to the gray clade with high bootstrap and UFBoot support, which together are sister to the red clade (Fig. 2A, B). This topology has only been reported

Table 3. Introgression test statistics from the trimmed dataset.

Topology	P1	P2	P3	*D*	f_4-ratio	Significance
Both	*M. jonahi*	*M. lehilahytsara*	*M. griseorufus*	0.13	0.03	**
Both	*M. murinus*	*M. griseorufus*	*M. jonahi*	0.17	0.03	***
ML	*M. griseorufus*	*M. ravelobensis*	*M. johani*	0.13	0.04	***
Network	*M. johani*	*M. ravelobensis*	*M. griseorufus*	0.36	0.13	***
Both	*M. johani*	*M. tavaratra*	*M. griseorufus*	0.07	0.02	n.s.
Both	*M. murinus*	*M. griseorufus*	*M. lehilahytsara*	0.27	0.05	***
ML	*M. griseorufus*	*M. ravelobensis*	*M. lehilahytsara*	0.12	0.04	***
Network	*M. lehilahytsara*	*M. ravelobensis*	*M. griseorufus*	0.28	0.10	***
Both	*M. tavaratra*	*M. lehilahytsara*	*M. griseorufus*	0.08	0.01	n.s.
Both	*M. murinus*	*M. griseorufus*	*M. ravelobensis*	0.23	0.04	***
Both	*M. murinus*	*M. griseorufus*	*M. tavaratra*	0.22	0.03	***
ML	*M. griseorufus*	*M. ravelobensis*	*M. tavaratra*	0.12	0.03	**
Network	*M. tavaratra*	*M. ravelobensis*	*M. griseorufus*	0.30	0.12	***
Both	*M. jonahi*	*M. lehilahytsara*	*M. murinus*	0.06	0.01	n.s.
Both	*M. jonahi*	*M. lehilahytsara*	*M. ravelobensis*	0.13	0.03	**
Both	*M. tavaratra*	*M. lehilahytsara*	*M. jonahi*	0.12	0.03	**
ML	*M. murinus*	*M. ravelobensis*	*M. jonahi*	0.22	0.06	***
Network	*M. jonahi*	*M. ravelobensis*	*M. murinus*	0.34	0.11	***
Both	*M. jonahi*	*M. tavaratra*	*M. murinus*	0.03	0.01	n.s.
Both	*M. jonahi*	*M. tavaratra*	*M. ravelobensis*	0.06	0.02	n.s.
ML	*M. murinus*	*M. ravelobensis*	*M. lehilahytsara*	0.27	0.09	***
Network	*M. lehilahytsara*	*M. ravelobensis*	*M. murinus*	0.32	0.10	***
Both	*M. tavaratra*	*M. lehilahytsara*	*M. murinus*	0.04	0.01	n.s.
Both	*M. tavaratra*	*M. lehilahytsara*	*M. ravelobensis*	0.08	0.01	n.s.
ML	*M. murinus*	*M. ravelobensis*	*M. tavaratra*	0.24	0.07	***
Network	*M. tavaratra*	*M. ravelobensis*	*M. murinus*	0.32	0.11	***

The *D* and *f*4-ratio statistics are reported for each rooted trio that is consistent with the maximum likelihood (ML), major network (network), or both (Both) tree topologies. A significant test indicates significant signature of introgression between P2 and P3. Significance codes: n.s., not significant; *, adjusted P < 0.05; **, adjusted P < 0.01; ***, adjusted P < 0.001.

twice in the literature: first by Weisrock et al. (2012) who leveraged twelve nuclear and two mitochondrial loci and second by Everson et al. (2023) [Preprint] with 334 nuclear loci. Weisrock et al. (2012) also reported strong gene-tree discordance within *Microcebus* which they attributed to ILS. The second topology we recovered using SVDQuartets placed the *M. ravelobensis* clade sister to both the red and gray clades with high bootstrap support (Fig. 2C). This topology was reported in the literature only once by Louis et al. (2006). In that study, Louis et al. (2006) used parsimony and recovered low bootstrap support for this relationship.

The utility of the bootstrap as a measure of support in large phylogenomic datasets has recently undergone increasing scrutiny (see Simon, 2022; Thomson & Brown, 2022). Several studies leveraging large, multi-locus phylogenomic datasets to attempt to resolve historically difficult nodes have demonstrated that several conflicting topologies can be obtained that are all maximally supported using metrics such as bootstrap proportions and Bayesian posterior probabilities (J. M. Brown & Thomson, 2017; Li et al., 2021). Thus, it is becoming increasingly important to evaluate additional support metrics with genomic data. To this end, we calculated gCFs and sCFls on each of our recovered species trees. On the SVDQuartets tree, we obtained strikingly low gCF and sCFl values on the node placing *M. ravelobensis* sister to the rest of *Microcebus* (Fig. 2C). Here, less than 12% of genes and less than 24% of sampled sites support this relationship. In both of our distinct recovered topologies (concatenation+ASTRAL versus SVDQuartets), we find that gCF's and sCFl's are decreased on nodes that give rise to the *M. ravelobensis* clade compared to other splits (Fig. 2). This suggests that the position of the *M. ravelobensis* clade varied greatly among gene trees and that previous studies were susceptible to sampling error.

Our gene tree discordance analyses carried out in DiscoVista revealed similar wide-spread disagreement among gene trees regarding the earliest divergences in *Microcebus*.

For example, over half of gene trees disagree with the ASTRAL/concatenation placement of *M. ravelobensis* sister to the gray clade, with just under one-fourth of gene trees

strongly (⩾ 70% bootstrap support) disagreeing with this relationship (Fig. 3). Furthermore, the earliest split recovered by SVDQuartets is supported by relatively few gene trees, with the majority of gene trees showing preference for a different topology (Fig 3).

By comparing different species topologies, it became apparent that the bifurcating species tree that best fit the data is one that places *M.ravelobensis* sister to the gray clade, as was recovered by ASTRAL and concatenation (Fig. 2A, B). A relative frequency analysis showed that this topology is the most common topology found amongst the gene trees (Fig. 3B). The topology recovered by SVDQuartets that places *M. ravelobensis* sister to the rest of *Microcebus* is the least common amongst the gene trees, appearing in just over 20% of them (Fig. 3B). This topology is also significantly rejected in favor of the ASTRAL/concatenation topology by all of the tree topology tests we conducted (Table 1). A third topology that places *M. ravelobensis* sister to the red clade is the most common topology reported in the literature (Fig. 1). However, none of our species tree methods recover this relationship (Fig. 2), and it is the second-most common relationship among the gene trees, being supported by just under one-third of the gene trees (Fig. 3B).

Both ILS and introgression in the history of Microcebus

Gene tree discordance is now well-accepted as a feature of interesting data rather than a problem for evolutionary biology (Bravo et al., 2019); although, we expect some gene tree discordance is caused by short alignments, orthology errors, model misspecification, and a general lack of information to recover accurate gene trees. The observed variation, however, is consistent with previous variation in analyses and even the disagreement between species tree methods used here. Why SVDQuartets estimated a somewhat unexpected well-supported species tree topology is unclear, but it is important to note that, after optimizing branch lengths with qAge (Peng et al., 2022), the node representing the common ancestor of *M. ravelobensis* and the rest of *Microcebus* is effectively zero (Fig. 2C). The gene trees suggested there was little information supporting the SVDQuartets topology (Fig. 3), and this may also suggest that representing *Microcebus* with a strictly bifurcating tree is inappropriate.

In some scenarios a phylogenetic network, as opposed to a strictly bifurcating tree, may be a better representation of the evolutionary history for a given clade. Here, we estimated phylogenetic networks for *Microcebus* using both a topology-based pseudolikelihood and site-pattern method and evaluated the strength of evidence for the inferred network over alternative trees or more complex models with Bayes factors. Our network hypothesis suggested some degree of gene flow between the common ancestors of the *M. ravelobensis* and gray clade (Fig. 4). Present-day distributions are not necessarily reflective of past distributions, but Madagascar was roughly in its present-day ecological configuration since the late Miocene (Wells, 2003), while *Microcebus* diversified between 10 Ma (dos Reis et al., 2018) and 1.5 Ma ago (Poelstra et al., 2020). Thus, there was likely opportunity for introgressive hybridization to occur in the past, following some period of isolation prior to the present-day sympatry of *M. ravelobensis* and *M. murinus* (Fig. 1).

Reinterpreting mouse lemurs through phylogenetic networks

Previous studies have suggested origins of *Microcebus* in southwestern Madagascar (Blair et al., 2014). Our findings imply origins in northwestern Madagascar, in the dry deciduous forest. The highest mountain in Madagascar, Mount Tsaratanana, and the ecological transition from dry to wet forest has not been an impermeable barrier to gene flow, at least in the past. The *M. ravelobensis* lineage has shared ancestry with the gray clade and the wet forest-distributed species from the red clade. This presents a new lens through which to interpret the processes driving speciation within the group, such that diversification occurred in a ring pattern around Madagascar and that species boundaries are reinforced in current areas of sympatry between gray and red clade members (Hotaling et al., 2016; Yoder et al., 2016). Conspecific recognition is likely achieved through some combination of olfactory sensation (Hunnicutt et al., 2020) and auditory cues (Hasiniaina et al., 2020), so prezygotic reproductive barriers may have been strong enough to prevent gene flow in cases of secondary contact.

Conclusions

The charismatic, though poorly understood, mouse lemurs (*Microcebus*) of Madagascar represent a recent radiation of approximately 26 morphologically cryptic taxa. Here, we provide the first phylogenomic investigation into the early-diverging backbone of the group. Past studies have failed to satisfactorily resolve the evolutionary history of the genus due to insufficient data and restricted analytical tools. We found that both ILS and introgression likely underly previous disagreements among phylogenetic studies. The placement of the *M. ravelobensis* clade has been especially perplexing. Using recently developed phylogenetic network analyses coupled with genomic data, we inferred historical introgression between this lineage and the gray clade. Our analyses demonstrate the importance of phylogenetic networks for understanding patterns of gene tree discordance as well as the diversification of diverse groups.

Funding

BDF is supported by a Graduate Research Fellowship awarded by the National Science Foundation. BP was supported by funding from the Laboratory Directed Research and Development Program of Oak Ridge National Laboratory, managed by UT-Battelle, LLC, for the U.S. Department of Energy under contract no. DE-AC05–00OR22725. Data generation and manuscript preparation were also supported by Duke University research funds and National Science Foundation Grant DEB-1354610 to ADY. This project

has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No. 101026923, awarded to GPT.

Acknowledgements

The authors thank the labs of Paul Manos and John Shaw for helpful discussions in the project development. Two anonymous reviewers and the editor Claudia Solís-Lemus provided feedback that improved the quality of the manuscript.

Supporting Information

Data available from the Dryad Digital Repository: https://doi.org/10.5061/dryad.xgxd254nj

Submitted: November 28, 2022 EDT, Accepted: November 17, 2023 EDT

References

Baudry, J.-P., Maugis, C., & Michel, B. (2012). Slope heuristics: overview and implementation. *Statistics and Computing*, *22*(2), 455–470. https://doi.org/10.1007/s11222-011-9236-1

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Blair, C., Heckman, K. L., Russell, A. L., & Yoder, A. D. (2014). Multilocus coalescent analyses reveal the demographic history and speciation patterns of mouse lemur sister species. *BMC Evolutionary Biology*, *14*(1), 57. https://doi.org/10.1186/1471-2148-14-57

Borowiec, M. L. (2016). AMAS: a fast tool for alignment manipulation and computing of summary statistics. *PeerJ*, *4*, e1660. https://doi.org/10.7717/peerj.1660

Bravo, G. A., Antonelli, A., Bacon, C. D., Bartoszek, K., Blom, M. P. K., Huynh, S., Jones, G., Knowles, L. L., Lamichhaney, S., Marcussen, T., Morlon, H., Nakhleh, L. K., Oxelman, B., Pfeil, B., Schliep, A., Wahlberg, N., Werneck, F. P., Wiedenhoeft, J., Willows-Munro, S., & Edwards, S. V. (2019). Embracing heterogeneity: coalescing the Tree of Life and the future of phylogenomics. *PeerJ*, *7*, e6399. https://doi.org/10.7717/peerj.6399

Brown, J. L., Cameron, A., Yoder, A. D., & Vences, M. (2014). A necessarily complex model to explain the biogeography of the amphibians and reptiles of Madagascar. *Nature Communications*, *5*(1), 5046. https://doi.org/10.1038/ncomms6046

Brown, J. M., & Thomson, R. C. (2017). Bayes Factors Unmask Highly Variable Information Content, Bias, and Extreme Influence in Phylogenomic Analyses. *Systematic Biology*, *66*(4), 517–530. https://doi.org/10.1093/sysbio/syw101

Brown, J. W., Walker, J. F., & Smith, S. A. (2017). Phyx: phylogenetic tools for unix. *Bioinformatics*, *33*(12), 1886–1888. https://doi.org/10.1093/bioinformatics/btx063

Cai, R., & Ané, C. (2021). Assessing the fit of the multi-species network coalescent to multi-locus data. *Bioinformatics*, *37*(5), 634–641. https://doi.org/10.1093/bioinformatics/btaa863

Callmander, M. W., Phillipson, P. B., Schatz, G. E., Andriambololonera, S., Rabarimanarivo, M., Rakotonirina, N., Raharimampionona, J., Chatelain, C., Gautier, L., & Lowry, P. P. (2011). The endemic and non-endemic vascular flora of Madagascar updated. *Plant Ecology and Evolution*, *144*(2), 121–125. https://doi.org/10.5091/plecevo.2011.513

Capella-Gutiérrez, S., Silla-Martínez, J. M., & Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, *25*(15), 1972–1973. https://doi.org/10.1093/bioinformatics/btp348

Chifman, J., & Kubatko, L. (2014). Quartet inference from SNP data under the coalescent model. *Bioinformatics*, *30*(23), 3317–3324. https://doi.org/10.1093/bioinformatics/btu530

Degnan, J. H., & Rosenberg, N. A. (2006). Discordance of species trees with their most likely gene trees. *PLoS Genetics*, *2*(5), e68. https://doi.org/10.1371/journal.pgen.0020068

dos Reis, M., Gunnell, G. F., Barba-Montoya, J., Wilkins, A., Yang, Z., & Yoder, A. D. (2018). Using phylogenomic data to explore the effects of relaxed clocks and calibration strategies on divergence time estimation; primates as a test case. *Systematic Biology*, *67*(4), 594–615. https://doi.org/10.1093/sysbio/syy001

Dunning, L. T., Olofsson, J. K., Parisod, C., Choudhury, R. R., Moreno-Villena, J. J., Yang, Y., Dionora, J., Quick, W. P., Park, M., Bennetzen, J. L., Besnard, G., Nosil, P., Osborne, C. P., & Christin, P.-A. (2019). Lateral transfers of large DNA fragments spread functional genes among grasses. *Proceedings of the National Academy of Sciences*, *116*(10), 4416–4425. https://doi.org/10.1073/pnas.1810031116

Everson, K. M., Pozzi, L., Barrett, M. A., Blair, M. E., Donohue, M. E., Kappeler, P. M., Kitchener, A. C., Lemmon, A. R., Lemmon, E. M., Pavón-Vázquez, C. J., Radespiel, U., Randrianambinina, B., Rasoloarison, R. M., Rasoloharijaona, S., Roos, C., Salmona, J., Yoder, A. D., Zenil-Ferguson, R., Zinner, D., & Weisrock, D. W. (2023). Not one, but multiple radiations underlie the biodiversity of Madagascar's endangered lemurs. *bioRxiv*. https://doi.org/10.1101/2023.04.26.537867

Felsenstein, J. (1981). Evolutionary trees from DNA sequences: A maximum likelihood approach. *Journal of Molecular Evolution*, *17*(6), 368–376. https://doi.org/10.1007/bf01734359

Flouri, T., Jiao, X., Rannala, B., & Yang, Z. (2018). Species tree inference with BPP using genomic sequences and the multipspecies coalescent. *Molecular Biology and Evolution*, *35*(10), 2585–2593. https://doi.org/10.1093/molbev/msy147

Flouri, T., Jiao, X., Rannala, B., & Yang, Z. (2020). A Bayesian implementation of the multispecies coalescent model with introgression for phylogenomic analysis. *Molecular Biology and Evolution*, *37*(4), 1211–1223. https://doi.org/10.1093/molbev/msz296

Hasiniaina, A. F., Radespiel, U., Kessler, S. E., Evasoa, M. R., Rasoloharijaona, S., Randrianambinina, B., Zimmermann, E., Schmidt, S., & Scheumann, M. (2020). Evolutionary significance of the variation in acoustic communication of a cryptic nocturnal primate radiation (*Microcebus* spp.). *Ecology and Evolution*, *10*(8), 3784–3797. https://doi.org/10.1002/ece3.6177

Heckman, K. L., Mariani, C. L., Rasoloarison, R., & Yoder, A. D. (2007). Multiple nuclear loci reveal patterns of incomplete lineage sorting and complex species history within western mouse lemurs (Microcebus). *Molecular Phylogenetics and Evolution*, *43*(2), 353–367. https://doi.org/10.1016/j.ympev.2007.03.005

Herrera, J. P., & Dávalos, L. M. (2016). Phylogeny and divergence times of lemurs inferred with recent and ancient fossils in the tree. *Systematic Biology*, *65*(5), 772–791. https://doi.org/10.1093/sysbio/syw035

Hibbins, M. S., & Hahn, M. W. (2022). Phylogenomic approaches to detecting and characterizing introgression. *Genetics*, *220*(2), 173. https://doi.org/10.1093/genetics/iyab173

Hotaling, S., Foley, M. E., Lawrence, N. M., Bocanegra, J., Blanco, M. B., Rasoloarison, R., Kappeler, P. M., Barrett, M. A., Yoder, A. D., & Weisrock, D. W. (2016). Species discovery and validation in a cryptic radiation of endangered primates: coalescent-based species delimitation in Madagascar's mouse lemurs. *Molecular Ecology*, *25*(9), 2029–2045. https://doi.org/10.1111/mec.13604

Hudson, R. R. (1983). Testing the constant-rate neutral allele model with protein sequence data. *Evolution*, *37*(1), 203–217. https://doi.org/10.2307/2408186

Hunnicutt, K. E., Tiley, G. P., Williams, R. C., Larsen, P. A., Blanco, M. B., Rasoloarison, R. M., Campbell, C. R., Zhu, K., Weisrock, D. W., Matsunami, H., & Yoder, A. D. (2020). Comparative genomic analysis of the pheromone receptor class 1 family (V1R) reveals extreme complexity in mouse lemurs (Genus, Microcebus) and a chromosomal hotspot across mammals. *Genome Biology and Evolution*, *12*(1), 3562–3579. https://doi.org/10.1093/gbe/evz200

Jones, M. R., Mills, L. S., Alves, P. C., Callahan, C. M., Alves, J. M., Lafferty, D. J. R., Jiggins, F. M., Jensen, J. D., Melo-Ferreira, J., & Good, J. M. (2018). Adaptive introgression underlies polymorphic seasonal camouflage in snowshoe hares. *Science*, *360*(6935), 1355–1358. https://doi.org/10.1126/science.aar5273

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods*, *14*(6), 587–589. https://doi.org/10.1038/nmeth.4285

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alingnment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, *30*(4), 772–780. https://doi.org/10.1093/molbev/mst010

Kishino, H., & Hasegawa, M. (1989). Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. *Journal of Molecular Evolution*, *29*(2), 170–179. https://doi.org/10.1007/bf02100115

Kishino, H., Miyata, T., & Hasegawa, M. (1990). Maximum likelihood inference of protein phylogeny and the origin of chloroplasts. *Journal of Molecular Evolution*, *31*(2), 151–160. https://doi.org/10.1007/bf02109483

Kubatko, L. S., & Degnan, J. H. (2007). Inconsistency of phylogenetic estimates from concatenated data under coalescence. *Systematic Biology*, *56*(1), 17–24. https://doi.org/10.1080/10635150601146041

Larget, B. R., Kotha, S. K., Dewey, C. N., & Ané, C. (2010). BUCKy: Gene tree/species tree reconciliation with Bayesian concordance analysis. *Bioinformatics*, *26*(22), 2910–2911. https://doi.org/10.1093/bioinformatics/btq539

Larsen, P. A., Harris, R. A., Liu, Y., Murali, S. C., Campbell, C. R., Brown, A. D., Sullivan, B. A., Shelton, J., Brown, S. J., Raveendran, M., Dudchenko, O., Machol, I., Durand, N. C., Shamim, M. S., Lieberman Aiden, E., Muzny, D. M., Gibbs, R. A., Yoder, A. D., Rogers, J., & Worley, K. C. (2017). Hybrid de novo genome assembly and centromere characterization of the gray mouse lemur (Microcebus murinus). *BMC Biology*, *15*(1), 110. https://doi.org/10.1186/s12915-017-0439-6

Li, H.-T., Luo, Y., Gan, L., Ma, P.-F., Gao, L.-M., Yang, J.-B., Cai, J., Gitzendanner, M. A., Fritsch, P. W., Zhang, T., Jin, J.-J., Zeng, C.-X., Wang, H., Yu, W.-B., Zhang, R., van der Bank, M., Olmstead, R. G., Hollingsworth, P. M., Chase, M. W., … Li, D.-Z. (2021). Plastid phylogenomic insights into relationships of all flowering plant families. *BMC Biology*, *19*(1), 232. https://doi.org/10.1186/s12915-021-01166-2

Louis, E. E., Jr., Coles, M. S., Andriantompohavana, R., Sommer, J. A., Engberg, S. E., Zaonarivelo, J. R., Mayor, M. I., & Brenneman, R. A. (2006). Revision of the mouse lemurs (Microcebus) of eastern Madagascar. *International Journal of Primatology*, *27*(2), 347–389. https://doi.org/10.1007/s10764-006-9036-1

Maddison, W. P. (1997). Gene Trees in Species Trees. *Systematic Biology*, *46*(3), 523–536. https://doi.org/10.1093/sysbio/46.3.523

Mai, U., & Mirarab, S. (2018). Treeshrink: fast and accurate detection of outlier long branches in collections of phylogenetic trees. *BMC Genomics*, *19*(5), 23–40. https://doi.org/10.1186/s12864-018-4620-2

Malinsky, M., Matschiner, M., & Svardal, H. (2021). Dsuite - Fast *D*-statistics and related admixture evidence from VCF files. *Molecular Ecology Resources*, *21*(2), 584–595. https://doi.org/10.1111/1755-0998.13265

Malinsky, M., Svardal, H., Tyers, A. M., Miska, E. A., Genner, M. J., Turner, G. F., & Durbin, R. (2018). Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. *Nature Ecology & Evolution*, *2*(12), 1940–1955. https://doi.org/10.1038/s41559-018-0717-x

Mallet, J. (2005). Hybridization as an invasion of the genome. *Trends in Ecolology and Evolution*, *20*(5), 229–237. https://doi.org/10.1016/j.tree.2005.02.010

Matsumotot, K., & Burney, D. A. (1994). Late Holocene environments at Lake Mitsinjo, northwestern Madagascar. *Holocene*, *4*(1), 16–24. https://doi.org/10.1177/095968369400400103

Meier, J. I., Marques, D. A., Mwaiko, S., Wagner, C. E., Excoffier, L., & Seehausen, O. (2017). Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nature Communications*, *8*(1), 14363. https://doi.org/10.1038/ncomms14363

Meng, C., & Kubatko, L. S. (2009). Detecting hybrid speciation in the presence of incomplete lineage sorting using gene tree incongruence: A model. *Theoretical Population Biology*, *75*(1), 35–45. https://doi.org/10.1016/j.tpb.2008.10.004

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., & Lanfear, R. (2020). IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution*, *37*(5), 1530–1534. https://doi.org/10.1093/molbev/msaa015

Mo, Y. K., Lanfear, R., Hahn, M. W., & Minh, B. Q. (2023). Updated site concordance factors minimize the effects of homoplasy and taxon sampling. *Bioinformatics*, *39*(1), 741. https://doi.org/10.1093/bioinformatics/btac741

Olivieri, G., Zimmermann, E., Randrianambinina, B., Rasoloharijaona, S., Rakotondravony, D., Guschanski, K., & Radespiel, U. (2007). The ever-increasing diversity in mouse lemurs: Three new species in north and northwestern Madagascar. *Molecular Phylogenetics and Evolution*, *43*(1), 309–327. https://doi.org/10.1016/j.ympev.2006.10.026

Page, A. J., Taylor, B., Delaney, A. J., Soares, J., Seemann, T., Keane, J. A., & Harris, S. R. (2016). SNP-sites: Rapid efficient extraction of SNPs from multi-FASTA alignments. *Microbial Genomics*, *2*(4), e000056. https://doi.org/10.1099/mgen.0.000056

Pamilo, P., & Nei, M. (1988). Relationships between gene trees and species trees. *Molecular Biology and Evolution*, *5*(5), 568–583. https://doi.org/10.1093/oxfordjournals.molbev.a040517

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., & Reich, D. (2012). Ancient Admixture in Human History. *Genetics*, *192*(3), 1065–1093. https://doi.org/10.1534/genetics.112.145037

Peng, J., Swofford, D. L., & Kubatko, L. (2022). Estimation of speciation times under the multispecies coalescent. *Bioinformatics*, *38*(23), 5182–5190. https://doi.org/10.1093/bioinformatics/btac679

Poelstra, J. W., Montero, B. K., Lüdemann, J., Yang, Z., Rakotondranary, S. J., Hohenlohe, P., Stetter, N., Ganzhorn, J. U., & Yoder, A. D. (2022). RADseq data reveal a lack of admixture in a mouse lemur contact zone contrary to previous microsatellite results. *Proceedings of the Royal Society B*, *289*(1980), 20220596. https://doi.org/10.1098/rspb.2022.0596

Poelstra, J. W., Salmona, J., Tiley, G. P., Schüßler, D., Blanco, M. B., Andriambeloson, J. B., Bouchez, O., Campbell, C. R., Etter, P. D., Hohenlohe, P. A., Hunnicutt, K. E., Iribar, A., Johnson, E. A., Kappeler, P. M., Larsen, P. A., Manzi, S., Ralison, J. M., Randrianambinina, B., Rasoloarison, R. M., … Yoder, A. D. (2020). Cryptic patterns of speciation in cryptic primates: microendemic mouse lemurs and the multispecies coalescent. *Systematic Biology*, *70*(2), 203–218. https://doi.org/10.1093/sysbio/syaa053

Radespiel, U., Rakotondravony, R., & Chikhi, L. (2008). Natural and anthropogenic determinants of genetic structure in the largest remaining population of the endangered golden-brown mouse lemur, *Microcebus ravelobensis*. *American Journal of Primatology*, *70*(9), 860–870. https://doi.org/10.1002/ajp.20574

Railsback, L. B., Dupont, L. A., Liang, F., Brook, G. A., Burney, D. A., Cheng, H., & Edwards, R. L. (2020). Relationships between climate change, human environmental impact, and megafaunal extinction inferred from a 4000-year multi-proxy record from a stalagmite from northwestern Madagascar. *Quaternary Science Reviews*, *234*, 106244. https://doi.org/10.1016/j.quascirev.2020.106244

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology*, *67*(5), 901–904. https://doi.org/10.1093/sysbio/syy032

Rannala, B., & Yang, Z. (2003). Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. *Genetics*, *164*(4), 1645–1656. https://doi.org/10.1093/genetics/164.4.1645

Rannala, B., & Yang, Z. (2017). Efficient Bayesian species tree inference under the multispecies coalescent. *Systematic Biology*, *66*(5), 823–842. https://doi.org/10.1093/sysbio/syw119

Rasoloarison, R. M., Weisrock, D. W., Yoder, A. D., Rakotondravony, D., & Kappeler, P. M. (2013). Two New Species of Mouse Lemurs (Cheirogaleidae: Microcebus) from Eastern Madagascar. *International Journal of Primatology*, *34*(3), 455–469. https://doi.org/10.1007/s10764-013-9672-1

Rieseberg, L. H. (1997). Hybrid origins of plant species. *Annual Review of Ecology and Systematics*, *28*(1), 359–389. https://doi.org/10.1146/annurev.ecolsys.28.1.359

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M. A., & Huelsenbeck, J. P. (2012). MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, *61*(3), 539–542. https://doi.org/10.1093/sysbio/sys029

Rosenberg, N. A. (2002). The probability of topological concordance of gene trees and species trees. *Theoretical Population Biology*, *61*(2), 225–247. https://doi.org/10.1006/tpbi.2001.1568

Sayyari, E., Whitfield, J. B., & Mirarab, S. (2018). DiscoVista: interpretable visualizations of gene tree discordance. *Molecular Phylogenetics and Evolution*, *122*, 110–115. https://doi.org/10.1016/j.ympev.2018.01.019

Schneider, N., Chikhi, L., Currat, M., & Radespiel, U. (2010). Signals of recent spatial expansions in the grey mouse lemur (Microcebus murinus). *BMC Evolutionary Biology*, *10*(105). https://doi.org/10.1186/1471-2148-10-105

Shimodaira, H. (2002). An Approximately Unbiased Test of Phylogenetic Tree Selection. *Systematic Biology*, *51*(3), 492–508. https://doi.org/10.1080/10635150290069913

Shimodaira, H., & Hasegawa, M. (1999). Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution*, *16*(8), 1114–1116. https://doi.org/10.1093/oxfordjournals.molbev.a026201

Simon, C. (2022). An Evolving View of Phylogenetic Support. *Systematic Biology*, *71*(4), 921–928. https://doi.org/10.1093/sysbio/syaa068

Solís-Lemus, C., & Ané, C. (2016). Inferring phylogenetic networks with maximum pseudoliklihood under incomplete lineage sorting. *PLoS Genetics*, *12*(3), e1005896. https://doi.org/10.1371/journal.pgen.1005896

Springer, M. S., Meredith, R. W., Gatesy, J., Emerling, C. A., Park, J., Rabosky, D. L., Stadler, T., Steiner, C., Ryder, O. A., Janečka, J. E., Fisher, C. A., & Murphy, W. J. (2012). Macroevolutionary dynamics and historical biogeography of primate diversification inferred from a species supermatrix. *PLoS One*, *7*(11), e49521. https://doi.org/10.1371/journal.pone.0049521

Strimmer, K., & Rambaut, A. (2002). Inferring confidence sets of possibly misspecified gene trees. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *269*(1487), 137–142. https://doi.org/10.1098/rspb.2001.1862

Stull, G. W., Soltis, P. S., Soltis, D. E., Gitzendanner, M. A., & Smith, S. A. (2020). Nuclear phylogenomic analyses of asterids conflict with plastome trees and support novel relationships among major lineages. *American Journal of Botany*, *107*(5), 790–805. https://doi.org/10.1002/ajb2.1468

Swofford, D. L. (2002). *PAUP: Phylogenetic Analysis Using Parsimony (and Other Methods)*. Sinauer Associates.

Tattersall, I. (2007). Madagascar's Lemurs: Cryptic diversity or taxonomic inflation? *Evolutionary Anthropology: Issues, News, and Reviews*, *16*(1), 12–23. https://doi.org/10.1002/evan.20126

Thomson, R. C., & Brown, J. M. (2022). On the Need for New Measures of Phylogenomic Support. *Systematic Biology*, *71*(4), 917–920. https://doi.org/10.1093/sysbio/syac002

Tiley, G. P., Poelstra, J. W., dos Reis, M., Yang, Z., & Yoder, A. D. (2020). Molecular Clocks without Rocks: New Solutions for Old Problems. *Trends in Genetics*, *36*(11), 845–856. https://doi.org/10.1016/j.tig.2020.06.002

Tiley, G. P., van Elst, T., Teixeira, H., Schüßler, D., Salmona, J., Blanco, M. B., Ralison, J. M., Randrianambinina, B., Rasoloarison, R. M., Stahlke, A. R., Hohenlohe, P. A., Chikhi, L., Louis, E. E., Radespiel, U., & Yoder, A. D. (2022). Population genomic structure in Goodman's mouse lemur reveals long-standing separation of Madagascar's Central Highlands and eastern rainforests. *Molecular Ecology*, *31*(19), 4901–4918. https://doi.org/10.1111/mec.16632

Vanderpool, D., Minh, B. Q., Lanfear, R., Hughes, D., Murali, S., Harris, R. A., Raveendran, M., Muzny, D. M., Hibbins, M. S., Williamson, R. J., Gibbs, R. A., Worley, K. C., Rogers, J., & Hahn, M. W. (2020). Primate phylogenomics uncovers multiple rapid radiations and ancient interspecific introgression. *PLoS Biology*, *18*(12), e3000954. https://doi.org/10.1371/journal.pbio.3000954

Vences, M., Wollenberg, K. C., Vieites, D. R., & Lees, D. C. (2009). Madagascar as a model region of species diversification. *Trends in Ecology & Evolution*, *24*(8), 456–465. https://doi.org/10.1016/j.tree.2009.03.011

Weisrock, D. W., Rasoloarison, R. M., Fiorentino, I., Ralison, J. M., Goodman, S. M., Kappeler, P. M., & Yoder, A. D. (2010). Delimiting species without nuclear monophyly in Madagascar's mouse lemurs. *PLoS One*, *5*(3), e9883. https://doi.org/10.1371/journal.pone.0009883

Weisrock, D. W., Smith, S. D., Chan, L. M., Biebouw, K., Kappeler, P. M., & Yoder, A. D. (2012). Concatentation and concordance in the reconstruction of mouse lemur phylogeny: An empirical demonstration of the effect of allele sampling in phylogenetics. *Molecular Biology and Evolution*, *29*(6), 1615–1630. https://doi.org/10.1093/molbev/mss008

Wells, N. A. (2003). *Some hypotheses on the Mesozoic and Cenozoic paleoenvironmental history of Madagascar* (G. S & B. J, Eds.; pp. 16–34). The University of Chicago Press.

Xie, W., Lewis, P. O., Fan, Y., Kuo, L., & Chen, M.-H. (2010). Improving marginal likelihood estimation for Bayesian phylogenetic model selection. *Systematic Biology*, *60*(2), 150–160. https://doi.org/10.1093/sysbio/syq085

Yang, Y., & Smith, S. A. (2014). Othology inference in nonmodel organisms using transcriptomes and low-coverage genomes: improving accuracy and matrix occupancy for phylogenomics. *Molecular Biology and Evolution*, *31*(11), 3081–3092. https://doi.org/10.1093/molbev/msu245

Yoder, A. D., Campbell, C. R., Blanco, M. B., dos Reis, M., Ganzhorn, J. U., Goodman, S. M., Hunnicutt, K. E., Larsen, P. A., Kappeler, P. M., Rasoloarison, R. M., Ralison, J. M., Swofford, D. L., & Weisrock, D. W. (2016). Geographic patterns in mouse lemurs (genus *Microcebus*) reveal the ghosts of Madagascar's forests past. *Proceedings of the National Academy of Sciences*, *113*(29), 8049–8056. https://doi.org/10.1073/pnas.1601081113

Yoder, A. D., Rasoloarison, R. M., Goodman, S. M., Irwin, J. A., Atsalis, S., Ravosa, M. J., & Ganzhorn, J. U. (2000). Remarkable species diversity in Malagasy mouse lemurs (primates, *Microcebus*). *Proceedings of the National Academy of Sciences*, *97*(21), 11325–11330. https://doi.org/10.1073/pnas.200121897

Yu, Y., Degnan, J. H., & Nakhleh, L. (2012). The probability of a gene tree topology within a phylogenetic network with applications to hybridization detection. *PLoS Genetics*, *8*(4), e1002660. https://doi.org/10.1371/journal.pgen.1002660

Zhang, C., Rabiee, M., Sayyari, E., & Mirarab, S. (2018). ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics*, *19*(153), 15–30. https://doi.org/10.1186/s12859-018-2129-y

Zhou, B.-F., Yuan, S., Crowl, A. A., Liang, Y.-Y., Shi, Y., Chen, X.-Y., An, Q.-Q., Kang, M., Manos, P. S., & Wang, B. (2022). Phylogenomic analyses highlight innovation and introgression in the continental radiations of Fagaceae across the Northern Hemisphere. *Nature Communications*, *13*(1320), 1–14. https://doi.org/10.1038/s41467-022-28917-1

Zimmermann, E., & Radespiel, U. (2014). Species concepts, diversity, and evolution in primates: lessons to be learned from mouse lemurs. *Evolutionary Anthropology: Issues, News, and Reviews*, *23*(1), 11–14. https://doi.org/10.1002/evan.21388

Supplementary Materials

Supplementary Information for: Ancient introgression in mouse lemurs (*Microcebus*: Cheirogaleidae) explains 20 years of phylogenetic uncertainty

Download: https://ssbbulletin.scholasticahq.com/article/94553-ancient-introgression-in-mouse-lemurs-_microcebus_-cheirogaleidae-explains-20-years-of-phylogenetic-uncertainty/attachment/198738.docx?auth_token=SS6_6VQX6pwwWLZQmf-6