

## RESEARCH ARTICLE

## Primate phylogenomics uncovers multiple rapid radiations and ancient interspecific introgression

Dan Vanderpool<sup>1\*</sup>, Bui Quang Minh<sup>2,3</sup>, Robert Lanfear<sup>3</sup>, Daniel Hughes<sup>4</sup>, Shwetha Murali<sup>4</sup>, R. Alan Harris<sup>4,5</sup>, Muthuswamy Raveendran<sup>4</sup>, Donna M. Muzny<sup>4,5</sup>, Mark S. Hibbins<sup>1</sup>, Robert J. Williamson<sup>6</sup>, Richard A. Gibbs<sup>4,5</sup>, Kim C. Worley<sup>4,5</sup>, Jeffrey Rogers<sup>4,5</sup>, Matthew W. Hahn<sup>1</sup>

**1** Department of Biology and Department of Computer Science, Indiana University, Bloomington, Indiana, United States of America, **2** Research School of Computer Science, Australian National University, Canberra, Australian Capital Territory, Australia, **3** Research School of Biology, Australian National University, Canberra, Australian Capital Territory, Australia, **4** Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas, United States of America, **5** Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas, United States of America, **6** Department of Computer Science and Software Engineering and Department of Biology and Biomedical Engineering, Rose-Hulman Institute of Technology, Terre Haute, Indiana, United States of America

\* [danvand@indiana.edu](mailto:danvand@indiana.edu)



## OPEN ACCESS

**Citation:** Vanderpool D, Minh BQ, Lanfear R, Hughes D, Murali S, Harris RA, et al. (2020) Primate phylogenomics uncovers multiple rapid radiations and ancient interspecific introgression. *PLoS Biol* 18(12): e3000954. <https://doi.org/10.1371/journal.pbio.3000954>

**Academic Editor:** Chris D. Jiggins, University of Cambridge, UNITED KINGDOM

**Received:** April 15, 2020

**Accepted:** November 2, 2020

**Published:** December 3, 2020

**Peer Review History:** PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pbio.3000954>

**Copyright:** © 2020 Vanderpool et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** The relevant assembly accessions and associated references used in this study are provided in [S1 Table](#). All raw data, assemblies, and annotation information used

## Abstract

Our understanding of the evolutionary history of primates is undergoing continual revision due to ongoing genome sequencing efforts. Bolstered by growing fossil evidence, these data have led to increased acceptance of once controversial hypotheses regarding phylogenetic relationships, hybridization and introgression, and the biogeographical history of primate groups. Among these findings is a pattern of recent introgression between species within all major primate groups examined to date, though little is known about introgression deeper in time. To address this and other phylogenetic questions, here, we present new reference genome assemblies for 3 Old World monkey (OWM) species: *Colobus angolensis* ssp. *palliatu*s (the black and white colobus), *Macaca nemestrina* (southern pig-tailed macaque), and *Mandrillus leucophaeus* (the drill). We combine these data with 23 additional primate genomes to estimate both the species tree and individual gene trees using thousands of loci. While our species tree is largely consistent with previous phylogenetic hypotheses, the gene trees reveal high levels of genealogical discordance associated with multiple primate radiations. We use strongly asymmetric patterns of gene tree discordance around specific branches to identify multiple instances of introgression between ancestral primate lineages. In addition, we exploit recent fossil evidence to perform fossil-calibrated molecular dating analyses across the tree. Taken together, our genome-wide data help to resolve multiple contentious sets of relationships among primates, while also providing insight into the biological processes and technical artifacts that led to the disagreements in the first place.

in these analyses are available through each species' NCBI BioProject link available in the relevant assembly accessions and associated references used in this study are provided in [S1 Table](#). The Data Dryad repository associated with this study can be accessed through the following link: <https://doi.org/10.5061/dryad.rfj6q577d> [22]. The repository contains the following files and archives: • 1730\_ALIGNMENT\_CONCAT.paup.nex – Concatenated alignment with PAUP block of commands used to generate the parsimony concatenated tree. • 1730\_Alignments\_FINAL.tar.gz: 1,730 single-copy ortholog alignments. • 1730\_ML\_GENETREEs.treefile: Maximum likelihood gene trees estimated from the 1,730 ortholog alignments. • ASTRAL\_Tree\_AVGdates.tre: The ASTRAL topology ([Fig 1](#)) with average dates from 10 independent datasets. • All\_Dating\_Datasets\_DRYAD.tar.gz: The concatenated alignments used for dating analyses. • PARSIMONY\_1730\_Gene\_Trees.tre: All 1,730 parsimony gene trees from MPBoot. • Supp\_fig4A\_F\_s6\_b1\_v2\_1.slim-SLiM3 recipe for [S4A Fig](#) simulation • Supp\_fig4B\_F\_s6\_b1\_v2\_highmut\_1.slim-SLiM3 recipe for [S4B Fig](#) simulation • All\_1735\_UNALIGNED\_Seqs.tar.gz: All unaligned single-copy gene sequences. • plot\_DATING.R: R script used for plotting [Fig 3](#).

**Funding:** Funding for this study was provided by grants from the National Science Foundation, grant numbers: DBI-1564611 and DEB-1936187 awarded to M.W.H. Salary was provided to D.V. and M.W.H. by grant number: DBI-1564611. Additional salary was provided to M.W.H. by grant number: DEB-1936187. The authors received no specific funding for this work. Additional funding was provided by the Chan-Zuckerberg Initiative grant for Essential Open Source Software for Science (<https://chanzuckerberg.com/eoss/>) awarded to B.Q.M. and R.L. The authors received no specific funding for this work. Additional funding was provided by the Australian Research Council under grant number: DP-200103151 awarded to R.L., B.Q.M., and M.W.H. The authors received no specific funding for this work. Additional funding was provided by a Australian National University (<https://www.anu.edu.au/>) Futures grant awarded to R.L., which paid salary for B.Q.M. The sequencing and assembly of the colobus, pig-tailed macaque, and drill genomes was funded by National Institutes of Health grant number: U54-HG006484 awarded to R.G. Salaries from this award were received by R.G., J.R., K.W., S.M., D.H., and D.M. The funders had no role in

## Introduction

Understanding the history of individual genes and whole genomes is an important goal for evolutionary biology. It is only by understanding these histories that we can understand the origin and evolution of traits—whether morphological, behavioral, or biochemical. Until recently, our ability to address the history of genes and genomes was limited by the availability of comparative genomic data. However, genome sequences are now being generated extremely rapidly. In primates alone, there are already 23 species with published reference genome sequences and associated annotations ([S1 Table](#)), as well as multiple species with population samples of whole genomes [1–11]. These data can now be used to address important evolutionary questions.

Several studies employing dozens of loci sampled across broad taxonomic groups have provided rough outlines of the evolutionary relationships and divergence times among primates [12,13]. Due to the rapid nature of several independent radiations within primates, these limited data cannot resolve species relationships within some clades [12–14]. For instance, the New World monkeys (NWM) experienced a rapid period of diversification approximately 15 to 18 million years ago (mya) [15] ([Fig 1](#)), resulting in ambiguous relationships among the 3 Cebidae subfamilies (Cebinae = squirrel monkeys and capuchins, Aotinae = owl monkeys, and Callitrichinae = marmosets and tamarins) [12–14,16–18]. High levels of incomplete lineage sorting (ILS) driven by short times between the divergence of distinct lineages have led to a large amount of gene tree discordance in the NWM, with different loci favoring differing relationships among taxa. Given the known difficulties associated with resolving short internodes [19–21], as well as the multiple different approaches and datasets used in these analyses, the relationships among cebid subfamilies remain uncertain.

In addition to issues of limited data and rapid radiations, a history of hybridization and subsequent gene flow between taxa means that there is no single dichotomously branching tree that all genes follow. Although introgression once was thought to be relatively rare (especially among animals [23]), genomic studies have uncovered widespread patterns of recent introgression across the tree of life [24]. Evidence for recent or ongoing gene flow is especially common among the primates (e.g., [9,25–27]), sometimes with clear evidence for adaptive introgression (e.g., [28–30]). Whether widespread gene flow among primates is emblematic of their initial radiation (which began 60 to 75 mya [13,31–33]) or is a consequence of current conditions—which include higher environmental occupancy and more secondary contact—remains an open question [34].

Here, we report the sequencing and annotation of 3 new primate genomes, all Old World monkey (OWM) species: *Colobus angolensis* ssp. *palliatu*s (the black and white colobus), *Macaca nemestrina* (southern pig-tailed macaque), and *Mandrillus leucophaeus* (the drill). Together with the published whole genomes of extant primates, we present a phylogenomic analysis including 26 primate species and several closely related non-primates. Incorporating recently discovered fossil evidence [35], we perform fossil-calibrated molecular dating analyses to estimate divergence times, including dates for the crown primates as well as the timing of more recent splits. Compared to recent hybridization, introgression that occurred between 2 or more ancestral lineages (represented by internal branches on a phylogeny) is difficult to detect. To get around this limitation, we modify a previously proposed method for detecting introgression [36] and apply it to our whole-genome datasets, finding additional evidence for gene flow among ancestral primates. Finally, we closely examine the genealogical patterns left behind by the NWM radiation, as well as the biases of several methods that have been used to resolve this topology. We use multiple approaches to provide a strongly supported history of

study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

**Abbreviations:** BUSCO, Benchmarking Universal Single-Copy Orthologs; CDS, coding sequences; gCF, gene concordance factor; HGCS, Human Genome Sequencing Center; ILS, incomplete lineage sorting; ML-CONCAT, maximum likelihood concatenated; my, million years; mya, million years ago; NWM, New World monkey; OWM, Old World monkey; PAUP\*, Phylogenetic Analysis Using Parsimony\*; sCF, site concordance factor; SRA, Short Read Archive.

the NWM and primates in general, while also highlighting the large amounts of gene tree discordance across the tree caused by ILS and introgression.

## Results and discussion

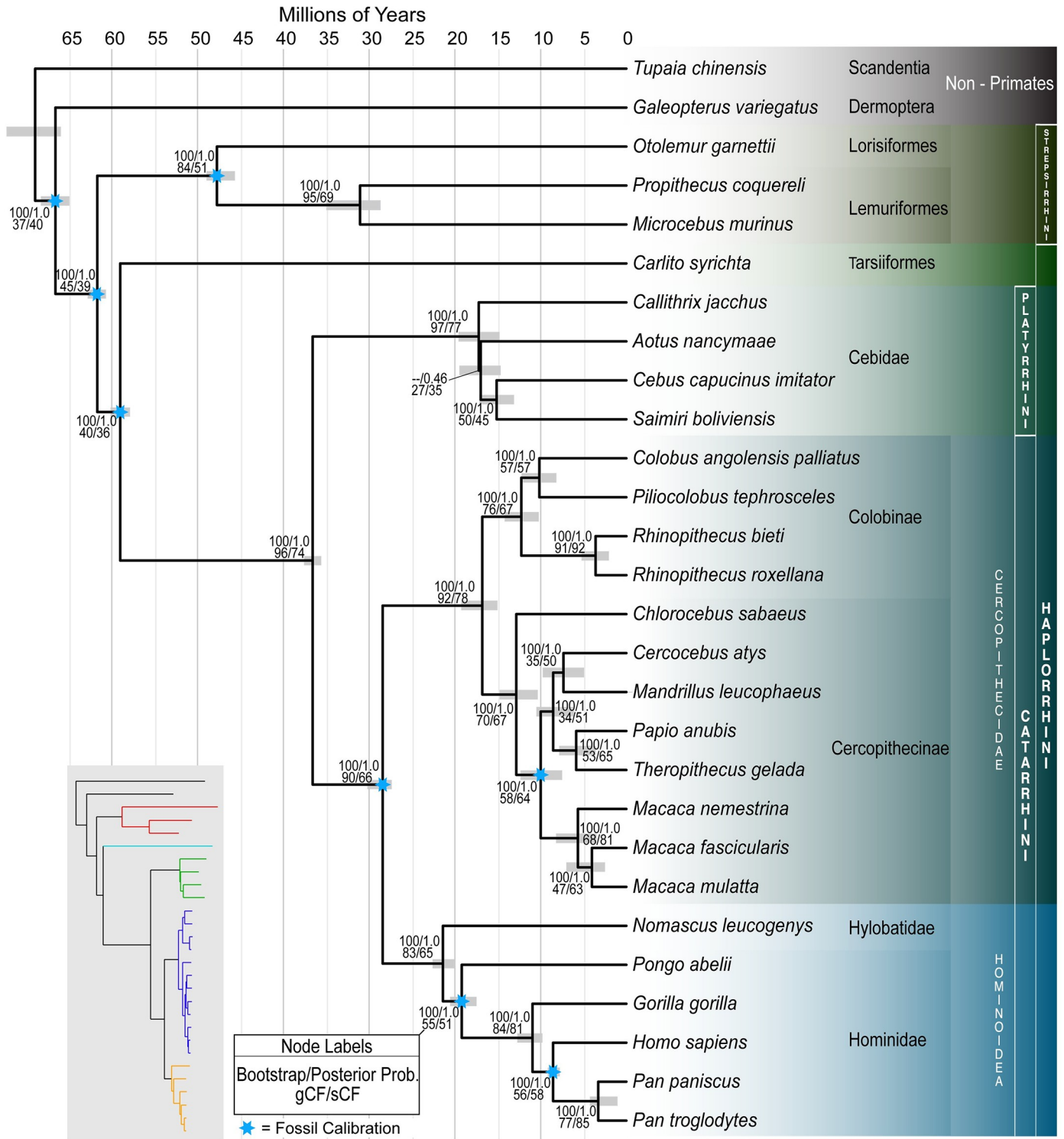
### Primate genome sequencing

The 3 species sequenced here are all OWMs, and each is closely related to an already-sequenced species. This sampling scheme provides us increased power to detect introgression among each of the sub-clades containing these species. The assembly and annotation of each of the 3 species sequenced for this project are summarized here, with further details listed in [Table 1](#). A summary of all published genomes used in this study, including links to the assemblies and NCBI BioProjects, is available in [S2 Table](#). All species were sequenced using standard methods according to Illumina (San Diego, California, United States of America) Hi-seq protocols. Additional long-read sequencing was performed using Pacific Biosciences (Menlo Park, California, USA) technology for *M. nemestrina*.

The sequencing effort for *C. angolensis* ssp. *palliatus* produced 514 Gb of data, which are available in the NCBI Short Read Archive (SRA) under the accession SRP050426 (BioProject PRJNA251421). The biological sample used for sequencing was kindly provided by Dr. Oliver Ryder (San Diego Zoo). Assembly of these data resulted in a total assembly length of 2.97 Gb in 13,124 scaffolds (NCBI assembly Cang.pa\_1.0; GenBank accession GCF\_000951035.1) with an average per base coverage of 86.8X. Subsequent annotation via the NCBI Eukaryotic Genome Annotation Pipeline (annotation release ID: 100) resulted in the identification of 20,222 protein-coding genes and 2,244 noncoding genes. An assessment of the annotation performed using Benchmarking Universal Single-Copy Orthologs (BUSCO) 3.0.2 [37] in conjunction with the Euarchontoglires ortholog database 9 ([https://busco-archive.ezlab.org/v3/datasets/euarchontoglires\\_odb9.tar.gz](https://busco-archive.ezlab.org/v3/datasets/euarchontoglires_odb9.tar.gz)) indicated that 95.82% single-copy orthologs (91.68% complete and 4.13% fragmented) were present among the annotated protein-coding genes. Comprehensive annotation statistics for *C. angolensis* ssp. *palliatus* with links to the relevant annotation products available for download can be viewed at [https://www.ncbi.nlm.nih.gov/genome/annotation\\_euk/Colobus\\_angolensis\\_palliatus/100/](https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Colobus_angolensis_palliatus/100/).

For *M. nemestrina*, 1,271 Gb of data were produced (SRA accession SRP045960; BioProject PRJNA251427), resulting in an assembled genome length of 2.95 Gb in 9,733 scaffolds (Mnem\_1.0; GenBank accession GCF\_000956065.1). This corresponds to an average per base coverage of 113.1X when both short- and long-read data are combined (Materials and methods). The biological sample used for sequencing was kindly provided by Drs. Betsy Ferguson and James Ha (Washington National Primate Research Center). The NCBI annotation resulted in 21,017 protein-coding genes and 13,163 noncoding genes (annotation release ID: 101). A BUSCO run to assess the completeness of the annotation (as above) indicated that 95.98% single-copy orthologs (92.23% complete and 3.75% fragmented) were present among the annotated protein-coding genes. Comprehensive annotation statistics for *M. nemestrina* with links to the relevant annotation products available for download can be viewed at [https://www.ncbi.nlm.nih.gov/genome/annotation\\_euk/Macaca\\_nemestrina/101/](https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Macaca_nemestrina/101/).

Sequencing of *M. leucophaeus* libraries resulted in 334.1 Gb of data (SRA accession SRP050495; BioProject PRJNA251423) that once assembled resulted in a total assembly length of 3.06 Gb in 12,821 scaffolds (Mleu.le\_1.0; GenBank accession GCF\_000951045.1) with an average coverage of 117.2X per base. The biological sample used for sequencing was kindly provided by Dr. Oliver Ryder (San Diego Zoo). The NCBI annotation produced 20,465 protein-coding genes and 2,300 noncoding genes (annotation release ID: 100). A BUSCO run to assess the completeness of the annotation (as above) indicated that 95.45% single-copy



**Fig 1. Species tree estimated using ASTRAL III with 1,730 gene trees (the *Mus musculus* outgroup was removed to allow for a visually finer scale).** Common names for each species can be found in [S1 Table](#). Node labels indicate the bootstrap value from a maximum likelihood analysis of the concatenated dataset as well as the local posterior probability from the ASTRAL analysis. gCFs and sCFs are also reported. Eight fossil calibrations (blue stars; [S6 Table](#)) were used to calibrate node ages. Gray bars indicate the minimum and maximum mean age from independent dating estimates. The inset tree with colored branches shows the maximum likelihood branch lengths estimated using a partitioned analysis of the concatenated alignment. Colors correspond to red = Strepsirrhini, cyan = Tarsiiformes, green = Platyrrhini (NWMs), blue = Cercopithecoidea (OWMs), and orange = Hominoidea (Apes). All alignments used for phylogenomic analyses (1730\_Alignments\_FINAL.tar.gz) and dating analyses (All\_Dating\_Datasets\_DRYAD.tar.gz) are available via Data Dryad: <https://doi.org/10.5061/dryad.rfj6q574d> [22]. gCF, gene concordance factor; NWM, New World monkey; OWM, Old World monkey; sCF, site concordance factor.

<https://doi.org/10.1371/journal.pbio.3000954.g001>

**Table 1. Genomes sequenced in this study and associated assembly and annotation metrics.**

Species name	Assembly accession	Assembly total length	No. of scaffolds	Scaffold N50 (mb)	Contig N50 (kb)	Protein-coding genes	BUSCO
<i>Colobus angolensis</i> ssp. <i>palliatu</i> s (the black and white colobus)	GCF_000951035.1	2,970,124,662	13,124	7.84	38.36	20,222	95.82%
<i>Macaca nemestrina</i> (pig-tailed macaque)	GCF_000956065.1	2,948,703,511	9,733	15.22	106.89	21,017	95.98%
<i>Mandrillus leucophaeus</i> (drill)	GCF_000951045.1	3,061,992,840	12,821	3.19	31.35	20,465	95.45%

BUSCO percentages reflect the complete and fragmented genes relative to the Euarchontoglires ortholog database v9.

BUSCO, Benchmarking Universal Single-Copy Orthologs.

<https://doi.org/10.1371/journal.pbio.3000954.t001>

orthologs (91.38% complete, 4.07% fragmented) were present among the annotated protein-coding genes. The full annotation statistics with links to the associated data can be viewed at [https://www.ncbi.nlm.nih.gov/genome/annotation\\_euk/Mandrillus\\_leucophaeus/100/](https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Mandrillus_leucophaeus/100/).

### Phylogenetic relationships among primates

To investigate phylogenetic relationships among primates, we selected the longest isoform for each protein-coding gene from 26 primate species and 3 non-primate species (S1 Table). After clustering, aligning, trimming, and filtering (Materials and methods), there were 1,730 single-copy orthologs present in at least 27 of the 29 species (see S3 Table for the orthogroup, protein name, chromosome, and location of each single-copy ortholog in the human genome). The cutoffs used to filter the dataset ensure high species coverage while still retaining a large number of orthologs. The coding sequences (CDS) of these orthologs have an average length of 1,018 bp and 178 parsimony informative characters per gene. Concatenation of these loci resulted in an alignment of 1,761,114 bp, with the fraction of gaps/ambiguities varying from 4.04% (*Macaca mulatta*) to 18.37% (*Carlito syrichta*) (S4 Table).

We inferred 1,730 individual gene trees from nucleotide alignments using maximum likelihood in IQ-TREE 2 [38] and then inferred a species tree using these gene tree topologies as input to ASTRAL III ([39]; Materials and methods). We used the mouse, *Mus musculus*, as an outgroup to root the species tree. This approach resulted in a topology (which we refer to as “ML-ASTRAL”; Fig 1) that largely agrees with previously published phylogenies [12,13]. We also used IQ-TREE to carry out a maximum likelihood analysis of the concatenated nucleotide alignment (a topology we refer to as ML-CONCAT). This analysis resulted in a topology that differed from the ML-ASTRAL tree only with respect to the placement of *Aotus nancymaae* (owl monkey), rather than sister to the *Saimiri+Cebus* clade (as in Fig 1), the ML-CONCAT tree places *Aotus* sister to *Callithrix jacchus*, a minor rearrangement around a very short internal branch (Fig 1). All branches of the ML-ASTRAL species tree are supported by maximum local posteriors, the default support values provided by ASTRAL III [40], except for the branch that defines *Aotus* as sister to the *Saimiri+Cebus* clade (0.46 local posterior probability). Likewise, each branch in the ML-CONCAT tree is supported by 100% bootstrap values, including the branch uniting *Aotus* and *Callithrix*. We return to this conflict in the next section.

There has been some contention as to the placement of the mammalian orders Scandentia (treeshrews) and Dermoptera (colugos) [41–50]. The controversy concerns whether Dermoptera is sister to Primates, Scandentia is sister to Primates, or Dermoptera and Scandentia are sister groups. As expected, both the ML-ASTRAL and ML-CONCAT trees place these 2 groups outside the Primates with maximal statistical support (i.e., local posterior probabilities of 1.0 and bootstrap values of 100%; Fig 1); they also both point to Dermoptera as the closest sister lineage to the Primates [12,51–53]. However, while support values such as the bootstrap

or posterior probability provide statistical confidence in the species tree topology, there can be large amounts of underlying gene tree discordance even for branches with 100% support (e.g., [54–56]). To assess discordance generally, and the relationships among the Primates, Scandentia, and Dermoptera in particular, we used IQ-TREE to calculate both the gene concordance factor (gCF) and site concordance factor (sCF) [57] for each internal branch of the topology in Fig 1. These 2 measures represent the fraction of genes and sites, respectively, which are in agreement with the species tree for any particular branch.

Examining concordance factors helps to explain previous uncertainty in the relationships among Primates, Scandentia, and Dermoptera (Fig 1). Although the bootstrap support is 100% and the posterior probability is 1.0 on the branch leading to the Primate common ancestor, the gCF is 45%, and the sCF is 39%. These values indicate that, of decisive gene trees ( $n = 1,663$ ), only 45% of them contain the branch that is in the species tree; this branch reflects the Primates as a single clade that excludes Scandentia and Dermoptera. While the species tree represents the single topology supported by the most gene trees (hence the strong statistical support for this branch), the concordance factors also indicate that a majority of gene tree topologies differ from the estimated species tree. In fact, the gCF value indicates that 55% of trees do not support a monophyletic Primate order, with either Dermoptera, Scandentia, or both lineages placed within Primates. Likewise, the sCF value indicates that only 39% of decisive sites in the total alignment support the branch uniting all primates, with 30% favoring Dermoptera as sister to the Primate suborder Strepsirrhini and 31% placing Dermoptera sister to the Primate suborder Haplorrhini. Similarly, only a small plurality of genes and sites have histories that place Dermoptera as sister to the Primates rather than either of the 2 alternative topologies (gCF = 37, sCF = 40; Fig 1), despite the maximal statistical support for these relationships. While discordance at individual gene trees can result from technical problems in tree inference (e.g., long-branch attraction, low phylogenetic signal, poorly aligned sequences, or model misspecification), it also often reflects biological causes of discordance such as ILS and introgression. We further address the possible role of technical errors in generating patterns of discordance in the section entitled “Sources of gene tree discordance” below.

Within the Primates, the phylogenetic affiliation of tarsiers (represented here by *C. syrichta*) has been debated since the first attempts by Buffon (1765) and Linnaeus (1767 to 1770) to systematically organize described species [58]. Two prevailing hypotheses group tarsiers (Tarsiiformes) with either lemurs and lorises (the “prosimian” hypothesis [59]) or with Simiiformes (the “Haplorrhini” hypothesis [60], where Simiiformes = Apes+OWM+NWM). The ML-ASTRAL and ML-CONCAT analyses place Tarsiiformes with Simiiformes, supporting the Haplorrhini hypothesis (Fig 1). The Strepsirrhines come out as a well-supported group sister to the other primates. Again, our inference of species relationships is consistent with previous genomic analyses [61,62] but also highlights the high degree of discordance in this part of the tree. The rapid radiation of mammalian lineages that occurred in the late Paleocene and early Eocene [32] encompassed many of the basal primate branches, including the lineage leading to Haplorrhini. The complexity of this radiation is likely the reason for low gCF and sCFs (39.5% and 36%, respectively) for the branch leading to Haplorrhini and perhaps explains why previous studies recovered conflicting resolutions for the placement of tarsiers [31,63,64].

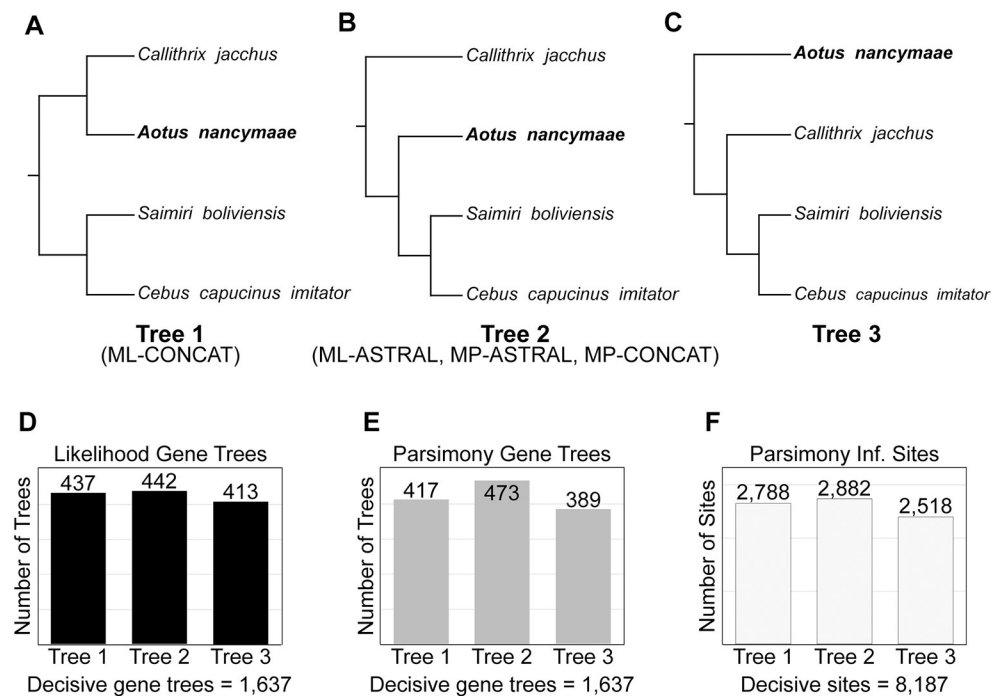
The remaining branches of the species tree that define major primate clades all have remarkably high concordance with the underlying gene trees (gCF >80%), though individual branches within these clades do not. The gCFs for the branches defining these clades are Strepsirrhini (lemurs+lorises) = 84.5, Catarrhini (OWM+Apes) = 90.0, Platyrrhini (NWM) = 96.6, Hominoidea (Apes) = 82.7, and Cercopithecoidea (OWM) = 92.3 (Fig 1). High gene tree/species tree concordance for these branches is likely due to a combination of more recent divergences (increasing gene tree accuracy) and longer times between branching events [65].

Within these clades, however, we see multiple recent radiations. One of the most contentious has been among the NWMs, a set of relationships we address next.

### ML concatenation affects resolution of the New World monkey radiation

Sometime during the mid to late Eocene (approximately 45 to 34 mya), a small number of primates arrived on the shores of South America [15,66]. These monkeys likely migrated from Africa [66] and on arrival underwent multiple rounds of extinction and diversification [15]. Three extant families from this radiation now make up the NWMs (Platyrrhini; Fig 1). Because of the rapidity with which these species spread and diversified across the new continent, relationships at the base of the NWM have been hard to determine [12–14,16–18].

As reported above, the concatenated analysis (ML-CONCAT) gives a different topology than the gene tree-based analysis (ML-ASTRAL). Specifically, the ML-CONCAT analysis supports a symmetrical tree, with *Aotus* sister to *Callithrix* (Fig 2A). In contrast, ML-ASTRAL supports an asymmetrical (or “caterpillar”) tree, with *Aotus* sister to a clade comprised of *Saimiri*+*Cebus* (Fig 2B). There are reasons to have doubts about both topologies. It is well known that carrying out maximum likelihood analyses of concatenated datasets can result in incorrect species trees, especially when the time between speciation events is short [67,68]. In fact, the specific error that is made in these cases is for ML concatenation methods to prefer a



**Fig 2. The 3 most frequent topologies of NWMs.** (A) Tree 1 is the symmetrical topology inferred by the ML-CONCAT analysis of 1,730 loci (1.76 Mb). (B) Tree 2 is the asymmetrical topology inferred by ASTRAL III using either maximum likelihood (ML-ASTRAL) or maximum parsimony (MP-ASTRAL) gene tree topologies. Using maximum parsimony on the concatenated alignment also returns this tree (MP-CONCAT). (C) Tree 3 is the alternative resolution recovered at high frequency in all gene tree analyses, though it is not the optimal species tree using any of the methods. (D) Number of gene trees supporting each of the 3 resolutions of the NWM clade when maximum likelihood is used to infer gene tree topologies. There are 1,637 decisive gene trees for these splits. (E) Gene tree counts when maximum parsimony is used to infer gene tree topologies. (F) Number of parsimony informative sites in the concatenated alignment supporting each of the 3 resolutions. ML-CONCAT, maximum likelihood concatenated; NWM, New World monkey.

<https://doi.org/10.1371/journal.pbio.3000954.g002>

symmetrical 4-taxon tree over an asymmetrical one, exactly as is observed here. Gene tree-based methods such as ASTRAL are not prone to this particular error, as long as the underlying gene trees are all themselves accurate [69,70]. However, if there is bias in gene tree reconstruction, then there are no guarantees as to the accuracy of the species tree. In addition, the ML-ASTRAL tree is supported by only a very small plurality of gene trees: There are 442 trees supporting this topology, compared to 437 supporting the ML-CONCAT topology and 413 supporting the third topology (Fig 2D). This small excess of supporting gene trees also explains the very low posterior support for this branch in the species tree (Fig 1). Additionally, a polytomy test [71], implemented in ASTRAL and performed using ML gene trees, failed to reject the null hypothesis of “polytomy” for the branch uniting *Aotus*+(*Saimiri*,*Cebus*) ( $P = 0.47$ ).

To investigate these relationships further, we carried out additional analyses. The trees produced from concatenated alignments can be biased in situations with high ILS when maximum likelihood is used for inference, but this bias does not affect parsimony methods [21,72]. Therefore, we analyzed exactly the same concatenated 1.76 Mb alignment used as input for ML but carried out a maximum parsimony analysis in PAUP\* [73]. As would be expected given the known biases of ML methods, the maximum parsimony tree (which we refer to as “MP-CONCAT”) returns the same tree as ML-ASTRAL, supporting an asymmetric topology of NWMs (Fig 2B). Underlying this result is a relatively large excess of parsimony informative sites supporting this tree (Fig 2F), which results in maximal bootstrap values for every branch. The 2 most diverged species in this clade (*Saimiri* and *Callithrix*) are only 3.26% different at the nucleotide level, so there should be little effect of multiple substitutions on the parsimony analysis.

As mentioned above, gene tree-based methods (such as ASTRAL) are not biased when accurate gene trees are used as input. However, in our initial analyses, we used maximum likelihood to infer the individual gene trees. Because protein-coding genes are themselves often a combination of multiple different underlying topologies [74], ML gene trees may be biased and using them as input to gene tree-based methods may still lead to incorrect inferences of the species tree [75]. Therefore, we used the same 1,730 loci as above to infer gene trees using maximum parsimony with MPBoot [76]. Although the resulting topologies still possibly represent the average over multiple topologies contained within a protein-coding gene, using parsimony ensures that this average tree is not a biased topology. These gene trees were used as input to estimate a species tree using ASTRAL; we refer to this as the “MP-ASTRAL” tree. Once again, the methods that avoid known biases of ML lend further support to an asymmetric tree, placing *Aotus* sister to the *Saimiri*+*Cebus* clade (Fig 2B). In fact, the gene trees inferred with parsimony now show a much greater preference for this topology, with a clear plurality of gene trees supporting the species tree (473 versus 417 supporting the second most common tree; Fig 2E). As a consequence, the local posterior for this branch in the MP-ASTRAL tree is 0.92, and the polytomy test performed using MP gene trees rejects ( $P = 0.037$ ) the null hypothesis of “polytomy” for the branch uniting *Aotus*+(*Saimiri*,*Cebus*). The increased number of concordant gene trees using parsimony suggests that the gene trees inferred using ML may well have been suffering from the biases of concatenation when multiple trees are brought together (as observed in the Great Apes [74]), reducing the observed levels of concordance.

A recent analysis of NWM genomes found *Aotus* sister to *Callithrix*, as in the ML-CONCAT tree, despite the use of gene trees to build the species tree [18]. However, the outgroup used in this analysis is a closely related species (*Brachyteles arachnoides*) that diverged during the NWM radiation and that shares a recent common ancestor with the ingroup taxa [12,13]. If the outgroup taxon used to root a tree shares a more recent common ancestor with subsets of ingroup taxa at an appreciable number of loci, the resulting tree topologies will be biased. A similar problem likely arose in previous studies that have used the Scandentia or Dermoptera



as outgroups to Primates. In general, this issue highlights the difficulty in choosing outgroups: Though we may have 100% confidence that a lineage lies outside our group of interest in the species tree, a reliable outgroup must also not have any discordant gene trees that place it inside the ingroup.

### Sources of gene tree discordance

As previously mentioned, there are both biological reasons for gene tree discordance (e.g., ILS or introgression) and technical reasons (e.g., long-branch attraction, homoplasy, low phylogenetic signal, poorly aligned sequences, or model misspecification). All of these phenomena may be reflected in gCFs and sCFs, but the proportion of discordance attributable to biological versus technical factors is often difficult to ascertain. We therefore performed additional analyses to assess the impact of error on estimates of concordance factors.

In order to determine the degree to which short alignments or genes with low phylogenetic signal contribute to inaccurate gene trees, we recalculated gCFs and sCFs using the genes with the 200 longest alignments in our dataset (lengths ranging from 1,640 bp to 6,676 bp, with 116 to 2,101 parsimony informative sites). The resulting gCFs for the branch leading to the Primate common ancestor increases from 45% to 66%, while the sCFs remain unchanged (S1 Fig). For the branch placing Dermoptera sister to Primates, using trees estimated from the 200 longest alignments resulted in a modest increase in gCFs from 37% to 45%. Overall, the gCFs for the 200 longest genes were higher for all branches in the tree, with the average gCF increasing from 65.18% to 79.74%. The consistent increase in gCF but not sCF when using longer genes points to errors in gene tree inference as a small, but significant, factor in our dataset.

Using a single outgroup (mouse) could potentially lead to biases such as long-branch attraction near the base of the tree. To ameliorate these concerns, we performed an additional analysis using 150 randomly chosen single-copy orthologs, with pika (*Ochotona princeps*) included as a second outgroup. As in the full dataset, maximum likelihood and parsimony were both applied to a concatenated dataset, and gene trees were also inferred via both ML and parsimony. Parsimony analysis of the concatenated alignment resulted in the same topology as in Fig 1, while a maximum likelihood analysis produced the same topology as the full ML-CONCAT tree from 1,730 loci, preferring a symmetric tree for the NWM clade. To assess the effect of including an additional outgroup on concordance factors, we calculated gCFs and sCFs using the 150 single-copy orthologs both with (S2A Fig) and without (S2B Fig) pika (using ML gene trees). In contrast to expectations about any error introduced by long-branch attraction, we observe slightly lower gCFs near the base of the tree when pika is included (S2A Fig). sCFs are not affected by the inclusion of pika. These analyses indicate that including additional outgroups when analyzing the full dataset is unlikely to reduce concordance factors or to change inferences of the species tree.

Technical errors leading to discordance should be more prominent deeper in the tree, as there is more opportunity for long-branch attraction, homoplasy, poor alignments, or model misspecification to cause problems. To determine whether concordance factors for deep branches in the primate tree are disproportionately affected by error, we looked for a correlation between concordance factors and the age of each bifurcation in the tree. For gCFs, we found no correlation with node age ( $r^2 = 0.0094$ ), while sCFs were slightly negatively correlated ( $r^2 = 0.2998$ ; S3 Fig). The negative correlation found between sCFs and node age is consistent with the expectation that substitutions occurring on deeper branches of the tree are more likely to suffer from the effects of multiple substitutions (homoplasy). While there may still be technical factors affecting gCFs, true discordance throughout the tree is high enough to mask any such effect.

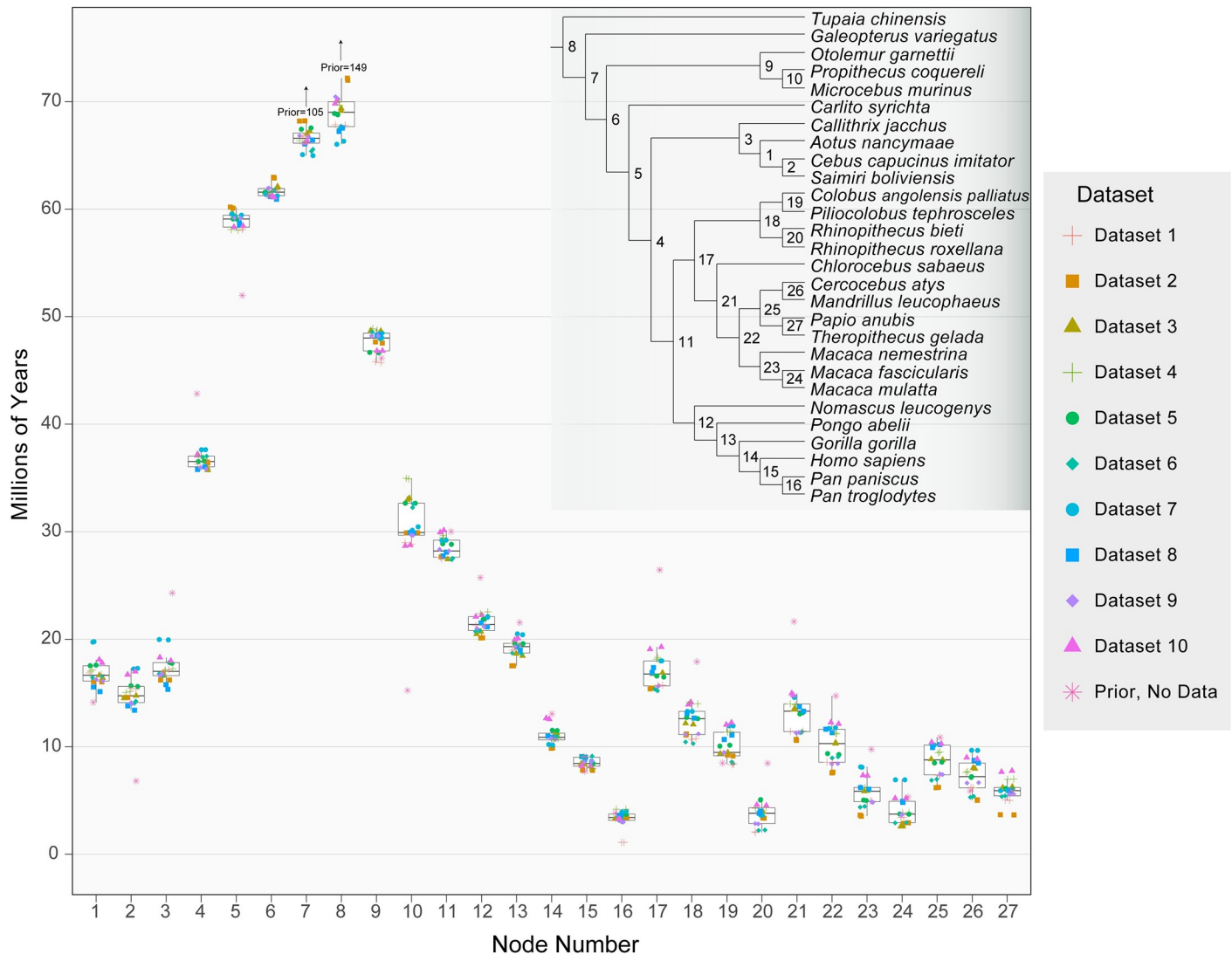
A recent simulation study [77] reported that negative selection, in combination with large differences in effective population size, can generate strong enough asymmetries in gene tree topologies that the most common topology does not match the species tree. Such an effect, if real, would mislead both gene tree-based and concatenation-based approaches to species tree inference. However, previous theoretical results predict that there should be no effect of negative selection on the distribution of tree topologies [78–81], and the new results were obtained using custom simulation software. To clarify this issue, we used the open-source simulator SLiM [82] to study non-recombining loci under the most extreme parameters used by He and colleagues [77] (see [Materials and methods](#)). We found no evidence for the bias in gene tree frequencies recently reported (S4A Fig). However, we observed fewer than 1 mutation per locus at the end of our simulations under the parameters exactly replicating He and colleagues [77], suggesting we may not have generated sufficient deleterious variation to observe the effect. To address this, we simulated the same conditions but with the deleterious mutation rate increased by 2 orders of magnitude and still did not observe a bias in topology frequencies (S4B Fig). Our results therefore indicate that weak negative selection does not generate gene tree discordance, consistent with population genetic theory [78–81].

### Strongly supported divergence times using fossil calibrations

Fossil-constrained molecular dating was performed using 10 independent datasets, each of which consisted of 40 protein-coding genes randomly selected (without replacement) and concatenated. The resulting datasets had an average alignment length of 39,374 bp (SD =  $2.6 \times 10^3$ ; S5 Table). Although individual discordant trees included in this analysis may have different divergence times, the difference in estimates of dates should be quite small [83]. We used 8 dated fossils (blue stars in Fig 1) from 10 studies for calibration (S6 Table). The most recent of these fossils is approximately 5.7 mya [84], while the most ancient is 55.8 mya [85]. Each separate dataset and the same set of “soft” fossil constraints, along with the species tree in Fig 1, were used as input to PhyloBayes 3.3 [86], which was run twice to assess convergence (Materials and methods).

We observed tight clustering of all estimated node ages across datasets and independent runs of PhyloBayes (Fig 3 and S6 Table). In addition, the ages of most major crown nodes estimated here are largely in agreement with previously published age estimates (Table 2). Some exceptions include the age of the crown Strepsirrhini (47.4 mya) and Haplorrhini (59.0 mya), which are more recent than many previous estimates for these nodes (range in the literature is Strepsirrhini = 51.6 to 68.7, Haplorrhini = 60.6 to 81.3; see Table 2). The crown nodes for Catarrhini, Hominoidea, and Cercopithecoidea (28.4, 21.4, and 16.8 mya, respectively) all fall within the range of variation recovered in previous studies (Table 2).

Our estimate for the most recent common ancestor of the extant primates (i.e., the last common ancestor of Haplorrhini and Strepsirrhini) is 61.7 mya, which is slightly more recent than several studies [13,31,33,88] and much more recent than other studies [12,87,89] (Table 2). However, our estimate is in good agreement with Herrera and colleagues [32], who used 34 fossils representing extinct and extant lineages (primarily Strepsirrhines) to infer divergence times among primates, concluding that the split occurred approximately 64 mya. Despite limited overlap in taxon sampling, 1 similarity between our study and that of Herrera and colleagues is that we have both used the maximum constraint of 65.8 million years (my) on the ancestral primate node suggested by Benton and colleagues [90], which likely contributes to the more recent divergence. It is worth noting that the soft bounds imposed in our analysis permit older ages to be sampled from the Markov chain, but these represented only a small fraction (median 3.37%) of the total sampled states after burn-in (S6 Table). To



**Fig 3. Mean node ages for independent Phylobayes dating runs.** Box plots show the median, interquartile range, and both minimum and maximum values of the mean nodes ages for 10 different datasets (with each dataset run twice). An additional run was performed with no sequence data to ascertain the prior on node divergence times in the presence of fossil calibrations (pink asterisks). Some prior ages were too large to include in the plot while still maintaining detail; these ages are given as numeric values. The species tree topology is from Fig 1; 95% HPD intervals for each node are reported in S7 Table. Node age estimates for each independent PhyloBayes run are provided in S1 Data. HPD, highest posterior density.

<https://doi.org/10.1371/journal.pbio.3000954.g003>

determine the effects of imposing the 65.8 my maximum constraint on the Primate node, we analyzed all 10 datasets for a third time with this constraint removed and report the divergence time of major primate clades in Table 2 (“No Max” entries). However, it may be that using genes that have gene trees most similar to the topology being dated will reduce bias caused by concatenation [74]. To determine whether using concordant loci has an impact on the estimated dates, we constructed an 11th dataset consisting of approximately 43 kb from the 20 loci most similar to the species tree in Fig 1 (as determined by Robinson–Foulds distances). There was no consistent difference in the dates estimated with this dataset (“Concord” entries in Table 2).

There are several caveats to our age estimates that should be mentioned. Maximum age estimates for the crown node of any given clade are defined by the oldest divergence among

Table 2. Mean crown node divergence times estimated in this study compared with mean divergences times estimated by 8 prior studies.

Node	This study	This study, no max*	This study, concord†	Herrera et al. [32]	Kistler et al. [33]	Perez et al. [17]	Springer et al. [13]	Meredith et al. [45]	Perelman et al. [12]	Wilkinson et al. [87]	Chatterjee et al. [31]
Primates	61.7	67.5	63	63.9	68	NA	67.8	71.5	87.2	84.5	63.7
Strepsirrhini	47.4	50.2	48.4	61.4	59	NA	54.2	55.1	68.7	49.8	51.6
Haplorrhini	59.0	63.8	59.8	61.9	67	60.6	61.2	62.4	81.3	NA	NA
Catarrhini	28.4	29.0	27.2	32.1	33	27.8	25.1	20.6	31.6	31.0	29.3
Hominoidea	21.4	21.6	19.9	NA	21	18.44	17.4	14.4	20.3	NA	21.5
Cercopithecidae	16.8	16.9	14.2	NA	24	13.4	13.2	NA	17.6	14.1	23.4

Estimates were calculated by averaging the mean times across all runs for 10 independent datasets.

\*Refers to the average divergence time of the crown node for the indicated taxonomic group when the 65.8 my maximum constraint was removed from the Primate node.

†Refers to the average divergence time of the crown node for the indicated taxonomic group when divergence times were estimated using the most concordant gene trees. Datasets used in all dating analyses are available via Data Dryad in the archive All\_Dating\_Datasets\_DRYAD.tar.gz, <https://doi.org/10.5061/dryad.rfj6q577d> [22]. my, million years; NA, not applicable.

<https://doi.org/10.1371/journal.pbio.3000954.t002>

sampled taxa in the clade. This limitation results in underestimates for nearly all crown node ages as, in practice, complete taxon sampling is difficult to achieve. Fossil calibrations are often employed as minimum constraints in order to overcome the limitations imposed by taxon sampling, allowing older dates to be estimated more easily. On the other hand, the systematic underestimation of crown node ages due to taxon sampling is somewhat counteracted by the overestimation of speciation times due to ancestral polymorphism. Divergence times estimated from sequence data represent the coalescence times of sequences, which are necessarily older than the time at which 2 incipient lineages diverged [91,92]. This overestimation will have a proportionally larger effect on recent nodes (such as the *Homo/Pan* split; Fig 3, node 15), but the magnitude can be no larger than the average level of polymorphism in ancestral populations and will be additionally reduced by post-divergence gene flow.

### Introgression during the radiation of primates

There is now evidence for recent interspecific gene flow between many extant primates, including introgression events involving humans [25], gibbons [93,94], baboons [9,27], macaques [95,96], and vervet monkeys [10], among others. While there are several widely used methods for detecting introgression between closely related species (see chapters 5 and 9 in [97]), detecting ancient gene flow is more difficult. One of the most popular methods for detecting recent introgression is the *D* test (also known as the “ABBA-BABA” test; [98]). This test is based on the expectation that, for any given branch in a species tree, the 2 most frequent alternative resolutions should be present in equal proportions. However, the *D* test uses individual SNPs to evaluate support for alternative topologies and explicitly assumes an infinite sites model of mutation (i.e., no multiple hits). As this assumption will obviously not hold the further back in time one goes, a different approach is needed.

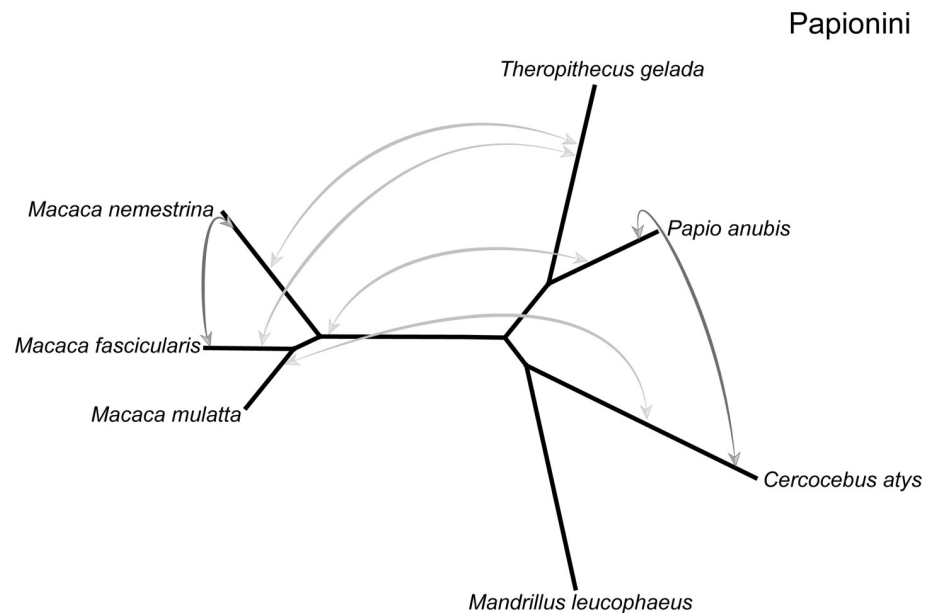
Fortunately, Huson and colleagues [36] described a method that uses gene trees themselves (rather than SNPs) to detect introgression. Using the same expectations as in the *D* test, these authors looked for a deviation from the expected equal numbers of alternative tree topologies using a test statistic they refer to as  $\Delta$ . As far as we are aware,  $\Delta$  has only rarely been used to test for introgression in empirical data, possibly because of the large number of gene trees needed to assess significance or the assumptions of the parametric method proposed to obtain *P*

values. Here, given our large number of gene trees and large number of internal branches to be tested, we adapt the  $\Delta$  test for genome-scale data.

To investigate patterns of introgression within primates, we used 1,730 single-copy loci to test for deviations from the null expectation of  $\Delta$  on each of the 24 internal branches of the primate phylogeny (Materials and methods). To test whether deviations in  $\Delta$  were significant (i.e.,  $\Delta > 0$ ), we generated 2,000 resampled datasets of 1,730 gene tree topologies each.  $P$  values were calculated from  $Z$ -scores generated from these resampled datasets. Among the 17 branches where at least 5% of topologies were discordant, we found 7 for which  $\Delta$  had  $P < 0.05$ .

To further verify these instances of potential introgression, for each of these 7 branches we increased the number of gene trees used, as well as the alignment length for each locus, by subsampling a smaller set of taxa. We randomly chose 4 taxa for each internal branch tested that also had this branch as an internal branch and then aligned all orthologs present in a single copy in each taxon. These steps resulted in approximately 3,600 to 6,400 genes depending on the branch being tested (S8 Table). Additionally, because instances of hybridization and introgression are well documented among macaques [96,99,100], we similarly resampled orthologs from the 3 *Macaca* species in our study.

We recalculated  $\Delta$  using the larger gene sets and found significant evidence (after correcting for  $m = 17$  multiple comparisons by using a cutoff of  $P = 0.00301$ ) for 6 introgression events, all of which occurred among the Papionini (Fig 4 and see next paragraph). Within the Hominoidea, we found  $\Delta = 0.0518$  for the branch leading to the great apes ( $P = 0.030$ ). The asymmetry in gene tree topologies here suggests that gene flow may have happened between gibbons (represented by *Nomascus*) and the ancestral branch leading to the African hominoids



**Fig 4. Introgression among Papionini taxa (the species tree is unrooted for clarity).** Arrows indicate that a significant  $\Delta$  was found in our 4 taxon tests and identify the 2 lineages inferred to have exchanged genes (values underlying these tests are listed in S8 Table). Among the Papionini, there was evidence of introgression between African taxa (*Papio*, *Theropithecus*, and *Cercocebus*) and Asian *Macaca* species (light gray arrows). Introgression events likely occurred between African taxa and the ancestral *Macaca*, which had a wide distribution across Northern Africa prior to the radiation throughout Asia 2–3 mya [103]. More recent instances of introgression are inferred between macaque species and among the African Papionini (dark gray arrows). mya, million years ago.

<https://doi.org/10.1371/journal.pbio.3000954.g004>

(humans, chimpanzees, and gorillas), but, like the  $D$  test,  $\Delta$  cannot tell us the direction of introgression. Although currently separated by significant geographic distances (African apes south of the Sahara Desert and gibbons all in Southeast Asia), it is worth noting that fossil hominoids dating from the early to late Miocene had a broad distribution extending from Southern Africa to Europe and Asia [101]. Support for introgression between ancestral hominins and ancestral chimpanzees has been previously reported [102]; our 4-taxon analyses found marginal support for this conclusion ( $\Delta = 0.0917$ ,  $P = 0.055$ ).

Within the OWM, approximately 40% of Cercopithecine species are known to hybridize in nature [34]. Consistent with this, *M. nemestrina* and *Macaca fascicularis* showed a strong signature of gene flow in our data ( $\Delta = 0.1761$ ,  $P = 1.377e-09$ ). These 2 species have ranges that currently overlap (S5 Fig). In contrast to the clear signal of recent gene flow in the macaques, we detected a complex pattern of ancient introgression between the African Papionini (*Cercocebus*, *Mandrillus*, *Papio*, and *Theropithecus*) and the Asian Papionini (*Macaca*) (Fig 4). The  $\Delta$  test was significant using multiple different subsamples of 4 taxa, suggesting multiple ancestral introgression events. An initial attempt to disentangle these events using Phylonet v3.8.0 [104] with the 7 Papionini species and an outgroup was unsuccessful, as Phylonet failed to converge on an optimal network for these taxa. An attempt to infer the network with SNaQ [105] gave similarly ambiguous results. When there are multiple episodes of gene flow within a clade, even complex computational machinery may be unable to infer the correct combination of events.

As an alternative approach, we used 4-taxon trees to estimate  $\Delta$  for each *Macaca* species paired with 2 African Papionini (1 from the *Papio*+*Theropithecus* clade and 1 from the *Mandrillus*+*Cercocebus* clade; see S8 Table) and an outgroup. Significant introgression was detected using each of the *Macaca* species and 3 of the 4 African Papionini species (*Cercocebus*, *Theropithecus*, and *Papio*). These results suggest gene flow between the ancestor of the 3 *Macaca* species in our analysis and the ancestors of the 3 African Papionini in our analysis, or 1 introgression event involving the ancestor of all 4 African species coupled with a second event that masked this signal in *Mandrillus*. This second event may have been either biological (additional introgression events masking the signal) or technical (possibly the lack of continuity or completeness of the *Mandrillus* reference genome sequence), but in either case, we could not detect introgression in the available drill sequence. The latter scenario would fit better with the current geographic distributions of these species, as they are on 2 different continents. However, the fossil record indicates that by the late Miocene to late Pleistocene, the ancestral distribution of the genus *Macaca* covered all of North Africa, into the Levant, and as far north as the United Kingdom (S5 Fig; [106]). The fossil record for *Theropithecus* indicates that several species had distributions that overlapped with *Macaca* during this time, including in Europe and as far east as India (S5 Fig; [107,108]). Ancestral macaques and ancestral Papionini may therefore have come into contact in the area of the Mediterranean Sea. The Sahara Desert is also responsible for the current disjunct distributions of many of these species. However, this region has experienced periods of increased rainfall or “greenings” over the past several million years [109–111]. Faunal migration through the Sahara, including by hominins, is hypothesized to have occurred during these green periods [110,112,113], resulting in successive cycles of range expansion and contraction [114]. Hybridization and introgression could have occurred between the ancestors of these groups during 1 of these periods.

Our results on introgression come with multiple caveats, both about the events we detected and the events we did not detect. As with the  $D$  test, there are multiple alternative explanations for a significant value of  $\Delta$  besides introgression. Ancestral population structure can lead to an asymmetry in gene tree topologies [115] though it requires a highly specific, possibly unlikely population structure. For instance, if the ancestral population leading to *M. nemestrina* was

more closely related to *M. fascicularis* than was the ancestral population leading to its sister species, *M. mulatta* (Fig 4), then there could be an unequal number of alternative topologies. Similarly, any bias in gene tree reconstruction that favors 1 alternative topology over the other could potentially lead to a significant value of  $\Delta$ . While this scenario is unlikely to affect recent divergences using SNPs, well-known biases that affect topology reconstruction deeper in the tree (such as long-branch attraction) could lead to gene tree asymmetries. However, we did not observe any significant  $\Delta$ -values for branches more than approximately 10 my old. One alternative approach to avoid biases in reconstruction could be the use of transposon insertions or other rare genomic changes (cf. [116,117]). Future analyses that compare these different approaches to detecting introgression would be especially useful.

There are also multiple reasons why our approach may have missed introgression events, especially deeper in the tree. All methods that use asymmetries in gene tree topologies miss gene flow between sister lineages, as such events do not lead to changes in the proportions of underlying topologies. Similarly, equal levels of gene flow between 2 pairs of non-sister lineages can mask both events, while even unequal levels will lead one to miss the less frequent exchange. More insidiously, especially for events further back in time, extinction of the descendants of hybridizing lineages will make it harder to detect introgression (though extinction of donor lineages is less of a problem than extinction of lineages receiving migrants). Internal branches closer to the root will be on average longer than those near the tips because of extinction [118], and therefore, introgression between non-sister lineages would have to occur longer after speciation in order to be detected. For instance, gene flow among Strepsirrhine species has been detected in many previous analyses of more closely related species (e.g., [119–122]), but the deeper relationships among the taxa sampled here may have made it very difficult to detect introgression. Nevertheless, our analyses were able to detect introgression between many primate species across the phylogeny.

## Conclusions

Several previous phylogenetic studies of primates have included hundreds of taxa, but fewer than 70 loci [12,13]. While the species tree topologies produced by these studies are nearly identical to the one recovered in our analysis, the limited number of loci meant that it was difficult to assess gene tree discordance accurately. By estimating gene trees from 1,730 single-copy loci, we were able to assess the levels of discordance present at each branch in the primate phylogeny. Understanding discordance helps to explain why there have been long-standing ambiguities about species relationships near the base of primates and in the radiation of NWMs. Our analyses reveal how concatenation of genes—or even of exons—can mislead maximum likelihood phylogenetic inference in the presence of discordance, but also how to overcome these biases. Discordance also provides a window into introgression among lineages, and here, we have found evidence for exchange among several species pairs. Each instance of introgression inferred from the genealogical data is plausible insofar as it can be reconciled with current and ancestral species distributions.

## Materials and methods

### Source material and sequencing

The San Diego Zoo and the Washington National Primate Research Center provided biomaterials to the Human Genome Sequencing Center (HGSC), Baylor College of Medicine under an agreement that granted permission to the HGSC to use the biomaterials for academic scientific research. This is the standard agreement between these institutions which regularly provide this service and the academic community that uses their biomaterials for various types of

analyses. The HGSC does not pay for biomaterials but does cover the costs of shipping the biomaterials from the provider Baylor.

For the sequencing of the *C. angolensis palliatus* genome, paired-end (100 bp) libraries were prepared using DNA extracted from heart tissue (isolate OR3802 from the San Diego Zoo). Sequencing was performed using 9 Illumina Hi-seq 2000 lanes and 4 Illumina Hi-seq 2500 lanes with subsequent assembly carried out using ALLPATHS-LG software (v. 48744) [123]. Additional scaffolding and gap-filling was performed using Atlas-Link v. 1.1 (<https://www.hgsc.bcm.edu/software/atlas-link>) and Atlas-GapFill v. 2.2. (<https://www.hgsc.bcm.edu/software/atlas-gapfill>), respectively. Annotation for all 3 species was carried out using the NCBI Eukaryotic Genome Annotation Pipeline. A complete description of the pipeline can be viewed at [https://www.ncbi.nlm.nih.gov/genome/annotation\\_euk/process/](https://www.ncbi.nlm.nih.gov/genome/annotation_euk/process/).

For the sequencing of the *M. nemestrina* genome, DNA was extracted from a blood sample (isolate M95218 from the Washington National Primate Research Center). Paired-end libraries were prepared and sequenced on 20 Illumina Hi-Seq 2000 lanes with the initial assembly performed using ALLPATHS-LG as above. Scaffolding was conducted using Atlas-Link v. 1.1. Additional gap-filling was performed using the original Illumina reads and Atlas-GapFill v. 2.2, as well as long reads generated using the Pacific Biosciences RS (60 SMRT cells) and RSII (50 SMRT cells) platforms. The PacBio reads were mapped to scaffolds to fill remaining gaps in the assembly using PBJelly2 (v. 14.9.9) [124].

For the sequencing of the *M. leucophaeus* genome, DNA was extracted from heart tissue (isolate KB7577 from the San Diego Zoo). Paired-end libraries were prepared and sequenced on 9 Illumina Hi-Seq 2000 lanes with the initial assembly performed using ALLPATHS-LG as above. Additional scaffolding was completed using Atlas-Link v. 1.1, and additional gap-filling in scaffolds was performed using the original Illumina reads and Atlas-GapFill v. 2.2.

## Phylogenomic analyses

The full set of protein-coding genes for 26 primates and 3 non-primates were obtained by combining our newly sequenced genomes with already published data (see S1 Table for references and accessions and Table 1 and S2 Table for genome statistics). Ortholog clustering was performed by first executing an all-by-all BLASTP search [125,126] using the longest isoform of each protein-coding gene from each species. The resulting BLASTP output was clustered using the mcl algorithm [127] as implemented in FastOrtho [128] with various inflation parameters (the maximum number of clusters was obtained with *inflation* = 5). Orthogroups were then parsed to retain those genes present as a single copy in all 29 taxa (1,180 genes), 28 of 29 taxa (1,558 genes), and 27 of 29 taxa (1,735 genes). We chose to allow up to 2 missing species per alignment to maximize the data used in our phylogenomic reconstructions while maintaining high taxon occupancy in each alignment.

CDS for each single-copy orthogroup were aligned, cleaned, and trimmed via a multistep process: First, sequences in each orthogroup were aligned by codon using GUIDANCE2 [129] in conjunction with MAFFT v7.407 [130] with 60 bootstrap replicates. GUIDANCE2 uses multiple bootstrapped alignments to generate quality scores for each column in the final alignment as well as for each taxon sequence in each alignment. Sequence residues in the resulting MAFFT alignment with GUIDANCE scores <0.93 were converted to gaps, and sites with >50% gaps were removed using Trimal v1.4.rev22 [131]. Alignments shorter than 200 bp (full dataset) or 300 bp (4-taxon tests for introgression), and alignments that were invariant or contained no parsimony informative characters, were removed from further analyses. Alignments with high numbers of discordant sites were further inspected for errors and removed from the analysis when warranted. This resulted in 1,730 loci for the full analysis (see S8 Table for gene counts used in 4-taxon tests).



IQ-TREE v2-rc1 was used with all 1,730 aligned loci to estimate a maximum likelihood concatenated tree with an edge-linked, proportional-partition model, and 1,000 ultrafast bootstrap replicates [132,133]. This strategy uses ModelFinder [134] to automatically find the best-fit model for each ortholog alignment (partition). Branch lengths are shared between partitions, with each partition having its own rate that rescales branch lengths, accommodating different evolutionary rates between partitions. The full IQ-TREE command line used was “iqtree -p Directory\_of\_Gene\_Alignments--prefix -m MFP -c 8 -B 1000”. Maximum likelihood gene trees were estimated for each alignment with nucleotide substitution models selected using ModelFinder [134] as implemented in IQ-TREE. The full IQ-TREE command line used was “iqtree -s Directory\_of\_Gene\_Alignments--prefix -m MFP -c 8”. We used the resulting maximum likelihood gene trees to estimate a species tree using ASTRAL III (ML-ASTRAL) [39]. Parsimony gene trees were generated using MPboot [76] and used to estimate a species tree using ASTRAL III (MP-ASTRAL), while PAUP\* [73] was used to estimate the concatenated parsimony tree (MP-CONCAT) with 500 bootstrap replicates. IQ-TREE was used to calculate both gCFs and sCFs, with sCFs estimated from 300 randomly sampled quartets using the command line “iqtree -cf-verbose--gcf 1730\_GENETREE.treefile -t Species\_tree\_file--df-tree--scf 300 -p Directory\_of\_Gene\_Alignments -c 4”.

### Effects of selection on gene tree distributions

We performed 100 replicate simulations for each mutation rate condition using SLiM version 3.3.1 [82], with tree sequence recording turned on and no neutral mutations. Each replicate simulation consisted of 50 non-recombining loci of 1 kb each, with free recombination between loci, for 3 populations with the phylogenetic relationship ((p2,p3),p1). These simulations closely match the population genetic parameters under the most extreme asymmetry condition reported in He and colleagues [77], with population sizes, selection coefficients, and mutation rates rescaled 2 orders of magnitude for performance (SLiM recipes are available via Data Dryad: <https://doi.org/10.5061/dryad.rfj6q577d> [22]). These parameters include a per-locus deleterious mutation rate of  $3 \times 10^{-7}$  per generation; a population-scaled selection coefficient ( $Ns$ ) of  $-7.5$ ; an internal branch subtending p2 and p3 of  $0.01N$  generations (where  $N$  is the population size of p1 and p2); a population size for p3 that is 0.04 times that of p1 and p2; and tip branch lengths of  $8N$  generations. In the higher mutation rate condition, the per-locus rate was increased to  $3 \times 10^{-5}$  per generation. We randomly sampled 1 chromosome from each population at each locus at the end of the simulation and obtained the genealogy of these samples recorded in the tree sequence at the locus.

### Introgression analyses

For each internal branch of the primate tree where the proportion of discordant trees was  $>5\%$  of the total, concordance factors were used to calculate the test statistic  $\Delta$ , where

$$\Delta = \frac{\text{Number of } DF1 \text{ trees} - \text{Number of } DF2 \text{ trees}}{\text{Number of } DF1 \text{ trees} + \text{Number of } DF2 \text{ trees}},$$

where  $DF1$  trees represent the most frequent discordant topology, and  $DF2$  trees are the second most frequent discordant topology. This is a normalized version of the statistic proposed by Huson and colleagues [36], which only included the numerator of this expression. Note also that, by definition,  $\Delta$  here is always equal to or greater than 0. To test whether deviations from zero were significant (i.e.,  $\Delta > 0$ ), we calculated  $\Delta$  for 2,000 pseudo-replicate datasets generated by resampling gene trees with replacement. The resulting distribution was used to calculate  $Z$ -scores and the resulting  $P$  values for the observed  $\Delta$  value associated with each branch

tested [135]. Of the 17 internal branches where >5% of topologies were discordant, 7 were significant at  $P < 0.05$ , and selected for more extensive testing. For each of the 7 significant branches in the all-Primates tree, 4 taxa were selected that included the target branch as an internal branch. Single-copy genes present in each taxon were aligned as previously described. Alignments with no variant or parsimony informative sites were removed from the analysis, and gene trees were estimated using maximum likelihood in IQ-TREE 2. The test statistic,  $\Delta$ , was calculated, and significance was again determined using 2,000 bootstrap replicates with the  $P$  value threshold for significance corrected for multiple comparisons ( $m = 17$ ) using the Dunn–Šidák correction [136,137].

## Molecular dating

Molecular dating analyses were performed on 10 datasets consisting of 40 CDS alignments each sampled randomly without replacement from the 1,730 loci used to estimate the species tree. Gene alignments were concatenated into 10 supermatrices ranging from 36.7 kb to 42.7 kb in length (see S5 Table for the length of each alignment). Each dataset was then analyzed using PhyloBayes 3.3 [86] with sequences modeled using a site-specific substitution process with global exchange rates estimated from the data (CAT-GTR; [138]). Among-site rate variation was modeled using a discrete gamma distribution with 6 rate categories. A relaxed molecular clock [139] with 8, soft-bounded, fossil calibrations (see S6 Table) was used to estimate divergence times on the fixed species tree topology (Fig 1); the analyses were executed using the following command line: `pb -x 1 15000 -d Alignment.phy -T Tree_file.tre -r outgroup_file.txt -cal 8_fossil.calib -sb -gtr -cat -bd -dgam 6 -ln -rp 90 90`. Each dataset was analyzed for 15,000 generations, sampling every 10 generations, with 5,000 generations discarded as burn-in. Each dataset was analyzed twice to ensure convergence of the average age estimated for each node (Fig 3 shows the node age for both runs). To determine the effect of including a maximum constraint on the root of the Primates, we analyzed each dataset a third time with this constraint removed. Both the constrained and unconstrained node ages for major groups within the Primates are reported in Table 2.

Single-copy CDS gene alignments, gene trees, dating datasets, SLiM3 recipes, unaligned gene sequences, and PAUP commands can be accessed via the Data Dryad repository located at <https://doi.org/10.5061/dryad.rfj6q577d> [22].

## Supporting information

**S1 Fig. Concordance factors for the species tree in Fig 1 calculated using maximum likelihood gene trees and site patterns from the 200 longest single-copy loci alignments used in the 1,730-gene analysis.** In general, gCFs increase, while the sCFs remain the same, indicating that gene tree error is a likely source of some discordance. gCF, gene concordance factor; sCF, site concordance factor.

(PDF)

**S2 Fig. Concordance factors calculated using 150 randomly chosen single-copy orthologs, with pika (*Ochotona princeps*) included as an additional outgroup to mouse.** (A) gCFs and sCFs for these 150 genes when pika is included. (B) gCFs and sCFs for these same genes when pika is not included. We observe slightly higher gCFs near the base of the tree with pika excluded (red boxes). Note that these species trees use unit-length branch lengths for readability of branch labels. gCF, gene concordance factor; sCF, site concordance factor.

(PDF)

**S3 Fig. Gene and site concordance factors plotted as a function of node depth (in millions of years).** No correlation was found between gCFs and node depth, whereas a slightly negative correlation was found between sCFs and node depth. This relationship indicates that homoplasy may act to slightly reduce sCFs deeper in the tree. The data underlying mean node ages are provided in [S1 Data](#). gCF, gene concordance factor; sCF, site concordance factor. (PDF)

**S4 Fig. Forward simulations using SLiM3 with the most extreme parameters used by He et al. (2020):** population size combination “F” with  $s = -7.5 \times 10^{-6}$  and  $\Delta\tau = 2,000$ . Our results show no significant difference in the distribution of gene tree topologies in the presence of negative selection (A). This result holds for simulations in which we increased the per-locus mutation rate by 2 orders of magnitude (B). SLiM3 recipes are available via Data Dryad at <https://doi.org/10.5061/dryad.rfj6q577d> [22]. Gene tree counts for both simulations, A and B, are available in [S1 Data](#). (PDF)

**S5 Fig. Present-day species distributions for 4 African Papionini (*Papio*, *Theropithecus*, *Mandrillus*, and *Cercocebus*) and 3 Asian *Macaca* species included in the introgression analysis.** The ancestral *Macaca* distribution (gray shading) is inferred from *Macaca* fossil localities in Africa and Europe as reviewed in Roos et al. [106]. The ancestral *Macaca* distribution likely represents only a fraction of the species range from the late Miocene to the late Pleistocene in Africa and Europe. The contemporary distribution of the African *Macaca sylvanus* (bright green) is included for reference; the current distribution of *Macaca nemestrina* is completely contained within that of *Macaca fascicularis*. Fossil localities for *Theropithecus* species hypothesized to overlap contemporaneously with various ancestral *Macaca* are included. Citations for spatial data of extant species: *M. nemestrina* (Richardson et al., 2008), *M. fascicularis* (Ong and Richardson, 2008), *M. sylvanus* (Butynski et al., 2008), *Macaca mulatta* (Timmins et al., 2008), *Theropithecus gelada* (Gippoliti et al., 2019), *Papio anubis* (Kingdon et al., 2008), *Cercocebus atys* (Oates et al., 2016), and *Mandrillus leucophaeus* (Oates and Butynski, 2008). Base map was obtained from the public domain map database Natural Earth (<http://www.naturalearthdata.com/downloads/>). (PDF)

**S1 Table. Genomes analyzed in this study with the original NCBI release date, the publication for the reference used, and the accession number for the assembly.** When possible, the most recent version for each genome was used. (DOCX)

**S2 Table. All published genomes used in this study, including links to the assemblies and NCBI BioProjects.** Annotation information is included for each genome at the time of download. (XLSX)

**S3 Table. Orthogroup, protein name, human chromosome number, and coordinates for the single-copy human orthologs used in the 1,730 gene analysis.** Alignment files are named by orthogroup, allowing the use of this table to identify the protein in each alignment. (XLSX)

**S4 Table. Gaps/ambiguities by species and as a percentage of total alignment length.** \* denotes species sequenced this study. (DOCX)

**S5 Table. Lengths for each 40-locus concatenated alignment used in the molecular dating analyses.** Each dataset was analyzed twice until node age estimates converged (15–25k steps) using a log-normal auto-correlated model [139]. Datasets are available via Data Dryad at <https://doi.org/10.5061/dryad.rfj6q577d> [22].  
(DOCX)

**S6 Table. Fossil calibrations employed in this study.** Node numbering corresponds to the numbering in Fig 3. Median underflow/overflow for each calibration was calculated from 20 independent runs performed on 10 datasets (2 runs per dataset).  
(DOCX)

**S7 Table. Mean node age for 20 independent PhyloBayes dating runs.** Node numbers correspond to the numbering in Fig 3. The 95% HPD intervals were calculated by averaging the minimum and maximum of the 95% HPD interval for each dating run. HPD, highest posterior density.  
(DOCX)

**S8 Table. Quartets used to test for significant  $\Delta$  values for internal branches of the primate tree.** Branches tested correspond to the labeled branches in Fig 3. After correcting for multiple comparisons (Dunn–Šidák,  $P = 0.00301$ ), 3 internal branches and 8 quartets were found to have significant  $\Delta$  values, indicating a likely introgression event.  
(DOCX)

**S1 Data. The Excel workbook contains 5 different tabs.** Tab 1, Fig 3 Data: consists of the node age estimates for all 20 independent PhyloBayes dating analyses as well as the run used to determine the prior for each node; each estimate is plotted separately in Fig 3. Tab 2, Fig 3 Data for R: the same data as in tab 1, but formatted for analysis with the accompanying R script “plot\_DATING.R” available via Data Dryad: <https://doi.org/10.5061/dryad.rfj6q577d> [22]. Tab 3, S3\_Fig\_Data: the data used to generate S3 Fig. The average node ages estimated in tab 1 are used here to plot age vs. concordance factors estimated for each node in IQ-TREE. Tab 4, S4\_Fig\_PanelA\_Data: contains the tree counts that resulted from the SLiM3 simulation conditions pictured in S4A Fig. Tab 5, S4\_Fig\_PanelB\_Data: contains the tree counts for the SLiM3 simulation conditions pictured in S4B Fig. SLiM3 recipes for both simulations are available via Data Dryad at <https://doi.org/10.5061/dryad.rfj6q577d> [22].  
(XLSX)

## Acknowledgments

We thank Yue Liu for assistance in assembling the genomes, and Fábio Mendes and Gregg Thomas for helpful advice.

## Author Contributions

**Conceptualization:** Dan Vanderpool, Matthew W. Hahn.

**Data curation:** Daniel Hughes, Shwetha Murali, R. Alan Harris, Muthuswamy Raveendran, Donna M. Muzny, Richard A. Gibbs.

**Formal analysis:** Dan Vanderpool.

**Funding acquisition:** Jeffrey Rogers, Matthew W. Hahn.

**Investigation:** Dan Vanderpool.

**Methodology:** Dan Vanderpool, Bui Quang Minh, Robert Lanfear, Kim C. Worley, Matthew W. Hahn.

**Project administration:** Dan Vanderpool, Jeffrey Rogers, Matthew W. Hahn.

**Resources:** Daniel Hughes, Shwetha Murali, R. Alan Harris, Muthuswamy Raveendran, Donna M. Muzny, Richard A. Gibbs.

**Software:** Dan Vanderpool, Bui Quang Minh, Kim C. Worley.

**Supervision:** Jeffrey Rogers, Matthew W. Hahn.

**Visualization:** Dan Vanderpool.

**Writing – original draft:** Dan Vanderpool, Matthew W. Hahn.

**Writing – review & editing:** Dan Vanderpool, Jeffrey Rogers, Matthew W. Hahn.

## References

1. The 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature*. 2011; 473:544–544. <https://doi.org/10.1038/nature09991>
2. Fawcett GL, Raveendran M, Deiros DR, Chen D, Yu F, Harris RA, et al. Characterization of single-nucleotide variation in Indian-origin rhesus macaques (*Macaca mulatta*). *BMC Genomics*. 2011; 12:311. <https://doi.org/10.1186/1471-2164-12-311> PMID: 21668978
3. Higashino A, Sakate R, Kameoka Y, Takahashi I, Hirata M, Tanuma R, et al. Whole-genome sequencing and analysis of the Malaysian cynomolgus macaque (*Macaca fascicularis*) genome. *Genome Biol*. 2012; 13:R58. <https://doi.org/10.1186/gb-2012-13-7-r58> PMID: 22747675
4. Kuhlwilm M, Han S, Sousa VC, Excoffier L, Marques-Bonet T. Ancient admixture from an extinct ape lineage into bonobos. *Nat Ecol Evol*. 2019; 3:957–965. <https://doi.org/10.1038/s41559-019-0881-7> PMID: 31036897
5. Locke DP, Hillier LW, Warren WC, Worley KC, Nazareth LV, Muzny DM, et al. Comparative and demographic analysis of orang-utan genomes. *Nature*. 2011; 469:529–533. <https://doi.org/10.1038/nature09687> PMID: 21270892
6. de Manuel M, Kuhlwilm M, Frandsen P, Sousa VC, Desai T, Prado-Martinez J, et al. Chimpanzee genomic diversity reveals ancient admixture with bonobos. *Science*. 2016; 354:477–481. <https://doi.org/10.1126/science.aag2602> PMID: 27789843
7. Prado-Martinez J, Sudmant PH, Kidd JM, Li H, Kelley JL, Lorente-Galdos B, et al. Great ape genetic diversity and population history. *Nature*. 2013; 499:471–475. <https://doi.org/10.1038/nature12228> PMID: 23823723
8. Prüfer K, Munch K, Hellmann I, Akagi K, Miller JR, Walenz B, et al. The bonobo genome compared with the chimpanzee and human genomes. *Nature*. 2012; 486:527–531. <https://doi.org/10.1038/nature11128> PMID: 22722832
9. Rogers J, Raveendran M, Harris RA, Mailund T, Leppälä K, Athanasiadis G, et al. The comparative genomics and complex population history of *Papio* baboons. *Sci Adv*. 2019; 5:eaau6947. <https://doi.org/10.1126/sciadv.aau6947> PMID: 30854422
10. Svardal H, Jasinska AJ, Apetrei C, Coppola G, Huang Y, Schmitt CA, et al. Ancient hybridization and strong adaptation to viruses across African vervet monkey populations. *Nat Genet*. 2017; 49:1705–1713. <https://doi.org/10.1038/ng.3980> PMID: 29083404
11. Zhou X, Wang B, Pan Q, Zhang J, Kumar S, Sun X, et al. Whole-genome sequencing of the snub-nosed monkey provides insights into folivory and evolutionary history. *Nat Genet*. 2014; 46:1303–1310. <https://doi.org/10.1038/ng.3137> PMID: 25362486
12. Perelman P, Johnson WE, Roos C, Seuánez HN, Horvath JE, Moreira MAM, et al. A molecular phylogeny of living Primates. *PLoS Genet*. 2011; 7:e1001342. <https://doi.org/10.1371/journal.pgen.1001342> PMID: 21436896
13. Springer MS, Meredith RW, Gatesy J, Emerling CA, Park J, Rabosky DL, et al. Macroevolutionary dynamics and historical biogeography of primate diversification inferred from a species supermatrix. *PLoS ONE*. 2012; 7:e49521. <https://doi.org/10.1371/journal.pone.0049521> PMID: 23166696
14. Wang X, Lim BK, Ting N, Hu J, Liang Y, Roos C, et al. Reconstructing the phylogeny of new world monkeys (Platyrrhini): evidence from multiple non-coding loci. *Curr Zool*. 2019; 65:579–588. <https://doi.org/10.1093/cz/zoy072> PMID: 31616489

15. Silvestro D, Tejedor MF, Serrano-Serrano ML, Loiseau O, Rossier V, Rolland J, et al. Early arrival and climatically-linked geographic expansion of New World monkeys from tiny African ancestors. *Syst Biol*. 2018; 68:78–92. <https://doi.org/10.1093/sysbio/syy046> PMID: 29931325
16. Jameson Kiesling NM, Yi SV, Xu K, Gianluca Sperone F, Wildman DE. The tempo and mode of New World monkey evolution and biogeography in the context of phylogenomic analysis. *Mol Phylogenet Evol*. 2015; 82(Pt B):386–399. <https://doi.org/10.1016/j.ympev.2014.03.027> PMID: 24792088
17. Perez SI, Tejedor MF, Novo NM, Aristide L. Divergence times and the evolutionary radiation of New World monkeys (Platyrrhini, Primates): An analysis of fossil and molecular data. *PLoS ONE*. 2013; 8: e68029. <https://doi.org/10.1371/journal.pone.0068029> PMID: 23826358
18. Schrago CG, Seuáñez HN. Large ancestral effective population size explains the difficult phylogenetic placement of owl monkeys. *Am J Primatol*. 2019; 55:e22955. <https://doi.org/10.1002/ajp.22955> PMID: 30779198
19. Degnan JH, Rosenberg NA. Discordance of species trees with their most likely gene trees. *PLoS Genet*. 2006; 2:e68. <https://doi.org/10.1371/journal.pgen.0020068> PMID: 16733550
20. Huang H, Knowles LL. What is the danger of the anomaly zone for empirical phylogenetics? *Syst Biol*. 2009; 58:527–536. <https://doi.org/10.1093/sysbio/syp047> PMID: 20525606
21. Mendes FK, Hahn MW. Why concatenation fails near the anomaly zone. *Syst Biol*. 2018; 67:158–169. <https://doi.org/10.1093/sysbio/syx063> PMID: 28973673
22. Vanderpool D. Primate phylogenomics uncovers multiple rapid radiations and ancient interspecific introgression. *Dryad Digital Repository* [Internet]. 2020. <https://doi.org/10.5061/dryad.rfj6q577d>
23. Mallet J. Hybridization as an invasion of the genome. *Trends Ecol Evol*. 2005; 20:229–237. <https://doi.org/10.1016/j.tree.2005.02.010> PMID: 16701374
24. Mallet J, Besansky N, Hahn MW. How reticulated are species? *BioEssays*. 2016; 38:140–149. <https://doi.org/10.1002/bies.201500149> PMID: 26709836
25. Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, et al. A draft sequence of the Neanderthal genome. *Science*. 2010; 328:710–722. <https://doi.org/10.1126/science.1188021> PMID: 20448178
26. Lima MGM, de Sousa E Silva J, Černý D, Buckner JC, Aleixo A, Chang J, et al. A phylogenomic perspective on the robust capuchin monkey (*Sapajus*) radiation: First evidence for extensive population admixture across South America. *Mol Phylogenet Evol*. 2018; 124:137–150. <https://doi.org/10.1016/j.ympev.2018.02.023> PMID: 29545109
27. Wall JD, Schlebusch SA, Alberts SC, Cox LA, Snyder-Mackler N, Nevenon KA, et al. Genomewide ancestry and divergence patterns from low-coverage sequencing data reveal a complex history of admixture in wild baboons. *Mol Ecol*. 2016; 25:3469–3483. <https://doi.org/10.1111/mec.13684> PMID: 27145036
28. Huerta-Sanchez E, Jin X, Asan, Bianba Z, Peter BM, Vinckenbosch N, et al. Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature*. 2014; 512:194–197. <https://doi.org/10.1038/nature13408> PMID: 25043035
29. Racimo F, Sankararaman S, Nielsen R, Huerta-Sanchez E. Evidence for archaic adaptive introgression in humans. *Nat Rev Genet*. 2015; 16:359–371. <https://doi.org/10.1038/nrg3936> PMID: 25963373
30. Racimo F, Gokhman D, Fumagalli M, Ko A, Hansen T, Moltke I, et al. Archaic adaptive introgression in *TBX15/WARS2*. *Mol Biol Evol*. 2017; 34:509–524. <https://doi.org/10.1093/molbev/msw283> PMID: 28007980
31. Chatterjee HJ, Ho SYW, Barnes I, Groves C. Estimating the phylogeny and divergence times of primates using a supermatrix approach. *BMC Evol Biol*. 2009; 9:259. <https://doi.org/10.1186/1471-2148-9-259> PMID: 19860891
32. Herrera JP, Dávalos LM. Phylogeny and divergence times of lemurs inferred with recent and ancient fossils in the tree. *Syst Biol*. 2016; 65:772–791. <https://doi.org/10.1093/sysbio/syw035> PMID: 27113475
33. Kistler L, Ratan A, Godfrey LR, Crowley BE, Hughes CE, Lei R, et al. Comparative and population mitogenomic analyses of Madagascar's extinct, giant 'subfossil' lemurs. *J Hum Evol*. 2015; 79:45–54. <https://doi.org/10.1016/j.jhevol.2014.06.016> PMID: 25523037
34. Tung J, Barreiro LB. The contribution of admixture to primate evolution. *Curr Opin Genet Dev*. 2017; 47:61–68. <https://doi.org/10.1016/j.gde.2017.08.010> PMID: 28923540
35. Stevens NJ, Seiffert ER, O'Connor PM, Roberts EM, Schmitz MD, Krause C, et al. Palaeontological evidence for an Oligocene divergence between Old World monkeys and apes. *Nature*. 2013; 497:611–614. <https://doi.org/10.1038/nature12161> PMID: 23676680

36. Huson DH, Klöpper T, Lockhart PJ, Steel MA. Reconstruction of reticulate networks from gene trees. *Proceedings of RECOMB 2005: The 9th Annual International Conference Research in Computational Molecular Biology*. Berlin: Springer; 2005. pp. 233–249. [https://doi.org/10.1007/11415770\\_18](https://doi.org/10.1007/11415770_18)
37. Waterhouse RM, Seppey M, Simão FA, Manni M, Ioannidis P, Klioutchnikov G, et al. BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol Biol Evol*. 2018; 35:543–548. <https://doi.org/10.1093/molbev/msx319> PMID: 29220515
38. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol*. 2020. <https://doi.org/10.1093/molbev/msaa015> PMID: 32011700
39. Zhang C, Rabiee M, Sayyari E, Mirarab S. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics*. 2018; 19:153. <https://doi.org/10.1186/s12859-018-2129-y> PMID: 29745866
40. Sayyari E, Mirarab S. Fast coalescent-based computation of local branch support from quartet frequencies. *Mol Biol Evol*. 2016; 33:1654–1668. <https://doi.org/10.1093/molbev/msw079> PMID: 27189547
41. Adkins RM, Honeycutt RL. Molecular phylogeny of the superorder Archonta. *Proc Natl Acad Sci U S A*. 1991; 88:10317–10321. <https://doi.org/10.1073/pnas.88.22.10317> PMID: 1658802
42. Arnason U, Adegoke JA, Bodin K, Born EW, Esa YB, Gullberg A, et al. Mammalian mitogenomic relationships and the root of the eutherian tree. *Proc Natl Acad Sci U S A*. 2002; 99:8151–8156. <https://doi.org/10.1073/pnas.102164299> PMID: 12034869
43. Bloch JI, Boyer DM. Grasping Primate origins. *Science*. 2002; 298:1606–1610. <https://doi.org/10.1126/science.1078249> PMID: 12446906
44. Madsen O, Scally M, Douady CJ, Kao DJ, DeBry RW, Adkins R, et al. Parallel adaptive radiations in two major clades of placental mammals. *Nature*. 2001; 409:610–614. <https://doi.org/10.1038/35054544> PMID: 11214318
45. Meredith RW, Janečka JE, Gatesy J, Ryder OA, Fisher CA, Teeling EC, et al. Impacts of the Cretaceous terrestrial revolution and KPg extinction on mammal diversification. *Science*. 2011; 334:521–524. <https://doi.org/10.1126/science.1211028> PMID: 21940861
46. Murphy WJ, Eizirik E, Johnson WE, Zhang YP, Ryder OA, O'Brien SJ. Molecular phylogenetics and the origins of placental mammals. *Nature*. 2001; 409:614–618. <https://doi.org/10.1038/35054550> PMID: 11214319
47. Murphy WJ, Eizirik E, O'Brien SJ, Madsen O, Scally M, Douady CJ, et al. Resolution of the early placental mammal radiation using Bayesian phylogenetics. *Science*. 2001; 294:2348–2351. <https://doi.org/10.1126/science.1067179> PMID: 11743200
48. Novacek MJ. Mammalian phylogeny: shaking the tree. *Nature*. 1992; 356:121–125. <https://doi.org/10.1038/356121a0> PMID: 1545862
49. O'Leary MA, Bloch JI, Flynn JJ, Gaudin TJ, Giallombardo A, Giannini NP, et al. The placental mammal ancestor and the post-K-Pg radiation of placentals. *Science*. 2013; 339:662–667. <https://doi.org/10.1126/science.1229237> PMID: 23393258
50. Poux C, Douzery EJP. Primate phylogeny, evolutionary rate variations, and divergence times: a contribution from the nuclear gene *IRBP*. *Am J Phys Anthropol*. 2004; 124:1–16. <https://doi.org/10.1002/ajpa.10322> PMID: 15085543
51. Janečka JE, Miller W, Pringle TH, Wiens F, Zitzmann A, Helgen KM, et al. Molecular and genomic data identify the closest living relative of primates. *Science*. 2007; 318:792–794. <https://doi.org/10.1126/science.1147555> PMID: 17975064
52. Mason VC, Li G, Minx P, Schmitz J, Churakov G, Doronina L, et al. Genomic analysis reveals hidden biodiversity within colugos, the sister group to primates. *Sci Adv*. 2016; 2:e1600633. <https://doi.org/10.1126/sciadv.1600633> PMID: 27532052
53. Schmitz J, Ohme M, Suryobroto B, Zischler H. The colugo (*Cynocephalus variegatus*, Dermoptera): the primates' gliding sister? *Mol Biol Evol*. 2002; 19:2308–2312. <https://doi.org/10.1093/oxfordjournals.molbev.a004054> PMID: 12446821
54. Jarvis ED, Mirarab S, Aberer AJ, Li B, Houde P, Li C, et al. Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science*. 2014; 346:1320–1331. <https://doi.org/10.1126/science.1253451> PMID: 25504713
55. Pease JB, Haak DC, Hahn MW, Moyle LC. Phylogenomics reveals three sources of adaptive variation during a rapid radiation. *PLoS Biol*. 2016; 14:e1002379. <https://doi.org/10.1371/journal.pbio.1002379> PMID: 26871574
56. Salichos L, Rokas A. Inferring ancient divergences requires genes with strong phylogenetic signals. *Nature*. 2013; 497:327–331. <https://doi.org/10.1038/nature12130> PMID: 23657258

57. Minh BQ, Hahn MW, Lanfear R. New methods to calculate concordance factors for phylogenomic datasets. *bioRxiv*. 2018;487801. <https://doi.org/10.1101/487801>
58. Yoder AD. The phylogenetic position of genus *Tarsius*: whose side are you on? In: Wright PC, Simons EL, Gursky S, editors. *Tarsiers Past, Present, and Future*. Rutgers University Press; 2003.
59. Gregory WK. On the classification and phylogeny of the Lemuroidea. *Bull Geol Soc Am*. 1915:426–446.
60. Pocock RI. On the external characters of the lemurs and of *Tarsius*. *Proc Zool Soc Lond* 1918; 88: 19–53. <https://doi.org/10.1111/j.1096-3642.1918.tb02076.x>
61. Hartig G, Churakov G, Warren WC, Brosius J, Makalowski W, Schmitz J. Retrophylogenomics place tarsiers on the evolutionary branch of anthropoids. *Sci Rep*. 2013; 3:1756. <https://doi.org/10.1038/srep01756> PMID: 23629008
62. Jameson NM, Hou Z-C, Sterner KN, Weckle A, Goodman M, Steiper ME, et al. Genomic data reject the hypothesis of a prosimian primate clade. *J Hum Evol*. 2011; 61:295–305. <https://doi.org/10.1016/j.jhevol.2011.04.004> PMID: 21620437
63. Hayasaka K, Gojobori T, Horai S. Molecular phylogeny and evolution of primate mitochondrial DNA. *Mol Biol Evol*. 1988; 5:626–644. <https://doi.org/10.1093/oxfordjournals.molbev.a040524> PMID: 3146681
64. Jaworski CJ. A reassessment of mammalian alpha A-crystallin sequences using DNA sequencing: implications for anthropoid affinities of tarsier. *J Mol Evol*. 1995; 41:901–908. <https://doi.org/10.1007/BF00173170> PMID: 8587135
65. Whitfield JB, Lockhart PJ. Deciphering ancient rapid radiations. *Trends Ecol Evol*. 2007; 22:258–265. <https://doi.org/10.1016/j.tree.2007.01.012> PMID: 17300853
66. Bond M, Tejedor MF, Campbell KE, Chornogubsky L, Novo N, Goin F. Eocene primates of South America and the African origins of New World monkeys. *Nature*. 2015; 520:538–541. <https://doi.org/10.1038/nature14120> PMID: 25652825
67. Kubatko LS, Degnan JH. Inconsistency of phylogenetic estimates from concatenated data under coalescence. *Syst Biol*. 2007; 56:17–24. <https://doi.org/10.1080/10635150601146041> PMID: 17366134
68. Roch S, Steel M. Likelihood-based tree reconstruction on a concatenation of aligned sequence data sets can be statistically inconsistent. *Theor Popul Biol*. 2015; 100C:56–62. <https://doi.org/10.1016/j.tpb.2014.12.005> PMID: 25545843
69. Warnow T. Concatenation analyses in the presence of incomplete lineage sorting. *PLoS Curr*. 2015; 7. <https://doi.org/10.1371/currents.tol.8d41ac0f13d1abedf4c4a59f5d17b1f7> PMID: 26064786
70. Bryant D, Hahn MW. The concatenation question. In: Scornavacca C, Delsuc F, Galtier N, editors. *Phylogenetics in the genomic era*. No commercial publisher | Authors open access book; 2020. p. 3.4:1–3.4:23. Available from: <https://hal.archives-ouvertes.fr/hal-02535651>.
71. Sayyari E, Mirarab S. Testing for polytomies in phylogenetic species trees using quartet frequencies. *Genes*. 2018; 9:132. <https://doi.org/10.3390/genes9030132> PMID: 29495636
72. Liu L, Edwards SV. Phylogenetic analysis in the anomaly zone. *Syst Biol*. 2009; 58:452–460. <https://doi.org/10.1093/sysbio/syp034> PMID: 20525599
73. Swofford DL. *PAUP\*: Phylogenetic Analysis Using Parsimony (\*and other methods)*, Version 4. Sinauer Assoc Sunderland Mass. 2002.
74. Mendes FK, Livera AP, Hahn MW. The perils of intralocus recombination for inferences of molecular convergence. *Philos Trans R Soc Lond Ser B Biol Sci*. 2019; 374:20180244. <https://doi.org/10.1098/rstb.2018.0244> PMID: 31154973
75. Springer MS, Gatesy J. The gene tree delusion. *Mol Phylogenet Evol*. 2016; 94:1–33. <https://doi.org/10.1016/j.ympev.2015.07.018> PMID: 26238460
76. Hoang DT, Vinh LS, Flouri T, Stamatakis A, von Haeseler A, Minh BQ. MPBoot: fast phylogenetic maximum parsimony tree inference and bootstrap approximation. *BMC Evol Biol*. 2018; 18:11–11. <https://doi.org/10.1186/s12862-018-1131-3> PMID: 29390973
77. He C, Liang D, Zhang P. Asymmetric distribution of gene trees can arise under purifying selection if differences in population size exist. *Mol Biol Evol*. 2020; 37:881–892. <https://doi.org/10.1093/molbev/msz232> PMID: 31638156
78. Golding GB. The effect of purifying selection on genealogies. In: Donnelly P, Tavaré S, editors. *Progress in population genetics and human evolution*. New York: Springer Verlag; 1997. pp. 271–285.
79. Przeworski M, Charlesworth B, Wall JD. Genealogies and weak purifying selection. *Mol Biol Evol*. 1999; 16:246–252. <https://doi.org/10.1093/oxfordjournals.molbev.a026106> PMID: 10084898
80. Slade PF. Most recent common ancestor probability distributions in gene genealogies under selection. *Theor Popul Biol*. 2000; 58:291–305. <https://doi.org/10.1006/tpbi.2000.1488> PMID: 11162788



81. Williamson S, Orive ME. The genealogy of a sequence subject to purifying selection at multiple sites. *Mol Biol Evol.* 2002; 19:1376–1384. <https://doi.org/10.1093/oxfordjournals.molbev.a004199> PMID: 12140250
82. Haller BC, Messer PW. SLiM 3: Forward genetic simulations beyond the Wright–Fisher model. *Mol Biol Evol.* 2019; 36:632–637. <https://doi.org/10.1093/molbev/msy228> PMID: 30517680
83. Mendes FK, Hahn MW. Gene tree discordance causes apparent substitution rate variation. *Syst Biol.* 2016; 65:711–721. <https://doi.org/10.1093/sysbio/syw018> PMID: 26927960
84. Brunet M, Guy F, Pilbeam D, Mackaye HT, Likius A, Ahounta D, et al. A new hominid from the Upper Miocene of Chad. *Nature.* 2002; 418:145–151. <https://doi.org/10.1038/nature00879> PMID: 12110880
85. Sigé B, Jaeger J-J, Sudre J, Vianey-Liaud M. *Altiatlasius koulchii* n. gen. et sp., primate omomyidé du Paléocène supérieur du Maroc, et les origines des euprimates. *Palaeontogr Abt A.* 1990:31–56.
86. Lartillot N, Lepage T, Blanquart S. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics.* 2009; 25:2286–2288. <https://doi.org/10.1093/bioinformatics/btp368> PMID: 19535536
87. Wilkinson RD, Steiper ME, Soligo C, Martin RD, Yang Z, Tavaré S. Dating primate divergences through an integrated analysis of palaeontological and molecular data. *Syst Biol.* 2011; 60:16–31. <https://doi.org/10.1093/sysbio/syq054> PMID: 21051775
88. Goodman M, Porter CA, Czelusniak J, Page SL, Schneider H, Shoshani J, et al. Toward a phylogenetic classification of Primates based on DNA evidence complemented by fossil evidence. *Mol Phylogenet Evol.* 1998; 9:585–598. <https://doi.org/10.1006/mpev.1998.0495> PMID: 9668008
89. Yoder AD, Yang Z. Divergence dates for Malagasy lemurs estimated from multiple gene loci: geological and evolutionary context. *Mol Ecol.* 2004; 13:757–773. <https://doi.org/10.1046/j.1365-294x.2004.02106.x> PMID: 15012754
90. Benton MJ, Donoghue PCJ, Asher RJ, Friedman M, Near TJ, Vinther J. Constraints on the timescale of animal evolutionary history. *Palaeontol Electron.* 2015; 18.1.1FC:1–106. <https://doi.org/10.26879/424>
91. Edwards SV, Beerli P. Perspective: gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. *Evolution.* 2000; 54:1839–1854. <https://doi.org/10.1111/j.0014-3820.2000.tb01231.x> PMID: 11209764
92. Rogers J. Levels of the genealogical hierarchy and the problem of hominoid phylogeny. *Am J Phys Anthropol.* 1994; 94:81–88. <https://doi.org/10.1002/ajpa.1330940107> PMID: 8042707
93. Carbone L, Harris RA, Gnerre S, Veeramah KR, Lorente-Galdos B, Huddleston J, et al. Gibbon genome and the fast karyotype evolution of small apes. *Nature.* 2014; 513:195–201. <https://doi.org/10.1038/nature13679> PMID: 25209798
94. Veeramah KR, Woerner AE, Johnstone L, Gut I, Gut M, Marques-Bonet T, et al. Examining phylogenetic relationships among gibbon genera using whole genome sequence data using an approximate bayesian computation approach. *Genetics.* 2015; 200:295–308. <https://doi.org/10.1534/genetics.115.174425> PMID: 25769979
95. Hamada Y, San AM, Malaivijitnond S. Assessment of the hybridization between rhesus (*Macaca mulatta*) and long-tailed macaques (*M. fascicularis*) based on morphological characters. *Am J Phys Anthropol.* 2016; 159:189–198. <https://doi.org/10.1002/ajpa.22862> PMID: 26354607
96. Osada N, Uno Y, Mineta K, Kameoka Y, Takahashi I, Terao K. Ancient genome-wide admixture extends beyond the current hybrid zone between *Macaca fascicularis* and *M. mulatta*. *Mol Ecol.* 2010; 19:2884–2895. <https://doi.org/10.1111/j.1365-294X.2010.04687.x> PMID: 20579289
97. Hahn MW. *Molecular population genetics.* 1st ed. Oxford, New York: Oxford University Press; 2018.
98. Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, et al. Ancient admixture in human history. *Genetics.* 2012; 192:1065–1093. <https://doi.org/10.1534/genetics.112.145037> PMID: 22960212
99. Fan Z, Zhao G, Li P, Osada N, Xing J, Yi Y, et al. Whole-genome sequencing of Tibetan macaque (*Macaca thibetana*) provides new insight into the macaque evolutionary history. *Mol Biol Evol.* 2014; 31:1475–1489. <https://doi.org/10.1093/molbev/msu104> PMID: 24648498
100. Yan G, Zhang G, Fang X, Zhang Y, Li C, Ling F, et al. Genome sequencing and comparison of two nonhuman primate animal models, the cynomolgus and Chinese rhesus macaques. *Nat Biotechnol.* 2011; 29:1019–1023. <https://doi.org/10.1038/nbt.1992> PMID: 22002653
101. Koufos GD. Potential hominoid ancestors for Hominidae. In: Henke W, Tattersall I, editors. *Handbook of Paleoanthropology.* Berlin, Heidelberg: Springer; 2007. pp. 1761–1790. [https://doi.org/10.1007/978-3-642-39979-4\\_44](https://doi.org/10.1007/978-3-642-39979-4_44)
102. Patterson N, Richter DJ, Gnerre S, Lander ES, Reich D. Genetic evidence for complex speciation of humans and chimpanzees. *Nature.* 2006; 441:1103–1108. <https://doi.org/10.1038/nature04789> PMID: 16710306

103. Fleagle JG. Apes and humans. *Primate Adaptation & Evolution*. Elsevier; 2013. pp. 151–168. <https://doi.org/10.1016/B978-0-12-378632-6.00007-0>
104. Than C, Ruths D, Nakhleh L. PhyloNet: a software package for analyzing and reconstructing reticulate evolutionary relationships. *BMC Bioinformatics*. 2008; 9:322. <https://doi.org/10.1186/1471-2105-9-322> PMID: 18662388
105. Solís-Lemus C, Ané C. Inferring phylogenetic networks with maximum pseudolikelihood under incomplete lineage sorting. *PLoS Genet*. 2016; 12:e1005896. <https://doi.org/10.1371/journal.pgen.1005896> PMID: 26950302
106. Roos C, Kothe M, Alba DM, Delson E, Zinner D. The radiation of macaques out of Africa: Evidence from mitogenome divergence times and the fossil record. *J Hum Evol*. 2019; 133:114–132. <https://doi.org/10.1016/j.jhevol.2019.05.017> PMID: 31358175
107. Belmaker M. The presence of a large cercopithecine (cf. *Theropithecus* sp.) in the ‘Ubeidiya formation (Early Pleistocene, Israel). *J Hum Evol*. 2010; 58:79–89. <https://doi.org/10.1016/j.jhevol.2009.08.004> PMID: 19875150
108. Hughes JK, Elton S, O’Regan HJ. *Theropithecus* and “Out of Africa” dispersal in the Plio-Pleistocene. *J Hum Evol*. 2008; 54:43–77. <https://doi.org/10.1016/j.jhevol.2007.06.004> PMID: 17868778
109. Larrasoaña JC, Roberts AP, Rohling EJ, Winkelhofer M, Wehausen R. Three million years of monsoon variability over the northern Sahara. *Clim Dyn*. 2003; 21:689–698. <https://doi.org/10.1007/s00382-003-0355-z>
110. Larrasoaña JC, Roberts AP, Rohling EJ. Dynamics of green Sahara periods and their role in hominin evolution. *PLoS ONE*. 2013; 8:e76514. <https://doi.org/10.1371/journal.pone.0076514> PMID: 24146882
111. Vaks A, Woodhead J, Bar-Matthews M, Ayalon A, Cliff RA, Zilberman T, et al. Pliocene–Pleistocene climate of the northern margin of Saharan–Arabian Desert recorded in speleothems from the Negev Desert, Israel. *Earth Planet Sci Lett*. 2013; 368:88–100. <https://doi.org/10.1016/j.epsl.2013.02.027>
112. Coulthard TJ, Ramirez JA, Barton N, Rogerson M, Brücher T. Were rivers flowing across the Sahara during the last interglacial? Implications for human migration through Africa. *PLoS ONE*. 2013; 8:e74834. <https://doi.org/10.1371/journal.pone.0074834> PMID: 24040347
113. Sahnouni M, Parés JM, Duval M, Cáceres I, Harichane Z, van der Made J, et al. 1.9-million- and 2.4-million-year-old artifacts and stone tool-cutmarked bones from Ain Boucherit, Algeria. *Science*. 2018; 362(6420):1297–1301. <https://doi.org/10.1126/science.aau0008> PMID: 30498166
114. de Menocal PB. African climate change and faunal evolution during the Pliocene–Pleistocene. *Earth Planet Sci Lett*. 2004; 220:3–24. [https://doi.org/10.1016/S0012-821X\(04\)00003-2](https://doi.org/10.1016/S0012-821X(04)00003-2)
115. Slatkin M, Pollack JL. Subdivision in an ancestral species creates asymmetry in gene trees. *Mol Biol Evol*. 2008; 25:2241–2246. <https://doi.org/10.1093/molbev/msn172> PMID: 18689871
116. Kuritzin A, Kischka T, Schmitz J, Churakov G. Incomplete lineage sorting and hybridization statistics for large-scale retroposon insertion data. *PLoS Comput Biol*. 2016; 12:e1004812. <https://doi.org/10.1371/journal.pcbi.1004812> PMID: 26967525
117. Springer MS, Molloy EK, Sloan DB, Simmons MP, Gatesy J. ILS-aware analysis of low-homoplasy retroelement insertions: Inference of species trees and introgression using quartets. *J Hered*. 2020; 111:147–168. <https://doi.org/10.1093/jhered/esz076> PMID: 31837265
118. Gernhard T. New analytic results for speciation times in neutral models. *Bull Math Biol*. 2008; 70:1082–1097. <https://doi.org/10.1007/s11538-007-9291-0> PMID: 18172736
119. Gligor M, Ganzhorn JU, Rakotondravony D, Ramilijaona OR, Razafimahatratra E, Zischler H, et al. Hybridization between mouse lemurs in an ecological transition zone in southern Madagascar. *Mol Ecol*. 2009; 18:520–533. <https://doi.org/10.1111/j.1365-294X.2008.04040.x> PMID: 19161471
120. Pastorini J, Zaramody A, Curtis DJ, Nievergelt CM, Mundy NI. Genetic analysis of hybridization and introgression between wild mongoose and brown lemurs. *BMC Evol Biol*. 2009; 9:32. <https://doi.org/10.1186/1471-2148-9-32> PMID: 19196458
121. Williams RC, Blanco MB, Poelstra JW, Hunnicutt KE, Comeault AA, Yoder AD. Conservation genomic analysis reveals ancient introgression and declining levels of genetic diversity in Madagascar’s hibernating dwarf lemurs. *Heredity*. 2020; 124:236–251. <https://doi.org/10.1038/s41437-019-0260-9> PMID: 31435007
122. Wyner YM, Johnson SE, Stumpf RM, Desalle R. Genetic assessment of a white-collared×red-fronted lemur hybrid zone at Andringitra, Madagascar. *Am J Primatol*. 2002; 57:51–66. <https://doi.org/10.1002/ajp.10033> PMID: 12111681
123. Gnerre S, MacCallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, et al. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc Natl Acad Sci U S A*. 2011; 108:1513–1518. <https://doi.org/10.1073/pnas.1017351108> PMID: 21187386

124. English AC, Richards S, Han Y, Wang M, Vee V, Qu J, et al. Mind the gap: upgrading genomes with Pacific Biosciences RS long-read sequencing technology. *PLoS ONE*. 2012; 7:e47768. <https://doi.org/10.1371/journal.pone.0047768> PMID: 23185243
125. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990; 215:403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2) PMID: 2231712
126. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. *BMC Bioinformatics*. 2009; 10:421. <https://doi.org/10.1186/1471-2105-10-421> PMID: 20003500
127. van Dongen S. Graph clustering by flow simulation. PhD thesis, University of Utrecht. 2000. Available from: <http://www.library.uu.nl/digiarchief/dip/diss/1895620/full.pdf>.
128. Wattam AR, Abraham D, Dalay O, Disz TL, Driscoll T, Gabbard JL, et al. PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic Acids Res*. 2014; 42:D581–D591. <https://doi.org/10.1093/nar/gkt1099> PMID: 24225323
129. Sela I, Ashkenazy H, Katoh K, Pupko T. GUIDANCE2: accurate detection of unreliable alignment regions accounting for the uncertainty of multiple parameters. *Nucleic Acids Res*. 2015; 43:W7–W14. <https://doi.org/10.1093/nar/gkv318> PMID: 25883146
130. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013; 30:772–780. <https://doi.org/10.1093/molbev/mst010> PMID: 23329690
131. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. 2009; 25:1972–1973. <https://doi.org/10.1093/bioinformatics/btp348> PMID: 19505945
132. Chernomor O, von Haeseler A, Minh BQ. Terrace aware data structure for phylogenomic inference from supermatrices. *Syst Biol*. 2016; 65:997–1008. <https://doi.org/10.1093/sysbio/syw037> PMID: 27121966
133. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol*. 2018; 35:518–522. <https://doi.org/10.1093/molbev/msx281> PMID: 29077904
134. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods*. 2017; 14:587–589. <https://doi.org/10.1038/nmeth.4285> PMID: 28481363
135. Eaton DAR, Ree RH. Inferring phylogeny and introgression using RADseq data: an example from flowering plants (*Pedicularis*: Orobanchaceae). *Syst Biol*. 2013; 62:689–706. <https://doi.org/10.1093/sysbio/syt032> PMID: 23652346
136. Dunn OJ. Confidence intervals for the means of dependent, normally distributed variables. *J Am Stat Assoc*. 1959; 54:613–621. <https://doi.org/10.2307/2282541>
137. Sidak Z. Rectangular confidence regions for the means of multivariate normal distributions. *J Am Stat Assoc*. 1967; 62:626–633. <https://doi.org/10.2307/2283989>
138. Lartillot N, Philippe H. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol Biol Evol*. 2004; 21:1095–1109. <https://doi.org/10.1093/molbev/msh112> PMID: 15014145
139. Thorne JL, Kishino H, Painter IS. Estimating the rate of evolution of the rate of molecular evolution. *Mol Biol Evol*. 1998; 15:1647–1657. <https://doi.org/10.1093/oxfordjournals.molbev.a025892> PMID: 9866200