Received: 2 December 2020 | Revised: 4 October 2021 | Accepted: 12 January 2022

DOI: 10.1002/evan.21940

REVIEW ARTICLE

Evolutionary Anthropology ISSUES, NEWS, AND REVIEWS WILEY

Rocks and clocks revised: New promises and challenges in dating the primate tree of life

Luca Pozzi | Anna Penna

Department of Anthropology, The University of Texas at San Antonio, San Antonio, Texas, USA

Correspondence
Luca Pozzi, Department of Anthropology, The University of Texas at San Antonio, San Antonio, TX 78249, USA.
Email: luca.pozzi@utsa.edu

Funding information
Division of Behavioral and Cognitive Sciences, NSF, Grant/Award Number: BCS-1926105

Abstract

In recent years, multiple technological and methodological advances have increased our ability to estimate phylogenies, leading to more accurate dating of the primate tree of life. Here we provide an overview of the limitations and potentials of some of these advancements and discuss how dated phylogenies provide the crucial temporal scale required to understand primate evolution. First, we review new methods, such as the *total-evidence dating* approach, that promise a better integration between the fossil record and molecular data. We then explore how the ever-increasing availability of genomic-level data for more primate species can impact our ability to accurately estimate timetrees. Finally, we discuss more recent applications of mutation rates to date divergence times. We highlight example studies that have applied these approaches to estimate divergence dates within primates. Our goal is to provide a critical overview of these new developments and explore the promises and challenges of their application in evolutionary anthropology.

KEYWORDS
fossil record, molecular clock, node dating, phylogenomics, total-evidence dating

1 | INTRODUCTION

The applications of dated phylogenetic trees to evolutionary biology and anthropology are multi-fold.[1–5] Because dated phylogenies represent both the historical relationships between taxa and the timing of cladogenetic events, they are critical for understanding patterns and processes of species diversification and adaptation. Phylogenies are thus commonly used to address questions in biogeography, comparative analyses, behavioral ecology, and many other fields.[1,3,6,7] For instance, divergence times between clades are pivotal to understanding the role of geological or climatic events in the diversification of a taxonomic group because these age estimates provide a timeline for biological events.[6] Similarly, dated phylogenies can provide essential evidence in understanding the timing of colonization of a particular island[8,9] or the impact of human activity in species extinction.[10,11]

Over the last decade, several technological and methodological advances in the field of phylogenetics have opened up new opportunities to study phylogenies. These advances are leading to a better understanding of the primate tree of life and more precise and (arguably) more accurate time estimates within the major primate clades. Two advancements are specifically worth noticing. First, new computational and modeling approaches have allowed researchers to better integrate information from the fossil record with molecular data, mitigating some of the main shortcomings of past studies.[12–16] Second, the promises of more efficient and affordable sequencing technologies have contributed to the rapid accumulation of a vast amount of genetic data for a growing number of primate species.[17–20] Now, evidence from hundreds or even thousands of loci across the genome, not only from one or a few genes, can be combined with morphological and fossil data to simultaneously estimate tree topology and divergence times.[21–23] These recent developments generated great excitement in the field of molecular anthropology, phylogenetics, and paleontology. Here we present a review of some of these new techniques and explore the promises and challenges of their application in primatology and evolutionary anthropology, in general.

2 | HOW TO DATE A TREE: THE INTEGRATION OF FOSSILS AND MOLECULAR DATA

2.1 | Rocks, clocks, and molecules

Historically, morphological evidence dominated taxonomic and phylogenetic studies, and divergence time was mainly inferred from the fossil record or geological events. With the advent of molecular techniques and DNA sequencing in the late 1980s and early 1990s, integrating molecular and paleontological data became standard practice in phylogenetics. DNA sequence data are used to estimate the degree of genetic similarity between taxa, and paleontological evidence can then be used to transform these genetic differences in absolute dates.[5,24–26] Molecules can serve as a clock, as the difference—that is, number of substitutions—between a pair of sequences is expected to increase with their divergence time.[25,26] To convert the level of genetic divergence between taxa into a chronological scale, the nucleotide substitution rate must be adjusted using external evidence about the time since these two taxa shared a common ancestor. Fossils represent the only direct evidence of past lifeforms and some specimens can be dated with confidence using different geologic methods (e.g., radiometric techniques, luminescence dating, stratigraphy). Therefore, paleontological evidence has classically been used as the external time information to convert the relative time of divergence given by the rate of molecular substitution into absolute time.[5,24–26]

The most common method to integrate fossil data in molecular phylogenies is the "node-dating" approach (Figure 1a–c). This method combines taxonomic and dating interpretations of the fossil record to inform the broad intervals that determine the age of some internal nodes in a tree.[27–29] Node dating consists of a stepwise analysis to obtain divergence times: first, calibration priors are derived from the fossil record and associated with specific nodes[5,27,28]; then, these priors are combined with molecular data to estimate the ages of internal nodes of a phylogenetic tree.[16] However, it is misleading to assume that the age of a given fossil represents the exact timing of the splitting event between two lineages—even when evidence suggests it was their direct ancestor. Therefore, only the oldest discovered representative of a clade is used to constrain the minimum age of a node (Figure 1a). In this case, the oldest-secure record should be preferred over the oldest possible date of a fossil.[28] Although fossil

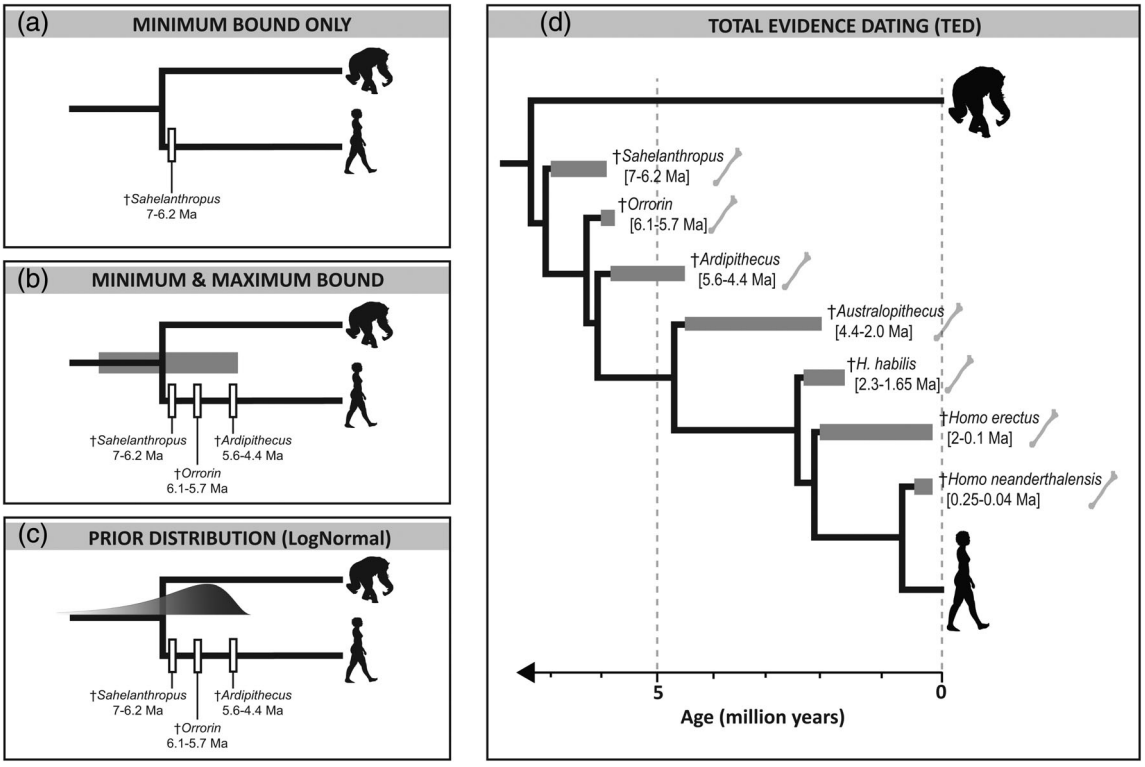

FIGURE 1 Examples of various methods to calibrate a tree. On the *left* (a–c) three examples of node dating calibrations using (a) only the oldest known fossil assigned to a particular node (minimum bound only); (b) all the information available to that node and assigning both a hard minimum and maximum bound; and (c) a prior probability distribution (e.g., lognormal) that includes not only the minimum estimate for a particular node but also an estimate of the likelihood that the real divergence falls within this range of dates. On the *right* (d), an example of total evidence dating for the split between humans and chimps. Chronologic information from all known fossil taxa is incorporated into a single total-evidence dating (TED) analysis. Both extant and extinct taxa represent tips of the tree. This graph is only meant to represent the concept behind TED analyses, and it is not supposed to represent an accurate and/or complete representation of the hominin fossil record. Only a few lineages are reported for clarity

data can inform the most recent age of a clade,[3,25,28,30–32] there is little evidence that this information can be used to conclusively identify the maximum age of a clade.[3,25,31,33]

The node-dating approach relies on a series of a priori assumptions about the accuracy of relationships inferred in phylogenetic reconstructions, the age, and phylogenetic placement of fossils. Consequently, topological uncertainties and alternative interpretations of the fossil record can impact the calibration of molecular trees.[3,5,15,34] Among those, the selection of inappropriate fossils (e.g., not the oldest-secure record) and the phylogenetic misplacement of a fossil (e.g., placing a stem taxon in a crown group) can lead to broader confidence intervals of node ages and topological congruence.[3,5,29] By its nature, the paleontological record is highly incomplete and often biased toward certain body parts (e.g., craniodental remains), environmental conditions (e.g., river sediments), geological periods, and even taxa. For instance, while the fossil record of Miocene apes from Europe and Africa is relatively abundant,[35] lemurs are entirely missing from the paleontological record in Madagascar between the Late Cretaceous and the end of the Quaternary—the only exception being the very recent subfossil taxa (17 species, <4000 years old).[36] Such incompleteness poses critical challenges in identifying appropriate fossils for node calibration, and many studies have employed either questionable fossils or unreliable dates based on secondary calibration points (e.g., estimates obtained by other molecular studies).[31] While only a few fossil specimens are used in the node calibration method, the fossil record can provide richer information than just a constraint for the minimum age of a given node—even if incomplete. Nevertheless, neither diversity of stem fossils nor morphological data are accounted for in the node-dating method. Consequently, neontological data coming from molecular evidence dominate tree topology and age calibration results.

2.2 | A new way to calibrate trees: The total-evidence dating approach

The node-dating approach has several limitations. First, only a small fraction of the fossil record is used to inform the minimum age of a clade (Figures 1 and 2). Second, it uses a fixed tree topology derived from molecular data alone[16]: calibration points must be associated with specific nodes in the tree, although we cannot be absolutely certain of the accuracy of any of those nodes. One way to alleviate the first problem is to incorporate additional paleontological information from multiple fossil specimens to draw more informative prior distributions for a given node. Although no real fossil evidence can provide information about the oldest age of a node,[33] researchers can assign a very low probability to divergence ages that seems unrealistically old based on the current knowledge of the group of interest (e.g., earliest stem member of a clade, abundance over time, etc.).[16,25,31] Such soft-bounded calibration can also be useful in addressing uncertainties regarding the age of a specific node.[25,31] For instance, the paleontological record indicates that the most recent common ancestor of humans and chimpanzees lived prior to ~5 Ma (e.g., *Ardipithecus*), and there are no fossils that suggest a common ancestor prior to about 7 Ma (e.g., *Sahelanthropus tchadensis*). Based on this knowledge, many researchers have employed 95% prior distributions that range between ~5 and 10 Ma.[31] These probability distributions represent the likelihood that the real divergence falls within this date range (Figure 1b,c). Although the lower bound can be justified by the presence of several fossils currently assigned to the hominin lineage, the upper soft bound is based on the absence of evidence in the paleontological record that indicates the split between *Homo* and *Pan* to be unlikely older than 10 Ma.[3,31] However, these prior distributions are usually drawn based on subjective or arbitrary criteria and not necessarily informed by biological processes.[12,25,37–39] Indeed, remarkably different calibration distributions used in various studies are often supported by the same paleontological information.[31]

An additional limitation of using only a few fossils to inform the most recent age of a clade is related to the incomplete nature of the fossil record: since the oldest fossil species recovered for any taxonomic group is likely much younger than the emergence of the lineage, a large percentage of the true stratigraphic range of the group is missing.[40–42] One of the first attempts to address this challenge was proposed by Tavare et al.[41] and then further developed by Wilkinson et al.[2] These authors combined the number of preserved primate species discovered in the fossil record and their geological age distribution, together with the number of extant primate species to provide estimates of the primate divergence times.[2] This approach aimed to better account for the rates of fossil preservation and discovery when calibrating molecule-based trees.

To incorporate other potentially useful temporal, topological, and morphological information from younger fossils other than a clade's oldest assignable fossil, an alternative method was recently proposed: the *total-evidence dating* (TED).[15,38] Total-evidence approaches have long been used in phylogenetics to analyze fossil and extant taxa together, combining morphological and molecular data. These methods have allowed researchers to identify the placement and evaluate the impact of fossil taxa in phylogenetic reconstructions (e.g., References 43–46). Although the resulting trees have also been used in classical node-dating analyses, until recently, no studies have directly used combined dataset to infer fossil placement and calibrate the tree at the same time. The TED approach—first proposed by Pyron[47] and then further developed by Ronquist et al.[16]—holds the promise to overcome some of the limitations in node-dating by incorporating a wide range of age and taxonomically informative data into a single analysis[15,16,47] (Figure 2). TED uses morphological data from the whole dataset (fossil and living species) and applies models of morphological evolution to estimate the relative phylogenetic relationship and time of divergence between living and extinct clades[13,15,16] (Figure 1d). Compared to node-dating, TED approaches are meant to better account for datasets in which the fossil record is poorly preserved and of difficult phylogenetic interpretation.[15,46]

There are two main models in total evidence approaches, the tip-dating[15] and the fossilized birth-death process.[12] While the tip-dating approach uses a total evidence approach to jointly infer the topology and divergence times of living and extinct species (accounting for both

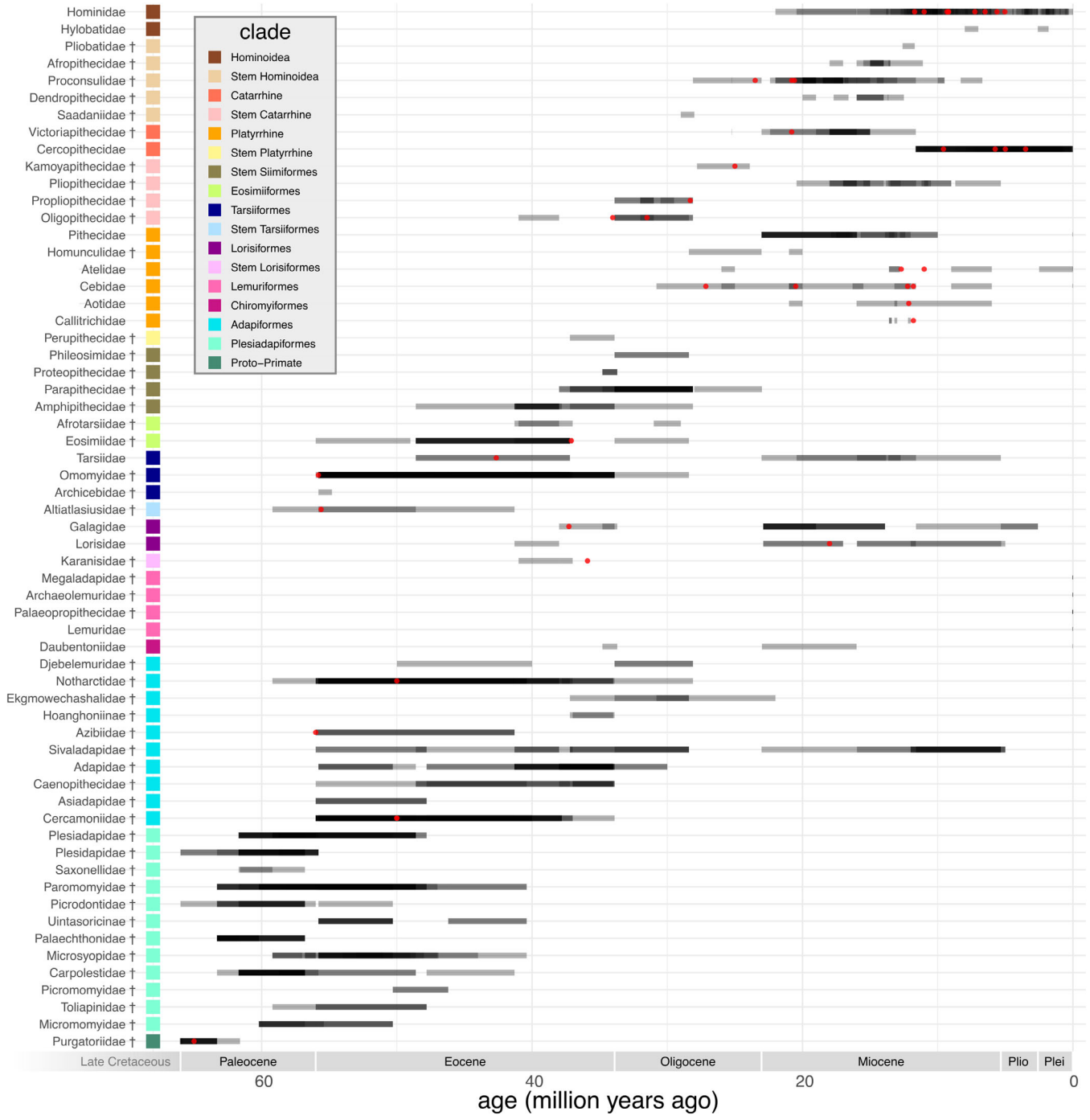

FIGURE 2 Age of the primate fossil record. Each row indicates the diversity of fossil for a given family, † indicates those known from extinct taxa only. Gray bars represent the age interval (minimum and maximum estimate in millions of years) for a given fossil species. Darker areas in the bars indicate higher overlap between the age ranges of fossil species. Red dots indicate values commonly used as lower bounds in node calibrations based on the age of a given fossil taxa. Fossil age ranges were compiled from the main literature, Fleagle (2013),[48] and the Paleobiology Database (https://paleobiodb.org/, downloaded in August 2020). Node calibration information summarized from tables reported by Raaum,[31] Pozzi et al.,[3] and dos Reis et al. (2018)[21]

morphological and molecular data), the fossilized birth-death approach parameterizes the branching process of the phylogeny based on speciation and extinction rates estimated from the fossil record. Because TED approaches account for both stem and crown fossils, they have the potential to provide more reliable age estimates and to reduce the level of discordance between molecular and paleontological estimates observed in many node-dating studies.[16,32]

In the tip-dating approach, both extant and extinct species are considered as terminal taxa (Figure 1d), allowing researchers to incorporate chronologic information from all taxa into a single analysis.[15,47]

This method calibrates the branch lengths by analyzing the morphological dataset for all taxa under a morphological character evolution model and molecular data—when available—usually only for living or recently extinct taxa. Heath et al.[12] introduced the fossilized birth-death process, a Bayesian approach that explicitly integrates the fossils and extant species under the same macroevolutionary framework. The fossilized birth-death estimates the tree topology and divergence times by simultaneously modeling fossil distribution over time and lineage divergence, thus, eliminating the need for ad hoc calibrations (such as in the node-dating). The parameters of this model are the rates of speciation, extinction, preservation, discovery, and sampled species, which interact to inform the uncertainty around the estimates of phylogenetic affinity and lineage divergence time.[12,13,16] The fossilized birth-death model assumes that fossil and molecular data are different observations from the same diversification process, and a model of speciation, fossilization, and preservation is then used as a prior to estimate the divergence times. The fossilized birth-death can also account for uncertainty in the relative positioning of fossils, as it assigns them to clades with varying precisions. The original implementation of fossilized birth-death was performed with molecular data alone and only considered the age of each fossil.[12] In this case, the relationships between fossil and their extant relatives were given a priori, leading to limitations similar to the node dating approach. More recent implementations of the method combined the fossilized birth-death process with tip dating, allowing the integration of morphological and molecular dating, as well as the incorporation of different speciation models (anagenesis and budding evolution).[37,49,50] One advantage of these modifications is the ability to integrate uncertainties about the fossil topology and divergence time, thus accounting for the possibility that a given fossil is indeed the ancestor of an extant lineage (i.e., the timing of fossil divergence is equal to its age).[12,13,16,51] Besides illustrating how previously reported incongruence between fossils and molecular phylogenies can be informative, these more recent implementations of the fossilized birth-death process are contributing to our understanding of the role of speciation and extinction in evolution.[40,50]

2.3 | Challenges of the TED approach

The major advantage of TED is the ability to incorporate more comprehensive information from the paleontological record[13,15,16] (Figure 2). Ideally, this information should minimize uncertainty about time estimates, and mitigate arbitrariness in building prior distributions when constraining node ages. Another promise of TED is to reduce the gap between age estimates drawn by the paleontological record and inferred from molecular analyses. However, the uncertainty around divergence times obtained using TED is often higher than estimates from the classic node-dating approach. More specifically, TED tends to estimate much older and often very imprecise divergence times.[37,52,53] This phenomenon—called by Ronquist et al. "deep root attraction"[16]—is probably due to inherent challenges of effectively modeling morphological evolution. Some of these problems can be alleviated by introducing an extra penalty for ghost lineages (i.e., a hypothesized ancestor in a species lineage that has left no fossil evidence),[16] but more realistic models of morphological evolution are yet challenging to implement.[54]

One of TED's major challenges is due to the fragmentary nature of the fossil record. Although morphology can be extremely useful in placing taxa on a phylogenetic tree, total evidence matrices usually have high levels of missing data biased toward extinct taxa. Most fossils are only known from a few morphological traits, often with little overlap across different taxa. For instance, while some fossils might only be represented by dental traits, others might only have postcranial information. While missing data have been indicated as a possible source of noise in phylogenetic analyses (see below), some recent studies suggest that the proportion of missing data for a taxon not necessarily impact its accurate phylogenetic placement, as long as enough morphological characters are included in the analysis to accurately place the incomplete taxa on the tree.[55] Moreover, increasing the morphological sampling for various living taxa can potentially circumvent topological uncertainty, as it integrates the phylogenetic signal arising from the morphological data together with that coming from molecular data.[56]

Another concern about the application of morphological data to estimate phylogenetic relationships is the presence of homoplastic traits (i.e., similar due to convergent ecological adaptations rather than genealogical relationships[16,37]). Within placental mammals, morphological analyses are known to identify groupings that reflect convergent adaptation to similar behaviors, ecologies, or life histories and have weak phylogenetic signal.[16,37,57] For instance, once considered a monophyletic clade, the Insectivora order encompassed a variety of small and relatively unspecialized mammals that feed upon insects. However, molecular analyses showed that these taxa belong to multiple independent lineages within the Afrotheria and the Laurasiatheria. Likewise, xenarthrans and pangolins were once considered to belong to the same group due to morphological similarities, the same misconceptions that once lead to clustering bats and flying lemurs together.[58] More careful evaluation of morphological traits often resolves some of these issues, as different traits carry different levels of homoplasy (e.g., dental, cranial, postcranial[59]). Notwithstanding, it was the integration of molecular data and morphological traits that has alleviated several of the phylogenetic relationship artifacts caused by morphological convergence.

Inferring phylogenetic relationships from morphological data alone can also be problematic because of the challenge of assigning different weights to character states. A major consideration to determine if a fossil is informative to calibrate a given node is the presence of one or more synapomorphies—shared derived traits—that characterize a particular clade. For instance, the best estimates of the divergence of the cercopithecoids and hominoids range from 20 to 28 Ma. The oldest fossil belonging to Hominoidea is still controversial: multiple fossils, including *Kamoyapithecus*, *Morotopithecus*, *Ugandapithecus*, *Proconsul*, and *Rukwapithecus*, date between 19 and 28 Ma. Similarly, some early stem cercopithecoids such as *Alophe and Nsungwepithecus* date between 22 and 25 Ma.[60–62] The fossil record is fragmentary

and taxonomic assignments of many of these specimens are often debated. Still, multiple victoriapithecid remains, including the best-known taxon *Victoriapithecus macinnesi*,[63] represent the earliest unambiguous members of the Cercopitheoidea, suggesting that the split between Cercopithecoidea and Hominoidea took place before ~20 Ma.[3,31,39] Victoriapithecids are unambiguously classified as crown cercopithecoids based on several derived traits (e.g., lower bilophodont molars with low cusps). However, assigning different weights to morphological characters is ambiguous and often subjective. One alternative is to input all characters with equal weighting. Nevertheless, this solution assigns synapomorphies the same weight as other less informative characters (e.g., homoplasy), which can impact tree topology. By combining information from extant and extinct taxa, TED analyses have the high potential to revisit the phylogenetic placement of the fossils included in the dataset. For instance, Herrera and Davalos[14] applied tip-dating and a fossilized birth-death process to infer Strepsirrhini phylogeny and divergence times. In their reconstruction, the extinct *Saharagalago mirensis*, originally described as a stem Galagidae based on several dental traits in both upper and lower dentition,[64] was placed as sister to Lorisiformes, a result also recovered by Gunnell et al.[65] The main consequence of this "alternative placement" is younger estimates for the origin of both Lorisidae and Galagidae than regular node calibration studies.[14,66,67] It is unclear if the placement of *Saharagalago* in their analyses is a possible artifact of morphological analyses or might represent a new interpretation of the fossil record for this primate group.

The lack of robust models of morphological evolution is—most likely—one of the major concerns about TED analyses. While we have a decent understanding of the chemistry behind DNA mutations required to implement complex models of nucleotide substitution rates, the same cannot be said for the patterns and rate of morphological evolution.[16] Until now, there are no robust models to account for variation in the rate of morphological evolution across multiple characters, and statistical methods often require discretizing and subjectively orienting characters. However, new methods are becoming available to address this limitation (see References 68,69). Furthermore, current methods use the Mk model to describe symmetric transitions between character states—which fails to accommodate directional evolution and it is of difficult biological interpretation.[70–73] Comparisons between molecular and morphological clocks showed that the morphological clock estimates older ages than the molecular clock.[52,53] Multiple studies have indicated issues in using the Mk model in Bayesian morphological phylogenetics suggesting that the Mk model might be too unrealistic and inadequate for the analysis of most morphological data sets.[74] The same authors have also suggested that weighted parsimony might outperform Bayesian analyses run under the Mk model[75] (but also see O'Reilly et al.[76]). Recent studies have shown that using continuous characters can overcome the problems related to the subjectiveness behind weighting discretized morphological traits, while preserving phylogenetic information present in interspecific variation, thus potentially improving morphological phylogenetics.[77,78] Advances in the use of probabilistic models for analyzing morphological data that relax the assumptions of the Mk model (e.g., to incorporate heterogeneity in substitution rates)[79,80] or that deal with the interpretation of missing data[81] have also been recently developed and hold promises to better model morphological evolution.

A final consideration about the limitation of many morphological analyses regards the assumption that traits evolve independently; however, correlated evolution is known to impact morphological change over time. Failing to account for interdependence and coevolution among morphological characters is likely to lead to phylogenetic inference errors mainly due to the effect of long-branch attraction,[16] in which similar morphologies will tend to be recovered as more closely related to each other. However, removing correlated characters tends to affect mainly fossil taxa, which often have fewer synapomorphic traits linking them to other lineages.[14] Recent methods are allowing the implementation of more complex models that account for the correlation between multiple continuous traits to estimate divergence times,[82] holding promises to overcome some of the above-mentioned challenges.

It is important, however, to keep in mind that node dating and TED approaches are not necessarily mutually exclusive.[83,84] As mentioned above, the node-dating approach is limited by the uncertainty in fossil evidence, especially in establishing maximum constraints and prior distributions for a node age. In contrast, TED can often produce unlikely young or old age estimates that violate fossil-based minima. The so-called "tip-and-node dating" approach combines tip and node calibrations in a single analysis: node minimum boundaries are used to enforce realistic ages, and fossil tips interact with node calibrations to estimates maximum age constraints.[83] This approach has been argued to produce more precise, accurate and realistic age estimates than either node-dating or tip-dating alone.[83,84]

2.4 | The application of TED to the primate tree

To date, only a few studies have used TED to date the primate tree. Dembo et al. were among the first ones to apply TED to assess the phylogenetic placement of hominin fossils, including the controversial *Australopithecus sediba* and *Homo naledi*.[85,86] Based on a large set of craniodental traits, the authors used tip-dating analyses to jointly estimate the divergence time and relationships between living and extinct species. Their analyses supported the hypothesis that both *A. sediba* and *H. naledi* belong to a clade with the other *Homo* species. Furthermore, their results suggest that *H. naledi* has originated much earlier than previously thought, around 900 ka.[85,86] More recently, Püschel et al.[87] used a TED approach combining mitochondrial and craniodental data to estimate divergence times under alternative topological hypotheses of controversial hominin taxa. Their results dated the origin of the genus *Homo* between 4.30 and 2.56 Ma.

One of the first studies to integrate non-human primate molecular and morphological data into a Bayesian TED approach was performed by Herrera and Davalos[14] and aimed to estimate divergence times and reconstruct the phylogenetic history of the Malagasy lemurs. This primate radiation represents an ideal example in which a

TED approach could help in obtaining more reliable age estimates due to the absence of appropriate lemur fossils to calibrate internal nodes. No true fossil lemur is currently known and even within the strepsirrhine clade as a whole, only two fossils can be reliably used for node-dating: *Saharagalago mirrensis* (~36.9–42 Ma)[64] and *Wadilemur elegans* (~35 Ma),[88] both recognized as stem galagids. However, at least 17 species of recently extinct subfossil lemurs dating between 400 and 20,000 years ago exist,[89] which could provide important information to date the lemur tree. The recent recovery of ancient DNA information from some of these subfossil species[90,91] made them even more valuable for TED analyses. In their study, Herrera and Davalos[14] used 421 morphological and 5767 protein-coding molecular characters for 148 taxa. Their study was the first to include most of the living and extinct species of strepsirrhine primates in a comprehensive analysis. Interestingly, their results recovered some alternative placement of fossil taxa. Among those, the subfossil genus *Megaladapis* was recovered as sister to all lemurs other than *Daubentonia*; *Plesiopithecus* as sister to *Daubentonia*; *Saharagalago* as sister to all lorisiformes instead of being a stem galagid; and the *Komba* sp. as the sister taxon of the extant needle-clawed galago (*Euoticus* spp.) instead of being a stem galagid. While this study is not the only one to recover these unusual results,[65,92] it is yet unclear if such incongruences are artifacts of the TED approach (see above), of the limited molecular data for subfossils (mtDNA only), or indeed represent a more accurate interpretation of the phylogenetic placement of fossil strepsirrhines. In a recent study, Marciniak et al. were able to obtain whole genomic data for *Megaladapis* and obtained more support for its placement as sister taxon to *Eulemur*.[91]

The application of TED has challenged our view of the phylogenetic and biogeographic history of other parts of the primate tree of life. For instance, TED methods have been used to classify a new adapiform primate, *Masradapis tahai*, from the late Eocene of Egypt,[92] to reclassify the enigmatic *Propotto* from Kenya as the sister taxon of the extant aye-aye (*Daubentonia*) from Madagascar,[65] and to study the diversification of Platyrrhini.[93] The study conducted by Silvestro et al.[93] on platyrrhines is of particular interest because it did not apply a regular tip-dating approach in which fossils were included in the analyses together with extant species. Instead, the authors used a fossilized birth-death process with constrained fossil placements on a molecular dataset that only included living species. Their results indicated earlier age estimates than previously suggested, with the origin of the clade dating around 43 Ma and the crown age of Platyrrhini at ~33 Ma. These results striking contrast with previous studies that used either morphological analyses[94] or a node-dating approach,[95,96] indicating an origin for the crown group around 20–24 Ma. According to Silvestro et al.,[93] these differences are likely a result of using a larger and more complete fossil data set and a more realistic approach to calibrate the tree (i.e., fossilized birth-and-death approach).

Despite some important caveats and limitations, the TED approach provides an exciting opportunity for evolutionary anthropologists to integrate morphological evidence of both extant and extinct species in dating analyses. Furthermore, the integration between molecular and morphological analyses was shown to be of particular relevance to resolving deep divergences, rapid radiations, and the accurate phylogenetic reconstruction of clades with few extant members, especially if fossils are incorporated.[97] As shown in the examples above, by better integrating the paleontological record and molecular data, new studies are suggesting significantly different age estimates compared to node-dating analyses and new hypotheses about the interpretations of the fossil record, resulting in different evolutionary and biogeographical scenarios of the primate diversification.

3 | THE GENOMIC ERA AND THE NEW FRONTIERS OF PHYLOGENOMICS

3.1 | From a single locus to a genomic approach

The last two decades have been characterized by remarkable progress in both computational power and sequencing capabilities. Anthropologists can now obtain large-scale genomic data in relatively easy and affordable ways, even for non-model organisms.[17–20] While the number of primate species for which we have high-quality annotated genomes is still relatively low (Figure 3), various types of genomic data are now available for many taxa.[17,19,20,98] These data range from hundreds or thousands of loci obtained using sequence capture techniques (e.g., exome, ultra-conserved regions)[21] to thousands of single nucleotide polymorphisms obtained using reduced representation libraries (e.g., dRAD or ddRAD)[99–101] or whole-genome sequencing.[22,23,102,103] Although these datasets do not necessarily include the whole genome, they represent a dramatic increase in the amount of data available to reconstruct phylogenetic relationships among primates. Additionally, it is now possible to obtain genomic data from low-quality sources of DNA, such as museum specimens[104,105] and fecal material.[106–108] For instance, van der Valk et al.[105] recently recovered whole genomes from modern and century-old historical museum specimens to study changes in the historical demography of Grauer's and mountain gorillas.

Another example of the growing application of massive parallel sequencing is paleogenomics. In addition to the many studies on ancient human and archaic hominins, such as Neanderthal and Denisovan populations,[109] ancient DNA has now been recovered from multiple primate species as well. Kistler et al. generated complete or near-complete mitochondrial genomes for five extinct species to elucidate the phylogenetic placement of giant subfossil lemurs.[90] In the same study, population-level data for two subfossil species were analyzed, suggesting that low populations sizes might have contributed to the extinction of some subfossil lemurs.[90] More recently, Marciniak et al.[91] sequenced a low-coverage whole genome for the recently extinct koala lemurs (*Megaladapis edwardsi*). Using a comparative genomics approach these authors found novel genetic evidence for leaf-eating specializations, supporting morphological interpretations of the ecology of this extinct species. Ancient DNA was also recently obtained from an extinct Caribbean primate, *Xenothrix mcgregori*[110] (mitochondrial genome and a few nuclear regions) and a 5800-year-old baboon from Lesotho (low coverage genome sequence[111]).

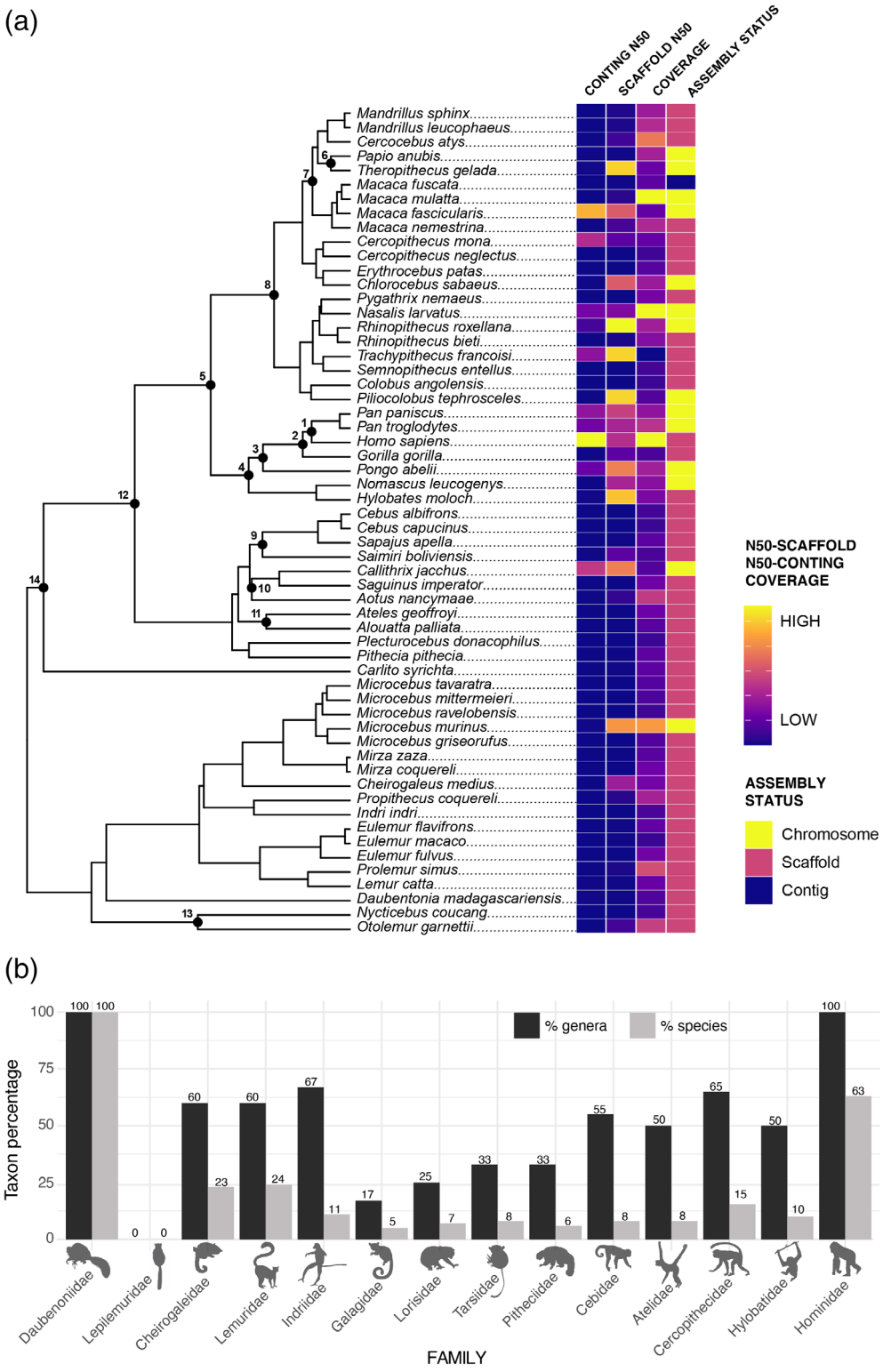

FIGURE 3 (a) Phylogenetic tree including species for which a draft genome is currently available on GenBank (updated September 2021). On the right we have details about the overall "quality" of the current draft (N50 contig [range: 9.79E + 03–5.64E + 07], N50 scaffold [range: 1.03E + 04–1.45E + 08], coverage [range: 1× to 290×], and assembly status [contig, scaffold, chromosome]). The figure also includes the main nodes that can be used as calibration points in the primate tree (see Table 1). Phylogenetic relationships are based on Springer et al.[1] (b) Percentage of primate species and genera with available genome resources organized by family. A whole genome for at least one member of the family Lepilemuridae, *Lepilemur mustelinus* has been recently sequenced (Marceniak et al.[91]), but it is currently unavailable on GenBank (September 2021)

The large amount of genomic data available represents a unique opportunity for anthropologists to explore phylogenetic relationships and understand the various mechanisms of adaptation within non-human primates.[17] However, they also pose fundamental challenges especially in our ability to deal with large—and often incomplete—datasets.

3.2 | The use of genomic data in dating the primate tree

The increasing availability of genomic data for several primate taxa has the potential to resolve some outstanding questions about the primate radiation. The limited amount of phylogenetic signal in

TABLE 1 Commonly used node calibrations for dating the primate tree with minimum bound (informed by the oldest known fossils for a specific clade) and suggested 95% prior distribution intervals

Node	MRCA	Minimum bound	Suggested 95% prior distribution	Informative fossils	Dates (Ma)
1	*Homo-Pan*	5	10	*Sahelanthropus tchadensis*	7–6.2
				Orrorin tugenensis	6.25.7
				Ardipithecus kadabba	5.77–5.54
				Ardipithecus ramidus	4.4
2	*Homo-Gorilla*	10	18	*Chororapithecus abyssinicus*[a]	10.5–10.0
				Nakalipithechus nakayami	12–5.33
3	*Homo-Hylobates*	12.5	18	*Kapi ramnagarensis*	13.8–12.5
				Yuanmoupithecus	9–7
4	*Homo-Pongo*	12.7	21	*Sivapithecus indicus*	12.7–8.5
				Khoratpithecus chiangmuanensis	13.5–10
5	Crown Catarrhini	21	34	*Proconsul* spp.	20–19
				Morotopithecus bishop	>20.6
				Victoriapithecus macinnesi	19–12.5
				Nsungwepithecus gunnelli	25.2
				Rukwapithecus fleaglei	25.2
				Alophe metios	22
				Catopithecus browni	35.9–35.6
6	*Theropithecus-Papio*	4.2	6.5	*Theropithecus oswaldi*	3.89–3.4
7	*Macaca*	5.3	10	*Macaca libyca*	6–5.5
				cf. *Macaca* sp.	6.1–5.3
8	Colobinae	12.5	15	*Microcolobus tugenensis*	11–10
9	*Saimiri-Cebus*	12.5	20.5	*Neosaimiri fieldsi*	12.5
				Dolichocebus gaimanensis	20.5
10	Crown Aotinae	12.5	20.5	*Aotus dindensis*	12.5
11	Crown Atelidae	12.5	20.5	*Stirtontia victoriae*	13.6
				Stirtonia tatacoensis	12.5
12	Crown Anthropoidea	34	56	*Catopithecus browni*	35.9–35.6
				Perupithecus ucayaliensis	37.2
				Oligopithecus savage	35.1–34.0
				Propliopithecus ankeli	33.4–33.1
				Aegyptopithecus zeuxis	34.0–33.8
13	Crown Lorisiformes	36.9	47	*Saharagalago misrensis*	41.2–36.
				Wadilemur elegans	35
14	Crown Haplorhini	55.8	70	*Teilhardina*	65.8–55.8
				Archicebus achilles	55.8–54.8

Note: The list was compiled following indications by Raaum,[31] and Pozzi et al.,[3] and updated with more recent findings. This table is not meant to provide an exhaustive list of all the fossils that can be used as calibration points, but a summary of the most commonly used calibration intervals for some of the major primate clades. Node numbers as indicated in Figure 3. Nodes indicate the divergence between most recent common ancestor between indicated taxa (MRCA).

[a]Disputed relationships.

traditional markers (e.g., mitochondrial DNA or small multilocus datasets) is one of the main factors affecting the lack of resolution in several parts of the primate tree. This might be particularly relevant when multiple lineages rapidly originate (e.g., adaptive radiations) leaving behind little trace of their evolutionary past. One of the most extraordinary examples of adaptive radiation within primates is the Malagasy lemurs that adapted to occupy a myriad of environments and dietary niches.[14,112] Although the sister-taxon position of aye-ayes in relation to all other lemurs is well established, the relationships among the other four families (Lemuridae, Cheirogaleidae, Indriidae, and Lepilemuridae) are still unresolved. Multiple recent studies have suggested a sister-taxon relationship between Cheirogaleidae and

Lepilemuridae; however, the relationships with indriids and lemurids are far from resolved.[1,3,14,67,90] The inadequate level of phylogenetic signal in the molecular data used in these studies (either mitochondrial DNA or a few nuclear loci) was identified as a possible reason for the lack of phylogenetic resolution. A similar argument has been used to explain the complicated relationships among galagids, African and Asian lorisids[66,67,113] or interrelationships within platyrrhines.[101] In most cases, the increase of genetic data has been proposed as a possible solution to the problem.

The use of genomic-level data can provide higher resolution in the tree topology. More data can increase the support of specific relationships among taxa, reducing the level of homoplasy present in the dataset.[114,115] For instance, phylogenomic approaches have clarified relationships within butterflies,[116] birds,[117,118] and mammals,[119] among others. Additionally, several new techniques that can reconstruct species trees from multiple independent loci have been recently developed. Both theoretical and empirical studies have demonstrated that the power of resolving phylogenetic relationships increases with the number of loci used.[120–122] Another advantage of phylogenomics is a significant improvement in the accuracy of branch length estimates, especially when combined with sophisticated molecular evolutionary models.[21] Branch lengths are particularly relevant in dating analyses since they provide the level of genetic divergence among taxa that will be converted into a chronological scale using an external source of information, as previously described.

Although genomic data can undoubtedly help in resolving the tree of life, the massive amount of data poses some analytical challenges. Thanks to advances in DNA sequencing technology, we now have publicly available whole-genome sequences for every family within the primate clade, the only exception being the family Lepilemuridae (Figure 3) but see Marceniak et al.[91] Although this trend is rapidly accelerating, the availability of genomic datasets is taxonomically biased and absent for the vast majority of the over 500 primate species commonly recognized.[20] Obtaining high-quality samples (e.g., blood or tissue) to use for genomic analyses can be challenging for many primate species due to wildlife regulations and/or ethical reasons (e.g., critically endangered species). Recent advancements in obtaining genomic data from low-quality specimens such as museum samples, hair, and fecal samples will be critical to expand the amount of genomic data for non-human primates.[104–107] Until then, many phylogenetic datasets will be largely incomplete, with multiple taxa represented by only a few loci.

Researchers are faced with a difficult choice between increasing taxon representation or minimizing the amount of missing data in their analyses. On one hand, taxon representation plays a major role in phylogenetic reconstructions at both the biological and methodological level.[115,123] At the biological level, poorly known taxa that require more phylogenetic studies are the ones that have limited (if any) genetic resources available. At the methodological level, missing taxa can potentially impact the accuracy and precision of both topology and branch length estimates. On the other hand, increasing the number of taxa likely results in more missing data (see below). For instance, Upham et al.[4] recently performed a comprehensive phylogeny for over 6000 mammal species. Although the dataset used was relatively small compared to genomic-level studies (only 31 genes), the level of missing data was extremely high (mean = 88.1% per species). The effect of missing data in molecular phylogenetic analyses is still debated. Some studies have suggested that missing data can yield inaccurate topology, node support, and branch length estimates.[124] Other studies have indicated that missing data might have little impact and that maximizing taxon representation—even though highly incomplete (<90%)—is beneficial to phylogenetic analyses.[55,123,125]

Another potential challenge in phylogenomic analyses is the computational power needed to build and date large phylogenies. First, rates of molecular evolution vary both among loci and species and are correlated to life-history traits, environment, and patterns of speciation. Therefore, complex models of molecular evolution are computationally demanding to be implemented for a high number of loci and taxa. Second, although parallelization can significantly reduce running time, the joint estimation of branch lengths and age estimates at the genomic scale for hundreds or thousands of species is still likely beyond the capacity of computational power available to most researchers.[4,21,126] Two possible solutions that can help minimizing the computational burden of molecular dating analyses have been suggested in the literature. In a series of papers, dos Reis et al.[126] proposed a two-step "sequential Bayesian analyses." In this approach, the posterior age estimates derived from a genomic-scale small taxonomic dataset (*step 1*) are used as priors in subsequent analyses with substantially more species but much-reduced nucleotide sample (*step 2*). Among some of its advantages, this method bypasses the computational challenges of a typical combined molecular analysis. Also, problems related to high level of missing data are minimized, as the divergence time estimate analyses conducted in stage one (reduced dataset) is obtained from a complete (or nearly complete) dataset.

The second approach, usually referred as "backbone-and-patch," was first implemented by Jetz et al. in large phylogenetic study of all living birds.[127] This approach splits the phylogenetic analysis of a large dataset into two nonoverlapping levels of analysis, each computationally feasible. The first level is the so-called estimate of the "backbone" divergences among major lineages (e.g., orders and families). The second level is represented by the dating of species-level "patch" clades, each corresponding to one representative tip on the backbone tree. The two levels of analyses (backbone and patch) are therefore nonoverlapping except at one shared node at the root of each patch clade. The final step of this approach is to generate sets of full-sized trees (all patches plus their backbone) in a common evolutionary timescale by rescaling the relative-time patches to absolute using the node that each patch clade shares on the dated backbone. The "backbone-and-patch" has been successfully used in many taxa, including birds, reptiles, amphibians, and mammals.[4,128,129] This approach minimizes the computational burden of dating large datasets and can reduce the level of missing data, at least in the backbone analyses. One advantage of this approach is the ability to integrate both living and fossil taxa by using a TED approach, as it is possible to implement some of the TED methods mentioned in the previous section.

4 | DATING WITHOUT FOSSILS

The use of fossils as external sources of absolute time is the most common approach in calibrating the tree of life.[16,25,31] However, incorporating fossil information can have major drawbacks related to errors in their phylogenetic placement and/or geochronological information, or due to the scarcity of fossil record for some groups. For instance, as discussed above, Malagasy lemurs lack a reliable fossil record that can be used to date this radiation. In these situations, distantly related fossils can be used (e.g., lorisiform fossils), however, this approach can increase the challenges in modeling the variation of molecular rates across phylogeny.[130]

BOX 1 GLOSSARY

Phylogenetics: A subfield of systematics that addresses the inference of the evolutionary history and relationships among or within groups of taxa (e.g., species), as represented by the tree topology. Phylogenetic analyses can be run using morphological, behavioral and/or molecular data.

Phylogenomics: A field that employs phylogenetic analyses using genomic-level data. Phylogenomics draws information by comparing large portions of genomes (e.g., reduced representation of genomic datasets), or entire genomes.

Gene-tree: A representation of the evolutionary history of one specific gene (or locus). (see Species-tree).

Species-tree: A tree representing the "true" relationships among evolutionary lineages. Internal nodes of a species tree represent speciation events.

Dated phylogeny (or timetrees): A phylogenetic tree in which the relative branch lengths indicating differences between lineages were converted into units of absolute time.

Clock models: Techniques that use a rate of evolution to convert molecular or morphological distances between lineages into timescales. More complex molecular clock models allow substitution rates to vary through time and among lineages (relaxed clock) or not (strict clock). Although the focus of intense debate, models of morphological evolution for continuous and discreate traits have also been developed.

Node-dating: A method that integrates age and a-priory taxonomic interpretations of the fossil record to enforce age constraints to some internal nodes of an evolutionary tree.

Total-evidence dating: An approach where morphological and temporal data from fossils are analyzed together with morphological and DNA sequence data from extant and—in some rare cases—extinct species to simultaneously infer the phylogenetic relationship and divergence times between extinct and extant lineages.

Tip-dating: A technique that allows the inference of time-calibrated phylogenetic trees by using both extant and extinct species as terminal taxa. This method calibrates the branch lengths by analyzing the morphological dataset for all taxa under a morphological character evolution model and molecular data, if available.

Fossilized birth-death: A model that allows the estimation of species divergence times from molecular and fossil information under the same macroevolutionary framework, including diversification and fossil sampling. The model includes multiple parameters, such as the rates of speciation, extinction, preservation, discovery, and sampled species.

Crown group: In phylogenetics, a crown group is a collection of species, composed of the most recent shared common ancestor of all extant members of a clade, as well as all of the descendants of that common ancestor, whether they are still living or are extinct.

Stem group: In phylogenetics, a stem group consists of a paraphyletic grouping of extinct species that are positioned outside a given crown group (on its "stem"). A stem group is more closely related to its corresponding crown group than to the extant sister clade of that crown group.

Exome: The exome is the term used to refer to all the exons within the genome collectively. Exons are all the DNA sequences that code for proteins.

Ultra-conserved regions: Ultra-conserved elements (UCE) are highly conserved regions of organismal genomes shared among evolutionary distant taxa.

Restriction site-associated DNA sequencing (RAD-seq): This technique is a fractional genome sequencing strategy, designed to reduce the amount of the genome investigated (0.1%–10%) by digesting the genome with a restriction enzyme and sequencing DNA next to those restriction sites. Double digest restriction-site associated DNA (ddRADseq) is a variation on the RAD sequencing protocol, in which the fragment shearing is replaced with a second restriction digestion.

Contig, short for contiguous sequence: A contig generated by the overlap between ends of several reads. These sequences can then be collapsed into a single non-redundant sequence in which the order of bases is known to a high confidence level.

Scaffold: A scaffold is a set of contigs that have been ordered and oriented. Scaffolds are composed of contigs (see above) and gaps (non-sequenced regions).

Genome coverage (or depth): The number of unique reads that include a given nucleotide in the underlying consensus sequence.

N50 statistic (contig/scaffold): It defines assembly quality in terms of contiguity. Given a set of contigs/scaffolds, the N50 is defined as the sequence length of the shortest contig/scaffold at 50% of the total genome length. It can be interpreted as the median contig size of your genomic assembly: at least half of the nucleotides in the assembly belongs to contigs/scaffolds with the N50 length or longer.

Recently, new approaches that bypass the use of fossils have been developed; however, their development is still at its infancy and their reliability on different parts of the primate tree is still poorly explored.[130] These methods rely on direct estimates of the mutation rate. Several studies have estimated the mutation rate for different primate species using pedigree trios.[131–137] This approach uses genomic data from three individuals in a known pedigree to identify variants in the offspring that are different from their parents. Researchers can estimate the number of "de novo mutations" and obtain a mutation rate for the focal species.[130] These mutation rates can then be used in multispecies coalescent approaches to jointly estimate divergence times and rates of evolution while accounting for incomplete lineage sorting, that is, the level of gene tree discordance.[130]

Mutation rates have been mostly applied to study the divergence times within great apes. In 2012, Langergraber et al. employed mutation rate calculated for humans[133,138] and the generation times within wild populations of chimpanzees and gorillas to estimate divergence times among humans, chimpanzees, bonobos, and gorillas.[139] Their estimates for the human–chimpanzee split was dated around 7–8 million years, in line with multiple studies that applied node-dating techniques.[1,3] More recently, Besenbacher et al. inferred de novo mutations in chimpanzee, gorilla, and orangutan parent-offspring trios, indicating a remarkable slowdown in the human mutation rate.[131] Using the nonhuman rates instead of the human rate, they estimated divergence times more in line with the fossil record.

These methods are relatively sensitive to the mutation rate estimates. For instance, Moorjani et al. applied the human mutation rate across the primate tree and recovered a divergence date between cercopithecoids and hominoids of ~62 Mya.[140] This estimate is twice as old as estimates from node-dating studies that indicate divergence ages ranging between 23 and 34 Mya.[3,21,31] The probable explanation for such a remarkable discrepancy is the slower mutation rate in hominoids compared to cercopithecoids. The application of these methods can be particularly problematic, especially across large phylogenies, when mutation rates and generation times can dramatically change across different lineages.[130] Additionally, this approach is currently not applicable to most species given the lack of genomic information and the complexity of obtaining pedigree data for many taxa. Nevertheless, thanks to the continuous advancement of computational techniques paired with the generation of new genomic data for an increasing number of primate species, it is likely that these methods will represent an important new avenue in estimating divergence times across primates. Additional studies should be conducted to compare traditional fossil-based approaches to molecular rates-based approaches to evaluate the level of concordance in dating the primate tree.[130]

5 | CONCLUSION

The recent ability to produce large-scale data through high-throughput sequencing technologies and the development of more sophisticated methods to integrate molecular and paleontological data offer an exciting opportunity for anthropologists. Several consortia are currently working in generating genomic data for a large number of species using similar protocols or by targeting similar genomic regions.[141] These technological developments can be applied to advance our understanding of the timing and pattern of evolutionary relationships in the primate tree and inform us on the processes of speciation, extinction, morphological and behavioral trait evolution, biogeography, and many other critical aspects of primate biodiversity.

We have identified three main future directions that will improve our understanding of the primate tree of life. First, future studies should explore the use of "tip-and-node dating" approaches to incorporate temporal, topological, and morphological information from the paleontological record and, at the same time, improve the reliability of age estimates. Second, future research should aim at generating genomic data obtained from museum specimens or low-quality samples (e.g., feces and hair) to expand our understanding of the primate tree of life. This will provide more genomic resources for taxa for which high-quality biological samples are currently lacking. Moreover, the generation of genomic data from fossil and subfossil samples (e.g., Neanderthals, Denisovans, subfossil lemurs) will also provide critical information that will increase our ability to link molecular and morphological evidence, thus increasing the accuracy and taxa representation of dated primate trees. Finally, new genomic resources will contribute to more accurate estimates of mutation rates across different primate lineages. These estimates, combined with multispecies coalescent techniques, will contribute to a better understanding of divergence times within the primate radiation.

Evolutionary anthropologists and primatologists have a unique opportunity to integrate genomic data and fossil-based approaches in a single framework to obtain not only better age estimates of divergence in the primate tree but also a better understanding of the paleontological record. More accurate age estimates will be essential to interpret climatic, geological, and biogeographical data and to explore diversification processes within primates.

ACKNOWLEDGMENTS

This work was supported by the NSF grant BCS-1926105 (to L.P.). The authors would like to thank the Editor-in-Chief Jason Kamilar for the opportunity to submit this article. Thanks to four anonymous reviewers for providing extremely valuable feedback to early versions of the manuscript. We would also like to thank Siobhan Cooke for providing important information about the fossil record in the platyrrhines and James Herrera for aiding with the verification of hierarchical taxonomic arrangement as well as age range information used to produce Figure 2.

CONFLICT OF INTEREST

The authors declared no potential conflicts of interest.

DATA AVAILABILITY STATEMENT

Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

ORCID

Luca Pozzi https://orcid.org/0000-0001-7916-0805
Anna Penna https://orcid.org/0000-0002-5147-317X

REFERENCES

[1] Springer MS, Meredith RW, Gatesy J, et al. 2012. Macroevolutionary dynamics and historical biogeography of primate diversification inferred from a species supermatrix. PLoS One 7:e49521.

[2] Wilkinson RD, Steiper ME, Soligo C, et al. 2011. Dating primate divergences through an integrated analysis of palaeontological and molecular data. Syst Biol 60:16–31.

[3] Pozzi L, Hodgson JA, Burrell AS, et al. 2014. Primate phylogenetic relationships and divergence dates inferred from complete mitochondrial genomes. Mol Phylogenet Evol 75:165–183.

[4] Upham NS, Esselstyn JA, Jetz W. 2019. Inferring the mammal tree: species-level sets of phylogenies for questions in ecology, evolution, and conservation. PLoS Biol 17:e3000494.

[5] Parham JF, Donoghue PCJ, Bell CJ, et al. 2012. Best practices for justifying fossil calibrations. Syst Biol 61:346–359.

[6] Pozzi L. 2016. The role of forest expansion and contraction in species diversification among galagos (primates: Galagidae). J Biogeogr 43:1930–1941.

[7] Landis MJ. 2016. Biogeographic dating of speciation times using paleogeographically informed processes. Syst Biol 66:syw040–syw144.

[8] García-Verdugo C, Caujapé-Castells J, Sanmartín I. 2019. Colonization time on Island settings: lessons from the Hawaiian and Canary Island floras. Bot J Linn Soc 191:155–163.

[9] Emerson BC. Evolution on oceanic islands: molecular phylogenetic approaches to understanding pattern and process. Mol Ecol. 2002; 11:951–966. https://doi.org/10.1046/j.1365-294x.2002.01507.x

[10] Estrada A, Garber PA, Rylands AB, et al. 2017. Impending extinction crisis of the world's primates: why primates matter. Sci Adv 3: e1600946.

[11] Pavoine S, Bonsall MB, Davies TJ, Masi S. 2019. Mammal extinctions and the increasing isolation of humans on the tree of life. Ecol Evol 9:914–924.

[12] Heath TA, Huelsenbeck JP, Stadler T. 2014. The fossilized birth-death process for coherent calibration of divergence-time estimates. Proc Natl Acad Sci USA 111:E2957–E2966.

[13] Zhang C, Stadler T, Klopfstein S, et al. 2016. Total-evidence dating under the fossilized birth–death process. Syst Biol 65: 228–249.

[14] Herrera JP, Dávalos LM. 2016. Phylogeny and divergence times of lemurs inferred with recent and ancient fossils in the tree. Syst Biol 65:772–791.

[15] Ronquist F, Klopfstein S, Vilhelmsen L, et al. 2012. A total-evidence approach to dating with fossils, applied to the early radiation of the hymenoptera. Syst Biol 61:973–999.

[16] Ronquist F, Lartillot N, Phillips MJ. 2016. Closing the gap between rocks and clocks using total-evidence dating. Philos Trans R Soc B 371:20150136.

[17] Grogan KE, Perry GH. 2020. Studying human and nonhuman primate evolutionary biology with powerful in vitro and in vivo functional genomics tools. Evol Anthropol 29:143–158.

[18] Rogers J, Gibbs RA. 2014. Comparative primate genomics: emerging patterns of genome content and dynamics. Nat Rev Genet 15: 347–359.

[19] Kuderna LF, Esteller-Cucala P, Marques-Bonet T. 2020. Branching out: what omics can tell us about primate evolution. Curr Opin Genet Dev 62:65–71.

[20] Orkin JD, Kuderna LFK, Marques-Bonet T. 2020. The diversity of primates: from biomedicine to conservation genomics. Annu Rev Anim Biosci 9:103–124.

[21] Reis MD, Gunnell GF, Barba-Montoya J, et al. Using phylogenomic data to explore the effects of relaxed clocks and calibration strategies on divergence time estimation: primates as a test case. Syst Biol. 2018;67:594–615. https://doi.org/10.1093/sysbio/syy001

[22] Rogers J, Raveendran M, Harris RA, et al. 2019. The comparative genomics and complex population history of *Papio* baboons. Sci Adv 5:eaau6947.

[23] Carbone L, Alan Harris R, Gnerre S, et al. 2014. Gibbon genome and the fast karyotype evolution of small apes. Nature 513:195–201.

[24] Steiper ME, Young NM. 2006. Primate molecular divergence dates. Mol Phylogenet Evol 41:384–394.

[25] Ho SYW, Phillips MJ. 2009. Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. Syst Biol 58:367–380.

[26] Ho SYW, Duchêne S. 2014. Molecular-clock methods for estimating evolutionary rates and timescales. Mol Ecol 23:5947–5965.

[27] Benton MJ. 2003. Dating the tree of life. Science 300:1698–1700.

[28] Donoghue PCJ, Benton MJ. 2007. Rocks and clocks: calibrating the tree of life using fossils and molecules. Trends Ecol Evol 22:424–431.

[29] Parham JF, Irmis RB. 2008. Caveats on the use of fossil calibrations for molecular dating: a comment on near et al. Am Nat 171: 132–136.

[30] Blair Hedges S, Kumar S. 2004. Precision of molecular time estimates. Trends Genet 20:242–247.

[31] Raaum RL. 2015. Molecular evidence on primate origins and evolution. In: Henke W, Tattersall I, editors. Handbook of paleoanthropology, Berlin: Springer Berlin Heidelberg. p 1083–1135.

[32] Steiper ME, Young NM. 2008. Timing primate evolution: lessons from the discordance between molecular and paleontological estimates. Evol Anthropol 17:179–188.

[33] Pozzi L, Hodgson JA, Burrell AS, Disotell TR. 2011. The stem catarrhine Saadanius does not inform the timing of the origin of crown catarrhines. J Hum Evol 61:209–210.

[34] Ksepka DT, Parham JF, Allman JF, et al. 2015. The fossil calibration database—a new resource for divergence dating. Syst Biol 64: 853–859.

[35] Begun DR. 2015. 4 fossil record of miocene hominoids. In: Handbook of paleoanthropology, Berlin: Springer Berlin Heidelberg. p 1261–1332.

[36] Godfrey LR. 2017. Subfossil lemurs. In: Bezanson M, MacKinnon KC, Riley E, et al., editors. The international encyclopedia of primatology, Hoboken, NJ: John Wiley & Sons. p 1–5.

[37] Springer MS, Foley NM, Brady PL, et al. 2019. Evolutionary models for the diversification of placental mammals across the KPg boundary. Front Genet 10:1241.

[38] Arcila D, Alexander Pyron R, Tyler JC, et al. 2015. An evaluation of fossil tip-dating versus node-age calibrations in tetraodontiform fishes (Teleostei: Percomorphaceae). Mol Phylogenet Evol 82: 131–145.

[39] Benton MJ, Donoghue PCJ. 2006. Paleontological evidence to date the tree of life. Mol Biol Evol 24:26–53.

[40] Marshall CR. 2019. Using the fossil record to evaluate timetree timescales. Front Genet 10:1049.

[41] Tavare S, Marshall CR, Will O, et al. 2002. Using the fossil record to estimate the age of the last common ancestor of extant primates. Nature 416:726–729.

[42] Martin RD. 1993. Primate origins: plugging the gaps. Nature 363: 223–234.

[43] Eernisse D, Kluge A. 1993. Taxonomic congruence versus total evidence, and amniote phylogeny inferred from fossils, molecules, and morphology. Mol Biol Evol 10:1170–1195.

[44] O'Leary MA. 1999. Parsimony analysis of total evidence from extinct and extant taxa and the cetacean-artiodactyl question (Mammalia, Ungulata). Cladistics 15:315–330.

[45] Wiens JJ, Kuczynski CA, Townsend T, et al. 2010. Combining phylogenomics and fossils in higher-level squamate reptile phylogeny: molecular data change the placement of fossil taxa. Syst Biol 59: 674–688.

[46] Koch NM, Garwood RJ, Parry LA. 2020. Fossils improve phylogenetic analyses of morphological characters. Proc Biol Sci 288: 20210044.
[47] Pyron RA. 2011. Divergence time estimation using fossils as terminal taxa and the origins of Lissamphibia. Syst Biol 60:466–481.
[48] Fleagle JG. Primate Adaptation and Evolution. Academic Press; 2013.
[49] Gavryushkina A, Heath TA, Ksepka DT, et al. 2016. Bayesian total-evidence dating reveals the recent crown radiation of penguins. Syst Biol 66:syw060–syw073.
[50] Silvestro D, Warnock RCM, Gavryushkina A, Stadler T. 2018. Closing the gap between palaeontological and neontological speciation and extinction rate estimates. Nat Commun 9:5237.
[51] Gavryushkina A, Welch D, Stadler T, Drummond AJ. 2014. Bayesian inference of sampled ancestor trees for epidemiology and fossil calibration. PLoS Comput Biol 10:e1003919.
[52] Puttick MN, Thomas GH, Benton MJ. 2016. Dating placentalia: morphological clocks fail to close the molecular fossil gap. Evolution 70: 873–886.
[53] Caldas IV, Schrago CG. 2019. Data partitioning and correction for ascertainment bias reduce the uncertainty of placental mammal divergence times inferred from the morphological clock. Ecol Evol 9: 2255–2262.
[54] Felsenstein J. 2012. A comparative method for both discrete and continuous characters using the threshold model. Am Nat 179:145–156.
[55] Wiens JJ. 2003. Missing data, incomplete taxa, and phylogenetic accuracy. Syst Biol 52:528–538.
[56] Guillerme T, Cooper N. 2016. Effects of missing data on topological inference using a total evidence approach. Mol Phylogenet Evol 94: 146–158.
[57] O'Leary MA, Bloch JI, Flynn JJ, et al. 2013. The placental mammal ancestor and the post-K-Pg radiation of placentals. Science 339: 662–667.
[58] Springer M. 2004. Molecules consolidate the placental mammal tree. Trends Ecol Evol 19:430–438.
[59] Sánchez-Villagra MR, Williams BA. 1998. Levels of homoplasy in the evolution of the mammalian skeleton. J Mamm Evol 5(2): 113–126.
[60] Leakey MG, Ungar PS, Walker A. 1995. A new genus of large primate from the late Oligocene of Lothidok, Turkana District, Kenya. J Hum Evol 28:519–531.
[61] Stevens NJ, Seiffert ER, O'Connor PM, et al. 2013. Palaeontological evidence for an Oligocene divergence between Old World monkeys and apes. Nature 497:611–614.
[62] Rasmussen DT, Friscia AR, Gutierrez M, et al. 2019. Primitive Old World monkey from the earliest Miocene of Kenya and the evolution of cercopithecoid bilophodonty. Proc Natl Acad Sci USA 116: 6051–6056.
[63] Pilbeam D, Walker A. 1968. Fossil monkeys from the Miocene of Napak, North-East Uganda. Nature 220:657–660.
[64] Seiffert ER, Simons EL, Attia Y. 2003. Fossil evidence for an ancient divergence of lorises and galagos. Nature 422:421–424.
[65] Gunnell GF, Boyer DM, Friscia AR, et al. 2018. Fossil lemurs from Egypt and Kenya suggest an African origin for Madagascar's aye-aye. Nat Commun 9:3193.
[66] Pozzi L, Disotell TR, Masters JC. 2014. A multilocus phylogeny reveals deep lineages within African galagids (primates: Galagidae). BMC Evol Biol 14:72.
[67] Pozzi L, Nekaris KAI, Perkin A, et al. 2015. Remarkable ancient divergences amongst neglected lorisiform primates: phylogeny of lorisiform primates. Zool J Linn Soc 175:661–674.
[68] Lanfear R, Frandsen PB, Wright AM, et al. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. Mol Biol Evol. 2016;34: 772–773. https://doi.org/10.1093/molbev/msw260
[69] Rosa BB, Melo GAR, Barbeitos MS. 2019. Homoplasy-based partitioning outperforms alternatives in Bayesian analysis of discrete morphological data. Syst Biol 68:657–671.
[70] Pagel M. 1994. Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. Proc R Soc Lond B 255:37–45.
[71] Pagel M, Meade A. 2006. Bayesian analysis of correlated evolution of discrete characters by reversible-jump Markov chain Monte Carlo. Am Nat 167(6):808–825.
[72] Pagel M, Meade A, Barker D. 2004. Bayesian estimation of ancestral character states on phylogenies. Syst Biol 53:673–684.
[73] Lewis PO. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. Syst Biol 50: 913–925.
[74] Goloboff PA, Pittman M, Pol D, Xu X. 2018. Morphological data sets fit a common mechanism much more poorly than DNA sequences and call into question the Mkv model. Syst Biol 68:20160081.
[75] Goloboff PA, Torres A, Arias JS. 2018. Weighted parsimony outperforms other methods of phylogenetic inference under models appropriate for morphology. Cladistics 34:407–437.
[76] O'Reilly JE, Puttick MN, Pisani D, Donoghue PCJ. 2018. Empirical realism of simulated data is more important than the model used to generate it: a reply to Goloboff et al. Palaeontology 61:631–635.
[77] Parins-Fukuchi C. 2018. Use of continuous traits can improve morphological phylogenetics. Syst Biol 67:328–339.
[78] Slater GJ, Harmon LJ, Wegmann D, et al. 2012. Fitting models of continuous trait evolution to incompletely sampled comparative data using approximate Bayesian computation: evolutionary rates from incompletely sampled data. Evolution 66:752–762.
[79] Wright AM, Lloyd GT, Hillis DM. 2016. Modeling character change heterogeneity in phylogenetic analyses of morphology through the use of priors. Syst Biol 65:602–611.
[80] Gavryushkina A, Zhang C. 2020. Total-evidence dating and the fossilized birth–death model. In: Ho SYW, editor. The molecular evolutionary clock, Cham, Switzerland: Springer International Publishing. p 175–193. https://doi.org/10.1007/978-3-030-60181-2_11
[81] Brazeau MD, Guillerme T, Smith MR. 2019. An algorithm for morphological phylogenetic analysis with inapplicable data. Syst Biol 68: 619–631.
[82] Álvarez-Carretero S, Goswami A, Yang Z, Dos Reis M. Bayesian estimation of species divergence times using correlated quantitative characters. Syst Biol. 2019;68:967–986. https://doi.org/10.1093/sysbio/syz015
[83] O'Reilly JE, Donoghue PCJ. 2016. Tips and nodes are complementary not competing approaches to the calibration of molecular clocks. Biol Lett 12:20150975.
[84] Kealy S. 2017. Total evidence phylogeny and evolutionary timescale for Australian faunivorous marsupials (Dasyuromorphia). BMC Evol Biol 17(1):1–23.
[85] Dembo M, Radovčić D, Garvin HM, et al. 2016. The evolutionary relationships and age of *Homo naledi*: an assessment using dated Bayesian phylogenetic methods. J Hum Evol 97:17–26.
[86] Dembo M, Matzke NJ, Mooers AØ, Collard M. 2015. Bayesian analysis of a morphological supermatrix sheds light on controversial fossil hominin relationships. Proc R Soc B 282:20150943.
[87] Püschel HP, Bertrand OC, O'Reilly JE, et al. 2021. Divergence-time estimates for hominins provide insight into encephalization and body mass trends in human evolution. Nat Ecol Evol 5: 808–819.
[88] Seiffert ER, Simons EL, Ryan TM, Attia Y. 2005. Additional remains of *Wadilemur elegans*, a primitive stem galagid from the late Eocene of Egypt. Proc Natl Acad Sci USA 102:11396–11401.
[89] Godfrey LR, Jungers WL, Burney DA. 2010. Subfossil lemurs of Madagascar. In: Werdelin L, editor. Cenozoic mammals of Africa, Berkeley (CA): University of California Press. p 351–368.

[90] Kistler L, Ratan A, Godfrey LR, et al. 2015. Comparative and population mitogenomic analyses of Madagascar's extinct, giant 'subfossil' lemurs. J Hum Evol 79:45–54.
[91] Marciniak S, Mughal MR, Godfrey LR, et al. 2021. Evolutionary and phylogenetic insights from a nuclear genome sequence of the extinct, giant, "subfossil" koala lemur *Megaladapis edwardsi*. Proc Natl Acad Sci USA 118:e2022117118.
[92] Seiffert ER, Boyer DM, Fleagle JG, et al. 2018. New adapiform primate fossils from the late Eocene of Egypt. Hist Biol 30:204–226.
[93] Silvestro D, Tejedor MF, Serrano-Serrano ML, et al. 2019. Early arrival and climatically-linked geographic expansion of New World monkeys from tiny African ancestors. Syst Biol 68:78–92.
[94] Kay RF. Biogeography in deep time – What do phylogenetics, geology, and paleoclimate tell us about early platyrrhine evolution?. Mol Phylogenet Evol. 2015;82:358–374. https://doi.org/10.1016/j.ympev.2013.12.002
[95] Perez SI, Tejedor MF, Novo NM, Aristide L. 2013. Divergence times and the evolutionary radiation of New World monkeys (Platyrrhini, primates): an analysis of fossil and molecular data. PLoS One 8: e68029.
[96] Hodgson JA, Sterner KN, Matthews LJ, et al. 2009. Successive radiations, not stasis, in the South American primate fauna. Proc Natl Acad Sci USA 106:5534–5539.
[97] Smith ND, Turner AH. 2005. Morphology's role in phylogeny reconstruction: perspectives from paleontology. Syst Biol 54:166–173.
[98] Anderson JA, Vilgalys TP, Tung J. 2020. Broadening primate genomics: new insights into the ecology and evolution of primate gene regulation. Curr Opin Genet Dev 62:16–22.
[99] Boubli JP, da Silva MNF, Rylands AB, et al. 2018. How many pygmy marmoset (Cebuella Gray, 1870) species are there? A taxonomic reappraisal based on new molecular evidence. Mol Phylogenet Evol 120:170–182.
[100] Poelstra JW, Salmona J, Tiley GP, et al. Cryptic patterns of speciation in cryptic primates: microendemic mouse lemurs and the multispecies coalescent. Syst Biol. 2021;70:(2):203–218. https://doi.org/10.1093/sysbio/syaa053
[101] Valencia LM, Martins A, Ortiz EM, di Fiore A. 2018. A RAD-sequencing approach to genome-wide marker discovery, genotyping, and phylogenetic inference in a diverse radiation of primates. PLoS One 13:e0201254.
[102] Vanderpool D, Minh BQ, Lanfear R, et al. 2020. 2020 Apr. Primate phylogenomics uncovers multiple rapid radiations and ancient interspecific introgression. PLoS Biol 18(12):e3000954.
[103] van der Valk T, Gonda CM, Silegowa H, et al. 2020. The genome of the endangered dryas monkey provides new insights into the evolutionary history of the vervets. Mol Biol Evol 37:183–194.
[104] Burrell AS, Disotell TR, Bergey CM. 2015. The use of museum specimens with high-throughput DNA sequencers. J Hum Evol 79: 35–44.
[105] van der Valk T, Díez-del-Molino D, Marques-Bonet T, et al. 2019. Historical genomes reveal the genomic consequences of recent population decline in eastern gorillas. Curr Biol 29:165–170.e6.
[106] Chiou KL, Bergey CM. 2018. Methylation-based enrichment facilitates low-cost, noninvasive genomic scale sequencing of populations from feces. Sci Rep 8:1975.
[107] Orkin JD, Montague MJ, Tejada-Martinez D, et al. The genomics of ecological flexibility, large brains, and long lives in capuchin monkeys revealed with fecalFACS. Proc Natl Acad Sci. 2021;118: e2010632118. https://doi.org/10.1073/pnas.2010632118
[108] Snyder-Mackler N, Majoros WH, Yuan ML, et al. 2016. Efficient genome-wide sequencing and low-coverage pedigree analysis from noninvasively collected samples. Genetics 203:699–714.
[109] Marciniak S, Perry GH. 2017. Harnessing ancient genomes to study the history of human adaptation. Nat Rev Genet 18:659–674.
[110] Woods R, Turvey ST, Brace S, et al. 2018. Ancient DNA of the extinct Jamaican monkey *Xenothrix* reveals extreme insular change within a morphologically conservative radiation. Proc Natl Acad Sci USA 115:12769–12774.
[111] Mathieson I, Abascal F, Vinner L, et al. 2020. An ancient baboon genome demonstrates long-term population continuity in Southern Africa. Genome Biol Evol 12:407–412.
[112] Herrera JP. 2017. Testing the adaptive radiation hypothesis for the lemurs of Madagascar. R Soc Open Sci 4(1):161014.
[113] Pozzi L, Roos C, Blair ME. 2020. Molecular advances in lorisid taxonomy and phylogeny. In: Nekaris KAI, Burrows AME, editors. Evolution, ecology and conservation of lorises and pottos, Cambridge: Cambridge University Press. p 57–66.
[114] Rokas A, Williams BL, King N, Carroll SB. 2003. Genome-scale approaches to resolving incongruence in molecular phylogenies. Nature 425:798–804.
[115] Rokas A, Carroll SB. 2005. More genes or more taxa? The relative contribution of gene number and taxon number to phylogenetic accuracy. Mol Biol Evol 22:1337–1344.
[116] Espeland M, Breinholt J, Willmott KR, et al. 2018. A comprehensive and dated phylogenomic analysis of butterflies. Curr Biol 28:770–778.e5.
[117] Jarvis ED, Mirarab S, Aberer AJ, et al. 2014. Whole-genome analyses resolve early branches in the tree of life of modern birds. Science 346:1320–1331.
[118] Cloutier A, Sackton TB, Grayson P, et al. 2019. Whole-genome analyses resolve the phylogeny of flightless birds (Palaeognathae) in the presence of an empirical anomaly zone. Syst Biol 68:937–955.
[119] Kumar V, Hallström BM, Janke A. 2013. Coalescent-based genome analyses resolve the early branches of the Euarchontoglires. PLoS One 8:e60019.
[120] Song S, Liu L, Edwards SV, Wu S. 2012. Resolving conflict in eutherian mammal phylogeny using phylogenomics and the multispecies coalescent model. Proc Natl Acad Sci USA 109:14942–14947.
[121] Zhang C, Rabiee M, Sayyari E, Mirarab S. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinform 19:153.
[122] Mirarab S, Bayzid MS, Warnow T. 2016. Evaluating summary methods for multilocus species tree estimation in the presence of incomplete lineage sorting. Syst Biol 65:366–380.
[123] Wiens JJ, Tiu J. 2012. Highly incomplete taxa can rescue phylogenetic analyses from the negative impacts of limited taxon sampling. PLoS One 7:e42925.
[124] Lemmon AR, Brown JM, Stanger-Hall K, Lemmon EM. 2009. The effect of ambiguous data on phylogenetic estimates obtained by maximum likelihood and Bayesian inference. Syst Biol 58:130–145.
[125] Wiens JJ, Morrill MC. 2011. Missing data in phylogenetic analysis: reconciling results from simulations and empirical data. Syst Biol 60: 719–731.
[126] dos Reis M, Inoue J, Hasegawa M, et al. 2012. Phylogenomic datasets provide both precision and accuracy in estimating the timescale of placental mammal phylogeny. Proc R Soc B 279:3491–3500.
[127] Jetz W, Thomas GH, Joy JB, et al. 2012. The global diversity of birds in space and time. Nature 491:444–448.
[128] Tonini JFR, Beard KH, Ferreira RB, et al. 2016. Fully-sampled phylogenies of squamates reveal evolutionary patterns in threat status. Biol Conserv 204:23–31.
[129] Jetz W, Pyron RA. 2018. The interplay of past diversification and evolutionary isolation with present imperilment across the amphibian tree of life. Nat Ecol Evol 2:850–858.
[130] Tiley GP, Poelstra JW, dos Reis M, et al. 2020. Molecular clocks without rocks: new solutions for old problems. Trends Genet 36: 845–856.

[131] Besenbacher S, Hvilsom C, Marques-Bonet T, et al. 2019. Direct estimation of mutations in great apes reconciles phylogenetic dating. Nat Ecol Evol 3:286–292.
[132] Wu FL, Strand AI, Cox LA, et al. 2020. A comparison of humans and baboons suggests germline mutation rates do not track cell divisions. PLoS Biol 18:e3000838.
[133] Scally A, Durbin R. 2012. Revising the human mutation rate: implications for understanding human evolution. Nat Rev Genet 13: 745–753.
[134] Chintalapati M, Moorjani P. 2020. Evolution of the mutation rate across primates. Curr Opin Genet Dev 62:58–64.
[135] Wang RJ, Thomas GWC, Raveendran M, et al. 2020. Paternal age in rhesus macaques is positively associated with germline mutation accumulation but not with measures of offspring sociability. Genome Res 30(6):826–834.
[136] Pfeifer SP. 2017. Direct estimate of the spontaneous germ line mutation rate in African green monkeys. Evolution 71:2858–2870.
[137] Campbell CR, Tiley GP, Poelstra JW, et al. Pedigree-based and phylogenetic methods support surprising patterns of mutation rate and spectrum in the gray mouse lemur. Heredity. 2021;127:233–244. https://doi.org/10.1038/s41437-021-00446-5
[138] Roach JC, Glusman G, Smit AFA, et al. 2010. Analysis of genetic inheritance in a family quartet by whole-genome sequencing. Science 328:636–639.
[139] Langergraber KE, Prufer K, Rowney C, et al. 2012. Generation times in wild chimpanzees and gorillas suggest earlier divergence times in great ape and human evolution. Proc Natl Acad Sci U S A 109: 15716–15721.
[140] Moorjani P, Gao Z, Przeworski M. 2016. Human germline mutation and the erratic evolutionary clock. PLoS Biol 14:e2000744.
[141] Zoonomia Consortium. 2020. A comparative genomics multitool for scientific discovery and conservation. Nature 587:240–245.

AUTHOR BIOGRAPHIES

Luca Pozzi is an assistant professor in the Department of Anthropology at the University of Texas at San Antonio. His research mostly focuses on primate phylogenetics, species boundaries, genomics, biogeography, and the evolution of morphological and behavioral traits.

Anna Penna is a PhD candidate in the Department of Anthropology at the University of Texas at San Antonio. Her research focuses on primate trait evolution and species diversification, integrating comparative macroevolutionary approaches and collection-based genomics. Her PhD research investigates the convergent evolution of miniaturization in primates.

How to cite this article: Pozzi L, Penna A. Rocks and clocks revised: New promises and challenges in dating the primate tree of life. Evolutionary Anthropology. 2022;1–16. https://doi.org/10.1002/evan.21940