Journal of Human Evolution 66 (2014) 18–28

Contents lists available at ScienceDirect

Journal of Human Evolution

journal homepage: www.elsevier.com/locate/jhevol

Molecular phylogenetic analysis of the Papionina using concatenation and species tree methods

Elaine E. Guevara[a,1], Michael E. Steiper[a,b,c,d,*]

[a] Department of Anthropology, Hunter College, City University of New York, 695 Park Avenue, New York, NY 10065, USA
[b] Program in Anthropology, The Graduate Center, City University of New York, 365 5th Avenue, New York, NY 10016, USA
[c] Program in Biology, The Graduate Center, City University of New York, 365 5th Avenue, New York, NY 10016, USA
[d] New York Consortium in Evolutionary Primatology (NYCEP), New York, NY, USA

ARTICLE INFO

Article history:
Received 19 December 2012
Accepted 13 September 2013
Available online 23 October 2013

Keywords:
Gene tree heterogeneity
Supermatrix
Multispecies coalescent
Bayesian concordance
Primate evolution
Mangabeys

ABSTRACT

The Papionina is a geographically widespread subtribe of African cercopithecid monkeys whose evolutionary history is of particular interest to anthropologists. The phylogenetic relationships among arboreal mangabeys (*Lophocebus*), baboons (*Papio*), and geladas (*Theropithecus*) remain unresolved. Molecular phylogenetic analyses have revealed marked gene tree incongruence for these taxa, and several recent concatenated phylogenetic analyses of multilocus datasets have supported different phylogenetic hypotheses. To address this issue, we investigated the phylogeny of the *Lophocebus* + *Papio* + *Theropithecus* group using concatenation methods, as well as alternative methods that incorporate gene tree heterogeneity to estimate a 'species tree.' Our compiled DNA sequence dataset was ~56 kb pairs long and included 57 independent partitions. All analyses of concatenated alignments strongly supported a *Lophocebus* + *Papio* clade and a basal position for *Theropithecus*. The Bayesian concordance analysis supported the same phylogeny. A coalescent-based Bayesian method resulted in a very poorly resolved species tree. The topological agreement between concatenation and the Bayesian concordance analysis offers considerable support for a *Lophocebus* + *Papio* clade as the dominant relationship across the genome. However, the results of the Bayesian concordance analysis indicate that almost half the genome has an alternative history. As such, our results offer a well-supported phylogenetic hypothesis for the *Papio*/*Lophocebus*/*Theropithecus* trichotomy, while at the same time providing evidence for a complex evolutionary history that likely includes hybridization among lineages.

Introduction

The evolutionary history of the subtribe Papionina (sensu Strasser and Delson, 1987) is of particular interest because of its geographic, temporal, and ecological parallels to hominin evolution (Jolly, 2001; Holliday, 2003). Although the biology, paleontology, and behavior of these taxa are well studied, phylogenetic relationships within the Papionina have been the subject of long standing debate due to apparent incongruence between molecular and morphological data (Disotell, 1994; Collard and Wood, 2000; Harris, 2000; Collard and O'Higgins, 2001). Early taxonomies divided the Papionina into a small-bodied mangabey group and a larger-bodied baboon group (Booth, 1956; Napier and Napier, 1967; Osman Hill, 1972; Szalay and Delson, 1979; Strasser and Delson, 1987). Later, strong molecular and morphological evidence for mangabey diphyly prompted the replacement of earlier hypotheses with the grouping of terrestrial mangabeys (*Cercocebus*) with drills and mandrills (*Mandrillus*) and the placement of *Lophocebus*, the arboreal mangabeys or crested monkeys, in a second clade including the savannah baboons (*Papio*) and geladas (*Theropithecus*) (Barnicot and Wade, 1970; Cronin and Sarich, 1976; Groves, 1978; Disotell, 1992, 1994; van der Kuyl et al., 1995; Harris and Disotell, 1998; Fleagle and McGraw, 1999, 2002; Page et al., 1999; Harris, 2000; Page and Goodman, 2001; Gilbert and Rossie, 2007; Gilbert et al., 2009). However, the relationships among *Papio*, *Lophocebus*, and *Theropithecus* have proven extremely difficult to resolve.

In molecular phylogenetic studies, different loci have supported all possible relationships for this clade (Disotell, 1992, 1994; van der Kuyl et al., 1995; Page et al., 1999; Page and Goodman, 2001; Tosi et al., 2003). The apparent *Papio* + *Lophocebus* + *Theropithecus* (*P*/*L*/*T*)

---

* Corresponding author.
E-mail addresses: elaine.guevara@yale.edu (E.E. Guevara), msteiper@hunter.cuny.edu (M.E. Steiper).

[1] Current address: Department of Anthropology, 10 Sachem Street, New Haven, CT 06511, USA.

0047-2484/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.jhevol.2013.09.003

trichotomy is strongly reminiscent of the hominoid trichotomy. However, 'gene tree/species tree' approaches have failed to resolve these intergeneric relationships (Harris and Disotell, 1998; Harris, 2000). Recently, several studies compiling large concatenated DNA sequence datasets for primates have addressed relationships within the Papionina (Chatterjee et al., 2009; Fabre et al., 2009; Perelman et al., 2011; Springer et al., 2012; Pozzi et al., 2013). These studies supported different phylogenetic hypotheses for this group (Fig. 1), demonstrating that the results of concatenated analyses are highly dependent on the dataset and method used.

Concatenation, or supermatrix, approaches analyze multiple independent loci as a single data partition and their effectiveness at producing well-resolved and highly supported phylogenies has led to their widespread application (Edwards, 2009). The success of concatenation derives from its capacity to maximize what may be weak and widely dispersed phylogenetic signal from across loci to overcome background noise (Kluge, 1989; de Queiroz and Gatesy, 2007). However, the recognition that gene trees do not necessarily match species trees (Takahata and Nei, 1985; Wu, 1991; Ruvolo, 1997), has caused concern that concatenation falsely assumes that partitions share a single phylogenetic history (Degnan and Rosenberg, 2009) and thus reconstructs potentially misleading phylogenies or support values (Mossel and Vigoda, 2005; Kubatko and Degnan, 2007; Leaché and Rannala, 2011; Song et al., 2012). For example, concatenated analyses may be disproportionately influenced by one or a few particularly informative partitions, leading to the construction of a phylogeny that reflects only one aspect of a clade's evolutionary history, and which may not match the true predominant history across the genome (Maddison, 1997; Pabijan et al., 2012).

Increasing attention to the pervasiveness of gene tree discordance has prompted the recent development of several methods that account for this incongruence (see Edwards, 2009 for a review). These methods directly estimate a species tree, or a phylogeny of species, rather than an individual gene tree. Species tree methods analyze phylogenetic signal from many loci because each genome possesses many genes with independent histories. The multispecies coalescent is a phylogenetic approach developed from population genetics that models the stochastic variation that exists in gene coalescent times to reconstruct relationships by minimizing deep coalescence, or loci that do not coalesce prior to species divergence (Kingman, 1982; Maddison, 1997; Edwards et al., 2007). In its implementation using Bayesian inference, the multispecies coalescent utilizes multiple independent loci to simultaneously estimate probability distributions of independent gene trees and the correlated species tree given those gene trees (Edwards et al., 2007; Liu and Pearl, 2007; Liu et al., 2009). This method has been shown in simulation studies to recover correct species trees even when branch lengths are very short (Leaché and Rannala, 2011).

Another species tree method, Bayesian concordance analysis (BCA), estimates a primary concordance tree by adding clades to the tree in order of the frequency of their occurrence across a set of posterior distributions of genes trees (Ané et al., 2007). Analysis involves only one prior, an estimate of expected gene tree discordance, and no a priori assumptions about the underlying cause of gene tree discordance are made (Baum, 2007). Pozzi et al. (2013) recently compared supermatrix methods to BCA in a reanalysis of data from Perelman et al. (2011) and found that these methods provided different estimates of Papionina phylogeny. The present analysis also includes the multispecies coalescent method and it includes additional data to that of Perelman et al. (2011).

Given the observed high levels of gene tree heterogeneity and incompatibility of concatenated results within the Papionina *P*/*L*/*T* trichotomy (Fig. 1), we investigated whether species tree methods can yield improved phylogenetic resolution in this group. We compiled a 56 kilobase pair (kb) DNA sequence dataset for these taxa comprised of 57 independent mitochondrial, autosomal, and sex chromosome sequence partitions. To compare species tree methods to concatenation, we first phylogenetically analyzed this dataset as a single concatenated alignment using parsimony, likelihood, and Bayesian methods. Next, we applied species tree methods to these taxa using Bayesian multispecies coalescent (Liu, 2008) and Bayesian concordance (Ané et al., 2007). The multispecies coalescent can accommodate sequence data for more than one individual per species, allowing for the modeling of the allelic polymorphism within populations that causes incomplete lineage sorting (Maddison and Knowles, 2006; Liu et al., 2008). We gathered sequence data for additional individuals of our species of interest when available for inclusion in the multispecies coalescent analysis.

Materials and methods

Sequence data

We compiled a DNA sequence dataset following the 11 species papioninan taxonomy of Perelman et al. (2011). *Macaca mulatta* sequences were used as the outgroup for all analyses. *Rungwecebus kipunji* was not included in this analysis because too few sequence data are available for this species. Approximately 35 kb (54 loci) of these data are from Perelman et al. (2011) and approximately 21 kb (11 loci) were obtained from GenBank (see Table 1). Loci were considered independent if >25,000 bp apart, based on an estimated N_e of >10,000 and recombination rate of >0.025% following Ruvolo (1997). Distances between loci were estimated from their positions in the human genome, according to the UCSC Genome Browser (Kent et al., 2002). Four pairs of non-independent loci were combined (NPAS3 and NPAS3.2, RAG1 and RAG2, LRPPRC_169 and LRPPRC_171, and AFF2 and AFF2.2), as were all Y chromosome loci as none fell within the Human Pseudoautosomal Region, the small recombining portion of Y chromosome (Mangs and Morris, 2007). Each locus was aligned individually using ClustalW 1.83 (Thompson et al., 1994) and refined by eye. The total length of the alignment was 56,322 bp, and included 57 independent partitions.

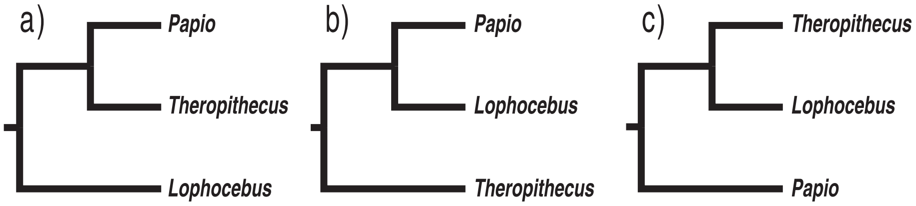

Figure 1. Alternative hypotheses supported by previous concatenated phylogenetic analyses. Hypothesis a) was supported by the concatenated analysis of Fabre et al. (2009) (63% ML bootstrap value). Hypothesis b) was supported by the concatenated analysis of Chatterjee et al. (nuclear DNA tree, 2009) (35% ML bootstrap value, 0.44 posterior probability), Perelman et al. (2011) (65% ML bootstrap value, 1.0 posterior probability), Pozzi et al. (2013) (81% ML bootstrap value, 1.0 posterior probability), and Springer et al. (2012) (50–70% ML bootstrap). Hypothesis c) was supported by the concatenated analysis of Chatterjee et al. (mitochondrial tree, 2009) (52% ML bootstrap value, 0.95 posterior probability) (and the non-concatenated BCA of Pozzi et al. (2013)).