

Taxon Combinations, Parsimony Analysis (PAUP*), and the Taxonomy of the Yellow-Tailed Woolly Monkey, *Lagothrix flavicauda*

Luke J. Matthews^{1*} and Alfred L. Rosenberger^{2,3}

¹Department of Anthropology, New York University, New York Consortium in Evolutionary Primatology (NYCEP), New York, NY 10003

²Department of Anthropology and Archaeology, The Graduate Center, Brooklyn College, The City University of New York, Brooklyn, NY 11210

³Department of Mammalogy, The American Museum of Natural History, New York Consortium in Evolutionary Primatology (NYCEP), New York, NY 10024

KEY WORDS cladistics; taxonomy; taxon combinations; PAUP; parsimony; tree inference; New World monkeys; ateline systematics

ABSTRACT The classifications of primates, in general, and platyrrhine primates, in particular, have been greatly revised subsequent to the rationale for taxonomic decisions shifting from one rooted in the biological species concept to one rooted solely in phylogenetic affiliations. Given the phylogenetic justification provided for revised taxonomies, the scientific validity of taxonomic distinctions can be rightly judged by the robusticity of the phylogenetic results supporting them. In this study, we empirically investigated taxonomic-sampling effects on a cladogram previously inferred from craniodental data for the woolly monkeys (*Lagothrix*). We conducted the study primarily through much greater sampling of species-level taxa (OTUs) after improving some character codings and under a variety of outgroup choices. The

results indicate that alternative selections of species subsets from within genera produce various tree topologies. These results stand even after adjusting the character set and considering the potential role of interobserver disagreement. We conclude that specific taxon combinations, in this case, generic or species pairings, of the primary study group has a biasing effect in parsimony analysis, and that the cladistic rationale for resurrecting the *Oreonax* generic distinction for the yellow-tailed woolly monkey (*Lagothrix flavicauda*) is based on an artifact of idiosyncratic sampling within the study group below the genus level. Some recommendations to minimize the problem, which is prevalent in all cladistic analyses, are proposed. *Am J Phys Anthropol* 137:245–255, 2008. © 2008 Wiley-Liss, Inc.

There have been significant shifts in the practice of primate systematics during the last decade. One involves the use of parsimony algorithms (often implemented in PAUP*) for reconstructing cladistic relationships. This approach has become ubiquitous. Another is the tolerance for highly split taxonomies (of the modern forms) that have been put forth on the basis of limited study, for example, naming newly discovered populations as distinct species or even genera, elevating the rank of recognized subspecies to the species level, reclassifying congeners into distinct genera, and multiplying the number of families within adaptive radiations (exemplary cases include Rylands et al., 2000; Groves, 2001). The rationale for these taxonomic revisions is often the phylogenetic position and distinctiveness of a taxon or taxa as inferred by parsimony analysis of characters. The scientific justification for such cladistic taxonomic philosophy is well articulated in Cracraft (1983, 2002). This unitary justification—phylogenetics—departs from more traditional taxonomic decisions, which have been based on the biological species concept in the designation of species and on the balance between phylogeny and adaptation in the case of the genus and suprageneric taxa. Given this more restricted rationale, new taxonomies pivot on the phylogenetic work that they represent, or the nonambiguous nature of the results.

In this work, we present a critique of one project that reflects the intersection of these trends, Groves' (2001)

analysis and interpretation of the alpha taxonomy of the New World woolly monkeys. From a taxonomic perspective, Groves' (2001) book, *Primate Taxonomy*, is revolutionary, reflecting all of the aforementioned types of revised taxonomic thinking, so his assessment deserves close scrutiny. For the platyrrhine primates, for example, he recognized 112 species. By comparison, two prior authoritative classifications of the previous generation, by Napier (1976) and Napier and Napier (1967), recognized 64 and 67 species, respectively.

Our main purpose is to increase awareness regarding the effect of incomplete taxonomic sampling on phylogenetic results for ateline primates presented in Groves (2001). Taxonomic sampling effects have been noted previously in theoretical (Siddall, 1995; Graybeal, 1998;

Grant sponsor: Tow Faculty Travel Fellowship, Brooklyn College (City University of New York), and a PSC CUNY Research Award.

*Correspondence to: Luke J. Matthews, Department of Anthropology, New York University, 25 Waverly Place, 4th Floor, New York, NY 10003. E-mail: ljm244@nyu.edu

Received 20 August 2007; accepted 18 March 2008

DOI 10.1002/ajpa.20859

Published online 23 May 2008 in Wiley InterScience (www.interscience.wiley.com).

Hawks, 2004), genetic (Collins, 2001, 2003; Hillis et al., 2003), and morphological (Rosenberger and Kearney, 1995; Sargis, 2007; Silcox, 2007) studies, but we show they operate quite powerfully on the sort of morphological data coded in discrete character states that can be scored relatively quickly from museum specimens and that form the basis for important reassessments of primate taxonomy (e.g., Groves, 2001). We show that the cladistic position obtained by Groves (2001) for the yellow-tailed woolly monkey (*Lagothrix flavicauda*), which is the sole justification for the resurrection of the generic nomen *Oreonax*, is likely an artifact of incomplete taxonomic sampling. The taxonomic-sampling effects remain even after we corrected inaccurate delimitations of metric character states, controlled for the logical interdependence of some characters, and ruled out interobserver reliability as an alternative cause of discrepancies in tree topologies. At a more theoretical level, our study addresses a somewhat different point than previous studies that focused on the importance of the number of taxa sampled (Siddall, 1995; Graybeal, 1998; Hawks, 2004; Hillis et al., 2003) in that we assess how the information content present in specific taxon combinations influences phylogenetic results, although Graybeal (1988) addresses this latter issue in the context of "long branch attraction" in genetic studies. Furthermore, the results of our study have obvious methodological relevance to the study of fossil primates, which are pseudorandomly sampled because of taphonomic constraints, and for which algorithmic parsimony has become a preferred method of inferring relationships (e.g., Kay et al., 1997; Strait et al., 1997; Strait and Grine, 1999; Seiffert et al., 2005). We conclude by making several suggestions on how to fix the problem, one of which is simply to sample more taxa when they are available, preferably of forms that are confidently known to be cladistically relevant.

Historical background

Fooden (1963) was the first to revise woolly monkey taxonomy during the modern era. He recognized two species, *L. lagotricha*, generally known as the (common) woolly monkey, and *L. flavicauda*, the (Peruvian) yellow-tailed woolly monkey. Unlike the familiar and widely distributed *L. lagotricha*, *L. flavicauda* is rare in two ways. First, having a very restricted, montane distribution in the wild, it was rarely seen or collected during much of the 20th century. In 1974, however, after a 50-year hiatus, Mittermeier et al. (Mittermeier et al., 1977; de Macedo Ruiz and Mittermeier, 1979) found that a population of these monkeys did exist in the Peruvian provinces of Amazonas and San Martín. Second, *L. flavicauda* is poorly represented in museum collections. Apart from 10 skins, only six adult skulls (and one juvenile skull) are catalogued, two each in the American Museum of Natural History (AMNH), the Natural History Museum (BMNH, London), and in the Museo de Historia Natural "Javier Prado," Lima, Peru. The latter were obtained by Mittermeier in the course of the 1974 expedition.

Groves (2001) raised the rank of *L. flavicauda* to the genus level, resurrecting the nomen *Oreonax* Thomas, 1927. Simultaneous with this assessment was a complementary raise in rank of the other woolly monkeys. Groves interpreted Fooden's (1963) four subspecies of woolly monkeys as distinct species (*L. lagotricha*, *L. cana*, *L. lugens*, *L.*

poepigii). As an arbitrary matter of convenience, we follow Groves' usage. Previously, Rosenberger et al. (1996), tending to agree with C. Groves, in part, because of our mutual familiarity with the specimens in the AMNH, also raised the rank of *L. flavicauda*. They proposed *Oreonax* as a subgenus of *Lagothrix*. Others have also begun to recognize *Oreonax* as a full genus (Rylands, 2000; Rylands et al., 2000) based on Groves' work. Given the narrow scope of this study, we use the nomina *Oreonax* and *L. flavicauda* interchangeably.

Groves' judgment to raise the rank of *L. flavicauda* was based on a parsimony analysis of cranial characters using PAUP 3.1.1 (Phylogenetic Analysis Under Parsimony; Swofford, 1993) to generate a hypothesis of cladistic interrelationships among several ateline species. He performed two sets of studies. One focused on the four modern ateline genera, *Alouatta*, *Lagothrix*, *Ateles*, and *Brachyteles*. Excluding *L. flavicauda* and *L. lagotricha*, the species used in this analysis were not identified but were "... picked at random from the AMNH collection," according to Groves (2001, p 93). A second study added two subfossil species (*Protopithecus* and *Caipora*; see Hartwig and Cartelle, 1996) to the data matrix. In both cases, he found that the species *L. flavicauda* was more closely linked with *Ateles* than with *Lagothrix lagotricha*. Arguing that the full assemblage of woolly monkeys cannot, therefore, be shown to be monophyletic, Groves opted to place the species *L. flavicauda* in a genus of its own.

MATERIALS AND METHODS

Materials

The samples forming the basis of this study are from collections of the American Museum of Natural History (AMNH), although we also examined the two adult and one juvenile *L. flavicauda* specimens in the British Museum. Our samples represent all the modern ateline genera and perhaps one-third or more of their species diversity, as well as three pitheciines, which were included as out-groups for the parsimony analyses. For genus *Lagothrix*, we used four of the recognized taxa, *L. lagotricha*, *L. cana*, *L. lugens*, and *L. poepigii*. We used only adult skulls and attempted to examine two females and two males for each species, choosing samples from a single locality whenever possible. The anatomical observations were scored by Matthews in consultation with Rosenberger. The following lists the 17 ateline and pitheciine taxa included in our study and their abbreviations as employed in the figures below: *Lagothrix lagotricha* (L. la.), *Lagothrix poepigii* (L.p.), *Lagothrix lugens* (L. lu.), *Lagothrix cana* (L. c.), *Lagothrix flavicauda* (L. f.); *Ateles chamek* (At. c.), *Ateles paniscus* (At. p.), *Ateles geoffroyi* (At. g.), *Ateles fusciceps* (At. f.), *Ateles belzebuth* (At. b.), *Ateles marginatus* (At. m.); *Brachyteles arachnoides* (B. a.), *Alouatta palliata* (Al. p.), *Alouatta seniculus* (Al. s.); *Pithecia pithecia* (P. p.); *Callicebus hoffmannsi* (C. h.), and *Callicebus donacophilus* (C. d.).

METHODS

Characters and character coding

Table 1 lists the 20 characters and character states as used by Groves (2001; Table 15, p 193), and Table 2 shows the original Groves data matrix. As part of our reanalysis, we assessed possible difficulties presented

TABLE 1. Character state definitions (from Groves 2001)

Character	Character state
1	Superolateral angle of orbit: rounded; squared
2	Lower margin of malar: evenly rounded; notched
3	Nasal profile: slightly concave; snub-nosed
4	Depth of zygomatic process of temporal: very shallow (4 mm or less); deeper (3.5 mm or more)
5	Length of zygomatic process of malar: restricted (21.5 mm or less); long (21.8 mm or more)
6	Glabella: flat; prominent
7	Postorbital constriction, viewed from above: smoothly rounded; angular
8	Pterionic foramen: tiny or virtually absent; large
9	Frontomalar sutural crest (in pterionic region): absent, or virtually so; present
10	Frontal bone at pterion: restricted; extends down to pterionic foramen
11	Internal nares: high, wide; narrow and angular
12	Median incisive foramen: absent; present
13	Foramen lacerum: sphenoid; on spheno-petrous suture
14	Glenoid fossa: mediolaterally concave; nearly flat
15	Vomer on floor of mesopterygoid fossa: flattened; raised
16	Medial pterygoid: free from lateral pterygoid; small, mainly fused to lateral pterygoid
17	Jaw angle: not enlarged; somewhat enlarged; very enlarged (ordered)
18	Cheekteeth: small; medium; enlarged [from Zingesser (1973)]
19	Alouattine dentition: no; yes [from Zingesser (1973)]
20	I1 size compared with M1: small; large [from Zingesser (1973)]

TABLE 2. Groves data matrix (from Groves, 2001)

Genus	Characters and states																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
<i>Oreonax</i>	1	1	0	0	0	0	1	1	1	1	0	1	0	1	1	1	1	2	0	1
<i>Lagothrix</i>	0	0	1	1	1	1	0	0	1	0	1	1	1	0	0	0	1	1	0	1
<i>Ateles</i>	0	1	1	1	0	0	0	1	1	1	0	0	0	{0,1}	1	1	0	0	0	2
<i>Brachyteles</i>	0	1	0	1	0	0	1	0	0	0	0	1	0	0	1	1	2	2	1	0
<i>Alouatta</i>	{0,1}	{0,1}	0	1	1	{0,1}	1	0	0	0	1	0	1	{0,1}	0	1	2	3	1	0

by each character. Thus, while we have hewed closely to Groves in replicating his methods and results to elaborate the essentials of his study we also justify why we treated some of the characters differently during our tree searching with PAUP*, either by eliminating some features altogether or by coding or weighting several in our own way. Ultimately, we removed two metric characters, ordered two more than Groves did initially, and reduced the weight of three others to minimize redundancy. In the end, we used 18 of his original 20 characters.

In the original study, all but one of the characters were unordered (Groves, 2001). However, in another departure from his protocol, we decided to order Character 18—Cheek teeth and Character 20—I1 size compared with M1. Our rationale is that they are multistate characters that reflect size categories of a morphological feature in the same way as Character 17—Jaw angle—which was ordered by Groves (2001).

Groves (2001) also weighted all characters equally. We chose to devalue the weight of Character 18—Cheek teeth, Character 19—Alouattine dentition, and Character 20—I1 size compared with M1, ascribing to each a value of 0.33. Our reason for doing this is that it is difficult to make the case that these three features are logically, that is, definitionally, independent. We would argue that one or more of these traits are overlapping as defined. For example, the “Alouattine dentition” is a characterization that relates to a number of resemblances shared by *Alouatta* and *Brachyteles* that were pointed out by Zingesser (1973), resemblances that include both absolute and relative cheek tooth size, and

incisor proportions. Thus, character 19, Alouattine dentition, is logically dependent upon characters 18 and 20. Its very definition overlaps with that of characters 18 and 20. Furthermore, the definition of character 18, which scores cheek tooth size, is clearly not logically independent from that of character 20, which scores the size of I1 relative to the size of M1. If character 20 had been defined as the size of I1, then it would be logically independent from character 18, but that was not the definition given in Groves (2001). All other characters were allotted a weight of 1.00, as in the original study, and so Characters 18, 19, and 20 considered together have a weight equivalent to each of the other characters singly. It should be noted that our decision about character weights was based on their logical independence and not on their developmental or ontogenetic independence. The importance of logical independence of characters used in phylogenetic analysis is established elsewhere (Pleijel, 1995; Lee and Bryant, 1999).

Parsimony analyses and the test of taxonomic-sampling bias

Our implementation of PAUP* intentionally follows Groves (2001) closely, although we used PAUP*4.0 10b (Swofford, 2002), a more recent version of the program. Using the same living taxa and data matrix as Groves, we successfully replicated his tree topology with an exhaustive search (see Fig. 3). The tree rooted with *Pithecia* was also inferred from the Groves data with all the *Pithecia* states set to 0 and states that Groves communicated to us personally. We performed a bootstrap analy-

sis of the same data matrix by resampling characters with replacement for 2000 iterations. We resampled only 15 of 20 characters to control for possible nonindependence (in this case developmental not logical) of characters from one another (Felsenstein, 2004). We did not attempt to replicate the result Groves obtained with the subfossils, which was an identical branching sequence for the moderns but with *Protopithecus* and *Caipora* inserted between the *Brachyteles* and *Lagothrix* nodes.

Having replicated the Groves' (2001) results, we used our newly collected data from the slightly modified character set to test how altering the composition of taxa forming the study group would change the resultant tree topology. We did this by increasing the number of species and also by constraining our taxonomic samples to combinations of five, as Groves had done. The goal of altering the taxonomic composition of the study group was to test whether the result of Groves could have been influenced by the small subset of extant atelines that he sampled. That is, was the result an artifact of the particular five species that were chosen "randomly"; would a different result have been obtained had four other species been pulled from the museum drawers? We conducted exhaustive parsimony searches in PAUP* with all possible combinations of five taxa that included *Oreonax* as well as one other sample from each genus. We also submitted the complete data matrix in Table 3 to an exhaustive parsimony search.

A note-regarding outgroups

Groves (2001) specified *Pithecia* and *Chiropotes* as outgroups, but scored the character states for both as 0 for all characters. As pitheciins, these genera are relatively divergent from atelines anatomically, and also from the ancestral pitheciine and atelid patterns, while also being quite homogeneous in sharing various derived dental and gnathic features. In this analysis, as in Groves' original study, they would effectively be replicates. Given these considerations, we included two species of *Callicebus* as outgroup taxa (which proved to have different character states) in addition to retaining *Pithecia pithecia*. We did so with the expectation that an "average" outgroup condition might more closely approximate the ancestral condition of the in-group than does the condition for any single outgroup taxon. Nevertheless, we were unable to replicate Groves' tree topology when using any outgroup unless it was specified as *Alouatta*. Thus, our tests for taxonomic sampling effects on the Groves (2001) result were carried out with *Alouatta* designated as the outgroup.

Interobserver validation of objectivity

We also attempted to rule out the possibility that unreliable character state coding, either on our part or by Groves (2001), rather than taxonomic sampling, was driving the results. To rule out such coding effects, we asked an impartial set of qualified observers to score them. We asked three other morphologists, two advanced graduate students in biological anthropology, and one practicing systematist who works on platyrrhines and other mammals, to score a subsample of the original taxa and characters. The subsample included *L. lagothricha*, *L. flavicauda*, *At. chamek*, *B. arachnoides*, *Al. seniculus*; that is, one form from each of the five genera at the core of Groves' study, and the features we asked our colleagues to score included all Groves' characters except

TABLE 3. Taxonomic distribution of character states scored by Matthews & Rosenberger

Taxon	Sample size	Taxon abbreviations	1	2	3	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
<i>Lagothrix lagothricha</i>	4	L. la.	0	{0,1}	1	1	{0,1}	0	{1,0}	0	1	1	1	{0,1}	0	0	{0,1}	{1,2}	0	1
<i>Lagothrix poeppigii</i>	4	L. p.	{0,1}	{0,1}	1	1	0	{0,1}	1	{0,1}	{0,1}	{0,1}	{0,1}	{0,1}	0	0	{1,2}	{1,2}	0	1
<i>Lagothrix lugens</i>	3	L. lu.	0	1	1	1	{0,1}	0	{0,1}	{0,1}	{0,1}	1	{0,1}	{0,1}	0	{0,1}	{1,2}	{1,2}	0	1
<i>Lagothrix cana</i>	4	L. c.	{0,1}	{0,1}	1	1	{0,1}	{0,1}	{0,1}	{0,1}	{0,1}	{0,1}	{0,1}	0	0	0	1	1	0	1
<i>Lagothrix flavicauda</i>	2	L. f.	1	1	0	0	1	1	1	1	0	1	0	1	1	1	1	2	0	1
<i>Ateles chamek</i>	4	At. c.	0	{0,1}	{0,1}	0	{0,1}	1	1	1	0	0	{0,1}	1	1	1	0	0	0	2
<i>Ateles paniscus</i>	4	At. p.	{0,1}	{0,1}	0	0	{0,1}	1	1	{0,1}	{0,1}	{0,1}	0	{0,1}	1	{0,1}	0	0	0	2
<i>Ateles goefferyi vellerosus</i>	4	At. g.	0	{0,1}	0	0	1	1	1	1	0	0	{0,1}	1	1	0	0	0	0	2
<i>Ateles fusciceps</i>	4	At. f.	0	{0,1}	{0,1}	0	1	{0,1}	{0,1}	1	{0,1}	{0,1}	{0,1}	{0,1}	{0,1}	{0,1}	{0,1}	{0,1}	0	2
<i>Ateles marginatus</i>	4	At. m.	0	0	{0,1}	{0,1}	{0,1}	{0,1}	1	1	0	0	{0,1}	{0,1}	{0,1}	1	{0,1}	0	0	2
<i>Ateles belzebuth</i>	4	At. b.	0	{0,1}	{0,1}	{0,1}	0	{0,1}	{0,1}	{0,1}	0	0	{0,1}	{0,1}	1	{0,1}	0	0	0	2
<i>Brachyteles arachnoides</i>	2	B. a.	0	{0,1}	0	0	1	0	{0,1}	{0,1}	0	1	0	1	1	1	2	{2,3}	1	0
<i>Alouatta palliata</i>	4	Al. p.	0	0	0	1	1	{0,1}	{0,1}	{0,1}	1	0	1	{0,1}	{0,1}	1	2	3	1	0
<i>Alouatta seniculus</i>	4	Al. s.	{0,1}	1	0	{0,1}	{0,1}	{0,1}	{0,1}	0	1	{0,1}	{0,1}	1	{0,1}	{0,1}	2	3	1	0
<i>Pithecia pithecia pithecia</i>	2	P. p.	0	0	0	0	1	{0,1}	0	1	0	0	0	0	1	0	0	0	0	0
<i>Callicebus hoffmannsi</i>	2	C. h.	{0,1}	1	0	0	{0,1}	0	0	0	0	0	1	0	0	0	0	1	1	0
<i>Callicebus donacophilus</i>	2	C. d.	0	1	0	0	1	0	0	0	0	0	{0,1}	0	1	0	1	1	0	0

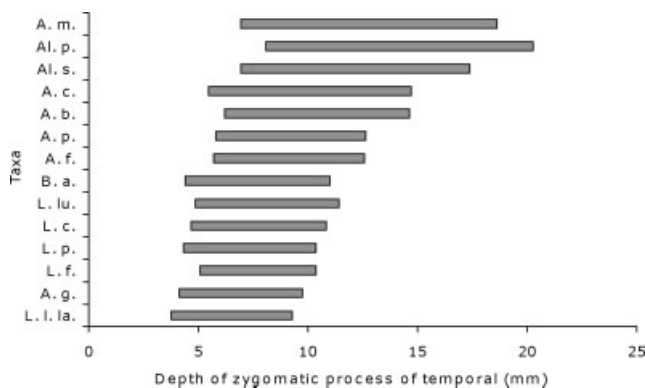


Fig. 1. Ranges of continuous character 4.

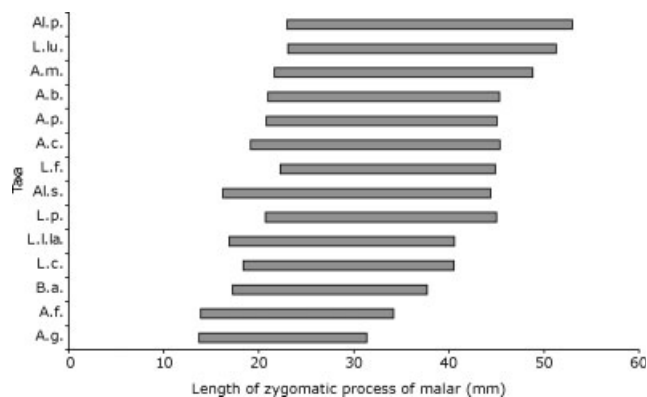


Fig. 2. Ranges of continuous character 5.

4, 5 (metric characters we eliminated on the basis of our previously collected measurements) and Character 19—Alouattine dentition. The latter was not included, because recognizing it presupposes familiarity with the particulars of a published work (Zingesser, 1973) as opposed to a general competence in primate morphology. We calculated pair-wise Kendall correlation coefficients between the coding produced by each of the three new observers, the scoring done by Matthews, and the data matrix of Groves to quantify interobserver agreement, using the computer program "R." Of course, complete agreement is desirable but not necessary for a robust phylogenetic investigation. We used the same basic conventions of PAUP* to generate parsimony trees based on the data from each observer in order to assess what influence, if any, interobserver disagreement might have on final tree topology.

RESULTS

Excluded characters

Two of Groves' (2001) characters seemed problematic after measurements were collected to better characterize them. Both were presented as having two discrete character states derived from continuously scaled quantitative measurements. Discrete character states are required for phylogenetic analysis in PAUP*. The problematic characters are Character 4—depth of zygomatic process of temporal and Character 5—length of zygomatic process of malar. Groves identified two states for each one (Table 1). We measured two males and two females of each taxon included in our study, with the exceptions of *L. lugens*, *L. flavicauda*, and *B. arachnoides*. Only one male and two females were measured in the case of *L. lugens*, and one of each sex was measured for the later two taxa. The availability of specimens in the museum collections restricted our sampling in these cases. The ranges for both characters exhibit substantial overlap when plotted by taxon. There are, in fact, no taxa with nonoverlapping ranges (Figs. 1 and 2).

Although both characters exhibit statistically significant differences in group means overall ($P < 0.01$), no individual pairs were significantly different for Character 4 (one way ANOVA followed by Games-Howell *post hoc* test performed in SPSS). Character 5 exhibited significant differences between pairs including *A. geoffroyi* and one of the following: *L. poeppigii*, *L. flavicauda*, *A. paniscus*, or *A. belzebuth*, but the latter four taxa are not significantly different from each other or from any

other taxa in the study. Thus, there are no homogeneous subsets that can be formed, as is recommended when continuously scaled character data are converted into a set of discrete states (Simon, 1983; Rae, 1998). Furthermore, ranges for the species do not exhibit nonoverlapping regions. As the data stand, we could not construct justifiable discrete categories for these characters, so we excluded them from the phylogenetic analysis.

Parsimony variations and taxonomic-sampling bias

With our more recent version of PAUP*, we successfully replicated the most parsimonious consensus tree found by Groves (2001) using his published data, with the same tree length, 31 steps (Fig. 3A), using *Alouatta* as the outgroup. However, our bootstrap analysis revealed that the nodes linking *Oreonax* with *Ateles* had relatively low support (Fig. 3B). When we rooted the tree with *Pithecia* instead of *Alouatta*, the topology shifted, which further implies that this cladogram is not especially robust. The dichotomous branching sequence devolved into a trichotomy of ateline lineages. Although the *Pithecia*-rooted topology still produced the *Oreonax*-*Ateles* clade to the exclusion of *Lagothrix*, it also shows a switch in the linkage of *Brachyteles* and *Alouatta* into a new clade (Fig. 3B), demonstrating that the cladistic position of *Lagothrix lagotricha* with respect to the two definable clades is notably unresolved (Fig. 3B).

When our 48 tree searches sampled an enlarged assortment of ateline species (or OTUs, Operational Taxonomic Units) as the study group, using *Alouatta* as the outgroup to be consistent with Groves (2001), new topologies were produced. Four examples of these five-taxon combinations are shown in Figure 4. It shows that *L. flavicauda* links with *Ateles* in a trichotomy when *L. lugens* is included in the study group and *L. lagotricha* is eliminated. When *L. cana* is combined in a matrix with *Al. seniculus*, *L. cana* is grouped with *Ateles* and *L. flavicauda* splits to become a sister taxon. When these taxa are combined with *Al. palliata*, *Brachyteles* is more closely linked with *At. paniscus* than *L. flavicauda*, and *L. lagotricha* becomes a sister group to the other three (see Fig. 4).

Overall, we found that Groves' (2001) tree topology was obtained primarily when certain species were combined in the matrix, mainly when *L. lagotricha* was used in conjunction with *Al. seniculus* (Table 4). Whereas matrices that included both *Al. palliata* and *Al. senicu-*

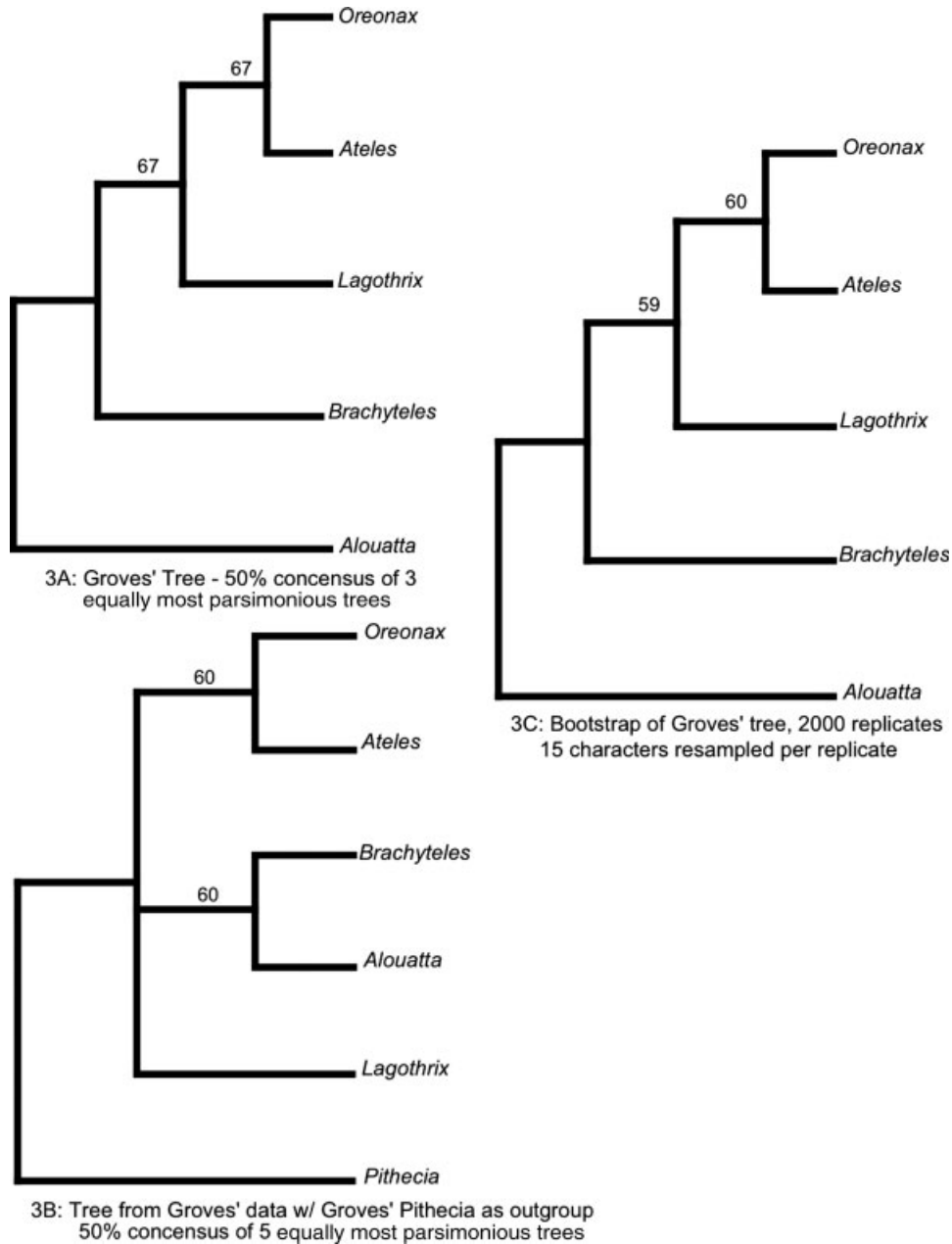


Fig. 3. Replication of Groves (2001) topology and modifications by bootstrap and outgroup.

lus generated an equal number of robust cladograms resolving to one most parsimonious tree, 31% of the *Al. seniculus* trees returned an *Oreonax-Ateles* clade but only 21% of the *Al. palliata* trees produced the same result. When a *Lagothrix* OTU other than *L. lagotricha* was used, the Groves tree is not a frequent outcome, and other fully resolved topologies occur when different combinations of species of *Lagothrix*, *Alouatta*, and *Ateles* are used in the analysis.

We also conducted a parsimony analysis with 14 ateline taxa plus two species of *Callicebus* and one species of *Pithecia* (see Fig. 5) as outgroups. The data produced 45,015 equally parsimonious trees, because all the possible permutations at the tips of the tree are unresolved. Given that *Alouatta* is extraordinarily divergent and

represented by two anatomically highly distinct species, it is not surprising that howlers appear as a monophyletic set of species. Other congeneric species are clustered as monophyletic polytomies, with one exception, the splitting of the genus *Callicebus*. The tree does link *L. flavicauda* with *Ateles*, but it also generates an *Alouatta-Brachyteles* clade. Moreover, it moves the other four *Lagothrix* species toward a node that includes *Callicebus hoffmannsi*, with *Callicebus donacophilus* placed outside the cluster (see Fig. 5).

Results from multiple observers

The scoring data collected by multiple observers was significantly correlated. Table 5 shows the pair-wise

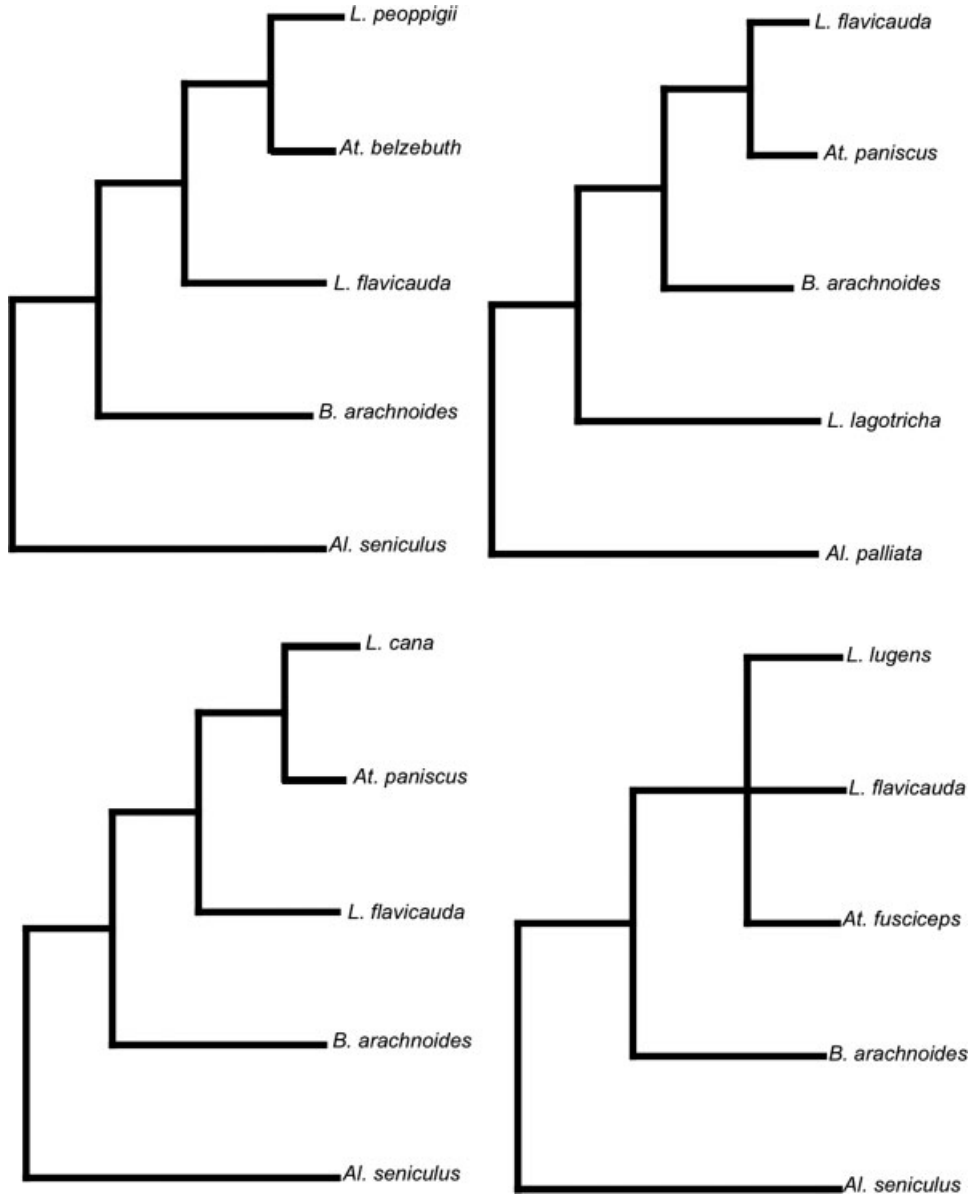


Fig. 4. Example parsimony trees from 5-taxon searches.

Kendall correlations for all observers, which were all significant at the 0.001 level after the Bonferroni correction. The Matthews-Groves pair had the highest correlation value ($\tau = 0.77$) and produced the same phylogenetic tree as a consequence (see Fig. 6). However, the data matrix generated from the other observers, when submitted to PAUP*, produced different tree topologies. Two thirds of the resulting most parsimonious trees

produced different tree topologies. Two thirds of the resulting most parsimonious trees

TABLE 4. Results of 5-taxon searches

Have one mpt	Have L.f. and Ateles clade	
Outgroup = <i>Alouatta seniculus</i> ^a		
16/24	7/24	note: 5 of these 7 are when L. la. is used. This indicates that if L. la is used it results in the Groves' tree with 5 of 6 <i>Ateles</i> species Ave. number parsimony informative characters = 4.21
66.70%	31%	
Outgroup = <i>Alouatta palliata</i> ^b		
16/24	5/24	note: none of these 5 are when L. la. is used Ave. number parsimony informative characters = 5.96
66.70%	20.80%	

MPT, most parsimonious tree.

^a Five of these seven are when L. la. is used. This indicates that if L. la is used it results in the Groves' tree with 5 of 6 *Ateles* species average number parsimony informative characters = 4.21.

^b None of these five are when L. la. is used. Average number parsimony informative characters = 5.96.

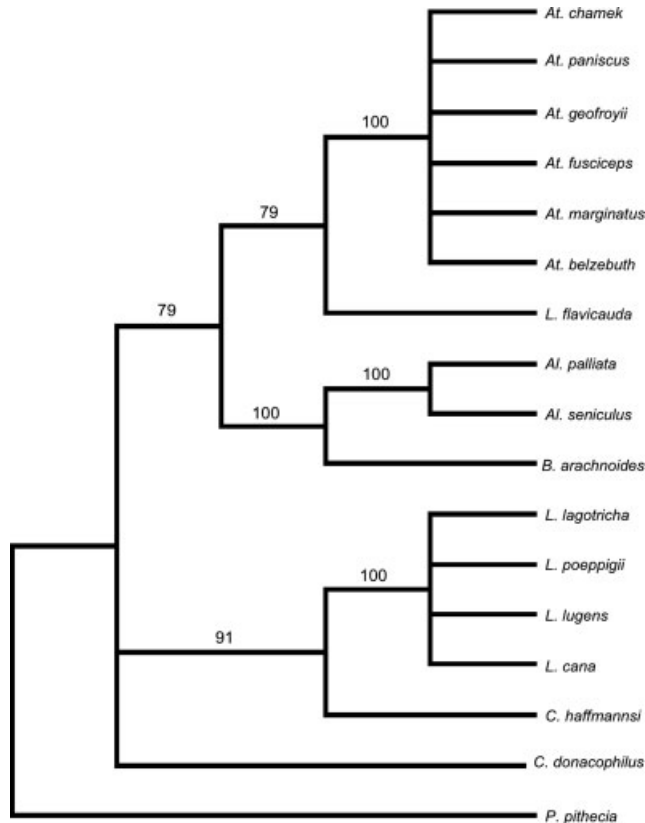


Fig. 5. 50% Majority-rule consensus tree of 45,015 equally parsimonious trees (numbers at nodes represent percentage of trees exhibiting that clade, numbers are *not* bootstrap values).

shared a polytomous clade of *Ateles* and the two *Lagothrix* species. No other clades were shared between more than half the trees based on these different observers. One tree linked *L. flavicauda* and *L. lagotricha* in a monophyletic group.

DISCUSSION

In seeking to optimize the results of algorithmic cladistic analyses, various studies have debated the value of increasing the sample size of a study group by adding taxa (e.g., Siddall, 1995; Graybeal, 1998; Hillis et al., 2003), as opposed to adding characters. We address a finer point than sheer sample size, the impact of the *information content* of the taxa sampled. Our study suggests that this is a significant issue for it appears that the *taxon combinations* that comprise the study group have a strong influence on the results of a parsimony investigation. Evidently, when certain ateline taxa are selected for analysis, namely *Alouatta seniculus* and *Lagothrix lagotricha*, the Groves (2001) topology is obtained. Groves did not specify which species of *Alouatta* he used in his study, but he did specify that *Lagothrix lagotricha* was the only species of woolly monkey included besides *Lagothrix flavicauda*. When other *Lagothrix* species are used as OTUs, it is common for one of them to link with a species of *Ateles*, but that species is not necessarily *L. flavicauda*. As shown in Figure 4, it may be *L. cana*, *L. poeppigii* or *L. lugens* + *L. flavicauda* in a trichotomy. This suggests us that the specificity of the cladistic link that undergirds Groves' move to

TABLE 5. Kendall correlations between data from different observers

	Observer A	Observer B	Observer C	Matthews	Groves
Observer A	1.000	0.342	0.511	0.551	0.438
Observer B	–	1.000	0.377	0.460	0.392
Observer C	–	–	1.000	0.365	0.354
Matthews	–	–	–	1.000	0.770
Groves	–	–	–	–	1.000

$P < 0.001$ after Bonferroni correction.

revive *Oreonax* as a genus, the alleged monophyly of *L. flavicauda* and *Ateles*, is an artifact of taxonomic sampling deriving from the limited set of in-group taxa that was included in his study.

The potential for artifactual results arising from sampling error in the form of study group composition have also been noticed by Collins (2001, 2003) in molecular phylogenetic studies within the genus *Ateles* and across the four modern ateline genera, indicating that our finding may have bearing upon genotypic as well as phenotypic data (see also Siddall, 1995; Hillis et al., 2003). At the ordinal taxonomic level, other recent studies of primate relationships (e.g. Sargis, 2007; Silcox, 2007) have also shown that taxon combinations have a similar effect, as might be expected, because PAUP* treats each terminal taxon as an equivalent OTU, irrespective of taxonomic rank. If the *L. flavicauda* cladistic linkage with *Ateles* is not regarded as an artifact, then neither can the results for *L. cana* and *L. poeppigii* for the same data set. Thus, one might argue that those who accept the *L. flavicauda* result of Groves as justification for reviving *Oreonax* are logically required to assign unique genus names to *L. cana* and *L. poeppigii*, because the same set of characters equally supports the generic distinctiveness of these taxa as well.

The difficulty of treating *L. flavicauda* as its own genus is further emphasized by the results from our parsimony search including all taxa that we studied. At first glance, the consensus tree of the 45,015 equally parsimonious trees could be interpreted to support the *Oreonax* hypothesis because *L. flavicauda* is linked with *Ateles* (see Fig. 5), and the other *Lagothrix* species form a monophyletic clade. However, the tree also links *Alouatta* and *Brachyteles*, a result that is in conflict with most morphological and molecular phylogenies of the atelines (e.g. Ford, 1986; Kay, 1990; Schneider, 2000; Schneider et al., 2001; Rosenberger, 2002). The choice of which presumptive clade one accepts or rejects from the same tree thus becomes entirely arbitrary. This same result is obtained when using Groves' (2001) own *Pithecia* character states as an outgroup (Fig. 3C). In other words, a stronger case for generic status of *Oreonax* would be evident if the analysis was consistent with well-established relationships of taxa *within the study group*.

One factor that could explain the differences between our results and Groves (2001) relates to the greater intraspecific variation exhibited in our data set as compared to the Groves data (Tables 2 and 3). One might be concerned that the concomitant reduction of parsimoniously informative characters in our five taxon trees caused by intraspecific polymorphism invalidates our taxon selection test of the Groves hypothesis (Table 4:

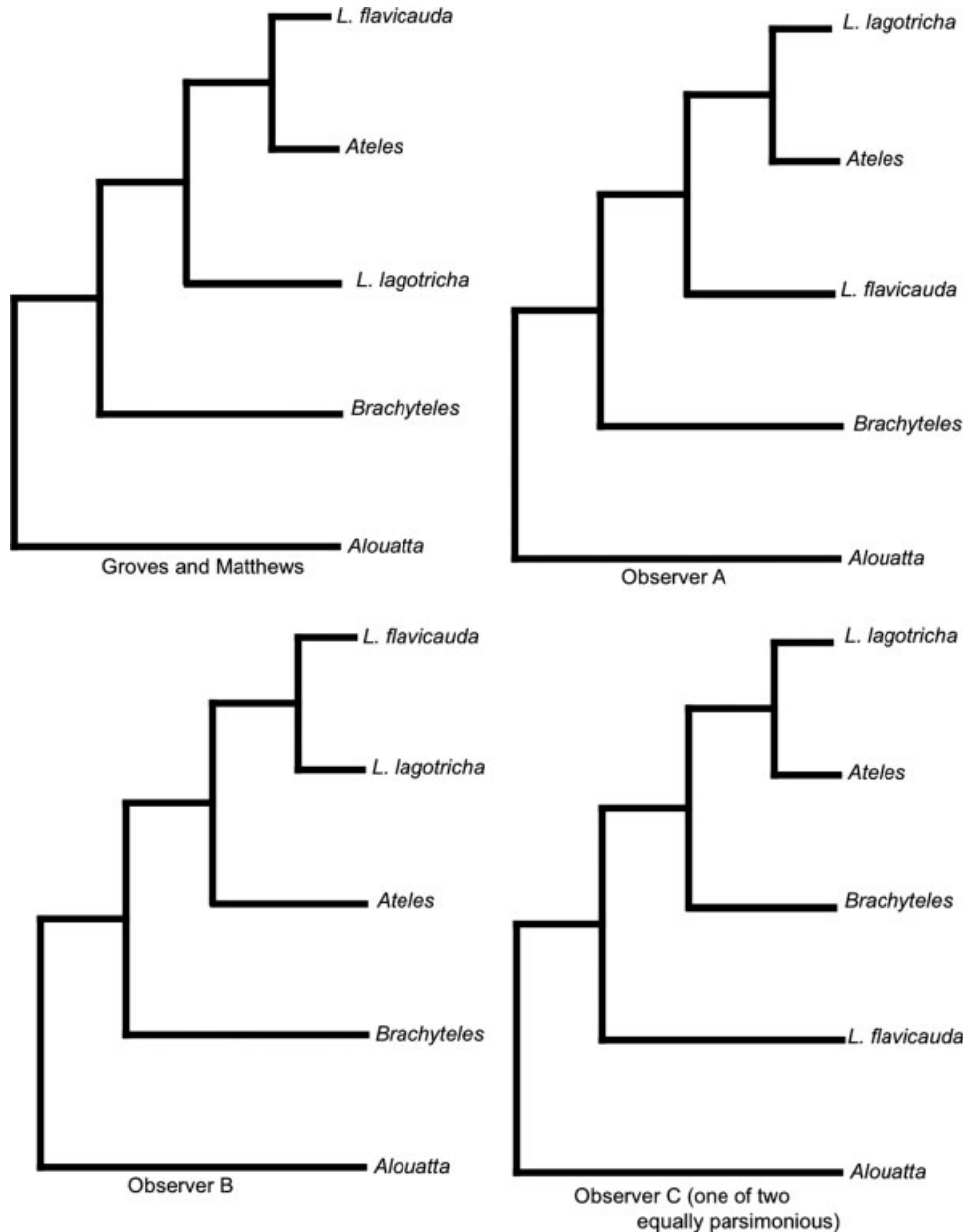


Fig. 6. Parsimony trees resulting from multiple observers.

our average number of parsimony informative characters [PIC] ranged from 4.21 to 5.96, while Groves had 13 PIC). The extensive intraspecific character polymorphism recorded by Matthews did not, however, in and of itself, cause our five taxon trees to differ from those of Groves, because his topology was consistently obtained when *Lagothrix lagotricha* and *Alouatta seniculus* were selected for analysis. *Alouatta seniculus* was likely the howler monkey species that Groves used in devising his study group. Furthermore, when our analysis included all 14 ateline taxa, 17 of the 18 characters were parsimoniously informative. Additionally, while our data set shows more intraspecific variation, it also better resembles the Groves matrix when compared with data from three other qualified observers (see below). Thus, we feel our observed level of intraspecific variation is justifiable,

and additionally that it does not confound our taxon selection test of the Groves hypothesis.

The results of our correlation analysis revealed that the scores from the independent observers were significantly correlated with one another (Table 5). Matthews and Groves had the highest correlation value ($\tau = 0.77$), and their respective data matrices result in identical tree topologies. This suggests that the differences between our results and Groves (2001) are not the product of imprecise character scoring by either party. Significant character coding subjectivity is indicated, however, by the result of the independent observers. The next most highly correlated pair of observers, Matthews and Observer A ($\tau = 0.55$), resulted in different tree topologies. Thus, for this study's characters and taxa, somewhere between 0.55 and 0.77 agreement is required for

the obtainment of tree topologies that are insensitive to interobserver differences. Data from different characters or taxa may require higher or lower levels of interobserver correlation. The most objective characters appear to be 1 (Superolateral angle of orbit), 2 (Lower margin of malar), and 11 (Internal nares). For these characters, no two observers disagreed regarding more than one taxon. Our result is both comforting, because complete interobserver agreement was not needed for robust phylogenetic results, and cautionary, because, in this case, data from a majority of qualified observers produced little consensus on tree topology.

CONCLUSIONS

Our study reveals possibly underappreciated idiosyncrasies of modern algorithm-driven parsimony analyses as exemplified by PAUP*, potential confounds that are more acute in algorithm driven studies than in conventional character analyses due to the contained nature of computer-based protocols. It suggests that many cladistic solutions can be returned in a study using a reasonable number of characters and taxa without any objective way of determining which tree, if any, or which parts of trees, are phylogenetically sensible. Having said that, we also emphasize that we are not surprised by this finding for, even though our study group is a relatively small collection of taxa—sometimes numbering only four or five taxonomic units—the underlying data are morphology, which means that the data are complex and have probably evolved in a complex fashion.

Although the substance of our study is geared to the cladistics of a living radiation, the same principles apply to fossils, only more so (Hawks, 2004). Although systematists have implicitly operated under the supposition that there is less of a danger of inadvertent sample bias in forming study groups than out-groups, there is also no way of really knowing what measure of biodiversity has been lost due to extinction, or unavailability of information for myriad other reasons, when one composes a target sample for analysis. Our study, coupled with others conducted on higher taxa and more divergent lineages (e.g., Sargis, 2007; Silcox, 2007), shows that OTU combinations of actual taxa and of clades are a confounding variable that potentially skews results.

Regarding the affinities and classification of *L. flavicauda*, the evidence we have found does not endorse either of Groves' (2001) interpretations, that (1) the yellow-tailed woolly monkey is more closely related to spider monkeys than any of the other woolly monkeys are; and (2) that *L. flavicauda* ought to be removed to a separate genus. In the matter of classification alone, even if one accepts for sake of argument that the cladistic position of *L. flavicauda* remains somewhat unsettled and the monophyly of genus *Lagothrix* is not effectively corroborated by this study, there seems to be no sufficient reason to separate *L. flavicauda* generically. The root cause behind all of these uncertainties may be that the characters selected are not appropriate to the question at hand. Moreover, given the absence of cladistic knowledge as we see it, there is little phenetic evidence to justify separating *L. flavicauda* at the genus level.

Regarding the more general issues pertaining to the use of PAUP* in morphological studies, we make the following points. There is an obvious tension between the mechanistic objectivity of an algorithm that manages large calculations and the mindful, experience-based

data constructions of a systematist who supplies the machine its information. Our study relates equally to the intentional taxonomic sampling strategies, which are aspects of the project formulation in neontological studies and to the quasi-random sampling that is inevitable in paleontology—both are impacted by the normal consequences of evolutionary processes as well as being intrinsically difficult steps in any systematics research program. Both are prey to a structural fact of phylogenetic inference under parsimony, the unanticipated taxon combinations, which drive results. There may be no easy way of getting around this type of bias procedurally, and the authors themselves are divided on this although broad sampling of available and cladistically pertinent specimens (populations) at the phylogenetic species level may ameliorate the bias. Another possible solution to minimize the influence of taxon combinations that might be modeled in further investigations is proportional character weighting, especially on a scale that is more trenchant than the small numerical differences usually applied to ordered character states. This may have the effect of shifting the center of gravity of a study from taxon to character. The logic would be similar to the “molecular scaffold” approach that was broadly sketched out by Pilbeam (1996) (under different terminology) and is now popular for rooting secure monophyletic groups based on molecular results. Certain relationships could be set on the basis of molecular data, and morphological characters then weighted, ordered, and polarized to maximize the probability of obtaining these groups. The phylogenetic positions of taxa could then be investigated within the scaffold with morphological data using the parameters so obtained. If certain characters are allowed to drive the parsimony analysis overall—specifically for the groups that are part of the study sample—this might compensate for the loss of information arising from the chaotic absence of taxa in different segments of the tree.

ACKNOWLEDGMENTS

LJM thanks the New York Consortium in Evolutionary Primatology for academic support of this project. We thank T Disotell for use of NYU's PAUP* program. J. Brinkworth, L. Tallman, M. Tejedor, and S. Wang all gave generously their time and expertise early in the project. S. Worthington was particularly helpful through his advice on homogeneous subset coding and by reading a draft of the manuscript. We also thank C. Groves for his helpful correspondence. We thank the editor-in-chief, the associate editor, and the one anonymous reviewer for their helpful comments on the manuscript.

LITERATURE CITED

- Collins AC. 2001. The importance of sampling for reliable assessment of phylogenetics and conservation among neotropical primates: a case study in spider monkeys (*Ateles*). *Primate Rep* 61:9–30.
- Collins AC. 2003. Atelinae phylogenetic relationships: the trichotomy revived? *Am J Phys Anthropol* 124:285–296.
- Cracraft J. 1983. Species concepts and speciation analysis. In: Johnston RF, editor. *Current ornithology*, Vol. 1. New York: Plenum. p 159–187.
- Cracraft J. 2002. The seven great questions of systematic biology: an essential foundation for conservation and the sustainable use of biodiversity. *Ann Missouri Bot Garden* 89:127–144.

- de Macedo Ruiz H, Mittermeier RA. 1979. Rediscovery of the Peruvian primate *Lagothrix flavicauda* (Humboldt 1812) and first observations on its biology. *Rev Ciencias Univ Nacional Mayor San Marcos* 71:78–92.
- Felsenstein J. 2004. Inferring phylogenies. Sunderland, MA: Sinauer Associates. 664 p.
- Fooden J. 1963. A revision of the woolly monkeys (genus *Lagothrix*). *J Mammal* 44:213–247.
- Ford SM. 1986. Systematics of the New World monkeys. In: Swindler DR, Erwin J, editors. *Comparative primate biology, Vol. 1: systematics, evolution and anatomy*. New York: Alan R. Liss. p 73–135.
- Graybeal A. 1998. Is it better to add taxa or characters to a difficult phylogenetic problem? *Syst Biol* 47:9–17.
- Groves C. 2001. *Primate taxonomy*. Washington: Smithsonian Institution Press. 350 p.
- Hartwig WC, Cartelle C. 1996. A complete skeleton of the giant South American primate *Protopithecus*. *Nature* 381:307–311.
- Hawks J. 2004. How much can cladistics tell us about early hominid relationships? *Am J Phys Anthropol* 125:207–219.
- Hillis DM, Pollock DD, McGuire JA, Zwickl DJ. 2003. Is sparse taxon sampling a problem for phylogenetic inference? *Syst Biol* 52:124–126.
- Kay RF. 1990. The phyletic relationships of extant and fossil Pitheciinae (Platyrrhini, Anthropoidea). *J Hum Evol* 19:175–208.
- Kay RF, Ross C, Williams BA. 1997. Anthropoid origins. *Science* 275:797–804.
- Lee DC, Bryant HN. 1999. A reconsideration of the coding of inapplicable characters: assumptions and problems. *Cladistics* 15:373–378.
- Mittermeier RA, de Macedo-Ruiz H, Luscombe BA, Cassidy J. 1977. Rediscovery and conservation of the Peruvian yellow-tailed woolly monkey (*Lagothrix flavicauda*). In: Prince Rainier III, Bourne GH, editors. *Primate Conservation*. New York: Academic Press. p 95–116.
- Napier JR, Napier PH. 1967. *A handbook of living primates*. London: Academic Press.
- Pilbeam D. 1996. Genetic and morphological records of the Hominoidea and hominid origins: a synthesis. *Mol Phylog Evol* 5:155–168.
- Pleijel F. 1995. On character coding for phylogeny reconstruction. *Cladistics* 11:309–315.
- Rae TC. 1998. The logical basis for the use of continuous characters in phylogenetic systematics. *Cladistics* 14:221–228.
- Rosenberger AL. 2002. Platyrrhine paleontology and systematics: the paradigm shifts. In: Hartwig WC, editor. *The primate fossil record*. Cambridge: Cambridge University Press. p 151–159.
- Rosenberger AL, Kearney M. 1995. The power of fossils, the pitfalls of parsimony—platyrrhine phylogeny. *Am J Phys Anthropol* 38(S20):130.
- Rosenberger AL, Norconk MA, Garber PA. 1996. New perspectives on pitheciines. In: Norconk MA, Rosenberger AL, Garber PA, editors. *Adaptive radiations of neotropical primates*. New York: Plenum. p 329–333.
- Rylands A. 2000. An assessment of the diversity of New World primates. *Neotropical Primates* 8:61–63.
- Rylands A, Members of the Primate Specialist Group. 2000. *Oreonax flavicauda*. In: IUCN 2004. 2004 IUCN Red List of Threatened Species. <www.redlist.org>.
- Sargis EJ. 2007. The postcranial morphology of *Ptilocercus lowii* (Scandenti, Tupaiidae) and its implications for primate supra-ordinal relationships. In: Ravosa MJ, Dagosto M, editors. *Primate origins*. New York: Springer. p 51–82.
- Schneider H. 2000. The current status of the New World monkey phylogeny. *Ana Acad Brasil Cie* 72:165–172.
- Schneider H, Canavez FC, Sampaio I, Moreira MA, Tagliaro CH, Seuánez HN. 2001. Can molecular data place each neotropical monkey in its own branch? *Chromosoma* 109:515–523.
- Seiffert ER, Simons EL, Clyde WC, Rossie JB, Attia Y, Bown TM, Chatrath P, Mathison ME. 2005. Basal anthropoids from Egypt and the antiquity of Africa's higher primate radiation. *Science* 310:300–304.
- Siddall ME. 1995. Another monophyly index: revisiting the jack-knife. *Cladistics* 11:33–56.
- Silcox MT. 2007. Primate taxonomy, Plesiadapiformes, and approaches to primate origins. In: Ravosa MJ, Dagosto M, editors. *Primate origins*. New York: Springer. p 143–178.
- Simon C. 1983. A new coding procedure for morphometric data with an example from periodical cicada wing veins. In: Felsenstein J, editor. *Numerical taxonomy*. New York: Springer-Verlag. p 378–382.
- Simpson GG. 1945. The principles of classification and a classification of mammals. *Bull Am Museum Nat Hist* 85:1–XVI, 1–350.
- SPSS 13.0 for Windows. 2004. Copyright: SPSS Inc.
- Strait DS, Grine FE. 1999. Cladistics and early hominid phylogeny. *Science* 285:1209.
- Strait DS, Grine FE, Moniz MA. 1997. A reappraisal of early hominid phylogeny. *J Hum Evol* 32:17–82.
- Swofford DL. 1993. PAUP: Phylogenetic analysis using parsimony. Version 3.1.1. Washington: Smithsonian Institution.
- Swofford DL. 2002. PAUP*: Phylogenetic analysis using parsimony (*and other methods). Version 4.0 10b. Sunderland, MA: Sinauer Associates.
- The R Foundation. 2005. The R Foundation for statistical computing. Version 2.1.1.
- Zingesser MR. 1973. Dentition of *Brachyteles arachnoides* with reference to alouattine and ateline affinities. *Fol Primatol* 20:351–390.