Original Article

Folia Primatol 2005;76:207–221
DOI: 10.1159/000086022

Received: June 11, 2004
Accepted after revision: February 10, 2005

Impact of Methodological Choices on Assessments of the Reliability of Fossil Primate Phylogenetic Hypotheses

Marcos Nadal-Roberts[a] Mark Collard[b, c]

[a] Department of Philosophy and Laboratory of Human Systematics, University of the Balearic Islands, Palma de Mallorca, Spain; [b] Department of Anthropology and Sociology, University of British Columbia, Vancouver, Canada; [c] AHRB Centre for the Evolutionary Analysis of Cultural Behaviour, University College London, London, UK

Key Words
Phylogeny · Cladistics · Fossil primates · Hominoids · Papionins · Isometric size correction · Allometric size correction · Outgroup composition · Character correlation

Abstract

It has been argued in several recent studies that conventional craniodental characters cannot be assumed to be reliable for the purposes of reconstructing primate phylogenetic relationships and that as a consequence little confidence can be invested in published fossil primate phylogenies. Here, we evaluate this claim by revisiting the analyses reported in one of these studies [Collard and Wood, 2000]. Specifically, we investigate whether the use of alternative methodological procedures would have altered their findings. We focus on three key issues: (1) size correction, (2) outgroup composition and (3) non-phylogenetic correlation among characters. Our analyses suggest that the results of Collard and Wood [2000] were not affected by the size correction method they used or by the outgroup they employed. Our analyses also suggest that their results were not affected by their decision to ignore developmental, functional and other non-phylogenetic correlations among the characters in their data sets. Accordingly, our study supports the assertion that conventional craniodental characters cannot be assumed to be reliable for reconstructing primate phylogenetic relationships. This in turn suggests that many published fossil primate phylogenies may be unreliable.

KARGER
Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

©2005 S. Karger AG, Basel
0015–5713/05/0764–0207$22.00/0

Accessible online at:
www.karger.com/fpr

Mark Collard, Department of Anthropology and Sociology, University of British Columbia
6303 NW Marine Drive, Vancouver, V6T 1Z1 (Canada)
Tel. +1 604 822 4845, Fax +1 604 822 6161
E-Mail mark.collard@ubc.ca

Introduction

In the last few years, a number of researchers have suggested that our grasp of fossil primate phylogeny may be much more tenuous than is generally appreciated. Lieberman [1995, 1997, 1999, 2000; Lieberman et al., 1996], for example, has criticised the widespread assumption that all skeletal resemblances among primates are heritable and therefore potentially informative regarding phylogeny. This assumption is problematic, according to Lieberman, because bone size and shape are now known to be greatly influenced by interactions between the skeleton and its mechanical environment. The corollary of this, Lieberman suggests, is that many cranial resemblances are likely to be a result of similar behaviours rather than descent from a common ancestor and are therefore not useful for phylogenetic reconstruction.

Other researchers have questioned our knowledge of fossil primate phylogeny in light of phylogenetic analyses of extant primate craniodental morphology. Hartman [1988], for instance, sought to determine whether or not primate teeth are phylogenetically informative by carrying out a cladistic analysis of extant hominoid molar morphology and then comparing the resulting phylogeny with the well-supported consensus molecular phylogeny for the group. He found that the phylogeny yielded by the molar data differed from the molecular phylogeny and concluded as a consequence that teeth and probably other morphological features cannot be assumed to be informative with regard to primate phylogeny. In a similar vein, Pilbeam [1996] argued that the existence of discrepancies between the hard-tissue and molecular phylogenies that have been produced for the extant hominoids means that hard-tissue characters cannot simply be assumed to be a good guide to the phylogenetic relationships of the Miocene apes. More recently, Collard and Wood [2000] applied Hartman's protocol to cranial and dental data from the extant hominoids and papionins with a view to assessing the likely reliability of published phylogenetic hypotheses for the early hominids. They found that the craniodental data sets produced strongly supported phylogenies that are incongruent with the groups' consensus molecular phylogenies. Based on this, Collard and Wood [2000] argued that conventional craniodental characters cannot be assumed to be reliable for the purposes of reconstructing primate phylogenetic relationships. The corollary of this, they suggested, is that existing hypotheses concerning the phylogenetic relationships of the fossil hominids may be unreliable.

Given that a robust phylogeny is crucial for progress in understanding the evolution of humans and other primates, the suggestion that conventional craniodental characters cannot be assumed to be reliable for the purposes of reconstructing primate phylogenetic relationships and that published fossil primate phylogenies are therefore probably not reliable warrants careful scrutiny. In this paper, we report the results of a study that examined the possibility that methodological aspects of the analyses of Collard and Wood [2000] confounded their results. Our main goal was to determine whether the use of alternative procedures would have produced morphological phylogenies that were congruent with the molecular phylogenies in contrast to the morphological phylogenies that Collard and Wood [2000] obtained, which, as noted above, were markedly incongruent with the molecular phylogenies. We focused on three key issues: (1) size correction, (2) outgroup composition and (3) non-phylogenetic correlation among characters.

Analysis 1: Size Correction[1]

Collard and Wood [2000] reported the results of three analyses. One analysis employed a qualitative data set for the extant hominoids. The other two analyses utilised quantitative data sets, one for the extant hominoids and one for the extant papionins. In order to employ quantitative data in a cladistic analysis, it is necessary to adjust them to counter the confounding effects of the body size differences among the taxa. The size correction method employed by Collard and Wood [2000] entails dividing each value for a specimen by the geometric mean of all the values for that specimen [Mosimann, 1970; Jungers et al., 1995]. This is an isometric method, which equalizes the volumes of the specimens while maintaining their original shapes [Jungers et al., 1995]. Unfortunately, as Jolly [2001] has recently reminded his colleagues, the geometric mean method of size correction does not remove size-related shape differences among taxa. Thus, as Jolly [2001] went on to aver, it is possible that the failure of the analyses of Collard and Wood [2000] to return morphological cladograms that are congruent with the molecular phylogenies is a result of the presence of phylogenetically misleading size-related shape similarities in the two metric data sets.

In order to evaluate this possibility, we carried out an analysis in which phylogenies were derived from two character state data matrices that differed only in the manner in which the data were size corrected. One of the matrices was generated from data that had been size corrected with the geometric mean method, while the other was derived from data that had been size corrected with a version of the method that palaeoanthropologists frequently use to correct for allometric effects, regression analysis. There are a number of reasons to doubt the efficacy of regression-based size correction. For example, it is heavily dependent on the line-fitting technique and the data set employed to generate the regression equation [Aiello, 1992; Falsetti et al., 1993; Martin, 1993; Jungers et al., 1995]. More problematically, Jungers et al. [1995] have shown that allometric methods of size correction can fail to identify specimens of the same shape. Nevertheless, given that Jolly [2001] has cast doubt on the significance of the results of Collard and Wood [2000] on the grounds that they used an isometric method of size correction, we felt it was appropriate to compare the performance of the isometric method employed with the performance of an allometric method.

The data used in this analysis were obtained from the same sources as the hominoid metric data of Collard and Wood [2000], namely a morphometric database maintained by Bomard Wood of George Washington University and Chamberlain [1987]. From the former we obtained values for 76 measurements recorded on 41 *Pongo pygmaeus* (20 males, 21 females), 37 *Gorilla gorilla* (20 males, 17 females), 35 *Pan troglodytes* (13 males, 22 females), 75 *Homo sapiens* (40 males, 35 females) and 24 *Colobus guereza* (12 males and 12 females). From Chamberlain [1987] we acquired values for 52 additional measurements. These were recorded on 14 *P. pygmaeus* (5 males and 9 females), 14 *G. gorilla* (4 males, 10 females), 19 *P. troglodytes* (10 males, 9 females), 20 *H. sapiens* (10 males, 10 females) and 20 *C.*

[1] The character state data matrices discussed in this and the other sections have been submitted to the online phylogenetic database Treebase, which can be accessed at www.treebase.org.

Table 1. Measurements used in analyses

Code	Definition	Analysis
P1	I^1 labiolingual diameter	1, 2, 3
P2	I^1 mesiodistal diameter	1, 2, 3
P3	I^2 labiolingual diameter	1, 2, 3
P4	I^2 mesiodistal diameter	1, 2, 3
P5	C^1 mesiodistal diameter	1, 2, 3
P6	C^1 labiolingual diameter	1, 2, 3
P7	C^1 labial height	1, 2, 3
P8	P^3 buccolingual diameter	1, 2, 3
P9	P^3 mesiodistal diameter	1, 2, 3
P10	P^4 buccolingual diameter	1, 2, 3
P11	P^4 mesiodistal diameter	1, 2, 3
P12	M^1 buccolingual diameter	1, 2, 3
P13	M^1 mesiodistal diameter	1, 2, 3
P14	M^2 buccolingual diameter	1, 2, 3
P15	M^2 mesiodistal diameter	1, 2, 3
P16	M^3 buccolingual diameter	1, 2, 3
P17	M^3 mesiodistal diameter	1, 2, 3
P18	Outer alveolar breadth at M^3	1, 2, 3
P19	Inter-upper-canine breadth	1, 2, 3
P20	Palate length	1, 2, 3
P21	Inner alveolar breadth at M^3	1, 2, 3
P22	Palate depth at M^1	1, 2, 3
P23	Prosthion to plane of M^3	1, 2, 3
P24	Maxillo-alveolar breadth (M^2B–M^2B)	1, 3
P25	Breadth between upper second molars (M^2L–M^2L)	1, 3
P26	Palate depth at incisive fossa	1, 3
P27	Palate depth at upper second molars	1, 3
P28	Maxillary alveolar subtense	1, 3
P29	Upper incisor alveolar length	1, 3
P30	Upper premolar alveolar length	1, 3
P31	Upper molar alveolar length	1, 3
M1	I_1 labiolingual diameter	1, 2, 3
M2	I_1 mesiodistal diameter	1, 2, 3
M3	I_2 labiolingual diameter	1, 2, 3
M4	I_2 mesiodistal diameter	1, 2, 3
M5	C_1 labiolingual diameter	1, 2, 3
M6	C_1 mesiodistal diameter	1, 2, 3
M7	C_1 labial height	1, 2, 3
M8	P_3 buccolingual diameter	1, 2, 3
M10	P_4 buccolingual diameter	1, 2, 3
M11	P_4 mesiodistal diameter	1, 2, 3
M12	M_1 buccolingual diameter	1, 2, 3
M13	M_1 mesiodistal diameter	1, 2, 3
M14	M_2 buccolingual diameter	1, 2, 3
M15	M_2 mesiodistal diameter	1, 2, 3
M16	M_3 buccolingual diameter	1, 2, 3
M17	M_3 mesiodistal diameter	1, 2, 3
M18	Maximum cusp height	1, 2, 3
M19	Condylar height	1, 2, 3
M20	Bicondylar breadth	1, 2, 3
M21	Coronoid height	1, 2, 3
M22	Bicoronoid breadth	1, 2, 3
M23	Right condylar head width	1, 2, 3
M24	Right condylar head anterior-posterior breadth	1, 2, 3
M25	Ramal breadth	1, 2, 3
M26	Bigonial width	1, 2, 3
M27	Height of mandibular body at M_1	1, 2, 3
M28	Thickness of mandibular body of M_1	1, 2, 3
M29	Symphyseal height	1, 2, 3
M30	Symphyseal thickness	1, 2, 3
M31	Inner alveolar breadth at M_3	1, 2, 3
M32	Maximum mandibular length	1, 2, 3
M33	Inter-lower-canine distance	1, 2, 3
M34	Mandibular corpus height at M_3	1, 3
M35	Height of foramen spinosum	1, 3
M36	Height of mental foramen	1, 3
M37	Breadth between lower second molars	1, 3
M38	Lower incisor alveolar length	1, 3
M39	Lower premolar alveolar length	1, 3
M40	Lower molar alveolar length	1, 3
F1	Right orbital breadth	1, 3
F2	Right orbital height	1, 2, 3
F3	Interorbital breadth	1, 2, 3
F4	Bi-orbital breadth	1, 2, 3
F5	Nasion-rhinion	1, 2, 3
F6	Nasion-nasospinale	1, 2, 3
F7	Maximum nasal width	1, 2, 3
F8	Nasospinale-prosthion	1, 2, 3
F9	Bijugal breadth	1, 2, 3
F10	Bizygomatic breadth	1, 2, 3
F11	Upper facial breadth	1, 3
F12	Lower facial breadth	1, 3
F13	Breadth between infra-orbital foramina	1, 3
F14	Lower nasal bone breadth	1, 3
F15	Facial height	1, 3
F16	Height of infra-orbital foramen	1, 3
F17	Height of orbital margin	1, 3
F18	Upper malar height	1, 3
F19	Lower malar height	1, 3
F20	Upper facial prognathism	1, 3
F21	Lower facial prognathism	1, 3
F22	Malar prognathism	1, 3
F23	Nasofrontal subtense	1, 3
F24	Maxillary subtense	1, 3
C1	Glabella-opisthocranion	1, 2, 3
C2	Minimum postorbital breadth	1, 2, 3
C3	Basion-bregma	1, 2, 3
C4	Maximum biparietal breadth	1, 2, 3
C5	Biporionic width	1, 2, 3
C6	Mastoid length	1, 2, 3
C7	Coronale-coronale	1, 2, 3
C8	Opisthion-inion	1, 2, 3
C9	Bimastoid width	1, 2, 3
C10	Posterior skull length	1, 2, 3
C11	Breadth across tympanic plates	1, 3
C12	Breadth between carotid canals	1, 3
C13	Breadth between petrous apices	1, 3
C14	Breadth between foramen ovale	1, 3
C15	Breadth between infratemporal crests	1, 3
C16	Breadth of mandibular fossa	1, 3
C17	Length of tympanic plate	1, 3
C18	Length of petrous temporal	1, 3

Table 1 (continued)

Code	Definition	Analysis
C19	Position of foramen ovale	1, 3
C20	Position of infratemporal crest	1, 3
C21	Length of foramen magnum	1, 3
C22	Breadth of foramen magnum	1, 3
C23	Length of infratemporal fossa	1, 3
C24	Breadth of infratemporal fossa	1, 2, 3
C25	Opisthion-infratemporal subtense	1, 3
C26	Basio-occipital length	1, 3
C27	Parietal thickness at lambda	1, 3
C28	Frontal sagittal chord	1, 3
C29	Parietal sagittal chord	1, 3
C30	Parietal coronal chord	1, 3
C31	Occipital sagittal chord	1, 3
C32	Frontal sagittal arc	1, 3
C33	Occipital sagittal arc	1, 3
C34	Auricular height	1, 3

guereza (10 males, 10 females). The resulting 128-measurement data set was similar to the one employed by Collard and Wood [2000]. The differences are that we did not include specimens with missing data, and we excluded a measurement (P_3 mesiodistal diameter) for which no data were available for *Colobus*. The measurements are listed in table 1.

Two copies of the dataset were created. One was subjected to the isometric size correction method used by Collard and Wood [2000]. To reiterate, this technique entails dividing each value for a specimen by the geometric mean of all the values for that specimen. The other copy of the dataset was size corrected with an allometric size correction method. The procedure we used involves three steps. First, the geometric mean of all the characters for each specimen in the complete sample is computed. Next, a series of least squares regression analyses in which all the values for one of the characters were regressed against the geometric means are undertaken. Lastly, a new character state data matrix is assembled from the residuals, which are considered to be the size-corrected values. Because the degree of specimen overlap between Wood's data and Chamberlain's data is unknown at this point in time, the two parts of the dataset were size corrected separately.

Subsequently, both datasets were converted into discrete character states using divergence coding [Thorpe, 1984]. This technique proceeds by calculating the mean values for the taxa, and testing the differences among them for statistical significance. The means are then ranked in ascending order, and a taxon-by-taxon matrix compiled. Each cell in the top row of the matrix is filled with a taxon name such that the rank of the taxa decreases from left to right. The cells of the first column of the matrix are also filled with the names of the taxa on the basis of their rank, with the highest ranked taxon being placed in the top cell, and the lowest ranked taxon in the bottom cell. Thereafter, each cell in the matrix is assigned a score of –1, +1 or 0. A cell is scored as +1 if the mean of the taxon in the column is significantly greater than the mean of the taxon in the row. A cell is scored with a –1 if the mean of the taxon in the column is significantly lower than the mean of the taxon in the row. If the difference between the mean of the taxon in the column and the mean of the taxon in the row is not significant, the cell is filled with a 0. Once the matrix is completely filled, the total score of each column is calculated. Lastly, in order to obtain positive figures suitable for use in the phylogeny reconstruction computer programs that are currently available, an integer is added to each column's value. In converting the data set, Student's t-test (two-tailed) was used to test for statistical significance ($p \leq 0.05$), and four was added to each taxon total.

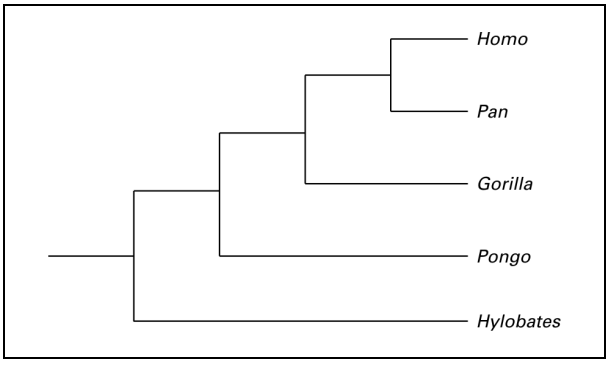


Fig. 1. Hominoid molecular cladogram.

Once the two character state data matrices had been generated, they were subjected to the two forms of analysis employed by Collard and Wood [2000], namely parsimony analysis and the phylogenetic bootstrap. Parsimony analysis identifies the cladogram requiring the smallest number of ad hoc hypotheses of homoplasy to account for the observed character state distribution. Each matrix was subjected to parsimony analysis using the branch-and-bound search routine of PAUP* 4.0 [Swofford, 1998], with the characters treated as linearly ordered and freely reversible variables. The phylogenetic bootstrap is a method of assessing confidence interval associated with a given clade. Using PAUP* 4.0 again, 10,000 replicates were generated from each data matrix by sampling characters with replacement and then subjected to parsimony analysis, and consensus trees were computed using a confidence region of 70% in line with Hillis and Bull [1993].

Lastly, the cladograms favoured in the parsimony analyses and the clades returned in the bootstrap analyses were compared with each other. They were also compared with the most parsimonious cladograms and bootstrap clades reported by Collard and Wood [2000], and with the consensus molecular phylogeny for the hominoids (fig. 1) [Ruvolo, 1997; Deinard et al., 1998; Deinard and Kidd 1999; Shi et al., 2003; Wildman et al., 2003].

The parsimony analysis of the data set corrected using the isometric method yielded a single most parsimonious cladogram whose branching pattern was the same as the one obtained by Collard and Wood [2000] (fig. 2a). As such, it was incongruent with the consensus molecular phylogeny for the extant hominoids. It suggested that *Homo* is the sister taxon of a *(Gorilla, Pongo, Pan)* clade and that *Pan* is the sister taxon of a *(Gorilla, Pongo)* clade. The most parsimonious cladogram was 72 steps shorter than a cladogram with the same topology as the molecular phylogeny (1,130 vs. 1,202 steps). The analysis of the data set corrected with the allometric method produced a most parsimonious cladogram that was different from the one obtained by Collard and Wood [2000] but was nevertheless still incongruent with the consensus molecular phylogeny. It suggested that *Gorilla* is the sister taxon of a *(Homo, Pan, Pongo)* clade and that *Homo* is the sister taxon of a *(Pan, Pongo)* clade (fig. 2b). The most parsimonious cladogram was 18 steps

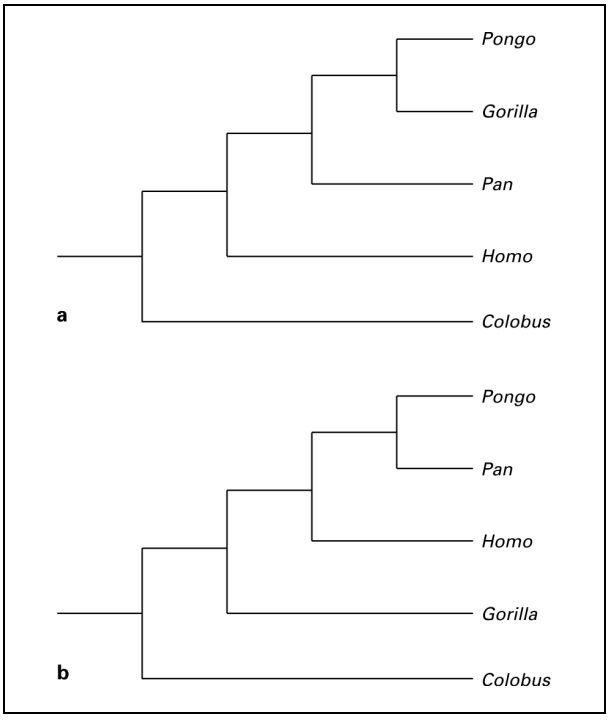


Fig. 2. Most parsimonious morphological cladograms obtained in the assessment of impact of different size correction methods. **a** Cladogram obtained using the isometric size correction method. **b** Cladogram obtained using the allometric size correction method.

shorter than a cladogram with the same topology as the molecular phylogeny (1,140 vs. 1,158 steps).

The bootstrap analysis of the isometrically corrected data set showed that a *(Gorilla, Pongo, Pan)* clade was supported by 87% of the replicates and a *(Gorilla, Pongo)* clade by 72% of the replicates. These clades are the same as the ones returned in Collard and Wood's [2000] bootstrap analysis of their hominoid metric data set. There were no clades supported by 70% or more of the bootstrap samples in the analysis of the data set corrected using the allometric method.

Based on these results, it seems unlikely that the findings of Collard and Wood [2000] would have been different if they had employed an allometric size correction method rather than the isometric one they chose to use. Allometric and isometric techniques give different results, as anticipated by Jolly [2001]. However, the difference is relatively modest and, more significantly, there is no evidence that allometric methods result in more accurate estimates of phylogeny than their isometric counterparts. It is worth noting in passing that the results of this part of our study are consistent with the results of analyses reported by Creel [1986] and Singleton [1996]. These authors also found that choice of size correction technique does not greatly influence cladogram topology.

Analysis 2: Outgroup Composition

In the second part of the study, we sought to determine whether the results of Collard and Wood [2000] would have been different if they had employed different outgroups in their analyses. As noted above, Collard and Wood [2000] reported the results of three analyses, two of which focused on the extant hominoids and one of which focused on the extant papionins. Collard and Wood [2000] employed *Colobus* as the outgroup in the extant hominoid analyses. In the extant papionin analysis, they used *Cercopithecus*. Collard and Wood's [2000] use of these taxa was justifiable since they are closely related to their respective ingroups. However, recent studies have demonstrated that the results of cladistic analyses can be sensitive to the choice of outgroup [e.g. Masters and Brothers, 2002]. Thus, it seemed reasonable to examine whether changing outgroup taxa would have affected the results of Collard and Wood [2000].

The data set comprised values for 76 measurements recorded on males and females of *G. gorilla* (20 males, 17 females), *Homo* (40 males, 35 females), *Pan* (13 males, 22 females), *Pongo* (20 males, 21 females) and two outgroup taxa, *Colobus* (12 males and 12 females) and *Papio* (15 males, 17 females). *Papio* was selected as a second outgroup because its craniofacial morphology is markedly different from that of *Colobus*. The data for all the taxa were taken from the morphometric database discussed in the previous section. The measurements are once again listed in table 1.

To counter the confounding effects of the body size differences among the taxa, we subjected the data to the isometric size correction method used by Collard and Wood [2000]. We then used divergence coding to assign character states, according to the procedure described in the preceding section, except that in this case the integer added to each taxon total at the end was 5.

Subsequently, a four-part procedure was carried out. The first part of the procedure mirrored the analyses carried out by Collard and Wood [2000]. *Papio* was removed from the character state data matrix, and the remaining taxa – *Colobus*, *Gorilla*, *Homo*, *Pan* and *Pongo* – were subjected to maximum parsimony analysis and bootstrapping in PAUP* 4.0. Next, the analyses were repeated with *Papio* as the outgroup rather than *Colobus*. Thereafter, the analyses were re-run with both *Colobus* and *Papio* included as outgroups. Lastly, the phylogenetic hypotheses obtained in the three preceding parts of the analysis were compared with each other, with the results of Collard and Wood [2000] and with the consensus molecular phylogeny for the extant hominoids.

The analysis of the data set including *Colobus* as an outgroup produced a single most parsimonious cladogram that had the same ingroup topology as the cladogram returned by the hominoid metric data of Collard and Wood [2000] (fig. 3a). Accordingly, it was incongruent with the consensus molecular phylogeny for the extant hominoids. It suggested that the first branching event in the evolution of the ingroup separated the *Homo* lineage from the common ancestor of a clade comprising *Gorilla*, *Pongo* and *Pan*, while the second branching event separated the *Pan* lineage from the common ancestor of a clade consisting of *Gorilla* and *Pongo*. The most parsimonious cladogram was 44 steps shorter than a cladogram with the same topology as the molecular phylogeny (779 vs. 823 steps). The analysis that included *Papio* as the single outgroup also produced a single most parsimonious cladogram

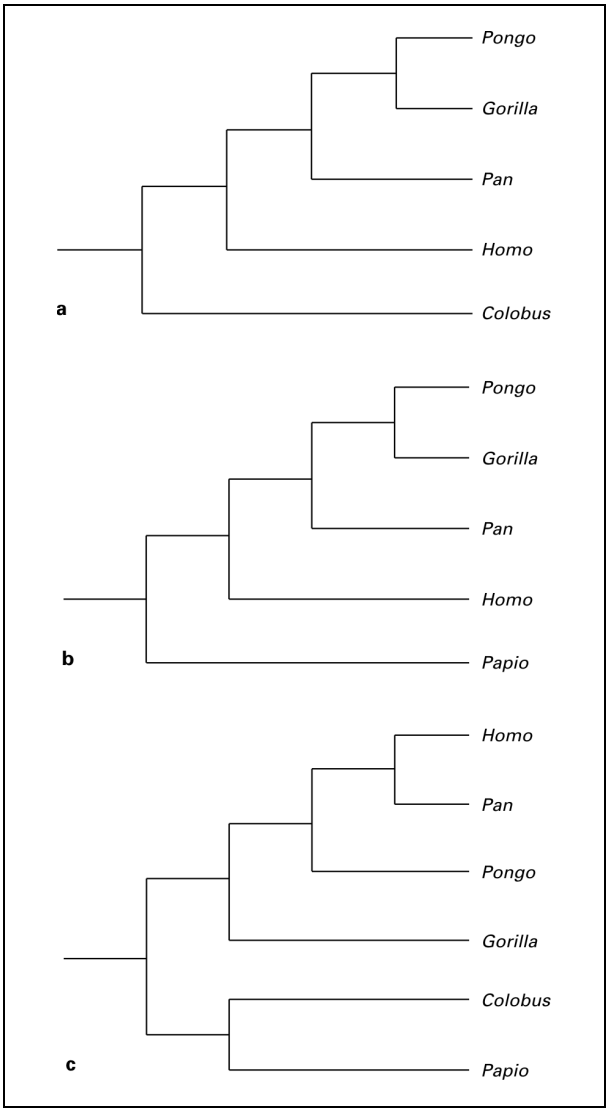


Fig. 3. Most parsimonious morphological cladograms obtained in the assessment of impact of different outgroups. **a** Cladogram obtained using *Colobus* as outgroup. **b** Cladogram obtained using *Papio* as outgroup. **c** Cladogram obtained using *Colobus* and *Papio* as outgroups.

that had the same ingroup topology as the one returned by the hominoid metric data of Collard and Wood [2000] (fig. 3b). The most parsimonious cladogram was 42 steps shorter than a cladogram with the same topology as the molecular phylogeny (811 vs. 853 steps). The analysis that included both *Colobus* and *Papio* as outgroups yielded a single most parsimonious cladogram that was different from the one ob-

tained by Collard and Wood [2000] but also different from the consensus molecular phylogeny for the extant hominoids. It suggested that *Gorilla* is the sister taxon of a *(Homo, Pan, Pongo)* clade and that *Pongo* is the sister taxon of a *(Homo, Pan)* clade (fig. 3c). The most parsimonious cladogram was 27 steps shorter than a cladogram with the same topology as the molecular phylogeny (1,007 vs. 1,034 steps).

The bootstrap analyses of the data set including *Colobus* as the outgroup showed that a *(Gorilla, Pan, Pongo)* clade was supported by 70% of the bootstrap samples. For the data set that included *Papio* as the outgroup, a *(Gorilla, Pongo, Pan)* clade was supported by 74% of the bootstrap samples. No clades were supported by 70% or more of the replicates in the bootstrap analysis in which both *Colobus* and *Papio* were included as outgroups.

The fact that the analysis in which *Papio* was used as the outgroup produced the same phylogenetic relationships as the analysis in which *Colobus* was used as the outgroup indicates that the two taxa polarise the character state transformation series in similar ways. Given that *Colobus* and *Papio* differ considerably in craniodental morphology, this suggests that outgroup choice did not greatly influence the results of Collard and Wood [2000]. Based on these results, therefore, there is no reason to think that the findings of Collard and Wood [2000] would have been substantively different if they had employed different outgroups. Apparently, employing more than one outgroup would have reduced the statistical support for erroneous clades, but it seems unlikely that it would have resulted in the recovery of craniodental cladograms with the same topologies as the molecular phylogenies.

Analysis 3: Non-Phylogenetic Correlation among Characters

Several authors have criticised the study of Collard and Wood [2000] on the grounds that many of the characters they utilised are likely to be correlated either developmentally or functionally, and that any such correlation among characters violates one of the main tenets of cladistics, which is that characters should be independent [McCollum and Sharpe, 2001; Rae, 2002; Miller, 2003]. With this criticism in mind, in the third section of the study we carried out an analysis that compared phylogenies derived from matrices compiled in the manner outlined by Collard and Wood [2000] with phylogenies derived from matrices from which correlated characters had been removed.

The data set employed in this part of the study was the same as that used to investigate the impact of different size correction methods. Four copies of the data set were created. Two of these were size corrected with the isometric method described earlier, and two were size corrected with the allometric method outlined in the same section. Subsequently, one of the isometrically corrected data sets and one of the allometrically corrected data sets were screened for non-phylogenetically correlated characters. Correcting for character correlation in the context of phylogenetic analysis is not a straightforward exercise. This is because development and function are not the only reasons why characters will correlate. Phylogenesis is also expected to be a cause of character correlation. Indeed, cladistics relies on the presence of correlation among characters – correlation due to shared phylogenetic history. Thus, when attempting to remove functionally or developmentally correlated characters, care must be taken not to also remove phylogenetically correlated characters. In

order to identify non-phylogenetic correlations among the characters in our data sets, we used the Excel correlation tool to examine the relationships among all the characters for each taxon. A pair of characters was deemed to be correlated for reasons other than phylogeny if they were found to be significantly correlated ($p \leq 0.05$) in all 5 taxa included in the study. We used this criterion on the grounds that a lack of correlation among a pair of characters in one or more taxa indicates that the characters are free to evolve separately and therefore do not violate the character independence requirement of cladistics. Bonferroni correction was not employed in order to increase the probability of identifying non-phylogenetically correlated characters. Our analysis indicated that 29 characters needed to be removed from the allometric data set (P2, P5, P6, P7, P8, P12, P13, P14, P15, P16, P19, P21, P25, M1, M3, M10, M13, M14, M21, M22, M29, M32, F1, F5, F9, F17, C17, C32, C33) and 30 from the isometric data set (P3, P5, P7, P8, P10, P12, P14, P15, P16, P19, P20, P21, P25, M1, M6, M13, M14, M21, M22, M29, F1, F6, F10, F17, C1, C2, C4, C22, C32, C33). Twenty of the deleted characters were the same in both data sets. After deleting the non-phylogenetically correlated characters, we used divergence coding to assign character states to all four data sets. Thereafter, the four matrices were subjected to parsimony analysis and the phylogenetic bootstrap in the manner described in the preceding two sections.

The isometrically corrected data set that had not been screened for correlated characters yielded a single most parsimonious cladogram whose branching pattern was the same as the one obtained by Collard and Wood [2000] (fig. 4a). As such, it was incongruent with the consensus molecular phylogeny for the extant hominoids. It suggested that *Homo* is the sister taxon of a *(Gorilla, Pongo, Pan)* clade and that *Pan* is the sister taxon of a *(Gorilla, Pongo)* clade. The most parsimonious cladogram was 72 steps shorter than a cladogram with the same topology as the molecular phylogeny (1,130 vs. 1,202 steps). The bootstrap analysis indicated that the *(Gorilla*, *Pongo*, *Pan)* clade was supported by 87% of the replicates and that 72% of the replicates supported the *(Gorilla*, *Pongo)* clade. These clades are the same as the ones returned in the bootstrap analysis of Collard and Wood [2000] of their metric hominoid data set.

The allometrically corrected data set that had not been screened for correlated characters produced a most parsimonious cladogram that was different from the one obtained by Collard and Wood [2000] but was nonetheless still incongruent with the consensus molecular phylogeny. It suggested that *Gorilla* is the sister taxon of a *(Homo, Pan, Pongo)* clade and that *Homo* is the sister taxon of a *(Pan*, *Pongo)* clade (fig. 4b). The most parsimonious cladogram was 18 steps shorter than a cladogram with the same topology as the molecular phylogeny (1,140 vs. 1,158 steps). None of the clades was supported by 70% or more of the replicates in the bootstrap analysis.

The analysis of the isometrically corrected, correlation-adjusted data set produced a single most parsimonious cladogram with the same ingroup topology as the cladogram returned by the hominoid metric data set of Collard and Wood [2000] (fig. 4c). Accordingly, it was not congruent with the consensus molecular phylogeny for the extant hominoids. It suggested that the first branching event in hominoid evolution separated the *Homo* lineage from the common ancestor of a clade comprising *Gorilla*, *Pongo* and *Pan*, while the second branching event separated the *Pan* lineage from the common ancestor of a clade consisting of *Gorilla* and

Pongo. The most parsimonious cladogram was 35 steps shorter than a cladogram with the same topology as the molecular phylogeny (894 vs. 929 steps). The bootstrap analysis for the isometrically corrected, correlation-adjusted data set showed that there were no clades supported by 70% or more of the replicates.

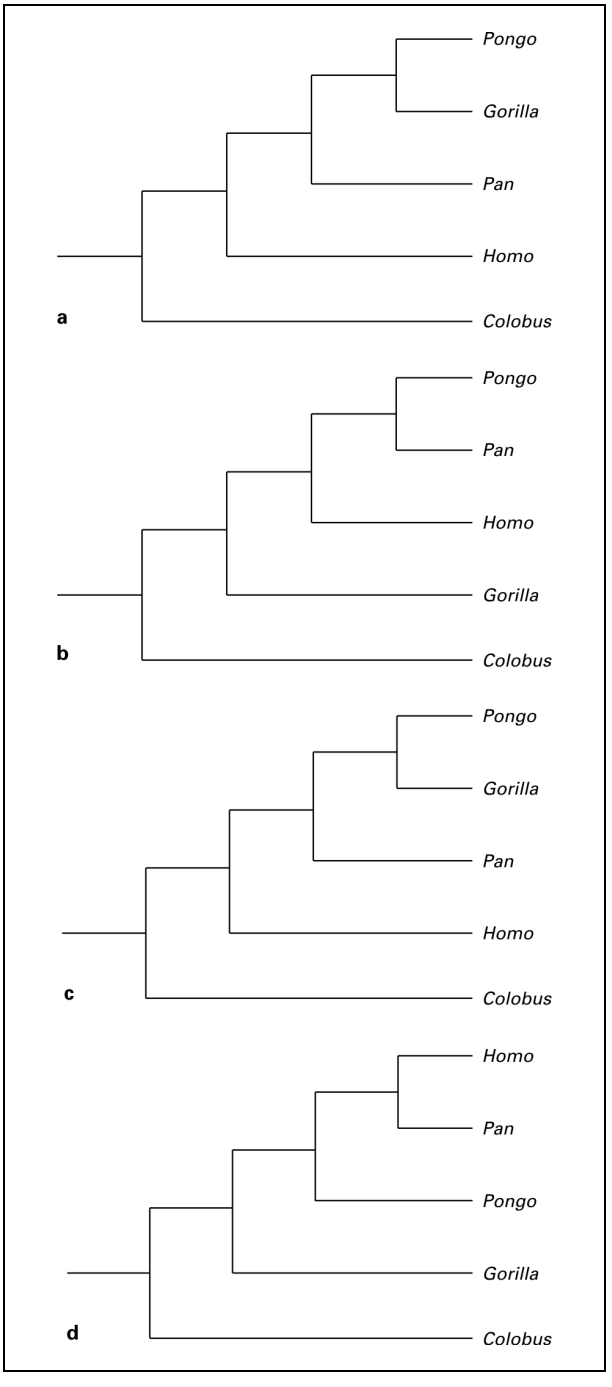
The analysis of the correlation-adjusted data set that was size corrected using the allometric method produced a single most parsimonious cladogram that was not compatible with Collard and Wood [2000] but was not compatible with the consensus molecular phylogeny for the extant hominoids either. It suggested that *Gorilla* is the sister taxon of a *(Homo, Pan, Pongo)* clade and that *Pongo* is the sister taxon of a *(Homo, Pan)* clade (fig. 4d). The most parsimonious cladogram was 17 steps shorter than a cladogram with the same topology as the molecular phylogeny (891 vs. 908 steps). The bootstrap analysis of the allometrically corrected, correlation-adjusted data set yielded a clade that did not agree with the molecular phylogeny. The clade consisted of *Homo*, *Pan* and *Pongo*, and was supported by 70% of the replicates.

Given that neither of the correlation-adjusted data sets yielded phylogenetic hypotheses that were compatible with the hominoid molecular phylogeny, it appears that, contrary to what some have argued, the findings of Collard and Wood [2000] would not have been substantively different if they had removed functionally or developmentally correlated characters from their data sets. Removing correlated characters from the isometrically corrected data set reduces support for false clades, but it does not result in the recovery of what is widely accepted to be the correct phylogeny for the extant hominoids. Removing correlated characters from the allometrically corrected data set does not reduce support for false clades either. On the contrary, false clades received more support in the analysis of the allometrically corrected correlation-adjusted data set than in the analysis of the allometrically corrected data set that had not been screened for correlated characters. These results suggest that the impact of correlated characters on phylogenetic analyses of primate morphological data is much less significant than has been claimed by McCollum and Sharpe [2001], Rae [2002] and Miller [2003].

Conclusions

The study reported here suggests that the results of Collard and Wood [2000] were not influenced by the size correction method they elected to use or by the outgroup they employed. Our study also suggests that Collard and Wood's [2000] results were not affected by their decision to ignore developmental, functional and other non-phylogenetic correlations among the characters in their data sets. Accordingly, our study indicates that Collard and Wood's [2000] main conclusion, which was that published phylogenies for the early hominids may be unreliable, cannot be dismissed on methodological grounds. It also supports the broader suggestion that our grasp of fossil primate phylogeny may be much more tenuous than is generally

Fig. 4. Most parsimonious morphological cladograms obtained in the assessment of impact of non-phylogenetic character correlation. **a** Cladogram obtained using the isometric size correction method with no adjustment for character correlation. **b** Cladogram obtained using the allometric size correction method with no adjustment for character correlation.



c Cladogram obtained using the isometric size correction method and removing character correlation. **d** Cladogram obtained using the allometric size correction method and removing character correlation.

appreciated [Cartmill 1994a, b; Hartman, 1988; Lieberman, 1995, 1997, 1999, 2000; Pilbeam, 1996; Collard and Wood, 2001].

How can the reliability of fossil primate phylogenetic hypotheses be improved? One strategy is to devise techniques for characterising primate craniodental morphology that are more sensitive to any phylogenetic signal than the methods that are currently in use. A recent study suggests that three-dimensional morphometrics may be one such technique [Lockwood et al., 2004]. Since exogenetic stimuli can be expected to confound phylogenetic reconstruction [Lieberman, 1995, 1997, 1999, 2000], another effective approach may be to focus on characters that are known to be minimally affected by such stimuli, for example, dental enamel and the structures of the middle and inner ear. A third strategy is to develop rigorous comparative methods for discriminating between phylogenetically informative and phylogenetically misleading craniodental similarities. For example, pursuit of detailed information about the ontogeny of characters may help identify convergences, parallelisms and reversals [Wood, 1988; Lieberman et al., 1996; Collard and O'Higgins, 2002]. Lastly, we suggest that more attention should be paid to non-morphological lines of evidence that may have a bearing on the phylogenetic relationships of fossil primates, such as biogeography, stratigraphy and behavioural indicators [Turner and Wood, 1993; Agustí et al., 1996].

Acknowledgments

We are grateful to Bernard Wood for allowing us to use his data and for helpful comments on the manuscript. We also benefited from the advice of Nina Jablonski, Stephen Lycett, Daniele Serdoz and three anonymous reviewers. The analyses were carried out while M.C. was a faculty member and M.N.-R. was a visiting researcher in M.C.'s lab in the Department of Anthropology, Washington State University. M.N.-R.'s visit was supported by a grant from the Conselleria d'Innovació i Energia, Govern Balear (FPI01-43098605).

References

Aiello LC (1992). Allometry and the analysis of size and shape in human evolution. *Journal of Human Evolution* 22: 127–147.

Agustí J, Kohler M, Moyà-Solà S, Cabrera L, Garcés M, Parés JM (1996). Can Llobateres: The pattern and timing of the Vallesian hominoid radiation reconsidered. *Journal of Human Evolution* 31: 143–155.

Cartmill M (1994a). A critique of homology as a morphological concept. *American Journal of Physical Anthropology* 94: 115–123.

Cartmill M (1994b). Anatomy, antinomies, and the problem of anthropoid origins. In *Anthropoid Origins* (Fleagle JG, Kay RF, eds.), pp 549–566. New York, Plenum Press.

Chamberlain AT (1987). *A Taxonomic Review and Phylogenetic Analysis of* Homo habilis. PhD thesis, University of Liverpool.

Collard M, O'Higgins P (2002). Ontogeny and homoplasy in the papionin monkey face. *Evolution and Development* 3: 322–331.

Collard M, Wood BA (2000). How reliable are human phylogenetic hypotheses? *Proceedings of the National Academy of Sciences of the USA* 97: 5003–5006.

Collard M, Wood BA (2001). Reliability of craniodental evidence in fossil catarrhine phylogenetics. In *Phylogeny of Neogene European Hominoid Primates* (de Bonis L, Koufous G, Andrews P, eds.), pp 118–150. Cambridge, Cambridge University Press.

Creel N (1986). Size and phylogeny in hominoid primates. *Systematic Zoology* 35: 81–99.

Deinard A, Kidd K (1999). Evolution of a HOXB6 intergenic region within the great apes and humans. *Journal of Human Evolution* 36: 687–703.

Deinard A, Sirugo G, Kidd K (1998). Hominoid phylogeny: Inferences from a subterminal minisatellite analyzed by repeat expansion detection (RED). *Journal of Human Evolution* 35: 313–317.

Falsetti AB, Jungers WL, Cole TM III (1993). Morphometrics of the callitrichid forelimb: A case study in size and shape. *International Journal of Primatology* 14: 551–572.
Hartman SE (1988). A cladistic analysis of hominoid molars. *Journal of Human Evolution* 17: 489–502.
Hillis DM, Bull JJ (1993). An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Systematic Biology* 42: 182–192.
Jolly CJ (2001). A proper study for mankind: Analogies from the papionin monkeys and their implications for human evolution. *Yearbook of Physical Anthropology* 44: 177–204.
Jungers WL, Falsetti AB, Wall CE (1995). Shape, relative size, and size-adjustments in morphometrics. *Yearbook of Physical Anthropology* 38: 137–161.
Lieberman DE (1995). Testing hypotheses about recent human evolution from skulls: Integrating morphology, function, development, and phylogeny. *Current Anthropology* 36: 159–197.
Lieberman DE (1997). Making behavioral and phylogenetic inferences from hominid fossils: Considering the developmental influence of mechanical forces. *Annual Review of Anthropology* 26: 185–210.
Lieberman DE (1999). Homology and hominid phylogeny: Problems and potential solutions. *Evolutionary Anthropology* 7: 142–151.
Lieberman DE (2000). Ontogeny, homology, and phylogeny in the hominid craniofacial skeleton: The problem of the browridge. In *Development, Growth and Evolution: Implications for the Study of the Hominid Skeleton* (O'Higgins P, Cohn MJ, eds.), pp 85–122. London, Academic Press.
Lieberman DE, Wood BA, Pilbeam DR (1996). Homoplasy and early *Homo*: An analysis of the evolutionary relationships of *H. habilis* sensu stricto and *H. rudolfensis*. *Journal of Human Evolution* 30: 97–120.
Lockwood CA, Kimbel WH, Lynch JM (2004). Morphometrics and hominoid phylogeny: Support for a chimpanzee-human clade and differentiation among great ape subspecies. *Proceedings of the National Academy of Sciences of the USA* 101: 4356–4360.
Martin RD (1993). Allometric aspects of skull morphology in *Theropithecus*. In Theropithecus*: The Rise and Fall of a Primate Genus* (Jablonski NG, ed.), pp 273–298. Cambridge, Cambridge University Press.
Masters JC, Brothers DJ (2002). Lack of congruence between morphological and molecular data in reconstructing the phylogeny of the Galagonidae. *American Journal of Physical Anthropology* 117: 79–93.
McCollum MA, Sharpe PT (2001). Developmental genetics and early hominid craniodental evolution. *Bioessays* 23: 481–493.
Miller E (2003). Review of De Bonis L, Koufous G, Andrews P (eds.) (2001) *Hominid Evolution and Climate Change in Europe: Phylogeny of the Neogene Hominid Primates of Eurasia,* volume 2. *American Journal of Physical Anthropology* 121: 390–391.
Mosimann JE (1970). Size allometry: Size and shape variables with characteristics of the log normal and generalized gamma distributions. *Journal of the American Statistical Association* 65: 930–945.
Pilbeam DR (1996). Genetic and morphological records of the Hominoidea and hominid origins. *Molecular Phylogenetics and Evolution* 5: 155–168.
Rae T (2002). Review of De Bonis L, Koufous G, Andrews P (eds.) (2001) *Hominid Evolution and Climate Change in Europe: Phylogeny of the Neogene Hominid Primates of Eurasia,* volume 2. *Journal of the Royal Anthropological Institute* 8: 590–591.
Ruvolo M (1997). Molecular phylogeny of the hominoids: Inferences from multiple independent DNA data sets. *Molecular Biology and Evolution* 14: 248–265.
Shi J, Xi H, Wang Y, Zhang C, Jiang Z, Zhang K, Shen Y, Jin L, Zhang K, Yuan W, Wang Y, Lin J, Hua Q, Wang F, Xu S, Ren S, Xu S, Zhao G, Chen Z, Jin L, Huang W (2003). Divergence of the genes on human chromosome 21 between human and other hominoids and variation of substitution rates among transcription units. *Proceedings of the National Academy of Sciences of the USA* 100: 8331–8336.
Singleton M (1996). Quantitative character coding in hominoid phylogeny reconstruction. Poster presented at the 66th annual meeting of the American Association of Physical Anthropology, 2–5 April 1996, St. Louis, Mo.
Swofford DL (1998). *PAUP*: Phylogenetic Analysis Using Parsimony (* and other methods) Version 4.0*. Sunderland, Sinauer Associates.
Thorpe RS (1984). Coding morphometric characters for constructing distance Wagner networks. *Evolution* 38: 244–255.
Turner A, Wood BA (1993). Comparative palaeontological context for the evolution of the early hominid masticatory system. *Journal of Human Evolution* 24: 301–318.
Wildman DE, Uddin M, Liu G, Grossman LI, Goodman M (2003). Implications of natural selection in shaping 99.4% nonsynonymous DNA identity between humans and chimpanzees: Enlarging genus *Homo*. *Proceedings of the National Academy of Sciences of the USA* 100: 7181–7188.
Wood BA (1988). Are the 'robust' australopithecines a monophyletic group? In *Evolutionary History of the 'Robust' Australopithecines* (Grine FE, ed.), pp 269–284. New York, Aldine de Gruyter.