

# How reliable are human phylogenetic hypotheses?

Mark Collard\*<sup>†</sup> and Bernard Wood\*

\*Department of Anthropology, University College London, London WC1E 6BT, United Kingdom; and †Department of Anthropology, George Washington University, Washington, DC 20052

Communicated by David Pilbeam, Harvard University, Cambridge, MA, January 28, 2000 (received for review May 27, 1999)

**Cladistic analysis of cranial and dental evidence has been widely used to generate phylogenetic hypotheses about humans and their fossil relatives. However, the reliability of these hypotheses has never been subjected to external validation. To rectify this, we applied identical methods to equivalent evidence from two groups of extant higher primates for whom reliable molecular phylogenies are available, the hominoids and papionins. We found that the phylogenetic hypotheses based on the craniodental data were incompatible with the molecular phylogenies for the groups. Given the robustness of the molecular phylogenies, these results indicate that little confidence can be placed in phylogenies generated solely from higher primate craniodental evidence. The corollary of this is that existing phylogenetic hypotheses about human evolution are unlikely to be reliable. Accordingly, new approaches are required to address the problem of hominin phylogeny.**

The upsurge in paleoanthropological field research over the past quarter century has resulted in the recognition of many new hominin species, including *Australopithecus afarensis* (1), *Paranthropus aethiopicus* (2), *Ardipithecus ramidus* (3, 4), *Australopithecus anamensis* (5), *Australopithecus bahrelghazali* (6), *Homo antecessor* (7), and *Australopithecus garhi* (8). This has led to commensurate interest in the generation of reliable hypotheses about human phylogeny (8–14). Without a reliable phylogeny, little confidence can be placed in hypotheses of ancestry, or in scenarios linking events in human evolution with environmental and ecological influences. However, the phylogenetic relationships of the dozen, or so, species whose remains comprise the hominin fossil record are far from certain. Despite, in paleontological terms, a relative abundance of fossil evidence, cladistic analyses of the hominins have so far yielded conflicting and weakly supported hypotheses of relationships (9–14, 15, 16). Conventionally, this state of affairs has been attributed to poor character choice, taxonomic disagreements, or flaws in the available analytical methods (10, 12, 14). A fourth possibility, namely, that the type of qualitative and quantitative craniodental characters normally used to reconstruct the phylogenetic relationships of hominin species and genera are not reliable for this purpose, has only recently been entertained (8, 13, 17–21).

To assess the likely reliability of standard hominin cranial and dental characters for interspecific and intergeneric phylogenetic reconstruction, we used hominin cladistic methods to analyze comparable characters from two groups of extant primates: the hominoids, the higher primate group most closely related to the fossil hominins, and the papionins, the Old World monkey tribe comprising the baboons, mangabeys, and macaques. We then judged the resulting cladograms against the groups' consensus molecular phylogenies (22, 23). This approach, which is similar to Hartman's (24), assumes that congruence between the morphological and molecular phylogenies indicates that equivalent hominin fossil evidence yields reliable phylogenies, whereas incongruence indicates the converse. The assumption that any conflict between the molecular and morphological cladograms results from limitations of the morphological evidence can be justified on several counts. First, phylogenetic relationships are genetic relationships. Thus, in phylogenetics, morphology can never be more than a proxy for molecular data. Second, because osseous and other morphological characters can be highly influenced by external stimuli, such as the forces generated by

habitual activities (13, 19–21), they can be expected to provide misleading information about phylogeny more frequently than genetical characters. Third, both molecular cladograms used in this study are supported by multiple lines of independent biomolecular and karyological evidence (22, 23). Congruence among multiple lines of independent evidence is the strongest support possible for a phylogenetic hypothesis. Lastly, molecular phylogenetic methods have been successfully tested on taxa of known phylogeny whereas comparable tests of morphological phylogenetic methods have not been successful (25–27).

In the present study, we concentrated on testing the phylogenetic utility of quantitative rather than qualitative craniodental data. Although more hominin cladistic analyses have been based on qualitative characters than on quantitative characters, there is in fact no intrinsic difference between qualitative and quantitative characters as far as the cladistic methodology is concerned (28). Moreover, qualitatively scored characters are necessarily more subjective than measurements, and the few studies that have examined the reproducibility of qualitative assessments of morphology have shown that many of them are impossible to replicate (14, 29). Furthermore, unlike discrete character states, morphometric data can be adjusted to account for any intertaxon body size differences, thereby reducing a further potential source of confusion. Nonetheless, to relate our results to all of the published cladistic analyses of the hominins, we also included an assessment of the phylogenetic utility of qualitatively derived character states.

## Materials and Methods

Standard paleoanthropological measurements of the cranium, mandible, and dentition were used to compile quantitative data sets for the extant hominoids and papionins. The hominoid data set comprised values for 129 measurements recorded on males and females of *Gorilla*, *Homo*, *Pan*, *Pongo*, and an Old World monkey, *Colobus*, which was included as an outgroup. Seventy-seven of the one hundred and twenty-nine measurements were recorded on 37 *Gorilla gorilla*, 75 *Homo sapiens*, 35 *Pan troglodytes*, 41 *Pongo pygmaeus*, and 24 *Colobus guereza* (30). The other 52 measurements were taken on 20 *G. gorilla*, 20 *H. sapiens*, 20 *P. troglodytes*, 20 *P. pygmaeus* and 20 *C. guereza* (31).

The papionin data set consisted of values for 62 measurements recorded on mixed sex samples of *Macaca*, *Cercocebus*, *Lophocebus*, *Mandrillus*, *Papio*, *Theropithecus*, and an outgroup, *Pan*. All 62 measurements were recorded on 26 *Cercocebus galeritus/torquatus*, 40 *Lophocebus albigena/atterimus*, 40 *Macaca fascicularis/mulatta*, 62 *Mandrillus leucophaeus/sphinx*, 39 *Papio hamadryas*, 44 *Theropithecus gelada*, and 17 *P. troglodytes* (32). Fifty-five of the measurements were taken on a further 14 *C. torquatus* and 12 *P. troglodytes*. These data were kindly provided by A. T. Chamberlain (University of Sheffield).

<sup>†</sup>To whom reprint requests should be addressed. E-mail: m.collard@ucl.ac.uk.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

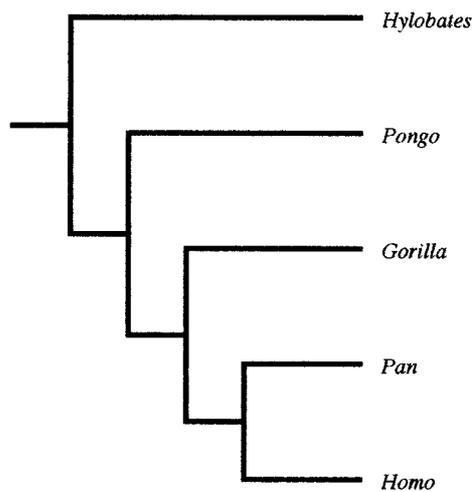


Fig. 1. Hominoid molecular relationships.

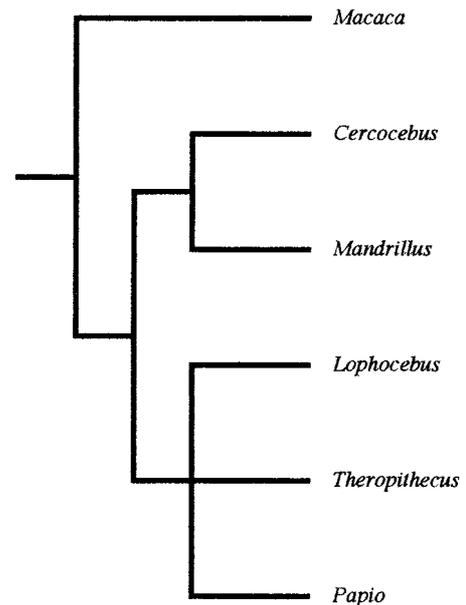


Fig. 2. Papionin molecular relationships.

A character state data matrix was derived from each data set. The confounding effects of the body size differences between the taxa were minimized by dividing the values for each specimen by their geometric mean (33). The size-adjusted data were converted into discrete character states using divergence coding (34). Alternative size-adjustment methods and coding procedures either give similar results to the techniques used here (35) or are inappropriate for analyses of fossil taxa (36).

A qualitative data matrix was also generated for the hominoids. This consisted of the states of 96 craniodental characters recorded on specimens of *Gorilla*, *Homo*, *Hylobates*, *Pan*, *Pongo*, and an outgroup. Sixty-two characters were taken from Shoshani *et al.* (37), two from Braga (38), six from Andrews (39), four from Schwartz (40), and two from Delson and Andrews (41). The remainder were the cranial and dental characters from Groves (42) that were excluded, without explanation, from the data set of Shoshani *et al.* (37). (This dataset is published as supplemental data on the PNAS web site, [www.pnas.org](http://www.pnas.org).)

The matrices were used to perform two tests of the hypothesis that craniodental characters are reliable for reconstructing the phylogenetic relationships of fossil primate species and genera. The first test was based on parsimony analysis, which identifies the cladogram/s requiring the smallest number of ad hoc hypotheses of character state change to account for the distribution of character states among the taxa. Each matrix was subjected to parsimony analysis by using PAUP 3.0s (43), and the shortest cladogram compared with the appropriate consensus molecular cladogram (Figs. 1 and 2). We reasoned that, for the hypothesis to be supported, the favored cladogram should not contain any “false” (i.e., nonmolecular) clades, since it has been standard practice in hominin cladistic analyses to assume that all clades of a cladogram are equally reliable (9–14). Thus, the hypothesis would have been supported if an analysis favored a fully resolved cladogram that matched the molecular cladogram, or a partially resolved cladogram that comprised only molecular clades. Alternatively, the hypothesis would have been supported had the analysis yielded multiple equally parsimonious cladograms whose strict consensus comprised only clades that were compatible with the molecular cladogram.

The second test of the hypothesis was based on the phylogenetic bootstrap, which is a method of assessing the confidence interval associated with a given clade (44, 45). Using PAUP 3.0s, 10,000 matrices were derived from each matrix by sampling with replacement. The new matrices were subjected to parsimony analysis, and a consensus cladogram of the most parsimonious

cladograms was computed using a confidence region of 70% (46). Thereafter, the clades of the consensus cladogram were compared with the appropriate molecular cladogram. In this test, it was reasoned that, for the hypothesis to be supported, the best supported clades should not be false clades, since it is commonly assumed in hominin analyses that the better the bootstrap support for a clade, the more likely the clade is to be “true” (15, 16).

In both the parsimony and bootstrap analyses, the quantitative characters were treated as linearly ordered and freely reversing variables (47). Twenty-seven of the qualitative characters were also treated as linearly ordered, freely reversing variables, because their states clearly formed linear transformation series. The other 69 qualitative characters were treated as unordered variables. Cladograms were obtained using PAUP’s branch-and-bound search routine.

## Results

The hypothesis was not supported by the parsimony analyses of the two quantitative data sets. The hominoid analysis yielded a most parsimonious cladogram whose branching pattern differed from that of the hominoid consensus molecular cladogram. It suggested that *Homo* is the sister taxon of a (*Gorilla*, *Pan*, *Pongo*) clade, and that *Pan* is the sister taxon of a (*Gorilla*, *Pongo*) clade. The papionin analysis also favored a single most parsimonious cladogram that was incompatible with the group’s molecular cladogram. It suggested that the first branching event in papionin evolution separated the *Lophocebus* lineage from the common ancestor of *Cercocebus*, *Macaca*, and the baboons, while the second branching event separated the *Cercocebus* lineage from the common ancestor of *Macaca* and the baboons. The third branching event, according to this cladogram, separated the macaque lineage from the common ancestor of the baboons, while the final phylogenetically informative branching event separated the *Theropithecus* lineage from the common ancestor of *Papio* and *Mandrillus*. The hypothesis was also not supported by the parsimony analysis of the hominoid qualitative matrix. The cladogram agreed with the hominoid molecular cladogram in locating *Hylobates* as the basal hominoid. However, it differed from the molecular cladogram in grouping *Homo* and *Pongo* in one clade and *Pan* and *Gorilla* in a second, the latter being

supported by four more putative synapomorphies than the *Pan-Homo* clade of the molecular cladogram.

The bootstrap analyses similarly failed to uphold the hypothesis. None of the clades supported by 70% or more of the bootstrap samples were compatible with the consensus molecular cladograms. The hominoid quantitative analysis supported a (*Gorilla, Pan, Pongo*) clade at 95%, and a (*Gorilla, Pongo*) clade at 73%. The papionin quantitative analysis supported a (*Cercocebus, Macaca*, baboon) clade at 98%, a (*Macaca*, baboon) clade at 78%, a baboon clade at 97%, and a (*Mandrillus, Papio*) clade at 73%. The analysis of the hominoid qualitative data yielded one clade, which incorrectly linked *Gorilla* and *Pan* to the exclusion of the other taxa (92%).

## Discussion

The results of the parsimony and bootstrap tests indicate that cladistic analyses based on standard craniodental characters cannot be relied on to reconstruct the phylogenetic relationships of the hominoids, papionins, and, by extension, the fossil hominins. More problematically, the tests suggest that such analyses can strongly support phylogenetic hypotheses that are misleading. For example, the bootstrap-based tests indicate that craniodental data can return impressive levels of statistical support (e.g., 97%) for patterns of phylogenetic relationship that are most likely incorrect. In other words, cladistic analyses of higher primate craniodental morphology may yield not only “false-positive” results, but false-positive results that pass, by a substantial margin, the statistical test favored by many researchers. It should be noted that the strong bootstrap support for the clades undermines the objection that, because they were designed to recover hominin phylogeny, standard paleoanthropological measurements cannot be expected to recover the phylogenetic relationships of other taxa. Even if this were true, which we dispute, the logic of cladistics is such that the measurements should be ambiguous with regard to the relationships of the extant hominoids and papionins; they most certainly should not provide strong support for “false” clades. It should also be noted that the results of the quantitative analyses cannot be dismissed as artifacts of size-related shape change, since the most parsimonious cladograms do not group the taxa in the manner that would be expected if skull size is the principal signal in the transformed data.

The results of our analyses are in line with Hartman’s (24) finding that hominoid dental characters are not reliable for phylogenetic reconstruction. More importantly, they indicate that the problem is not confined to the dentition of the hominoids. It afflicts other groups of higher primates, and extends to quantitative and qualitative cranial characters. Our results suggest that paleoanthropologists should not rely on phylogenetic hypotheses for fossil hominins that are based solely on craniodental evidence. Most likely, these hypotheses reflect a mixture of the true phylogeny and the phylogenetically misleading effects of convergence, parallelism, reversal, and/or behaviorally induced morphogenesis. If anything, it is likely that extrapolating from the results of the present study to the results of hominin cladistic studies overestimates the reliability of the latter, since we did not account for two factors that routinely complicate

analyses of the hominin fossil record: namely, uncertain species identification and intraspecific morphological change through time. We stress that these results do not indicate that the cladistic methodology is flawed, or that primate craniodental data are problematic for phylogenetic reconstruction at all taxonomic levels. Rather, our results show that the type of craniodental characters that have hitherto been used in hominin phylogenetics are probably not reliable for reconstructing the phylogenetic relationships of higher primate species and genera, including those among the hominins. Although this study addressed the reliability of hominin phylogenetic hypotheses, it is noteworthy that our results are in line with recent observations regarding the use of osteological data to reconstruct the phylogenetic relationships of fossil catarrhines (48), Miocene hominoids (49), and fossil and extant colobines (50). There is, we suggest, a pressing need to undertake comparable analyses to the one reported here for other mammalian groups whose fossil members are also known principally from craniodental evidence.

How can the reliability of human phylogenetic hypotheses be improved? One strategy is to devise techniques for analyzing hominin craniodental morphology that are more sensitive to any phylogenetic signal than the methods presently in use (51, 52). Another approach is to develop rigorous comparative methods for discriminating between phylogenetically informative and phylogenetically misleading craniodental similarities. For example, the pursuit of detailed information about the ontogeny of characters may help distinguish homologous and homoplastic resemblances (53–56). Likewise, functional analyses may enable researchers to predict where similarities resulting from behavior-induced morphogenesis are likely to occur in the hominin cranium (13, 19–21). A third strategy is to develop techniques for assigning postcranial specimens to taxa in the absence of associated skeletons, thereby overcoming the taphonomy-imposed focus on craniodental morphology and enabling hominin cladistic analyses to be based on a wider sample of the phenotype (57). Significantly, recent studies suggest that data sets based on cranial, postcranial, and soft-tissue characters can yield relationships that are compatible with low-level primate molecular phylogenies (37, 58–60). Last, we suggest that more attention should be paid to nonmorphological lines of evidence that may have a bearing on hominin phylogenetic relationships, such as the temporal distribution of the species (61).

We thank the following institutions for access to specimens in their care: the Natural History Museum (London), the Anthropologisches Institut und Museum (Universität Zürich-Irchel, Zurich), the Muséum d’Histoire Naturelle (Geneva), the Muséum d’Histoire Naturelle (Paris), and the Museum für Naturkunde (Humboldt-Universität zu Berlin, Berlin). We also thank A. T. Chamberlain for allowing us to use his data. This study benefited from the advice of L. C. Aiello, A. Bilsborough, A. T. Chamberlain, M. C. Dean, D. E. Lieberman, P. O’Higgins, C. R. C. Paul, D. R. Pilbeam, M. Ruta, D. S. Strait, and several anonymous reviewers. M.C. was funded by a Natural Environment Research Council studentship and is now funded by a Wellcome Trust Bioarchaeology Fellowship. B.W. was funded by The Leverhulme Trust and is now funded by The Henry Luce Foundation.

1. Johanson, D. C., White, T. D. & Coppens, Y. (1978) *Kirtlandia* **28**, 1–14.
2. Walker, A. C., Leakey, R. E. F., Harris, J. M. & Brown, F. H. (1986) *Nature (London)* **322**, 517–522.
3. White, T. D., Suwa, G. & Asfaw, B. (1994) *Nature (London)* **371**, 306–312.
4. White, T. D., Suwa, G. & Asfaw, B. (1995) *Nature (London)* **375**, 88.
5. Leakey, M. G., Feibel, C. S., McDougall, I. & Walker, A. C. (1995) *Nature (London)* **376**, 565–571.
6. Brunet, M., Beauvilain, A., Coppens, Y., Heintz, E., Moutaye, A. H. E. & Pilbeam, D. R. (1996) *C. R. Acad. Sci. Paris* **322**, 907–913.
7. Bermudez de Castro, J. M., Arsuaga, J. L., Carbonell, E., Rosas, A., Martinez, I. & Mosquera, M. (1997) *Science* **276**, 1392–1395.
8. Asfaw, B., White, T. D., Lovejoy, O., Latimer, B., Simpson, S. & Suwa, G. (1999) *Science* **284**, 629–635.
9. Stringer, C. B. (1987) *J. Hum. Evol.* **16**, 135–146.
10. Chamberlain, A. T. & Wood, B. A. (1987) *J. Hum. Evol.* **16**, 119–133.
11. Wood, B. A. (1991) *Koobi Fora Research Project, Volume 4: Hominid Cranial Remains* (Oxford Univ. Press, Oxford).
12. Skelton, R. R. & McHenry, H. M. (1992) *J. Hum. Evol.* **23**, 309–349.
13. Lieberman, D. E., Pilbeam, D. R. & Wood, B. A. (1996) *J. Hum. Evol.* **17**, 503–511.
14. Strait, D. S., Grine, F. E. & Moniz, M. A. (1997) *J. Hum. Evol.* **32**, 17–82.
15. Corruccini, R. S. (1994) in *Integrative Approaches to the Past: Paleoanthropological Advances in Honor of F. Clark Howell*, eds. Corruccini, R. S. & Ciochon, R. L. (Prentice-Hall, Englewood Cliffs, NJ), pp. 167–183.
16. Wood, B. A. & Collard, M. (1999) *Science* **284**, 65–71.
17. Trinkaus, E. (1990) *Am. J. Phys. Anthropol.* **83**, 1–11.

18. Trinkaus, E. (1992) in *Continuity or Replacement: Controversies in Homo sapiens Evolution*, eds. Bräuer, G. & Smith, F. H. (Balkema, Rotterdam, the Netherlands), pp. 1–7.
19. Lieberman, D. E. (1995) *Curr. Anthropol.* **36**, 159–197.
20. Lieberman, D. E. (1997) *Annu. Rev. Anthropol.* **26**, 185–210.
21. Lieberman, D. E. (1999) *Evol. Anthropol.* **7**, 142–151.
22. Ruvolo, M. (1997) *Mol. Biol. Evol.* **14**, 248–265.
23. Harris, E. E. & Disotell, T. R. (1998) *Mol. Biol. Evol.* **15**, 892–900.
24. Hartman, S. E. (1988) *J. Hum. Evol.* **17**, 489–502.
25. Fitch, W. M. & Atchley, W. R. (1987) in *Molecules and Morphology in Evolution: Conflict or Compromise?*, ed. Patterson, C. (Cambridge Univ. Press, Cambridge, U.K.), pp.203–216.
26. Atchley, W. R. & Fitch, W. M. (1991) *Science* **254**, 554–558.
27. Hillis, D. M., Bull, J. J., White, M. E., Badgett, M. R. & Molineux, I. J. (1992) *Science* **255**, 589–592.
28. Rae, T. C. (1998) *Cladistics* **14**, 221–228.
29. Ahern, J. H. (1998) *Am. J. Phys. Anthropol.* **105**, 461–480.
30. Wood, B. A., Li, Y. & Willoughby, C. (1991) *J. Anat.* **174**, 185–205.
31. Chamberlain, A. T. (1987) Ph.D. thesis (Univ. of Liverpool, Liverpool, U.K.).
32. Collard, M. (1998) Ph.D. thesis (Univ. of Liverpool, Liverpool, U.K.).
33. Jungers, W. L., Falsetti, A. B. & Wall, C. E. (1995) *Yearbook Phys. Anthropol.* **38**, 137–161.
34. Thorpe, R. S. (1984) *Evolution (Lawrence, Kans.)* **38**, 244–255.
35. Creel, N. (1986) *Syst. Zool.* **35**, 81–99.
36. Strait, D. S., Moniz, M. & Strait, P. (1996) *Syst. Biol.* **45**, 67–78.
37. Shoshani, J., Groves, C. P., Simons, E. L. & Gunnell, G. F. (1996) *Mol. Phyl. Evol.* **5**, 102–154.
38. Braga, J. (1995) *Folia Primatol.* **65**, 144–153.
39. Andrews, P. (1987) in *Molecules and Morphology in Evolution: Conflict or Compromise?*, ed. Patterson, C. (Cambridge Univ. Press, Cambridge, U.K.), pp. 23–53.
40. Schwartz, J. H. (1984) *Curr. Anthropol.* **25**, 655–672.
41. Delson, E. & Andrews, P. (1975) in *Phylogeny of the Primates: A Multidisciplinary Approach*, eds. Lockett, W. P. & Szalay, F. S. (Plenum, New York), pp. 405–446.
42. Groves, C. P. (1986) in *Comparative Primate Biology, Volume 1: Systematics, Evolution and Anatomy*, eds. Swindler, D. R. & Erwin, J. (Liss, New York), pp. 187–217.
43. Swofford, D. L. (1991) *Phylogenetic Analysis Using Parsimony Version 3.0s* (Illinois Natural History Survey, Champaign, IL).
44. Felsenstein, J. (1985) *Evolution (Lawrence, Kans.)* **39**, 783–791.
45. Sanderson, M. J. (1995) *Syst. Biol.* **44**, 299–320.
46. Hillis, D. M. & Bull, J. J. (1993) *Syst. Biol.* **42**, 182–192.
47. Slowinski, J. (1993) *Syst. Biol.* **42**, 155–165.
48. Harrison, T. (1993) in *Species, Species Concepts and Primate Evolution*, eds. Kimbel, W. H. & Martin, L. B. (Plenum, New York), pp. 345–371.
49. Pilbeam, D. R. (1996) *Mol. Phylogenet. Evol.* **5**, 155–168.
50. Jablonski, N. (1999) *Curr. Biol.* **9**, 122.
51. Cheetham, A. H. (1987) *Paleobiology* **13**, 286–296.
52. Budd, A. F. & Coates, A. G. (1992) *Paleobiology* **18**, 425–446.
53. Wood, B. A. (1988) in *Evolutionary History of the “Robust” Australopithecines*, ed. Grine, F. E. (Aldine de Gruyter, New York), pp. 269–284.
54. Bromage, T. G. (1989) *J. Hum. Evol.* **18**, 751–773.
55. McCollum, M. (1999) *Science* **284**, 301–305.
56. Lovejoy, C. O., Cohn, M. J. & White, T. D. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 13247–13252.
57. Aiello, L. C. & Wood, B. A. (1994) *Am. J. Phys. Anthropol.* **95**, 409–426.
58. Begun, D. R., Ward, C. V. & Rose, M. D. (1997) in *Function, Phylogeny, and Fossils: Miocene Hominoid Evolution and Adaptations*, eds. Begun, D. R., Ward, C. V. & Rose, M. D. (Plenum, New York), pp. 389–415.
59. Groves, C. P. (2000) in *Old World Monkeys*, eds. Whitehead, P. F. & Jolly, C. J. (Cambridge Univ. Press, Cambridge, U.K.), in press.
60. Gibbs, S. (1999) Ph.D. thesis (Univ. of Liverpool, Liverpool, U.K.).
61. Paul, C. R. C. (1982) in *Problems of Phylogenetic Reconstruction*, eds. Joysey, K. A. & Friday, A. E. (Academic, London), pp. 75–117.