

On the Reliability of Recent Tests of the Out of Africa Hypothesis for Modern Human Origins

GÜNTER BRÄUER,^{1*} MARK COLLARD,² AND CHRIS STRINGER³

¹Institute of Human Biology, University of Hamburg, Hamburg, Germany

²Department of Anthropology, Washington State University, Pullman, Washington

³Department of Palaeontology, Natural History Museum, London, United Kingdom

ABSTRACT

In this paper we critique two recent studies that have been claimed to disprove the Out of Africa hypothesis for modern human origins (Hawks et al., 2000; Wolpoff et al., 2001). We show that the test prediction employed by Hawks et al. (2000) and Wolpoff et al. (2001) is not relevant to many versions of the Out of Africa hypothesis, and that the key specimens they used are problematic in terms of morphological representativeness. We also show that there are significant problems with the character state datasets employed in the studies. Lastly, we highlight evidence that the main method used in the studies (pairwise difference analysis) is not reliable when applied to the type of data employed by Hawks et al. (2000) and Wolpoff et al. (2001). In view of the foregoing, we contend that Hawks et al.'s (2000) and Wolpoff et al.'s (2001) claim to have disproved the Out of Africa hypothesis cannot be sustained.

© 2004 Wiley-Liss, Inc.

Key words: African replacement hypothesis; Out of Africa; Neanderthal; Australasia

It is obvious that a scientific hypothesis can only be considered disproved if a valid prediction of the hypothesis has been tested and falsified, and if the data and methods employed in the test are both suitable and reliable. Here we argue that recent claims that the Out of Africa hypothesis for modern human origins (also called the Recent African Origin hypothesis, African Replacement hypothesis, or simply the Replacement hypothesis) has been disproved (Hawks et al., 2000; Wolpoff et al., 2001) are invalid because the foregoing conditions were not satisfied. The prediction tested by Hawks et al. (2000) and Wolpoff et al. (2001) is incorrectly formulated, the data they employ are flawed, and there are substantial problems with the main method they use, pairwise difference analysis, which is a phenetic clustering technique that groups taxa on the basis of the number of differences the taxa exhibit. Thus, their claim to have disproved the Out of Africa hypothesis for modern human origins is without foundation. We begin with a brief description of the pairwise difference analyses carried out by Hawks et al. (2000) and Wolpoff et al. (2001), then discuss the predictions tested in the analyses, drawing attention to their shortcomings. Thereafter, we highlight problems with Hawks et al.'s (2000) and Wolpoff et al.'s (2001) selection of specimens and their choice of characters and character states. Lastly, we discuss a study that has cast doubt on Hawks et al.'s (2000) and Wolpoff et al.'s (2001) application of pairwise difference analysis to hominid cranial data.

TESTS OF THE OUT OF AFRICA HYPOTHESIS

Hawks et al. (2000) investigated the affinities of WLH-50 with the aid of pairwise difference analysis. WLH-50 is a modern human fossil from the Willandra Lakes area of Australia that has been dated to 15–13 ka BP (Simpson and Grün, 1998). Using 16 qualitative cranial characters, Hawks et al. (2000) compared WLH-50 with 10 Late Pleistocene *H. sapiens* fossils from Africa and western Asia and 7 specimens from the Javanese site of Ngandong that are Pleistocene in date, and which are usually classified as *Homo erectus* (Rightmire, 1990). The comparisons revealed that WLH-50 was closer to six of the seven specimens from Ngandong and was only separated

The current affiliation of Mark Collard is Department of Anthropology and Sociology, University of British Columbia, Vancouver, Canada.

*Correspondence to: Dr. Günter Bräuer, Institut für Humanbiologie, Universität Hamburg, Allende-Platz 2, 20146 Hamburg, Germany. Fax: 49-40-4320-599833.

E-mail: guenter.braeuer@uni-hamburg.de

Received 5 February 2004; Accepted 5 May 2004

DOI 10.1002/ar.a.20064

Published online 9 July 2004 in Wiley InterScience (www.interscience.wiley.com).

from the seventh Ngandong specimen by one western Asian specimen (Skhul 9). Hawks et al. (2000) inferred from this that the Ngandong hominids were most probably one of the ancestors of WLH-50, and that the Out of Africa hypothesis can therefore be discounted.

Wolpoff et al. (2001) replicated the WLH-50 analysis carried out by Hawks et al. (2000) and also outlined a pairwise difference analysis that focused on the affinities of two early modern human fossils from the Moravian Upper Paleolithic site of Mladeč (Mladeč 5 and Mladeč 6). In the latter analysis, the Mladeč crania were compared on the basis of 30 and 22 qualitative cranial characters, respectively, to early modern human crania from Skhul and Qafzeh, Israel, and to Neanderthal crania from a number of European sites. Although specimens from Skhul and Qafzeh were found to have the smallest number of differences to both Mladeč 5 and 6, the average number of differences between Mladeč 5 and 6 and the Neanderthal specimens was similar to the average number of differences between Mladeč 5 and 6 and the early modern human specimens from Skhul and Qafzeh. Wolpoff et al. (2001) argued that the results of their pairwise difference analyses falsified the Out of Africa hypothesis for modern human origins and instead supported the multiregional evolution hypothesis.

PREDICTIONS

The tests of the Out of Africa hypothesis conducted by Hawks et al. (2000) and Wolpoff et al. (2001) focused on what these authors evidently considered to be the main prediction of the hypothesis, namely, that there should be no genetic continuity or interbreeding between early modern humans and archaic hominids in Europe, East Asia, and Southeast Asia [see also Thorne and Wolpoff (2003: p. 48)]. However, this prediction is misleading. As we have pointed out elsewhere (Stringer and Bräuer, 1994; Bräuer and Stringer, 1997; Stringer, 2002), the idea of complete replacement without any interbreeding is one variant of the Out of Africa hypothesis, often identified with a particular interpretation of the mitochondrial DNA data, dubbed the Eve theory (Frayser et al., 1993). Complete replacement does not feature in all versions of the Out of Africa hypothesis, and therefore we contend that it would not be falsified by demonstration of some modern-archaic human gene flow outside of Africa. Both before and after the pioneering study by Cann et al. (1987), proponents of the Out of Africa hypothesis accepted the possibility of a certain amount of gene flow between the migrating early anatomically modern humans and the non-African archaic groups (Bräuer, 1984: p. 395, 1992: p. 95, 2001; Cann, 1992: p. 71; Stringer, 1992: p. 20, 2001; Stringer and Bräuer, 1994: p. 416). Thus, even if the results of Hawks et al.'s (2000) and Wolpoff et al.'s (2001) analyses were reliable, they would only shed light on the accuracy of the extreme Eve theory. They would not disprove the Out of Africa hypothesis.

SPECIMENS

Hawks et al.'s (2000) and Wolpoff et al.'s (2001) decision to focus on WLH-50 in their Australasian tests of the African replacement hypothesis is questionable for two reasons. First, WLH-50 cannot be considered to be representative of early Australians in terms of morphology. Not only is it the most robust and supposedly archaic-looking

of the Willandra Lakes specimens (Wolpoff, 1999: p. 736), but also its cranial vault has apparently been modified by severe pathological hyperostosis (Webb, 1990; Brown, 1992). Second, recent dating work suggests that WLH-50 is much younger than previously thought. Although it was a surface find, it had been argued that the specimen could be older than 35 ka and was thus representative of the earliest Australians, from initial ESR dates on bone material, and from its extreme morphology (Wolpoff, 1999: p. 736). However, uranium-series dates reported by Simpson and Grün (1998) indicate that the calvaria is only around 14 ka in age. Thus, WLH-50 fits with the great morphological variation in robustness seen among final Pleistocene/early Holocene Australians, which is likely due to drift effects as well as adaptation to changing climatic conditions and diet (Brown, 1992; Klein, 1999, Stone and Cupper, 2003). Together, the pathological alterations and age of WLH-50 make it a poor specimen on which to base a test of the African Replacement hypothesis since it is hardly representative of early Australian *H. sapiens*.

There are alternative specimens for such a test. For example, there is another similarly preserved possible male specimen from the Willandra Lakes, Lake Mungo 3 (LM 3) (Bowler and Thorne, 1976), which according to several absolute dates of 43 ka up to ~ 60 ka (Thorne et al., 1999; Bowler and Magee, 2000) is more likely to represent the original inhabitants of Australia and would thus have been a much better focus for Hawks et al.'s (2000) and Wolpoff et al.'s (2001) tests. According to Wolpoff (1999: p. 736–738), this fossil does not reflect Indonesian *H. erectus*/Australian *H. sapiens* continuity because it is gracile and has a well-rounded forehead, thin vault bone, weak muscle attachments, and weak supraorbital development. Thus, it seems likely that if Hawks et al. (2000) and Wolpoff et al. (2001) had used LM 3 rather than WLH-50 in their tests, their results would not have supported continuity between Indonesian *H. erectus* and Australasian *H. sapiens*. This is probably also true for several other early Australian fossils, such as Mungo 1 and Keilor.

In our view, Wolpoff et al.'s (2001) European test is as dubious as their Australasian test in terms of specimen selection. In their European test, Wolpoff et al. (2001) focused on the male specimens from Mladeč, Mladeč 5 and 6, even though more complete female specimens, Mladeč 1 and 2, have also been recovered. Wolpoff et al. (2001: p. 294) justified this decision on the grounds that their "identification of males in Mladeč and in the earlier samples of potential ancestors is reliable." This statement is questionable even for the Mladeč sample, where Mladeč 1 [omitted by Wolpoff et al. (2001)] has sometimes been identified as male, and this is also true for the comparative sample since they included Qafzeh 9, thought to be a female by other workers (Vandermeersch, 1981). Because they focused on the male specimens, Wolpoff et al. (2001) inevitably excluded facial morphology from their analyses, which is unfortunate for two reasons: the face is undoubtedly a key area in identifying modern or Neanderthal affinities, and a recent metric study of facial shape in the Mladeč specimens revealed no evidence of Neanderthal affinities (Bräuer et al., 2004). For example, the nasiofrontal angles of the Mladeč 1, 2, and 5 specimens were found to diverge strongly from the smaller values for the Neanderthals and to fall instead among the values for other Upper Paleolithic European crania and for the Skhul and

Qafzeh specimens. The same holds true for the zygomaxillary angle, which can only be determined in Mladeč 1 (Bräuer et al., 2004). A principal-components analysis of facial shape and projection conducted by Bräuer et al. (2004) supported a distinction between Neanderthal and early modern humans in facial morphology and highlighted the differences between Mladeč 1 and the Neanderthals. This is also in agreement with an analysis of vault morphology of the inferred female specimens Mladeč 1 and 2 (Bräuer and Broeg, 1998). Thus, there is good reason to believe that the results of Wolpoff et al.'s (2001) European test would have been different if they had included both the male and the female crania from the site rather than limiting their sample to the former.

CHARACTERS AND CHARACTER STATES

Evaluating Hawks et al.'s (2000) and Wolpoff et al.'s (2001) character data is difficult because their matrix of character scores is not available for examination (M. Wolpoff, personal communication). However, some major problems are obvious. We will deal first with the data employed in Hawks et al.'s (2000) and Wolpoff et al.'s (2001) Australasian tests, then examine the data used in Wolpoff et al.'s (2001) European test.

Australasian Test

The nonmetric character state data set employed by Hawks et al. (2000) and Wolpoff et al. (2001) is questionable for several reasons. First, it only contains a small selection of the characters that have featured in the modern human origins debate to date. For example, Stringer et al. (1984) highlight 10 characters that they believe are synapomorphic in archaic Africans and all modern humans and therefore support the Out of Africa hypothesis. None of these characters is among the nonmetric characters employed by Hawks et al. (2000) and Wolpoff et al. (2001). Similarly, Hawks et al.'s (2000) and Wolpoff et al.'s (2001) data set lacks a number of characters that were used in Frayer et al.'s (1993) widely cited contribution to the debate. Although Hawks et al. (2000: p. 4) claim that their "nonmetric traits represent an exhaustive list of all characteristics that could be scored on WLH-50," and Wolpoff et al. (2001: p. 294) state that the 16 characters they employed were the only ones that could be scored unambiguously on WLH-50, Frayer et al. (1993: p. 24) presented several additional characters that they argued linked WLH-50 and the Ngandong *H. erectus* crania to the exclusion of the African Ngaloba specimen, LH 18. These features include location of greatest cranial breadth, parietal boss, form of the external mastoid process, sagittal length of the frontal relative to the vault length, and sagittal curvature of the frontal just above the supraorbital region. Hawks et al. (2000) and Wolpoff et al. (2001) do not explain why they excluded so many of the characters examined in earlier studies. In our view, this undermines confidence in the results of their analyses.

Second, Hawks et al. (2000) and Wolpoff et al. (2001) assign different states to some of the characters that were employed in earlier studies. For example, Hawks et al. (2000: p. 7) describe parietal sagittal keel as a thickening along the sagittal suture and suggest that this character is "scored as present . . . as long as it can be identified anywhere along the suture." The same character also appears in Frayer et al. (1993: p. 24), but here it is deemed present

only if it is "marked on the anterior half of the parietals." A similar situation is seen regarding the character "temporal line forms a ridge." Hawks et al. (2000: p. 7) indicate that this character is "scored as present . . . if the temporal lines form a ridge along the frontal bone, posterior to the orbital constriction." In contrast, Frayer et al. (1993: p. 24) score this character as present if the ridge for the temporal extends across the entire frontal bone, and absent if the ridge only encompasses the temporal notch. In both of these cases, Hawks et al. (2000) and Wolpoff et al. (2001) employ less restrictive state definitions than Frayer et al. (1993), and in both cases they fail to justify this course of action. These examples highlight a fundamental problem with the characters used in such studies. While some are metric and can be determined relatively objectively by measurement, others are nonmetric and potentially observer-dependent. There is no evidence that they employed an overt, standardized, and reproducible method of scoring, such as the one developed by Lahr (1996), and this raises questions about both the internal consistency of the data and their reliability.

Third, the phylogenetic utility of a number of characters employed by Hawks et al. (2000) and Wolpoff et al. (2001) is questionable. One of these is the prebregmatic eminence. This character varies greatly within and among human groups. For instance, Lahr (1994: p. 35) found a prebregmatic eminence in only one specimen of her recent Australo-Melanesian sample ($n = 25$). In contrast, Bräuer and Mbua (1992: p. 86) demonstrated the frequent occurrence of this variably expressed character in African Middle Pleistocene specimens as well as in African and Asian *H. erectus*. In addition, it is possible that cultural practices and/or disease can affect whether an individual exhibits a prebregmatic eminence or not. Brown (1981), for example, found that the presence of this character in Australian human crania is often due to artificial deformation of the vault. More recently, Antón and Weinstein (1999: p. 195) have argued that the high frequency of hyperostotic characters, including bregmatic eminences, in crania from the Australian sites of Coobool Creek, Kow Swamp, and Narrurie is influenced by both artificial deformation and pathological hypervascularity/hyperostosis. Given its variability and apparent ready susceptibility to nongenetic influences, the prebregmatic eminence is unlikely to provide reliable phylogenetic information (Brown, 1981).

The postlambdoidal eminence (Hawks et al., 2000; Wolpoff et al., 2001) is another character whose phylogenetic utility is dubious. Following Weidenreich (1951), Wolpoff et al. (1984) claim that this character links Ngandong *H. erectus* and Australasian *H. sapiens*. In support of this contention, they cite a study by Larnach and Macintosh (1974) in which postlambdoidal eminence frequency was examined in samples of human crania from different continents. Wolpoff et al. (1984) draw attention to Larnach and Macintosh's (1974) finding that their Australian and New Guinean samples exhibit the highest frequencies of postlambdoidal eminences and contrast this with Larnach and Macintosh's (1974) European and Asian samples, in which postlambdoidal eminences were not found. However, as Groves (1989) points out, Wolpoff et al. (1984) fail to mention that the Australian and New Guinean specimens that exhibit a postlambdoidal eminence only represent 1.4% and 1.2% of the Australian ($n = 207$) and New Guinean ($n = 80$) samples, respectively. By any standards, this is weak evidence, and we contend that in view of its

infrequent occurrence in any human group, the postlambdoidal eminence must be considered to be a poor character for phylogenetic analysis. It is also unfortunate that, while Wolpoff et al. (1984) selectively used the data of Larnach and Macintosh (1974), they failed to acknowledge their final conclusions (Larnach and Macintosh, 1974: p. 102): "The scores completely and unequivocally discriminate Solo crania from those of the other racial groups used in this comparison and they identify Solo skulls specifically. . . . The occurrence of traits unique to Solo man is considered and the difficulties they introduce to any hypothesis which would derive Australians from Solo man are discussed."

Several of the other characters used by Hawks et al. (2000) and Wolpoff et al. (2001) suffer from similar shortcomings to the prebregmatic eminence and the postlambdoidal eminence. For example, it is not clear that presence/absence is an adequate scoring system for the blunt superior margin of the orbit, the projectinginion, the mastoid crest, and the supramastoid crest. Likewise, the states of several of the characters seem likely to be greatly affected by muscle mass and activity. The characters in question include transversely extensive nuchal torus, suprainsiac fossa, temporal line forms a ridge, projectinginion, angular torus, strongly developed linea obliquus, lateral frontal trigone, mastoid crest, and supramastoid crest. As Lieberman (1995) has pointed out, characters whose expression is heavily influenced by environmental factors such as muscle activity should be avoided in phylogenetic analyses because their states in different taxa have a high probability of being homoplastic. It would seem, therefore, that most of the 16 characters employed by Hawks et al. (2000) and Wolpoff et al. (2001) are unlikely to yield reliable results when employed in a pairwise difference analysis or any other phylogenetic analysis for that matter.

In sum, Hawks et al.'s (2000) and Wolpoff et al.'s (2001) Australasian data set is deficient in a number of respects. It contains few characters that have previously been deemed to be relevant to the modern origins debate. Furthermore, in several cases in which Hawks et al. (2000) and Wolpoff et al. (2001) have employed characters that have featured in previous modern human origins-oriented studies, they have used character states that are less rigorously defined and done so without providing any justification. In addition, the majority of the characters included in the Hawks et al.'s (2000) and Wolpoff et al.'s (2001) Australasian data set can be shown to be either of questionable phylogenetic significance or unreliable for phylogenetic reconstruction. The corollary of these criticisms is that Hawks et al.'s (2000) and Wolpoff et al.'s (2001) conclusion that their Australasian test disproves the Out of Africa hypothesis cannot be sustained.

European Test

There are also several problems with the data set employed by Wolpoff et al. (2001) in their European test. One is that several of their "nonmetric traits" characters are not nonmetric at all. Rather, they are metric characters that have been divided into arbitrarily defined qualitative character states. These include thick parietal at "asterion > 9 mm," "broad frontal > 125 mm," "broad occiput > 120 mm," "long glabella-bregma chord > 113 mm," "thick occipital at lambda > 8 mm," "medially tall supraorbital > 19 mm," and "long occipital plane > 60 mm." There is no

evidence that Wolpoff et al. (2001) size-corrected their metric data, or that they employed a recognized method for converting metric data into qualitative character states (Thorpe, 1984; Baum, 1988). In addition, Wolpoff et al. (2001) did not offer any justification for the character states into which they divide characters. Accordingly, these measurement-based characters cannot be accepted as reliable data, since there is no reason to believe that their states accurately reflect the similarities and differences among the specimens.

A second substantial problem with the data that Wolpoff et al. (2001) employ in their European test is that many of the qualitative characters are the subject of debate regarding their expression in early modern humans and Neanderthals. This problem can be illustrated with reference to the occipital bun. Wolpoff et al. (2001) treat the occipital bun as a homologous character that can be scored unproblematically as present or absent in the specimens in their sample. However, it is not at all clear that this is a defensible course of action. Several researchers have pointed out that the Neanderthal bun or "chignon" differs in a number of respects from the projecting occipital area seen in some Upper Paleolithic crania (Smith, 1982; Trinkaus and Le May, 1982; Caspari, 1991; Churchill and Smith, 2000). Among the key Neanderthal characteristics are extensive lambdoidal flattening, more acute angulation between the nuchal plane and the general orientation of the occipital plane, and laterally wider bun. The differences between the conditions seen in Neanderthals and early modern humans led Smith (1982) to distinguish the morphology exhibited by the latter as a "hemibun." More recently, Lieberman et al. (2000: p. 291) have raised the possibility that the specialized Neanderthal bun and the early modern human hemibun may not be homologous.

If Lieberman et al. (2000) are right, then counting the buns in Neanderthals and the hemibuns in Mladeč as occipital buns will lead to pairwise difference scores that do not correctly represent the specimens' affinities. However, even if Lieberman et al. (2000) are wrong, it is not obvious how the specimens in Wolpoff et al.'s (2001) sample should be scored for this character. For example, Frayer (1992a: p. 22) argues in relation to the bun-like structure exhibited by Mladeč 6 that "no specimens from the Near East or Africa provide a reasonable ancestral state to Mladeč 6 or virtually any of the other early Upper Paleolithic material," whereas Smith et al. (1995) and Bräuer and Broeg (1998) concluded that the bun-like structures exhibited by several early modern European specimens, including those from Mladeč, are similar to the bun-like structures exhibited by the specimens from Jebel Irhoud, north Africa. There is a comparable diversity of opinion regarding the specimens from Skhul and Qafzeh. Frayer (1992a: p. 21) avers that the "specimens from Skhul/Qafzeh lack development of either lambdoidal flattening or occipital buns." This conclusion is also supported by Smith et al. (1995: p. 201), who argue that "buns are not present in the Near Eastern Neanderthals or in the Skhul/Qafzeh hominids." However, Wolpoff (1999) arrived at a different assessment. Commenting on the specific form of occipital bunning in the Jebel Irhoud specimens, he noted that "it is ironic that only the earlier non-Neanderthals of the Levant are bunned in a similar manner" (Wolpoff, 1999: p. 613). Most recently, Bräuer (2003) demonstrated close similarities in the parieto-occipital con-

tours of Mladeč 5 and 6 to those of Qafzeh 6 and Skhul 5 and concluded that if the conditions in the Mladeč crania are considered as hemibuns, there is no reason to classify the morphology in the two Skhul/Qafzeh specimens differently. In fact, hemibun-like conditions occur in great variability not only in Europe, the Near East, and Africa but also in the Far East, as for example in the late Pleistocene Liujiang cranium of China (Trinkaus and Le May, 1982; Bräuer, 1992: p. 93; Stringer, 1992: p. 18; Bräuer and Broeg, 1998; Wu, 1998: p. 282). Since Wolpoff et al.'s (2001) data set is not available for inspection by other researchers, it is not possible to compare their codes for occipital bun with those employed in earlier studies or with our own observations. Nevertheless, it should be obvious from the foregoing that Wolpoff et al.'s (2001) codes cannot be anything other than contentious.

Wolpoff et al.'s (2001) use of controversial qualitative characters can be further illustrated by consideration of the elliptical suprainiac fossa. As with the character occipital bun, Wolpoff et al. (2001) treat elliptical suprainiac fossa as a homologous character that can be scored as present or absent in the specimens in their sample without any problem. However, the assessment of this character is also clouded by different definitions and views, which make any simple scoring highly disputable. In his detailed treatment of the suprainiac fossa, Nara (1994) demonstrated its development from Preneanderthal conditions up to the most derived conditions in late Neanderthals, in which the suprainiac fossa is generally large, wide, and either rectangular or triangular in shape with a horizontal base. In contrast to this shape, Nara (1994) describes the depressions found in early modern Europeans as different in forming an inverse triangle accompanied by a downward curving V-shaped crest connected with the external occipital protuberance. Frayer (1992b: p. 182) basically described the same differences and added: "Thus, while present in a similar position in post-Neanderthal Europeans, the fossa is not identical and seems to be more clearly an artifact of the very strong supreme nuchal lines and generally prominent external inion." According to Caspari (1991: p. 184), the suprainiac fossa in Neanderthals is a characteristically wide and oval-shaped resorptive surface, filling the area above the occipital transverse torus, whereas in other groups, resorptive surfaces in the same area have a different morphology. In view of these differences, Caspari (1991: p. 184) raised the issue of whether it is actually valid to compare Neanderthal suprainiac fossae with the resorptive pits in this area found in early modern humans [see also Lieberman (1995: p. 170)].

However, once again, even if the suprainiac fossae exhibited by Neanderthals and the pits in early modern humans are determined to be homologous, it is not clear how the specimens examined by Wolpoff et al. (2001) should be scored. Frayer (1986: p. 251) opined that "neither Mladeč 6 nor 5 have a distinct suprainiac fossa." This assessment is in line with those of Gambier (1992: p. 279) and Bräuer and Broeg (1998), who concluded that the resorptive pits of Mladeč 5 and 6 are not Neanderthal-like. In contrast, Caspari (1991: p. 152–153) described the small depression in Mladeč 6 as being "very similar to the suprainiac fossa of Neanderthals, though much less clearly expressed," even though elsewhere she noted that "the shapes of these Upper Paleolithic suprainiac fossae are quite different from those of Neanderthals, corre-

sponding to the differences in the medial portions of the torus" (p. 185). In 1992, Frayer changed his earlier assessment and recognized the presence of a suprainiac fossa in Mladeč 6. The Mladeč 6 resorptive pit, he suggested, was "the only possible exception" in having a Neanderthal-like morphology among the early modern Europeans (Frayer, 1992a: p. 22). Authors also disagree over the states exhibited by the Skhul and Qafzeh specimens in Wolpoff et al.'s (2001) sample. Several researchers have concluded that weak or stronger supranuchal depressions are present in Qafzeh 6, Skhul 5, and Skhul 9 (McCown and Keith, 1939: p. 280; Vandermeersch, 1981: p. 106; Gambier, 1992: p. 280; Nara, 1994: p. 54), and Tillier (1999: p. 141) has suggested that the juvenile specimen Qafzeh 11 exhibits an incipient fossa. Yet Frayer (1992b: p. 181) observed fossae in only two of the crania, Skhul 9 and Qafzeh 6. Thus, once again, it is clear that the scores for an elliptical suprainiac fossa are of questionable consistency. In view of the problems shown with regard to the bunning and suprainiac fossa, it is doubtful that features like "arched frontonasal suture," "glabellar depression," "well-separated mastoid-supramastoid crests," or "sagittal groove along vault posterior" can reasonably be used in Wolpoff et al.'s (2001) test.

It goes without saying that the results of any phylogenetic analysis can only be as good as the data employed in the analysis. Little confidence can be placed in the results of phylogenetic analyses of characters whose homology is questionable and/or whose states are controversial. Given the substantial problems with the data Wolpoff et al. (2001) employed in their Mladeč analysis, there is good reason to question the phylogenetic relationships they proposed. Therefore, we do not accept Wolpoff et al.'s (2001) assertion that the results of their Mladeč analysis disprove the Out of Africa hypothesis.

METHOD

As noted earlier, Hawks et al. (2000) and Wolpoff et al. (2001) employed pairwise difference analysis (PDA) in their tests of the African replacement hypothesis. To reiterate, PDA is a phenetic method that groups taxa on the basis of the number of differences they exhibit. Thus, in PDA, taxa that exhibit few differences are deemed to be more closely related than taxa that exhibit many differences because they must share a common ancestor that lived more recently. Hawks et al. (2000) and Wolpoff et al. (2001) justified their use of PDA on the grounds that it is often applied to genetic data to address demographic and phylogenetic questions. This is indeed the case, since PDA underpins the molecular clock (Hedrick, 2000). However, a recent study suggests that the application of PDA to hominid morphological data is not defensible (Collard and Franchino, 2002). The authors of this study evaluated the likely ability of PDA to recover phylogenetic information from hominid morphological data by conducting four pairwise difference analyses of craniodental and soft tissue data for the extant hominoids. These primates have been the focus of considerable molecular phylogenetic research, and there is now widespread agreement that their phylogenetic relationships can be considered known (Ruvolo, 1997; Gagneux and Varki, 2001; Page and Goodman, 2001). Accordingly, Collard and Franchino (2002) assumed that, if the phylogenetic hypotheses yielded by pairwise difference analyses of the hominid morphological data match the group's molecular phylogeny, pairwise

difference analysis can be relied on to recover phylogenetic information from primate morphological data. They also assumed that, if pairwise difference analyses of the hominoid morphological data yield phylogenetic hypotheses that disagree with the group's molecular phylogeny, pairwise difference analysis cannot be relied on to recover phylogenetic information from primate morphological data.

In all four of the analyses conducted by Collard and Franchino (2002), the phylogenetic hypotheses returned by pairwise difference analysis were incompatible with the molecular phylogeny for the extant hominoids. Given the robustness of the hominoid molecular phylogeny, these results indicate that pairwise difference analysis cannot be relied on to recover phylogenetic information from primate morphological data sets. It is possible that future studies will show that pairwise difference analysis is able to recover phylogenetic information from some data sets, but the results of Collard and Franchino's (2002) analyses indicate that it is unable to recover phylogenetic information from all primate morphological data sets. Thus, it cannot be assumed that a phylogeny yielded by a pairwise difference analysis of primate morphological data is accurate. The corollary of this is that, even if there were no problems with the data employed by Hawks et al. (2000) and Wolpoff et al. (2001), the results of their pairwise difference analyses would still not support their claim to have disproved the African replacement hypothesis.

Besides these basic problems regarding the application of PDA, a mere perusal of the calculated differences shown in Wolpoff et al. (2001)'s Tables 1 to 3 reveals that the method is problematic and inadequate, not even providing rough impressions of similarities between the specimens. For example, the Qafzeh 9 cranium, which is widely considered to be a female *H. sapiens*, is apparently more similar to the *H. erectus* specimen Ngandong 6 (six differences) than it is to WLH-50 (seven differences). Similarly, specimens from Skhul and Qafzeh show both the smallest number of differences (8 and 6) and the largest number (23 and 16) from the Mladeč 5 and 6 crania, even though the Skhul and Qafzeh specimens are usually regarded as being closely related to each other (Vandermeersch, 1981), and the Skhul, Qafzeh, and Mladeč fossils are all generally regarded as belonging to the same species. There would thus seem to be yet more reasons to question the reliability of Hawks et al.'s (2000) and Wolpoff et al.'s (2001) pairwise difference analyses.

CONCLUSIONS

In the present article we have highlighted what we consider to be some major shortcomings of two recent studies that have claimed to have disproved the Out of Africa hypothesis for modern human origins (Hawks et al., 2000; Wolpoff et al., 2001). These shortcomings can be summarized as follows. One, the prediction on which the studies focussed is not relevant to many versions of the Out of Africa hypothesis. Two, the key specimens employed in the studies are problematic in terms of morphological representativeness. Three, many relevant characters are ignored, and many of those that are examined are of questionable phylogenetic utility. Four, little confidence can be placed in the character state assignments used in the studies. Five, the main method used in the studies is inappropriate because, when applied to comparable data

sets, it yields phylogenetic relationships that are most likely incorrect. In view of the foregoing, we aver that Hawks et al.'s (2000) and Wolpoff et al.'s (2001) claim to have disproved the Out of Africa hypothesis cannot be sustained.

ACKNOWLEDGMENTS

The authors thank Erin Wallace, Dagmar Walkusch, and three anonymous reviewers for their help with the final version of the manuscript.

LITERATURE CITED

- Antón SC, Weinstein KJ. 1999. Artificial cranial deformation and fossil Australians revisited. *J Hum Evol* 36:195–209.
- Baum BR. 1988. A simple procedure for establishing discrete characters from measurement data, applicable to cladistics. *Taxon* 37:63–70.
- Bowler JM, Thorne AG. 1976. Human remains from lake Mungo: discovery and excavation of Lake Mungo III. In: Kirk RL, Thorne AG, editors. *The origin of the Australians*. Canberra: Australian Institute of Aboriginal Studies. p 127–138.
- Bowler JM, Magee JW. 2000. Redating Australia's oldest human remains: a sceptic's view. *J Hum Evol* 38:719–726.
- Bräuer G. 1984. A craniological approach to the origin of anatomically modern *Homo sapiens* in Africa and implications for the appearance of modern Europeans. In: Smith FH, Spencer F, editors. *The origins of modern humans: a world survey of the fossil evidence*. New York: Liss. p 327–410.
- Bräuer G. 1992. Africa's place in the evolution of *Homo sapiens*. In: Bräuer G, Smith FH, editors. *Continuity or replacement: controversies in *Homo sapiens* evolution*. Rotterdam: A.A. Balkema. p 83–98.
- Bräuer G, Mbua E. 1992. *Homo erectus* features used in cladistics and their variability in Asian and African hominids. *J Hum Evol* 22:79–108.
- Bräuer, Stringer CB. 1997. Models, polarization, and perspectives on modern human origins. In: Clark GA, Willermet CM, editors. *Conceptual issues in modern human origins research*. New York: Aldine de Gruyter. p 191–201.
- Bräuer G, Broeg H. 1998. On the degree of Neanderthal: modern continuity in the earliest Upper Palaeolithic crania from the Czech Republic—evidence from non-metrical features. In: Omoto K, Tobias PV, editors. *The origins and past of modern humans: towards reconciliation*. Singapore: World Scientific. p 106–125.
- Bräuer G. 2001. The "out-of-Africa" model and the question of regional continuity. In: Tobias PV, Raath MA, Moggi-Cecchi J, Doyle GA, editors. *Humanity from African naissance to coming millennia*. Florence: Witwatersrand University Press. p 183–189.
- Bräuer G. 2003. The Neanderthal-modern transition: models, gene flow and the fossil evidence. *Am J Phys Anthropol* 36(Suppl):71.
- Bräuer G, Broeg H, Stringer CB. 2004. The earliest Upper Palaeolithic crania from the Czech Republic and the question of Neanderthal-modern continuity: metrical evidence from the fronto-facial region. In: Finlayson C, Stringer CB, editors. *Proceedings of the Calpe 2001 conference, Gibraltar*. Chicago: University of Chicago Press.
- Brown P. 1981. Artificial cranial deformation: a component in the variation in Pleistocene Australian Aboriginal crania. *Archaeol Oceania* 16:156–167.
- Brown P. 1992. Recent human evolution in East Asia and Australasia. In: Aitken MJ, Stringer CB, Mellars PA, editors. *The origin of modern humans and the impact of chronometric dating*. Princeton, NJ: Princeton University Press. p 217–233.
- Cann RL, Stoneking M, Wilson AC. 1987. Mitochondrial DNA and human evolution. *Nature* 325:31–36.
- Cann RL. 1992. A mitochondrial perspective on replacement or continuity in human evolution. In: Bräuer G, Smith FH, editors. *Continuity or replacement: controversies in *Homo sapiens* evolution*. Rotterdam: A.A. Balkema. p 65–73.
- Caspari RE. 1991. The evolution of the posterior cranial vault in the central European Upper Pleistocene. PhD dissertation. Ann Arbor, MI: University of Michigan.

- Churchill SE, Smith FH. 2000. Makers of the early Aurignacian of Europe. *Yearbook Phys Anthropol* 43:61–115.
- Collard M, Franchino N. 2002. Pairwise difference analysis in modern human origins research. *J Hum Evol* 43:323–352.
- Frayser DW. 1986. Cranial variation at Mladeč and the relationship between Mousterian and Upper Paleolithic hominids. *Anthropos (Brno)* 23:243–256.
- Frayser DW. 1992a. Evolution at the European edge: Neanderthal and Upper Paleolithic relationships. *Préhist Eur* 2:9–69.
- Frayser DW. 1992b. The persistence of Neanderthal features in post-Neanderthal Europeans. In: Bräuer G, Smith FH, editors. *Continuity or replacement: controversies in Homo sapiens evolution*. Rotterdam: A.A. Balkema. p 179–188.
- Frayser DW, Wolpoff MH, Thorne AG, Smith FH, Pope G. 1993. Theories of modern human origins: the paleontological test. *Am Anthropol* 95:14–50.
- Gagneux P, Varki A. 2001. Genetic differences between humans and great apes. *Mol Phylogenet Evol* 18:2–13.
- Gambier D. 1992. The origin of modern man in Europe: a comparison of cranial data in central and western Europe. In: Toussaint M, editor. *Cinq millions d'années: l'aventure humaine*. Liège: E.R.-A.U.L. p 269–284.
- Groves CP. 1989. A regional approach to the problem of the origin of modern humans in Australasia. In: Mellars P, Stringer CN, editors. *The human revolution: behavioural and biological perspectives on the origins of modern humans*. Edinburgh: Edinburgh University Press. p 274–285.
- Hawks J, Oh S, Hunley K, Dobson S, Cabana G, Dayalu P, Wolpoff MH. 2000. An Australasian test of the recent African origin theory using the WLH 50 calvarium. *J Hum Evol* 39:1–22.
- Hedrick PW. 2000. *Genetics of populations*. Sudbury, MA: Jones and Barlett.
- Klein RG. 1999. *The human career: human biological and cultural origins*, 2nd ed. Chicago: University of Chicago Press.
- Lahr MM. 1994. The multiregional model of modern human origins: a reassessment of its morphological basis. *J Hum Evol* 26:23–56.
- Lahr MM. 1996. *The evolution of human diversity*. Cambridge: Cambridge University Press.
- Larnach SL, Macintosh NWG. 1974. A comparative study of Solo and Australian Aboriginal crania. In: Elkin AP, Macintosh NWG, editors. *Grafton Elliot Smith: the man and his work*. Sydney: Sydney University Press. p 95–102.
- Lieberman DE. 1995. Testing hypotheses about recent human evolution from skulls. *Curr Anthropol* 36:159–197.
- Lieberman DE, Pearson OM, Mowbray KM. 2000. Basicranial influence on overall cranial shape. *J Hum Evol* 38:291–315.
- McCown I, Keith A. 1939. *The Stone Age of Mount Carmel: the fossil human remains from the Levallois-Mousterian*, vol. 2. Oxford: Clarendon Press.
- Nara MT. 1994. *Etude de la variabilité de certains caractères métriques et morphologiques des Néandertaliens*. Bordeaux: Thèse de Docteur.
- Page SL, Goodman M. 2001. Catarrhine phylogeny: non-coding DNA evidence for a diphyletic origin of mangabeys and for a human-chimpanzee clade. *Mol Phylogenet Evol* 18:14–25.
- Rightmire GP. 1990. *The evolution of Homo erectus*. Cambridge: Cambridge University Press.
- Ruvolo M. 1997. Molecular phylogeny of the hominoids: inferences from multiple independent DNA data sets. *Mol Biol Evol* 14:248–265.
- Simpson J, Grün R. 1998. Non-destructive gamma spectrometric U-series dating. *Quaternary Sci Rev (Quaternary Geochronol)* 17:1009–1022.
- Smith FH. 1982. Upper Pleistocene hominid evolution in south-central Europe. *Curr Anthropol* 23:667–703.
- Smith FH, Falsetti B, Simmons T. 1995. Circum-Mediterranean biological connections and the pattern of late Pleistocene human evolution. In: Ullrich H, editor. *Man and environment in the Paleolithic*. Liège: E.R.A.U.L. p 167–179.
- Stone T, Cupper ML. 2003. Last glacial maximum ages for robust humans at Kow Swamp, southern Australia. *J Hum Evol* 45:99–111.
- Stringer CB, Hublin JJ, Vandermeersch B. 1984. The origin of anatomically modern humans in Western Europe. In: Smith FH, Spencer F, editors. *The origins of modern humans: a world survey of the fossil evidence*. New York: Liss. p 51–135.
- Stringer CB. 1992. Replacement, continuity and the origin of *Homo sapiens*. In: Bräuer G, Smith FH, editors. *Continuity or replacement: controversies in Homo sapiens evolution*. Rotterdam: A.A. Balkema. p 9–24.
- Stringer CB, Bräuer G. 1994. Methods, misreading, and bias. *Am Anthropol* 96:416–424.
- Stringer CB. 2001. Modern human origins: distinguishing the models. *Afr Archaeol Rev* 18:67–75.
- Stringer CB. 2002. Modern human origins: progress and prospects. *Philosoph Transact Royal Soc London (B)* 357:563–579.
- Thorne A, Grün R, Mortimer G, Spooner NA, Simpson JJ, McCulloch M, Taylor L, Curnoe D. 1999. Australia's oldest remains: age of the Lake Mungo 3 skeleton. *J Hum Evol* 36:591–612.
- Thorne A, Wolpoff MH. 2003. The multiregional evolution of humans. *Scientific American special edition: New look at Human Evolution*. p 46–53.
- Thorpe RS. 1984. Coding morphometric characters for constructing distance Wagner networks. *Evolution* 38:244–255.
- Tillier AM. 1999. *Les enfants moustériens de Qafzeh*. Paris: CNRS.
- Trinkaus E, Le May M. 1982. Occipital bunning among later Pleistocene hominids. *Am J Phys Anthropol* 57:27–35.
- Vandermeersch B. 1981. *Les hommes fossiles de Qafzeh (Israel)*. Paris: CNRS.
- Webb S. 1990. Cranial thickening in an Australian hominid as a possible paleoepidemiological indicator. *Am J Phys Anthropol* 82:365–406.
- Weidenreich F. 1951. *Morphology of solo man*. New York: American Museum of Natural History.
- Wolpoff MH, Wu X, Thorne AG. 1984. Modern *Homo sapiens* origins: a general theory of hominid evolution involving the fossil evidence from East Asia. In: Smith FH, Spencer F, editors. *The origins of modern humans: a world survey of the fossil evidence*. New York: Liss. p 411–483.
- Wolpoff MH. 1999. *Paleoanthropology*, 2nd ed. New York: McGraw-Hill.
- Wolpoff MH, Hawks J, Frayer DW, Hunley K. 2001. Modern human ancestry at the peripheries: a test of the replacement theory. *Science* 291:293–297.
- Wu X. 1998. Origin of modern humans of China viewed from cranio-dental characteristics of late *Homo sapiens* in China. *Acta Anthropol Sinica* 17:276–284.