Using extant morphological variation to understand fossil relationships: a cautionary tale

Rebecca Rogers Ackermann*

RECENT STUDIES OF VARIATION IN LIVING monkeys, apes, and humans have produced a number of insights that are pertinent to how we evaluate relationships among our fossil human ancestors. Here I summarize four such insights. I then use a fossil hominid example to illustrate how our understanding of variation can alter our interpretation of the past. Results show that our assessments of the relationships among fossil hominids can differ depending on which extant model of variation is used as a variation 'yardstick.' Additionally, our interpretations of these relationships can be swayed considerably by how we evaluate significance.

When living species are used as analogues for fossil ones, an assumption is commonly made that the fossil species and the living species vary in the same way. Implicit in this assumption is the idea that closely related living species themselves vary in the same way. Such assumptions underlie many analyses,[1–14] despite a growing understanding that they are inaccurate.[5,15,16]

Recent research has involved investigating exactly whether and how variation and covariation patterns differ among populations in two living clades — the New World tamarins (genus *Saguinus*) and the African great apes and humans.[16–18] Such understanding allows us to evaluate the utility of using surrogate models of variation and covariation for evaluating fossil relationships. To date, a number of patterns have emerged from this work that have some bearing on our understanding of how variation changes through time and across space, and by extension on how we interpret phylogenetic relationships in the fossil record.

In the first part of this paper, I will summarize a few of these patterns as 'lessons' learned from studying tamarins and hominoids. These lessons are drawn from three manuscripts, which can be referred to for further details of each analysis.[16–18] The lessons themselves are not new, and previous research has supported different aspects of them, particularly in the context of understanding sexual dimorphism.[7,14,19–25] However, a detailed analysis of patterns of variance and covariance *per se* provides important additional insight into such issues. Furthermore, as a unit the lessons appear contradictory, and it is useful to summarize them and consider their implications.

In the second part of this paper, I will employ one fossil example and one method to illustrate precisely how assumptions of variance/covariance (V/CV) equality can bias our understanding about phylogeny and phylogenetic relationships in important ways, some of which may be predictable or patterned, and some of which may not. The main thrust of the paper is this: just acknowledging the biases inherent in assuming variation equality across populations — and then getting on with the analysis — is *not good enough,* because different causes of variation inequality can have different and often quite profound effects on the results.

Lesson 1: Patterns of variation and covariation are not equal, even among closely related, morphologically homogeneous primate species.

When patterns of variation and covariation were compared across extant Old World and New World primate taxa (see Table 1 for sample information) using a variety of quantitative methods, they were generally found to be different. This was particularly striking for the tamarins, where we compared patterns of variation in rather large samples, across the entire genus *Saguinus*.[17,18] None of the species' V/CV matrices could be considered equal, despite the fact that tamarins are morphologically a relatively homogeneous genus. This also holds for the African ape/human clade; gorillas, chimps, bonobos, and humans do not have equal patterns of variation.[16]

Lesson 2: Patterns of variation and covariation are similar among related primate populations.

A well-known phenomenon when working with large samples is that small differences can be significant without being particularly meaningful. So, while patterns of variation are not equal among tamarins or African apes and humans, they are nonetheless similar. Within both the tamarin and African ape/human clades there are significant similarities among species patterns of allometry, among eigenvectors and among eigenvalues. The species within each clade also consistently share patterns of morphological correlation and integration.

Lesson 3: Differences in variation and covariation patterns may be tied to phylogenetic divergence in some instances.

Why are patterns of variation simultaneously different and similar? One possibility is that they are tied to phylogenetic divergence and, indeed, may themselves be indicators of phylogenetic similarity. In many instances within these two clades, differences among variation patterns correspond to the phylogenetic relationships based on mitochondrial DNA among the species, indicating that patterns of variation may have diverged through time in both clades. For chimps, humans, and gorillas, divergence of morphological variation patterns follows genetically derived phylogenies, but it is important to remember that this is a not very robust sample of three species. So, given that there are only two other alternative scenarios, the possibility of this outcome based on chance alone is 1/3.[16] In the tamarin clade, this pattern shows up at shallower nodes in the evolutionary tree, and can be correlated to some extent with the postulated biogeographic dispersal events. However, this relationship starts to fall apart when non-random processes such as selection are likely to have influenced morphology, indicating

Table 1. Extant samples used to estimate variation in New World tamarins and Old World great apes and humans (compiled from refs 13, 16–18). Samples used in this analysis are marked with an asterisk.

New World monkeys (tamarins)	Old World apes and humans
Saguinus fuscicollis (*n* = 289)	*Homo sapiens* — worldwide sample (*n* = 360)*
Saguinus geoffroyi (*n* = 132)	*Homo sapiens* — sub-Saharan African sample (*n* = 347)*
Saguinus midas (*n* = 116)	*Pan troglodytes* (*n* = 65)*
Saguinus mystax (*n* = 72)	*Pan paniscus* (*n* = 21)
Saguinus nigricollis (*n* = 59)	*Gorilla gorilla* (*n* = 117)*
Saguinus oedipus (*n* = 180)	

*Department of Archaeology, University of Cape Town, Private Bag, Rondebosch 7701, South Africa.
E-mail: becky@science.uct.ac.za

that variation patterns do not always diverge through time in a regular, even manner.[18] Interestingly, this suggests that morphological regions that are highly subject to non-random forces should be avoided. It also suggests that patterns of variation may themselves reflect evolutionary processes.

Lesson 4: Differences in patterns of variation and covariation may be the result of other phenomena.

Differences in variation patterns may reflect other things besides phylogeny. As already mentioned, there is the possible influence of non-random evolutionary processes. As with most studies, sample bias could also influence the interpretation of morphological similarity and difference. Additionally, we are comparing populations with different histories and with different patterns of sexual dimorphism. All of these can affect the patterning of variation and covariation. Furthermore it can be difficult to tease apart how these factors might contribute to the differences that are seen among living species. Certainly the models palaeoanthropologists most often use for evaluating fossil hominids are built from samples of great apes and human populations that undoubtedly have different histories and different patterns of dimorphism. Any similarities and differences perceived among the fossils are therefore influenced by properties inherent in the extant samples, or by artifacts of their sampling.

Applying the lessons: a fossil comparison

To provide an example of the implications of these lessons for the analysis of fossil data, I will focus on the relationships among three fossil members of the genus *Homo:* KNM-ER 1470, KNM-ER 1813, and KNM-ER 3733. The relationship between the two *Homo habilis* (*sensu lato*) specimens — KNM-ER 1470 and KNM-ER 1813 — has been controversial for years, and many of the seminal papers splitting these fossils into two species or lumping them together into one rely on models of extant variation drawn from living ape and human populations.[1,26,27] KNM-ER 3733 is a member of a taxon traditionally called 'African *Homo erectus*.'

In this analysis, patterns of variation from living species stand as surrogates for fossil variation. Fossils are compared using the Mahalanobis distance statistic (D^2) — a measure that is frequently used in the palaeoanthropological literature to assess morphological distance. This statistic is chosen for this example specifically because it requires an explicit statement about patterns of variance and covariance in the sample, and as such the assumption of variance/covariance equality between fossil and analogue populations is fundamental to its application. Because we know from the lessons that variation and covariation patterns are not equal across extant species (and by extension across extinct ones), we can test the effect that such inequality has on the value of, and evaluation of, this statistic. The Mahalanobis distance statistic is defined as follows:

$$D^2 = (x_1 - x_2)' \, V^{-1}(x_1 - x_2).$$

Here, x_1 is the vector of Euclidean distances for the first individual, x_2 the vector for the second, and V is the V/CV matrix of the extant model population. Mahalanobis distances between the vectors of 13 landmark-based Euclidean distances shared among the three fossils are calculated, using variation surrogates drawn from human (both worldwide and sub-Saharan African samples), chimp, and gorilla models of variation (Tables 1 and 2; see also ref. 13). The facial distances used are chosen based on what data are shared by the fossils, and focus in the mid-face region (Table 2). Following Darroch and Mosimann,[28] all Euclidean distance data are adjusted to reduce the effects of size by dividing each variable by the geometric mean of all variables for each individual. V/CV matrices of the extant populations are obtained for each hominoid species, using the residual covariance matrix from a MANOVA with the 13 traits as dependent variables and subspecific affiliation as the independent variable, thus pooling the covariances across subspecies. For further discussion on sample structure for this group of hominoids, refer to Ackermann[15]).

In order to evaluate whether differences between the fossil hominids are comparable to what we see in the extant analogue species, a frequency distribution of expected pairwise distances (D^2) is created for each extant species. A randomization is performed by drawing 1000 pairs with replacement from the extant populations, calculating the D^2 value for each pair, and producing a frequency distribution of Mahalanobis distances. Fossil distances are considered outside of the typical range of pairwise distances when they exceed 95% of the values in the extant randomized distribution(s). Importantly, the distances between the fossil hominids are evaluated in two ways. First, frequency distributions are calculated using extant pairs (x_1 and x_2, above) from one species and an assumed pattern of variation (V, above) from the same species — that is, human pairs with human variation/covariation — this is designated 'same-species' evaluation. This is the approach commonly used for evaluating significance of such morphological distance values. Second, frequency distributions are calculated using extant pairs from one species and an assumed pattern of variation from a *different* extant species — that is, chimp pairs with an assumption of a human-like pattern of variation, human pairs with gorilla variation, etc. — that is designated 'different-species' evaluation. This approach *assumes* that variation will be different between the evaluated pairs and the analogue species variation — that the fossils are drawn from species that vary differently than the extant species — and factors that assumption into the evaluation of the results. (This was not done for humans with sub-Saharan African variation, and vice versa, as one is by definition a subset of the other).

Two main results emerge from the fossil calculations. One is that the distance values change dramatically depending on which living species is used as a variation model (Table 3). For example, the distance between KNM-ER 1813 and KNM-ER 1470 is highest when a sub-Saharan African model is used, and lowest when a human model is used, with the

Table 2. Euclidean distance data used in the analysis. See also: refs 13, 16–18. Each distance is calculated from 3-D coordinate data collected on sutures and suture intersections on the facial skeleton.

Variable	Description	Position
NA-NSL	Distance between nasion and nasale	Midline
NA-ANS	Distance between nasion and anterior nasal spine	Midline
NA-IS	Distance between nasion and intradentale superior	Midline
NA-FMN	Distance between nasion and frontal-maxillary-nasal suture	Left, right
NSL-ANS	Distance between nasale and anterior nasal spine	Midline
NSL-IS	Distance between nasale and intradentale superior	Midline
NSL-FMN	Distance between nasale and frontal-maxillary-nasal suture	Left, right
ANS-IS	Distance between anterior nasal spine and intradentale superior	Midline
ANS-FMN	Distance between anterior nasal spine and frontal-maxillary-nasal suture	Left, right
IS-FMN	Distance between intradentale superior and frontal-maxillary-nasal suture	Left

chimp and gorilla V/CV models giving similar values to the human one. In the KNM-ER 1813 versus KNM-ER 3733 comparison, there is a similar pattern, though the ape-based values are more intermediate, while in the KNM-ER 1470 vs. KNM-ER 3733 comparison, the Mahalanobis distance values are strikingly low using *either Homo sapiens* model, and relatively high with the other two.

This first result — that distances among the fossils differ depending on which extant model of variation is used — is not too surprising, at least in principle, because we know from the first lesson that living populations (the extant models) vary in different ways. But even though palaeoanthropologists increasingly acknowledge that living substitutes for fossil variation are imperfect,[5,15,16] that the precise quantification of this using a relatively simple, widely used multivariate statistic can produce such varied results is, to say the least, disturbing.

The second result to emerge from this analysis is that the interpretation of the morphological distance between the fossil hominids changes depending on whether you base your evaluation on 'same-species' or 'different-species' frequency distributions. When a 'same-species' evaluation is performed, all the distance values among the fossils fall beyond the 95% level for all species evaluations, with one exception (KNM-ER 1470 vs. KNM-ER 3733 under a human model of variation — see below), which could be interpreted as support for separating the three into three distinct species (see Table 3). In other words, KNM-ER 1813 and KNM-ER 1470 are more different than one would expect two individuals from one species to be, when this species' distribution of D^2 values has been calculated in a traditional manner — that is, using variance/covariance estimates from the same population. This is also true of KNM-ER 1813 and KNM-ER 3733. In fact, nearly all of the distances are outside of the *range* of what is seen in the living populations (Table 4). There are two exceptions: KNM-ER 1470 vs. KNM-ER 3733 falls at the 85% in the human frequency distribution, and the 96% in the sub-Saharan African (but well out of the range for the great apes).

Table 3. Mahalanobis distances (D^2) between fossil pairs, using the four different extant models of variation. Notation indicates when the distance values are larger than 95% of the values in the extant distributions calculated using V/CV from the same species (†) and from *all* of the different species (*).

	H. sapiens (WW) variation	*H. sapiens* (SSA) variation	*P. troglodytes* variation	*G. gorilla* variation
KNM-ER 1813 vs. 1470	244 †*	816 †*	268 †	255 †
KNM-ER 1813 vs. 3733	227 †*	777 †*	443 †	457 †
KNM-ER 1470 vs. 3733	46	57 †	660 †	504 †

However, when a 'different-species' evaluation is done, the results are quite different, and much more varied. Only distances between KNM-ER 1813 and each of the other two fossils fall beyond the 95% levels. This is the case when they are evaluated using human and sub-Saharan models of variation. In other words, KNM-ER 1813 and KNM-ER 1470 are more different than one would expect two individuals from the same (non-human) species to be using a surrogate model of variation/covariation from human populations. The same is true for the KNM-ER 1813 vs. KNM-ER 3733 comparison. Additionally, for both of these comparisons, while the distance values are within the range of what would be expected (below 95%) when human pairwise distances are generated using ape variation, or ape pairwise distances using human variation, they fall beyond the ninety-fifth percentile (but in some cases still within the range) of gorilla pairs evaluated using chimp variation or vice versa (Table 4). Interestingly, the distances between KNM-ER 3733 and KNM-ER 1470 were well within the range of what could be expected if these two specimens were from the same species, using ape pairs and human models of variation, and human pairs with ape models of variation, but not ape with ape (Table 4). These 'different-species' evaluations of the fossil distance provide more nuanced results than the 'same-species' evaluations, and might offer some insight into *how* the fossil pairs vary from each other, and which extant combinations might be more appropriate models.

The broader implications of this second result — that our interpretations can be swayed so drastically by how we choose to evaluate the results — are more alarming than the first. Too often standard evaluations (in the form of significance tests) are applied without serious thought being given to what they are implying. With the 'same-species' evaluation, one both assumes similar variation, and evaluates differences based on this similarity of variation. With the 'different-species' evaluation, one acknowledges that variation can differ between populations, and builds that into the evaluation of the results. In this example, the only proper evaluation can be a 'different-species' one, since we know that the fossils are not modern humans, or gorillas, or chimps, but rather something else.

So why does a change in assumed variation pattern alter the results so much in the first place? Probably for all of the reasons highlighted in the 'lessons.' Humans, gorillas, chimps, and hominids all vary in different ways, largely because they all have different evolutionary histories; no matter how much you try to make the sampling unbiased, it is still biased by those histories. And while we expect that some extant models will 'fit' fossils better than others, especially when the model and fossil species are closer phylogenetically (although this is clearly not the only

Table 4. Statistics of the frequency distributions of D^2 in living populations. For a description of same-species and different-species evaluations, see text. Only 5% of the values in each distribution falls beyond the '95% value'.

	Same-species evaluation				Different-species evaluation									
	Human V/CV	Sub-Saharan Africa V/CV	Chimp V/CV	Gorilla V/CV	Human V/CV		SS Africa V/CV		Chimp V/CV			Gorilla V/CV		
	Human pairs	SS Africa pairs	Chimp pairs	Gorilla pairs	Chimp pairs	Gorilla pairs	Chimp pairs	Gorilla pairs	Human pairs	SS Afr. pairs	Gorilla pairs	Human pairs	SS Afr. pairs	Chimp pairs
95% value	64	45	51	56	116	154	325	320	1696	757	137	2788	899	170
Median	27	24	24	24	38	44	77	78	377	201	36	531	270	42
Mean	31	25	26	27	49	57	111	111	552	271	53	873	349	57
Maximum	134	89	69	90	389	389	1130	1154	2786	1713	490	6263	2113	291

reason this could happen), the match will inevitably be imperfect (see also ref. 5).

In order to mitigate the effects of these biases and imperfections, researchers need to build awareness of how our methods and samples *explicitly* affect our interpretations. Simply using the most 'variable' group as a conservative standard is not good enough. An approach taken by some researchers has been to use the gorilla as a conservative model for assessing fossil differences.[1,3,4] But the results of the present analysis show that when the assessment of three hominid fossils is based on a gorilla 'same-species' evaluation they are split into at least two different species, while with a gorilla 'different-species' evaluation, they can be lumped into one. This throws some doubt on the effectiveness of this 'conservative' approach. Similarly, acknowledging the biases inherent in assuming variation equality across populations — and then getting on with the analysis — is also not good enough, because, as has been shown here, different causes of variation inequality can have different and often quite profound effects on the results. The interpretation of phylogenetic relationships in the fossil record is confounded by a lack of understanding of how variation changes through time and space, due to both random and non-random processes. For instance, there are differences in variation patterns between populations due to phylogenetic distance, due to evolutionary pressures, as a result of population structure, sample bias, and so on, and each of these can influence results and subsequent interpretations in different ways. Incorporating approaches that assume divergence in patterns of population variation into our models can alter our understanding of human evolution.

Finally, while this paper made a specific fossil comparison using a specific method, its implications have broader relevance. Clearly, when using a statistic such as the Mahalanobis distance statistic, which explicitly requires an assumed population variation and covariation pattern, divergence in variation and covariation patterns between species can affect the conclusions of an analysis. But even when the statistic involved does not require such assumptions, there is often still an assumption that fossil and living (or even living and living) species vary in the same way. In the simplest scenario, this analysis suggests that it is unclear whether even a single analysed fossil variable can be expected to be as variable (or more or less so) as its counterpart in the comparative sample. In order to understand better the complex morphologies we see in the fossil record, it is essential to ground our fossil comparisons within a firm understanding of how living morphologies vary.

An earlier version of this paper, presented at the 2002 American Association of Physical Anthropologists (AAPA) annual meeting in Buffalo, New York, was titled 'What can morphological variation tell us about phylogenetic divergence?' As co-organizer of this symposium, I thank my colleagues who participated in it, and one anonymous reviewer. In addition, special thanks go to Sheela Athreya for her assistance and support during the AAPA meeting, and to Meave Leakey and the National Museums of Kenya for allowing access to fossil material in their care.

1. Lieberman D.E. and Pilbeam D.R. (1988). A probabilistic approach to the problem of sexual dimorphism in *Homo habilis*: a comparison of KNM-ER 1470 and KNM-ER 1813. *J. Hum. Evol.* **17**, 503–511.
2. McHenry H.M. (1991). Sexual dimorphism in *Australopithecus afarensis. J. Hum. Evol.* **20**, 21–32.
3. Miller J.A. (1991). Does brain size variability provide evidence of multiple species in *Homo habilis*? *Am. J. Phys. Anthropol.* **84**, 385–398.
4. Miller J.M.A. (2000). Craniofacial variation in *Homo habilis:* an analysis of the evidence for multiple species. *Am. J. Phys. Anthropol.* **112**, 103–128.
5. Wood B. (1991). A palaeontological model for determining the limits of early hominid taxonomic variability. *Palaeontologia afr.* **28**, 69–75.
6. Wood B. (1992). Early hominid species and speciation. *J. Hum. Evol.* **22**, 351–365.
7. Wood B.A., Li Y. and Willoughby C. (1991). Intraspecific variation and sexual dimorphism in cranial and dental variables among higher primates and their bearing on the hominid fossil record. *J. Anat.* **174**, 185–205.
8. Grine F.E., Demes B., Jungers W.L. and Cole T.M. (1993). Taxonomic affinity of the early *Homo* cranium from Swartkrans, South Africa. *Am. J. Phys. Anthropol.* **92**, 411–426.
9. Rightmire G.P. (1993). Variation among early *Homo* crania from Olduvai Gorge and the Koobi Fora region. *Am. J. Phys. Anthropol.* **90**, 1–33.
10. Kramer A., Donnelly D., Kidder J., Ousley D. and Olah S. (1995). Craniometric variation in large-bodied hominoids: testing the single species hypothesis for *Homo habilis. J. Hum. Evol.* **29**, 443–462.
11. Richmond B.G. and Jungers W.L. (1995). Size variation and sexual dimorphism in *Australopithecus afarensis* and living hominoids. *J. Hum. Evol.* **29**, 229–245.
12. Lockwood C.A., Richmond B.G, Jungers W.L. and Kimbel W.H. (1996). Randomization procedures and sexual dimorphism in *Australopithecus afarensis. J. Hum. Evol.* **31**, 537–548.
13. Ackermann R.R. (1998). *A quantitative assessment of variability in the australopithecine, human, chimpanzee, and gorilla face.* Ph.D. thesis, Washington University, St Louis, MO.
14. Lockwood C.A. (1999). Sexual dimorphism in the face of *Australopithecus africanus. Am. J. Phys. Anthropol.* **108**, 97–127.
15. Collard M. and Wood B. (2000). How reliable are human phylogenetic hypotheses? *Proc. Natl Acad. Sci. USA* **97**(9), 5003–5006.
16. Ackermann R.R. (2002). Patterns of covariation in the hominoid craniofacial skeleton: implications for palaeoanthropological models. *J. Hum. Evol.* **42**, 167–187.
17. Ackermann R.R. and Cheverud J.M. (2000). Phenotypic covariance structure in tamarins (genus: *Saguinus*): a comparison of variation patterns using matrix correlation and common principal component analysis. *Am. J. Phys. Anthropol.* **111**, 489–501.
18. Ackermann R.R. and Cheverud J.M.(2002). Discerning evolutionary processes in patterns of tamarin (genus *Saguinus*) craniofacial variation. *Am. J. Phys. Anthropol.* **117**, 260–271.
19. Oxnard C. (1987). *Fossils, Teeth and Sex: New perspectives on human evolution.* University of Washington Press, Seattle.
20. O'Higgins P., Chadfield P. and N. Jones (2001). Facial growth and the ontogeny of morphological variation within and between the primates *Cebus apella* and *Cercocebus torquatus. J. Zool. Lond.* **254**, 337–357.
21. O'Higgins P., Moore W.J., Johnson D.R. and McAndrew T.J. (1990). Patterns of cranial sexual dimorphism in certain groups of extant hominoids. *J. Zool. Lond.* **222**, 399–420.
22. O'Higgins P. and Dryden I. (1993). Sexual dimorphism in hominoids: further studies of craniofacial shape differences in *Pan, Gorilla,* and *Pongo. J. Hum. Evol.* 24, 183–205.
23. Plavcan J.M. (2001). Sexual dimorphism in primate evolution. *Yrbk Phys Anthropol.* **44**, 25–53.
24. Plavcan J. (2002). Taxonomic variation in patterns of craniofacial dimorphism in primates. *J. Hum. Evol.* **42**, 579–608.
25. Plavcan J. (2003). Scaling relationships between craniofacial sexual dimorphism and body mass dimorphism in primates: implications for the fossil record. *Am. J. Phys. Anthropol.* **120**, 38–60.
26. Stringer C.B. (1986). The credibility of *Homo habilis.* In *Major Topics in Primate and Human Evolution,* eds B. Wood, L. Martin, and P. Andrews, pp. 266–294. Cambridge University Press, Cambridge.
27. Wood B.A. (1985). Early *Homo* in Kenya, and its systematic relationships. In *Ancestors: The Hard Evidence,* ed. E. Delson, pp. 206–214. Alan R. Liss, New York.
28. Darroch J N. and Mosimann J.E. (1985). Canonical and principle components of shape. *Biometrika* **72**, 241–252.