

An assessment of the likely impact of strain-related phenotypic plasticity on hominin fossil species identification

M. Collard^{a*} and S.J. Lycett^b

It has been proposed that strain-related phenotypic plasticity may be a major confounding factor in attributing hominin fossils to species. The study reported here tested this hypothesis with craniometric data from the great apes and *Colobus guereza*. We divided the measurements into three groups: measurements of features subject to high masticatory strain, measurements of features subject to low-to-moderate masticatory strain, and measurements of features that do not remodel and therefore are not prone to strain-related phenotypic plasticity. Next, we used the coefficient of variation and ANOVA to investigate whether masticatory strain is a cause of variability. These analyses partially supported the hypothesis. The predicted differences between the high-strain measurements and the other measurements were found in the majority of the species. However, the coefficient of variation values for the low-to-moderate strain and non-phenotypically plastic measurements were indistinguishable. Thereafter, we used discriminant function analysis to compare the ability of the three groups of measurements to assign specimens to species. This analysis did not support the hypothesis. The high-strain measurements were less effective than the other measurements, but the low-to-moderate strain measurements were more effective than the non-phenotypically plastic measurements. In addition, better discrimination was achieved when all the measurements were employed than when just the non-phenotypically plastic measurements were utilised. We conclude from this that strain-related phenotypic plasticity is unlikely to impede hominin alpha taxonomic research.

Key words: phenotypic plasticity, hominin, taxonomy, species identification, strain, mastication, hominoid

Introduction

Phenotypic plasticity is the expression by a genotype of different phenotypes in response to different environmental conditions.¹ Recently, Wood and Lieberman² hypothesised that this phenomenon may negatively impact efforts to delineate species in the hominin fossil record. Wood and Lieberman's² suggestion is based on work examining how mechanical loading affects bone. This work demonstrates that interactions between the skeleton and its mechanical environment can greatly influence bone size and shape.³⁻⁷ For example, mechanical loading experienced during development has been found to affect both the growth of cortical bone in diaphyses and the growth of trabecular bone in epiphyses.⁸ Likewise, studies of individuals experiencing lower than normal mechanical strains (e.g. those following denervation, bed-rest or exposure to gravity-free

environments) indicate that bone may resorb at rapid rates in many regions of the skeleton.⁹ One corollary of these observations, argued Wood and Lieberman,² is that skeletal features prone to strain-related phenotypic plasticity can be expected to be particularly poor indicators of species diversity because much of the variation they exhibit is likely to be related to environmental factors rather than a reflection of genetically-determined patterns.

Wood and Lieberman² tested their hypothesis with measurement data recorded on crania, mandibles and teeth of four hominoid species, *Gorilla gorilla*, *Homo sapiens*, *Pan troglodytes* and *Pongo pygmaeus*, and one cercopithecoïd species, *Colobus guereza*. They divided the measurements into three groups based on their likelihood of exhibiting phenotypic plasticity, and then computed the coefficient of variation of each measurement for each species. Subsequently, they compared the mean coefficient of variations for the three groups of measurements on a species-by-species basis. Wood and Lieberman's² analyses partially supported their hypothesis. The prediction that measurements with a moderate probability of exhibiting phenotypic plasticity should be more variable than measurements with a low probability of exhibiting phenotypic plasticity was fulfilled in all five cases. However, the prediction that measurements with a high probability of exhibiting phenotypic plasticity should be more variable than measurements with a moderate probability of exhibiting phenotypic plasticity and measurements with a low probability of exhibiting phenotypic plasticity was fulfilled in only two cases. In the other three cases, the average coefficient of variation of the high probability measurements was significantly higher than either the average coefficient of variation of the moderate probability measurements or the average coefficient of variation of the low probability measurements, but not both. Despite the fact that the analyses only partially supported the predictions of their hypothesis, Wood and Lieberman² concluded that there was reason for palaeoanthropologists to avoid phenotypically plastic features when assigning fossil hominin specimens to species.

Recently, Collard and Lycett¹⁰ have reported a study in which they used craniometric data recorded on seven Old World monkey species to test Wood and Lieberman's² hypothesis that characters exhibiting relatively high levels of phenotypic plasticity are less reliable for taxonomic discrimination than characters exhibiting lower levels of phenotypic plasticity. Collard and Lycett¹⁰ carried out their study because Wood and Lieberman² did not explicitly examine the relationship between phenotypic plasticity and taxonomic utility; they only dealt with the relationship between strain and phenotypic plasticity. Collard and Lycett¹⁰ used the coefficient of variation and one-way analysis of variance (ANOVA) to compare levels of phenotypic plasticity in three groups of craniodental measurements: measurements of features that are subject to high levels of strain during mastication, measurements of features that experience low-to-moderate levels of strain during mastication, and measurements of teeth, which do not remodel and therefore are not prone to strain-related phenotypic plasticity. Thereafter, Collard and Lycett¹⁰ used discriminant function analysis to compare the ability of the three groups of measurements to correctly assign specimens by species. For comparative purposes, they also carried out a discriminant function analysis with all the measurements. Collard and Lycett's¹⁰ coefficient of variation analysis partially supported Wood and Lieberman's² hypothesis. The predicted differences between the high-strain measurements and the low-to-moderate strain measurements were found to be both present and statistically significant in all seven species, as was the predicted differences between the

^aLaboratory of Human Evolutionary Studies, Department of Archaeology, Simon Fraser University, 8888 University Drive, Burnaby, British Columbia, V5A 1S6, Canada.

^bDepartment of Anthropology, University of Kent at Canterbury, U.K.

*Author for correspondence E-mail: mark.collard@sfu.ca

Table 1. Measurements employed in this study.

High strain	Low-to-moderate strain	Non-phenotypically plastic
Condylar height	Right orbital breadth	I ¹ labiolingual diameter
Coronoid height	Right orbital height	I ² labiolingual diameter
Right condylar head width	Interorbital breadth	Upper C buccolingual diameter
Right condylar head anterior–posterior breadth	Biorbital breadth	M ¹ buccolingual diameter
Ramal breadth	Nasion–Rhinion	M ² buccolingual diameter
Bigonial width	Nasion–nasospinale	M ³ buccolingual diameter
Height of mandibular body at M ₁	Glabella–Opisthocranion	I ₁ labiolingual diameter
Thickness of mandibular body of M ₁	Basion–Bregma	I ₂ labiolingual diameter
Symphyseal height	Maximum bi-parietal breadth	Lower C buccolingual diameter
Symphyseal thickness	Biporionic width	M ₁ buccolingual diameter
Inner alveolar breadth at M ₃	Coronale–Coronale	M ₂ buccolingual diameter
Inter-lower canine distance	Posterior skull length	M ₃ buccolingual diameter

high strain measurements and the non-phenotypically plastic measurements. However, the coefficient of variation values for the low-to-moderate strain measurements and the non-phenotypically plastic measurements were statistically indistinguishable.

Collard and Lycett's¹⁰ discriminant function analysis did not support Wood and Lieberman's² hypothesis. The high-strain measurements correctly classified 94.9% of specimens, the low-to-moderate strain measurements 97.5% and the non-phenotypically plastic measurements 97.4%. When all 60 craniodental measurements were included, 100% of specimens were correctly classified. Thus, the high strain measurements were less effective than the other two sets of measurements, as Wood and Lieberman's² hypothesis predicts. However, not only were the low-to-moderate strain measurements more effective than the non-phenotypically plastic measurements, but also better discrimination was achieved when all measurements were employed than when the non-phenotypically plastic measurements were utilised. Based on these results, Collard and Lycett¹⁰ concluded that, while phenotypic plasticity likely contributes to the variation observable in the hominin fossil record, controlling for phenotypic plasticity is probably an unnecessary course of action for researchers attempting to group fossil hominin specimens into species.

Here, we describe a second study that explicitly tests Wood and Lieberman's² hypothesis. The study employed measurement data from the five species that were used by Wood and Lieberman,² and was carried out to counter the main potential criticism of Collard and Lycett's¹⁰ study, namely that Wood and Lieberman's² hypothesis was framed in relation to fossil hominins and the taxa used by Collard and Lycett,¹⁰ the Old World monkeys, are too distantly related to hominins to represent a reasonable test case.

Materials and methods

The dataset comprised values for 36 measurements recorded on 37 *Gorilla gorilla* (20 males, 17 females), 75 *Homo sapiens* (40 males, 35 females), 35 *Pan troglodytes* (13 males, 22 females), 41 *Pongo pygmaeus* (20 males, 21 females) and 24 *Colobus guereza* (12 males, 12 females). The measurements were selected on the basis of current knowledge of their likely propensity to exhibit phenotypic plasticity as a result of mastication-related strain. Twelve were dental measurements. Since dental enamel does not remodel, these were designated non-phenotypically plastic characters. Labiolingual and buccolingual tooth crown dimensions were used in order to avoid the confounding effect of interstitial wear.¹¹ The other 24 measurements were cranial and mandibular measurements. Twelve of them were included because they relate to features that strain-gauge analyses indi-

cate experience strain of at least 1 000 $\mu\epsilon$ during mastication^{12–15}. These measurements were designated high-strain measurements, because strain in the order of 1 000 $\mu\epsilon$ is known to be capable of inducing bone growth.⁹ The remaining 12 measurements were included because strain-gauge studies indicate they are associated with features of the primate skull that experience strain of less than 1 000 $\mu\epsilon$ during mastication. These measurements were designated low-to-moderate strain measurements. Further details of the measurements are given in Table 1. The data were taken from Wood *et al.*¹⁶ This was also the source for the data used by Wood and Lieberman.² Thus, our dataset overlaps with the one they used. The cranial and mandibular measurements were rounded up to the nearest 1 mm, and the dental measurements to the nearest 0.1 mm.

We employed the dataset in two sets of analyses. These were similar to the ones carried out by Collard and Lycett.¹⁰ Thus, the first set of analyses focused on Wood and Lieberman's² suggestion that strain increases phenotypic variability. We used the coefficient of variation to assess phenotypic variability. A coefficient of variation was determined for each measurement, and then the mean coefficient of variation for each group of measurements was computed. Thereafter, ANOVA with post-hoc least significant difference pairwise comparisons was employed to test for statistically-significant differences among the mean coefficient of variation of the three groups of measurements ($\alpha \leq 0.05$). Since ANOVA assumes data are normally distributed,¹⁷ the coefficient of variation was logarithmically transformed ($\log e$) prior to analysis. We reasoned that, if Wood and Lieberman's² hypothesis is correct, the coefficient of variation for the high-strain measurements should be significantly higher than the coefficient of variation for the low-to-moderate strain measurements, and the coefficient of variation for the latter should be significantly higher than the coefficient of variation for the non-phenotypically plastic measurements. The second set of analyses focused on the second part of Wood and Lieberman's² hypothesis, that is the idea that phenotypic plasticity impedes species identification. This was accomplished by separately subjecting the three groups of measurements to discriminant function analysis. The form of discriminant function analysis employed separates groups on the basis of canonical discriminant functions.¹⁸ In addition, we carried out a discriminant function analysis in which all the measurements were included. We reasoned that, if Wood and Lieberman's² hypothesis is correct, the high-strain measurements should be worse at assigning specimens to species than the low-to-moderate strain measurements, and the latter should be worse at assigning specimens to species than the non-phenotypically plastic measurements. We also reasoned that, if Wood and Lieberman's² hypothesis is correct, the analysis in which all the measurements

were used should return more misclassified specimens than the analysis in which only the non-phenotypically plastic measurements were employed. Because Wood and Lieberman's² hypothesis is focused on fossil hominins, and fossil hominin specimens can rarely be sexed with confidence, we employed mixed sex samples in both analyses. The analyses were carried out in SPSS 12.0.1; the program's stepwise insertion option was used to conduct the discriminant function analysis.

Results

Table 2 shows the mean coefficient of variation of each variable group for each taxon as well as the results of the least significant difference pairwise comparisons following a one-way ANOVA. The mean coefficient of variation for the high-strain measurements was consistently higher than the mean coefficient of variation for the low-to-moderate strain measurements and for the non-phenotypically plastic measurements. However, the differences were only significant in *H. sapiens*, *P. troglodytes*, *P. pygmaeus* and *C. guereza*. In *G. gorilla* the high-strain measurements were more variable than the low-to-moderate strain measurements and the non-phenotypically plastic measurements, but the differences were not statistically significant at the 0.05 level. The low-to-moderate strain measurements were not significantly more variable than the non-phenotypically plastic measurements in any of the species. They were more variable than the non-phenotypically plastic measurements in *G. gorilla*, *P. troglodytes*, *P. pygmaeus* and *C. guereza*, but none of the differences were statistically significant. In the case of *H. sapiens* the non-phenotypically plastic measurements were more variable than the low-to-moderate strain measurements although the difference was not significant. Thus, the results of the coefficient of variation analysis were only partially consistent with the prediction that the high strain measurements should exhibit greater phenotypic plasticity than the low-to-moderate strain measurements, and that the latter should exhibit greater phenotypic plasticity than the non-phenotypically plastic measurements.

The first four discriminant functions were used in the discriminant function analyses. Plots of the first two discriminant functions from each analysis are presented in Figs 1-4. The high-strain measurements correctly classified 97.6% of specimens, the low-to-moderate strain measurements 100% and the non-phenotypically plastic measurements 98.6%. When all the measurements were analysed together, 100% of the specimens were correctly classified. Thus, the results of the discriminant function analysis did not support Wood and Lieberman's² hypothesis. The high-strain measurements were less effective at assigning specimens to species than the other two groups of measurements, as predicted. However, contrary to expectation, the low-to-moderate strain measurements were more effective at assigning specimens to species than the non-phenotypically plastic measurements. Also contrary to expectation, more specimens were correctly assigned to species when all the measurements were used than when the non-phenotypically plastic measurements were employed.

Discussion and conclusions

The results of the coefficient of variation analysis reported here were comparable to the results of the coefficient of variation analysis carried out by Collard and Lycett.¹⁰ As discussed earlier, the latter only partially supported Wood and Lieberman's² suggestion that strain increases variability. The predicted differences between the high-strain measurements and the low-to-moderate strain measurements were found to be both present

Table 2. Mean coefficient of variation values and results of one-way ANOVA.

Taxon	Coefficient of variation			ANOVA		
	High ¹	LM ²	NPP ³	High vs LM	High vs Dental	LM vs NPP
<i>Gorilla</i>	12.57	10.37	9.95	0.079	0.061	0.903
<i>Homo</i>	10.82	6.63	7.30	<0.001***	0.007**	0.196
<i>Pan</i>	9.12	7.04	7.03	0.021*	0.031*	0.861
<i>Pongo</i>	15.24	10.91	10.14	0.003***	0.004***	0.953
<i>Colobus</i>	9.66	6.79	6.65	0.012*	0.008**	0.878

* = significant at $P \leq 0.05$; ** = significant at $P \leq 0.01$; *** = significant at $P \leq 0.005$.

¹High = high-strain measurements. ²LM = low-to-moderate strain measurements; ³NPP = non-phenotypically plastic measurements.

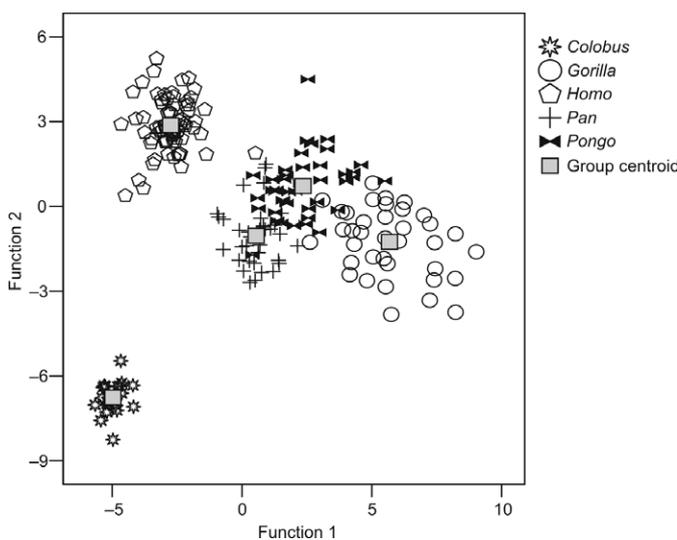


Fig. 1. Plot of the first two discriminant functions from discriminant function analysis of high-strain measurements. 97.6% of specimens were correctly classified to species. 2.4% of *Gorilla gorilla* and 7.3% of *Pan troglodytes* were misclassified as *Pongo pygmaeus*.

and statistically significant in all seven species, as were the predicted differences between the high-strain measurements and the non-phenotypically plastic measurements. However, the coefficient of variation values for the low-to-moderate strain measurements and the non-phenotypically plastic measurements were statistically indistinguishable. The results of the coefficient of variation analysis reported here also only partially supported Wood and Lieberman's² hypothesis. The predicted

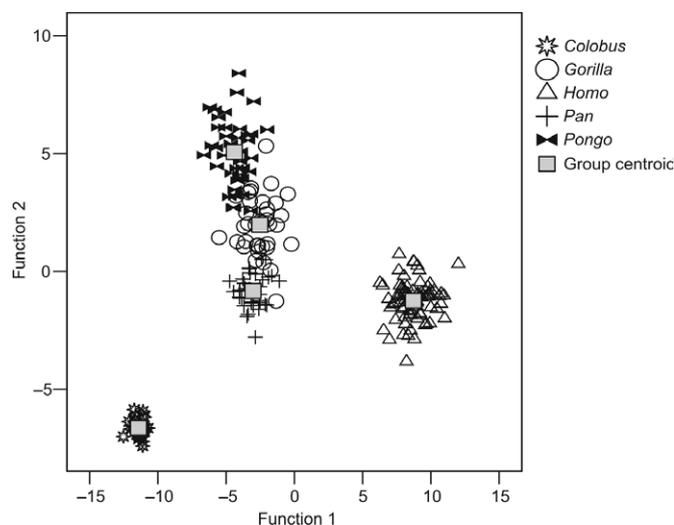


Fig. 2. Plot of the first two discriminant functions from discriminant function analysis of low-to-moderate strain measurements; 100% of specimens were correctly classified to species.

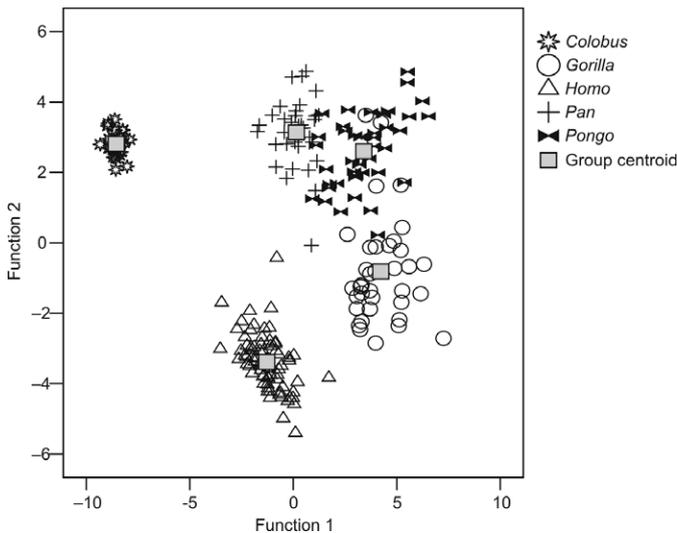


Fig. 3. Plot of the first two discriminant functions from discriminant function analysis of non-phenotypically plastic measurements. 98.6% of specimens were correctly classified to species. 2.4% of *Gorilla gorilla* and 2.4% of *Pan troglodytes* were misclassified as *Pongo pygmaeus*.

differences between the high-strain measurements on the one hand, and the low-to-moderate strain and non-phenotypically plastic measurements on the other, were found to be both present and statistically significant in only four of the five species. In the fifth species, *G. gorilla*, the predicted differences were present but did not reach statistical significance. In addition, the coefficient of variation values for the low-to-moderate strain measurements and the non-phenotypically plastic measurements were statistically indistinguishable.

The results of the discriminant function analysis reported here were also comparable to the discriminant function analysis results obtained by Collard and Lycett.¹⁰ To reiterate, the latter did not support Wood and Lieberman's² hypothesis because the low-to-moderate strain measurement were more effective than the non-phenotypically plastic measurements and the best discrimination was achieved when phenotypic plasticity was ignored and all measurements were employed. The discriminant function analysis reported here did not support Wood and Lieberman's² hypothesis for the same reasons. The high-strain measurements were less effective at assigning specimens to species than either the low-to-moderate strain or non-phenotypically plastic measurements, but the low-to-moderate strain measurements were more effective at allocating specimens to species than the non-phenotypically plastic measurements. In addition, better discrimination was achieved when all the measurements were employed than when the non-phenotypically plastic measurements were utilised.

Taken together, the results of this study and those obtained by Collard and Lycett¹⁰ provide partial support for Wood and Lieberman's² suggestion that mastication-related strain is capable of significantly increasing intra-specific variability in the bones of the hominin skull. However, they do not support Wood and Lieberman's² other suggestion—that strain-related phenotypic plasticity negatively impacts attempts to delineate species in the hominin fossil record. The studies indicate that discriminating among features of the cranium, mandible and dentition on the basis of their likelihood of exhibiting phenotypic plasticity not only does not automatically lead to significantly more reliable taxonomic hypotheses, but may in fact lead to less reliable taxonomic hypotheses. In our view, this finding argues against palaeoanthropologists adopting the course of action proposed by Wood and Lieberman.²

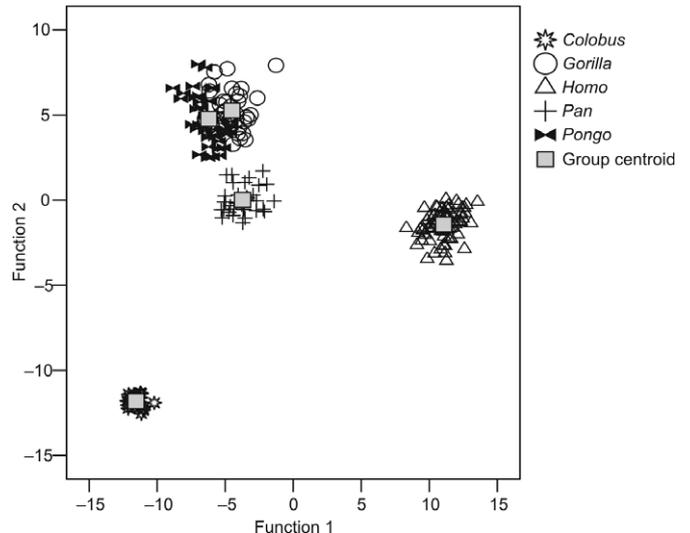


Fig. 4. Plot of the first two discriminant functions from discriminant function analysis of all measurements; 100% of specimens were correctly classified to species.

With regard to alternative strategies for improving understanding of the species-level diversity of the fossil hominins, it is noteworthy that in both this study and Collard and Lycett's¹⁰ study complete discrimination was achieved in the discriminant function analysis when all measurements were included. This suggests that simply maximising the number of characters may be an effective strategy to employ when attempting to delineate species in the hominin fossil record. Evaluating this possibility would seem to be an obvious next step. Given the fragmentary state of most hominin fossil specimens, key issues to address are how many characters it is necessary to use in order to be confident that specimens are correctly allocated to species, and whether that number is dependent on the skeletal element employed.

M.C. is supported by the Social Sciences and Humanities Research Council, the Canada Research Chairs Program, the Canada Foundation for Innovation, the British Columbia Knowledge Development Fund, and Simon Fraser University. S.J.L. is supported by the British Academy Centenary Research Project, *Lucy to Language*.

Received 5 May. Accepted 18 June 2009.

1. Reidl R.J. (1978). *Order in Living Organisms*. Wiley, New York.
2. Wood B.A. and Lieberman D.E. (2001). Craniodental variation in *Paranthropus boisei*: a developmental and functional perspective. *Am. J. Phys. Anthropol.* **116**, 13–25.
3. Lanyon L.E. and Rubin C.T. (1985). Functional adaptation in skeletal structures. In *Functional Vertebrate Morphology*, eds M. Hildebrand, D.M. Bramble, K.E. Liem and D.B. Wake, pp. 1–25. Harvard University Press, Cambridge, MA.
4. Frost H.M. (1998). From Wolff's law to the mechanostat: a new 'face' of physiology. *J. Ortho. Sci.* **3**, 282–286.
5. Herring S.W. (1993). Epigenetic and functional influences on skull growth. In *The Skull*, vol. 1, eds J. Hanken and B.K. Hall, pp. 153–206. Chicago University Press, Chicago.
6. Lieberman D.E. (1997). Making behavioral and phylogenetic inferences from hominid fossils: considering the developmental influence of mechanical forces. *Ann. Rev. Anthropol.* **26**, 185–210.
7. Skerry T. (2000). Biomechanical influences on skeletal growth and development. In *Development, Growth and Evolution: Implications for the Study of the Hominid Skeleton*, eds P. O'Higgins and M.J. Cohn, pp. 29–39. Academic Press, London.
8. Lieberman D.E. and Crompton A.W. (1998). Responses of bone to stress. In *Principles of Biological Design: the Optimization and Symmorphosis Debate*, eds E. Wiebel, C. Taylor and L. Bolis, pp. 78–86. Cambridge University Press, Cambridge.
9. Martin R.B., Burr D.B. and Sharkey N.A. (1998). *Skeletal Tissue Mechanics*. Springer, New York.
10. Collard M. and Lycett S.J. (2008). Does phenotypic plasticity confound attempts to identify hominin fossil species? An assessment using extant Old

- World monkey craniodental data. *Fol. Primatol.* **79**, 111–122.
11. Hinton R.J. (1982). Differences in interproximal and occlusal tooth-wear among prehistoric Tennessee Indians: implications for masticatory function. *Am. J. Phys. Anthropol.* **57**, 103–115.
 12. Hylander W.L., Picq P.G. and Johnson K.R. (1991). Masticatory stress hypotheses and the supraorbital region of primates. *Am. J. Phys. Anthropol.* **86**, 1–36.
 13. Hylander W.L. and Johnson K.R. (1994). Jaw muscle function and wishboning of the mandible during mastication in macaques and baboons. *Am. J. Phys. Anthropol.* **94**, 523–547.
 14. Hylander W.L. and Johnson K.R. (1997). In vivo bone strain patterns in the zygomatic arch of Macaques and the significance of these patterns for functional interpretations of craniofacial form. *Am. J. Phys. Anthropol.* **102**, 203–232.
 15. Ross C.E. (2001). *In vivo* function of the craniofacial haft: the interorbital ‘pillar’. *Am. J. Phys. Anthropol.* **116**, 108–139.
 16. 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.
 17. Sokal R.R. and Rohlf F.J. (1995). *Biometry*. Freeman, New York.
 18. Huberty C. (1994). *Applied Discriminant Analysis*. Wiley, New York.