Author's personal copy

Journal of Human Evolution 60 (2011) 240–243

Contents lists available at ScienceDirect

Journal of Human Evolution

journal homepage: www.elsevier.com/locate/jhevol

News and Views

Morphometrics and phylogenetics: Principal components of shape from cranial modules are neither appropriate nor effective cladistic characters

D.C. Adams [a,*], A. Cardini [b,c], L.R. Monteiro [b,d], P. O'Higgins [b], F.J. Rohlf [e]

[a] Department of Ecology, Evolution, and Organismal Biology, and Department of Statistics, Iowa State University, Ames, IA 50011 USA
[b] Anatomical and Human Sciences, Hull York Medical School, The University of York, York Y010 5DD, UK
[c] Museo di Paleobiologia e dell'Orto Botanico, Universitá di Modena e Reggio Emilia, via Università 4, 41100, Modena, Italy
[d] Department of Biological Sciences, The University of Hull, Hull HU6 7RX, UK
[e] Department of Ecology and Evolution, Stony Brook University, Stony Brook, NY 11794 USA

ARTICLE INFO

Article history:
Received 26 November 2008
Accepted 21 May 2009

Keywords:
Geometric morphometrics
Human origins

Introduction

Recently, González-José et al. (2008) proposed that phylogeny estimation using principal components (PCs) from shape analyses of cranial 'modules' (modular cladistic approach: MCL) is an advance that could provide a benchmark for future studies of evolutionary relationships. Unfortunately, MCL ignores serious practical and theoretical issues concerning the use of PCs as characters in cladistic analyses, and there is no evidence that the method is an improvement over alternative methods. Here we address three interrelated issues with the proposed approach. First, we show why rank-order scores along PC axes do not provide appropriate characters for the phylogenetic analyses used, question the rationale behind selecting a subset of PCs from each module, and consider why the use of maximum likelihood does not overcome these issues. Second, we evaluate the findings against those from methods that are not expressly cladistic, grouping by overall rather than shared derived similarity. Third, we consider the effects of subdivision into modules. Our reanalyses show that despite the fact that they take no account of synapomorphy, topologies of unweighted pair group method with arithmetic mean (UPGMA; Rohlf and Sokal, 1981) phenograms from the whole skull and the modules differ no more from the published maximum parsimony (MP) and maximum likelihood (ML) cladograms than these differ from each other. Likewise, the vast majority of analyses of arbitrary subsets of landmarks (i.e., not based on a priori biological knowledge) from whole skulls and from arbitrary subdivisions of the landmarks from the whole into four modules support a cluster including all representatives of *Homo*, the "most remarkable result" of MCL (González-José et al., 2008: 776). This arises because there is such a strong pattern of overall similarity that even approaches such as UPGMA of arbitrary 'modules' yield the same result. Therefore, the MCL approach has not been shown to be phylogenetically more informative than groupings based on overall similarity using arbitrary modules.

Rank-order scores along principal component axes do not provide appropriate characters for phylogenetic analysis

Geometric morphometric shape variables provide quantification of biological form that conserves the geometry of the landmark configuration. As such, it is natural to consider whether such variables can be used to estimate phylogenetic trees (Macleod and Forey, 2002). However, there are important methodological reasons to be cautious when using such variables for phylogenetic inference. The primary analytical problem is how to use cladistic linear parsimony (whose input is typically a set of variables represented by discrete character states) to examine the phylogenetic signal present in continuous, multivariate shape data. González-José et al. (2008) proposed reducing the shape data for anatomical regions to a set of variables based on the principal component axes that spanned at least 75% of the variation in shape space within each region. Subsequently, rank-ordered scores on each principal component axis for each species were used to generate a set of ordered character states for use in a cladistic parsimony analysis (as described in Goloboff et al., 2006). Unfortunately, like previous attempts at forcing multivariate shape variables into a form compatible with cladistic software (Fink and Zelditch, 1995; Zelditch et al., 1995), this method distorts the information

* Corresponding author.
E-mail addresses: dcadams@iastate.edu (D.C. Adams), alcardini@interfree.it (A. Cardini), l.monteiro@hull.ac.uk (L.R. Monteiro), paul.ohiggins@hyms.ac.uk (P. O'Higgins), rohlf@life.bio.sunysb.edu (F.J. Rohlf).

0047-2484/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jhevol.2010.02.003

Author's personal copy

present in such variables and this severely limits useful biological conclusions that can be made.

Using geometric morphometric methods, the shape of each species is represented by a point in the multivariate shape space based on the generalized Procrustes analysis (GPA) aligned coordinates or partial warp scores (Rohlf, 1999). Shape differences are described by the Euclidean distance between species in tangent space. Principal components analysis simply rotates this space; therefore, distances are preserved and no information is lost if all principal component scores are used. However, separate rank-ordering of objects along the individual axes does not preserve the distances among objects and character states generated from individual axes will change in complex ways if the axes are rotated differently (Adams and Rosenberg, 1998; Rohlf, 1998; Monteiro, 2000). In fact, any possible set of rank-orderings for the species can be found through some arbitrary rotation of the multivariate space (Bookstein, 1994; Naylor, 1996), which would result in a different set of ordered character states for each taxon.

When estimating phylogenies, any method relying on characters derived from individual axes of shape space is susceptible to arbitrary character state changes resulting from data rotations. This is the case for linear parsimony methods (including MCL), as these are based on Manhattan distances (Sneath and Sokal, 1973). This point is crucial, since different rotations will be obtained from principal component scores when another specimen is added to or deleted from the study (although the shape differences among existing species remain constant). While parsimony can yield different trees when additional Operational Taxonomic Units (OTUs) are incorporated, the data matrix for existing taxa should remain the same. Here it does not. In fact, even for the same set of taxa, the ordered character states generated from the GPA aligned coordinates, the partial warp scores, or the principal components will differ from one another, resulting in different phylogenetic estimates. Because all of these data sets contain identical shape information, the differences are due solely to the arbitrary procedure of generating character states for cladistic parsimony. The reason for this is clear: MCL requires rank-ordered characters that approximate the observed shape differences, yet these values change when the shape space is rotated. Further, while distances remain the same, changes in the rank-order along one PC axis are not compensated for by changes in another axis. Including more PC axes does not solve this problem. Consequently, the set of character states found through this procedure is inconsistent with the mathematical properties of shape space. Thus, using this approach, different phylogenetic estimates will be obtained from the same shape data.

An example is seen in Figure 1. Here, the rank-order of five species is seen along their two principal component axes of shape (the objects are triangles). When a sixth species is added, the order of the original specimens along all axes is altered. Thus, the MCL approach generates different sets of characters for cladistic parsimony from the same data set. Indeed, as acknowledged by the authors of earlier studies using partial warps (PWs) as cladistic characters: "PCA provides a coordinate system for shape analysis, and may be useful for finding characters, but individual PCs (like individual PWs) cannot be viewed as characters in their own right" (Zelditch et al., 2004).

Importantly, use of maximum likelihood (the alternative method used by González-José et al. [2008] and in other recent studies such as Cannon and Manos [2001], Polly [2003, 2008], Caumul and Polly [2005], and Cardini and Elton [2008]) does not overcome issues with using PCs of shape as characters for phylogenetic inference, because the use of maximum likelihood, such as implemented in the CONTML software, assumes that the characters evolve independently with equal variances following a Brownian motion model of evolutionary change. This assumption is not met because the PCs used were based on the wrong covariance matrix and weighting. The PCs have to be computed using a covariance matrix corresponding to what Martins and Garland (1991) call 'input correlations'—correlations of the change at each internal node of a phylogenetic tree rather than the among-species covariance matrix that depends on the choice of species included in a study. This is the matrix that Polly (2004; matrix P) used to create simulations, rather than an among species covariance matrix that one could have at the end of a simulation. The 'input correlations' matrix can also be estimated using the phylogenetic comparative method (see equation 3 of Rohlf [2006]), but that requires a strongly supported phylogeny, such as a molecular tree (Felsenstein, 2002).

A final point is that in the MCL analyses it is possible that if all PCs are used rather than those representing a certain proportion of total variance (e.g., 0.75 as in González-José et al. [2008]), the results could be different. While higher PCs often represent very small proportions of the total variance, they are arguably of relevance to phylogenetics and so should be included.

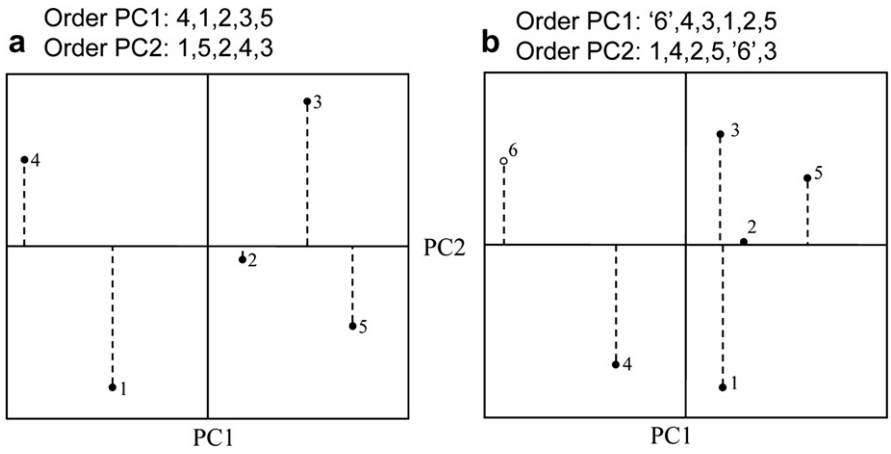


Fig. 1. (a) Five hypothetical species (analysis of triangles) as visualized along their two principal component axes. (b) The same data with a sixth species included. Notice that the rank-order of the original species along both axes is altered when another taxon is added. Thus, different estimates of character states (and therefore phylogeny) would be obtained from a cladistic parsimony analysis.

Author's personal copy

MCL trees arise from a strong pattern of superficial similarity

Beyond the theoretical considerations above, here we demonstrate the lack of evidence of efficacy of MCL through reanalyses of the data in González-José et al. (2008: their supplementary material). A UPGMA phenogram (Rohlf and Sokal, 1981; Rohlf, 2008) based on the whole set of landmarks, which ignores the concepts of 'character' and 'character polarity' and instead is based on shape distances between OTUs, shows groupings which are very similar to the clades apparent in the MP and ML trees (González-José et al., 2008). Further, quantitative comparisons of tree topologies (Page, 1993) fail to show any consistently greater differences between the phenogram based on Procrustes shape distances from the whole cranium and MCL trees, than those that are found between MCL trees. Very similar results arise from the UPGMA phenogram using Euclidean distances from combined module PCAs (see below; Table 1). This indicates that a purely phenetic clustering method (de Queiroz and Good, 1997) achieves exactly the same result as the new and purportedly promising MCL approach. This raises serious doubts about judging effectiveness of MCL in recovering the phylogenetic signal based on the observation that the "tree recovered the monophyletic status of *Homo* as well as some of the most undisputed internal relationships" (González-José et al., 2008: 777). MCL, cluster analysis, and previous studies show a good degree of agreement since they all reflect the same strong pattern of superficial similarity. This may be due to phylogeny or to other causes, but we cannot distinguish these possibilities from these data alone.

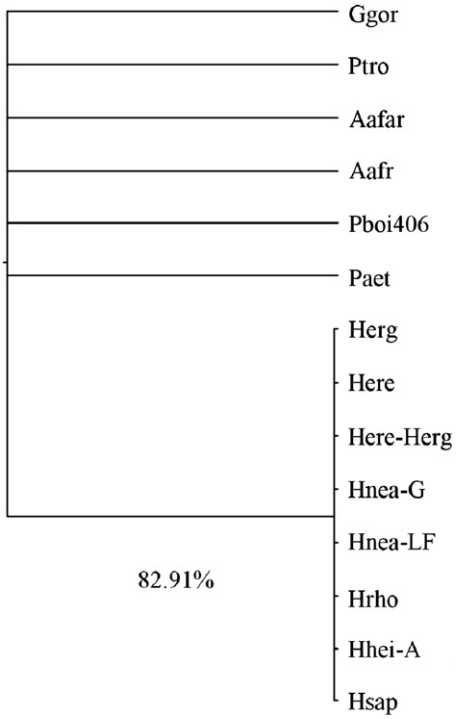


Fig. 2. Majority rule consensus tree of UPGMA phenograms from 10,000 random 'modules' comprising 13 landmarks. 82.91% percent of trees show a *Homo* cluster.

Modules are elusive and arbitrarily defined modules perform as well as those of González-José et al. (2008)

The MCL method is based on the identification of relatively independent modules. However, evolutionary covariation among face, cranial base, and vault (the independent modules postulated by González-José et al. [2008]) has been shown to be much stronger than previously expected by developmental modular organization (Bookstein et al., 2003; Lieberman et al., 2008; Mitteroecker and Bookstein, 2008; Perez and Monteiro, 2008). This is likely because of the elusive nature of evolutionary modules in relation to developmental integration (Monteiro et al., 2005; Mitteroecker and Bookstein, 2008; Zelditch et al., 2008) and the widespread effects of localized mutations causing integration among modules (Hallgrimsson et al., 2007). The division of cranial shape as proposed by González-José et al. (2008) does not reflect evolutionarily independent modules, nor does it generate independent characters for cladistic analysis.

To show that the proposed division of landmarks into cranial modules has little value for estimating phylogenies, we examined the effects of analyzing randomly defined modules (subsets of landmarks). Two analyses were carried out. First, we randomly selected 13 (of 43) landmarks to represent a 'module' (the mean number of landmarks across modules in González-José et al. [2008]). This was used to generate a matrix of Procrustes shape distances and a UPGMA phenogram was computed. The process was repeated 10,000 times. We found that *Homo* species group together to the exclusion of all other taxa in 82.91% of such phenograms (Fig. 2). Second, we divided the cranium into four arbitrary modules (each with approximately equal numbers of landmarks), computed PCs from shape coordinates of each of the four arbitrary modules, and built a UPGMA phenogram from the Euclidean distances combining all four PCAs. As in the previous randomization experiment, the process was repeated 10,000 times with the finding that some 99.64% of these resulted in an exclusively *Homo* containing cluster (results not shown). These findings indicate that because the skull is a very integrated structure, the proposed division into modules has no specific effect relative to using the entire skull or randomly generated modules.

Table 1
Measures of tree dissimilarity[a]

	MP	ML	$UPGMA_M$	$UPGMA_W$
MP	–	4	5	3
ML	10	–	5	4
$UPGMA_M$	10	12	–	4
$UPGMA_W$	12	14	12	–

[a] Partition measures and the number of leaves to prune in order to get the greatest agreement among subtrees were computed using unrooted trees (Page, 1993) and are shown, respectively, below and above the main diagonal. Probability that observed distances are more similar than expected by chance is always < 0.001 when estimated by measuring distances of pairs of trees randomly selected from the set of all possible binary trees (Page, 1993). MP = maximum parsimony, ML = maximum likelihood, $UPGMA_M$ = UPGMA from modules, $UPGMA_W$ = UPGMA from the whole cranium.

Conclusion

While theoretical considerations alone raise severe objections to the modular cladistic approach proposed by González-José et al. (2008), our reanalyses of their data confirm the lack of efficacy of the MCL approach. Our conclusion is that results based on the new method approximately agree with results from alternative approaches, and to some degree with previous phylogenetic hypotheses because of a strong pattern of superficial similarity. Indeed, the geometric morphometric analyses presented here give a statistically stable result of some interest in that the genus *Homo* as defined by this sample is morphologically distinct within the hominids. This said, the considerations and analyses presented here lead to the conclusion that the ability of the new method (MCL) to recover phylogenetic information is unproven. While the

Author's personal copy

trees produced are appealing, it should be recognized that the superiority of one method over another is not demonstrated using the present data set because the signal is so strong that even UPGMA produces similar trees.

The absence of evidence of efficacy and the severe methodological and theoretical issues surrounding the use of shape variables as cladistic characters should strongly caution researchers against the acceptance of the MCL approach.

References

Adams, D.C., Rosenberg, M.S., 1998. Partial-warps, phylogeny, and ontogeny: a comment on Fink and Zelditch (1995). Syst. Biol. 47, 168–173.

Bookstein, F.L., 1994. Can biometrical shape be a homologous character? In: Hall, B.K. (Ed.), Homology: The Hierarchical Basis of Comparative Biology. Academic Press, New York, pp. 197–227.

Bookstein, F.L., Gunz, P., Mitteroecker, P., Prossinger, H., Schaefer, K., Seidler, H., 2003. Cranial integration in *Homo*: singular warps analysis of the midsagittal plane in ontogeny and evolution. J. Hum. Evol. 44, 167–187.

Cannon, C.H., Manos, P.S., 2001. Combining and comparing continuous morphometric descriptors with a molecular phylogeny: the case of fruit evolution in the Bornean *Lithocarpus* (Fagaceae). Syst. Biol. 50, 860–880.

Cardini, A., Elton, S., 2008. Does the skull carry a phylogenetic signal? evolution and modularity in the guenons. Biol. J. Linn. Soc. Lond. 93, 813–834.

Caumul, R., Polly, D., 2005. Phylogenetic and environmental components of morphological variation: skull, mandible, and molar shape in marmots (*Marmota*, Rodentia). Evolution 59, 2460–2472.

Felsenstein, J., 2002. Quantitative characters, phylogenies, and morphometrics. In: MacLeod, N., Forey, P. (Eds.), Morphology, Shape and Phylogeny. Taylor and Francis, London, pp. 27–44.

Fink, W.L., Zelditch, M.L., 1995. Phylogenetic analysis of ontogenetic shape transformations: a reassessment of the piranha genus *Pygocentrus* (Teleostei). Syst. Biol. 44, 344–361.

Goloboff, P.A., Mattoni, C.I., Quinteros, A.S., 2006. Continuous characters analyzed as such. Cladistics 22, 589–601.

González-José, R., Escapa, I., Neves, W.A., Cúneo, R., Pucciarelli, H.M., 2008. Cladistic analysis of continuous modularized traits provides phylogenetic signals in *Homo* evolution. Nature 453, 775–779.

Hallgrimsson, B., Lieberman, D.E., Liu, W., Ford-Hutchinson, A.F., Jirik, F.R., 2007. Epigenetic interactions and the structure of phenotypic variation in the cranium. Evol. Dev. 9, 76–91.

Lieberman, D.E., Benedikt, H., Wei, L., Parsons, T.E., Jamniczky, H.A., 2008. Spatial packing, cranial base angulation, and craniofacial shape variation in the mammalian skull: testing a new model using mice. J. Anat. 212, 720–735.

Macleod, N., Forey, P. (Eds.), 2002. Morphology, Shape And Phylogenetics. Taylor and Francis, London.

Martins, E.P., Garland Jr., T., 1991. Phylogenetic analysis of the correlated evolution of continuous characters: a simulation study. Evolution 45, 534–557.

Mitteroecker, P., Bookstein, F.L., 2008. The evolutionary role of modularity and integration in the hominid cranium. Evolution 62, 943–958.

Monteiro, L.R., 2000. Why morphometrics is special? The problem with using partial warps as characters for phylogenetic inference. Syst. Biol. 49, 796–799.

Monteiro, L.R., Bonato, V., Reis, S.F., 2005. Evolutionary integration and morphological diversification in complex morphological structures: mandible shape divergence in spiny rats (Rodentia, Echimyidae). Evol. Dev. 7, 430–440.

Naylor, G.J.P., 1996. Can partial warp scores be used as cladistic characters? In: Marcus, L.F., Corti, M., Loy, A., Naylor, A., Slice, D.E. (Eds.), Advances in Morphometrics. Plenum, New York, pp. 519–530.

Page, R.D.M., 1993. Component: Tree Comparison Software for Microsoft Windows, Version 2.0. User's Guide. The Natural History Museum, London.

Perez, S.I., Monteiro, L.R., 2008. Non-random factors in modern human morphological diversification: a study of craniofacial variation in southern South American populations. Evolution 62(4), 978–993.

Polly, D.P., 2003. Paleophyloeography: the tempo of geographic differentiation in marmots (*Marmota*). J. Mammal. 84, 369–384.

Polly, P.D., 2008. Adaptive zones and the pinniped ankle: a 3d quantitative analysis of carnivoran tarsal evolution. In: Sargis, E., Dagosoto, M. (Eds.), Mammalian Evolutionary Morphology: A Tribute to Frederick S. Szalay. Springer, Dordrecht, The Netherlands, pp. 165–194.

de Queiroz, K., Good, D.A., 1997. Phenetic clustering in biology: a critique. Q. Rev. Biol. 72, 3–30.

Rohlf, F.J., 1998. On applications of geometric morphometrics to studies of ontogeny and phylogeny. Syst. Biol. 47, 147–158.

Rohlf, F.J., 1999. Shape statistics: procrustes superimpositions and tangent spaces. J. Classific. 16, 197–223.

Rohlf, F.J., 2006. A comment on phylogenetic correction. Evolution 60, 1509–1515.

Rohlf, F.J., 2008. NTSYSpc, Version 2.20U. Exeter Software, Setauket, New York.

Rohlf, F.J., Sokal, R.R., 1981. Comparing numerical taxonomic studies. Syst. Zool. 30 (4), 459–490.

Sneath, P.H.A., Sokal, R.R., 1973. Numerical Taxonomy. Freeman, London, UK.

Zelditch, M.L., Fink, W.L., Swiderski, D.L., 1995. Morphometrics, homology, and phylogenetics: quantified characters as synapomorphies. Syst. Biol. 44, 179–189.

Zelditch, M.L., Swiderski, D.L., Sheets, H.D., Fink, W.L., 2004. Geometric Morphometrics for Biologists: a Primer. Elsevier Academic Press, New York.

Zelditch, M.L., Wood, A.R., Bonett, R.M., Swiderski, D.L., 2008. Modularity of the rodent mandible: integrating bones, muscles and teeth. Evol. Dev. 10, 756–768.