doi:10.1038/nature06891

LETTERS

Cladistic analysis of continuous modularized traits provides phylogenetic signals in *Homo* evolution

Rolando González-José[1], Ignacio Escapa[2], Walter A. Neves[3], Rubén Cúneo[2] & Héctor M. Pucciarelli[4]

Evolutionary novelties in the skeleton are usually expressed as changes in the timing of growth of features intrinsically integrated at different hierarchical levels of development[1]. As a consequence, most of the shape-traits observed across species do vary quantitatively rather than qualitatively[2], in a multivariate space[3] and in a modularized way[4,5]. Because most phylogenetic analyses normally use discrete, hypothetically independent characters[6], previous attempts have disregarded the phylogenetic signals potentially enclosed in the shape of morphological structures. When analysing low taxonomic levels, where most variation is quantitative in nature, solving basic requirements like the choice of characters and the capacity of using continuous, integrated traits is of crucial importance in recovering wider phylogenetic information. This is particularly relevant when analysing extinct lineages, where available data are limited to fossilized structures. Here we show that when continuous, multivariant and modularized characters are treated as such, cladistic analysis successfully solves relationships among main *Homo* taxa. Our attempt is based on a combination of cladistics, evolutionary-development-derived selection of characters, and geometric morphometrics methods. In contrast with previous cladistic analyses of hominid phylogeny, our method accounts for the quantitative nature of the traits, and respects their morphological integration patterns. Because complex phenotypes are observable across different taxonomic groups and are potentially informative about phylogenetic relationships, future analyses should point strongly to the incorporation of these types of trait.

Cladistic analysis provides a solid framework to reconstruct phylogenetic relationships among taxa, because it identifies monophyletic groups by looking for shared derived characters. Theoretically, most cladistic methods need these characters to be discrete and independent, among other requirements. However, completion of cladistic analysis becomes problematic because, at lower taxonomic levels, most of observable variation is expressed as continuous changes of size and shape[2,3] rather than in discrete identifiable structures. Although some traits can be reasonably treated as discrete, it is also true that modern morphometrics provide a rich source of quantitative characters, which raises the question of how to use them in inferring phylogenies. This problem has some important implications. First, apart from the fact that discretization methods are still the subject of intense debate[2,7], these procedures disregard the continuous nature of many complex morphological traits. Note that discretization can be either explicit, through a broad spectrum of gap-weighting methods applied on an admittedly continuous trait, or implicit through arbitrary definition of discrete character states upon a complex trait. In this case, the morphological features can be described quantitatively, but are presented qualitatively (for example, low position of the infraorbital foramen). Second, discretization procedures force the multivariant nature of many complex phenotypes to be artificially treated as a collection of univariant measurements, disregarding the multivariant and geometric nature of form[3]. This is of crucial importance in cladistic practice, because homoplasy could be less likely in multivariate, complex phenotypes than in univariate traits[8]. Finally, at these low taxonomic levels, quantitative and developmental genetics can provide powerful additional tools to estimate degrees of character independence[5,9]. In this context, functional and developmental integration leads to the co-inheritance of character complexes, often called modules, which are then constrained to evolve in a coordinated, rather than independent, fashion[9]. In human palaeontology, for instance, this principle is routinely violated as functionally and developmentally linked traits are subdivided for analytical purposes[10]. A logical approach to this fact in phylogenetic systematics is to treat integrated features as a single phylogenetic complex, and to treat the complex as if it were an independent character[5,11].

Here we present a cladistic analysis of the most complete fossil specimens pertaining to the hominid lineage, which explicitly takes into account the above implications. Our analysis considers the genus *Homo* as the ingroup, and the remaining specimens assigned to the genera *Gorilla*, *Pan*, *Australopithecus* and *Paranthropus* as outgroups (Table 1). The choice of characters is based on the most conservative approach in terms of modularity. The classical bulk of characters used previously[12] is reduced to just four modular characters that condense the main craniofacial shape changes observed in the analysed taxa. Geometric morphometrics[13] methods are used to capture shape changes on these characters respecting both the geometric and the multivariate concept of shape. Finally, these geometric-morphometrics-derived, continuous, multivariant and modular characters are treated as such in a cladistic analysis.

Characters selected for analysis are flexure of the cranial base, facial retraction, neurocranial globularity, and shape and relative position of the masticatory apparatus. These characters reflect principal trends of variation on structures behaving as modules by varying somewhat independently[11,14]. Even when further localized modules can be detected or hypothesized, our attempt here is to evaluate the phylogenetic signal contained in this restricted modularity hypothesis. Furthermore, these traits reflect the major evolutionary trends that acted to differentiate the hominid lineage. Even though multiple characters reflecting aspects of these traits have been cladistically analysed in previous studies, they have not been used in a modular way in a phylogenetic–systematic framework. Three-dimensional landmarks reflecting the shape of the modular characters were

[1]Unidad de Investigación de Diversidad, Sistemática y Evolución, Centro Nacional Patagónico, Consejo Nacional de Investigaciones Científicas y Técnicas, CONICET, Boulevard Brown 2825, U9120ACF Puerto Madryn, Argentina. [2]Museo Paleontológico Egidio Feruglio, Consejo Nacional de Investigaciones Científicas y Técnicas, CONICET, Avenida Fontana 140, U9100GYO, Trelew, Argentina. [3]Laboratório de Estudos Evolutivos Humanos, Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, CP 11461, 05422.970 São Paulo, Brazil. [4]Departamento Científico de Antropología del Museo de La Plata, Universidad Nacional de La Plata, Consejo Nacional de Investigaciones Científicas y Técnicas, CONICET, 1900 La Plata, Argentina.

digitized on casts of the specimens listed in Table 1 (see Methods and Supplementary Information). Raw landmark coordinates were converted to shape coordinates by generalized Procrustes analysis. Generalized Procrustes analysis was performed for each module separately, and the aligned specimens were submitted to a principal component analysis[13]. Principal components can enable the pattern of variation between specimens described by many variables to be summarized by relatively few, when the data (for example, relative landmark locations) covary. The loading of each specimen on all the principal components necessary to achieve the 75% of variance explained was used as a continuous variable depicting the character state. Phylogenetic analysis of the resulting matrix was performed using the maximum parsimony algorithm[6] for additive characters implemented in TNT[15], and the maximum likelihood algorithm for quantitative traits developed by Felsenstein[16] available in Phylip[17] (see Methods).

One tree of maximum parsimony was obtained using a heuristic search, with 10,000 random addition sequences, saving 10 trees per replicate (Fig. 1a). In addition, one tree of maximum likelihood was also computed (Fig. 1b). The monophyletic status of the genus *Homo* is the most remarkable result in both analyses. Hypothetical ancestral character states for the *Homo* clade (node 8) are presented in Fig. 2 and Supplementary Information. With respect to the outgroups, this clade shows a more flexed cranial base, more retracted faces and an increase in the neurocranial globularity. This particular morphology is usually used to define the genus as well as to discuss the inclusion of some taxa in it[18,19].

Moreover, the internal relationships of the *Homo* specimens show a remarkable agreement with previews phylogenetic hypotheses[20]. Maximum parsimony and maximum likelihood analyses only differ in the relative position of *H. sapiens* in relation to the complex *H. erectus*, *H. ergaster*, *H. rhodesiensis*. Whereas the maximum parsimony cladogram places *H. sapiens* as a sister of a clade formed by specimens assigned to *H. erectus*, *H. ergaster* and *H. rhodesiensis*, maximum likelihood accommodates *H. sapiens* in a derived position relative to it. What is coincident in both analyses is the association of two controversial specimens (D2700 and Broken Hill) to a clade also formed by *H. erectus* and *H.ergaster*, as previously suggested[21–23]. *H. neanderthalensis* and *H. heidelbergensis* are represented in our analyses by several specimens, which form a single separate monophyletic group. Thus, our results are in agreement with previous assertions[23,24] recognizing *H. heidelbergenis* and *H. neanderthalensis* as chronological variants inside a single biological lineage. The fact that *H. neanderthalensis sensu stricto* does not form a monophyletic clade with *H. sapiens* reinforces the idea that they are separate species. Conceptually, this is a key support for the method presented here, because this observation is also defended by studies based in evidence other than skull shape, such as analyses of ancient DNA[25] and growth patterns[26].

Finally, *H. rudolfensis* is the sister group of all the previous clades, and *H. habilis* is at the base of the monophyletic *Homo* clade (Fig. 1a). Based on our matrix, and as previously stated[27], there is no convincing reason to remove *H. habilis* and *H. rudolfensis* from the genus[18], and their position goes counter to their inclusion in the *Australopithecus* genus[19]. In addition, our analysis supports the notion that *H. habilis* and *H. erectus* indeed represent different lineages even though they were recently reported as sympatric and contemporaneous forms[28].

Even when the focus of this work is on the relationships within *Homo*, the outgroup topology also shows some relevant points. For

Table 1 | Fossil (top) and recent (bottom) samples included in the analysis*

Specimen	Species assigned	Known age range (Myr)	Code
A.L. 444-2 (reconstruction)	*Australopithecus afarensis*	≥3.7–3.0	Aafa
Sts 5	*A. africanus*	*ca*. 3.0–2.5	Aafr
KNMER-406	*Paranthropus boisei*	≥2.3–1.4	Pboi-406
OH 5	*P. boisei*	≥2.3–1.4	Pboi-OH5
SK 48	*P. robustus*	*ca*. 1.5–2.0	Prob
WT 17000	*P. aethiopicus*	*ca*. 2.7–2.3	Paet
KNMER 1470	*Homo rudolfensis*	2.5–1.8	Hrud
KNMER 1813	*H. habilis*	2.1–1.5	Hhab
KNMER 3733	*H. ergaster*	2–1	Herg
Zhoukoudian†	*H. erectus*	1.8–0.03	Here
D2700	*H. erectus/H. ergaster*	1.8	Here-Herg
Steinheim	*H. heidelbergensis*	0.8–0.2	Hhei-S
Kabwe, Broken Hill 1	*H. rhodesiensis*	0.8–0.2	Hrho
Atapuerca 5	*H. heidelbergensis*	0.8–0.2	Hhei-A
Gibraltar 1, Forbes' Quarry	*H. neanderthalensis*	0.2–0.03	Hnea-G
La Chappelle-aux-Saints 1	*H. neanderthalensis*	0.2–0.03	Hnea-LC
La Ferrassie 1	*H. neanderthalensis*	0.2–0.03	Hnea-LF
CTL-004	*Gorilla gorilla*	9–0‡	Ggor
CTL-006	*Pan troglodytes*	8–0‡	Ptro
Patagonian, Río Negro #797	*H. sapiens*	0.2–0	Hsap

* All specimens are stored at the Laboratorio de Estudos Evolutivos Humanos, University of São Paulo, Brazil, except Prob, Hgeo and Hhei-A, which are stored at the Unitat d'Antropologia, Faculty of Biology, University of Barcelona, and Hsap, which is stored at the Museo de La Plata, Argentina.
† First-generation casts were used for the fossil specimens. Sawyer and Tattersall's reconstruction.
‡ Estimated (molecular) time of disruption from the lineage of *H. sapiens*.

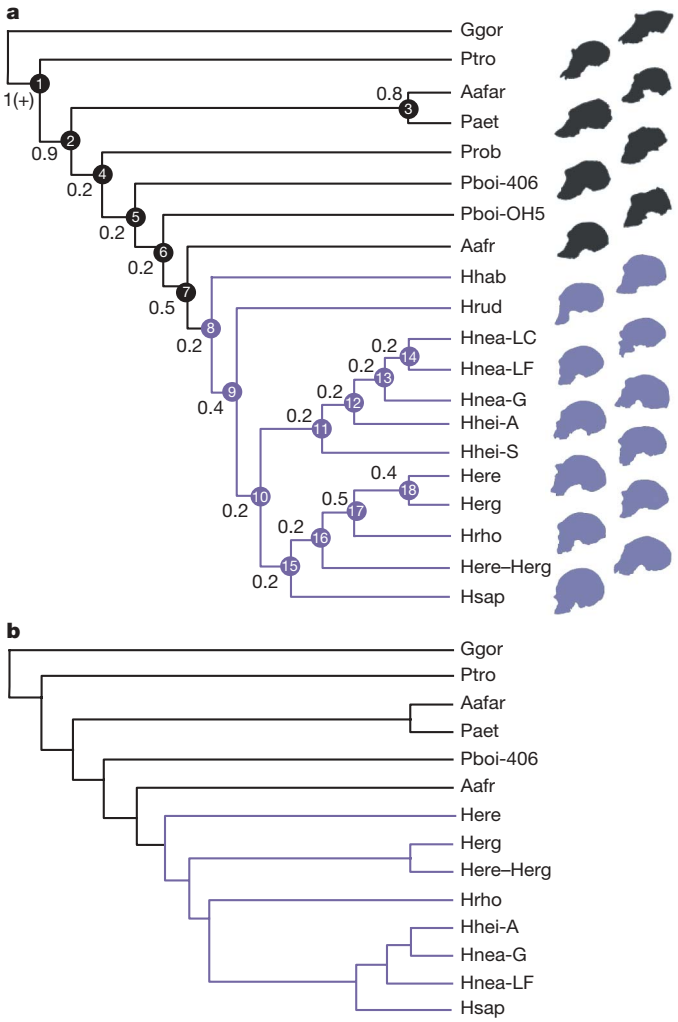

Figure 1 | Phylogenetic relationships among *Homo* species and other hominid taxa. a, Single tree obtained by equal weighted maximum-parsimony analysis based on morphological data of four cranial morphological modules. Bremer support values are displayed, as well as a numeric label for each node. The Bremer support values were determined by examination of the strict consensus of trees 0.01–0.12 steps longer than the shortest tree found for the data set (Supplementary Information). **b**, Single tree obtained by maximum likelihood. The ingroup (*Homo* specimens) and outgroups are displayed in purple and black respectively.

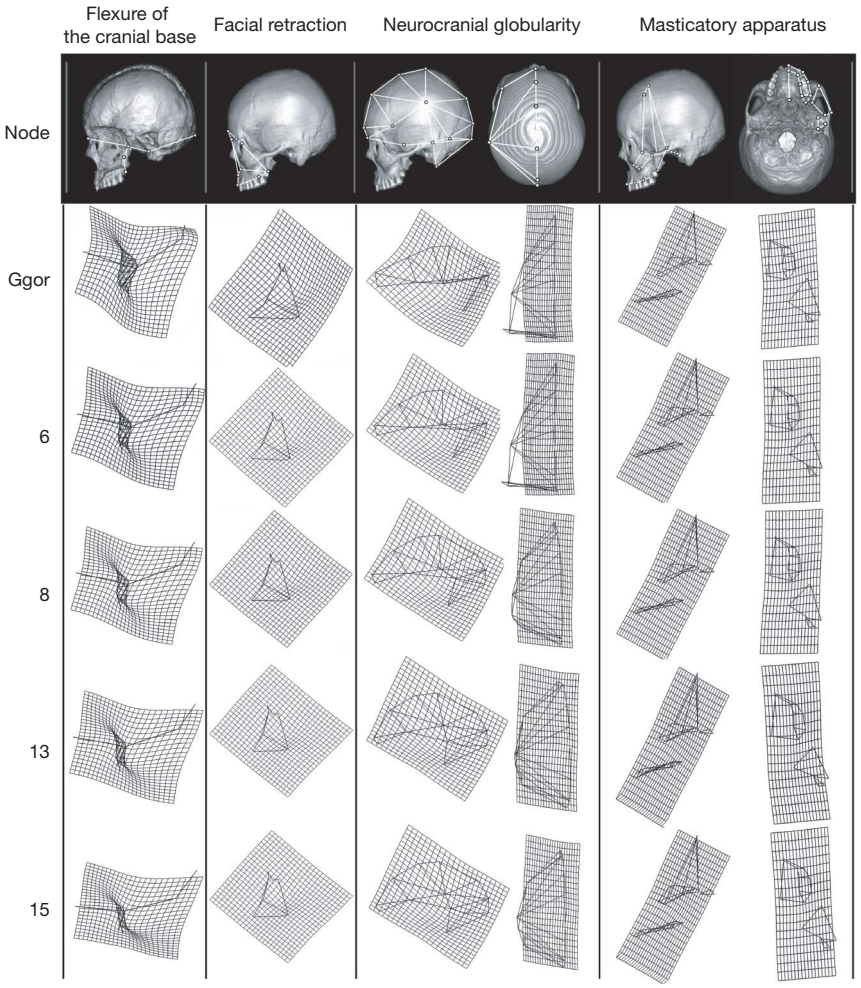

Figure 2 | Reconstruction of ancestral states corresponding to the root and the main nodes of the maximum parsimony cladogram. Ancestral states corresponding to the first principal component of each trait, estimated as values across the principal component of each character, and visualized as deformation grids from the reference (the origin of the first principal component) towards the estimated principal component score of each node. Variation and ancestral states corresponding to further principal components can be explored using Supplementary Information.

instance, *Australopithecus afarensis* and *Paranthropus aethiopicus* are in a basal branch of the hominid clade, in a derived position respecting *Pan*, which is placed at the base of the tree, over the root (*Gorilla*). Neither maximum parsimony nor maximum likelihood cladograms support the genus status of *Paranthropus* or *Australopithecus*, because the fossils classically included in each of these genera fail to form monophyletic clades. Note that the paraphyletism of *Paranthropus* has been previously reported[12]. Interestingly, our analyses place *Australopithecus africanus* as the sister group of the genus *Homo*.

In summary, our approach shows how classically disregarded information recovers significant evolutionary signals, using a new and promising methodology. Certainly, our tree recovered the monophyletic status of *Homo* as well as some of the most undisputed internal relationships (see a review in refs 20 and 22). However, our analysis is based on only four traits selected for previous knowledge about the relative independence of modules, and which treats the multivariant, geometric and continuous nature of skull shape as such.

The theoretical implications of the approach presented here are broad, but point in three main directions. First, reconstruction of phylogenetic relationships can be done by taking into account the modular development and evolution of complex phenotypes. In fact, this type of critical study on character independence should be the first phase of any cladistic analysis when no previous information is available[5]. Second, there is no reason to discard or force discretization of continuous multivariant traits. Conversely, our analysis shows that it is possible to recover relevant phylogenetic signals in characters previously ignored or arbitrarily discretized. Note, however, that even though there are relatively few truly discrete traits, many of them exhibit qualitatively different states that are very distinct in different groups of taxa, and thus are reasonably discrete. In this context, future work should emphasize the combined use of quantitative and qualitative traits. Finally, when geometric morphometrics methods are used to depict shape changes across a multivariate space, reconstruction of ancestral states in combination with the visualization of shape changes across phyletic lineages should be considered as a straightforward and common-sense procedure. In summary, valuable phylogenetic information is recovered from data sets that consider independence on a developmentally and functionally basis, and which preserve the multivariant and continuous nature of complex phenotypes.

METHODS SUMMARY

Three-dimensional craniofacial landmark coordinates were digitized on 17 fossil hominid specimens and *Pan paniscus*, *Gorilla gorilla* and *Homo sapiens* skulls (Table 1). Landmarks were divided into four different subsets describing flexure of the cranial base, facial retraction, neurocranial globularity and the masticatory apparatus. Each subset was superimposed using generalized Procrustes analysis to remove the effects of translation, rotation and scaling, and then submitted to a principal component analysis. Projection of each specimen on the principal components was used as the shape descriptor and considered as a continuous trait in the cladistic analysis. Trees were obtained using *Gorilla gorilla* as the root,

and using maximum parsimony as well as maximum likelihood criteria. Ancestral states obtained from the maximum parsimony algorithm were used to visualize the shape changes corresponding to each node on the tree.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 12 October 2007; accepted 5 March 2008.
Published online 4 May 2008.

1. Olson, E. C. & Miller, R. L. *Morphological Integration* (Univ. Chicago Press, Chicago, 1958).
2. Rae, T. C. The logical basis for the use of continuous characters in phylogenetic systematics. *Cladistics* **14,** 221–228 (1998).
3. MacLeod, N. & Forey, P. in *Morphology, Shape and Phylogeny* (eds MacLeod, N. & Forey, P.) 1–7 (Taylor & Francis, London, 2002).
4. Pigliucci, M. & Preston, K. *Phenotypic Integration. Studying the Ecology and Evolution of Complex Phenotypes* (Oxford Univ. Press, New York, 2004).
5. Strait, D. S. Integration, phylogeny, and the hominid cranial base. *Am. J. Phys. Anthropol.* **114,** 273–297 (2001).
6. Farris, J. S., Kluge, A. G. & Eckhardt, M. J. A numerical approach to phylogenetic systematics. *Syst. Zool.* **19,** 172–189 (1970).
7. Humphries, C. J. in *Morphology, Shape and Phylogeny* (eds MacLeod, N. & Forey, P.) 8–26 (Taylor & Francis, London, 2002).
8. Polly, P. D. On the simulation of the evolution of morphological shape under selection and drift. *Palaeontol. Electron.* **7,** 1–28 (2004).
9. Cheverud, J. M. Morphological integration in the saddle-back tamarin (*Saguinas fuscicolis*) cranium. *Am. Nat.* **145,** 63–89 (1995).
10. Curnoe, D. Problems with the use of cladistic analysis in palaeoanthropology. *Homo* **53,** 225–234 (2003).
11. Ackermann, R. R. & Cheverud, J. M. in *Phenotypic Integration: Studying the Ecology and Evolution of Complex Phenotypes* (eds Pigliucci, M. & Preston, K.) 302–319 (Oxford Univ. Press, Oxford, 2004).
12. Skelton, R. R. & McHenry, H. M. Evolutionary relationships among early hominids. *J. Hum. Evol.* **23,** 309–349 (1992).
13. Zelditch, M. L., Swiderski, D. L., Sheets, H. D. & Fink, W. L. *Geometric Morphometrics for Biologists* (Elsevier, London, 2004).
14. Lieberman, D. E., McBratney, B. M. & Krovitz, G. The evolution and development of cranial form in *Homo sapiens*. *Proc. Natl Acad. Sci. USA* **99,** 1134–1139 (2002).
15. Goloboff, P. A., Farris, J. S. & Nixon, K. TNT, a free program for phylogenetic analysis. *Cladistics*. (in the press).
16. Felsenstein, J. *Inferring Phylogenies* (Sinauer Associates, Sunderland, Massachusetts, 2004).
17. Felsenstein, J. PHYLIP (Phylogeny Inference Package) v.3.67 (Department of Genome Sciences, University of Washington, Seattle, 2007).
18. Wood, B. A. & Collard, M. The human genus. *Science* **284,** 65–71 (1999).
19. Collard, M. & Wood, B. A. in *Handbook of Paleoanthropology* (eds Henke, W., Hardt, T. & Tattersall, I.) 1575–1610 (Springer, Berlin and Heidelberg, 2007).
20. Strait, D. S., Grine, F. E. & Fleagle, J. G. in *Handbook of Paleoanthropology* (eds Henke, W., Hardt, T. & Tattersall, I.) 1782–1806 (Springer, Berlin and Heidelberg, 2007).
21. Vekua, A. *et al.* A new skull of early *Homo* from Dmanisi, Georgia. *Science* **297,** 85–89 (2002).
22. Rightmire, G. P. in *Handbook of Paleoanthropology* (eds Henke, W., Hardt, T. & Tattersall, I.) 1695–1715 (Springer, Berlin and Heidelberg, 2007).
23. Stringer, C. B. in *Paleoclimate and Evolution with Emphasis on Human Origins* (eds Vrba, E. S., Denton, G. H., Partridge, T. C. & Burckle, L. H.) 524–531 (Yale Univ. Press, New Haven, 1995).
24. Arsuaga, J. L., Martínez, I., García, A. & Lorenzo, C. The Sima de los Huesos crania (Sierra de Atapuerca, Spain). A comparative study. *J. Hum. Evol.* **33,** 219–281 (1997).
25. Lalueza-Fox, C. *et al.* Neandertal evolutionary genetics: mitochondrial DNA data from the Iberian Peninsula. *Mol. Biol. Evol.* **22,** 1077–1081 (2005).
26. Ramirez Rozzi, F. V. & Bermudez De Castro, J. M. Surprisingly rapid growth in Neanderthals. *Nature* **428,** 936–939 (2004).
27. Strait, D. S. & Grine, F. E. Inferring hominoid and early hominid phylogeny using craniodental characters: the role of fossil taxa. *J. Hum. Evol.* **47,** 399–452 (2004).
28. Spoor, F. *et al.* Implications of new early *Homo* fossils from Ileret, east of Lake Turkana, Kenya. *Nature* **448,** 688–691 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank P. Goloboff, M. Hernández, C. Lalueza, N. Martínez-Abadías, D. Pol and F. Ramírez-Rozzi for reading and discussing previous versions of this paper. We also thank R. E. Ambrosetto, R. Nicoletti and B. Nicoletti for their assistance during this work. The program TNT is freely available, thanks to a subsidy from the Willi Hennig Society.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to R.G.J. (rolando@cenpat.edu.ar).

METHODS

Data acquisition and geometric morphometrics. Data were collected as three-dimensional coordinates of anatomical landmarks on the casts of specimens listed in Table 1. The landmark used in each character is provided in the Supplementary Information. All crania were measured by one observer (R.G.J.) using a Microscribe G2X digitizer. Because many of the fossil specimens were incomplete, some landmarks were reconstructed using anatomical information from the preserved surrounding areas. Missing bilateral landmarks on one side only were estimated by reflection. Scores were standardized to a mean of zero and a variance equal to the proportion of the variance explained by the corresponding principal component[29]. Principal components have a biological meaning, as orthogonal dimensions of variance, even though that is not equivalent to the meaning of a character[13]. They are not likely to be characters in their own right because they are directions of variation that are constrained to be orthogonal (by definition), not directions of evolutionary change. However, data matrices submitted to cladistic analysis are not formed by characters represented by consecutive principal components of a single morphological structure, but by a collection of the first principal components representing the main trends of morphological change on four independent modules. Thus, overlapping of two or more specimens in a given-character principal component score does not mean that they are similar for a principal component on the remaining characters.

Cladistic analysis. The matrix is composed of 20 cranial specimens (Table 1) and 18 characters, which correspond to the firsts principal components of each module necessary to account for the 75% of explained variance. Maximum parsimony cladistic analysis used equal weighted maximum parsimony implemented in TNT[15,30], which allowed the use of continuous characters as such by optimizing them following the classical algorithm[6] for additive characters[31]. When using continuous characters, a synapomorphy could be just a subtle change in one character. Thus, using continuous characters implies that there are an infinite number of character states. The most parsimonious tree was obtained using a heuristic search with 10,000 random addition sequences followed by tree-branching-regrafting. Characters were polarized, using *G. gorilla* to root the tree. Branch support was estimated using the Bremer method, using the suboptimal trees from 0.1 to 1 additional steps. Ancestral shapes are the optimizations of the maximum parsimony tree using the Wagner algorithm[6]. When using continuous traits, the usual output of this algorithm is a range of ancestral states, according to the maximum parsimony principle. Thus, to visualize the ancestral shape, we mapped back the transformations by taking the central value of the range and back-standardizing the standardized scores to recover the raw score values. These raw scores were used as the 'target' to obtain the wireframe depicting shape changes from the reference. Alternatively, because quantitative traits are expected to reverse direction in a Brownian-motion-like manner, a maximum likelihood tree was computed following Felsenstein[16,17].

29. Polly, P. D. Paleophylogeography: the tempo of geographic differentiation in marmots (*Marmota*). *J. Mammal.* **84,** 369–384 (2003).
30. Goloboff, P. A., Farris, J. S. & Nixon, K. TNT: Tree analysis using New Technology v.1.1 (Willi Hennig Society, New York, 2008).
31. Goloboff, P. A., Mattoni, C. I. & Quinteros, A. S. Continuous characters analyzed as such. *Cladistics* **22,** 589–601 (2007).