



CORRESPONDENCE



Who is the closest extant cousin of humans? Total-evidence approach to hominid phylogenetics via simultaneous optimization

ABSTRACT

J. R. Grehan & J. H. Schwartz (*Journal of Biogeography*, 2009, 36, 1823–1844) argued that humans (*Homo*) are more closely related to orangutans (*Pongo*) than to chimpanzees (*Pan*), and used this scenario to draw biogeographical conclusions about human origins. They discussed a contradiction between phenotypical and molecular results that has led to a debate about the reliability of genetic versus phenotypic data. The main aim of our study is to test the conflicting phylogenetic hypotheses by a total-evidence analysis based on simultaneous optimization of extensive phenotypic and molecular data sets. Our results supported the human–chimpanzee clade, without any phenotypical–molecular data conflict, as the same phylogeny emerged both from the total analysis and when the molecular and phenotypic data were analysed separately. Sensitivity analyses showed that the result was not dependent on the parameters chosen for character weighting.

Keywords Cladistics, direct optimization, DNA, hominid evolution, morphology, phylogenetics, sensitivity analysis, total evidence.

INTRODUCTION

The evolutionary relationships among hominids have been debated for a long time. Since molecular evidence became available, the mainstream view has converged on the (*Homo–Pan*)–*Gorilla* clade being sister to *Pongo*, and *Hyllobates* being sister to all great apes and humans (e.g. Enard & Pääbo, 2004; Campbell *et al.*, 2008). The prevailing view

of a *Homo–Pan* clade has recently been challenged by Grehan & Schwartz (2009), whose phylogenetic analysis based on phenotypic characters supported a *Homo–Pongo* clade instead. Grehan & Schwartz (2009) questioned the reliability of molecular data, and targeted especially heavy criticism on the methods used to deduce phylogenetic relationships from those data. Although some discussion has followed (Schwartz & Grehan, 2009; Stoneking, 2009), the methodological doubt remains unaddressed.

We largely agree with the criticism that Schwartz & Maresca (2006) and Grehan & Schwartz (2009) presented on the analytical methods used to reconstruct the hominid phylogenetic tree. No molecular studies concerning hominid relationships have actually critically tested the assumed molecular homologies. Instead, DNA sequences have been aligned before the actual phylogenetic analysis, which means that one of several possible hypotheses on which nucleotides are homologous is chosen *a priori*. However, characters are defined as homologous when they share a common origin, so one cannot logically claim homology without referring to a phylogenetic tree. In principle, homology is hypothesized for molecular and morphological characters in the same way: character congruence is assessed under competing phylogenetic scenarios. Grehan & Schwartz (2009) claim that DNA sequences must always be aligned before analysis. It is true that most phylogenetic analyses have done so, but this is not necessary. In the approach known as direct optimization (DO), both the alignments and the phylogenetic trees are considered hypotheses, and the analysis searches for the most parsimonious combination of the two (Hein, 1989, 1990; Wheeler, 1996). In this context it is important to understand that DO optimizes all the data simultaneously, so if morphological characters are included in the analysis, even they will affect the final alignment of molecular data (Phillips, 2006). Traditionally, morphological variation is also ‘aligned’ as a static data matrix before phylogenetic

analysis, so that alternative homology schemes are not tested (Agolin & D’Haese, 2009). It is possible to optimize morphological homologies dynamically as well, but usually the characters chosen for coding are assumed to represent the most optimal homology correspondences (but see Schulmeister & Wheeler, 2004; Agolin & D’Haese, 2009).

In the present study, we assess whether the earlier molecular results supporting a *Homo–Pan* clade were dependent on methodological problems. Our approach is to perform a total-evidence analysis; that is, to incorporate both DNA and phenotypic data in the same analysis, without imposing any *a priori* hypotheses about the phylogenetic relationships among hominids. The potential usefulness of a total-evidence analysis for solving the apparent conflict between molecular and morphological data in hominid phylogenetics was recognized several years ago (Gura, 2000), but such an analysis has, to our knowledge, not yet been carried out. In an early study, Andrews & Martin (1987) combined morphological and molecular data in one analysis, but they were limited by the relatively simple analytical methods available at the time, and in addition had access to only a rather small data set in comparison to those available at present. Here we use a large compilation of data and the most rigorous test of homology possible by applying DO of DNA sequences (Wheeler, 1996), and test the sensitivity of the resulting phylogeny to assumptions made about sequence evolution (Giribet, 2003).

MATERIALS AND METHODS

Details of the data set and analytical methods used are given in Appendix S1 in the Supporting Information. The phenotypic data set (Appendix S2) is based on existing character matrices for craniodental (Strait & Grine, 2004) and soft-tissue (Gibbs *et al.*, 2002) data. Strait & Grine (2004) modified and expanded their craniodental data set from the one analysed by Collard & Wood

(2000). A third phenotypic data set in our analysis is the one presented by Grehan & Schwartz (2009) in favour of the *Homo–Pongo* clade (with some modifications, see Appendix S1). The DNA sequence sampling was based largely on the studies of Wildman *et al.* (2003) and Satta *et al.* (2000), but we used only those sequences that were available for all great apes. Our data set was enlarged by sampling additional coding and non-coding DNA sequences available for all great apes in GenBank (Appendix S3). We used DO to guarantee the most parsimonious homology scheme for the DNA sequences and the most rigorous test of homology. We ran four separate analyses: (1) molecular data including extant taxa only, (2) phenotypic data including extant taxa only, (3) molecular data + phenotypic data including extant taxa only, and (4) molecular data + phenotypic data including those fossil taxa that had a sufficient number of informative characters (as in analysis D of Grehan & Schwartz, 2009). The analyses were performed with POY4 (Varón *et al.*, 2010) using eight processors in the HP CP4000 BL ProLiant supercluster (Finnish IT Center for Science, CSC). We investigated the branch support of the obtained tree topologies by calculating dynamic jackknife support values (Farris *et al.*, 1996; Varón *et al.*, 2010). In addition, we evaluated the robustness of the phylogenetic tree by testing how sensitive each of the nodes of interest is to changes in the transformation cost parameters of the phylogenetic analysis (Appendix S1).

RESULTS AND DISCUSSION

The prevailing view of chimpanzees as the nearest living relatives of humans is supported by our analyses, and the *Homo–Pan* clade appears robust. Analysis 1 (DNA sequence data only) produced a single most parsimonious tree (tree length 32,639 steps and jackknife support values $\geq 99\%$ for every branch) supporting the *Homo–Pan* clade as sister to *Gorilla*, and *Pongo* as sister to African apes and humans (Fig. 1a). This tree proved to be very stable in the sensitivity analysis. Only one extremely unequal transformation cost regime (transversion cost 4, gap cost 16) resulted in a different tree, in which *Gorilla* was resolved as the sister of *Homo*.

Analysis 2 (phenotypic data only) resulted in a single most parsimonious tree with the same topology as that obtained with the molecular data, and with 79–100% jackknife support values (Fig. 1a; tree length 894 steps).

Analysis 3 (phenotypic and DNA evidence of extant taxa) produced a single most parsimonious tree (length 33,535 steps) with the same topology and 100% jackknife support value for every branch (Fig. 1a). In the sensitivity analysis, this topology was obtained under all transformation cost parameter combinations, which means that the inclusion of phenotypic data increased the stability and support of the *Homo–Pan* hypothesis. The molecular data consisted of a much larger number of characters than the phenotypic data, which may lead one to think that the result was in practice determined by the molecular evidence. However, even a modest number of characters can significantly direct the result of a phylogenetic analysis. For example, a study on mammal phylogeny based on a nucleotide data matrix of 16,397 bp resulted in a tree that detached the guinea-pig from the rodent clade (D'Erchia *et al.*, 1996). The addition of just 21 morphological characters was sufficient to return the guinea-pig to the rodents (Bergsten, 2005).

Finally, Analysis 4 (molecular and phenotypic data including sufficiently known fossils) resulted in a single most parsimonious tree (33,580 steps). This tree (Fig. 1b) contradicts the results of Grehan & Schwartz (2009), but is in agreement with the prevailing view on extant hominid relationships, with all African genera in one clade and the Eurasian genera in another. Although the jackknife support is generally rather low (probably because of the large amount of missing data in the fossil taxa), the topology was highly stable, as only one transformation cost regime (transversion cost 4, gap cost 16) resulted in a slightly different topology.

The overall conclusion from our analyses is that the prevailing view of chimpanzees as the nearest living relatives of humans is supported and appears robust when all existing evidence is analysed together. However, it cannot be denied that there are numerous similarities between *Homo* and *Pongo* in morphological, life history, physiological, behavioural and cultural traits (Schwartz, 1984a,b; van Schaik *et al.*, 2003; Grehan, 2006; O'Higgins & Elton, 2007; Thorpe *et al.*, 2007; Kelley & Schwartz, 2010). If the currently accepted hypothesis of hominid evolution is correct, these similarities must either be symplesiomorphies (shared primitive characters), convergences, or erroneous observations. If the similarities are symplesiomorphies, this would greatly affect the way we look at ourselves: many human features should be considered

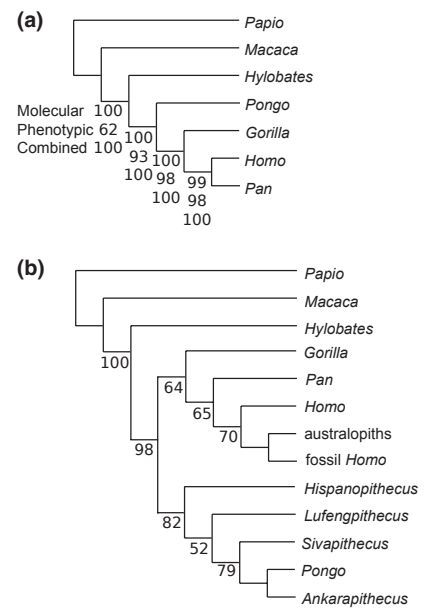


Figure 1 Phylogenetic relationships among extant great apes and some fossil taxa. (a) Cladogram of extant taxa based on either molecular, phenotypic, or simultaneous analysis of molecular and phenotypic data; jackknife support values for each node are given separately for each analysis. (b) Cladogram of both extant and fossil taxa based on simultaneous analysis of molecular and phenotypic data with jackknife support values.

'primitive', whereas the corresponding traits in gorillas and chimpanzees should be considered more 'derived'. It has already been suggested that lack of knuckle-walking is a primitive trait shared by *Homo* and *Pongo*, and that knuckle-walking itself is a derived trait that evolved independently in *Gorilla* and *Pan* (O'Higgins & Elton, 2007; Thorpe *et al.*, 2007).

The second possibility, that *Homo* and *Pongo* share many traits as a result of convergent evolution, is equally intriguing. Until now, most hypotheses of human evolution have attempted to explain how and why the human lineage diverged from that of chimpanzees (Schwartz, 2004). In order to understand human evolution, it might be equally important to explain how and why so many putatively convergent characteristics emerged in orangutans and humans. Convergent adaptations can be crucial for understanding what selective forces and evolutionary conditions have prevailed during the millions of years of lineage evolution (Pagel, 1994).

The third alternative for explaining the similarities between *Homo* and *Pongo* is

methodological rather than evolutionary. Erroneous character coding is a potential problem in all kinds of data, and this possibility should be carefully examined. Accumulation of analogies or substitution of synapomorphies (shared derived characters) will erode the phylogenetic signal (Wägele & Mayer, 2007), and the recognition of this requires multiple independent data sources (requirement of total evidence; Fitzhugh, 2006). Most molecular systematic studies are based on phenetically determined primary homologies, not on the maximization of homologies as is the case for DO. On the other hand, character state delimitation is much more straightforward in molecular than in morphological data. Most of the phenotypic characters used in hominid systematics are quantitative rather than qualitative (Curnoe, 2003). The delimitation of character states in quantitative characters is highly problematic (e.g. Farris, 1990; Stevens, 1991; Thiele, 1993; Rae, 1998; Curnoe, 2003; de Bivort *et al.*, 2010), which may lead to different codings of the same features by different authors, and to unjustified character state coding in general. One interesting approach to this problem is the Wagner optimization of continuous character state ranges without *a priori* character state delimitation, which is allowed by TNT (Goloboff *et al.*, 2008). This approach has been applied to clarify the evolutionary relationships among fossil and extant species of *Homo* (with *Gorilla* and *Pan* among the outgroups; González-José *et al.*, 2008). Once the character states have been defined, they can be ordered and weighted using different methods, each of which has its strengths and pitfalls (Kluge, 2004; Wägele & Mayer, 2007). We followed the rationale that both ordering and unequal character weighting are *ad hoc* hypotheses, and that the most parsimonious topology is obtained with equal character weights and unordered transformation series (Kluge, 2004).

SAMULI LEHTONEN*, ILARI E. SÄÄKSJÄRVI, KALLE RUOKOLAINEN AND HANNA TUOMISTO

Section of Biodiversity and Environmental Research, Department of Biology, University of Turku, FI-20014 Turku, Finland
*E-mail: samuli.lehtonen@utu.fi

ACKNOWLEDGEMENTS

S.L. received financial support from the Kone foundation.

REFERENCES

- Agolin, M. & D'Haese, C.A. (2009) An application of dynamic homology to morphological characters: direct optimization of setae sequences and phylogeny of the family Odontellidae (Poduromorpha, Collembola). *Cladistics*, **25**, 353–385.
- Andrews, P. & Martin, L. (1987) Cladistic relationships of extant and fossil hominoids. *Journal of Human Evolution*, **16**, 101–118.
- Bergsten, J. (2005) A review of long-branch attraction. *Cladistics*, **21**, 163–193.
- de Bivort, B., Clouse, R.M. & Giribet, G. (2010) A morphometrics-based phylogeny of the Gondwanan mite harvestmen (Opiliones, Cyphophthalmi, Pettalidae). *Journal of Zoological Systematics and Evolutionary Research*, doi: 10.1111/j.1439-0469.2009.00562.x.
- Campbell, N.A., Reece, J.B., Urry, L.A., Cain, M.L., Wasserman, S.A., Minorsky, P.V. & Jackson, R.B. (2008) *Biology*, 8th edn. Pearson – Benjamin Cummings, San Francisco.
- Collard, M. & Wood, B. (2000) How reliable are human phylogenetic hypotheses? *Proceedings of the National Academy of Sciences USA*, **97**, 5003–5006.
- Curnoe, D. (2003) Problems with the use of cladistic analysis in palaeoanthropology. *Homo*, **53**, 225–234.
- D'Erchia, A.M., Gissi, C., Pesole, G., Saccone, C. & Arnason, U. (1996) The guinea-pig is not a rodent. *Nature*, **381**, 597–600.
- Enard, W. & Pääbo, S. (2004) Comparative primate genomics. *Annual Review of Genomics and Human Genetics*, **5**, 351–378.
- Farris, J.S. (1990) Phenetics in camouflage. *Cladistics*, **6**, 91–100.
- Farris, J.S., Albert, V.A., Källersjö, M., Lipscomb, D. & Kluge, A.G. (1996) Parsimony jackknifing outperforms neighbor-joining. *Cladistics*, **12**, 99–124.
- Fitzhugh, K. (2006) The “requirement of total evidence” and its role in phylogenetic systematics. *Biology and Philosophy*, **21**, 309–351.
- Gibbs, S., Collard, M. & Wood, B. (2002) Soft-tissue anatomy of the extant hominoids: a review and phylogenetic analysis. *Journal of Anatomy*, **200**, 3–49.
- Giribet, G. (2003) Stability in phylogenetic formulations and its relationship to nodal support. *Systematic Biology*, **52**, 554–564.
- Goloboff, P.A., Farris, J.S. & Nixon, K.C. (2008) TNT, a free program for phylogenetic analysis. *Cladistics*, **24**, 774–786.
- 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**, 775–778.
- Grehan, J.R. (2006) Mona Lisa smile, the morphological enigma of human and great ape evolution. *The Anatomical Record (Part B: New Anatomist)*, **289B**, 139–157.
- Grehan, J.R. & Schwartz, J.H. (2009) Evolution of the second orangutan: phylogeny and biogeography of hominid origins. *Journal of Biogeography*, **36**, 1823–1844.
- Gura, T. (2000) Bones, molecules...or both? *Nature*, **406**, 230–233.
- Hein, J. (1989) A new method that simultaneously aligns and reconstructs ancestral sequences for any number of homologous sequences, when the phylogeny is given. *Molecular Biology and Evolution*, **6**, 649–668.
- Hein, J. (1990) Unified approach to alignment and phylogenies. *Methods in Enzymology*, **183**, 626–645.
- Kelley, J. & Schwartz, G.T. (2010) Dental development and life history in living African and Asian apes. *Proceedings of the National Academy of Sciences USA*, **107**, 1035–1040.
- Kluge, A.G. (2004) On total evidence: for the record. *Cladistics*, **20**, 205–207.
- O'Higgins, P. & Elton, S. (2007) Walking on trees. *Science*, **316**, 1292–1294.
- Pagel, M.D. (1994) The adaptationist wager. *Phylogenetics and ecology* (ed. by P. Eggleton and R.I. Vane-Wright), pp. 29–51. Linnean Society Symposium Series, Vol. 17. Academic Press, London.
- Phillips, A.J. (2006) Homology assessment and molecular sequence alignment. *Journal of Biomedical Informatics*, **39**, 18–33.
- Rae, T.C. (1998) The logical basis for the use of continuous characters in phylogenetic systematics. *Cladistics*, **14**, 221–228.
- Satta, Y., Klein, J. & Takahata, N. (2000) DNA archives and our nearest relative, the trichotomy problem revisited. *Molecular Phylogenetics and Evolution*, **14**, 259–275.
- van Schaik, C.P., Ancrenaz, M., Borgen, G., Galdikas, B., Knott, C.D., Singleton, I., Suzuki, A., Utami, S.S. & Merrill, M. (2003) Orangutan cultures and the evolution of material culture. *Science*, **299**, 102–105.
- Schulmeister, S. & Wheeler, W.C. (2004) Comparative and phylogenetic analysis of developmental sequences. *Evolution & Development*, **6**, 50–57.
- Schwartz, J.H. (1984a) The evolutionary relationships of man and orang-utans. *Nature*, **308**, 501–505.
- Schwartz, J.H. (1984b) Hominoid evolution: a review and a reassessment. *Current Anthropology*, **25**, 655–672.

- Schwartz, J.H. (2004) Trying to make chimpanzees into humans. *History and Philosophy of the Life Sciences*, **26**, 271–277.
- Schwartz, J.H. & Grehan, J. (2009) Reply to “Humans as second orangutans: sense or nonsense?”. *Bioessays*, **31**, 1263–1266.
- Schwartz, J.H. & Maresca, B. (2006) Do molecular clocks run at all? A critique of molecular systematics. *Biological Theory*, **1**, 357–371.
- Stevens, P.F. (1991) Character states, morphological variation, and phylogenetic analysis: a review. *Systematic Botany*, **16**, 553–583.
- Stoneking, M. (2009) Humans as second orangutans: sense or nonsense? *Bioessays*, **31**, 1010–1012.
- Strait, D.S. & Grine, F.E. (2004) Inferring hominoid and early hominid phylogeny using craniodental characters, the role of fossil taxa. *Journal of Human Evolution*, **47**, 399–452.
- Thiele, K. (1993) The holy grail of the perfect character: the cladistic treatment of morphometric data. *Cladistics*, **9**, 275–304.
- Thorpe, S.K.S., Holder, R.L. & Crompton, R.H. (2007) Origin of human bipedalism as an adaptation for locomotion on flexible branches. *Science*, **316**, 1328–1331.
- Varón, A., Vinh, L.S., Bomash, I. & Wheeler, W.C. (2010) POY version 4: phylogenetic analysis using dynamic homologies. *Cladistics*, **26**, 72–85.
- Wägele, J.-W. & Mayer, C. (2007) Visualizing differences in phylogenetic information content of alignments and distinction of three classes of long-branch effects. *BMC Evolutionary Biology*, **7**, 147.
- Wheeler, W. (1996) Optimization alignment, the end of multiple sequence alignment in phylogenetics? *Cladistics*, **12**, 1–9.
- Wildman, D.E., Uddin, M., Liu, G., Grossman, L.I. & Goodman, M. (2003) Implications of natural selection in shaping 99.4% nonsynonymous DNA identity between humans and chimpanzees: enlarging genus *Homo*. *Proceedings of the National Academy of Sciences USA*, **100**, 7181–7188.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Appendix S1 Background information on the phenotypic data, molecular data and phylogenetic analyses used to assess phylogenetic relationships among the great apes.

Appendix S2 Phenotypic data matrices used to analyse phylogenetic relationships among the great apes.

Appendix S3 DNA sequences used to analyse phylogenetic relationships among the great apes.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

Editor: Brett Riddle