Journal of Biogeography (J. Biogeogr.) (2011) 38, 2397–2404

CORRESPONDENCE

Response to Collins *et al.* (2011)

ABSTRACT

We note serious problems in Collins *et al.* (*Journal of Biogeography*, 2011, doi: 10.1111/j.1365-2699.2011.02506.x): failure to use over 80% of the available data; failure to use one of the two available archipelagoes; mistaken inclusion of four species; and reliance on a grossly inadequate number of null matrices. Curing the paper of these problems would have strengthened the evidence for checkerboards and the role of competition.

Keywords Birds, Bismarck Archipelago, checkerboards, competition, Solomon Islands, supertramps.

We are pleased that Collins *et al.* (2011) now agree that for Bismarck bird distributions 'there are excesses of checkerboard pairs within both genera and defined guilds' compared with random expectations, and that this is consistent with strong influence by interspecific competition. They also discuss the other hypotheses that we considered in a prior publication (Sanderson *et al.*, 2009). Here, we alert readers to problems in their paper that, if cured, would have considerably strengthened those conclusions.

1. We had already provided, in downloadable form, the complete database of 41 Bismarck islands and 142 Solomon islands, and 150 and 141 analysable bird species, respectively (Sanderson *et al.*, 2009). Collins *et al.* (2011) analysed only a small fraction of those data: 31 Bismarck islands and 154 bird species tabulated by Mayr & Diamond (2001). They ignored the Solomon Islands completely, thereby failing to notice the strikingly consistent behaviour of individual bird genera between the two archipelagoes that we discussed. Furthermore, as noted by Mayr & Diamond (2001) and by Sanderson *et al.* (2009), four of those 154 species (*Casuarius bennettii, Pelecanus conspicillatus, Scythrops novaehollandiae, Halcyon sancta*) should not have been included in the analysis because of uncertainty about their breeding status or native status.
2. Collins *et al.* (2011) employed only 1000 null matrices. We found (Sanderson *et al.*, 2009, p. 774, paragraph 2) that even 10,000 null matrices are insufficient to converge on consistent null distributions. We used 1,000,000 matrices for all our published analyses.
3. Collins *et al.* (2011) conducted three sets of analyses (community-wide, congeneric, and intraguild checkerboards), of which we had already published the first two. We did not do the third (intraguild analyses), because Connor & Simberloff (1979) had criticized them previously, and we note that Collins *et al.* (2011) criticized them again in their paper. We performed four additional analyses that Collins *et al.* (2011) omitted: positive associations (our p. 775), archipelago differences (p. 775), incidence effects (p. 777), and genus comparisons between archipelagoes (p. 778). Thus, Collins *et al.* (2011) report no new analyses: they merely repeat one-third of our analyses with one-sixth of our dataset and 1–1000th of our number of randomized matrices. Their discussion also adds nothing substantively new: it merely raises again, without the benefit of personal experience, the issues of supertramp distributions and historical biogeography that we had already discussed with knowledge of the species and archipelagoes involved.

James G. Sanderson[1], Jared Diamond[2] and Stuart L. Pimm[3]

[1]*Wildlife Conservation Network, Los Altos, CA, USA,* [2]*Geography Department, University of California, Los Angeles, CA, USA,* [3]*Nicholas School of the Environment, Duke University, Durham, NC, USA*
E-mail: stuartpimm@me.com

REFERENCES

Collins, M.D., Simberloff, D. & Connor, E.F. (2011) Binary matrices and checkerboard distributions of birds in the Bismarck Archipelago. *Journal of Biogeography*. doi:10.1111/j.1365-2699.2011.02506.x

Connor, E.F. & Simberloff, D. (1979) The assembly of species communities: chance or competition? *Ecology*, **60**, 1132–1140.

Mayr, E. & Diamond, J. (2001) *The birds of Northern Melanesia: speciation, ecology, and biogeography*. Oxford University Press, Oxford.

Sanderson, J.G., Diamond, J.M. & Pimm, S.L. (2009) Pairwise co-existence of Bismarck and Solomon landbird species. *Evolutionary Ecology Research*, **11**, 771–786.

Editor: Robert Whittaker

doi:10.1111/j.1365-2699.2011.02575.x

Evolution of human–ape relationships remains open for investigation

ABSTRACT

We demonstrate that much of the morphological and metrical data Lehtonen *et al.* (*Journal of Biogeography*, 2011, **38**, 805–808) present in support of a closer relationship between humans and chimpanzees than between humans and orangutans are faulty. When the numerous invalid features and suggested character states are excluded, as they should be, the most robust theory of relationship that can be generated is between humans and orangutans. With regard to the direct optimization method (DOM) resolving

Journal of Biogeography

problems in analysing molecular (sequence) data, the method, if valid, requires that demonstration of sequence similarity is per force a demonstration of synapomorphy. In brief, DOM fails both to test theories of relatedness and to take into account the fact that identification of shared similarity does not translate into demonstration of synapomorphy.

Keywords Cladistics, evolution, hominid, human evolution, morphology, orangutan, phylogenetics.

INTRODUCTION

We welcome Lehtonen *et al.*'s (2011) critique of our analysis of potential phylogenetic relationships among the four large-bodied hominoids, which yielded the formerly long-accepted sister grouping of chimpanzees and gorillas as well as the less well-received sister grouping of humans and orangutans (Grehan & Schwartz, 2009). At the very least, but significantly, their rebuttal focuses attention on a number of broadly held misconceptions about the 'doing' of systematics and phylogenetic reconstruction.

First, the goal of our article (Grehan & Schwartz, 2009) was to assess critically the accuracy of previously published morphological features and measurements, as well as their purported character states, presence in taxa, and unique possession by taxa that were then declared sister taxa. We did so because many of these features have been recycled, without critical evaluation, from one publication to the next [the latest being Lehtonen *et al.*'s (2011)] and then cited as demonstration of the hypothesis that, among large-bodied hominoids, humans and chimpanzees are sister taxa, with gorillas and orangutans increasingly distant relatives (Collard & Wood, 2000; Gibbs *et al.*, 2002; Strait & Grine, 2004). As witnessed in Lehtonen *et al.*'s reiteration of these features and their interpretation, the conclusion of these earlier publications of a close human–chimpanzee relationship has assumed the status of received wisdom and their lists of morphology and metrics embraced as accurate. As Gibbs *et al.* (2002) and Lockwood *et al.* (2004) could declare on the basis of their respective inferences from their soft-tissue and morphometric analyses, finally at least some morphology supported what we 'knew' all along from interpretation of molecular data: humans and chimpanzees are the most closely related of the large-bodied hominoids. In order to maintain analytical integrity, we (Grehan & Schwartz, 2009) similarly investigated features that had been cited as uniting humans and orangutans and chimpanzees and gorillas as respective sister groups (Schwartz, 1984a,b, 1988, 1997, 2004, 2005).

After these assessments, we discovered that many published features supporting a human–chimpanzee relationship were not as they were portrayed: that is, a feature was incorrectly presented in its primary description, character state portrayal, and/or taxic representation (including its restriction only to the taxa for which it is presented as synapomorphic). We considered a feature that failed corroboration even in one of these categories of presentation to be false and did not include it in our analysis. We therefore disagree with Lehtonen *et al.*'s (2011) suggestion that our rejection of previously published features (in whichever form they were found to be lacking or incomplete) constituted a conscious effort to include in our analysis only those features that would support the results of our phylogenetic analysis: namely, that the sister groups human–orangutan and chimpanzee–gorilla are highly corroborated theories of relatedness. As is eminently clear in our publication (Grehan & Schwartz, 2009), we scrutinized the character lists of those publications that claimed support for human–chimpanzee as well as human–orangutan and chimpanzee–gorilla groupings and applied the same critical criteria to all. That Lehtonen *et al.* (2011) were at least somewhat aware of our effort to assess the veracity of published features prior to incorporating any of them into our analysis would seem to be indicated by their inclusion in their analysis of morphology only those features that survived our investigation of Schwartz's (1984a,b, 1988, 1997, 2004, 2005) publications. It is thus a mystery to us why, in light of our demonstration of their incorrectness, Lehtonen *et al.* (2011) would then cite *in toto* the recycled features from other publications claiming support for a human–chimpanzee relationship.

We are also perplexed by Lehtonen *et al.*'s (2011) second insinuation of our being consciously selective in choosing characters that a priori would support a human–orangutan theory of relatedness. In this instance, their incorrect assumption was that features we hypothesized as being shared derived – i.e. synapomorphic – of humans and orangutans were chosen as such, with features that might have yielded a human–chimpanzee or any other relationship consequently regarded as primitive character states. It is true that in the domain of a cladistic, hypothetico-deductive approach to phylogenetic reconstruction, only those features identified as shared derived or synapomorphic are regarded as reflecting potential closeness of relatedness. But in contrast to Lehtonen *et al.*'s misconception, in a proper cladistic analysis, features that are shared by taxa are not identified as synapomorphies in light of a preferred theory of taxic relationship [unlike the features that have been cited in support of a human–chimpanzee relationship (see Grehan & Schwartz, 2009)].

Unfortunately, Lehtonen *et al.* (2011) confuse (1) the hypothesizing of relative states of primitiveness versus derivedness through a transparent procedure that involves comparison among a taxically broad outgroup with (2) a conscious selection of features that would support a preferred theory of relationship. We are particularly disappointed in Lehtonen *et al.*'s reiteration of the latter, especially in light of the fact that we painstakingly spelled out our criteria in non-technical language so that readers not familiar with, improperly informed about, or in need of a methodological reminder of the process, would understand from the outset our approach.

In an attempt to maintain methodological rigour, we (Grehan & Schwartz, 2009) stipulated three criteria that morphological data must satisfy in order to be understood unambiguously and to be useful in analysing relatedness: (1) regardless of how many times a feature has been used in other analyses, its purported details – anatomical characteristics and representation in taxa – must hold up under independent scrutiny; (2) shared similarities qualify as potential synapomorphies for the ingroup (in this case the large-bodied hominoids comprising living and extinct humans and great apes) *if and only if* these shared features are absent or at best very rarely present in outgroup taxa; and (3) the outgroup must be taxically broad and diverse in order to properly test a hypothesis of synapomorphy between taxa of the presumed ingroup. Satisfying these criteria, however, is only the first step in a rigorous cladistic analysis.

But even when characters are well documented, and claims of similarity unambiguous, not all demonstrations of shared similarity reflect phylogenetic propinquity. The reason is simple: the hypotheses 'derivedness' and 'primitiveness' are relative concepts (e.g. Eldredge & Cracraft, 1980).

Features hypothesized as being derived at one level in a hierarchy of nested clades (i.e. are synapomorphic of a hypothesized clade and autapomorphic for the last common ancestor of that clade) are *de facto* primitive retentions for the members of that hypothetical clade (i.e. taxa hypothetically united by virtue of inheriting their presumed ancestor's autapomorphy or autapomorphies). Thus, while all shared similarities retained from a common ancestor may be properly hypothesized as being homologous among the taxa that possess them, only features that are restricted to a subset of taxa under investigation can lend themselves to the argument that they would have been unique to the hypothetical common ancestor of these taxa. Consequently, only features that satisfy the latter criterion – being taxically restricted – can reasonably be hypothesized as synapomorphic for the taxa sharing them. And the only way in which one can suggest that a feature is uniquely possessed by two or more taxa is by demonstrating that the same feature is not present in members of a broader taxonomic group. Furthermore, although we would hope this would be self evident, one cannot know prior to pursuing the broader taxic comparison which or how many features will emerge as unique to a subset of that larger taxonomic assemblage (see review in Schwartz, 2008).

We are baffled by Lehtonen *et al.*'s (2011) claim of verifying a close human–chimpanzee relationship because, in addition to demonstrating in transparent detail that many of the features they recycled from earlier publications were incorrect and/or not fully described, we (Grehan & Schwartz, 2009) went to considerable lengths to illustrate that claims other authors made in their publications for determining primitive versus derived characters states were not grounded in a methodologically solid, taxically broad comparison (see review in Schwartz, 2008). Rather, statements of character polarity were often based solely on comparison of the target taxa (the large-bodied hominoids, and especially humans and chimpanzees) with gibbons (one of the genera of small-bodied hominoids, or hylobatids). Although the rooting of an algorithm-based phylogenetic analysis in a specific taxon, which is *de facto* taken as being wholly primitive relative to its presumed sister clade, is commonplace in molecular analyses (and replicated in Lehtonen *et al.*, 2011), the repetition of this assumption does not prove its validity.

The presumption was invoked to justify an extremely limited outgroup comparison: because small-bodied hominoids are commonly accepted as the sister taxon of large-bodied hominoids, the character states of hylobatid features must be primitive relative to alternative character states seen in large-bodied hominoids. But when Zuckerkandl & Pauling (1962) first articulated their assumption that overall molecular similarity reflected phylogenetic propinquity, they claimed that more distantly related taxa were molecularly different from more recently divergent taxa because the earlier divergences of the former provided more time over which molecular difference would accrue. Clearly, while both scenarios may rely on the theoretically and methodologically invalid assertion that greater overall similarity equates with closer evolutionary ties, these respective assumptions cannot both be true simultaneously (Schwartz, 2005, in press a). That is, outgroup taxa cannot be different because they accrued difference after diverging from a common ancestor with the ingroup and also because they did not change (i.e. remained primitive) while only ancestors of the ingroup did change.

We must here also re-address (following Grehan & Schwartz, 2009) the validity of 'total evidence' analysis: i.e. the combination of morphological features with comparisons of randomly sampled molecular (DNA) sequence comparisons. To date, we are unaware of any biological justification, including Lehtonen *et al.* (2011), for combining morphological and sequence data in the same analysis. Indeed, because metazoan morphology results from complexly integrated communication between genes and gene products that regulate development, and virtually all molecular sequence comparisons are restricted to small portions of the coding region of the genome (see review in Schwartz & Maresca, 2007) that has nothing to do with development but only physiological adaptation, it is difficult to understand in the context of developmental biology how combining these unrelated datasets in a single analysis is feasible, much less phylogenetically meaningful.

Nevertheless, and in spite of these pitfalls and lacunae, Lehtonen *et al.* (2011, p. 806) conclude 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'. Consequently, they also conclude that the numerous features that humans and orangutans uniquely shared must be interpreted as symplesiomorphies, convergences, or erroneous observations. The former situations they attribute to problems in character coding from which, they believe, molecular data are exempt. Lehtonen *et al.* (2011) also claim that our critiques of DNA sequence alignment can be solved by a direct optimization method (DOM) whereby homology is determined (a posteriori) by the parsimony analysis that generated the preferred phylogeny in the first place. Although this assertion is clearly tautological – the preferred phylogeny is then used as the structure in which to analyse other data brought to bear on that structure – if and only if Lehtonen *et al.*'s (2011) 'evidence' is accepted as they present it can a human–orangutan relationship be rejected.

While a purely phenetic comparison between humans and any or all great apes will certainly identify an incalculable number of shared similarities between dyadic pairs, two important factors must be taken into consideration [in this case, upon accepting the hypothesis that the four extant large-bodied hominoids constitute a clade (Schwartz, 1986)]. First, if humans are held as the 'constant' against which each ape is compared, it is self-evident, as Huxley (1863) was fully aware, that any ape that differs more from humans than the others because it embodies a greater number of autapomorphies will, on a superficial level, appear as less 'like' humans; and if degree of 'overall similarity' is taken wholesale as the arbiter of 'closeness of relatedness' the taxon that is distinguished from the other hominoids by its very own uniquenesses will incorrectly be assumed to be the most primitive. Also incorrectly, then, the generally more similar taxa (because unlike the more autapomorphic taxon they retain more primitive features from the last common ancestor of the four extant large-bodied hominoids) will be considered more closely related to the exclusion of the autapomorphic taxon. Secondly, if one strove to determine the nature of any dyadic set of shared similarities, and did so in the context of a taxically broad comparison, one would discover that only a relatively small number of shared similarities are unique to any set(s) of taxa.

Consequently even if, as did Lehtonen *et al.* (2011), one uses 11 species of gibbon as the outgroup, it remains the case that these virtually identical species represent the same genus, *Hylobates*. Effectively then, because all evidence robustly supports the hypothesis that the small-bodied hominoids (gibbons and siamangs = hylobatids) constitute the sister-group of large-bodied

hominoids (see review in Schwartz, 1986), the hylobatid outgroup consists of a small clade of taxa that differ primarily in pelage and sometimes slightly in size. Furthermore, while the unity of Hylobatidae is based on various craniodental and forearm morphologies that only gibbons and siamangs among anthropoids possess, the task of rooting a tree in an outgroup demands that any morphology found in the defined outgroup is *de facto* primitive relative to features possessed by ingroup taxa. Clearly, however, the autapomorphies that distinguish hylobatids from all other anthropoids – including the large-bodied hominoids – cannot also be primitive relative to large-bodied hominoid features.

The same contradiction pervades the analysis of molecular data. For how can the common ancestor of hylobatids both inherit a history of molecular change that accumulated after this 'outgroup' diverged from its common ancestor with the 'ingroup' and yet also remain primitive relative to the ingroup in the molecular sequences being compared (see discussion in Schwartz, in press b). It should also be self-evident that this contradiction exists regardless of the approach one uses to align sequences. Furthermore, it is no less important that if one defines as primitive a taxon that was selected a priori for outgroup comparison, and then identifies as derived character states of these features that differ in one or more members of the predetermined ingroup, one is not pursuing a cladistic analysis, no matter how many times one invokes the terms primitive and derived (see discussion in Schwartz, 2008). At the risk of being redundant, the only way in which one can sensibly hypothesize character states as either primitive or derived is through the broadest taxonomic comparison possible. Contrary to the claims reiterated by Lehtonen *et al.* (2011), a cladistic analysis is not based on an a priori identification of features that would support a preferred phylogeny. Nor is analysing a data set that does not include features used in other analyses in support of what is clearly a preferred phylogeny when, as we have done before and here, one can demonstrate that those features are erroneous in, for example, accurately representing that feature and its character states or its presence or absence in cited taxa.

The non-traditional hypothesis of large-bodied hominoid relationships – human–orangutan and chimpanzee–gorilla sister groups – that emerged from our morphological analysis (Grehan & Schwartz, 2009) implies that the accepted approach to interpreting molecular data (sequence, retrotransposons, insertions, etc.) is incorrect. Although it has become commonplace to assume that demonstration of molecular similarity falsifies a morphologically-based theory of relationship (e.g. see claims and citations in Ruvolo, 1997; Stauffer *et al.*, 2001; O'hUigin *et al.*, 2002; Perelman *et al.*, 2011), this assumption does not derive from empirical demonstration (see review in Schwartz, in press b), but most frequently from invocation of 'the law of large numbers' (Sibley & Ahlquist, 1984): i.e. because in sequence analyses more bases are compared than could ever be possible with morphological features, only the former provides reliable phylogenetic insight. This argument is purely phenetic, as are post-1960s algorithmically-based molecular analyses, which are claimed to be cladistic, but which in reality assume that synapomorphy is generated by the molecular assumption: namely, that greater similarity reflects continual molecular change in an ancestral lineage prior to the divergence of sister taxa.

In responding to Lehtonen *et al.* (2011) we will here demonstrate that even if one overlooks the fact that restricting their outgroup to gibbons alone violates cladistic methodology, many of their proposed ((human–*Pan*)–*Gorilla*) synapomorphies must be rejected either because they are not restricted to humans and one or both of these great apes, or because the publications they relied on for their data often reported erroneous information. We also will return to questions about the nature of DNA sequence comparisons that we hope will provoke the recognition and reconsideration of the basic assumptions that underlie molecular interpretation of human and great ape evolution as well as the evolution of life in general.

MATERIALS AND METHODS

In the study criticized by Lehtonen *et al.* (2011), we (Grehan & Schwartz, 2009) dealt only with characters that could be documented, not only for the large-bodied hominoids (humans, orangutans, and the African apes) or for them and the lesser apes (gibbons and siamangs), but also for a diverse array of Old World monkey species. Such comparative breadth is critical because it is only by surveying a broadly representative outgroup that one can confidently determine character polarity among the large-bodied hominoids. Lehtonen *et al.* (2011) did not adhere to this cladistic requirement, limiting their 'molecular' comparisons to gibbons and, for some characters, only one or another species of Old World monkey. For morphology, they relied on Gibbs *et al.* (2002) for soft tissue information, where the outgroup was restricted solely to the gibbon. In both analyses, and of course in the combined endeavor, Lehtonen *et al.* (2011) assumed from the outset that the single-taxon outgroup represented the primitive state relative to the others.

Our original analysis included 28 proposed synapomorphies for humans and orangutans. Although Lehtonen *et al.* (2011) stated in their appendix that 'many' of our character-state codings were arbitrary, they only excluded three (GS55, molar shape poorly defined; GS68, estriol level same as GS39; GS71, intelligence because it was not quantified). For the purposes of our re-analysis of the data presented by Lehtonen *et al.* (2011) we accommodated these exclusions and in addition removed GS39 because this character was not represented in *Hylobates*. This brought the total of remaining GS human–orangutan characters to 24 in our current analysis.

Given our commitment to rigorous systematic analysis, we were obliged in our review of Lehtonen *et al.*'s (2011) data (see Appendix S1 in Supporting Information) to restrict the dataset to character states or metric values that were not present in the outgroup, did not overlap with the outgroup, or were not autapomorphic. We applied this principle to all characters, regardless of the human–ape relationship they might support. Lehtonen *et al.* (2011) used Thiele's (1993) clustering algorithm to generate grouped, multistate characters in spite, as they acknowledge, of well-known problems with this methodology (see Appendix S1 of Lehtonen *et al.*, 2011). In view of those problems we retained the original values to provide individual character states for each taxon where their grouped character states resulted in loss of information as a result of merging taxa prior to the analysis (Appendix S2).

Although we would heartily embrace critical evaluation of our dataset and its analysis, Lehtonen *et al.* (2011) accepted, without question, the correctness of the other datasets (Gibbs *et al.*, 2002; Strait & Grine, 2004) that in turn had been accepted without question and included in earlier compilations (e.g. Collard & Wood, 2000 in Strait & Grine, 2004). Both of the former publications were based on accounts in other literature that, upon scrutiny, we often

found were not supported by the citations provided. We further pointed out (Grehan & Schwartz, 2009) that the characters presented by Gibbs *et al.* (2002) and Strait & Grine (2004) (especially those deemed apomorphic) were either often variable, only occasionally present, or present in more than two of the often four or five taxa included in the comparison. For our review here, we continued our scrutiny of features in Gibbs *et al.* (2002) that Lehtonen *et al.* (2011) used, and again found (Appendix S2) that the sources cited by Gibbs *et al.* (2002) often did not support their assertions of apomorphy and synapomorphy, either because the asserted derived character states were not present in some or all of the taxa listed, or because ingroup character states were also present in outgroup taxa. Because we are still in the process of reviewing the original citations for all characters Gibbs *et al.* (2002) cited, we focused here on a subset of these characters to test our contention that so many are either erroneous or unsupported that any analysis using them will yield spurious results.

We then re-analysed the partially corrected dataset giving equal weight to all characters; outgroup states were coded 0 (zero), and ingroup states 1 (one) or higher. Even though it is commonplace to use different outgroup codes and rely on the algorithm to assign (hopefully accurately) character states, for the sake of clarity and methodological transparency we used a single code for the outgroup. Multistate characters were recognized as individual values for each taxon and, for consistency with Lehtonen *et al.* (2011), we treated all character states as unordered. We used maximum parsimony analysis with PAUP* 4.0b10 (Swofford, 2005). An exhaustive search was followed by bootstrap (50% majority rule) analysis to assess the resulting phylogenetic hypothesis. We did not include our fossil dataset because Lehtonen *et al.* (2011) had not modified that information.

RESULTS

Only 12 of Strait & Grine's (2004) coded craniodental measurements could be used in our analysis because the remainder were either autapomorphic (and therefore uninformative about relationships between taxa) or were also represented in the outgroup. Of the 89 measurements provided by Collard & Wood (2000), only 32 did not overlap with values given for the outgroup and could therefore be considered as putatively representing derived states for humans and great apes. Of the 171 characters listed by Gibbs *et al.* (2002), we rejected 31, either because information was lacking for the gibbon (the sole outgroup taxon), or because the states of derivedness claimed to characterize various large-bodied hominoids were not supported by the references cited (Appendix S2). We treated the remaining characters as informative even though we had not yet verified all of them and in spite of the outgroup being represented solely by the gibbon.

An exhaustive search generated a single most parsimonious tree with tree statistics (excluding uninformative characters) of consistency index (CI) = 0.62, homoplasy index (HI) = 0.32, retention index (RI) = 0.38, and rescaled consistency index (RC) = 0.18. This result supported the monophyly of humans and orangutans followed by ((human–orangutan)–chimpanzee) and then (((human–orangutan)–chimpanzee)–gorilla). A bootstrap analysis also supported the human–orangutan relationship at 72% but collapsed relationships with the African apes into an unresolved trichotomy (Fig. 1).

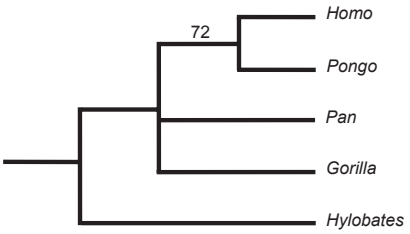

Figure 1 Bootstrap consensus tree for the evolutionary relationships among the large-bodied hominoids based on a character set extracted from the literature by Lehtonen *et al.* (2011) with removal of invalid characters where character states were unsubstantiated by the sources cited, contradicted by the sources cited, represented autapomorphic states, represented uniform character states for all ingroup taxa, or overlapped with the outgroup character states (and therefore provided no basis for determining whether the ingroup character state/s were primitive or derived). The 50% majority rule analysis gave a bootstrap value (of 72%) only for the human (*Homo*)–orangutan (*Pongo*) clade. This analysis shows that there was a sufficient number of problematic character states in the original data of Lehtonen *et al.* (2011) to change the result of their morphological analysis and to support the orangutan relationship. This result does not imply that any or all of the remaining characters added from the literature by Lehtonen *et al.* (2011) are necessarily valid or that their restriction of the outgroup to *Hylobates* alone provides a high level of confidence in any resulting phylogeny (see Grehan & Schwartz, 2009).

DISCUSSION

Character data presented in tabular form and with accompanying literature citation is often regarded as failsafe. Unfortunately it appears to be the case, especially when reaffirming the assumption that humans and chimpanzees are sister taxa, that such publications are too often taken literally and as reflecting reality. None of us can claim to be entirely exempt from critical review, regardless of the methodological approach we may embrace. As unpopular as the conclusions of our analyses may be, they result from efforts that are intended to be as unbiased and as transparent as possible, so that anyone taking the time to critique our data and its analysis can do so easily and with complete understanding of our methods.

Even though Lehtonen *et al.* (2011) did not question the validity of the published data they incorporated into their analysis, their criticism of our analytical approach caused us to test yet again the theory that humans and orangutans are sister taxa. In doing so, we have further demonstrated that the morphological (including metric) datasets Lehtonen *et al.* (2011) used are seriously flawed and in need of re-evaluation. Contrary to Lehtonen *et al.*'s (2011) contention, our reanalysis indicates that their characters do not have the veracity to support a robust hypothesis of relationship at this most critical node between humans and the great apes. Although the result will remain problematic because of restricted outgroup comparison, further evaluation of all characters in their dataset is not only desirable but necessary, because these features have been widely accepted as 'real' and the conclusion of studies based on them as 'true'.

We can, however, agree with Lehtonen *et al.* (2011) on another matter: namely, that the prevalent use in hominoid systematics of quantitative rather than qualitative traits is highly problematic and may result in inconsistent character coding. Indeed, demonstration of sub-equally long first and second lower premolars in humans and also in chimpanzees obscures the fact that in shape and morphological detail chimpanzee premolars do not look anything like human

premolars (Schwartz, 2005). Further, although Lehtonen *et al.* (2011) continue this practice, we must point out that one cannot embrace as synapomorphic for a chosen ingroup measurements that overlap with or subsume those of the outgroup, because it is not possible to identify what is derived for the ingroup state.

Because overlapping character states do not a priori provide evidence of being derived for an ingroup, we removed from our analysis the appropriate 47 of 89 instances of such ambiguity. Unfortunately we could not investigate these 47 measurements because our request for the original data was met with the authors' response that they had lost this information.

In addition to the morphological evidence that yielded a human–orangutan sister relationship we (Grehan & Schwartz, 2009) addressed conceptual and methodological assumptions that may lead to an erroneous molecular hypothesis of relationship even where there is a high degree of molecular similarity. In addition to the theoretical as well as methodological question of whether molecular data are sufficient to reject or even corroborate a theory of relationship derived from morphological data (also see Schwartz, 2005, in press a), we pointed specifically to the lack of outgroup sampling and exclusion of the orangutan in molecular analyses of human–great ape relationships. Further, we questioned whether interchangeable bases actually represent alternative character states and whether any method of sequence alignment, which creates non-empirically demonstrated homology via phenetics techniques of best fit, can generate a cladistically relevant result (see also Giribet *et al.*, 2002; De Laet, 2005; Redelings & Suchard, 2005; Phillips, 2006; Kjer *et al.*, 2007).

Lehtonen *et al.* (2011) responded to our molecular critique by asserting that we claimed DNA sequences must always be aligned before analysis. While acknowledging (as we did) that prior alignment was common practice, Lehtonen *et al.* (2011) presented a direct optimization method (DOM) as the solution to the alignment problem: that is, alignments and phylogenetic trees are considered to be hypotheses that are solved by searching for the most parsimonious combination of the two. With this assertion they proceeded to combine a molecular analysis that supported a human–chimpanzee relationship with their uncorrected morphological dataset, which not unexpectedly produced a human–chimpanzee sister grouping.

Perhaps the DOM may justify dismissing our previous critique of DNA sequence analyses if it indeed represents a cladistically valid solution. Some molecular theorists, however, have raised serious concerns about the basis of any approach to sequence alignment (e.g. Lake, 1991). In an appraisal of molecular homology and sequence alignment by direct optimization, Morgan & Kelchner (2010) recently concluded that the test of congruence alone cannot be used to determine homologous (historically identical) features. Acknowledging that earlier DOM critiques of the parsimony criterion rendered hypotheses of homology cladogram-dependent rather than grounded in causal relationships or phylogeny-independent factors (e.g. underlying molecular processes), Morgan & Kelchner (2010) conclude that the test of congruence alone cannot be used to determine homologous (historically identical) features and that character analysis and the proposition of primary homology [where a proposition of homology is first generated and then evaluated through the test of congruence – as in Grehan & Schwartz (2009)] is the only procedure that can identify correspondences between features that can be justified as retaining phylogenetic information. Morgan & Kelchner (2010) also conclude (p. 310) that 'the results of a DOM analysis should not be treated automatically as a phylogenetic estimate, but as one possible mathematical optimization of sequence data represented as a tree-like diagram that *may or may not resemble the historical relationships of the incorporated taxa* [our emphasis]'.

Although the molecular assumption – greater sequence similarity equates with closeness of relatedness – renders any generated hypothesis unfalsifiable (because it is internally consistent), it is also at odds with the assumptions that permit rooting a tree in a taxon that is a priori taken as both branching early and yet remaining primitively unchanged in the sequence(s) analysed (Schwartz, 2005, in press b). But while it seemed appropriate in the 1960s when Zuckerkandl & Pauling (1962) first articulated the molecular assumption to transfer observations of continual molecular change in bacteria to metazoans, it is now clear that there are profound differences between bacterial and metazoans genomes: namely, 97–98% of a bacterial genome is coding (i.e. codes for metabolically active proteins and enzymes) while only 2–3% is non-coding or regulatory (but able to mutate back to its original state), whereas the opposite describes metazoans (Eisen, 2000). Consequently, as the molecular literature makes clear (e.g. Ruvolo, 1997; Stauffer *et al.*, 2001; O'hUigin *et al.*, 2002; Perelman *et al.*, 2011), not only are molecular comparisons limited to minuscule portions of genomes, most are still confined to the coding region – which means that claims of '98% DNA similarity between humans and chimpanzees' is actually a claim of similarity in a small portion of 2–3% of the genome (see Schwartz, in press b) that codes for metabolically active proteins and enzymes, which, as in bacteria, reflect adaptation to environmental circumstances, not *de facto* synapomorphy (Schwartz & Maresca, 2007). Similarly, because mitochondria serve only metabolic function (Scheffler, 2000), demonstration of similarity between taxa in mtDNA sequence is also not necessarily a reflection of phylogenetic propinquity.

Comparisons, such as they are, that have been attempted in the non-coding regions where a protein product/transcription factor cannot corroborate the DNA sequence underlying it (e.g. in BRCA1 or FOXP2 genes) do not acknowledge additional factors that will confound strict DNA sequence comparisons, e.g. RNA-mediated intron splicing (Ast, 2005), transduction pathway signalling (e.g. Tarchini *et al.*, 2006), and the correlation between the three-dimensional structure of a non-coding region and it subsequent function (Parker *et al.*, 2009). This also underscores the too-long touted disconnect between 'molecular systematics' and the realities of biology, which continually demonstrate the developmental continuum from molecular interaction to the unfolding of morphological form (e.g. Davidson & Erwin, 2006).

There are evidently methodological and conceptual problems in molecular analyses that are recognized within the molecular research community, many of which are highlighted by the incongruence between the morphologically supported human–orangutan relationship and the molecular supported human–chimpanzee relationship. A hypothetico-deductive approach would require Lehtonen *et al.* (2011) to question both the morphological studies they cite (the 'data' as well as the analysis of it), and the 'molecular assumption' underlying the interpretation of DNA sequence data (cf. Schwartz & Maresca, 2007). Rethinking the latter assumption is particularly critical because, from the perspective of the philosophy of science, phylogenies that emerge from its application defy falsification (Schwartz, 2008; Grehan & Schwartz, 2009).

Much of the remainder of the non-coding region of the genome is involved in signalling pathways that govern development (i.e. developmentally regulated genes) (see Schwartz & Maresca, 2007). As such, sequencing exons or developmentally regulated genes gets us no closer to deciphering phylogenetic relationships because, while a 'gene' may have an identifiable 'start' and 'stop' codon, a gene comes into functional existence only after assuming a specific three-dimensional form (Parker *et al.*, 2009) after which RNA-mediated intron splicing and 'sense' and/or 'antisense' transcription has taken place (Ast, 2005; see also Rosenblatt *et al.*, 1997). Thus while humans and echinoderms may share similar regulatory molecules (e.g. distalless, engrailed, and orthodenticle), demonstration of this similarity is not in and of itself significant; rather the significance lies in how, when and where these regulatory genes are recruited, which leaves humans as bilaterally symmetrical organisms while transforming bilateral echinoderm larvae into radially symmetrical adults (Lowe & Wray, 1997).

Thus to demonstrate that chimpanzees and humans possess a FOXP2 gene does not demonstrate, as Lehtonen *et al.* (2011) suggest, a close relationship between these primates. Rather, FOXP2 is functional only in human development and not in chimpanzees. Until the presence/absence and/or active/inactive state of this or any other developmentally regulated gene can be compared across a wide array of taxa, just citing its presence in two taxa does not translate into a demonstration of biological (developmental) or phylogenetic relevance. As developmental biology is increasingly demonstrating, there is a continuum from the molecular biology of the zygote (even gametes) to that of the adult, rather than disconnected molecular and morphological realms (see Gerhart & Kirschner, 1997; Ronshaugen *et al.*, 2002; Davidson & Erwin, 2006; Stern *et al.*, 2006).

With these morphological and molecular issues in mind, we conclude that the human–orangutan relationship remains the most highly corroborated of any proposed morphological theory of human–ape relationship. Recognition of the developmental necessity of a continuum from the molecular to the morphological is crucial not only for its relevance to pursuing the systematics of metazoans with an imperfect or virtually unknown fossil record, but also for those taxa with a documented fossil record, because it is only through their morphology that living and extinct taxa can be directly compared. The orangutan evidence provides a coherent and integrated theory of relationship that is also consistent with the fact that hominid fossils, particularly the australopiths, display derived orangutan, rather than chimpanzee, features (Schwartz, 2004). In addition, the distributional record for humans, fossil hominids, and orangutans and their fossil relatives generates a theory of vicariant (or allopatric) differentiation that explains the origin of taxa between Africa, Europe and Asia without requiring pervasive migrations back and forth that belie the underlying vicariant origin of these taxa.

John R. Grehan[1] and Jeffrey H. Schwartz[2]

[1]*Buffalo Museum of Science, 1020 Humboldt Parkway, Buffalo, NY 14211-1293, USA,*
[2]*Departments of Anthropology, and History and Philosophy of Science, University of Pittsburgh, Pittsburgh, PA 15260, USA*
E-mail: jgrehan@sciencebuff.org

REFERENCES

Ast, G. (2005) The alternative genome. *Scientific American*, **April**, 59–65.

Collard, M. & Wood, B. (2000) How reliable are human phylogenetic hypotheses? *Proceedings of the National Academy of Sciences USA*, **97**, 5003–5006.

Davidson, E.H. & Erwin, D.H. (2006) Gene regulatory networks and the evolution of animal body plans. *Science*, **311**, 796–800.

De Laet, J.E. (2005) Parsimony and the problem of inapplicables in sequence data. *Parsimony, phylogeny and genomics* (ed. by V.A. Albert), pp. 81–116. Oxford University Press, Oxford.

Eisen, J.A. (2000) Assessing evolutionary relationships among microbes from whole-genome analysis. *Current Opinion in Microbiology*, **3**, 475–480.

Eldredge, N. & Cracraft, J. (1980) *Phylogenetic patterns and the evolutionary process*. Columbia University Press, New York.

Gerhart, J. & Kirschner, M. (1997) *Cells, embryos, and evolution: toward a cellular and developmental understanding of phenotypic variation and evolutionary adaptability*. Blackwell, Malden, MA.

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., Wheeler, W.C. & Muona, J. (2002) DNA multiple sequence alignments. *Molecular systematics and evolution: theory and practice* (ed. by R. DeSalle, G. Giribet and W. Wheeler), pp. 107–114. Birkhäuser Verlag, Switzerland.

Grehan, J.R. & Schwartz, J.H. (2009) Evolution of the second orangutan: phylogeny and biogeography of hominid origins. *Journal of Biogeography*, **36**, 1823–1844.

Huxley, T.H. (1863) *Man*'s place in nature. Appleton & Company, New York.

Kjer, K.M., Gillespie, J.J. & Ober, K.A. (2007) Opinions on multiple sequence alignment, and an empirical comparison of repeatability and accuracy between optimization and structural alignments. *Systematic Biology*, **56**, 133–146.

Lake, J.A. (1991) The order of sequence alignments can bias the selection of tree topology. *Molecular Biology and Evolution*, **8**, 378–385.

Lehtonen, S., Sääksjärvi, I., Ruokolainen, K. & Tuomisto, H. (2011) Who is the closest extant cousin of humans? Total-evidence approach to hominid phylogenetics via simultaneous optimization. *Journal of Biogeography*, **38**, 805–808.

Lockwood, C.A., Kimbel, W.H. & Lynch, J.M. (2004) Morphometrics and hominoid phylogeny: support for a chimpanzee–human clade and differentiation among great ape species. *Proceedings of the National Academy of Sciences USA*, **101**, 4356–4360.

Lowe, C.J. & Wray, G.A. (1997) Radical alterations in the roles of homeobox genes during echinoderm evolution. *Nature*, **389**, 718–729.

Morgan, M.J. & Kelchner, S.A. (2010) Inference of molecular homology and sequence alignment by direct optimization. *Molecular Phylogenetics and Evolution*, **56**, 305–311.

O'hUigin, C., Satta, Y., Takahata, N. & Klein, J. (2002) Contribution of homoplasy and of ancestral polymorphism to the evolution of genes in anthropoid primates. *Molecular Biology and Evolution*, **19**, 1501–1513.

Parker, S.C.J., Hansen, L., Abaan, H.O., Tullius, T.D. & Margulies, E.H. (2009) Local DNA topography correlates with functional noncoding regions of the human genome. *Science*, **324**, 389–392.

Perelman, P., Johnson, W.E., Roos, C., Seuánez, H.N., Horvath, J.E., Moreira, M.A.M., Kessing, B., Pontius, J., Roelke, M., Rumpler, Y., Schneider, M.P.C., Silva, A., O'Brien, S.J. & Pecon-Slattery, J. (2011) A molecular phylogeny of living primates. *PLoS Genetics*, **7**, 1–17.

Phillips, C.A. (2006) Homology assessment and molecular sequence alignment. *Journal of Biomedical Informatics*, **39**, 18–33.

Redelings, B.D. & Suchard, M.A. (2005) Joint Bayesian estimation of alignment and phylogeny. *Systematic Biology*, **54**, 401–418.

Ronshaugen, M., McGinnis, N. & McGinnis, S.W. (2002) Hox protein mutation and macroevolution of the insect body plan. *Nature*, **415**, 914–917.

Rosenblatt, K.P., Sun, Z.P., Heller, S. & Hudspeth, A.J. (1997) Distribution of Ca^{2+}-activated K^+ channel isoforms along the tonotopic gradient of the chicken's cochlea. *Neuron*, **19**, 1061–1075.

Ruvolo, M. (1997) Molecular phylogeny of the hominoids: inferences from multiple independent DNA sequence datasets. *Molecular Biology and Evolution*, **14**, 248–265.

Scheffler, I.E.A. (2000) A century of mitochondrial research: achievements and perspectives. *Mitochondrian*, **1**, 3.

Schwartz, J.H. (1984a) The evolutionary relationships of man and orang-utans. *Nature*, **308**, 501–505.

Schwartz, J.H. (1984b) Hominid evolution: a review and a reassessment. *Current Anthropology*, **25**, 655–672.

Schwartz, J.H. (1986) Primate systematics and a classification of the order. *Comparative primate biology,* Vol. 1: *Systematics, evolution, and anatomy* (ed. by D.R. Swindler and J. Erwin), pp. 1–41. Alan R. Liss, New York.

Schwartz, J.H. (1988) History, morphology, paleontology, and evolution. *Orang-utan biology* (ed. by J.H. Schwartz), pp. 59–85. Oxford University Press, New York.

Schwartz, J.H. (1997) *Lufengpithecus* and hominoid phylogeny: problems in delineating and evaluating phylogenetically relevant characters. *Function, phylogeny, and fossils: Miocene hominoid evolution and adaptations* (ed. by D.R. Begun, C.V. Ward and M.D. Rose), pp. 363–388. Plenum Press, New York.

Schwartz, J.H. (2004) Barking up the wrong ape – australopiths and the quest for chimpanzee characters in hominid fossils. *Collegium Antropologicum*, **28**(Suppl. 2), 87–101.

Schwartz, J.H. (2005) *The Red Ape: orang-utans and human origins*, 2nd revised edn. Westview Press, Boulder, CO.

Schwartz, J.H. (2008) Cladistics. *Icons of evolution* (ed. by B. Regal), pp. 517–544. Greenwood Press, Westport, CT.

Schwartz, J.H. (in press a) Molecular systematics and evolution (revised and updated). *Encyclopedia of molecular cell biology and molecular medicine (EMCBMM)* (ed. by R.A. Meyer). Wiley-VCH Verlag, Weinheim.

Schwartz, J.H. (in press b) Organismal biology, molecular systematics, and phylogenetic reconstruction. *Leaping ahead: advances in prosimian biology* (ed. by J. Masters, M. Gamba and F. Génin). Springer Science, New York.

Schwartz, J.H. & Maresca, B. (2007) Do molecular clocks run at all? A critique of molecular systematics. *Biological Theory*, **1**, 357–371.

Sibley, C.G. & Ahlquist, J.E. (1984) The phylogeny of the hominoid primates, as indicated by DNA-DNA hybridization. *Journal of Molecular Evolution*, **20**, 2–15.

Stauffer, R.L., Walker, A., Ryder, O.A., Lyons-Weiler, M. & Hedges, S.B. (2001) Human and ape molecular clocks and constraints on paleontological hypotheses. *The Journal of Heredity*, **92**, 469–474.

Stern, C.D., Charité, J., Deschamps, J., Duboule, D., Durston, A.J., Kmita, M., Nicolas, J.-F., Palmeirem, I., Smith, J.C. & Wolpert, L. (2006) Head-tail patterning of the vertebrate embryo: one, two or many unresolved problems? *International Journal of Developmental Biology*, **50**, 3–15.

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.

Swofford, D.L. (2005) *PAUP*: phylogenetic analysis using parsimony (*and other methods). Version 4*. Sinauer Associates, Sunderland, MA.

Tarchini, B., Duboule, D. & Kmita, M. (2006) Regulatory constraints in the evolution of the tetrapod limb anterior–posterior polarity. *Nature*, **443**, 985–988.

Thiele, K. (1993) The holy grail of the perfect character: the cladistic treatment of morphometric data. *Cladistics*, **9**, 275–304.

Zuckerkandl, E. & Pauling, L. (1962) Molecular disease, evolution and genetic heterogeneity. *Horizons in biochemistry* (ed. by M. Kasha and B. Pullman), pp. 189–225. Academic Press, New York.

SUPPORTING INFORMATION

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

Appendix S1 Phenotypic data matrix used to analyse phylogenetic relationships among the great apes.

Appendix S2 Explanation of the rationale for removal of characters from the original data matrix of Lehtonen *et al.* (2011).

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

doi:10.1111/j.1365-2699.2011.02577.x