6-7-1998

Enigmatic phylogeny of skuas: An alternative hypothesis

M. J. Braun
Smithsonian National Museum of Natural History

R. T. Brumfield
Smithsonian National Museum of Natural History

Recommended Citation

Braun, M., & Brumfield, R. (1998). Enigmatic phylogeny of skuas: An alternative hypothesis. *Proceedings of the Royal Society B: Biological Sciences, 265* (1400), 995-999. https://doi.org/10.1098/rspb.1998.0389

The Royal Society

Enigmatic phylogeny of skuas: an alternative hypothesis

Michael J. Braun* **and Robb T. Brumfield**

Laboratory of Molecular Systematics, National Museum of Natural History, Smithsonian Institution MRC 534, Washington, DC 20560, USA, and Department of Zoology, University of Maryland, College Park, MD 20742, USA

Last year, Cohen *et al*. presented molecular data suggesting the surprising result that both currently recognized genera of skuas, *Stercorarius* and *Catharacta* (Aves: Stercorariidae), are not monophyletic. However, the most enigmatic conclusion from their analysis, that *S. pomarinus* is sister to *C. skua*, rests solely on mtDNA sequence data. When the mtDNA data are analysed in a maximum likelihood framework that accounts for variation in evolutionary rates, *Catharacta* monophyly cannot be rejected. None of the best trees that can be derived from two nuclear data sets of Cohen *et al*. support the controversial *pomarinus*–*C. skua* node. We propose an alternative hypothesis, that *pomarinus* is sister to a monophyletic *Catharacta*, as the best explanation of the available molecular, morphological, and behavioural evidence.

Keywords: Stercorariidae; skua; phylogeny; lineage sorting; mtDNA; avian systematics

1. INTRODUCTION

Apparent conflicts between molecules and morphology in systematics have become almost commonplace (Balter 1997), sparking heated debate in many cases (e.g. Poe 1996; Lee *et al*. 1997; Sullivan & Swofford 1997). This debate is healthy because it stimulates additional research and new analyses that are generating fresh insights in a broad array of evolutionary disciplines.

A recent case in point involves the skuas, a distinctive family of predatory seabirds. Cohen *et al*. (1997) present molecular data bearing on skua phylogeny that pose a striking enigma. The data indicate that one of the three *Stercorarius* skuas, *S. pomarinus* (the Pomarine skua), falls within the clade comprising the *Catharacta* skuas, rendering both genera non-monophyletic (figure 1*a*). Yet these two genera are (i) quite different in outward appearance, (ii) defined by putative morphological synapomorphies (Brooke 1978; Furness 1987), and (iii) have been maintained as distinct genera in the overwhelming majority of modern treatises (e.g. AOU 1983). Cohen *et al*. (1997) suggest three possible explanations for the remarkable discordance between molecules and morphology in skuas, but recognize that all three explanations are 'far-fetched'. Our purpose here is to explore this apparent paradox, and to propose a fourth hypothesis. We believe this fourth explanation best accounts for the available data, which now include plumage and skeletal morphology, behaviour, ectoparasitic lice, and a variety of molecular evidence. This explanation has the further advantage that it is readily testable.

Cohen *et al*. (1997) present five data sets, four of which are molecular, the fifth of which is based on ectoparasitic lice. The data are consistent in many ways, and demonstrate convincingly that *pomarinus* is more closely related to *Catharacta* than to other *Stercorarius*. While this is surprising in the light of the general resemblance of *pomarinus* to other *Stercorarius*, this result does not actually contradict the morphological data. As Cohen *et al*. point out, the *Stercorarius* morphotype appears to be ancestral in the Stercorariidae; therefore, it does not provide cladistic information supporting monophyly of *Stercorarius*. The single putative morphological synapomorphy linking *pomarinus* with *Stercorarius*, barred juvenile plumage (Brooke 1978), may instead be a symplesiomorphy in the family that has been lost in *Catharacta*.

What is more difficult to accept is the idea that *Catharacta* is also not monophyletic. The phylogeny presented by Cohen *et al*. places *pomarinus* sister to *C. skua* (Great skua), to the exclusion of the other five *Catharacta* species (figure 1*a*). This aspect of their phylogeny forces the authors to propose remarkable convergence in the origins of either the *Catharacta* or *Stercorarius* morphology (their hypotheses (*a*) and (*b*)) or a bizarre inter-generic hybridization event resulting in a stable hybrid species (*pomarinus*) that is *Catharacta*-like in its mitochondrial and nuclear genomes but *Stercorarius*-like in its external appearance (their hypothesis (*c*)). It is this node which causes the bulk of the enigma they face.

Although Cohen *et al*. discuss five data sets, the phylogeny they present is actually derived from only one, the mitochondrial DNA (mtDNA) sequences. The mtDNA is a single, clonally inherited genetic unit. Any phylogeny based upon it must be regarded as the phylogeny of a single gene, which may or may not accurately track the species phylogeny (Nei 1987, p. 288; Maddison 1997). One set of conditions under which a gene phylogeny is likely to differ from the species phylogeny is when a series of speciation events occurs with insufficient time between them for the gene lineages to reach reciprocal monophyly. In this situation, random fixation of gene lineages in the daughter species can result in a gene tree that is incongruent with the

*Author for correspondence (braun@onyx.si.edu).

Received 26 January 1998 *Accepted* 3 March 1998

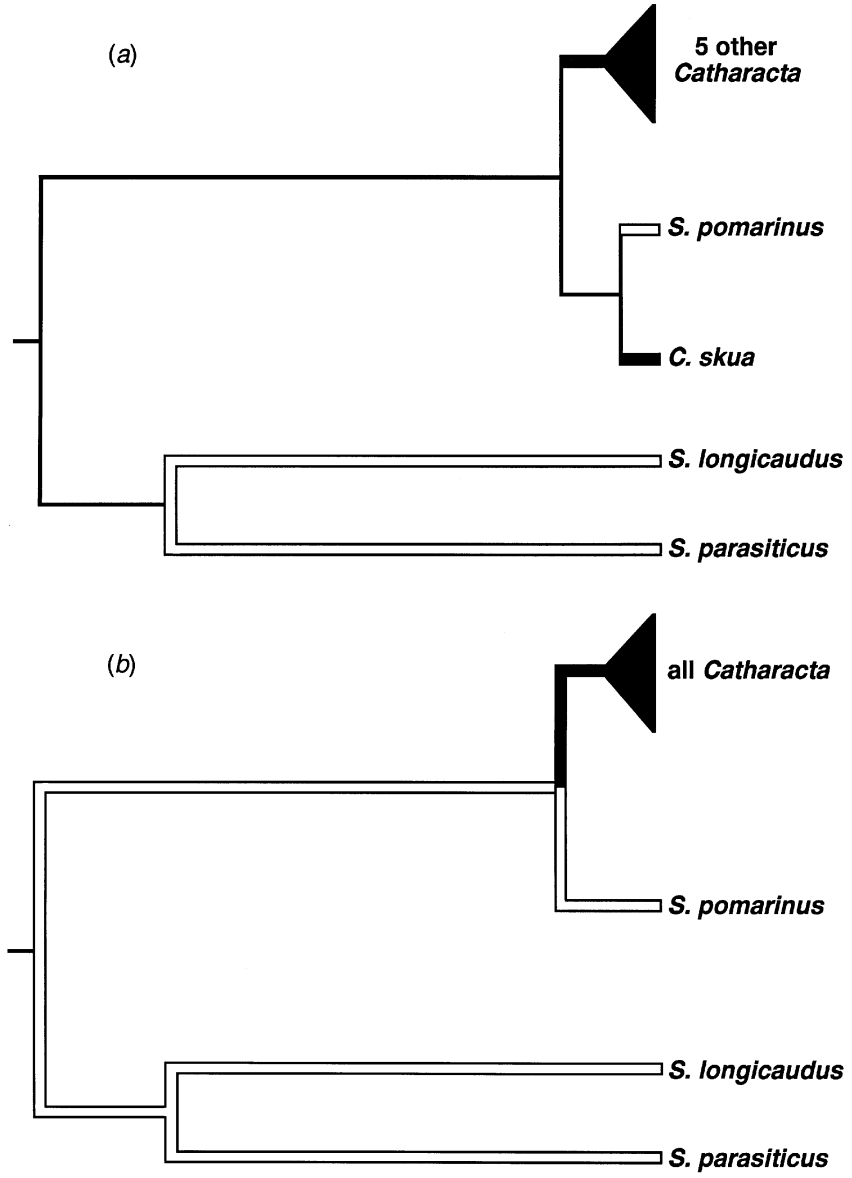

Figure 1. Hypotheses of phylogenetic relationship among skuas. Heavy black branches are those on which a *Catharacta*-like plumage morphology can be inferred. Open branches are those on which a *Stercorarius*-like plumage morphology can be inferred. Branch lengths are not proportional to evolutionary change. (*a*) Major features of the topology proposed by Cohen *et al.* (1997) based on mtDNA sequences. Either the *Stercorarius* or the *Catharacta* morphology must have evolved twice on this tree. (*b*) Alternative topology proposed herein. Each morphology need only evolve once on this tree.

species tree. The small divergences among *Catharacta* species and *pomarinus* in all four genetic data sets (see Cohen 1997; note branch lengths in figure 2 herein) indicate that speciation has been rapid and relatively recent in skuas. Thus, it seems plausible that reciprocal monophyly of mtDNA may not have been reached in some lineages of the group, and that the mtDNA tree, however well-resolved it might be, simply is not the same as the species tree.

We propose an alternative hypothesis of relationship that accommodates both the morphology and the data of Cohen *et al.* (figure 1*b*). In this hypothesis, *pomarinus* is sister to a monophyletic genus *Catharacta*, reflecting the close relationship so strongly indicated by a wealth of molecular evidence. The *pomarinus*–*C. skua* node of figure 1*a* is presumed incorrect, either due to the gene tree–species tree problem discussed above or to inadequate resolution of the mtDNA tree (see below). Although *Stercorarius* is not monophyletic, figure 1*b* only requires each morphotype to evolve once (contrary to hypotheses (*a*) and (*b*) of Cohen *et al.*; see figure 1*a*). The resemblance of *pomarinus* to other *Stercorarius* is explained by assuming that *pomarinus* has retained the ancestral *Stercorarius* morphotype.

This hypothesis also accounts for the similarity of *pomarinus* to *Catharacta* observed in a phenetic study of 50 skeletal characters (Schnell 1970), as well as behavioural similarities in calls and displays (Andersson 1973). In fact, Andersson seems to have proposed essentially the same hypothesis of relationship entailed in figure 1*b* when he wrote 'the most likely explanation for this [behavioural similarity] seems to be that the Pomarine and Great skuas diverged from each other at a time when the predecessor of the two smaller species had already branched from the common skua ancestor' (Andersson 1973, p. 14). Although he did not explicitly state that he considered *Catharacta* monophyletic, Andersson treated the South Polar skua (*C. s. maccormicki*) as a subspecies of the Great skua (*C. s. skua*) in the same section, so the idea seems implicit.

2. METHODS, RESULTS AND DISCUSSION

(a) *Mitochondrial data*

Given the two competing hypotheses of relationship contained in figures 1*a* and 1*b*, it is natural to ask whether the mtDNA sequence data support one hypothesis significantly more than

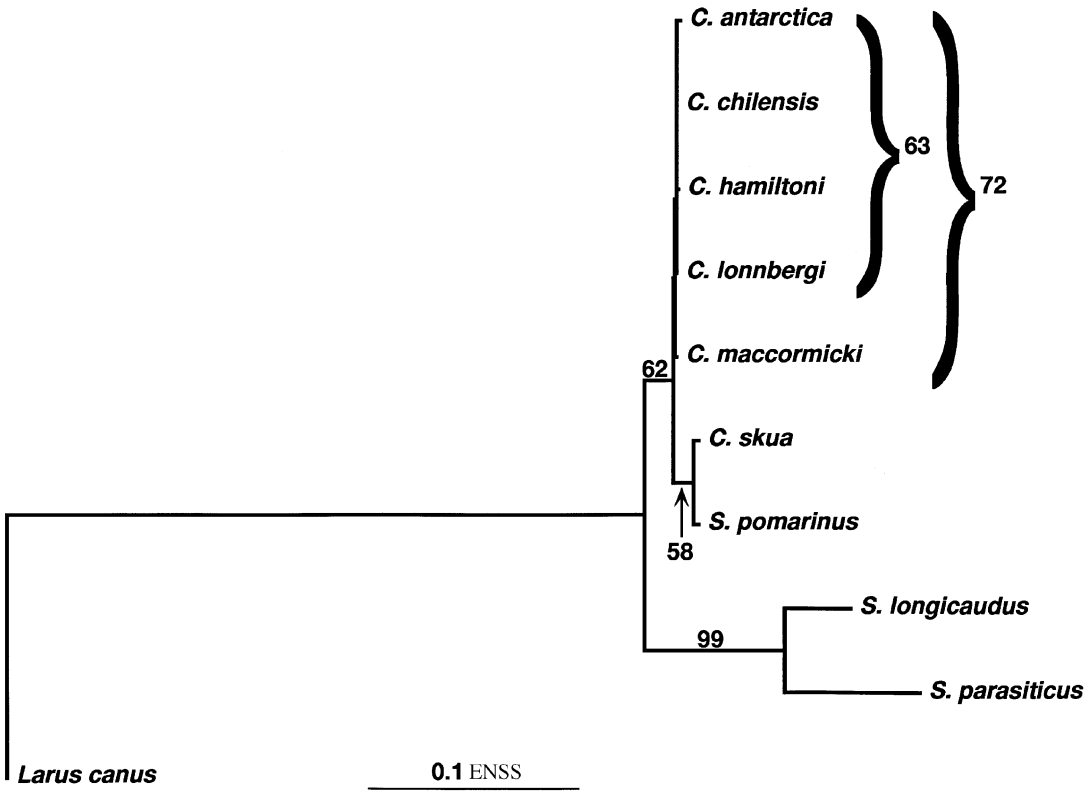

Figure 2. Maximum-likelihood estimate of skua phylogeny based on mtDNA sequence data of Cohen *et al.* (1997). Numbers refer to percentage of 100 unconstrained likelihood bootstrap replicates in which each node occurs. Only values greater than 50% are shown. Branch lengths are proportional to the expected number of substitutions per site (ENSS). Likelihood parameters for the general time reversible (GTR) substitution model (Lanave *et al.* 1984) with among-site rate heterogeneity following a gamma distribution (Γ; Yang 1993) were estimated separately on the cytochrome *b* (cyt *b*) and 12S ribosomal RNA (12S) data. The GTR+Γ likelihood model was used because it fit the data significantly better based on likelihood ratio tests (Goldman 1993) than other models available in PAUP* (e.g. GTR+Γ versus HKY85+Γ, $\chi^2=21.02$, d.f.$=5$, $p<0.05$). Because the estimated parameters (i.e. rate matrix and gamma-shape parameter) were very similar for both genes, cyt *b* and 12S sequences were combined. Using all 1415 bp of sequence data, likelihood parameters for the GTR+Γ model were estimated using the successive approximations approach suggested by Swofford *et al.* (1996, p. 445), beginning with a neighbour-joining tree (Saitou & Nei 1987) constructed from HKY85 (Hasegawa *et al.* 1995) genetic distances assuming no among- site rate heterogeneity. The parameter estimates used were rate matrix: A–C$=2.329\times10^{6}$, A–G$=1.826\times10^{7}$, A–T$=8.528\times10^{5}$, C–G$=7.613\times10^{-2}$, C–T$=3.455\times10^{7}$, G–T$=1$; gamma-shape parameter, $\alpha=0.0974665$.

the other. The advantages of model-based maximum-likelihood methods of phylogenetic analysis in this regard have been discussed (see Swofford *et al.* 1996; Huelsenbeck & Rannala 1997 and references therein). To assess whether the mitochondrial sequence data are significantly more likely under the topology of figure 1*a* than 1*b*, we performed a Kishino–Hasegawa test (KH test; Kishino & Hasegawa 1989) on maximum-likelihood trees with these topologies. Unless otherwise specified, all phylogenetic analyses were performed with PAUP* v. 4.0 (Swofford 1998). An unconstrained branch-and-bound search found a most likely tree (log-likelihood score$=-3259.52$; figure 2) identical in topology to the maximum-parsimony tree of Cohen *et al.* Using the same model and parameter estimates, a branch-and-bound search in which *Catharacta* was constrained to be monophyletic also found a most likely tree (score$=-3262.40$). A KH test indicates that this tree is not significantly worse than the most likely tree from the unconstrained search as an explanation of the data ($T=0.8883$, $p=0.3745$). One hundred unconstrained, likelihood bootstrap replicates using the above parameters (with heuristic search, addseq=as-is) also demonstrate less confidence in the enigmatic *pomarinus–C. skua* node than was found by Cohen *et al.* in an unweighted parsimony framework (58% versus 97%; figure 2). To understand why this node is less robust under maximum likelihood, it is important to note that the branch lengths in the *pomarinus–Catharacta* clade are all quite short. This results from the fact that the maximum-likelihood model accounts for the substantial variation among sites in evolutionary rates present in this data set (see figure 2 legend).

Although we believe model-based likelihood methods are better suited for phylogenetic analysis of sequence data, in the interest of completeness we also performed KH and Templeton (1983) tests on optimal trees from unconstrained and constrained unweighted parsimony analyses analogous to those presented by Cohen *et al.* Our analysis of 102 informative sites resulted in a single most parsimonious tree (145 steps) identical in topology to that presented by Cohen *et al.* Discrepancies in the number of informative sites and tree length from that presented in Cohen *et al.* (109 informative sites, tree length 148) are due to an improvement in the ability of PAUP* to detect uninformative characters in certain situations involving polymorphic terminal taxa. This improvement does not affect their phylogenetic conclusions. A search in which *Catharacta* was constrained monophyletic resulted

in a single most parsimonious tree (151 steps) that placed *pomarinus* basal to the *Catharacta* clade. The KH test found the unconstrained tree to be better than the constrained tree ($T=2.1591$, $p=0.0332$). There was no significant difference as judged by the Templeton test ($z=1.8904$, n.s.).

(b) *Nuclear data*

Thus, maximum-likelihood methods indicate that the tree of figure 1*a* is not a significantly better explanation of the mtDNA data than the tree of figure 1*b*, and parsimony methods indicate that it is at best marginally so. However, even if there was a significant difference, the gene tree–species tree problem would remain. Whether a particular gene tree is congruent with the species tree can best be tested by independent estimates of the species tree based on other data (i.e. other genes, morphology, behaviour, etc.). Cohen *et al.* describe two nuclear data sets (allozymes, RAPDs) that can be used for this purpose. However, they do not present trees based on these data, concluding instead that the nuclear gene data strongly confirm the close relationship of *pomarinus* to *Catharacta*, but 'provide no critical evidence about the ancestry of the Pomarine skua . . . '.

Reanalysis of the nuclear data sets reveals that the best trees that can be derived from either of them supports the close relationship of *pomarinus* to *Catharacta*, but contradicts the *pomarinus–C. skua* node of the mtDNA tree. That this might be true can be seen by inspection of the distance matrices (Cohen 1997). The RAPD distance matrix indicates that *C. skua* and *C. maccormicki* are more similar to each other than either is to *pomarinus* (*skua* and *maccormicki* are the only *Catharacta* taxa in the RAPD data matrix). Minimum evolution (Kidd & Sgaramella-Zonta 1971) and FM (Fitch & Margoliash 1967) analyses confirm this similarity, yielding a tree with *Catharacta* monophyletic. Parsimony analysis of the RAPD data yields four most parsimonious trees, all of which have *C. maccormicki* sister to *pomarinus*. The shortest trees (90 steps) are only one step shorter than a tree in which *Catharacta* is monophyletic. Thus, neither distance nor parsimony trees derived from the RAPD data contain a *pomarinus–C. skua* node.

Using BIOSYS (Swofford & Selander 1981), distance matrices (Cavalli-Sforza & Edwards 1967; Nei 1972) were calculated from the original 42-locus allozyme data set provided by A. Baker (including allele frequency data for the outgroups *Larus fuscus*, *L. novaehollandiae* and *Sterna hirundo*). In each case, *C. skua* and *C. antarctica* are more similar to one another than either is to *pomarinus* (*skua*, *antarctica* and *maccormicki* are the only *Catharacta* taxa in the allozyme data matrix). In fact, one allele at the guanine deaminase (GDA) locus represents an unambiguous synapomorphy for *C. skua* and *C. antarctica*. Given that the mtDNA data are also derived from a single genetic unit, we consider GDA by itself to be an important conflict for the mtDNA tree of figure 1*a*.

A frequency parsimony analysis of the allozyme data was performed by converting the BIOSYS file to a FREQPARS file using the FORTRAN program BIO2FREQ (Swofford & Berlocher 1987; program available via anonymous FTP at `onyx.si.edu`). The FREQPARS file was imported into PAUP* and a branch-and-bound search was performed (for details of the FREQPARS analysis utilizing PAUP*, see Berlocher & Swofford 1997). The three most parsimonious FREQPARS trees (49.916 steps) have *C. skua* sister to *C. antarctica*. A tree with *Catharacta* constrained monophyletic is only 1.1 steps longer (a FREQPARS step is equivalent to an allelic frequency change of 0.5). UPGMA (unweighted pair group method), neighbour-joining, and minimum evolution trees were also constructed using the chord distance (Cavalli-Sforza & Edwards 1967); all of these also have *C. skua* sister to *C. antarctica*.

3. CONCLUSIONS

(a) *Phylogeny*

In summary, none of the best trees derived from the nuclear data sets includes the *pomarinus–C. skua* node found in the mtDNA tree, and near optimal trees exist in each case in which *Catharacta* is monophyletic. While neither nuclear data set can be used to exclude confidently one or other of the topologies in question, the same can probably be said of the mtDNA data as explored above. Given the conflict between data sets and the gene tree–species tree problem, a more conservative interpretation of the total genetic evidence (mitochondrial and nuclear) would be that it supports a clade composed of *pomarinus* plus all *Catharacta*, but no further resolution within that clade is yet possible. When the several apparent morphological synapomorphies that unite *Catharacta* are considered (Furness 1987), it seems more reasonable to suppose that the group is monophyletic (figure 1*b*).

Distinguishing between these two hypotheses of relationship (figures 1*a* and 1*b*) should be straightforward. Any independent estimate of the species tree that resolves the nodes within the *pomarinus–Catharacta* clade will corroborate one of these hypotheses (or one of the other possible resolutions within the clade), and refute the rest. In principle, it would be possible to make such an estimate from morphological or behavioural data, if such information exists for all the forms of *Catharacta*. However, given that genetic samples from all named forms of *Catharacta* and *Stercorarius* are in hand in several laboratories, it seems most likely that independent estimates of the species tree will come from DNA sequences of nuclear genes.

(b) *Taxonomy*

The phylogeny in figure 1*b* makes *Stercorarius* non-monophyletic, while that of figure 1*a* renders both *Stercorarius* and *Catharacta* non-monophyletic. In either case, a different generic treatment of the group is required. One possibility is to treat all skuas in a single genus, *Stercorarius*, as recommended by Hartert (1912), Moynihan (1959) and Andersson (1973). However, if the topology of figure 1*b* proves correct, it would also be reasonable to retain *Stercorarius* and *Catharacta*, and place *pomarinus* in a separate genus. This treatment would have the advantage of recognizing the morphological distinctiveness that separates *pomarinus* from *Catharacta*, while highlighting the true phylogenetic structure of the group. The disadvantage of this arrangement is that it would erect separate genera for two groups (*pomarinus* and *Catharacta*) which, based on their genetic similarity, probably share a more recent common ancestor than do the two remaining species of *Stercorarius*. Accumulation of genetic and palaeontological data may make recency of common ancestry a desirable metric of categorical rank in future. At present, there is only a loose correlation between the two in avian taxonomy (e.g. Brumfield *et al.* 1997). The dual goals of recognizing phylogeny and morphological similarity take priority in current usage, and this second treatment is the

one we favour if figure 1*b* is the correct phylogeny. The generic name with priority for *pomarinus* under this scenario seems to be *Coprotheres* Reichenbach 1850.

We thank A. J. Baker, B. L. Cohen, P. Harvey, A. J. Helbig, J. P. Huelsenbeck, S. Steppan, and D. L. Swofford for helpful discussion and criticism of the manuscript. R. C. Banks advised us on the historical taxonomy and nomenclature of the Stercorariidae.

REFERENCES

AOU 1983 *Check-list of North American birds*, 6th edn. Washington, DC: American Ornithologists' Union.

Andersson, M. 1973 Behaviour of the Pomarine Skua, *Stercorarius pomarinus* Temm. with comparative remarks on Stercorariinae. *Ornis Scand.* **4**, 1–16.

Balter, M. 1997 Morphologists learn to live with molecular upstarts. *Science* **276**, 1032–1034.

Berlocher, S. H. & Swofford, D. L. 1997 Searching for phylogenetic trees under the frequency parsimony criterion: an approximation using generalized parsimony. *Syst. Biol.* **46**, 211–215.

Brooke, R. K. 1978 The *Catharacta* skuas (Aves: Laridae) occurring in South African waters. *Durban Mus. Nov.* **11**, 295–308.

Brumfield, R. T., Swofford, D. L. & Braun, M. J. 1997 Evolutionary relationships among the potoos (Nyctibiidae) based on isozymes. *Ornithol. Monogr.* **48**, 129–145.

Cavalli-Sforza, L. L. & Edwards, A. W. F. 1967 Phylogenetic analysis: models and estimation procedures. *Evolution* **32**, 550–570.

Cohen, B. L. 1997 Enigmatic phylogeny of skuas (Aves: Stercorariidae) (This is available at `http://www.ibls.-gla.ac.uk/ibls/staff/bl-cohen/skuas.html`. (October 22, 1997.)

Cohen, B. L. (and 15 others) 1997 Enigmatic phylogeny of skuas (Aves: Stercorariidae). *Proc. R. Soc. Lond.* B **264**, 181–190.

Fitch, W. M. & Margoliash, E. 1967 Construction of phylogenetic trees. *Science* **155**, 279–284.

Furness, R. W. 1987 *The skuas*. Staffordshire, England: T. & A. D. Poyser Ltd.

Goldman, N. 1993 Statistical tests of models of DNA substitution. *J. Molec. Evol.* **36**, 182–198.

Hartert, E. 1912 *Die Vogel der paläarktischen Fauna*. Berlin.

Hasegawa, M., Kishino, H. & Yano, T. 1985 Dating of the human–ape splitting by a molecular clock of mitochondrial DNA. *J. Molec. Evol.* **21**, 160–174.

Huelsenbeck, J. P. & Rannala, B. 1997 Phylogenetic methods come of age: testing hypotheses in an evolutionary context. *Science* **276**, 227.

Kidd, K. K. & Sgaramella-Zonta, L. A. 1971 Phylogenetic analysis: concepts and methods. *Am. J. Hum. Genetics* **23**, 235–252.

Kishino, H. & Hasegawa, M. 1989 Evaluation of the maximum-likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. *J. Molec. Evol.* **29**, 170–179.

Lanave, C., Preparata, G., Saccone, C. & Serio, G. 1984 A new method for calculating evolutionary substitution rates. *J. Molec. Evol.* **20**, 86–93.

Lee, K., Feinstein, J. & Cracraft, J. 1997 The phylogeny of ratite birds: resolving conflicts between molecular and morphological data sets. In *Avian molecular evolution and systematics* (ed. D. Mindell), pp. 173–208**.** San Diego, CA: Academic Press.

Maddison, W. P. 1997 Gene trees in species trees. *Syst. Biol.* **46**, 523–536.

Moynihan, M. 1959 A revision of the family Laridae (Aves). *Am. Mus. Novitates* **1928**, 1–42.

Nei, M. 1972 Genetic distance between populations. *Am. Nat.* **106**, 283–292.

Nei, M. 1987 *Molecular evolutionary genetics*. New York: Columbia University Press.

Poe, S. 1996 Data set incongruence and the phylogeny of the crocodilians. *Syst. Biol.* **45**, 393–414.

Saitou, N. & Nei, M. 1987 The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molec. Biol. Evol.* **4**, 406–425.

Schnell, G. D. 1970 A phenetic study of the suborder Lari (Aves), parts I–II. *Syst. Zool.* **19**, 35–57, 264–302.

Sullivan, J. & Swofford, D. L. 1997 Are guinea-pigs rodents? The importance of adequate models in molecular phylogenetics. *J. Mamm. Evol.* **4**, 77–86.

Swofford, D. L. 1998 *PAUP*, Beta test versions 56, 57, 59*. Sunderland, MA: Sinauer Associates.

Swofford, D. L. & Berlocher, S. H. 1987 Inferring evolutionary trees from gene frequencies under the principle of maximum parsimony. *Syst. Zool.* **36**, 293–325.

Swofford, D. L., Olsen, G. J., Waddell, P. J. & Hillis, D. M. 1996 Phylogenetic inference. In *Molecular systematics* (eds D. M. Hillis, C. Moritz & B. K. Mable), pp. 407–514. Sunderland, MA: Sinauer Associates.

Swofford, D. L. & Selander, R. B. 1981 BIOSYS-1: a FORTRAN program for the comprehensive analysis of electrophoretic data in population genetics and systematics. *J. Hered.* **72**, 281–283.

Templeton, A. R. 1983 Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the humans and apes. *Evolution* **37**, 221–244.

Yang, Z. 1993 Maximum likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites. *Molec. Biol. Evol.* **10**, 1396–1401.