



Forum

The AIDS Pandemic is New, but is HIV Not New?

Mark E. Siddall

Museum of Zoology, University of Michigan, Ann Arbor, Michigan 48109, U.S.A.

Accepted 17 June 1997

The determinations made by Mindell, Shultz and Ewald regarding the ancestral host for immunodeficiency retroviruses, and their conclusion that monkeys acquired their infections as a result of a host-switch from humans, do not withstand rigorous scrutiny. Their hypothesis requires the complete uniformity of third position transitions and of gapped regions in the alignment. When all of the data are permitted to corroborate or refute relationships, optimizing hosts on the viral phylogeny renders either equivocal statements or an unequivocal simian ancestry. However, merely optimizing hosts as characters on the viral phylogeny is illogical. Not only does this treat hosts as dependent on the viruses (instead of the reverse) but it ignores 15 years of methodological developments specifically designed to answer questions regarding cospeciation or host-switching. © 1997 The Willi Hennig Society

Acquired immunodeficiency syndrome (AIDS) was unknown prior to the early 1980s. When HIV was determined to be the cause of the disease, it was naturally assumed to be a “new virus” (new to infecting humans). It is, however, possible that HIV is an “old virus” which had been present in human populations and only recently became virulent. Using phylogenetic

methods, Mindell et al. (1995) concluded that HIV was not a “new virus” and, contrary to conventional belief, that monkeys acquired their retroviruses from humans. Knowledge about the history of the human immunodeficiency viruses (HIV1 and HIV2) can have a bearing on expected results for disease intervention strategies and for the expectations that vaccination regimes will be successful. Different assessment of the relative age of the origin of primate immunodeficiency viruses (PIV) will lead to different conclusions regarding mutation rates, and retroviruses with faster mutation rates are not good candidates for vaccine development (Atlan et al., 1994). Also there are sociological, anthropological and transmission considerations regarding the relative order of appearance of the immunodeficiency viruses. So far as is known, HIV transmission requires direct contact with virus-laden fluids (e.g. blood, blood products, and semen) for transmission to occur. The retroviral envelope, which is necessary for infection of cells, renders the virions susceptible to degradation. It is this Achilles heel, the envelope, which renders fomite transmission unlikely (because of desiccation) and vector-borne transmission routes nearly impossible (because of rapid digestion of lipids in the gastrointestinal tracts of blood-feeding dipterans or leeches). These elements of the host-parasite system are consistent with humans having recently acquired the virus from green monkeys. In light of this,

Correspondence to: M. E. Siddall, E-mail: msiddall@umich.edu,
Fax: (313) 663 4080.

the findings by Mindell et al. (1995: 83) are unusual in that “within the HIV2/SIVsm/SIVmac/SIVagm clade, human is also shown as the ancestral host species”, and that “the more parsimonious scenario indicates virus transmission *from* humans (HIV2s) *to* sooty mangabeys and macaques” (emphasis added). Humans do not regularly bite monkeys, or share needles with them, nor is there evidence of humans being hunted and eaten by arboreal green monkeys, macaques or mangabeys. This leaves us having to postulate alternative primate behavioural patterns regarding HIV transmission *from* humans to monkeys, or be concerned that mosquitoes could transmit the virus after all.

The conclusions drawn by Mindell et al. (1995) can be shown to be dependent on their perspective on the data and on the manner in which hosts were dealt with. Fig. 1 reproduces the most parsimonious tree found by Mindell et al. (1995) for PIV relationships. In it, HIV1 and HIV2 are hypothesized as having separate origins which is consistent with the present viral nomenclatural designations. The most parsimonious reconstruction of ancestral host type, when each host species is taken as a discrete characteristic of the virus (*sensu* Mindell et al., 1995) is that humans are the ancestral source of the virus. For each host species to be considered a characteristic of the virus requires strict host specificity. However, even the findings of Mindell et al. (1995) suggest a remarkable lack of host specificity. That is, with at least six host-switches (transformations on their tree), the viruses cannot be said to be very specific. Although the data were presented as individual isolates from individual species, there is no evidence that the three isolates from macaques (SIVstm, SIVmne, and SIVmm239) are necessarily restricted to the host species from which they were obtained. Nor has this been suggested for the five viruses infecting *Cercopithecus* species (SIVsyk, SIVagm677, SIVagm155, SIVagm3, and SIVagtyo). Supposing, for the moment, that the viruses at least are restricted by genus of host (Fig. 1A), the original ancestral host type becomes ambiguous. The available data, however, refute even genus-level specificity. Various viruses from *Cercopithecus* species are unconstrained by species boundaries. The 1967 outbreak of Marburg virus demonstrated that the Ebola-like filoviridae can kill humans infected from green monkey body fluids (Halstead, 1981). Similarly, a cytomegalovirus from green monkeys has been revealed as the etiological

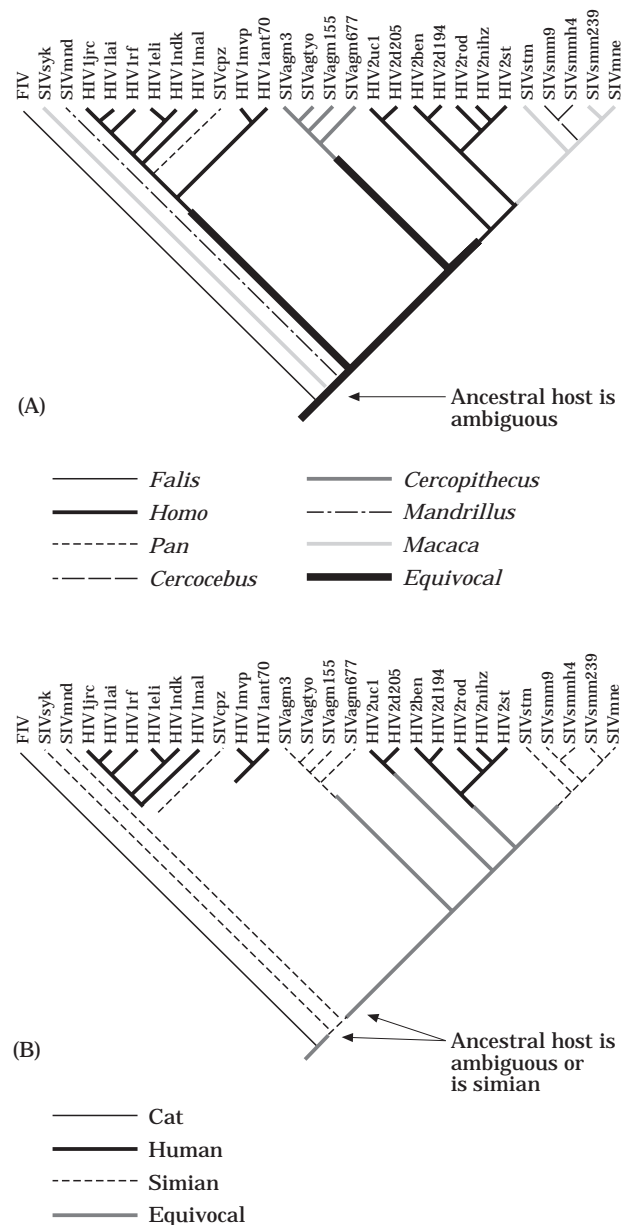


FIG. 1. The most parsimonious reconstructions of host characters when defined as genera (A) and when defined more generally (B) have different implications for determining whether or not humans are suggested as the ancestral host. In the latter perspective, the origins of the viruses are completely equivocal (tree redrawn from Mindell et al., 1995).

agent of human herpesviridae infection (Martin et al., 1995). The primate immunodeficiency virus of green monkeys is no exception. In natural populations, green monkey SIVs readily infect other cercopithecids (Bibolett-Ruche et al., 1996) and even baboons (Jin et

al., 1994). In this light the question should have been a more general human versus monkey dichotomy. However, this renders all of the ancestral host optimizations equivocal (Fig. 1B).

Mindell et al. (1995: 83) suggested that their topology was “preferable because of its incorporation of more character evidence”. However, they did not actually include *all* of the character evidence. Their hypothesis was predicated on the complete uninformative nature of transitions in the third positions of codons and also of all gapped regions. That is, third positions in codons received zero weight and gapped regions were excluded. This practice, though defended by Mindell et al. (1995) in terms of concerns regarding swamping of phylogenetic signal by noise, is a contentious issue in systematics (Bryant, 1989; Hillis, 1991; Collins et al., 1994) and it has yet to be shown that noise will always swamp a signal (Rieppel, 1989). Even if third-position transitions are expected to be problematic at ancient nodes (Mindell et al., 1995: table 1), those transitions may yet be expected to be informative regarding recent divergences. Insofar as the working hypothesis was that human infection has been the result of a recent divergence, data relevant to these recent phenomena should not have been excluded, or the difference that this makes in the resulting hypothesis should at least have been acknowledged. I have reanalysed the PIV data set¹ in light of these concerns. Use of Farris et al.’s (1994) incongruence measure indicated that none of the excluded portions of data were significantly heterogeneous with the included data. That is, for each of the four genes, treating gapped regions as a partition rendered a tail probability of incongruence between 0.70 and 0.18 and for third positions in codons versus first and second, tail probabilities were between 0.93 and 0.62 (using ARNIE.EXE [Siddall, 1996a]).

If one allows third positions to be informative (with or without the designated “poorly aligned” regions) the most parsimonious tree (Fig. 2) differs in certain significant relationships from that in Fig. 1. Mindell et al. (1995) compared their findings of HIV2 paraphyly² with previous reports of a single origin for HIV2 in humans (Myers et al., 1993). In contrast, reanalysis indicates that the hypothesis of HIV2 monophyly

¹ftp://ftp.cc.utexas.edu/pub/systbiol/44_1/44_1_mindell/mindelldata.txt

²This was mistakenly reported as polyphyly by Mindell et al. (1995).

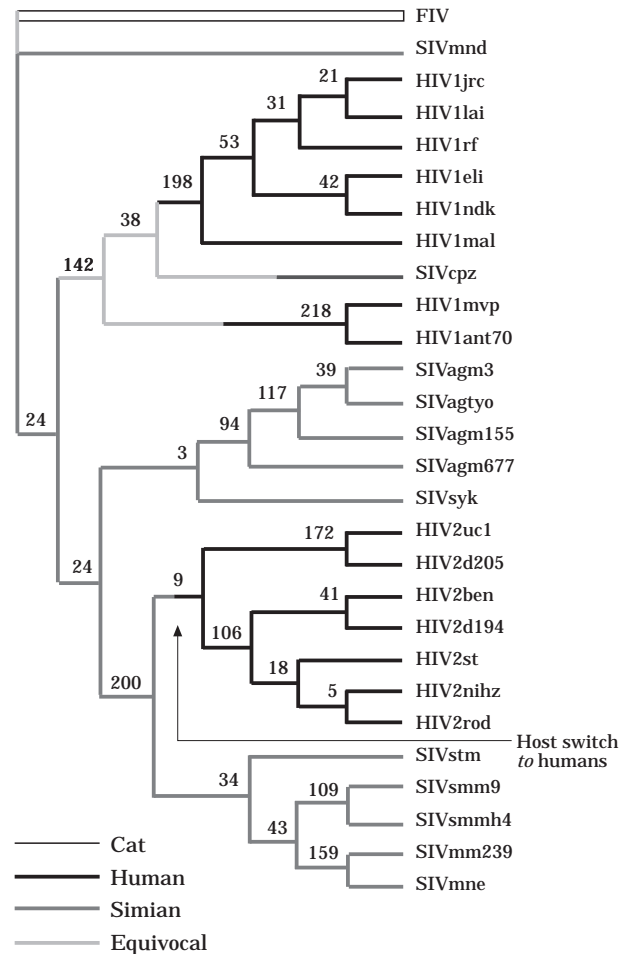


FIG. 2. The most parsimonious tree obtained using all of the data unweighted, all of the data with successive approximations, and all of the data with third-position transitions downweighted according to their estimated frequency. Bremer (1988) support indices are given at each node. Reconstruction of host type on this tree suggests an unequivocal simian origin for human HIV infections.

cannot be rejected. Moreover, whereas SIVsyk previously was not found to be monophyletic with the other viruses infecting *Cercopithecus* species (Fig. 1), it is when all of the data are considered (Fig. 2). Although I disagree with the relative weighting of character information, it turns out that it does not matter in this case. When all four genes are considered together, the transition: transversion ratios are: 1.20 for first positions, 0.97 for second positions, and 1.94 for third positions. Applying a 2:1 cost ratio for third positions, in accordance with these values, renders the same tree as was found when all characters are considered of equal

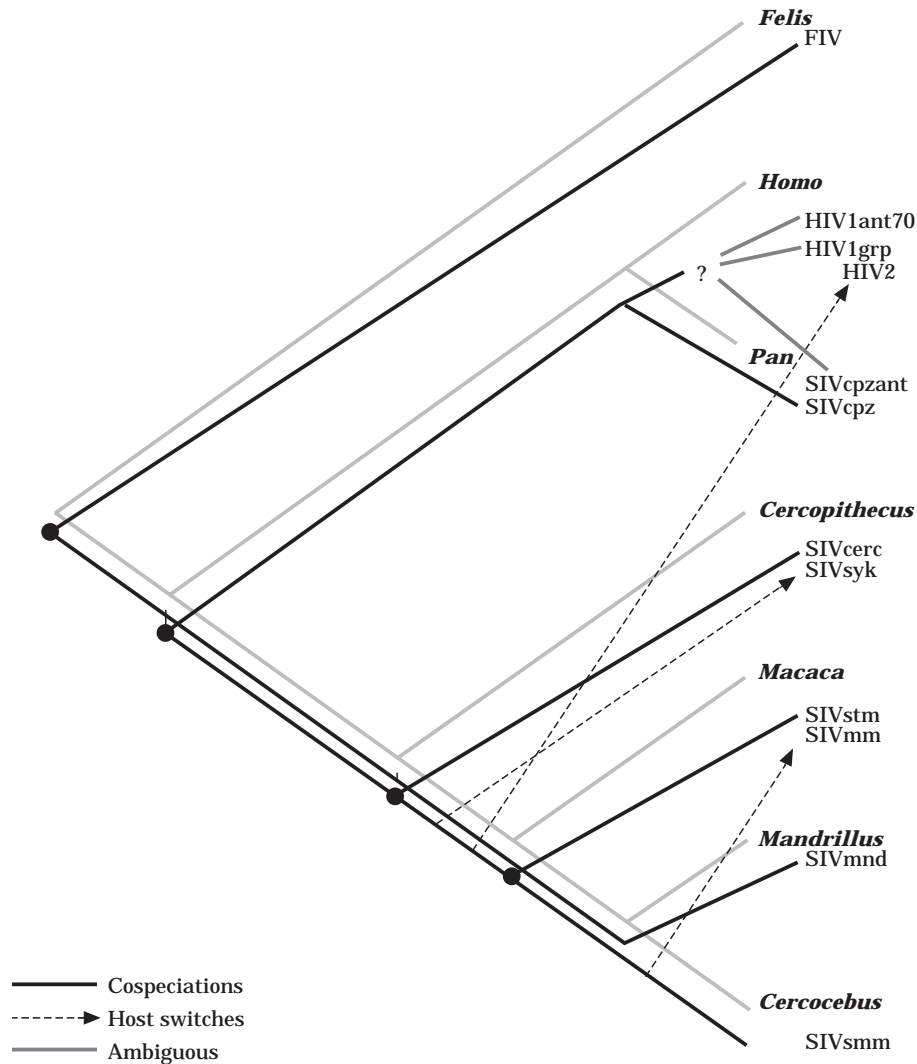


FIG. 3. A more appropriate macroevolutionary comparison of host and parasite phylogenies using TreeMap indicates host-switching of HIV2 viruses from monkeys to humans. HIV1 origins are ambiguous. **Hosts**; viruses.

weight (Fig. 2). If, however, one is going to weight according to the relative consistency of a character, it is inadvisable to assume that one particular class, like third positions, necessarily have changed according to a homogenous "rule". Use of successive approximations character weighting (Farris, 1969) applies weights in accordance with character consistency on a character-by-character basis, and yet this too yields the same optimal tree as was found using the unadulterated data (Fig. 2).

Using this total-evidence perspective on the data, which appears to result in a robust and well-corroborated tree, and the more realistic restatement of perspective on the hosts (monkey versus man), a very different hypothesis of HIV origins is obtained. Monkeys are indicated to be the ultimate ancestral hosts as well as the more immediate ancestral hosts for HIV2 infections in humans, whereas the origins of HIV1 are ambiguous (Fig. 2). On the whole, then, this hypothesis does, in fact, support the "new virus" hypothesis in

methods all are predicated on comparing a phylogenetic hypothesis of hosts with that of the parasites. Reconciled tree analysis (Page, 1994a), for example, seeks to separate three historical phenomena from each other in ecological associations: cospeciation (origins by descent), host-switching (origins by dispersal), and duplication (origins by paralogy). In order to make a determination on the patterns of historical association of the immunodeficiency viruses, I have compared the PIV phylogeny based on all of the data to a comprehensive phylogeny of primates based on seven protein coding genes (Goodman et al., 1982). An exhaustive search with TreeMap (Page, 1996) yielded 12 optimal reconstructions, each with five putative cospeciation events. These 12 reconstructions, however, have different numbers of sympatric speciation events, host-switching events and sorting events; nonetheless, all of them agreed that HIV2 resulted from a host-switch from cercopithecoid monkeys to humans (Fig. 3). When minimizing the number of sympatric speciation events (Fig. 3), there are at least two other host-switches involving the cercopithecoids. The solution to HIV1 origins is ambiguous (Fig. 3). Similarly, results of BPA analysis (Fig. 4) agree that HIV2 resulted from a host-switch from cercopithecoids, and also that HIV1 origins are ambiguous (two optimal reconstructions for viral-character 14). As was indicated using the more limited host framework in Fig. 2, the most corroborated hypothesis, for the time being, is that HIV2 infections originated in humans by virtue of a host-switch *from* monkeys; specifically from cercopithecoids, and that HIV1 infections could have arisen by virtue of a host-switch as well.

In light of the amount of host-switching from monkeys to other monkeys and man in the preceding analysis, any hypothesis of cospeciation of these viruses with their hosts is not significant. That is, neither random trees in TreeMap nor Siddall's (1996b) PCP test for BPA yields significant results (e.g. PCP=0.538). It is possible that an "old" coevolutionary origin of at least part of HIV infections could become compelling as more virus isolates from chimpanzees are added to analyses. Presently there are only two chimpanzee sequences. SIVcpz (used here and previously) as well as SIVcpzant (Genbank accession no. U42720). Adding the SIVcpzant data for the *gag* and *pol* sequences used by Mindell et al. (1995), and removing aligned positions 1764–1791 in the Mindell et al. (1995)

data set³ confirmed the findings of Myers and Korber (1994) that SIVcpzant is sister taxon to the SIVcpz/HIV1 clade. This relationship makes a cospeciation hypothesis even less tenable than it might have been above. Although the reconciled tree approach is more appropriate than merely optimizing hosts on a parasite tree, Fig. 2 is still simplistic. Because PIVs readily cross-infect distantly related primates (Jin et al., 1994; Bibolett-Ruche et al., 1996) and have entered human populations at least twice (HIV1 and HIV2), coevolution within primates is unlikely.

Mindell et al. (1995: 89) indicated that "the new virus hypothesis... devolves to unjustified assumptions". However, similar assumptions, like that of strict host-specificity, seem to have complicated their approach. Moreover, by eliminating data which could be relevant to recent events, such as third-position transitions and some of the faster evolving segments, Mindell et al. (1995) were prevented from recovering a single origin for HIV2 like that found by Myers et al. (1993). This alone would have pointed to an unequivocal simian origin for the human infections. These observations, as well as the more appropriate historical association and reconciled trees approach, strongly support the traditional view that human immunodeficiency viruses are "new", and have their origins in host-switches *from* monkeys. The findings of Mindell et al. (1995) would have us reconsider the mechanisms by which these viruses are transmitted, or have us markedly reconsider primate behavioural patterns. Notably, Mindell et al. (1995:89) made the prediction that "small human populations with dormant PIVs may readily have been missed". In light of the toll that these infections have exacted on the human community our hypotheses and our perspectives on data must be subjected to extraordinary rigour and must withstand the severest of tests. The conclusions drawn by Mindell et al. (1995) do not.

ACKNOWLEDGEMENTS

I thank David Mindell for inviting me to look at the HIV data and for stimulating discussions stemming from this issue. Steve Farris, Arnold Kluge and John Wenzel offered comments on earlier drafts of this manuscript. Fred Ronquist provided a constructive review.

³The data in positions 1764–1791 were duplicated erroneously at positions 1792–1819 in the data set used by Mindell et al. (1995). These two regions of the same data also were aligned differently.

REFERENCES

- Atlan, H. M., Gersten, M. J., and Salk, J. (1994). Mechanisms of autoimmunity and AIDS: Prospects for therapeutic intervention. *Res. Immunol.* **145**, 165–178.
- Bibollet-Ruche, F., Galat-Luong, A., Cuny, G., Sarni-Manchado, P., Galat, G., Durand, J. P., Pourrut, X., and Veas, F. (1996). Simian immunodeficiency virus infection in a patas monkey (*Erythrocebus patas*): Evidence for cross-species transmission from African green monkeys (*Cercopithecus aethiops sabaeus*) in the wild. *J. Gen. Virol.* **77**, 773–781.
- Bremer, K. (1988). The limits of amino-acid sequence data in angiosperm phylogenetic reconstruction. *Evolution* **42**, 795–803.
- Brooks, D. R. (1979). Testing the context and extent of host-parasite coevolution. *Syst. Zool.* **28**, 299–307.
- Brooks, D. R. (1981). Hennig's parasitological method: A proposed solution. *Syst. Zool.* **30**, 229–249.
- Brooks, D. R. (1990). Parsimony analysis in historical biogeography and coevolution: Methodological and theoretical update. *Syst. Zool.* **39**, 14–30.
- Bryant, H. N. (1989). An evaluation of cladistic and character analyses as hypothetico-deductive procedures, and the consequences for character weighting. *Syst. Zool.* **38**, 214–227.
- Collins, T. M., Kraus, F., and Estabrook, G. (1994). Compositional effects and weighting of nucleotide sequences for phylogenetic analysis. *Syst. Biol.* **43**, 449–459.
- Farris, J. S. (1969). A successive approximations approach to character weighting. *Syst. Zool.* **18**, 374–385.
- Farris, J. S., Källersjö, M., Kluge, A. J., and Bult, C. (1994). Testing significance of incongruence. *Cladistics* **10**, 315–320.
- Goodman, M., Romero-Herrera, A., Dene, H., Czelusniak, J., and Tashier, R. E. (1982). Amino acid sequence evidence on the phylogeny of primates and other eutherians. In "Macromolecular Sequences in Systematics and Evolutionary Biology" (M. Goodman, Ed.), pp. 115–191. Plenum Press, New York.
- Halstead, S. B. (1981). Viral hemorrhagic fevers. *J. Infect. Dis.* **143**, 127–138.
- Hillis, D. M. (1991). Discriminating between phylogenetic signal and random noise in DNA sequences. In "Phylogenetic Analysis of DNA Sequences". (M. M. Miyamoto, and J. Cracraft, Eds), pp. 278–294. Oxford Univ. Press, New York.
- Jin, M. J., Rogers, J., Phillips-Conroy, J. E., Allan, J. S., Desrosiers, R. C., Shaw, G. M., Sharp, P. M., and Hahn, B. H. (1994). Infection of a yellow baboon with simian immunodeficiency virus from African green monkeys: Evidence for cross-species transmission in the wild. *J. Virol.* **68**, 8454–8460.
- Kellogg, V. L. (1896). New Mallophaga. I. — with special reference made from maritime birds of the Bay of Monterey, California. *Proc. Cal. Acad. Sci.* **6**, 31–196.
- Martin, W. J., Ahmed, K. N., Cheng-Zeng, L., Olsen, J. -C., Seward, J. G., and Seehrai, I. S. (1995). African green monkey origin of the cytopathic "Stealth Virus" isolated from a patient with chronic fatigue syndrome. *Clin. Diag. Virol.* **4**, 93–103.
- Mindell, D. P., Shultz, J. W., and Ewald, P. W. (1995). The AIDS pandemic is new, but is HIV new? *Syst. Biol.* **44**, 77–92.
- Myers, G., and Korber, B. (1994). The future of human immunodeficiency virus. In "The Evolutionary Biology of Viruses" (S. S. Morse, Ed.), pp. 211–232. Raven Press, New York.
- Myers, G., Macinnes, K., and Myers, L. (1993). Phylogenetic moments in the AIDS epidemic. In "Emerging Viruses" (S. S. Morse, Ed.), pp. 120–137. Oxford University Press, New York.
- Page, R. D. M. (1988). Quantitative cladistic biogeography: Constructing and comparing area cladograms. *Syst. Zool.* **37**, 254–270.
- Page, R. D. M. (1990). Temporal congruence and cladistic analysis of biogeography and cospeciation. *Syst. Zool.* **39**, 205–226.
- Page, R. D. M. (1991). Clocks, clades, and cospeciation: Comparing rates of evolution and timing of cospeciation events in host-parasite assemblages. *Syst. Zool.* **40**, 188–198.
- Page, R. D. M. (1993). Genes, organisms, and areas: The problem of multiple lineages. *Syst. Biol.* **42**, 77–84.
- Page, R. D. M. (1994a). Maps between trees and cladistic analysis of historical associations among genes, organisms, and areas. *Syst. Biol.* **43**, 58–77.
- Page, R. D. M. (1994b). Parallel phylogenies: Reconstructing the history of host-parasite assemblages. *Cladistics* **10**, 155–173.
- Page, R. D. M. (1996). TreeMap. Division of Environmental and Evolutionary Biology, Institute of Biomedical and Life Sciences, University of Glasgow, Glasgow.
- Rieppel, O. (1989). Character incongruence: Noise or data? *Abh. Naturwiss. Ver. Hamburg (NF)* **28**, 53–62.
- Siddall, M. E. (1996a). ARNIE. EXE. In Random Cladistics (The Ohio Edition) Version 4. Department of Zoology, University of Toronto. Toronto, Ontario, Canada.
- Siddall, M. E. (1996b). Phylogenetic covariance probability: Confidence and historical associations. *Syst. Biol.* **45**, 48–66.