Pergamon

Zoologica Scripta, Vol. 26, No. 4, pp. 313–322, 1997
Elsevier Science Ltd

PII: S0300-3256(98)00008-7
0300-3256/97 $17.00+0.00

Phylogenetic approaches in coevolution and biogeography

FREDRIK RONQUIST

Accepted 27 *January* 1998

Ronquist, F. 1998. Phylogenetic approaches in coevolution and biogeography.—*Zool. Scr. 26*: 313–322.

I review phylogenetic approaches to problems in coevolution and biogeography, illustrating with case studies. In coevolution, genealogical trees are essential in differentiating between ancient and recent associations, in identifying cospeciation events, and in studying host-switching patterns. Cospeciating associations are of particular interest because they allow powerful tests of molecular clocks and accurate comparison of evolutionary rates across groups of organisms. In biogeography, phylogenies can help reconstruct the distribution history of individual groups and identify past geological events that have affected the evolution of entire communities. Parsimony analysis in coevolution and biogeography should be based on identification of different types of events, each of which is associated with a specific cost. Similar event-based methods are applicable to coevolutionary and biogeographic inference, as well as in the mapping of gene trees onto organism trees. The discussed examples span a variety of organisms and spatiotemporal scales: primate pin worms, HIV, pocket gophers and their lice, aphids and their bacterial symbionts, gall wasps and their host plants, the root of the tree of life, the historical biogeography of the Holarctic, and the geographical origin of our own species. © 1998 The Norwegian Academy of Science and Letters

Fredrik Ronquist, Department of Zoology, Uppsala University, Villavägen 9, SE-752 36 Uppsala, Sweden.

Introduction

The hierarchy of life is *the* general reference system in biology; thus, understanding genealogical relationships is fundamental to all biological research. In this paper, I will discuss some of the ways in which phylogenetic hypotheses can be used in the study of two, as we shall see, closely related biological disciplines: coevolution and biogeography. In particular, the basics of parsimony inference in these disciplines will be explained. I will also show how coevolutionary methods can be used in the mapping of gene trees onto organism trees. I will argue that parsimony methods in coevolutionary and biogeographic inference, like in phylogenetic inference, must be based on the identification of different types of events, each of which is assigned a cost related to the likelihood of that event. Hence, I will be focusing throughout the paper on event-based parsimony methods and neglect many of the other commonly used methods which cannot, or have not yet, been described in such terms.

Coevolution

Coevolution occurs when two species interact with each other intimately enough and sufficiently long to affect each other's evolution. Some familiar examples of such co-evolved species associations include moth larvae and the host plants they feed on, orchids and their pollinators, and termites and their intestinal symbionts that help them digest cellulose.

Even if we restrict coevolution to symbioses, in which one organism lives inside or on the surface of another, we are talking about an exceedingly common phenomenon. Take, for instance, a normal, healthy, adult human. He (or she) carries around about 1.5 kg of microorganisms, in number more than 10 times as many as the cells in the body (Grubb 1994). The microorganisms occur on the skin, in the oral cavity, in the outer parts of the urinary tract, but above all in the intestine. Of bacteria alone there are more than 400 species in the normal gut flora (Grubb 1994). We are so intimately adapted to our intestinal flora that our well-being is entirely dependent upon it. In addition to the normal flora, humans are associated with a number of pathogenic viruses, bacteria, and protozoans, as well as parasitic higher animals such as tapeworms, roundworms, pin worms, head-lice and itch-mites. We have no reasons to believe that humans are exceptional; other organisms undoubtedly carry similar communities of symbionts and parasites.

With some of our symbionts, we share a long evolutionary history, dating back long before the time when the human evolutionary lineage separated from that of the chimpanzees. Other organisms have only recently come to be associated with humans. How can we separate newcomers from veterans among our associates?

To answer this question we need phylogenetic trees. Take for instance the pin worm, a common human parasite in temperate areas, particularly among children. Let us compare current estimates of relationships among the human pin worm and its closest relatives and our own genealogical relationships (Fig. 1). If we disregard humans, all speciations, or branching points, in the pin-worm phy-

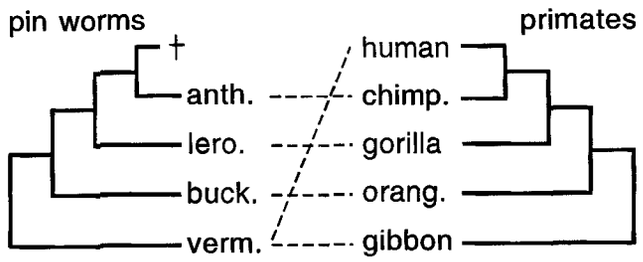

Fig. 1. Relationships among some primate pin worms and their hosts (Glen & Brooks 1985; Shoshani *et al.* 1996). Dashed lines connect parasites with their hosts. The parasite phylogeny mirrors the host phylogeny perfectly, if we assume that the pin worm of humans went extinct and was replaced by the pin worm of gibbons.

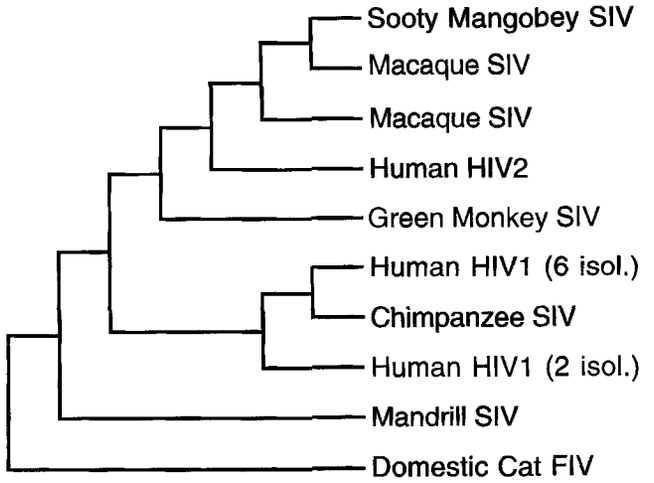

Fig. 2. Relationships among HIV isolates and related viruses (Siddall in press).

logeny correspond to speciations in the primate tree. The only reasonable explanation is that pin worms already plagued the common ancestor of humans and apes, and that they share our evolutionary history. When host populations became isolated and developed into separate species, the parasites were simultaneously isolated and responded by speciating. Such parasite-host cospeciation is of particular importance in the study of coevolving associations.

Humans deviate from the cospeciation pattern (Fig. 1). We can explain this by assuming that humans originally hosted a pin worm that was closely related to the pin worm of chimpanzees. Later, the pin worm of gibbons colonised humans and displaced the original human parasite. Alternatively, humans may have escaped pin-worm attack earlier, before being infected by the gibbon parasite, perhaps as early as the time when humans and chimpanzees became separately evolving lineages. In either case, phylogenetic trees can help us draw two conclusions. First, pin worms have been associated with our evolutionary lineage for a long time, and humans are therefore likely to have ancient defence mechanisms against pin worms. Second, the pin-worm species attacking humans today is a recent colonist in evolutionary terms, and we might therefore be less well equipped to deal with the unique features of this particular species.

Take another example, the human immunodeficiency virus (HIV). It was initially thought that HIV was transmitted from monkeys to humans in Africa some decades ago, and then rapidly spread over the globe, but recent trees of relationships among HIV isolates and related viruses give a more complex picture (Fig. 2) (Siddall 1997). There are two, fundamentally different types of HIV. Type 2 is only found in Africa, whereas type 1 is cosmopolitan. Phylogenetic analyses show that type 2 is closely related to viruses occurring in mangobeys. In this case, recent transmission to humans appears likely, perhaps through infected monkeys biting humans (Leigh Brown & Holmes 1994). Type 1, however, is closely related to a chimpanzee isolate, but not to type 2 or other monkey isolates. Thus, HIV type 1 cannot have evolved from HIV type 2, and it is quite possible that HIV 1 is a virus that has an ancient association with humans and only recently became strongly virulent (Mindell *et al.* 1995).

Parsimony methods

Associations differ in their tendency to cospeciate. Sometimes the common pattern is so strong that it is possible to use phylogenies for one group to elucidate relationships in the other group. Mitochondria, originally derived from free-living bacteria, is a group of symbionts that, in principle, show strict cospeciation with their host organisms. If this were not so, mitochondrial genes could not be used in the study of relationships among their host organisms.

When the patterns are less clear, quantitative methods are needed to show whether or not there is a nonrandom fit between the host and parasite phylogenies. For instance, there are many similarities between the phylogenies of pocket gophers, a North American group of ground-dwelling rodents, and their chewing lice (Fig. 3), but there are also important discrepancies. Can we safely draw the conclusion that the lice have cospeciated with their hosts?

A simple type of method we can use in this case is parsimony analysis. In its most generalised form, parsimony analysis is based on the identification of different types of events, each of which is associated with a cost inversely related to the likelihood of that event. In other words, when we combine events into a solution, rare events are costly and likely events cheap. The most parsimonious reconstruction is the cheapest combination of events that will solve the problem; if events are assigned costs appropriately, this is also the most likely explanation of the data.

The simplest type of parsimony analysis of phylogenetic relationships is based on binary characters. There are two different states (0 and 1) and two different types of events (a 0–>1 change and a 1–>0 change), each with the same cost. From this, we can calculate the most likely phylogenetic hypothesis, i.e., the tree requiring the smallest number of evolutionary changes.

In coevolutionary analysis we need to consider four different types of events: duplications, host shifts, sorting events, and cospeciations (Fig. 4) (for an overview of coevolutionary methods that are not event-based, see Brooks (1988)). Duplications occur when parasites speciate independently of their hosts, but remain associated with their ancestral host (Fig. 4A). Host shifts are often considered as being associated with parasite speciation, one daughter parasite lineage shifting to a new host (Fig. 4B). Sorting events occur when a parasite tracks a host lineage through a speciation event without speciating itself (Fig. 4C) or when a parasite and a host cospeciates, but

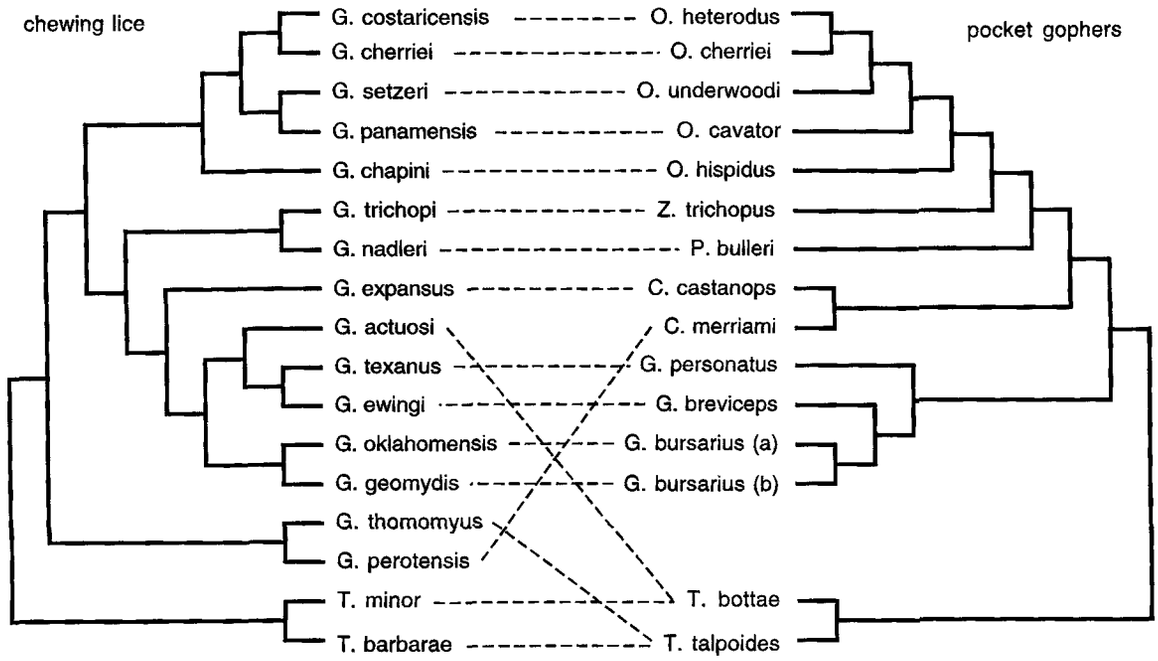

Fig. 3. Relationships among pocket gophers and their chewing lice (Hafner *et al.* 1994). Randomisation test using maximum cospeciation analysis indicates that there is significant congruence between the parasite and host phylogenies.

the parasites go extinct in one of the daughter host lineages. Cospeciations involve simultaneous host and parasite speciation (Fig. 4D).

Once the costs of different events have been defined, standard two-dimensional (Ronquist 1996) or recently developed three-dimensional cost matrix optimisation algorithms (Ronquist 1997; Ronquist in press) can be used to find the best reconstructions. The trick is to find the costs of the events. Ideally, the likelihood of different types of events would be estimated directly from the data, but as of yet, no good methods to achieve this have been developed (but see Ronquist 1996). Instead, we work with simple methods that focus on one or a few types of events. Perhaps the most frequently used is maximum cospeciation, presented about two years ago by Page (1995). It maximises the number of cospeciations, ignoring all other types of events. This means that cospeciations are given an arbitrary negative cost, i.e., a benefit value, whereas all other types of events have a cost of zero (Ronquist in press) (Table I).

Maximum cospeciation is useful for analysing whether or not there is more cospeciation in an association than expected by chance. This is done using randomisations. The number of cospeciations obtained with the original data is recorded and then the host associations of the parasites are shifted randomly, or the parasite or host trees are randomly rearranged (Page 1995). The maximum number of cospeciations for the randomised data sets is calculated, and the procedure is repeated a sufficient number of times. If the observed value is higher than that observed in, say, 95% of the random replications, we can reject the hypothesis of random match between parasite and host phylogenies.

With this randomisation method it is possible to show that the chewing lice of pocket gophers are likely to have

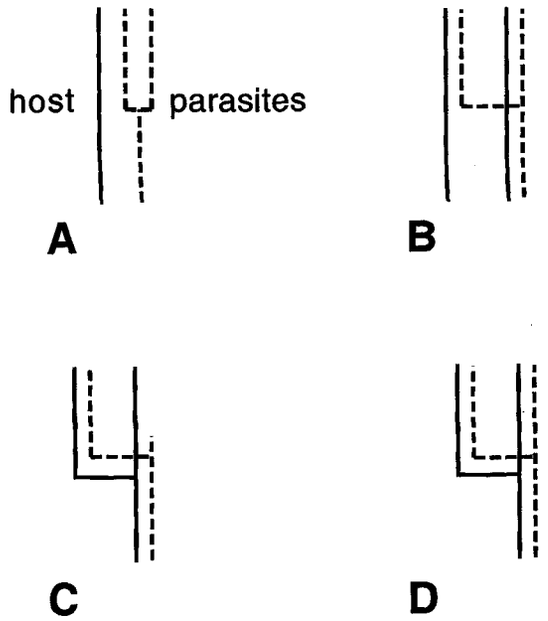

Fig. 4. Different types of coevolutionary events.—*A.* Duplication or independent parasite speciation.—*B.* Host shift associated with speciation.—*C.* Sorting event, in which the parasite disappears from or fails to occur on a host lineage.—*D.* Cospeciation, simultaneous host and parasite speciation.

Table I. *Comparison of event-cost assignments in maximum cospeciation analysis (Page 1995) and tracking/switching analysis (Ronquist 1996). In the latter method, the value of* s *may be determined by an optimality function which takes the relative fit of the tracking events (sorting and cospeciation events) to the host phylogeny into account (Ronquist 1996).*

	Cost	
Event	Max. cospeciation	Tracking/switching
Duplication	0	0
Host switch	0	s
Sorting event	0	1
Cospeciation	−1	1[1]

[1]Cost per daughter lineage.

shared a long history of speciation with their hosts (Hafner *et al.* 1994). The same method indicates that whydahs (*Vidua*), African brood-parasitic birds, have not cospeciated with the *Pytilia* finches they use as their hosts (Fig. 5). The whydahs have either shifted hosts frequently in their evolution, such that any historical pattern has been lost, or they have only recently come to be associated with *Pytilia* finches.

Cospeciation and dating

When does cospeciation occur? We still have data from relatively few associations, but some patterns are emerging. It is self-evident that cospeciation is more likely to occur in parasites that are specialised in their host choice than in generalists, but not all specialists cospeciate with their hosts (Brooks 1988). It has been assumed that intricate host-parasite coadaptation would favour cospeciation, but it appears now that the critical factor may be the rate at which the symbiont or parasite encounters potential new host species and not coadaptations per se. Associations in which the encounter rate is low, as in the case of endocellular symbionts, often show extensive cospeciation even though host-parasite coadaptations appear to be nonspecific. On the other hand, cospeciation is apparently rare in associations in which the symbiont or parasite has a free-living stage locating the host, such as phytophagous insects, regardless of the intricacy of host-parasite coadaptations.

Gall wasps, for instance, are highly specialised and rarely attack more than one or a few very closely related plant species. They induce the formation of extremely complex galls, which presumably requires intricate adaptations to their hosts. Yet, there is no evidence of insect-plant cospeciation, at least not during the early radiation of gall wasps (Ronquist & Liljeblad submitted.).

Gall wasps are useful for illustrating another point. Phylogenetic methods can be used to reconstruct ancestral hosts, even in the absence of cospeciation, if host choice is conservative enough. Maximum cospeciation analysis cannot be used in this case because it focuses entirely on cospeciation; instead, it is necessary to find reconstructions by minimising the number of host switches using a method such as tracking/switching analysis. This method, developed recently from ideas presented by Nylin and myself in 1990 (Ronquist & Nylin 1990; Ronquist 1996), finds a balance between the cost of host switching on one hand and sorting and cospeciation events, collectively called tracking events, on the other (Table I). An analysis of this kind can reconstruct the ancestral host-plant preferences of gall wasps (Fig. 6).

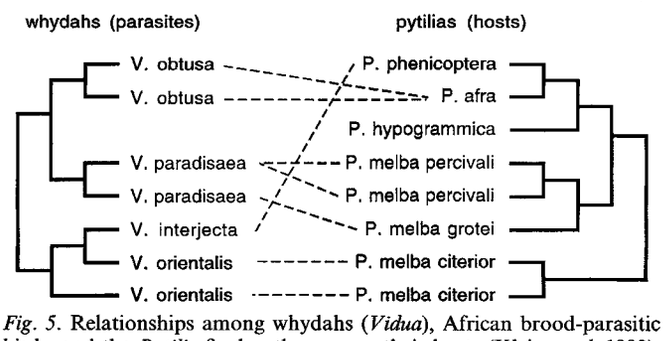

Fig. 5. Relationships among whydahs (*Vidua*), African brood-parasitic birds, and the *Pytilia* finches they use as their hosts (Klein *et al.* 1993). Randomisation test using maximum cospeciation suggests that there has not been cospeciation in this association.

An interesting property of the gall-wasp reconstruction is that it postulates certain relations between the timing of events in the evolution of the parasites and their hosts. For instance, if it is true that the ancestral gall wasp was associated with poppies (*Papaver*), gall wasps cannot be older than the genus *Papaver*. In cospeciating associations, evolutionary events are particularly closely tied in time. For instance, if there are fossils of one group, such that we can date some of the speciations in that phylogeny, the datings are directly applicable to the corresponding speciations in the associates.

Aphids feed on phloem sap, which is poor in essential nutrients. To complement their diet, they rely on endocellular symbiotic bacteria that synthesise amino acids for them. Special mechanisms have been developed to ensure that the symbionts are transferred from the mother to her offspring through the egg. The symbiotic bacteria show strict cospeciation with their host aphids (Moran & Baumann 1994). Because we can date some of the splits in the aphid phylogeny with the aid of fossils or biogeography, we can obtain absolute dates for the corresponding splits in the bacterial tree (Fig. 7). Bacterial fossils are few and difficult to place phylogenetically. Therefore, this is one of the best ways of dating events in bacterial evolution. The datings in turn allow accurate calibration of long-term rates of molecular evolution in bacteria (Moran *et al.* 1993).

Cospeciating associations can also be used to test molecular clocks, and to compare evolutionary rates in parasites and their hosts (Hafner & Page 1995; Page & Hafner 1996). Let us examine the synonymous substitution rates in the evolution of the mitochondrial COI genes of pocket gophers and their lice (Fig. 8). The rates are approximately clocklike, but the mitochondrial DNA of the lice evolves considerably faster than that of the gophers (Hafner *et al.* 1994; Page 1996). What causes the accelerated substitution rate in lice? One possibility is the difference in life span between the lice and their hosts. Because of their short generation time, the lice simply have to copy their DNA more often, which leads to more errors and more evolutionary change (Hafner *et al.* 1994). Hafner *et al.* originally estimated the rate difference to be ten-fold for synonymous changes (Fig. 8), corresponding to a similar difference in generation time between lice and their hosts. More recent reanalyses of the Hafner *et al.* data using more sophisticated techniques confirm that the lice gene evolves faster than the host gene, but suggest that the rate difference is less dramatic (Page *et al.* 1996; Huelsenbeck *et al.* 1997).

Rooting the tree of life

The same approach we have used to analyse host-parasite associations can be used to compare gene trees and organism trees. Suppose that we were faced with the task of rooting the tree of life. Many molecular analyses of

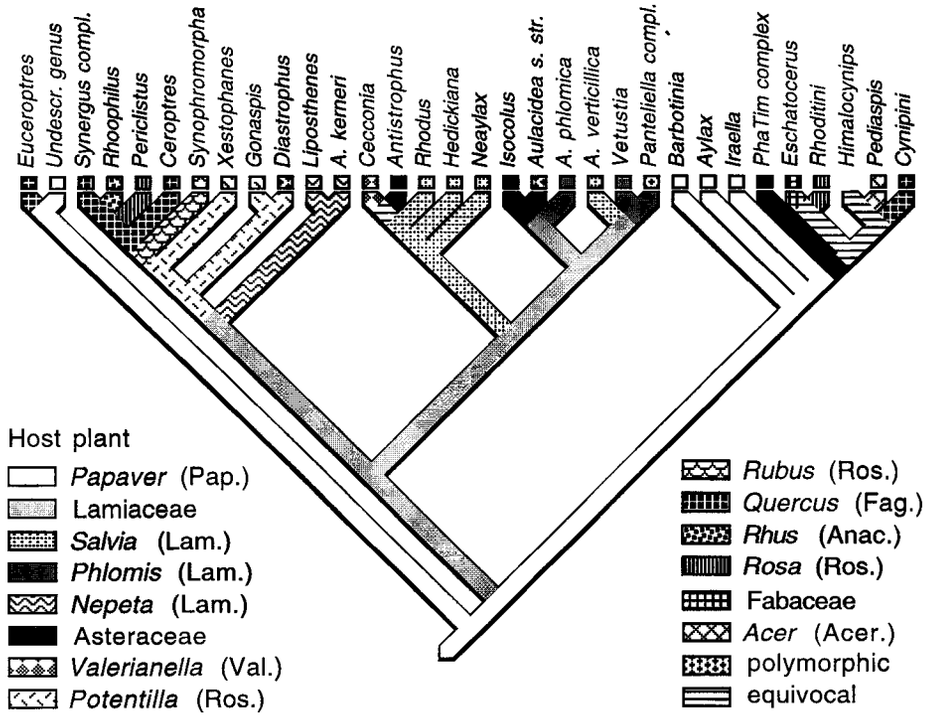

Fig. 6. Reconstruction of the ancestral host-plant preferences of gall wasps (Ronquist & Liljeblad submitted). Acer = Aceraceae, Anac = Anacardiaceae, Fag = Fagaceae, Lam = Lamiaceae, Pap = Papaveraceae, Ros = Rosaceae, Val = Valerianaceae.

relationships among all life forms have produced unrooted trees with a similar topology (Fig. 9). However, depending on where the tree is rooted, entirely different pictures of the evolution of life emerge. The tree can be rooted between the major groups but also within one of the groups. For instance, if we root the tree within the eukaryotes, we would have to conclude that prokaryotic organisms evolved by reduction of more complex life forms.

Ordinarily, a tree is rooted by reference to a more inclusive analysis. This can also be described as dividing the tree into an ingroup and an outgroup part. The tree of life cannot be rooted in this manner. We could find the root by making all distances from the root to the leaves of the tree, measured in terms of character changes, as equal as possible, but this method fails when evolutionary rates vary considerably among organisms, as they often do (cf. the mitochondrial genes of pocket gophers and their lice discussed previously).

A much more powerful method is to look at genes that were duplicated before the earliest split in the tree of life (Iwabe *et al.* 1989). In doing this, we can apply the same thinking we have used in analysing the evolution of host-parasite associations. If genes are viewed as parasites and organisms as their hosts, it is possible to work with the same types of events and the same parsimony methods as in coevolutionary analysis (Page 1993).

To the extent that this problem has been analysed quantitatively at all, it has been approached with a method called reconciliation, originally formulated by Goodman and colleagues in the late 70's for analysing the evolution

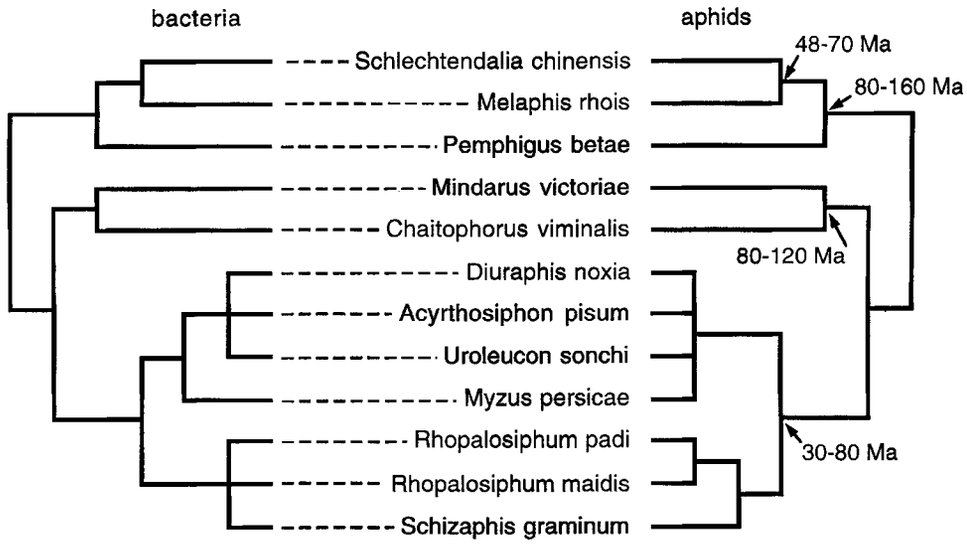

Fig. 7. Cospeciation between aphids and their endocellular symbionts (Moran *et al.* 1994). Dates estimated for branching events in the aphid phylogeny based on biogeographic and fossil data also apply to the corresponding events in the bacterial phylogeny.

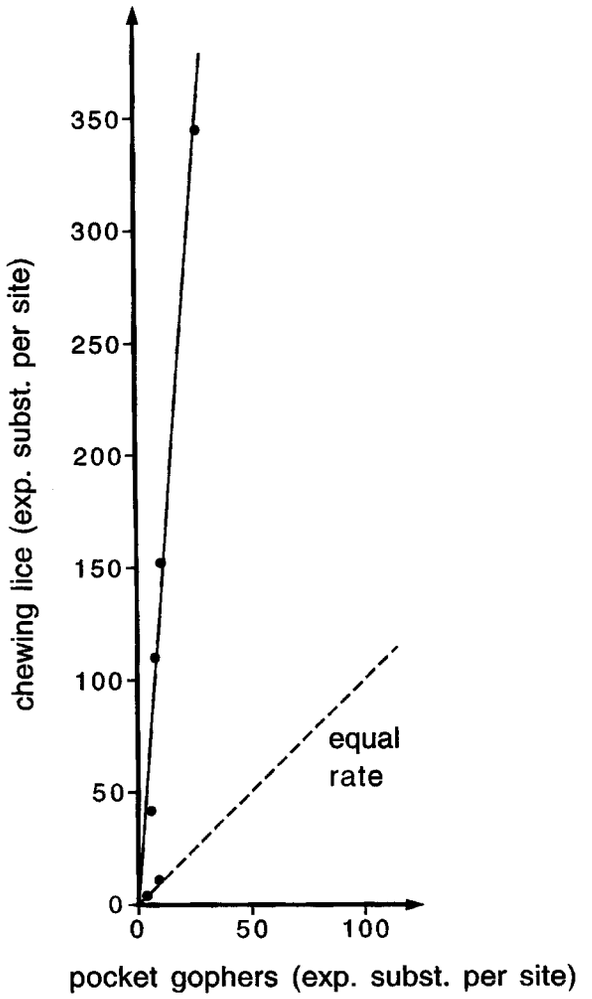

Fig. 8. Rates of evolution of mitochondrial COI genes of pocket gophers and their lice (Hafner *et al.* 1994) (expected number of substitutions at four-fold degenerate sites). The mitochondrial gene of the lice evolves much faster than the mitochondrial gene of the gophers (dashed line corresponds to equal rate).

of globin genes (Goodman *et al.* 1979) (Table II) and later formalised and developed by Page and others (Page 1994; Mirkin *et al.* 1995; Guigó *et al.* 1996; Page & Charleston 1997; Zhang 1997). The events correspond directly to coevolutionary events: duplication is gene duplication, host shift is horizontal transmission of the gene between unrelated organisms, sorting occurs when a gene disappears from a lineage for some reason, and cospeciation corresponds to organism speciation, leading to isolation of

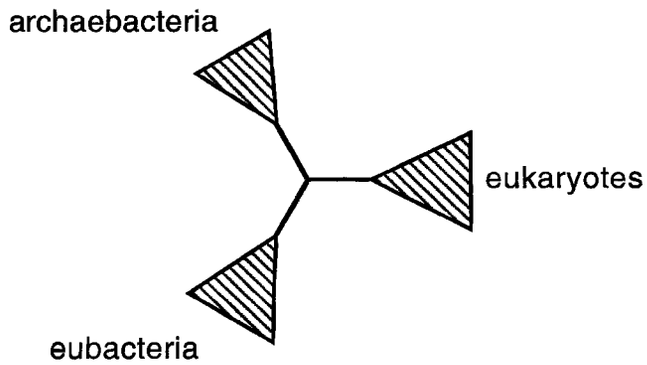

Fig. 9. Unrooted tree of life (supported by many molecular analyses, e.g., Iwabe *et al.* 1989; Woese *et al.* 1990; Brown & Doolittle 1995; De Rijk *et al.* 1995).

Table II. *Different event–cost assignments describing the method of fitting gene trees onto species trees developed by Goodman and colleagues (Goodman* et al. *1979). Both sets of cost assignments give the same optimal reconstructions as the original version of the method.*

	Cost	
Event	*Minimise dupl.*	*Maximise cospec.*
Duplication	*1*	*0*
Host switch	∞	∞
Sorting event	*0*	*0*
Cospeciation	*0*	*−1*

the gene into separately evolving lineages. Reconciliation is based on minimisation of the number of implied gene duplications, i.e., this type of event is associated with a positive cost. To simplify calculations, horizontal transmission is prohibited, i.e., this type of event is associated with an infinite cost so that no optimal explanation can include horizontal transmission.

It is equally possible to describe reconciliation as a method that maximises the number of cospeciations — the results of the analysis will be identical, which can be easily shown. In this case, the cost of duplications is set to 0 and cospeciations to −1. Now the similarities with maximum cospeciation become obvious (cf. Table I). The only difference is that maximum cospeciation allows horizontal transmission. Since we cannot exclude the possibility that horizontal transmission occurs occasionally, it would be better to use maximum cospeciation than reconciliation in the analysis of gene trees.

Consider the gene tree for elongation factor, a protein that occurs in two different copies in most organisms (Fig. 10A). It is possible to fit this gene tree onto an organism tree by introducing a single duplication event (Fig. 10B). The reconstruction that minimises the number of duplications also maximises the number of cospeciations. Any change in the topology or rooting of the organism tree would entail an increase in the number of duplications and a decrease in the number of cospeciations for elongation factor; thus, coevolutionary analysis fixes the root and topology of the organism tree.

Several other proteins give the same branching pattern as elongation factor, but not all of course. For instance, GAPDH (glycerylaldehyde-3-phosphate dehydrogenase) from archaebacteria differs substantially from the homologous enzyme in eubacteria and eukaryotes (Iwabe *et al.* 1989). This could be because of an unproportionally high rate of change on the branch leading to archaebacteria, but assume that there were a sibling gene suggesting a rooting between archaebacteria and the other groups (Fig. 10C). Then we could explain the evolution of GAPDH as the result of a gene duplication followed by sorting events (Fig. 10D), one gene copy ending up in eukaryotes and eubacteria and the other in archaebacteria. However, the number of cospeciations would be lower than for elongation factor, pointing out the mismatch between this hypothetical, duplicated GAPDH gene and the tree supported by elongation factor. To locate the root of the tree of life accurately, it is necessary to study a large number of duplicated genes and search for the tree that allows the maximum number of cospeciations. Thus far, relatively

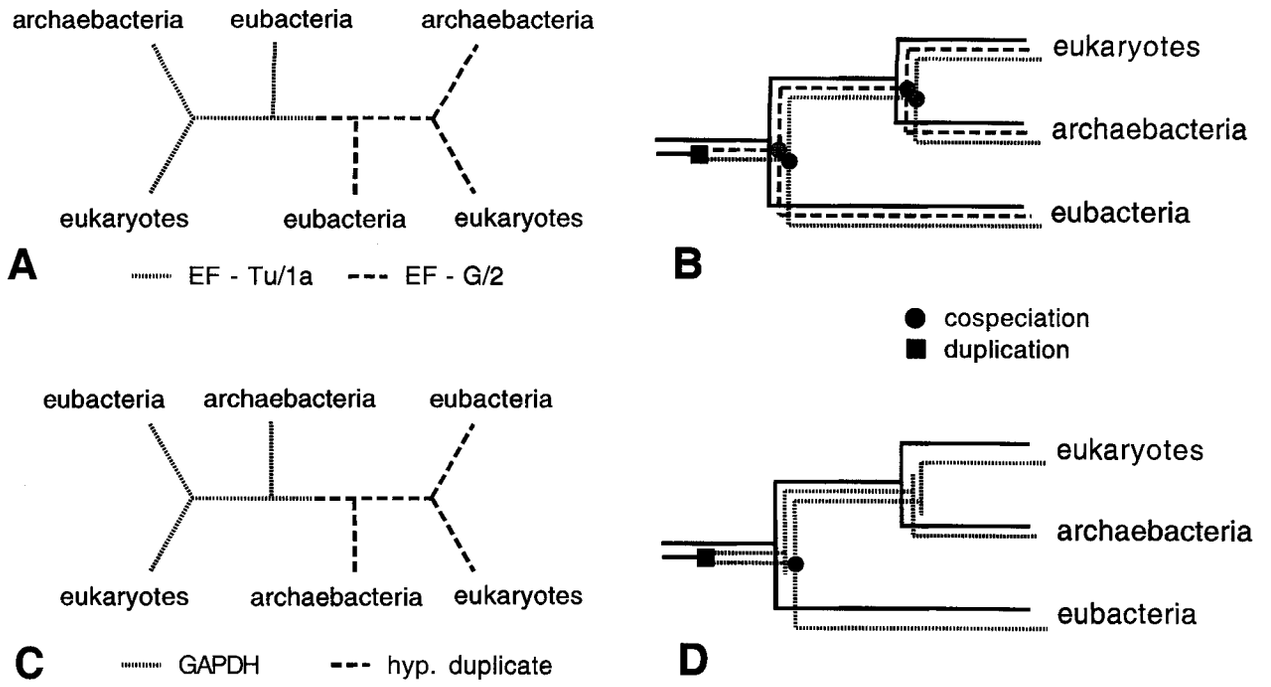

Fig. 10. Using coevolutionary analysis of genes duplicated in the ancestor of all life to root the tree of life.—*A*. An unrooted tree for the two paralogous copies of elongation factor (EF) (Iwabe *et al.* 1989).—*B*. The rooted tree of life explaining the evolution of EF best. No other rooted tree gives fewer duplications (squares) or more cospeciations (circles).—*C*. A tree of the enzyme GAPDH and a hypothetical duplicate gene.—*D*. Mapping GAPDH on the rooted tree of life suggested by EF indicates that GAPDH was duplicated in the ancestor of all life, and that one copy ended up in eubacteria and eukaryotes, while the other copy is exclusively present in archaebacteria (only the GAPDH half of the gene tree shown).

few duplicated genes have been studied, each in a small number of exemplar species. It is quite possible that future analyses will show that the root of the tree of life is located within one of the prokaryotic domains rather than between the domains as suggested by current analyses, or that the current division of life into three domains is incorrect (e.g., Gupta & Golding 1993).

Historical biogeography

Biogeography is the study of the distribution of organisms. The discipline is often divided into ecological biogeography and historical biogeography, depending partly on the time perspective and partly on the conceptual framework and methods used. The phylogenetic approach is appropriate in historical biogeography. Usually, the time scales are considerably longer in historical than in ecological biogeography, but the recent application of phylogenetic techniques to study population-level relationships and their biogeographic implications, sometimes called phylogeography, have made historical biogeography grade into ecological time scales (Avise *et al.* 1987; Avise 1994).

Phylogenetic trees can be used to infer the distribution history of organisms and to identify common geological events that have affected the evolution of many different groups of organisms inhabiting the same areas. In principle, it is possible to analyse biogeographic problems with the same parsimony techniques used in coevolutionary analysis if we replace hosts with geographical areas and parasites with organisms inhabiting these areas (for an overview of biogeographic inference methods that are not event-based, see Crisci & Morrone (1995)). The coevolutionary events translate directly to biogeographic events: Duplication corresponds to sympatric speciation: a species is divided into two separate lineages, both occurring in the same area. Alternatively, duplication may result from allopatric speciation in response to a temporary or partial barrier followed by secondary dispersal when the barrier disappears. Host shift is the same as dispersal between two separate, disjunct areas. Sorting events are equivalent to partial extinctions — a group of organisms is missing from an area where it is expected to occur. Cospeciation corresponds to vicariance, i.e., division of a continuous area into two or more subareas separated by dispersal barriers, with subsequent speciation in the organisms occurring in the original area.

However, there is an important difference between biogeography and coevolution: whereas hosts can be safely assumed to have a branching genealogy, areas are not always related hierarchically. In biogeography, a hierarchical history requires successive subdivision of a large ancestral area (Fig. 11A–B). This gives a sequence of vicariance events that affect all groups of organisms evolving in the area. This in turn results in phylogenetically ordered distribution patterns.

Compare this with a simplified model of the Cenozoic geological history of the Holarctic (Enghoff 1996) (Fig. 11C). Currently, the Holarctic is divided into a Nearctic and a Palaearctic land mass. However, both the Nearctic and Palaearctic were earlier separated into a western and eastern half by large epicontinental seaways. Before that, the areas again formed two large continents, but in a different configuration than today: the eastern Palaearctic was united with the western Nearctic via the Bering land bridge, and a similar land bridge across the Atlantic con-

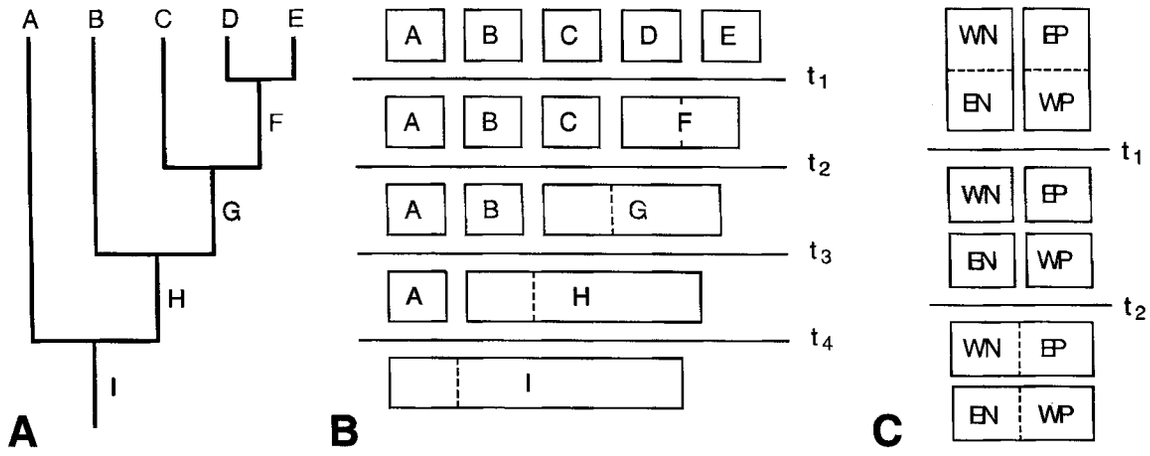

Fig. 11. Different types of biogeographic scenarios.—*A.* A branching scenario shown as a tree (corresponding to a host phylogeny).—*B.* The same scenario illustrated as maps of the areas involved and their connections at different times.—*C.* A simplified model of the Cenozoic geological history of the Holarctic (Enghoff 1996), exemplifying a reticulate biogeographic scenario that cannot be represented in the form of a tree. WN = western Nearctic, EN = eastern Nearctic, WP = western Palaearctic, EP = eastern Palaearctic.

nected Europe with the eastern Nearctic. This biogeographic scenario cannot be represented in the form of a branching diagram.

In analysing such reticulate biogeographic scenarios, we need to focus on two types of events that are predicted by the scenario and that may affect the distribution history of many groups of organisms simultaneously (Fig. 12). The first is vicariance, which we expect to see in the species inhabiting a large, continuous area when that area is divided into smaller parts. The second is dispersal in response to the disappearance of a previous dispersal barrier. It is important to separate this type of predicted dispersal from random colonisation of disjunct areas (cf. Ronquist 1997). Unlike vicariance, which corresponds to cospeciation, there is no direct coevolutionary counterpart to predicted dispersal.

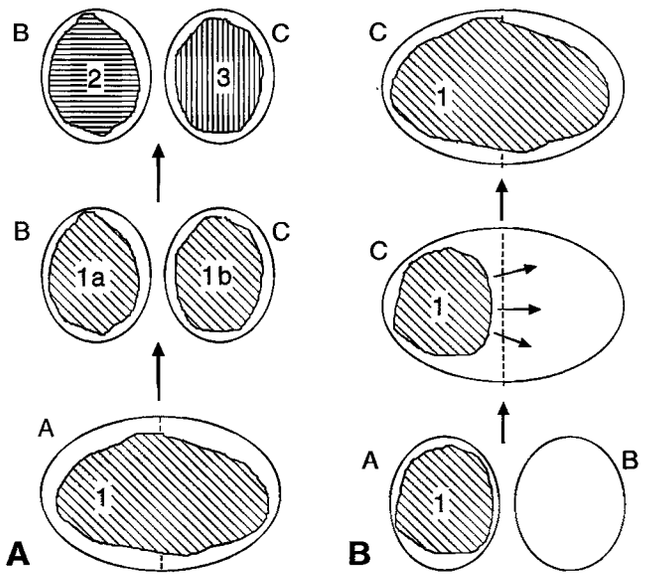

Fig. 12. Predicted events to consider in biogeographic analysis.—*A.* Vicariance, corresponding to cospeciation in coevolutionary analysis. A continuous area (A) is divided into two disjunct areas (B and C). Species (1) occurring in the original area become subdivided (1a and 1b), and eventually differentiate into separate lineages (2 and 3).—*B.* Dispersal in response to the disappearance of a previous dispersal barrier (no counterpart in coevolutionary analysis). Two separate areas (A and B) become united in a single area (C). In response, species (1) occurring in one of the initial areas tend to disperse over the previous dispersal barrier and occupy the entire continuous area.

Table III. *Suggested event-cost assignments of constrained dispersal-vicariance analysis.*

Event	Cost
Duplication	0
Random dispersal	1
Predicted dispersal	−1
Extinction	1
Vicariance	−1

Although this has not been attempted so far (but see Ronquist 1997), it is possible to construct a method analogous to maximum cospeciation by maximising the number of vicariance events and predicted dispersal events. However, all other types of events cannot be ignored (assigned zero cost) as in maximum cospeciation. Because extinctions can wipe out all traces of a predicted dispersal, spurious events will be included in optimal reconstructions unless extinctions carry a cost. A simple way of solving this dilemma is to assign random dispersals and extinctions a cost equivalent to the unit benefit of predicted dispersal and vicariance events (Table III). Since this type of method focuses on both dispersal and vicariance events, and constrain reconstructions according to a general biogeographic scenario, it may be appropriately termed constrained dispersal-vicariance analysis (cf. Ronquist 1997). Constrained dispersal-vicariance analysis or similar methods could be used in comparing the distribution history of Holarctic organisms against reticulate biogeographic scenarios (Fig. 11C).

Human origins

If a group of organisms is younger than any geological events affecting the areas they inhabit, it is possible to use very simple parsimony methods to infer ancestral areas and dispersal events. Similar simplifications apply to

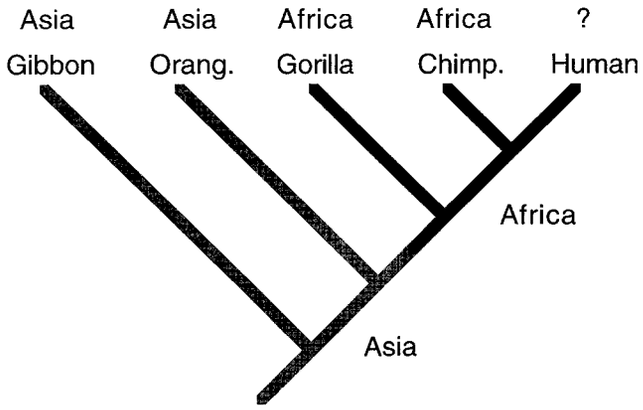

Fig. 13. A simple parsimony analysis of the geographical origin of humans employing Fitch optimisation of two distribution states, Southeast Asia and Africa.

coevolutionary problems where the parasites are much younger than their hosts.

Consider the geographical origin of humans as an example. Two areas are of interest, Southeast Asia and Africa. Assume that dispersal may occur between the areas, but that species distributions comprising both areas is a transitional phenomenon, at least in an evolutionary perspective. We can then find the most parsimonious biogeographic reconstruction by simply minimising the number of dispersals, i.e., the number of changes between the area "states". This is called Fitch optimisation. If we treat the original distribution of humans as unknown (or polymorphic), parsimony indicates that humans originally occurred in Africa (Fig. 13). Palaeontological data essentially confirm this scenario, except that the African ancestry may stretch further back in time in the hominoid-hylobatid lineage (Sola & Kohler 1993; Shoshani *et al.* 1996).

We know that hominids (*Homo erectus*) were widespread about 1.5 to 2 million years ago. Yet, several gene trees indicate that modern humans all descend from a small population that presumably lived in a confined area about 200 000 years ago and then spread over the globe, displacing other human populations (Ruvolo 1996). By inferring the distribution history of each of the gene trees using parsimony methods, it would be possible to locate the area in which this population lived. Each gene tree would have its own distribution history, but the ancestral area would be the same (as well as the coalescence time). In the first analyses of this kind, based on mitochondrial DNA, it was concluded that the ancestral area was Africa (e.g., Vigilant *et al.* 1991). More detailed analyses have shown that this conclusion is uncertain (Maddison *et al.* 1992). It is possible to root the mitochondrial tree in several different places, with insignificant differences in tree length. Some of these rootings place the origin of modern humans in Africa, others in Asia or Papua New Guinea. In other words, we cannot yet say with certainty where this ancestral population lived.

Conclusions

In summary, I have tried to point out the significant similarities between the phylogenetic perspectives on coevolution and biogeography, and that similar parsimony methods can be used in both disciplines. Phylogenetic approaches in coevolution and biogeography span a fascinating variety of important biological problems at different spatial and temporal scales, and I have attempted to give some examples of this. Methods for coevolutionary and biogeographic analysis are still being developed, and the coming decades will undoubtedly see rapid progress, not only in coevolution and biogeography but also in areas that will unexpectedly turn out to be amenable to similar types of analyses. A good example of this is how coevolutionary analysis can be applied to study the relation between gene trees and organism trees, an approach that, among other things, may help root the tree of life more accurately than has been possible before.

References

Avise, J. C. 1994. *Molecular Markers, Natural History and Evolution.* Chapman & Hall, New York.

Avise, J. C., Arnold, J., Ball, R. E., Bermingham, E., Lamb, T., Neigel, J. E., Reeb, C. A. & Saunders, N. C. 1987. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics.—*A. Rev. Ecol. Syst. 18:* 489–522.

Brooks, D. R. 1988. Macroevolutionary comparisons of host and parasite phylogenies.—*A. Rev. Ecol. Syst. 19:* 235–259.

Brown, J. R. & Doolittle, W. F. 1995. Root of the universal tree of life based on ancient aminoacyl-tRNA synthetase gene duplications.—*Proc. natn. Acad. Sci. U.S.A. 92:* 2441–2445.

Crisci, J. V. & Morrone, J. J. 1995. Historical biogeography: Introduction to methods.—*A. Rev. Ecol. Syst. 26:* 373–401.

De Rijk, P., Van de Peer, Y., Van den Broeck, I. & De Wachter, R. 1995. Evolution according to large ribosomal subunit RNA.—*J. molec. Evol. 41:* 366–375.

Enghoff, H. 1996. Historical biogeography of the holarctic: Area relationships, ancestral areas, and dispersal of non-marine animals.—*Cladistics 11:* 223–263.

Glen, D. R. & Brooks, D. R. 1985. Parasitological evidence pertaining to the phylogeny of the hominoid primates.—*Biol. J. Linn. Soc. 27:* 331–354.

Goodman, M., Czelusniak, J., Moore, G. W., Romero-Herrera, A. E. & Matsuda, G. 1979. Fitting the gene lineage into its species lineage: A parsimony strategy illustrated by cladograms constructed from globin sequences.—*Syst. Zool. 28:* 132–168.

Grubb, R. 1994. Normalflora. In (Engström, C. *et al.*, eds.), pp. 264. *Nationalencyklopedin* Bra Böcker, Höganäs.

Guigó, R., Muchnik, I. & Smith, T. F. 1996. Reconstruction of ancient molecular phylogeny.—*Molec. Phyl. & Evol. 6:* 189–213.

Gupta, R. S. & Golding, G. B. 1993. Evolution of HSP70 gene and its implications regarding relationships between Archaebacteria, Eubacteria, and Eukaryotes.—*J. molec. Evol. 37:* 573–582.

Hafner, M. S. & Page, R. D. M. 1995. Molecular phylogenies and host-parasite cospeciation: Gophers and lice as a model system.—*Phil. Trans. R. Soc. B 349:* 77–83.

Hafner, M. S., Sudman, P. D., Villablanca, F. X., Spradling, T. A., Demastes, S. W. & Nadler, S. A. 1994. Disparate rates of molecular evolution in cospeciating hosts and parasites.—*Science 265:* 1087–1090.

Huelsenbeck, J., Rannala, B. & Yang, Z. 1997. Statistical tests of host-parasite cospeciation.—*Evolution 51:* 410–419.

Iwabe, N., Kuma, K.-I., Hasegawa, M., Osawa, S. & Miyata, T. 1989. Evolutionary relationship of archaebacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes.—*Proc. natn. Acad. Sci. U. S. A. 86:* 9355–9359.

Klein, N. K., Payne, R. B. & Nilane, M. E. D. 1993. A molecular genetic perspecitve on speciation in the brood parasitic *Vidua* finches.—*Annls Mus. r. Afr. cent. 268:* 29–39.

Leigh Brown, A. J. & Holmes, E. C. 1994. Evolutionary biology of human immunodeficiency virus.—*A. Rev. Ecol. Syst. 25:* 127–165.

Maddison, D. R., Ruvolo, M. & Swofford, D. L. 1992. Geographic origins of human mitochondrial DNA: Phylogenetic evidence from control region sequences.—*Syst. Biol. 41:* 111–124.

Mindell, D. P., Schultz, J. W. & Ewald, P. W. 1995. The AIDS pandemic is new, but is HIV new?—*Syst. Biol. 44:* 77–92.

Mirkin, B., Muchnik, I. & Smith, T. F. 1995. A biologically consistent

model for comparing molecular phylogenies.—*J. comput. Biol. 2:* 493–507.

Moran, N. & Baumann, P. 1994. Phylogenetics of cytoplasmically inherited microorganisms of arthropods.—*Trends Ecol. Evol. 9:* 15–20.

Moran, N. A., Baumann, P. & Von Dohlen, C. 1994. Use of DNA sequences to reconstruct the history of the association between members of the Sternorrhyncha (Homoptera) and their bacterial endosymbionts.—*Eur. J. Ent. 91:* 79–83.

Moran, N. A., Munson, M. A., Baumann, P. & Ishikawa, H. 1993. A molecular clock in endosymbiotic bacteria is calibrated using the insect hosts.—*Proc. R. Soc. B 253:* 167–171.

Page, R. D., Clayton, D. H. & Paterson, A. M. 1996. Lice and cospeciation: A response to Barker.—*Int. J. Parasit. 26:* 213–218.

Page, R. D. M. 1993. Genes, organisms, and areas: the problem of multiple lineages.—*Syst. Biol. 42:* 77–84.

Page, R. D. M. 1994. Maps between trees and cladistic analysis of historical associations among genes, organisms, and areas.—*Syst. Biol. 43:* 58–77.

Page, R. D. M. 1995. Parallel phylogenies: Reconstructing the history of host-parasite assemblages.—*Cladistics 10:* 155–173.

Page, R. D. M. 1996. Temporal congruence revisited: Comparison of mitochondrial DNA sequence divergence in cospeciating pocket gophers and their chewing lice.—*Syst. Biol. 45:* 151–167.

Page, R. D. M. & Charleston, M. A. 1997. From gene to organismal phylogeny: Reconciled trees and the gene tree/species tree problem.—*Molec. Phyl. & Evol. 7:* 231–240.

Page, R. D. M. & Hafner, M. S. 1996. Molecular phylogenies and host-parasite cospeciation: Gophers and lice as a model system. In *New uses for new phylogenies* (eds. Harvey, P. H., Leigh Brown, A. J., Maynard Smith, J. & Nee, S.): 255–270. Oxford University Press, Oxford.

Ronquist, F. 1996. Reconstructing the history of host-parasite associations using generalised parsimony.—*Cladistics 11:* 73–89.

Ronquist, F. 1997. Dispersal-vicariance analysis: A new approach to the quantification of historical biogeography.—*Syst. Biol. 46:* 195–203.

Ronquist, F. in press. Three-dimensional cost matrix optimisation and maximum cospeciation.—*Cladistics.*

Ronquist, F. & Liljeblad, J. submitted. Evolution of the gall wasp-host plant association.—*Evolution.*

Ronquist, F. & Nylin, S. 1990. Process and pattern in the evolution of species associations.—*Syst. Zool. 39:* 323–344.

Ruvolo, M. 1996. A new approach to studying modern human origins: Testing with coalescence time distributions.—*Molec. Phyl. & Evol. 5:* 202–219.

Shoshani, J., Groves, C. P., Simons, E. L. & Gunnell, G. F. 1996. Primate phylogeny: Morphological vs molecular results.—*Molec. Phyl. & Evol. 5:* 102–154.

Siddall, M. E. 1997. The AIDS pandemic is new, but is HIV *not* new?—*Cladistics 13:* 267–273.

Sola, S. & Kohler, M. 1993. Recent discoveries of *Dryopithecus* shed new light on evolution of great apes.—*Nature 365:* 543–545.

Vigilant, L., Stoneking, M., Harpending, H., Hawkes, K. & Wilson, A. C. 1991. African populations and the evolution of human mitochondrial DNA.—*Science 253:* 1503–1507.

Woese, C. R., Kandler, O. & Wheelis, M. L. 1990. Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria and Eucarya.—*Proc. natn. Acad. Sci. U.S.A. 87:* 4576–4579.

Zhang, L. 1997. On a Mirkin–Muchnik–Smith conjecture for comparing molecular phylogenies.—*J. comput. Biol. 4:* 177–187.