

Report on NERC Grant GR3/1A095 “Molecular phylogenies and host-parasite cospeciation”

Roderic D. M. Page

Division of Environmental and Evolutionary Biology

Institute of Biomedical and Life Sciences

Graham Kerr Building, University of Glasgow, Glasgow G12 8QQ

1 Introduction

A key question in the study of coevolution is the tempo and mode of evolution of the interacting partners. For example, how old is the association between a given parasite and its host? Is the association an ancient one, reflecting a long and intimate interaction between the two organisms, or is it a recent event due perhaps to a parasite colonising a new host? Answering these questions requires the comparison of evolutionary trees (phylogenies) for host and parasite. To the extent that the trees match, host and parasite have cospeciated, that is, parasite and host speciate at the same time. Mismatches between the host and parasite trees implies processes other than cospeciation, such as host switching, speciation by parasites independently of their hosts, and parasite extinction. Documenting the extent of cospeciation is both a fundamental step in interpreting the history of a host-parasite assemblage, and also provides a temporal framework for comparative analyses of the evolution of host and parasite.

The aim of the work reported here was to “develop a general method for recovering the history of host-parasite associations.” This method was to be implemented in user-friendly software, then evaluated using both simulations and empirical data. Although the immediate goal was to develop a quantitative method for phylogenetic analysis of host-parasite associations, there are close parallels with the problem of comparing gene trees and species trees (Page, 1993). Indeed, the starting point for this research was a technique originally developed to reconcile incongruent gene trees with species trees (Goodman *et al.*, 1979). Reconciling trees (Page, 1994a) is a useful way of depicting the history of a gene tree embedded within a species tree, and have been applied to host-parasite studies (Page, 1990; Paterson *et al.*, 1993). However, this technique is limited by its failure to accommodate horizontal transfer. Whereas this is less of a problem in studying gene trees, it is a major handicap in systems where horizontal transfer is potentially common.

2 Horizontal transfer

2.1 *The problem*

Parasite phylogeny mirrors host phylogeny with a degree of fidelity that depends on the relative frequency of four basic categories of events: cospeciation, duplication, horizontal transfer, and sorting (Figure 1). **Cospeciation** is joint speciation of host and parasite. If the parasite speciates independently of the host there are two possibilities: both parasite lineages remain on the host (**duplication**), or one or both

lineages switch to new hosts (**horizontal transfer**). “**Sorting event**” is a generic term (Page, 1994a) for the apparent absence of a parasite from a host upon which it would have expected to occur because the host’s ancestor harboured either that parasite, or its ancestor. This absence could be due to at least three reasons: (1) extinction of the parasite; (2) failure of the parasite to be transmitted to the descendant host (“missing the boat,” see Paterson and Gray, 1997), and (3) the parasite is actually present but is as yet undiscovered.

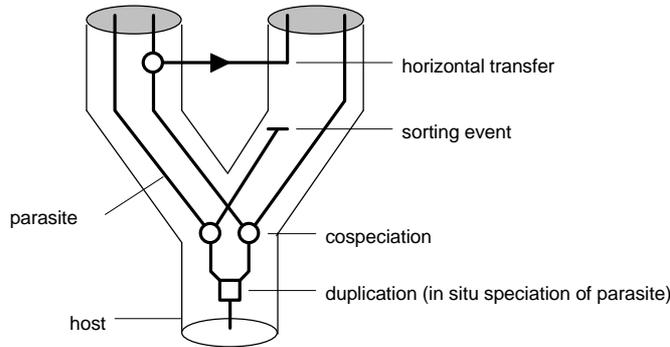


Figure 1. The four categories of event in a host-parasite association (after Page, 1994b).

Reconciled trees accommodate cospeciation, duplication, and sorting, but not horizontal transfer. Hence a major goal of this project was to develop a method that could incorporate horizontal transfer. I had earlier attempted this (Page, 1994b), but Ronquist (1995) showed that my method could fail in some cases. Because postulating a horizontal transfer requires that the **source** and **destination** hosts are contemporaneous, we have to consider the relative ages of different host lineages. Failure to take this constraint into account can result in postulating transfers that are mutually incompatible (Figure 2).

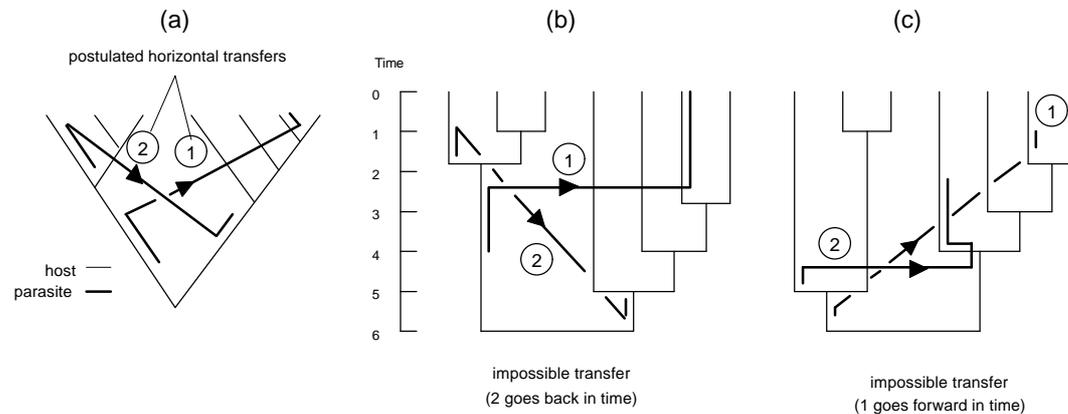


Figure 2. The problem of horizontal transfers.(a) Given this host phylogeny, the two postulated host switches (1 and 2) are incompatible. Ensuring that the source and destination hosts for switch 1 are contemporaneous (b) would require switch 2 to have gone backwards in time. Likewise, for switch 2 to occur on this tree, switch 1 would have to go forwards in time.

2.2 A solution (“jungles”)

The requirement that source and destination hosts are contemporaries greatly complicates the task of developing a method of recovering the history of the association between a parasite and its host. Prior to this project there was no satisfactory method for dealing with horizontal transfer. Solving this problem occupied most of the two years devoted to this project. The RA (Mike Charleston) developed a new mathematical structure called a “jungle” (Charleston, in press) that depicts all the possible associations between a given parasite and host phylogeny, subject to the constraint that no impossible host switches are postulated. Having enumerated all the possible solutions, a dynamic programming approach is used to find the optimal solution (which corresponds to a subtree of the jungle). A potentially very useful feature of jungles is that the four different categories of events can be assigned different costs. These costs could be assigned based on the relative probabilities of the different events determined from empirical studies.

Something of the complexity of jungles can be appreciated from the following example of a pair of host and parasite trees:

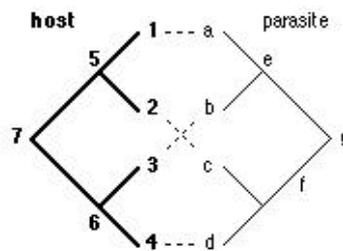


Figure 3. Two simple host and parasite phylogenies.

For these trees the corresponding jungle is:

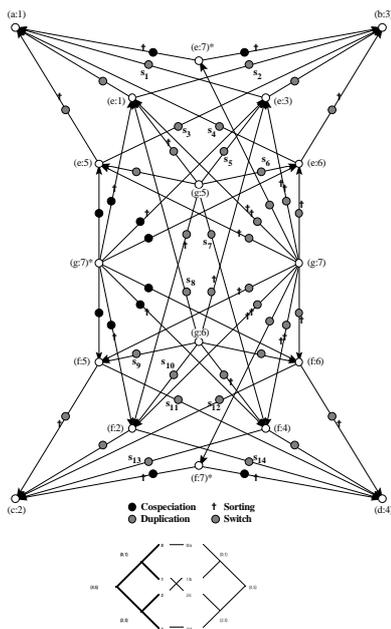


Figure 4. Jungle for the two trees shown in Figure 3. Each node in the jungle represents a possible association between host and parasite. Four of these associations (a:1, b:3, c:2, and d:4) are observed (i.e., parasites a-d on hosts 1-4) the remaining associations are between ancestral parasites and their hosts. A valid reconstruction of the history the host-parasite association is a subtree in the jungle that includes all nodes in the parasite phylogeny (see Figure 5). Any subtree has an associated “cost” in terms of events such as cospeciation and host switching. A reconstruction that minimises this cost is optimal and is considered the best reconstruction.

Details of how this jungle is constructed, and how optimal solutions are found are given in Charleston (in press), but as an example Figure 5 shows one subtree in the jungle and the corresponding reconstruction.

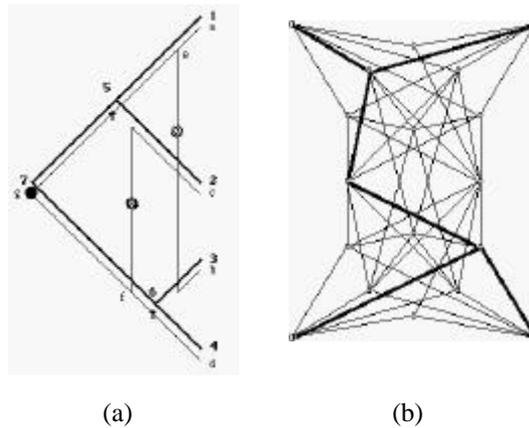


Figure 5. (a) One possible reconstruction of the history of the host-parasite assemblage shown in Figure 3, and (b) the corresponding pathway (heavy lines) in the jungle shown in Figure 4. This reconstruction requires two host shifts.

It should be noted that if we impose the constraint that there is no host switching then there is just a single optimal reconstruction, which corresponds to the reconciled tree. Hence, jungles encompass reconciled trees as a special case.

2.3 *Ongoing work on jungles*

At the time the grant ended (and the RA left for a post doc at Oxford) the software implementation of jungles was about 90% complete. The core code is running (on a Power Mac and a Sun workstation), however, the program lacks a useable interface. This will be added in the coming months by the PI. To date jungles have been applied to the classic Hafner and Nadler (1988) gopher-louse data set, as well as several hypothetical examples (Charleston, in press).

3 Simulations

Because of the amount of time developing jungles required the simulation part of the project did not receive as much attention as at first planned. Some simulations were run using a simple model that included rates of speciation and extinction, and the probability that a parasite when speciating would remain on its host (either cospeciation or duplication), or switch hosts. This work is in preparation, but the most striking preliminary outcome is that the same set of parameters can yield widely different results, and that even with a low probability of host switching the number of potential reconstructions is highly variable. This means that once host switching reaches a certain frequency it may be very hard to accurately reconstruct the history of a host-parasite assemblage. This work is ongoing.

4 Empirical studies

Suitable data sets for phylogenetic studies of host-parasite cospeciation are still rare. In part this is due to the demanding requirements for a good data set (Page *et al.*, 1996):

- Adequate alpha-taxonomy of both hosts and parasites.
- Accurate phylogenies of host and parasites
- Exhaustive sampling of clades of parasites
- Molecular phylogenies based on comparable genes
- Quantitative comparison of host and parasite phylogenies
- Host transfer experiments to test host specificity

Of the three data sets listed in the original proposal, the swiftlet-louse data set best met these criteria. This data set was developed by colleagues at Oxford University concurrently with this project. Given the need for data sets to test methods of cospeciation analysis, I spent some time and resources contributing to the analysis of these data, including alpha-taxonomy of the lice (Clayton *et al.*, 1996) and molecular phylogenetic analysis of the hosts (Lee *et al.*, 1996). As part of this project I compared the phylogeny of the lice with those of their hosts using the TREEMAP program as the Jungle software is still in development. This analysis (Page *et al.*, in press) provided evidence for cospeciation of the lice with their avian hosts (Figure 6).

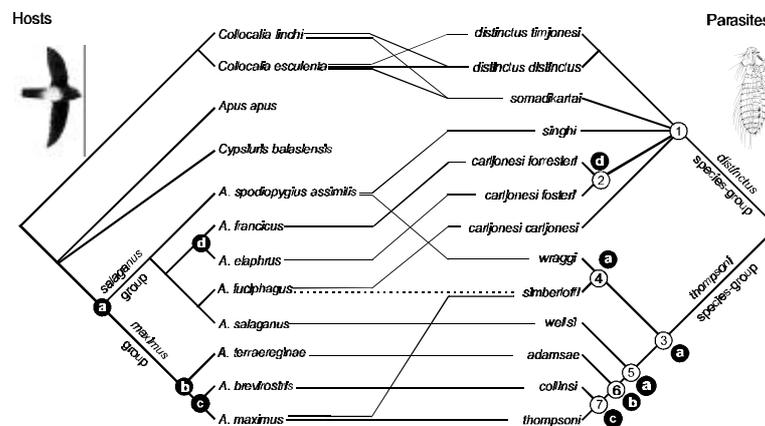


Figure 6. Phylogenies of swifts and swiftlets, and their parasitic lice. Thin lines connect the lice to their hosts. Both phylogenies were inferred from mitochondrial DNA cytochrome *b* sequences. Putative cospeciation events are indicated by labels in black circles (from Page *et al.*, in press).

Using these cospeciation events to compare relative rates of evolution in the birds and their lice demonstrated that the lice are evolving about 2.5 times more rapidly than their hosts at the cytochrome *b* locus. Comparisons with other insects suggests that louse mitochondrial DNA is among the most rapidly evolving in the Insecta. This rapidity of evolution in lice may reflect their small effective population size (Lee and Clayton, 1995), coupled with repeated founder events as new host individuals are colonised. These hypotheses are currently being tested in NERC funded work (GR3/11075 awarded to the PI and two co PIs) on seabirds and their lice.

5 Other work

In the course of this project two other areas of research were pursued, namely gene trees and spectral analysis.

5.1 *Gene trees*

The gene tree/species tree problem is intimately related to the study of host-parasite cospeciation — in both cases one entity (the gene or parasite) is tracking another (the species or host). Hence I thought it worthwhile to investigate reconciled trees in more detail (which are a special case of jungles, see section 2.2), especially as new mathematical work on reconciled trees was appearing (Guigó *et al.*, 1996; Mirkin *et al.*, 1995; Zhang, 1997), and the RA has expertise in heuristic search methods. Together we published two papers on constructing reconciled trees, and using them to infer organismal phylogeny (Page and Charleston, 1997a; 1997b).

5.2 *Spectrum*

The grant also funded the development of the program SPECTRUM by Mike Charleston. Spectral analysis (Hendy and Penny, 1993) is a useful tool for visualising signal in phylogenetic data, but prior to this project had not been available in a user friendly form. Anticipating that this approach would be useful in the analyses of the empirical data used in this study, and given that much of the code had already been developed by the RA in his PhD thesis (Charleston, 1994), it seemed sensible to devote some resources to writing this program¹. SPECTRUM is available over the Internet (<http://taxonomy.zoology.gla.ac.uk/~mac/spectrum/spectrum.html>), and was used in the analysis of the swiftlet louse DNA data (Page *et al.*, in press).

6 Summary and Future Directions

The major achievement of this research is a solution (jungles) to the problem of reconstructing the history of a host-parasite assemblage when horizontal transfer is allowed. Previous methods either prohibited horizontal transfer or attempted to incorporate it but failed (Page, 1994a). Jungles have considerable potential application to other problems, such as horizontal transfer of genes, and dispersal in biogeography. Because the approach enables weights to be assigned to each of the four types of event (Figure 1), the method can be generalised to produce maximum likelihood estimates of the history of parasites and their hosts. Jungles will be shortly included in a new version of the PI's program TREEMAP (Page, 1994b), the current version is available

¹ The other motivation was that I had just taken up the position at Glasgow a month before the RA started and was still settling in. Hence I did not have much time to devote to the project at the beginning of the RA's appointment. By writing SPECTRUM, the RA was contributing to the project, but without relying on me for direct guidance. The other benefit was that the RA used SPECTRUM to master the C++ programming language, which was used for the work on jungles.

on the Internet (<http://taxonomy.zoology.gla.ac.uk/rod/treemap.html>) and has been used in a number of cospeciation studies (e.g., Bécerra, 1997; Paterson and Gray, 1997).

Empirically, the lack of good data sets proved a problem. However, the swiftlet-louse cytochrome *b* data set provided the first evidence that louse mitochondrial DNA (mtDNA) is evolving more rapidly than bird mtDNA, and analysis of this data shows that lice are among the most rapidly evolving insects (Page *et al.*, in press). The reasons for this (and whether the observed increased rate in louse mtDNA also applies to nuclear DNA) are the subject of a NERC project just commencing (GR3/11075).

Lastly, the work on gene trees reflects the generality of the problem of historical associations, of which hosts and parasites are but one instance. The increasing attention being paid to gene trees and species trees (Doyle, 1997; Maddison, 1997) suggests that this is an area which deserves further attention. The inference of species trees from gene trees is an obvious further area of application for jungles, given the presence of horizontal gene transfer (e.g., Delwiche and Palmer, 1996).

6.1 Publications funded by this grant

Papers supported by GR3/1A095 are indicated in **boldface** in the list of references.

6.2 Resources on the Internet

The Taxonomy at Glasgow Web server (<http://taxonomy.zoology.gla.ac.uk/>) is maintained by the PI, and some work resulting from this project has been placed there, including:

- A preprint of Charleston's (in press) paper on jungles.
- A poster of the swiftlet-louse cospeciation work (Page *et al.*, in press).
- Preprints of the work on gene trees and species trees (Page and Charleston, 1997a; Page and Charleston, 1997b).
- The program SPECTRUM.

In the near future a program implementing jungles will be placed on this server.

7 References (papers directly funded by this grant in **boldface**)

- Bécerra, J. X. (1997). Insects on plants: macroevolutionary chemical trends in host use. *Science* **276**: 253-256.
- Charleston, M. A. (1994). "Factors affecting the performance of phylogenetic methods". PhD, Massey University.
- Charleston, M. A. (in press). Jungles: A new solution to the host/parasite phylogeny reconciliation problem. *Math. Biosci.***
- Clayton, D. H., Price, R. D. and Page, R. D. M. (1996). Revision of *Dennyus (Collodennyus)* lice (Phthiraptera: Menoponidae) from swiftlets, with descriptions of new taxa and a comparison of host-parasite relationships. *Syst. Entomol.* **21**: 179-204.**

- Delwiche, C. F. and Palmer, J. D. (1996). Rampant horizontal transfer and duplication of rubisco genes in Eubacteria and plastids. *Mol. Biol. Evol.* **13**: 873-882.
- Doyle, J. J. (1997). Trees within trees: Genes and species, molecules and morphology. *Syst. Biol.* **46**: 537-553.
- Goodman, M., Czelusniak, J., Moore, G. W., Romero-Herrera, A. E. and 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.
- Guigó, R., Muchnik, I. and Smith, T. F. (1996). Reconstruction of ancient molecular phylogeny. *Mol. Phylogenet. Evol.* **6**: 189-213.
- Hafner, M. S. and Nadler, S. A. (1988). Phylogenetic trees support the coevolution of parasites and their hosts. *Nature* **332**: 258-259.
- Hendy, M. D. and Penny, D. (1993). Spectral analysis of phylogenetic data. *J. Classif.* **10**: 5-24.
- Lee, P. L. M. and Clayton, D. H. (1995). Population biology of swift (*Apus apus*) ectoparasites in relation to host reproductive success. *Ecol. Entomol.* **20**: 43-50.
- Lee, P. L. M., Clayton, D. H., Griffiths, R. and Page, R. D. M. (1996). Does behaviour reflect phylogeny in swiftlets (Aves: Apodidae)? A test using cytochrome b mitochondrial DNA sequences. *Proc. Natl. Acad. Sci., USA* **93**: 7091-7096.**
- Maddison, W. P. (1997). Gene trees in species trees. *Syst. Biol.* **46**: 523-536.
- Mirkin, B., Muchnik, I. and Smith, T. F. (1995). A biologically consistent model for comparing molecular phylogenies. *J. Comput. Biol.* **2**: 493-507.
- Page, R. D. M. (1990). Temporal congruence and cladistic analysis of biogeography and cospeciation. *Syst. Zool.* **39**: 205-226.
- 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. and Charleston, M. A. (1997a). From gene to organismal phylogeny: Reconciled trees and the gene tree/species tree problem. *Mol. Phylogenet. Evol.* **7**: 231-240.**
- Page, R. D. M. and Charleston, M. A. (1997b). Reconciled trees and incongruent gene and species trees. In "Mathematical hierarchies in biology. DIMACS Workshop November 13-15, 1996" (B. Mirkin, F. R. McMorris, F. S. Roberts and A. Rzhetsky, Eds.), Vol. 37, pp. 57-70, American Mathematical Society.**
- Page, R. D. M., Clayton, D. H. and Paterson, A. M. (1996). Lice and cospeciation: A response to Barker. *Int. J. Parasitol.* **26**: 213-218.**
- Page, R. D. M., Lee, P. L. M., Becher, S. A., Griffiths, R. and Clayton, D. H. (in press). A different tempo of mitochondrial DNA evolution in birds and their parasitic lice. *Mol. Phylogenet. Evol.***
- Paterson, A. M. and Gray, R. D. (1997). Host-parasite cospeciation, host switching, and missing the boat. In "Host-parasite evolution: General principles and avian models" (D. H. Clayton and J. Moore, Eds.), pp. 236-250, Oxford University Press, Oxford.
- Paterson, A. M., Gray, R. D. and Wallis, G. P. (1993). Parasites, petrels and penguins: Does louse presence reflect seabird phylogeny? *Int. J. Parasitol.* **23**: 515-526.
- Ronquist, F. (1995). Reconstructing the history of host-parasite associations using generalised parsimony. *Cladistics* **11**: 73-89.
- Zhang, L. (1997). On a Mirkin-Muchnik-Smith conjecture for comparing molecular phylogenies. *J. Comput. Biol.* **177**-187.