Extending phylogenetic studies of coevolution: secondary Brooks parsimony analysis, parasites, and the Great Apes

Daniel R. Brooks* and Deborah A. McLennan

Department of Zoology, University of Toronto, 25 Harbord Street, Toronto, Ont., Canada M5S 3G5

Accepted 10 February 2003

Abstract

Dowling recently compared the empirical properties of Brooks parsimony analysis (BPA) and the leading method for studying phylogenetic aspects of coevolution, reconciliation tree analysis (using the computer program TreeMap), based on a series of simulations. Like the majority of authors who have compared BPA with other methods, however, Dowling considered only the form of BPA proposed in 1981 and did not take into account various modifications of the method proposed from 1986 to 2002. This leaves some doubt as to the robustness of his assessments of both the superiority of BPA and its shortcomings. We provide a précis of the principles of contemporary BPA, including ways to implement it algorithmically, using either Wagner algorithm-based or Hennigian argumentation-based approaches, followed by an empirical example. Our study supports Dowling’s fundamental conclusions about the superiority of primary BPA relative to TreeMap. However, his conclusions about the shortcomings of BPA due to inclusive ORing (i.e., the production of ghost taxa) are incorrect, as secondary BPA eliminates inclusive ORing from the method. Secondary BPA provides a more complete account of the evolutionary associations between the parasite groups and their hosts than does primary BPA, without sacrificing any indirectly generated information about host phylogeny. Secondary BPA of two groups of nematodes inhabiting Great Apes shows that TreeMap analysis underestimated the amount of cospeciation in the evolution of the nematode genus Enterobius.

© 2003 The Willi Hennig Society. Published by Elsevier Science (USA). All rights reserved.

Parasites became important model systems in evolutionary biology when orthogenetic evolution became fashionable in the early 20th century. Orthogeneticists believed that once a lineage became parasitic, it became so dependent on its host that it no longer had any independent evolutionary existence, becoming a wraith-like shadow of the host’s evolution. Even when NeoDarwinism replaced orthogenesis as the dominant theory of evolution, many biologists continued to assume that the phylogenetic relationships of parasites exhibiting pronounced host specificity must parallel the phylogenetic relationships of their hosts. Hennig (1950, 1966) referred to this assumption as “the parasitological method,” which he rejected on the basis that there were no objective criteria for distinguishing cases in which two or more hosts inhabited by a parasite species form a monophyletic group from those in which they do not. Brooks (1979) showed that parasites could have plesiomorphic, synapomorphic, autapomorphic, or even homoplasmous host associations, supporting Hennig’s objection. He coined the term cospeciation to refer to the congruent portions of host and parasite phylogenies. Although recognizing that a variety of processes could produce such macroevolutionary patterns, he tended to think that such congruence was most likely a by-product of vicariant cospeciation (see also Brooks and McLennan, 1991, 1993, 2002; Dowling, 2002).

Existing models of coevolution all deal with coaccommodation (Brooks, 1979), later called coadaptation (Brooks and McLennan, 1991, 1993), or various forms of mutual evolutionary modification between hosts and parasites which are anagenetic rather than cladogenetic processes. For this reason, Brooks (1979) proposed that assessments of host specificity (coaccommodation) were decoupled from assessments of the degree of phylogenetic association among host and parasite clades (co-
speciation); hence, one could not use degree of host specificity as a guide to determining which parasite clades were "likely" to have cospeciated with their hosts. Theoretical and experimental studies of coevolving systems of many kinds have supported this proposal (for a review and references, see Brooks and McLennan (2002)). Instead, Brooks (1979) supported the view that parasites could be treated as analogous to taxonomic characters of their hosts, with the proviso that, like all characters, not all parasites provided information about host phylogeny. Brooks (1981) extended this analogy, proposing a direct application of the inclusion/exclusion rule in phylogenetic systematics for studies of cospeciation: treat parasite phylogenetic trees as ordered and polarized multistate transformation series, treat the hosts as terminal taxa, run a phylogenetic systematic analysis of the data, and compare the resultant host cladogram (analogous to an area cladogram) with the host phylogenetic tree. Congruence between the two "trees" corroborates a hypothesis of cospeciation between the parasite clades, while incongruence falsifies it. Brooks also came to a radical conclusion: if parasites are evolutionarily independent entities, like free living organisms, then members of different parasite clades need not show identical patterns of evolutionary diversification with respect to host relationships, except in the default case of association by common ancestry (possibly indicating simply common episodes of vicariant speciation). He argued that even if each parasite clade included some species which did not cospeciate with their hosts, simultaneous analysis of multiple parasite clades inhabiting a given host group should provide evidence of host phylogeny as a byproduct of assessing the unique coevolutionary history of each parasite clade. Comparing the host cladogram produced by phylogenetic analysis of parasites with a host phylogeny generated using other data can test this hypothesis. Points of congruence between the two trees corroborate the hypothesis that the parasites and the hosts have cospeciated; points of incongruence falsify that hypothesis. This proposal echoed the then emerging phylogenetic systematic view that homology was determined a posteriori by character congruence (Patterson, 1981, 1982, 1987; Wiley, 1981). After this approach was applied to historical biogeography (Brooks, 1985), it was dubbed Brooks parsimony analysis or BPA (Wiley, 1986, 1988a,b).

Not all phylogeneticists are convinced that Brooks' Hennigian solution to the question of "the parasitological method" is adequate. For example, the "maximum cospeciation" research program (Page, 2002; Page and Charleston, 1998 and references therein) is based upon the traditional assumption that cospeciation lies at the core of many specialized associations, rejecting Brooks' (1979) argument that cospeciation and coaccommodation were decoupled evolutionarily. Advocates of this approach have developed a variety of methods designed to identify the parasite clades most likely to exhibit cospeciation patterns and then provide the maximum degree of fit between the phylogenies of those parasites and their hosts. Dowling (2002) recently compared the empirical properties of BPA and the leading method for determining maximum cospeciation, reconciled tree analysis (using the computer program TreeMap), based on a series of simulations. His results indicated that BPA was preferable to TreeMap for two reasons. "First, TreeMap grossly overestimates duplications and sorting events when widespread taxa due to host switching are present in the associations between the host and parasite phylogenies. Second, the ghost taxa that BPA mistakenly produces do not cause any topological changes in the tree, are readily recognizable, and are easy to interpret." Like the majority of authors who have compared BPA with other methods (e.g., Morrone and Carpenter, 1994; Morrone and Crisci, 1995; Page, 1990; Page and Charleston, 1998; Ronquist, 1995, 1996, 1997a,b, 1998), Dowling considered only the form of BPA proposed in 1981, and did not take into account modifications of the method (Brooks, 1990; Brooks and McLennan, 1991, 1993, 2002; Brooks et al., 2001; Green et al., 2002; Van Veller and Brooks, 2001; Wiley, 1986, 1988a,b). This leaves some doubt as to the robustness of Dowling's conclusions about both the superiority of BPA and its shortcomings. In this paper, therefore, we provide a précis of contemporary BPA, followed by an empirical example using pinworm–roundworm–Great Ape interactions, to assess Dowling's conclusions in light of contemporary BPA.

Methodology

Part 1: Starting points

(Assumption 0). All species and all host distributions in each input parasite phylogeny must be considered without modification, and the final analysis must be logically consistent with all input data. This assumption is important because it mandates a minimum of constraints on any analysis by prohibiting the researcher from eliminating, modifying, or weighting data a priori (Wiley, 1986, 1988a,b; Zandee and Roos, 1987). Page (1990) coined the term "Assumption 0 Analysis" for the methodological approach that codes hosts from which members of a parasite clade are missing as primitive lacking the parasite clade. This shortcoming in the original BPA formulation (Brooks, 1981) occurred because computer programs at the time did not accept missing data entries. Five years later, the computer programs had been modified and Wiley (1986, 1988a,b) solved this particular problem by suggesting that "absence" be coded a priori as "missing" (?) and interpreted a posteriori as either primitive absence or secondary loss
We term this the Missing Data Coding Protocol: all cases of “absence” should be coded a priori as “missing.” By coding all absences as “missing” a priori, BPA requires the investigator to account for such missing data codes a posteriori, i.e., after the host cladogram has been produced. This is done by character optimization (see Fig. 9 and associated discussion below). “Assumption 0 Analysis” of Page (1990), implemented in various computer programs, is thus not BPA, which relies on both the original Assumption 0 and the Missing Data Coding Protocol.

**Part 2: Distinguish general and special elements**

Cospeciation between hosts and parasites as a byproduct of vicariance should be the null hypothesis for studies of coevolution because it is the only mode of speciation that will always produce episodes of congruent host and parasite speciation (Brooks, 1981, 1985; Brooks and McLennan, 1993, 2002; Dowling, 2002). Other modes of parasite speciation (e.g., sequential colonization of hosts) may, but need not, produce cospeciation patterns (see Brooks and McLennan (2002)). Therefore, whenever we find pairs of sister species of parasites representing at least two clades that show the same host relationships (Fig. 1), we invoke cospeciation even in the absence of an independent host phylogeny, because we have evidence that the two parasite groups share a common history of speciation. Requiring two or more congruent speciation events is simply the methodological “cost” of invoking vicariance and does not imply a priori assumptions about the probability or likelihood of the occurrence of cospeciation in general or for the study groups in particular. Two additional evolutionary phenomena produce patterns that are always logically consistent with any general cospeciation pattern. These are (1) cases in which single species are widespread in two or more hosts because they did not respond to a vicariance event that produced sister species of hosts (Fig. 2) and (2) episodes of sympatric speciation (sometimes called “lineage duplication”), which produce replicate pairs of sister species of the same clade in the same host species (Fig. 3).

A number of evolutionary phenomena produce special or unique coevolutionary elements that do not conflict with the general cospeciation patterns but indicate novel host relationships; e.g., host switching by members of one clade to a host not inhabited by other members of its clade or by any members of the clades being analyzed. Such dispersal may take the form of host switching without speciation (postspeciation dispersal) (Fig. 4), which increases host range, or of host switching with speciation (peripheral isolates speciation) (Fig. 5), which does not increase host range (for a discussion of speciation by host switching as a form of peripheral isolates allopatric speciation see Brooks and McLennan, 1993, 2002; Funk et al., 1995; Funk, 1998). In the latter case, clade 1 is absent from host E because its members were never associated with that host. This means that we would not expect host E to be the sister species of host D, hence E is not represented in the host phylogeny (Fig. 5d). There is, however, an additional interpretation of Fig. 5. If species 11 and 12 have been produced as the result of a cospeciation event (meaning, among other things, that host E might be the sister species of host D), then it is possible that the same event

---

**Fig. 1.** Minimal support for cospeciation: two parasite clades supporting the same host relationships indicate episodes of mutual speciation by the parasite clades. The default expectation is that these host relationships are congruent with host phylogeny, which can be tested if an independently derived estimate of host phylogeny is available. Open circles highlight putative instances of vicariant speciation (the default speciation mode)/cospeciation events.
fragmented clade 1, producing species 4 (in host D) and its sister species (in host E), which later became extinct. If we examine only two clades we can identify the problem of missing taxa, but we cannot determine a solution to that problem because we do not know whether the ambiguity is caused by the presence or absence of species in the unusual host (in this case host E). In other words, we do not know which alternative represents the general pattern. This leads to the “Three’s Rule” rule: to distinguish between general and special patterns of historical association based solely on the data at hand, in particular to determine when absence from a host is due to secondary extinction or primitive absence, at least three clades inhabiting the same host group must be analyzed.

For example, suppose there is a third group of parasites with representatives in all five hosts shown in Fig. 5. The new host cladogram (Fig. 6) indicates that the presence of species 12 and species 21 in host E conforms to a general pattern. It is thus the absence of the sister to species 4 (clade 1) in host E that is the unique element, and it is most parsimonious to invoke an episode of secondary extinction to explain that absence. In the absence of other information, we would also expect that host E is the sister species of host D.
Part 3: Recognize hosts that have reticulate histories of association with their parasites

A number of evolutionary phenomena produce co-evolutionary elements that conflict with cospeciation patterns, e.g., host switching among members of one parasite clade to hosts already inhabited by other members of that parasite clade or by any members of the other parasite clades being analyzed. Such episodes of postspeciation expansion of host range and peripheral isolates speciation by host switching create more than one set of relationships for the same host species (i.e., some host species have reticulate histories with respect to their parasite faunas). The only way to satisfy Assumption 0 in such cases is to duplicate the hosts with reticulate histories until all such speciation events involving those hosts are accounted for (Figs. 7 and 8). This is called the Host Duplication Convention: whenever hosts have reticulate histories with respect to the parasite species inhabiting them, Assumption 0 will be violated.
unless those hosts are represented as separate entities for each separate historical episode. In this case the ambiguous hosts should be duplicated until Assumption 0 is satisfied.

The host duplication convention is an extension of the principles of total evidence, logical parsimony, and falsification embodied in Assumption 0. Our null hypothesis is that all hosts have a single history with respect to all the species of parasites that inhabit them.

When all available data (total evidence) falsify that null hypothesis, BPA duplicates hosts to indicate the exact departures from (falsifications of) the null hypothesis but does not multiply hosts beyond necessity (logical parsimony: Wiley, 1975; see also Kluge, 1997, 1998, 1999).

Brooks (1981) proposed that when two or more members of the same parasite clade occur in the same host, the binary codes representing each species and its
Fig. 8. Peripheral isolates speciation by host switching into a host utilized by another member of the parasite clade. Such reticulate host relationships can be detected using the Host Duplication Convention.

Fig. 9. A posteriori explanations for “missing” parasites. The absence of any member of clade 1 in host A is not due to extinction, because the most parsimonious interpretation of the data is that clade 1 did not become associated with the host clade until the common ancestor of BCD. The absence of any member of clade 3 in host C, however, is most parsimoniously explained as an extinction event, because clade 3 became associated with the host clade in the common ancestor of ABCD and because clades 1 and 2 have sister species occurring in sister hosts C and D.

phylogenetic relationships should be combined in the binary matrix used to generate the host cladogram. Cressey et al. (1983) first noted that this protocol, which they called inclusive ORing, produced the “ghost taxa” to which Dowling (2002) referred. Brooks (1990) proposed the Duplication Convention (in a biogeographic context) specifically to eliminate inclusive ORing from the method. BPA not using the host duplication convention has been called “primary BPA,” while that invoking the duplication convention has been called “secondary BPA” (Brooks and McLennan, 2002; Brooks et al., 2001; Green et al., 2002). Van Veller and Brooks (2001) suggested that reconciled tree analysis is more closely analogous to secondary BPA than to pri-
mary BPA (e.g., reconciled tree analysis permits clade duplications, whereas secondary BPA permits host duplications), and thus empirical comparisons between the methods in studies of historical biogeography should use secondary BPA. We suggest that the same holds true for comparisons between the methods in studies of coevolution.

Note also that we have used the “Three’s Rule” rule in Figs. 7 and 8 to determine which version of host A belongs to the general (cospeciation) pattern and which does not. BPA assesses cospeciation patterns among different clades of parasites. Only when different parasite clades exhibit cospeciation patterns due to a common history of speciation events also involving their hosts will those patterns correspond to host phylogeny.

Performing BPA

BPA is not a model, so there are no a priori costs, probabilities, or weights associated with different coevolutionary phenomena. Rather, BPA is a discovery-based approach, in which different coevolutionary phenomena are inferred from the analysis a posteriori, in accordance with the most parsimonious description of the data. This does not mean that BPA is nonalgorithmic, however, because BPA is a direct application of the principles of phylogenetic systematics. This can be done in two ways, which we outline below.

Wagner algorithm

1. Convert all parasite phylogenetic trees into parasite–host cladograms by replacing parasite identities with host identities on the terminal branches of each parasite phylogenetic tree and then numbering each branch of each parasite–host cladogram for binary coding.

2. Construct a data matrix in which the rows correspond to host taxa (each host taxon is listed once) and the columns correspond to the total number of branches in all parasite–host cladograms being considered.

3. Convert each parasite–host cladogram into additive binary codes and fill in the matrix: (a) for each case in which a host taxon is not inhabited by a member of a given parasite–host cladogram, list host taxon as “missing” (?) for the parasite clade; (b) for each case in which a parasite taxon occurs in more than one host taxon, combine the information from each of those cases (inclusive ORing); and (c) for each case in which more than one parasite taxon occurs in the same host, combine the binary codes representing each species and its phylogenetic relationships in the binary matrix (inclusive ORing).

4. Perform parsimony analysis on the data matrix (this is primary BPA): If the result is a single most parsimonious host cladogram in which there is no homoplasy, the default explanation is that the host cladogram is congruent with the host phylogeny, and cospeciation, nonresponse to host speciation, and sympatric speciation (=lineage duplication) explains the observed host–parasite associations. In other words, all notations for parasites on the host cladogram will correspond to the patterns shown in Figs. 1–3. If the results are otherwise, go to step 5.

5. Account for all host taxa coded as “missing” for particular parasites a posteriori. This means distinguishing cases in which the parasite group was never associated with particular host taxa from those in which the parasite group has been secondarily lost. This can be done by a simple optimization procedure—locate the code for the basal branch of each given parasite clade on the host cladogram (if primary BPA has produced more than one equally parsimonious tree, this part of the analysis can be done using any of the alternatives). This indicates the point at which the parasite clade became associated with the host clade (Fig. 9). Absence of the parasite clade in taxa above that point on the host cladogram is the result of extinction (Figs. 6 and 9). If the result of a primary BPA is a single most parsimonious host cladogram in which there is homoplasy due to inferred reversals (=loss of the parasite clade in a particular host taxon), the default explanation is that the host cladogram is congruent with the host phylogeny, and a single history of cospeciation (and possibly sympatric speciation [=lineage duplication]) plus parasite extinctions explains the observed host–parasite associations. In other words, all notations for parasites on the host cladogram will correspond to the patterns shown in Figs. 1–3 and 6.

6. If the result is one or more most parsimonious host cladograms that include nonreversal homoplasy, the default explanation is that the nonreversal homoplasy indicates episodes of host switching. In such cases, notations for parasites on the host cladogram will correspond to the patterns shown in Figs. 4 and 5.

7. Eliminate the ghost lineages from the host cladogram by duplicating hosts for each instance of nonreversal homoplasy (Figs. 7 and 8).

8. Check the analysis by producing a new binary matrix, this time including duplicated taxa as separate entries. The result should be a single host cladogram (which may include polytomies, because unique speciation by host switching events can produce hard polytomies: Brooks and McLennan, 2002) with no homoplasy except for inferred extinction events.
Hennig argumentation

Hennigian argumentation is an approach in which transformation series are combined sequentially, using the inclusion/exclusion rule (see Brooks and McLennan, 2002; Wiley et al., 1991, 2003). In standard phylogenetic analysis, homoplasy produces violations of the inclusion/exclusion rule. In BPA, each parasite–host cladogram is used as a transformation series, and host switching produces violations of the inclusion/exclusion rule.

1. Convert all parasite phylogenetic trees into parasite–host cladograms by replacing parasite identities with host identities on the terminal branches of each parasite phylogenetic tree, and then numbering each branch of each parasite–host cladogram for binary coding.

2. Begin with one of the parasite–host cladograms (Fig. 10a). Convert it to a binary matrix without inclusive ORing; this may produce duplicated hosts (Fig. 10b). Standard parsimony analysis of this matrix should reproduce the parasite–host cladogram exactly, and with a consistency index of 100%.

3. Take a second parasite–host cladogram (Fig. 11a). Convert to an additive binary matrix for combination with the data from the first parasite–host cladogram (Fig. 10b). There are two coding protocols that need to be observed at this point. First, if the second parasite–host cladogram has a species occurring in a host that has already been duplicated, treat the host for the second parasite clade as a separate occurrence and duplicate it (host A3 in Fig. 11b). [Note: when making a data matrix combining the information for both clades, in this case host A3 should be coded as “missing” for clade 1–11 and hosts A1 and A2 should be coded as “missing” for clade 12–20]. If the duplication is unnecessary, standard parsimony analysis will result in the same hosts appearing either as sister groups or in a polytomy with a third host (see hosts A2 and A3 in Fig. 11c). In these cases, combine the duplicated hosts (see host A2 in Fig. 11d).

4. Calculate a new matrix for Fig. 11d, and use it in combining the third parasite–host cladogram. Continue this process iteratively until all parasite–host cladograms are incorporated. This should result in a single host cladogram with a consistency index of 100%.

5. Assess missing data a posteriori as with the Wagner algorithm to obtain the final host cladogram. If there are absences most parsimoniously interpreted as secondary losses, the final host cladogram will have a consistency index of less than 100%.

Results

Brooks and Glen (1982) presented the first phylogenetic systematic analysis examining the relationship between the Great Apes and their parasites based on the pinworm (nematodes) genus Enterobius. Subsequently, a group of hookworms [Oesophagostomum (Conowieberia); Glen and Brooks (1985); see both Brooks and Glen (1985) and Glen and Brooks (1982) for discussions of host records including the extreme ambiguity of zoon-based data] was added to the database, followed by a new phylogenetic analysis of Enterobius and relatives (Hugot, 1999). Hugot (1999) examined the coevolutionary relationship between the Enterobiinae and their primate hosts. In this analysis we will focus on a subset of Hugot’s study, the relationship between the Enterobius pinworms and the Great Apes. Reanalysis of Hu-
got’s data matrix indicated that there were three equally parsimonious trees for *Enterobius*, differing only in the placement of *E. lerouxi* with respect to the *E. buckleyi* clade and *E. anthropopitheci*. In his study, Hugot used the topology that gave the best fit to the host phylogeny. We have used the consensus tree showing a polytomy among these three taxa. Our interpretation of coevolutionary relationships within this subclade may thus change if the polytomy is resolved by the addition of further characters. On the other hand, the polytomy may represent a “real” evolutionary phenomenon (e.g., peripheral isolates species; see discussion in Brooks and McLennan, 2002), in which case the addition of further data will corroborate the interpretation presented in this study. BPA does not differ from any other phylogenetic systematics-based method; its conclusions are based upon the most parsimonious interpretation of the data at hand and may thus change with the addition of new information.

Here we use both the pinworm and the hookworm phylogenetic trees to demonstrate how BPA is properly performed for coevolutionary studies. The use of only two trees violates the “Three’s Rule” rule, which means that some unique events may have two equally parsimonious explanations (e.g., extinction versus peripheral isolates speciation, see discussion of Fig. 5). Unfortunately there are currently only two resolved parasite clades for the Great Apes, so the results presented in this paper represent the best hypothesis that we can erect based on the data in hand. Bearing this caveat in mind, we compare both the results of primary BPA with those of secondary BPA and the results of the secondary BPA with those of the Tree map-based results discussed by Hugot (1999), for the relevant species.

1. Because we are interested only in the relationships between the worms and their Great Ape hosts, we have simplified the phylogenetic trees for the parasites by collapsing each monophyletic group that inhabits other hosts into a single lineage. This does not change the phylogenetic tree for the parasites and thus will not affect our interpretations about parasite and Great Ape coevolution. The phylogenetic tree for the pinworms (Fig. 12) depicts three such lineages: the *E. brevicauda* + *E. bipapillatus* + *E. macaci* clade (6) whose members inhabit cercopithecines, the *Colobenterobius* clade (7) whose members inhabit colobines, and the *Trypanoxyuris* clade (8) whose members plesiomorphically inhabit New World monkeys. The phylogenetic tree for the hookworms (Fig. 13) depicts one such lineage: *O. xeri* + *O. susannae* clade (20) whose members inhabit rodents.
2. The current estimate of anthropoid phylogeny (Goodman et al., 1998) is shown in Fig. 14. Primary BPA (Table 1) produces two equally parsimonious host cladograms (the consensus is shown in Fig. 15), which depict substantial, but not complete, congruence with the phylogenetic relationships among the Great Apes and between the Great Apes and their sister groups (Old and New World monkeys; compare the host cladogram with the host phylogeny shown in Fig. 14).

3. Secondary BPA (Table 2) produces one most parsimonious host cladogram with a CI of 97%, but that host cladogram places *O. bifurcum* (species 23) as the ancestor to species 31, rather than as its sister, violating Assumption 0. This misplacement occurs because of the widespread distribution of *O. bifurcum* (it is the only species to inhabit *Homo*, *Pan*, and the cercopithecines). Fig. 16 depicts the host cladogram derived by maintaining the phylogenetic relationships of *O. bifurcum* in accordance with the requirements of Assumption 0. It differs from the computer (i.e., Wagner algorithm)-generated host cladogram only in the placement of *Homo* and *Pan* (PAUP places these two taxa into a polytomy at the base of CO + CE) and in having a CI of 100%, which indicates the truly most parsimonious representation of the data.

**Discussion**

Hugot used Using TreeMap to fit the *Enterobius* phylogeny to the phylogeny of the Great Apes (missing *Hylobates*) + sister group (Old World monkeys).
Fig. 14. Phylogenetic tree for the Great Apes (Hominidae) and their relatives. The Old World monkey clade includes the Cercopithecinae (e.g., baboons, macaques) and the Colobinae (e.g., colobines, langurs). The New World monkeys include such famous representatives as the howler monkeys and tamarins.

Table 1
Primary matrix listing the distribution of pinworm and hookworm clades among their primate hosts, with the binary codes representing the phylogenetic relationships for both clades

<table>
<thead>
<tr>
<th>Host</th>
<th>Species</th>
<th>Binary code</th>
</tr>
</thead>
<tbody>
<tr>
<td>New World monkeys</td>
<td>8, 24</td>
<td>0000000100 0001000000 0001000000 001</td>
</tr>
<tr>
<td>Cercopithecines</td>
<td>6, 16, 21, 23</td>
<td>0000010000 1111010000 1010111111 111</td>
</tr>
<tr>
<td>Colobines</td>
<td>7, 21, 22, 23</td>
<td>0000001000 0011000000 1110000001 111</td>
</tr>
<tr>
<td>Hylobates</td>
<td>17, 18</td>
<td>?????????? ???001100 0000011111 111</td>
</tr>
<tr>
<td>Pongo</td>
<td>5, 19</td>
<td>0000100000 1111000010 0000000111 111</td>
</tr>
<tr>
<td>Gorilla</td>
<td>4, 15</td>
<td>0001000000 0111100000 0000111111 111</td>
</tr>
<tr>
<td>Pan</td>
<td>3, 15, 23</td>
<td>0010000001 0111100000 0010111111 111</td>
</tr>
<tr>
<td>Homo</td>
<td>1, 2, 15, 23</td>
<td>1100000011 0111100000 0010111111 111</td>
</tr>
<tr>
<td>Rodents</td>
<td>20</td>
<td>?????????? ???000001 0000000011 111</td>
</tr>
</tbody>
</table>

Species 1–8, Enterobius; species 15–24, Oesophagostomum; ?, species missing from the host group. For explanations of parasite species numbers see Figs. 12 and 13.

Fig. 15. Primary BPA of the parasite data. Bold lines, historical backbone of association between the primates and their pinworms and hookworms; dashed lines, host misplaced on the tree, indicating interactions other than cospeciation.
Map found 12 different reconstructions in which the number of posited extra evolutionary events (sorting, duplication, host switching) ranged from 11 to 6, indicating that the relationship between the Great Apes and their pinworms is an extremely complex one. Hugot preferred the hypothesis invoking 6 events because it was the scenario requiring the fewest events that was consistent with the assumption that the pinworms of Gorilla + Pan + Homo were monophyletic. This scenario posited one sorting event at the base of the Great Ape tree, two cospeciation events (the Gorilla + [Pan, Homo] and Pan + Homo splits), one host switch from Gorilla to Pongo, one host switch from Gorilla to the Old World monkeys, and one speciation event within the same host (Homo).

Secondary BPA analysis, on the other hand, posits that hookworms and pinworms were already jointly associated in the common ancestor of the anthropoid primates (ancestors 14 and 33 on Fig. 16). There is a strong historical backbone of cospeciation among these organisms (bold lines in Fig. 16). In terms of Enterobius, BPA corroborates Hugot’s proposal of a cospeciation event occurring with (1) the Pan + Homo bifurcation (ancestor 10 giving rise to species 3 and ancestor 9 in Fig. 16) and (2) the Gorilla + sister-group split (ancestor 12 giving rise to species 4 and ancestor 10 in Fig. 16). It

Table 2
Secondary BPA matrix listing the distribution of pinworm and hookworm clades among their primate hosts, with binary codes representing the phylogenetic relationships for both clades

<table>
<thead>
<tr>
<th>Host</th>
<th>Species</th>
<th>Binary code</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>New World monkeys</td>
<td>8, 24</td>
<td>0000000100 0001000000 0010000000 111</td>
<td></td>
</tr>
<tr>
<td>Cercopithecines1</td>
<td>21, 23</td>
<td>??????????? 0010001000 0010000000 0000111111 111</td>
<td></td>
</tr>
<tr>
<td>Cercopithecines2</td>
<td>6</td>
<td>0000100000 1111?????? ??????????? ??</td>
<td></td>
</tr>
<tr>
<td>Cercopithecines3</td>
<td>16</td>
<td>??????????? 0000100000 0010000000 0000111111 111</td>
<td></td>
</tr>
<tr>
<td>Colobines</td>
<td>7, 21, 22, 23</td>
<td>0000010000 0010000000 1110000001 0000001111 111</td>
<td></td>
</tr>
<tr>
<td>Hylobates1</td>
<td>18</td>
<td>??????????? 0000100000 0010000000 0000001111 111</td>
<td></td>
</tr>
<tr>
<td>Hylobates2</td>
<td>17</td>
<td>??????????? 0000100000 0010000000 0000001111 111</td>
<td></td>
</tr>
<tr>
<td>Pongo</td>
<td>5, 19</td>
<td>0000100000 0111000010 0000000011 111</td>
<td></td>
</tr>
<tr>
<td>Gorilla</td>
<td>4, 15</td>
<td>0000100000 0111000000 0000111111 111</td>
<td></td>
</tr>
<tr>
<td>Pan1</td>
<td>3, 15</td>
<td>0000100000 0111000000 0000111111 111</td>
<td></td>
</tr>
<tr>
<td>Pan2</td>
<td>23</td>
<td>??????????? 0000100000 0010000000 0000001111 111</td>
<td></td>
</tr>
<tr>
<td>Homo1</td>
<td>1, 2, 15</td>
<td>1100000011 0111000000 0000111111 111</td>
<td></td>
</tr>
<tr>
<td>Homo2</td>
<td>23</td>
<td>??????????? 0000100000 0010000000 0000001111 111</td>
<td></td>
</tr>
<tr>
<td>Rodents</td>
<td>20</td>
<td>??????????? 00000001 0010000000 0010000000 0000001111 111</td>
<td></td>
</tr>
</tbody>
</table>

The Cercopithecines are listed as three separate taxa, while Hylobates, Pan, and Homo are each listed as two separate taxa. Species 1–8; Enterobius; species 15–24, Oesophagostomum; ?, species missing from the host group. For explanations of parasite species numbers see Figs. 12 and 13.
is relevant to note at this point that BPA resolves the trichotomy of the E. lerouxi in favor of the co-speciation pattern adopted by Hugot, but does so without an a priori assumption of maximum co-speciation. BPA, however, places a third co-speciation event at the Pongo + sister group split. The presence of species 5 in Pongo is thus hypothesized to be the result of co-speciation and not of a host switch from Gorilla. There are three apparent cases of lineage duplication in the ancestor of the Old World monkeys + Great Apes (denoted by the joint occurrence of ancestors 30, 31, and 32). As with all lineage duplications, these may represent ancient cases of sympatric speciation. No lineage duplications are postulated for Enterobius. The secondary BPA suggests that there have been six episodes of host switching. The most obvious of these is the occurrence of the hookworms O. xeri + O. susanae (20) in South African rodents, which represents an episode of speciation by host switching from apes to rodents. This type of exchange is not unique; Hugot (1999) documented instances of other pinworms (Trypanoxyuris) speciating after they had transferred from New World monkeys to squirrels. The occurrence of O. pachycephalum (species 16) in cercopithecines and of O. ovatum (species 17) and O. raiilieti (species 18) in Hylobates also appear to represent episodes of peripheral isolates speciation by host switching from one or more species ancestral to the African Great Apes. The only pinworm to host switch was species 6 (recall that “6” actually refers to the ancestor of a monophyletic group of three species in cercopithecines; see Fig. 12). Ancestor 6 appears to have colonized the cercopithecines from the Pongo lineage and not from Gorilla as suggested by Hugot, where it subsequently diversified in the new host group.

The occurrence of O. bifurcum (species 23) in Pan and Homo represents postspeciation dispersal by the parasite from Old World monkeys (dotted lines + airplanes in Fig. 16). Because Homo and Pan are sister groups, it is actually most parsimonious to postulate that O. bifurcum colonized the common ancestor of Pan + Homo, in which case a single host switch explains the observed increase in host range. Once there, the hookworm did not speciate again, even though its new host did, producing the Pan and Homo lineages. The presence of another hookworm, O. stephanostomum (species 15) in Gorilla, Pan, and Homo also represents a case of the parasite failing to speciate with its host. In other words, the expanded host ranges of O. stephanostomum and O. bifurcum were gained not by rampant host switching but rather by the parasite species not speciating when its host did. O. bifurcum, O. brumpti (species 22), and O. aculeatum (species 21) inhabit various Old World monkeys. At this level of analysis we cannot tell whether this large host range was achieved through association by descent (hosts continued to speciate while the parasites did not) or widespread colonization. To investigate this problem further, we need a detailed phylogeny for the Old World monkeys and intensive inventories of the parasites of all primate species.

Because BPA treats hosts and parasites as directly analogous to geographic areas and the species living in them, it is relatively easy to add historical biogeographical information to coevolutionary findings. Geographic distribution data for these two parasite groups suggest complex, and as yet incompletely resolved, movements of primates throughout the tropical regions of the Old World. O. bifurcum and O. brumpti occur in both Africa and Asia, while O. aculeatum is endemic to Asia, possibly suggesting dispersal from Africa to Asia (although O. xeri + O. susanae are endemic to South Africa). O. blanchardi (species 19), O. raiilieti (species 17), and O. ovatum (species 18) are Asian endemics, while O. pachycephalum (species 16) and O. stephanostomum (species 15) are African endemics, suggesting dispersal from Asia to Africa by the common ancestor(s) of, not surprisingly, the African Great Apes (Gorilla, Pan, and Homo). As early hominids moved out of the African forest into the African savannah, our direct ancestors may have added O. bifurcum (remember, O. bifurcum may have been acquired by the ancestor of humans and chimpanzees) to their parasite repertoire through host switches from Old World monkeys. The parasites then indicate that Homo moved to Asia, leaving O. stephanostomum behind in Africa (humans in Asia do not have this hookworm). The occurrence of O. bifurcum in humans in Africa and Asia is not surprising; the host cladogram suggests that O. bifurcum occurred in Old World monkeys on both continents long before Homo arrived in Asia. This period may also have seen the differentiation of the pinworms E. vermicularis (species 2) and E. gregorii (species 1), sister species in the same host produced in the absence of host speciation.

This apparently is not an isolated case in the evolution of human parasites. Hoberg et al. (2000) recently discovered evidence that two cases of speciation by host switching, correlated with the shift from scavenging to predation (hunting) more than a million years ago in Africa, explained the association between humans and tapeworms in the genus Taenia that infect humans (including the beef tapeworm, T. saginata, and the pork tapeworm, T. solium). Furthermore, they found that T. saginata and T. asiatica, sister species occurring in Africa and Asia, respectively, diverged from each other between 750,000 and 1,700,000 years ago, perhaps at the same time that E. vermicularis and E. gregorii were differentiating (but in different parts of the host).

Conclusions

TreeMap, using the Enterobius + Great Ape data, identified one sorting event at the base of the Great Ape
tree, two cospeciation events (the Gorilla + [Pan, Homo] and Pan + Homo splits), one host switch from Gorilla to Pongo, one host switch from Gorilla to the Old World monkeys and one speciation event within the same host (Homo) (six events). Secondary BPA posited three cospeciation events, one host switch, and one speciation event within the same host (five events), providing a more parsimonious explanation of the data without having to invoke any a priori assumptions about the monophyly of the parasites inhabiting the hosts or the likelihood of cospeciation occurring. It is also noteworthy that, in this case, BPA postulates more cospeciation events than does TreeMap. Our study thus corroborates Dowling’s (2002) fundamental conclusions about the merits of primary BPA relative to TreeMap; if the only objective is to provide a parasite-based estimate of host phylogeny, primary BPA should be the preferred method of analysis. This parallels findings by Van Veller and Brooks (2001) that primary BPA is better than TreeMap at finding the general biogeographic pattern when there have been substantial amounts of dispersal. However, Dowling’s study, like most previous studies purportedly comparing BPA and other methods in a biogeographic context (Morrone and Carpenter, 1994; Morrone and Crisci, 1995; Page, 1990; Page and Charleston, 1998; Ronquist, 1995, 1996, 1997a,b, 1998), failed to consider secondary BPA. As a result, his conclusions about the shortcomings of BPA due to inclusive ORing (i.e., the production of ghost taxa) are incorrect, as secondary BPA eliminates inclusive ORing from the method (Van Veller and Brooks, 2001). In addition, secondary BPA provides a more complete account of the evolutionary associations between the parasite groups and their hosts than does either TreeMap or primary BPA, without sacrificing any indirectly generated information about host phylogeny. Our analysis of the two groups of nematodes inhabiting Great Apes uncovered a substantial backbone of vicariant cospeciation (as reported by Hugot for Enterobius) and some lineage duplication but also found five episodes of peripheral isolates speciation by host switching, one episode of postspeciation dispersal adding hosts, one episode in which the parasite species did not speciate when the host species did, and one episode in which the parasite speciated when the host species did not. This result supports Brooks’ (1979, 1981) initial recognition that parasites are not simply dependent variables on host phylogeny. Rather, they are independent variables on different host resources, each of which may be distributed in a variety of ways among host species of varying degrees of phylogenetic relatedness (Brooks and McLennan, 1993, 2002).

Acknowledgments

We thank Eric Hoberg, Marco Van Veller, Greg Klassen, and Ashley Dowling for advice and useful discussions about the issues addressed in this contribution. This research was supported by research funding from the Natural Sciences and Engineering Research Council (NSERC) of Canada to D.R.B. and to D.A.M.

References


Further reading