

VIMS Articles

---

1996

## Phylogenetic covariance probability: Confidence and historical associations

ME Siddall  
*Virginia Institute of Marine Science*

Follow this and additional works at: <https://scholarworks.wm.edu/vimsarticles>



Part of the [Other Ecology and Evolutionary Biology Commons](#)

---

### Recommended Citation

Siddall, ME, "Phylogenetic covariance probability: Confidence and historical associations" (1996). *VIMS Articles*. 1495.

<https://scholarworks.wm.edu/vimsarticles/1495>

This Article is brought to you for free and open access by W&M ScholarWorks. It has been accepted for inclusion in VIMS Articles by an authorized administrator of W&M ScholarWorks. For more information, please contact [scholarworks@wm.edu](mailto:scholarworks@wm.edu).

## PHYLOGENETIC COVARIANCE PROBABILITY: CONFIDENCE AND HISTORICAL ASSOCIATIONS

MARK E. SIDDALL

*Virginia Institute of Marine Science, School of Marine Science, The College of William and Mary, Gloucester Point, Virginia 23062, USA; E-mail: mes@vims.edu*

*Abstract.*—The correlation that exists among multiple cladograms is often taken as evidence of some underlying macroevolutionary phenomenon common to the histories of those clades and, thus, as an explanation of the patterns of association of the constituent taxa. Such studies have various forms, the most common of which are cladistic biogeography and host–parasite coevolution. The issue of confidence has periodically been a theoretical consideration of vicariance biogeographers but in practice has been largely ignored by others. Previous approaches to assessing confidence in historical associations are examined here in relation to the difference between simple-event and cumulative probabilities and in relation to the restrictiveness of joint hypothesis testing. The phylogenetic covariance probability (PCP) test, a novel approach to assessing confidence in hypotheses of historical association, employs the empirical protocol of Brooks parsimony analysis (BPA) in an iterative, computer-intensive randomization routine. The PCP value consists of the frequency with which a solution as efficient or more efficient than the observed hypothesis of correlated phylogeny is achieved with random associations (e.g., of parasites and hosts or of taxa and areas). Because only the associations, and not the contributing phylogenies, are subjected to randomization, the test is not prone to certain criticisms leveled at other cladistic randomization routines. The behavior of the PCP test is examined in relation to eight published studies of historical association. This test is appropriately sensitive to the degrees of freedom allowed by the number of contributing clades and the number of taxa in those clades, to the extent of noncorrelated associations in the observed hypothesis, and to the relative information content contributing to that hypothesis. [Biogeography; BPA; coevolution; confidence; historical associations; randomization.]

Although biogeography and host–parasite association studies are relatively common, rarely is there an attempt to quantify the strength of the support for historical association. Borrowing from correlation analysis, an  $r$  value provides an estimate of the goodness of fit of a function. Confidence in the relationship is expressed in terms of a  $P$  value, the probability that an equally efficient result could be achieved by chance. In questions of historical association, the notion that congruence between independently derived cladograms is fundamentally a probabilistic issue has been widely acknowledged (Metcalf, 1929: 4; Rosen, 1978:160; Brooks, 1979:303, 1981: 229; Nelson, 1979:8; Nelson and Platnick, 1981:312; Mitter and Brooks, 1983:85), as is the notion that discovering the extent of historical association is fundamentally an exercise in determining the degree of correlation between the associates' phylogenetic histories (Kellogg, 1913:158; Brooks, 1979:300, 1981:235, 1988:249; Nelson and

Platnick, 1981:10; Mitter and Brooks, 1983: 96)

Since Platnick and Nelson (1978) asserted that cladistic biogeography is “inherently statistical,” empirically meaningful measures of the relative nonrandomness of a particular hypothesis of historical association have not been ignored, although their application has been infrequent. The approaches used by Rosen (1978), Platnick and Nelson (1978), and Nelson (1979) are identical in the manner in which probability is determined, although they differ in their perspective on the data. Their approaches are addressed here only briefly; they have been soundly criticized elsewhere (Pielou, 1981; Simberloff et al., 1981; Cracraft, 1988; Simberloff, 1987). Simberloff (1987) and Page (1988) both employed a Monte Carlo approach to determining levels of significance. The new method proposed here for finding estimates of confidence on historical associations employs the empirical protocol detailed by Brooks

(1981, 1990; see also Wiley, 1988) in a computationally intensive approach that randomizes associations rather than trees. Eight examples of its application to data sets extracted from published literature are examined to consider a variety of data sets and biogeographic problems.

#### N-TREES

Rosen's (1978) biogeographic method involved superimposing two cladograms (in his example, area cladograms for species of *Heterandria* and *Xiphophorus*) and retaining in a reduced area cladogram those elements that were consistent for both. His estimate of significance for the resulting hypothesis of historical congruence involved calculating the probability that two labeled trees (i.e., n-trees sensu Margush and McMorris, 1981) would have the same topology. Thus, for a five-area reduced area cladogram, there are 105 possible n-trees, and the probability that one five-taxon cladogram will replicate one particular congruent five-taxon portion of another is  $1/105$  or  $P = 0.009$ . The only difference in this process introduced by Nelson (1979) was calculation of these probabilities in terms of replicated components of a particular term size in relation to the number of possible components of equivalent size for all replicated components of each size and accumulation of a final overall probability by multiplying all of the component partial probabilities. Apart from problems introduced by reduced area cladograms, widespread taxa, missing taxa, and redundant distributions, these methods were criticized for their assumption of a null distribution defined as all n-trees are equiprobable, the lack of knowledge regarding the probability that cladograms for which no associations exist would match, and the lack of consideration of the probability in relation to imperfectly resolved topologies (see Simberloff, 1987, for a more complete review). A serious criticism that can be leveled at the component-replication method (Nelson, 1979) is that the partial probability of one component is not independent either of other components or of their partial probabilities (Rohlf, 1982). This un-

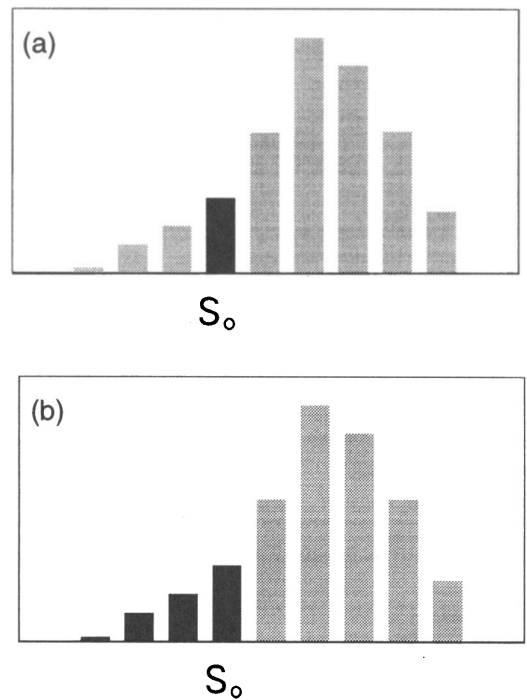


FIGURE 1. Hypothetical null distribution for observed tree length ( $S_o$ ). (a) Simple-event probability of occurrence of length  $S_o$  (■). (b) Cumulative probability of occurrence of length  $S_o$  or shorter (■).

realistic compounding error rate is equally problematic for Brown's (1994) method of calculating confidence in biogeography, in addition to the errors and information loss introduced by reliance on strict consensus.

The probability of a completely imperfect match (i.e., the minimum possible degree of historical association) is only represented by a few of all possible trees and, thus, is also unlikely. There is a difference between the simple-event probability of an observation's occurrence in a null distribution and the cumulative probability of that particular observation's occurrence as well as all other possible occurrences toward one extreme in the null distribution. In Figure 1, the parameter of interest is tree length (as in the permutation tail probability test; Faith and Cranston, 1991). The simple-event probability of a tree of some observed length  $S_o$  occurring is 0.080 (Fig. 1a), apparently significant at a critical level of 10%. Hypothesis testing, however,

relates to cumulative probability distributions, not simple-event probability. In Figure 1b, the cumulative probability of occurrence of a tree of length  $S \leq S_0$  is 0.165, which is not significant at the same critical level. An example of this difference involves the issue of adaptive radiation (see Farris, 1976; Slowinski and Guyer, 1989, 1993; Maddison and Slatkin, 1991; Felsenstein, 1992). With respect to the approaches taken by Rosen (1978), Platnick and Nelson (1978), Nelson (1979), and Brown (1994), although one can determine the simple-event probability of chance occurrence of a particular  $n$ -tree, there is no measured parameter against which one can determine a cumulative probability distribution. As such, these approaches do not lend themselves to hypothesis testing.

#### MARKOVIAN MONTE CARLO

Simberloff et al. (1981) and Simberloff (1987) suggested that a more appropriate null distribution for cladograms than the all  $n$ -trees equiprobable model is a Markovian model (Harding, 1971) that states that for any existing  $n$ -tree with  $n$  taxa, the next speciation event creating  $n + 1$  taxa is equally likely to occur at any of the  $n$  terminal nodes. Thus, the probability of finding a particular labeled  $n$ -tree is dependent on its structure to the point that all trees are not equiprobable (see also Slowinski, 1990). The protocol described by Simberloff et al. (1981) entailed randomly generating cladograms under the Markovian model many times and finding the number of times in those replicates that the observed similarity exactly matched a particular biogeographic hypothesis. This approach suffers from the same restriction as do simple-event probabilities. In Figure 2a, there is one "item of error" between host and associate phylogenies. The actual associate tree in Figure 2a is one particular topology that could result from a Markovian growth process with a particular frequency. What is left out of Simberloff's (1987) approach is the notion that a better match (e.g., Fig. 2b) will occur in the randomized data sets with a finite probability.

Page (1988) introduced a modification of

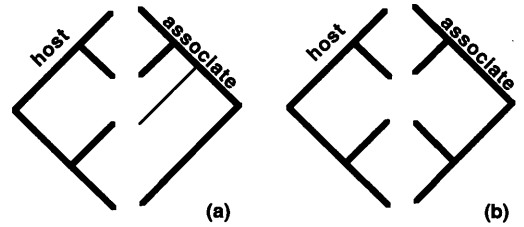


FIGURE 2. Simple-event probability for assessing coevolutionary confidence values. The frequency that randomized trees will match a hypothesis as well as does the observed tree (a) is less than the frequency that randomized trees will match as well or better than does the observed tree (a plus b).

the Markovian approach in relation to the problem of multiple equally parsimonious area cladograms under assumptions 1 and 2 of component analysis (Nelson and Platnick, 1981). A median binary tree (see Penny et al., 1982, for algorithms) is calculated for each suite of area cladograms hypothesized by each taxonomic group, and a minimal tree similarity value is calculated. Next, taxon cladograms of equal size are randomly generated and analyzed according to the same principles as used for the real cladograms. This process is repeated many times (e.g., 100), and the cumulative probability (number of iterations) for which as good or better similarity was achieved with random data is determined. From this method, Page (1988) concluded that assumption 2 performed better (i.e., nonrandom) with respect to Rosen's (1978) *Heterandria/Xiphophorus* data set than did either assumption 0 or assumption 1.

Although the computer-intensive approaches used by Simberloff et al. (1981), Simberloff (1987), and Page (1988) represent significant theoretical advancements in determining levels of confidence in questions of biogeographic congruence and historical associations, in terms of randomization, their methodologies focus on the structure of the trees themselves and not on the associations between them. Two items warrant consideration as a result: (1) the joint hypothesis testing that occurs with Monte Carlo-based approaches and (2) what substantive question is being asked of the data and what randomization

routine is appropriate for asking it. Question 2 ought to be, "For two or more associated clades, are their historical cladogenetic events correlated?" Recognition of this approach as a correlation analysis should govern the choice of randomization method.

Addressing first the problem of joint hypothesis testing, the characteristics of the Monte Carlo method of resampling are such that the structure of the null hypothesis is "the data are a random sample from a particular population." These joint hypotheses are used for testing questions concerning the particular model being used to generate the null distribution with which an observation is being compared. Thus, the underlying assumption in the methods proposed by Simberloff et al. (1981), Simberloff (1987), Page (1988), and others is that the Markovian model of tree growth is an appropriate one. Rejection of the null hypothesis might result from a real phylogenetic phenomenon (i.e., congruence) or from inappropriateness of the Markovian model. Both Simberloff (1987) and Page (1988) argued for the applicability of this null distribution on the grounds that empirical data compiled by Savage (1983) for 594 cladograms indicated that the topological distributions were not significantly different than would be expected by chance under the Markovian model as opposed to the all-trees-equiprobable model. The argument that the Markovian growth model is more realistic may be true on the whole for a collection of many trees, yet there is no reason to assume a priori that it need be true for a particular tree being assessed in a particular biogeographic comparison. For example, if one or more clades have undergone an episode of adaptive radiation, deviations from this stochastic model are predicted (Slowinski and Guyer, 1989; Slowinski, 1990).

Another approach to assessing confidence in these questions of historical association is possible when one recognizes that this question is fundamentally one of correlation. Tests of significance provide information about one of two distinct random influences. One influence is con-

cerned with a characteristic of the population from which a random sample is drawn. Conventional parametric tests such as *t*-tests are of this type, as are Monte Carlo and bootstrap methods. The other random influence concerns the relationship among variables. Significance tests used in correlation analyses (i.e.,  $P[r = 0]$ ) are of this type, as is the approximate randomization method of reassociation or permutation (Noreen, 1989:1–5). The shape of the host and associate trees is only one aspect of the evidentiary data in a test of coevolution. The other aspect is the number and complexity of the associations between the terminals in the associated trees, i.e., how the variables are related. Investigations of historical association, whether biogeographic or host–parasite cospeciation, are fundamentally questions of correlation. Were the cladogenetic events apparent in the associates' history correlated with those that occurred in the hosts'? The appropriate null distribution of a test statistic in the calculation of a level of significance in conventional correlation analysis is the distribution obtained when the values of the two (or more) parameters are associated randomly. In host–parasite associations, if the substantive theory is that the observed distributions of parasites and hosts (or taxa and areas) are explained by correlated cladogenesis (cospeciation, association by descent, or some other non-random influence), then the null distribution against which the observed data must be compared is that predicted by a random association of parasites with hosts (or taxa with areas). The advantages of approximate randomization procedures are that they can be used to test the significance of any correlation test statistic, the data can be drawn from any population, and, perhaps more importantly for data of a phylogenetic nature, the observations need not be a random sample.

The applicability of approximate randomization tests to coevolutionary questions can be readily visualized in comparison with the analogous mathematical correlation test of significance. Figure 3a shows a linear correlation for two points.

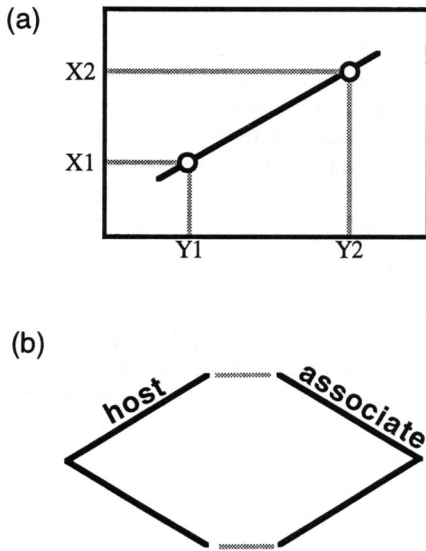


FIGURE 3. Linear (a) and coevolutionary (b) correlation of two points representing two pairs for associated variables. Although there is a perfect fit in both, this high efficiency is trivial because any conceivable association of the observations is equally good (i.e.,  $P = 1.00$ ).

These two points represent two associations between two pairs of observations of two variables, i.e.,  $X_1$  is associated with  $Y_1$ , and  $X_2$  is associated with  $Y_2$ . The efficiency of the line connecting the two points is perfect (i.e.,  $r = 1.00$ ). The comparable correlated phylogeny example is given in Figure 3b. There are two associations between two pairs of observations (taxa) of two variables (trees) and the degree of coevolution is perfect. In both of these examples, the correlations are trivially perfect. Two points in a linear correlation are guaranteed to yield a perfect fit but a perfect lack of confidence (i.e.,  $P = 1.00$ ) because any confidence value is dependent on degrees of freedom that are two less than the number of associations. Similarly, were the associations between the host and associate any different, an equally perfect coevolutionary scenario would result. Tests of significance in coevolution are also dependent on degrees of freedom that are two less than the number of associations. The relevance of this minimal value of 2 becomes apparent in more complex questions.

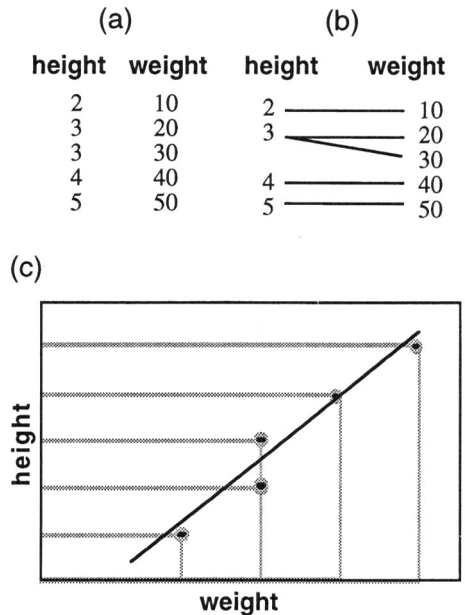


FIGURE 4. Representation of a series of correlated values for height and weight. (a) Tabular form. (b) Association form. (c) Graphical form. The appropriate procedure for determining a confidence value in the best linear fit is accomplished by randomly changing the associations (b) and recalculating a best linear fit to discover the frequency with which a fit as good or better can be achieved.

An approximate randomization test looking at the correlation between, for example, height and weight (Fig. 4) proceeds through the following steps: a correlation coefficient ( $r_o$ ) is calculated for the "observed" associations of height and weight; then the heights and weights are randomly reassociated and a new correlation coefficient ( $r_r$ ) is calculated; the frequency with which  $r_r$  matches or exceeds  $r_o$  is the tail distribution or  $P$  value. The values of the associated variables are not randomly generated (as they would be in a Monte Carlo approach) because random values would have no bearing on the relationship between the observed values. Rather, the likelihood of achieving a particular association in a randomized data set, and thus the overall null distribution, is governed by the number and structure of associations in the original sample. For example, in Figure 4, in any randomized data set the

probability that a height of 2 will be associated with any particular weight is 0.20, whereas the probability that a height of 3 will be associated with any particular weight is 0.40 because there are two associations involving height = 3 in the original sample. The correlation that exists between historical associates is treated similarly. Consider two cladograms of associated taxa in which there are six taxa each and six associations (Fig. 5). Maximal congruence, or maximal correlation among the phylogenies, is shown in Figure 5a; all 10 lineages (black internodes) and five cladogenetic events (black nodes) in one tree are correlated with those in the other tree, i.e., there is a perfect fit and all six associations (black lines between trees) are evidentiary of coevolution. In Figure 5b, the associations have been arranged in such a way as to give maximally uncorrelated phylogeny, yet it is impossible to have a complete absence of implied coevolution. Two lineages (black internodes) and one correlated event (black nodes) are correlated by virtue of two associations (black lines between trees). Figure 5c is exactly the same as Figure 5b with respect to the shape of the trees and the associations between them, but the implied coevolutionary event (black nodes) is different because a different pair of associations has been arbitrarily selected. Thus, not only must there always be at least two associations (i.e., two pairs of taxa) that suggest at least one correlated evolutionary event, but as the amount of correlated phylogeny diminishes, it becomes more and more arbitrary which nodes in the associated trees are correlated. This lack of stability rests solely on the apparent associations, irrespective of the shape of the trees. It is not possible, therefore, to state that one has confidence in a particular portion of the associated topologies as indicative of coevolution and not in some other portion. Rather, the question and any resulting statement must be expressed in terms of the entirety of the hypothesis. Given a group of associated taxa, their phylogenetic histories, and the observed associations, is there evidence to suggest that there has been correlated phy-

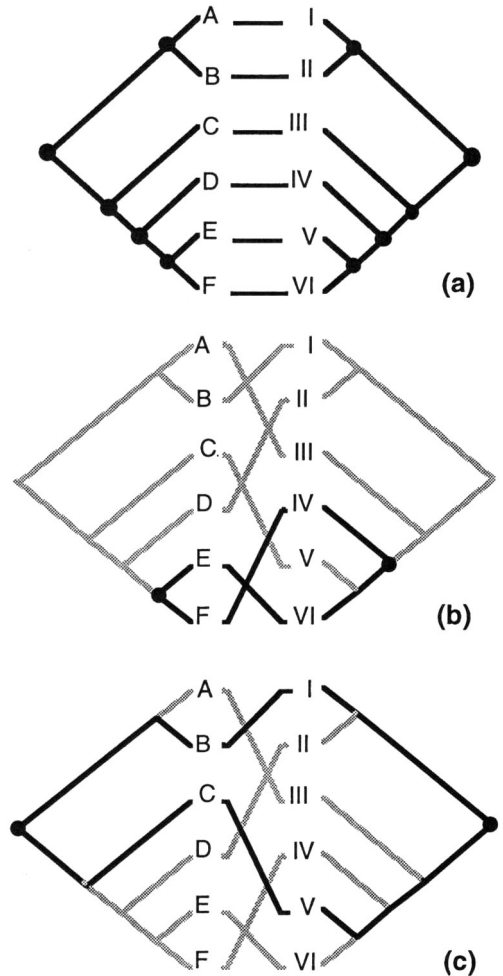


FIGURE 5. Juxtaposition of associated topologies and how the taxa are associated. Correlated cladogenetic events are represented by black circles at nodes, correlated lineages by black internodes, and associations evidentiary of coevolution by black lines between the two trees. (a) Perfect match between the two phylogenies. (b) Perfect mismatch between the two phylogenies, but at least one event is correlated. (c) Associations and trees are the same as in b, but a different pair of associations is arbitrarily chosen as evidence of coevolution.

logeny of the taxa? If so, how much evidence?

#### PHYLOGENETIC COVARIANCE PROBABILITY

The empirical protocol outlined by Brooks (1981) has undergone some modification (Brooks, 1990) to accommodate

problems associated with deviations from a strict 1:1 host-parasite relationship, including widespread taxa ( $n:1$ ), redundant distributions ( $1:n$ ), and missing taxa ( $1:0$ ) (the 0:1 relationship is impossible because parasites must have hosts and taxa must reside in areas). What constitutes a best estimate in Brooks parsimony analysis (BPA) is either the most-parsimonious optimization of parasite characters on an established host phylogeny or the most-parsimonious resolution of a host phylogeny using data derived from multiple parasite clades. Brooks (1988:243) argued that "standard goodness of fit measures, like the consistency index, F-ratio or D-measure, may be used to give a quantitative estimate of the relative proportion of association by descent and association by colonization for any host-parasite assemblage." Whatever the measure or the corrections made for biases introduced by the size of the data set (Archie, 1989; Klassen et al., 1991; Meier et al., 1991), all measures of goodness of fit are dependent on the observed number of steps required by the hypothesis. In BPA, as the number of steps increases (i.e., character distribution becomes less parsimonious) there is less implied coevolution as a result of the extra steps imposed by dispersal or local extinction. Thus, in estimating the relative randomness of a hypothesis, one need only calculate the number of steps as a test statistic. In the phylogenetic correlation probability (PCP) test, the null hypothesis is "the tree length for the observed associations is not significantly shorter than is expected by random association."

For an approximate randomization test to be valid, the observed associations must be randomly reassociated in the context of the legitimate sample space reflected in the original data. That is, if redundant distributions, widespread taxa, and missing taxa are allowable outcomes in the real data (which we know them to be), then they must be possible in the randomized data sets, whether or not they were realized in the original data. The population consists of all possible permutations of that number of associations, of which the ob-

served associations represent one possible outcome.

A stepwise protocol is as follows.

1. Following the methods of BPA (Brooks, 1990), using additive binary codes for parasite relationships and inclusive oring, construct the most-parsimonious solution for cospeciation.
2. Calculate the number of steps for the observed correlation ( $S_o$ ).
3. Randomly reassociate parasites with hosts and repeat steps 1 and 2 to find the number of steps for a randomized data set ( $S_r$ ).
4. If  $S_r \leq S_o$ , add 1 to a running tally,  $\gamma$ .
5. Repeat steps 3 and 4  $R$  times.
6. Calculate  $PCP = (\gamma + 1)/(R + 1)$ ; 1 is added to both  $\gamma$  and  $R$  (the number of iterations) because the observed hypothesis represents one of the possible outcomes.

In practice, these calculations are facilitated by software (DANRAN.EXE) that I have coded and compiled for an MS-DOS operating system environment (DANRAN.EXE is a subroutine of Random Cladistics 3.0, available at no cost from the author or via Internet file transfer protocol to zoo.toronto.edu and stored as a self-extracting compressed utility in pub/random.exe). Step 3 is accomplished by accepting an input file of a specified structure (see Appendix) containing two matrices, one representing the additive binary codes for parasites (or biogeographic associates) and the other representing the association of parasites and hosts (or taxa and areas). Elements of each column representing a parasite in the second matrix are randomly reallocated to taxa according to appropriate approximate randomization for any test of correlation. That is, if an associate is distributed across  $n$  hosts (or areas) in the original data, it will also be distributed across  $n$  hosts (or areas) in the randomized data sets, i.e., the number of associations is not altered. For each randomized set of associations, inclusive oring is applied to redundant distributions, and an additive binary code is applied to each of the hosts (or areas). Output is di-



rected to a file that contains all of the randomized data sets in a series, separated by a set of Hennig86 commands. If the associations are being compared with a defined host cladogram (type II BPA), these commands consist of a "tread" statement followed by a user-defined tree in parenthetical notation. If a host (or area) cladogram is unknown and the correlation under consideration involves one or more parasite clades (type I BPA), the commands consist of a user-defined tree calculating algorithm (e.g., "mhennig;bb;", "ie;", etc). In either case, these commands are automatically followed by "tchoose; xsteps;" to return one tree length value for each randomized data set in the series. The entire file of R randomized data sets plus the original "observed" BPA matrix is automatically run in batch mode through Hennig86 (which is not provided with Random Cladistics) as a daughter process. A logfile is automatically analyzed for the frequency distribution of tree lengths achieved by each randomized data set and the original "true" data set, and the PCP value is calculated.

Below are examples of the application of the PCP test to questions of historical association. In the first, Hafner and Nadler's (1988, 1990) analysis of chewing lice and pocket gopher associations, the protocol and data matrices are detailed. The next four examples are representative of the differences in the various applications of BPA where an independent host (or area) cladogram may or may not be available (types II and I BPA, respectively), where more than one associate clade may or may not be available, where host-parasite or area-associate relationships may be under consideration, and where associates may be missing from hosts (or areas).

#### *Pocket Gophers and Their Lice*

Interest in the coevolution of lice and their hosts originated with Kellogg's (1896a, 1896b) consideration of mallophagan ectoparasites on birds, wherein he considered extant distributions to be indicative of a "common history of genealogy" (Kellogg, 1913), although he concluded

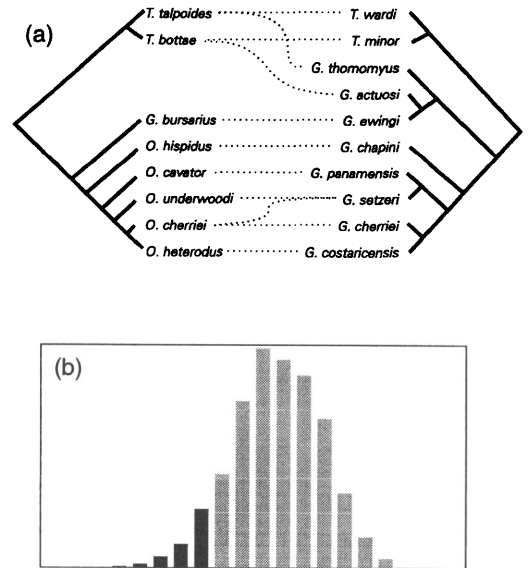


FIGURE 6. (a) Host and parasite cladograms with the respective host-parasite associations indicated (.....) for pocket gophers (left) and lice (right) (redrawn from Hafner and Nadler, 1990). (b) Frequency histogram (■ = tail distribution) of steps for randomized associations and optimization onto host phylogeny (PCP = 0.080).

that parasite divergence proceeded at a slower pace than host divergence (Kellogg, 1896a). Hafner and Nadler (1988, 1990) examined the relationships of pocket gophers (Geomidae) and the chewing lice (Trichodectidae) found associated with them. The high level of congruence observed for host and parasite phylogenies was considered by these authors a prerequisite for their subsequent determination of whether or not cladogenesis for lice and gophers was contemporaneous or proceeded at disparate rates, an issue they investigated using genetic distance measures derived from electrophoretic data. Although the significance of support for correlated evolution was considered an important foundation for the exploration of the timing of cospeciation, Hafner and Nadler (1988, 1990) relied on Nelson and Platnick's (1981) component-replication method.

Hafner and Nadler's parasite phylogeny (Fig. 6a, right side) was converted into additive binary codes (Table 1) representative of the terminal branches and internodes

TABLE 1. Additive binary codes representing the chewing louse phylogeny in Figure 6.

Parasite	Codes																			
<i>Thomomydoecus wardi</i>	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	1	
<i>T. minor</i>	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1
<i>Geomydoecus thomomyus</i>	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	1	1	
<i>G. actuosi</i>	0	0	0	0	0	0	1	0	0	0	0	0	1	0	1	0	0	1	1	
<i>G. ewingi</i>	0	0	0	0	0	1	0	0	0	0	0	0	1	0	1	0	0	1	1	
<i>G. chapini</i>	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1	1	
<i>G. panamensis</i>	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	1	1	1	1	
<i>G. setzeri</i>	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	1	1	1	1	
<i>G. cherriei</i>	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	1	1	
<i>G. costaricensis</i>	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	1	1	

relevant to each parasite. The host-parasite associations were then converted into an association matrix (Table 2). The combined information of the matrices in Tables 1 and 2 allow creation of the BPA matrix (Table 3), in which additive binary codes (rows) for parasites (Table 1) are applied to their respective hosts (Table 2) and in which redundant distributions of more than one parasite on a host are accommodated by inclusive or-ing the combined codes for the parasites. Although the matrix in Table 3 is the classic BPA matrix outlined by Brooks (1981, 1990), to be meaningfully analyzed by numerical cladistic software such as Hennig86, the states in each column require polarization by the addition of a root taxon of all 0's (Table 4). The rationale for this requirement is that if a parasite (or ancestor) is plesiomorphically absent (0) from a host, no steps will be added to the hypothesis, yet if a parasite is apomorphically absent, a step will be added with the caveat that an apomorphic com-

plete absence of parasites will not unrealistically cost multiple steps (see Brooks's [1990] use of missing data). The matrix in Table 4 was examined in relation to the independently derived phylogeny for the pocket gophers (Fig. 6a, left side) by optimizing the characters on the host topology (i.e., using "tread (0((1 2)((3(4 5))(6(7 8)))));" in Hennig86), providing a most-parsimonious length of 26 steps. This approach is typical of type II BPA, in which the phylogenetic data for the hosts are independent of data available for the parasites, i.e., host and parasite phylogenies are both "known." Of the 1,000 randomized association matrices (random shuffling of data in columns of Table 2), 47 of the resulting BPA matrices yielded the same number of steps and 32 were shorter (Fig. 6b), resulting in a PCP value of 0.080. Thus, Hafner and Nadler's premise for secondary determination of correlated timing appears to have had a reasonable basis.

TABLE 2. Association matrix of pocket gopher hosts and chewing louse parasites.

Host	Parasites*									
	1	2	3	4	5	6	7	8	9	10
<i>Thomomys talpoides</i>	1	0	1	0	0	0	0	0	0	0
<i>T. bottae</i>	0	1	0	1	0	0	0	0	0	0
<i>Geomys bursarius</i>	0	0	0	0	1	0	0	0	0	0
<i>Orthogeomys hispidus</i>	0	0	0	0	0	1	0	0	0	0
<i>O. cavator</i>	0	0	0	0	0	0	1	0	0	0
<i>O. underwoodi</i>	0	0	0	0	0	0	0	1	0	0
<i>O. cherriei</i>	0	0	0	0	0	0	0	1	1	0
<i>O. heterodus</i>	0	0	0	0	0	0	0	0	0	1

\* 1 = *Thomomydoecus wardi*; 2 = *T. minor*; 3 = *Geomydoecus thomomyus*; 4 = *G. actuosi*; 5 = *G. ewingi*; 6 = *G. chapini*; 7 = *G. panamensis*; 8 = *G. setzeri*; 9 = *G. cherriei*; 10 = *G. costaricensis*.

TABLE 3. BPA matrix for cospeciation analysis of pocket gophers and chewing lice (as listed in Table 1).

Hosts	BPA codes																		
<i>Thomomys talpoides</i>	0	0	0	0	0	0	0	1	0	1	0	0	0	1	1	0	0	1	1
<i>T. bottae</i>	0	0	0	0	0	0	1	0	1	0	0	0	1	1	1	0	0	1	1
<i>Geomys bursarius</i>	0	0	0	0	0	1	0	0	0	0	0	0	1	0	1	0	0	1	1
<i>Orthogeomys hispidus</i>	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1	1
<i>O. caxator</i>	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	1	1	1	1
<i>O. underwoodi</i>	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	1	1	1	1
<i>O. cherriei</i>	0	1	1	0	0	0	0	0	0	0	1	1	0	0	0	1	1	1	1
<i>O. heterodus</i>	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	1	1

Primates and Pinworms

Brooks and Glen (1982) noted that the phylogeny of parasitic nematodes in the genus *Enterobius* (i.e., pinworms) mirrored that of their primate hosts. Although they interpreted this association as support for coevolution, Simberloff (1987:191) criticized their work in that "they do not explicitly discuss just how improbable a match as close as the observed one would be if parasites colonized hosts independently rather than co-specified with them." Addressing the relative probability that the host and parasite cladograms would match by chance, Simberloff (1987) applied the Markovian randomly generated trees approach and concluded that the probability of two five-taxon cladograms matching was 0.244, thus providing relatively little support for the hypothesis of coevolution. Glen and Brooks (1986) added to the complexity of this particular host-parasite system by factoring the phylogenetic relationships of hookworms (*Oesophagostum* species) and their associations with primates into the equation. In this assessment of coevolution, congruence among

independent parasite phylogenies is taken as support for cospeciation with their hosts. Cladograms for the phylogenetic relationships of the two nematode clades are numbered for cospeciation analysis in Figures 7a (hookworms) and 7b (pinworms). BPA yielded two equally parsimonious solutions for host phylogeny (Figs. 7c and 7d), each with 18 steps. The goodness of fit is indicated by a consistency index (CI) of 0.89. Matrices representing the parasites' relationships and their associations were input to DANRAN.EXE, and instead of including a topology for optimization, the "ie;" command was entered to assure that Hennig86 would find the most-parsimonious hypothesis of relationships of host taxa based on parasite characters. This approach is typical of type I BPA, in which there are no available host phylogenetic data and in which host phylogeny is being estimated on the basis of parasite phylogeny and association. The result from 1,000 random associations (Fig. 7e) was a PCP value of 0.333, indicating lack of support for a hypothesis of cospeciation.

TABLE 4. Data file for Hennig86 analysis of BPA matrix (Table 3) for pocket gophers and chewing lice.

xread																			
19 9																			
ROOT	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ttal	0	0	0	0	0	0	0	1	0	1	0	0	0	1	1	0	0	1	1
Tbot	0	0	0	0	0	0	1	0	1	0	0	0	1	1	1	0	0	1	1
Gbur	0	0	0	0	0	1	0	0	0	0	0	1	0	1	0	0	0	1	1
Ohis	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	1	1	1
Ocav	0	0	0	1	0	0	0	0	0	0	1	0	0	0	1	1	1	1	1
Ound	0	0	1	0	0	0	0	0	0	0	1	0	0	0	1	1	1	1	1
Oche	0	1	1	0	0	0	0	0	0	1	1	0	0	0	1	1	1	1	1
Ohet	1	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	1	1	1;

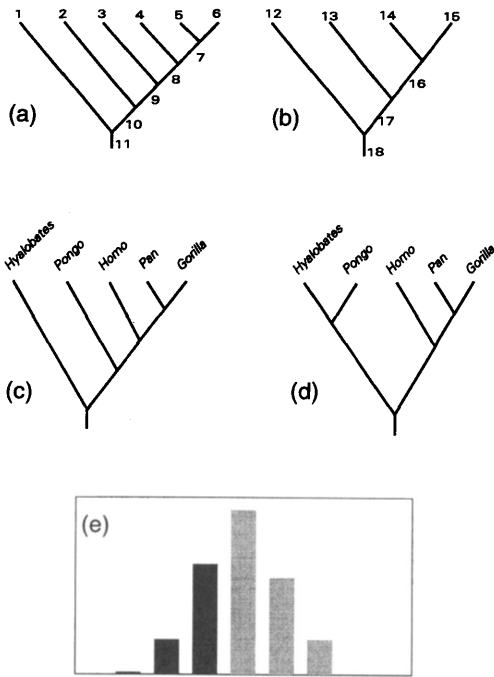


FIGURE 7. Cladograms for hookworm (a) and pinworm (b) parasites of primates (redrawn from Brooks and McLennan, 1991) and the two equally parsimonious solutions (c, d) for the host phylogeny based on parasite relationships. The phylogenies of these two parasite groups were found to not be significantly correlated ( $PCP = 0.333$ ), as indicated in the tail distribution (■) (e).

#### Rosen's Fishes

The historical biogeography of poeciliid fishes of Middle America was the subject of the first attempt to quantify the probabilistic nature of a hypothesis of congruent cladograms (Rosen, 1978). With respect to a five-part taxon-area relationship in the reduced area cladogram, Rosen concluded (in relation to the "probability" that a particular five-member cladogram would result twice) that the coincidence of *Heterandria* and *Xiphophorus* was expected to occur randomly with a frequency of  $P = 0.01$ . Page (1988) reviewed this data set to compare the relative performance of component analysis applied to biogeography (i.e., assumptions 1 and 2). His results indicated that only assumption 2 yielded results that were significantly different from random evolution of the two clades (as op-

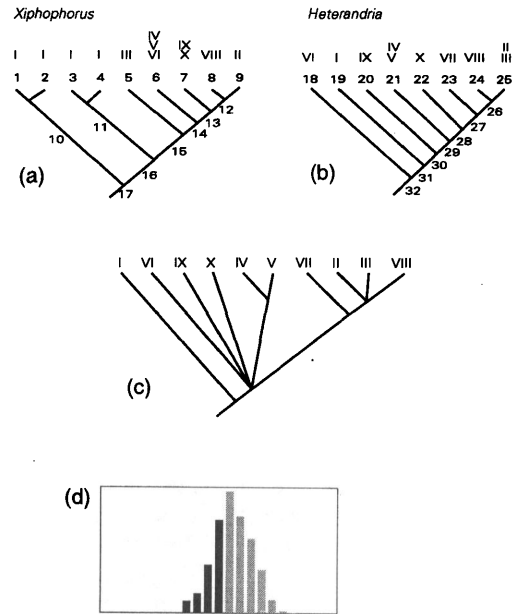


FIGURE 8. (a) Phylogeny of *Xiphophorus* species (arabic numerals) with their associated areas superimposed (roman numerals). (b) Phylogeny of *Heterandria* species (arabic numerals) with their associated areas superimposed (roman numerals). (c) Strict consensus of 10 equally parsimonious solutions for the area relationships for *Xiphophorus* and *Heterandria* resulting from BPA. (d) Frequency histogram of steps found for random associations of taxa and areas in which the probability (PCP; ■) of random association in the observed hypothesis was 0.331.

posed to random association). In Figures 8a and 8b, the cladograms for *Xiphophorus* and *Heterandria*, respectively, are numbered for cospeciation analysis, and the areas in which taxa occur are superimposed in roman numerals above each terminal. BPA resulted in 40 steps for 10 equally parsimonious trees. A strict consensus is given in Figure 4d. The PCP test applied in the same manner as for the primate/pinworm data set (i.e., type I BPA) yielded a value of 0.331 for 1,000 random associations with areas (Fig. 8d), far from an indication that the observed hypothesis differs significantly from what would be obtained by random association of poeciliid fishes to areas and rather different from the level of confidence previously suggested for this example (Rosen, 1978; Page, 1988).

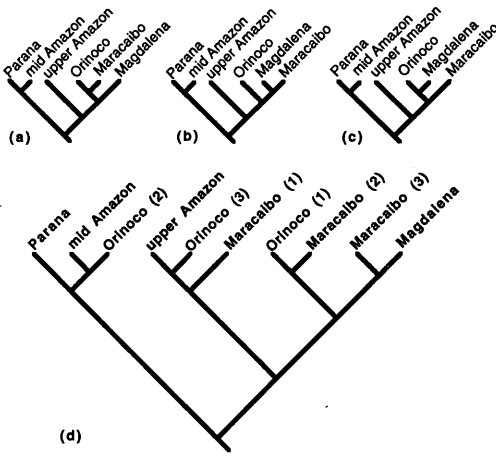


FIGURE 9. Three equally parsimonious solutions (a, b, c) for area relationships of major South American river systems based on the phylogenetic relationships of helminth clades parasitizing freshwater stingrays in those areas (PCP = 0.503). (d) Composite historical hypothesis (redrawn from Brooks and McLennan, 1991); PCP = 0.016.

#### *Parasites of South American Freshwater Stingrays*

In an investigation of the historical biogeography of the major river systems of South America, Brooks et al. (1981) examined these area relationships in the context of the helminth fauna of endemic freshwater stingrays. Their results included a single area cladogram (Fig. 9a) with a CI of 0.80, from which they concluded that the faunal distribution was strongly suggestive of area relationships dominated by vicariant patterns consistent with the Cretaceous to mid-Miocene Andean orogeny. The data matrix used represents a departure from the standard application of type I BPA in that although the first 23 characters are representative of two clades of helminths and their phylogenetic histories, the last 10 characters represent an additional 10 parasites for which there was no phylogenetic information. These 10 characters were coded as presence/absence of parasites and might not, then, add strength to any hypothesis of covarying phylogeny. The effect of these 10 parasites on the hypothesis is seen in differences observed when they are included in or delet-

ed from the BPA data matrix (see Appendix). Removal of these parasites from the analysis had no effect on the resulting biogeographic hypothesis. Both analyses (with or without these 10 parasites) produced three equally parsimonious area cladograms (Fig. 9), reflecting the three possible resolutions of the Orinoco/Maracaibo/Magdalena clade. Inclusion of these 10 parasites raised the CI from 0.79 to 0.80 but had the opposite effect on the PCP value, which was 0.503 for the two parasite clades alone and 0.683 when the additional 10 were included. Neither evaluation suggests a nonrandom association of parasites and areas.

The macroevolutionary history of species assemblages in areas is not always so straightforward. Such assemblages can result from composite histories of clades in the same areas. Brooks (1990) addressed this issue by replicating areas a posteriori and subdividing associations to reflect this. The PCP test can be applied just as easily to these a posteriori interpretations of the data to determine if they are more or less compelling than the original unmodified hypothesis. When the secondary hypothesis of composite area relationships of South American river systems (Fig. 9d) was analyzed (see Appendix, "stingray paras—composite areas"), the results were highly significant (PCP = 0.016).

#### *Liolopids and Vertebrates*

Brooks and McLennan (1991) suggested that the digenean family Liolopidae was typical of what would be expected of numerical relicts (sensu Brooks and Bandoni, 1988). Contributing to this perception was the lack of diversity in liolopids, as compared with that of their sister group (the strigeoids), their overall congruence with their vertebrate hosts' phylogenies, and their obvious lack of associations with a variety of vertebrate clades. This last element has bearing on how the missing taxon problem is addressed by BPA. In its first form (Brooks, 1981), BPA treated situations in which members of an associate clade were completely missing from a host (or area) as all 0's (Table 5). Because the

TABLE 5. Alternative methods of coding liolipid data for missing associates in BPA.

Hosts	Brooks, 1981	Brooks, 1990	New method
Sarcopterygii	0000000000	—————	—————0
Anura	0000000000	—————	—————0
Caudata	1000000001	1000000001	1000000001
Gymnophiona	0000000000	—————	—————0
Mammalia	0100000011	0100000011	0100000011
Chelonia	0010000011	0010000011	0010000011
Aves	0000000000	—————	—————0
Crocodylia	0001000111	0001000111	0001000111
Rhynchocephalia	0000000000	—————	—————0
Ophidia	0000101111	0000101111	0000101111
Sauria	0000011111	0000011111	0000011111

additive binary codes contributed unrealistic additional tree length (i.e., a single exclusionary event could cost multiple steps [see Cracraft, 1988]), Brooks (1990) later modified the protocol to represent this situation as all missing data (Table 5). Considering how BPA is employed in the PCP test, this approach would tend to under-

estimate the number of steps. I propose a slight modification in which the ancestral node shared by all members of an associate clade is coded as 0 if there is no member of that particular clade in that particular host (Table 5). The rationale for this modification (Fig. 10) is that if the absence of members of that clade from that host is explained by the fact that the host clade was colonized subsequent to the divergence of that particular host (i.e., plesiomorphic absence) then no steps will be added (Fig. 10a), whereas only one step will be added (Fig. 10b) if, and only if, the absence of parasites is most parsimoniously explained as an exclusionary event (i.e., apomorphic absence). (DANRAN.EXE prompts the user to override this default option and calculate PCP according to Brooks's [1990] treatment of the missing taxon problem.) Reexamination of the associations between liolipids and vertebrates in which the parasite phylogenetic data were optimized onto the vertebrate phylogeny (i.e., type II BPA) yielded a PCP value of 0.057, indicating significantly non-random associations between hosts and parasites.

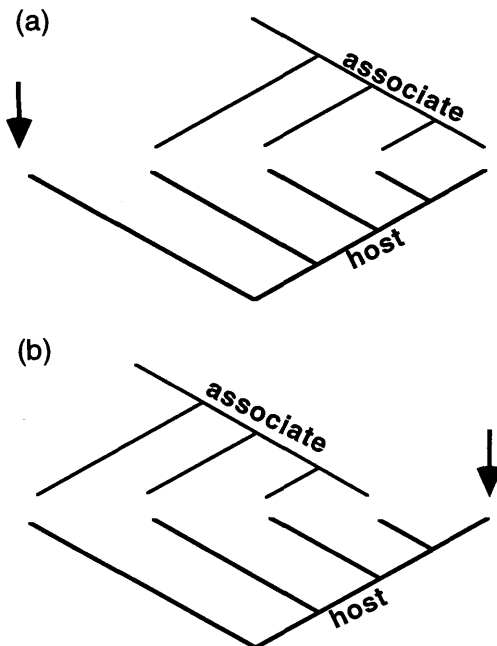


FIGURE 10. Implications of a single plesiomorphic code for wholly missing associates (arrows), as indicated in Table 5. (a) If the associate is plesiomorphically absent, i.e., associations began subsequent to the divergence of the host, no steps will be added. (b) If the associate is apomorphically absent, i.e., by virtue of a sorting event, an extra step will result.

## DISCUSSION

When the preceding five examples are considered in combination with the additional three summarized in Table 6 and detailed in the Appendix, a number of properties of the PCP test are apparent.

1. As the amount of information contributing to the original observed hypoth-

TABLE 6. Summary of eight examples of cospeciation (BPA) analysis and their respective PCP values.

Reference	Hosts/areas	Associates	BPA type	PCP
Bandoni and Brooks, 1987a	fish/turtles	amphilinids	II	0.261
Brooks and McLennan, 1991	vertebrates	liolopids	II	0.057
Brooks et al., 1981	South America	helminths	I	0.016
Glen and Brooks, 1986	primates	nematodes	I	0.333
Hafner and Nadler, 1988	pocket gophers	mallophagans	II	0.080
Hoberg, 1986	alcid seabirds	<i>Alcataenia</i> spp.	II	0.034
Klassen and Beverly-Burton, 1987	ictalurid catfishes	<i>Ligictaluridis</i> spp.	II	0.722
Rosen, 1978	Central America	poeciliids	I	0.331

esis increases, the likelihood of rejecting the null hypothesis of random association seems to increase. For example, in the case of type II BPA in which there is more phylogenetic information (i.e., that of host phylogeny), three of five applications resulted in significant PCP values, whereas when host phylogenetic information is not independently available (type I BPA), only one application yielded a significant PCP value.

- As the amount of implied random dispersal (or host switching) enters into the observed hypothesis, the tendency is, as expected, toward nonsignificant PCP values. For example, Klassen and Beverly-Burton (1987) suggested that there was not evidence of cospeciation between catfish and their monogenean parasites; the PCP value for this association was not significant. Host switching in and of itself is not necessarily going to lead to poor fits in BPA, as has been emphasized with respect to associations of *Alcataenia* species with their avian hosts (Hoberg, 1986; Brooks and McLennan, 1991). In this particular case, the associations are explained in terms of sequential and nonrandom host switching.
- The significant PCP value demonstrates that this test is sensitive to any nonrandom influence on the relative correlation of host and parasite (or area and associate) phylogenies.
- The PCP test is sensitive ultimately to what is possible given the structure of the question. For example, although the associations of gyrocotylids and ratfish

as presented by Bandoni and Brooks (1987b) had a perfect fit (i.e., CI = 1.00), so would any other possible pattern of association (i.e., PCP = 1.000). Where one is invoking type I BPA (estimating host phylogeny from parasites) and there is only one associated clade in the analysis, the question is trivial and has no degrees of freedom.

The nature of the PCP test, being an approximate randomization procedure as opposed to a Monte Carlo method, is that it precludes extending interpretation beyond the observable data. If the null hypothesis that the observed historical associations are not correlated is rejected, there is no expressed implication (although it may be likely) that a more inclusive hypothesis with more putative correlates (e.g., hosts and parasites) will also yield a significant PCP value.

#### ACKNOWLEDGMENTS

In the time this paper was in review, Joe Slowinski communicated to me that he had started developing a similar approach using randomized associations and BPA. My understanding is that Slowinski's method is confined to a 1:1 host:associate pattern in comparing pairs of trees and may be superior for those particular situations. I thank Dan Brooks, Sherwin Desser, Diana Lipscomb, Bob Murphy, and Kevin Doyle for their critical evaluations of the earliest drafts of this manuscript. The pointed comments of Rod Page and an anonymous reviewer contributed significantly to the final version. Brad Anholt and Nick Collins provided much of the encouragement to pursue a computer-intensive approach. Henry Spencer and Liz Seus enabled me to circumvent periodic difficulties encountered in software programming. The Department of Zoology at the University of Toronto has generously allowed the Random Cladistics package to remain on the zoo.toronto.edu ftp site indefinitely.

## REFERENCES

- ARCHIE, J. W. 1989. A randomization test for phylogenetic information in systematic data. *Syst. Zool.* 38:239-252.
- BANDONI, S. M., AND D. R. BROOKS. 1987a. Revision and phylogenetic analysis of the Amphilinidea Poche, 1922 (Platyhelminthes: Cercomeria: Cercomeromorpha). *Can. J. Zool.* 65:1110-1128.
- BANDONI, S. M., AND D. R. BROOKS. 1987b. Revision and phylogenetic analysis of the Gyrocotylidea Poche, 1926 (Platyhelminthes: Cercomeria: Cercomeromorpha). *Can. J. Zool.* 65:2369-2389.
- BROOKS, D. R. 1979. Testing the context and extent of host-parasite coevolution. *Syst. Zool.* 28:299-307.
- BROOKS, D. R. 1981. Hennig's parasitological method: A proposed solution. *Syst. Zool.* 30:229-249.
- BROOKS, D. R. 1988. Macroevolutionary comparisons of host and parasite phylogenies. *Annu. Rev. Ecol. Syst.* 19:235-259.
- BROOKS, D. R. 1990. Parsimony analysis in historical biogeography and coevolution: Methodological and theoretical update. *Syst. Zool.* 39:14-30.
- BROOKS, D. R., AND S. M. BANDONI. 1988. Coevolution and relicts. *Syst. Zool.* 37:19-33.
- BROOKS, D. R., AND D. R. GLEN. 1982. Pinworms and primates: A case study in coevolution. *Proc. Helminthol. Soc. Wash.* 49:76-85.
- BROOKS, D. R., AND D. A. McLENNAN. 1991. Phylogeny, ecology, and behavior: A research program in comparative biology. Univ. Chicago Press, Chicago.
- BROOKS, D. R., T. B. THORSON, AND M. A. MAYES. 1981. Freshwater stingrays (Potamotrygonidae) and their helminth parasites: Testing hypotheses of evolution and coevolution. Pages 147-175 in *Advances in cladistics* (V.A. Funk and D. R. Brooks, eds.). New York Botanical Garden, New York.
- BROWN, J. K. M. 1994. Probabilities of evolutionary trees. *Syst. Biol.* 43:78-91.
- CRACRAFT, J. 1988. Deep-history biogeography: Retrieving the historical pattern of evolving continental biotas. *Syst. Zool.* 37:221-236.
- FAITH, D. P., AND P. S. CRANSTON. 1991. Could a cladogram this short have arisen by chance alone?: On permutation tests for cladistic structure. *Cladistics* 7:1-28.
- FARRIS, J. S. 1976. Expected asymmetry of phylogenetic trees. *Syst. Zool.* 25:196-198.
- FELSENSTEIN, J. 1992. Estimating effective population size from samples of sequences: Inefficiency of pairwise and segregating sites as compared to phylogenetic estimates. *Genet. Res.* 49:139-147.
- GLEN, D. R., AND D. R. BROOKS. 1986. Parasitological evidence pertaining to the phylogeny of the hominoid primates. *Biol. J. Linn. Soc.* 27:331-354.
- HAFNER, M. S., AND S. A. NADLER. 1988. Phylogenetic trees support the coevolution of parasites and their hosts. *Nature* 332:258-259.
- HAFNER, M. S., AND S. A. NADLER. 1990. Cospeciation in host-parasite assemblages: Comparative analysis of rates of evolution and timing of cospeciation events. *Syst. Zool.* 39:192-204.
- HARDING, E. F. 1971. The probabilities of rooted tree shapes generated by random bifurcation. *Adv. Appl. Probab.* 3:44-77.
- HOBERG, E. P. 1986. Evolution and historical biogeography of a host-parasite assemblage: *Alcataenia* spp. (Cyclophyllidea: Dilepididae) in Alcidae (Charadriiformes). *Can. J. Zool.* 64:2576-2589.
- KELLOGG, V. L. 1896a. New Mallophaga. I—With special reference to a collection made from maritime birds of the Bay of Monterey, California. *Proc. Calif. Acad. Sci.* 6:31-196.
- KELLOGG, V. L. 1896b. New Mallophaga. II—From land birds; together with an account of the mallophagous mouth-parts. *Proc. Calif. Acad. Sci.* 6:431-548.
- KELLOGG, V. L. 1913. Distribution and species-forming of ectoparasites. *Am. Nat.* 47:129-158.
- KLASSEN, G. J., AND M. BEVERLY-BURTON. 1987. Phylogenetic relationships of *Ligictaluridus* spp. (Monogenea: Ancyrocephalidae) and their ictalurid (Siluriformes) hosts: An hypothesis. *Proc. Helminthol. Soc. Wash.* 54:84-90.
- KLASSEN, G. J., R. D. MOOI, AND A. LOCKE. 1991. Consistency indices and random data. *Syst. Zool.* 40:446-457.
- MADDISON, W. P., AND M. SLATKIN. 1991. Null models for the number of evolutionary steps in a character on a phylogenetic tree. *Evolution* 45:1184-1197.
- MARGUSH, T., AND F. R. MCMORRIS. 1981. Consensus n-trees. *Bull. Math. Biol.* 43:239-244.
- MEIER, R., P. KORES, AND S. DARWIN. 1991. Homoplasy slope ratio: A better measurement of observed homoplasy in cladistic analyses. *Syst. Zool.* 40:74-88.
- METCALF, M. M. 1929. Parasites and the aid they give in problems of taxonomy, geographical distribution, and palaeogeography. *Smithson. Misc. Collect.* 81:1-36.
- MITTER, C., AND D. R. BROOKS. 1983. Phylogenetic aspects of coevolution. Pages 65-98 in *Coevolution* (D. J. Futuyma and M. Slatkin, eds.). Sinauer, Sunderland, Massachusetts.
- NELSON, G. 1979. Cladistic analysis and synthesis: Principles and definitions, with a historical note on Adanson's *Familles des plantes* (1763-1764). *Syst. Zool.* 28:1-21.
- NELSON, G., AND N. PLATNICK. 1981. Systematics and biogeography: Cladistics and vicariance. Columbia Univ. Press, New York.
- NOREEN, E. W. 1989. Computer-intensive methods for testing hypotheses: An introduction. John Wiley and Sons, New York.
- PAGE, R. D. M. 1988. Quantitative cladistic biogeography: Constructing and comparing area cladograms. *Syst. Zool.* 37:254-270.
- PENNY, D., L. R. FOULDS, AND M. D. HENDY. 1982. Testing the theory of evolution by comparing phylogenetic trees constructed from five different protein sequences. *Nature* 297:197-200.
- PIELOU, E. C. 1981. Crosscurrents in biogeography. *Science* 213:324-325.
- PLATNICK, N. I., AND G. NELSON. 1978. A method of analysis for historical biogeography. *Syst. Zool.* 27:1-16.



ROHLF, F. J. 1982. Consensus indices for comparing classifications. *Math. Biosci.* 59:131–144.

ROSEN, D. E. 1978. Vicariant patterns and historical explanation in biogeography. *Syst. Zool.* 27:159–188.

SAVAGE, H. M. 1983. The shape of evolution: Systematic tree topology. *Biol. J. Linn. Soc.* 20:225–244.

SIMBERLOFF, D. 1987. Calculating probabilities that cladograms match: A method of biogeographical inference. *Syst. Zool.* 36:175–195.

SIMBERLOFF, D., K. L. HECK, E. D. MCCOY, AND E. F. CONNOR. 1981. There have been no statistical tests of cladistic biogeographical hypotheses. Pages 40–63 in *Vicariance biogeography: A critique.* (G. Nelson and D. E. Rosen, eds.). Columbia Univ. Press, New York.

SLOWINSKI, J. B. 1990. Probabilities of n-trees under two models: A demonstration that asymmetrical interior nodes are not improbable. *Syst. Zool.* 39:89–94.

SLOWINSKI, J. B., AND C. GUYER. 1989. Testing null models in questions of evolutionary success. *Syst. Zool.* 38:189–191.

SLOWINSKI, J. B., AND C. GUYER. 1993. Testing whether certain traits have caused amplified diversification: An improved method based on a model of random speciation and extinction. *Am. Nat.* 142:1019–1024.

WILEY, E. O. 1988. Parsimony analysis and vicariance biogeography. *Syst. Zool.* 37:271–290.

*Received 12 December 1994; accepted 18 September 1995*  
*Associate Editor: James Liebherr*

APPENDIX

Input file structure and output from DANRAN.EXE. Numbers in the second line of input are number of hosts, number of parasites, and number of nodes in the parasite phylogeny (i.e., characters in the first matrix).

Input file

Output

```
'Rosen fish - area relationships'
10 17 32
one 10000000010000001-----
two 01000000010000001-----
thr 00100000001000011-----
fou 00010000001000011-----
fiv 00001000000000111-----
six 00000100000001111-----
sev 00000010000011111-----
eig 00000001000111111-----
nin 00000000100111111-----
etn -----100000000000001
ntn -----010000000000011
twt -----001000000000111
teo -----000100000001111
tet -----000010000011111
teh -----000001000111111
tef -----000000101111111
tev -----000000011111111

one 11110000001000000
two 00000000100000001
thr 00001000000000001
fou 00000100000010000
fiv 00000100000010000
six 00000100010000000
sev 00000000000000100
eig 00000001000000010
nin 00000010000100000
ten 00000010000001000
```

```
Phylogenetic Covariance Probability for:
'Rosen fish - area relationships

Using: ie;
1 at 47 steps
4 at 46 steps
25 at 45 steps
83 at 44 steps
141 at 43 steps
184 at 42 steps
232 at 41 steps
177 at 40 steps
92 at 39 steps
37 at 38 steps
23 at 37 steps
1 at 36 steps
0 at 35 steps
0 at 34 steps
0 at 33 steps
0 at 32 steps

<<<<<< OBSERVED LENGTH

Your p-value is: 0.330669
```

'Hafner and Nadler - lice on gophers'  
8 10 19

Tward 0000000001000100001  
Tmino 0000000010000100001  
Gthom 000000010000010011  
Gactu 0000001000001010011  
Gewin 0000010000001010011  
Gchap 0000100000000000111  
Gpana 0001000000010001111  
Gsetz 0010000000010001111  
Gcher 0100000000100001111  
Gcost 1000000000100001111

Ttal 1010000000  
Tbot 0101000000  
Gbur 0000100000  
Ohis 0000010000  
Ocav 0000001000  
Oumd 0000000100  
Oche 0000000110  
Ohet 0000000001

'Liolopids & Verts - Brooks & McLennan'

11 6 10  
one 1000000001  
two 0100000011  
thr 0010000011  
for 0001000111  
fiv 0000101111  
six 0000011111

A 000000  
B 000000  
C 100000  
D 000000  
E 010000  
F 001000  
G 000000  
H 000100  
I 000000  
J 000010  
K 000001

'stingray paras & areas - 1st 23'  
6 13 23

one 1000111-----  
two 0100111-----  
thr 0010011-----  
fou 0001001-----  
eig -----10011-----  
nin -----01011-----  
ten -----00101-----  
tht -----10000000111  
fot -----01000000111  
fit -----00100000011  
sit -----00010001001  
set -----00001011001  
eit -----00000111001

upar 0001000000100  
mana 0000000000010  
leti 0010001001000  
orin 0100011010101  
mara 1000001010000  
magd 1000100100000

Phylogenetic Covariance Probability for:  
'Hafner and Nadler - lice on gophers

Using: tread (0((1 2)((3(4 5))(6(7 8)))));  
7 at 35 steps  
25 at 34 steps  
60 at 33 steps  
120 at 32 steps  
155 at 31 steps  
168 at 30 steps  
177 at 29 steps  
134 at 28 steps  
75 at 27 steps  
47 at 26 steps <<<<<< OBSERVED LENGTH  
19 at 25 steps  
9 at 24 steps  
3 at 23 steps  
1 at 22 steps  
0 at 21 steps

Your p-value is: 0.079920

Phylogenetic Covariance Probability for:  
'Liolopids & Verts - Brooks & McLennan

Using: tread (0(1((2 3 4)(5 6 ((7 8)(9(10  
11)))))));  
210 at 14 steps  
288 at 13 steps  
299 at 12 steps  
147 at 11 steps  
56 at 10 steps <<<<<< OBSERVED LENGTH

Your p-value is: 0.056943

Phylogenetic Covariance Probability for:  
'stingray paras & areas - 1st 23

Using: ie;  
6 at 33 steps  
49 at 32 steps  
173 at 31 steps  
269 at 30 steps  
251 at 29 steps <<<<<< OBSERVED LENGTH  
162 at 28 steps  
65 at 27 steps  
21 at 26 steps  
4 at 25 steps  
0 at 24 steps  
0 at 23 steps

Your p-value is: 0.503497

```
'stingray paras & areas - all 33'
6 23 33
one 1000111-----
two 0100111-----
thr 0010011-----
fou 0001001-----
eig -----10011-----
nin -----01011-----
ten -----00101-----
tht -----10000000111-----
fot -----01000000111-----
fit -----00100000011-----
sit -----00010001001-----
set -----00001011001-----
eit -----00000111001-----
tfo -----1-----
tfi -----1-----
tsi -----1-----
tse -----1-----
tei -----1-----
tni -----1-----
tht -----1-----
tho -----1-----
thu -----1-----
tht -----1-----
```

```
upar 00010000001001100000011
mana 00000000000101100000000
leti 00100010010000011100000
orin 01000110101011100101100
mara 10000010100000000000000
magd 10001001000000000010000
```

```
'stingray paras - composite areas'
10 23 33
one 1000111-----
two 0100111-----
thr 0010011-----
fou 0001001-----
eig -----10011-----
nin -----01011-----
ten -----00101-----
tht -----10000000111-----
fot -----01000000111-----
fit -----00100000011-----
sit -----00010001001-----
set -----00001011001-----
eit -----00000111001-----
tfo -----1-----
tfi -----1-----
tsi -----1-----
tse -----1-----
tei -----1-----
tni -----1-----
tht -----1-----
tho -----1-----
thu -----1-----
tht -----1-----
```

```
upar 00010000001001100000011
mana 00000000000101100000000
leti 00100010010000011100000
orin1 01000110100000000001100
orin2 0000000001011100000000
orin3 00000010000000000100000
mara1 00000010000000000000000
mara2 10000000000000000000000
mara3 00000000100000000000000
magd 10001001000000000010000
```

Phylogenetic Covariance Probability for:  
'stingray paras & areas - all 33

Using: ie;  
6 at 45 steps  
26 at 44 steps  
76 at 43 steps  
209 at 42 steps  
241 at 41 steps <<<<<< OBSERVED LENGTH  
218 at 40 steps  
139 at 39 steps  
62 at 38 steps  
16 at 37 steps  
7 at 36 steps  
0 at 35 steps  
0 at 34 steps  
0 at 33 steps

Your p-value is: 0.683317

Phylogenetic Covariance Probability for:  
'stingray paras - composite areas

Using: ie;  
3 at 47 steps  
6 at 46 steps  
30 at 45 steps  
53 at 44 steps  
140 at 43 steps  
231 at 42 steps  
205 at 41 steps  
183 at 40 steps  
89 at 39 steps  
44 at 38 steps  
14 at 37 steps <<<<<< OBSERVED LENGTH  
2 at 36 steps  
0 at 35 steps  
0 at 34 steps  
0 at 33 steps

Your p-value is: 0.016983

'Brooks & Glen - Worms on primates'  
5 10 18

worm1 1000000001-----  
worm2 0100000011-----  
worm3 00100000111-----  
worm4 00010001111-----  
worm5 00001011111-----  
worm6 00000111111-----  
worm12 -----1000001  
worm13 -----0100011  
worm14 -----0010111  
worm15 -----0001111

Hyalo 1110001000  
Pongo 0010000100  
Homo 0000101000  
Pan 0000100010  
Gorri1 0000100001

Phylogenetic Covariance Probability for:  
'Brooks & Glen - Worms on primates

Using: ie;  
78 at 21 steps  
219 at 20 steps  
371 at 19 steps  
248 at 18 steps <<<<<< OBSERVED LENGTH  
79 at 17 steps  
5 at 16 steps

Your p-value is: 0.332667