Syst. Biol. 51(2):217–234, 2002

A Statistical Test for Host–Parasite Coevolution

PIERRE LEGENDRE,[1] YVES DESDEVISES,[1,2] AND ERIC BAZIN[1]

[1]*Département de sciences biologiques, Université de Montréal, C.P. 6128, succursale Centre-ville, Montréal, Québec, Canada H3C 3J7; E-mail: Pierre.Legendre@umontreal.ca*

[2]*Centre de Biologie et d'Écologie Tropicale et Méditerranéenne, UMR CNRS 5555, Université de Perpignan, 52 Avenue de Villeneuve, F-66860 Perpignan Cedex, France*

Abstract.—A new method, ParaFit, has been developed to test the significance of a global hypothesis of coevolution between parasites and their hosts. Individual host–parasite association links can also be tested. The test statistics are functions of the host and parasite phylogenetic trees and of the set of host–parasite association links. Numerical simulations are used to show that the method has correct rate of type I error and good power except under extreme error conditions. An application to real data (pocket gophers and chewing lice) is presented. [Coevolution; fourth-corner statistic; host-parasite; permutation test; phylogenetic analysis; power analysis; simulations; statistical test.]

Parasites generally form tight ecological associations with their hosts. Biologists have long assumed that the evolution of parasites is highly dependent on that of their hosts (Barrett, 1986; Klassen, 1992). This has led to the establishment of several "rules," such as Farenholz's Rule (1913)—"parasite phylogeny mirrors host phylogeny"—and Szidat's Rule (1940)—"primitive hosts harbour primitive parasites." Klassen (1992) summarized the history of host–parasite (H-P) coevolution studies through 1991. The term "coevolution" was introduced by Ehrlich and Raven (1964) in a study on butterflies and their plant hosts. In the present paper, coevolution is defined as the extent to which the host and parasite phylogenetic trees are congruent, where congruence refers to the degree to which parasites and their hosts occupy corresponding positions in the phylogenetic trees. Perfect congruence is a good indicator of host and parasite cospeciation; a total absence of congruence, on the other hand, indicates random associations in their evolutionary history. This definition corresponds to that of Brooks (1979, 1985) and Klassen (1992), which refers to the macroevolutionary context.

Until the work of Brooks (1977, 1981) at the end of the 1970s, no rigorous analytical method had been developed to study H-P coevolution. Since then, several methods for doing so have been developed: Brooks parsimony analysis (BPA; Brooks and McLennan, 1991), component analysis (Component; Page, 1993a), a method based on reconciled phylogenetic trees (TreeMap; Page, 1994), event-based methods (e.g., TreeFitter [Ronquist, 1995, 1997]; Jungles [Charleston, 1998]), and a maximum likelihood-based test (Huelsenbeck et al., 1997). This more rigorous framework led to the publication of several studies about host–parasite coevolution (e.g., Brooks and Glen, 1982; Hafner and Nadler, 1988, 1990; Klassen and Beverley-Burton, 1988; Demastes and Hafner, 1993; Page, 1993b; Paterson et al., 1993; Hafner et al., 1994; Page and Hafner, 1996; Boeger and Kritsky, 1997; Roy, 2001). This topic has gained considerable importance in evolutionary biology (Futuyma and Slatkin, 1983; Brooks and McLennan, 1991; Thompson, 1994; Page and Holmes, 1998).

The above-mentioned methods treat differently the different kinds of evolutionary events occurring in a host–parasite association (Ronquist, 1997; Page and Charleston, 1998). Consequently, they can produce different results. The simultaneous evolution of hosts and parasites can exhibit four main different kinds of events (Ronquist, 1997; Charleston, 1998; Page and Charleston, 1998): cospeciation (simultaneous speciation of a host and its parasite), duplication (independent parasite speciation), lineage sorting (disappearance of a parasite lineage on a host lineage), and host switching (colonization of a new host by a parasite). A drawback common to all the above-mentioned methods is that they are ideally designed for the one host-one parasite case; as the numbers of hosts and parasites increases, the problem becomes highly computer-intensive, making optimal solutions hard to find. These methods aim at reconstructing a putative history of the host–parasite association by adequately mixing the types of events and trying to minimize the overall cost of

the estimated evolutionary scenario. They attempt to answer the question, What is the most probable coevolutionary history of the host–parasite association, given the costs of the different events?

PRINCIPLE OF THE TEST

In the present paper, we want to know if the data agree with a model of coevolution of the hosts and parasites. The corresponding null hypothesis (H_0) is that the evolution of the two groups, as revealed by the two phylogenetic trees and the set of H-P association links, has been independent, which is the same as saying that one is random with respect to the other. Strict coevolution requires two conditions to be fulfilled:

1. Ideally, the "true phylogenies" of the hosts and parasites should be the same. In ectoparasites, for instance, the similarity may be attributable to the geographic separation of the host lineages, which leads to allopatric speciation of the hosts and parasites (e.g., Barker, 1994; Hafner et al., 1994). These phylogenies are known, however, only by the estimates that have been worked out by researchers; these estimates may be imperfect.
2. The hosts and parasites located in corresponding positions of their respective trees must be associated (linked).

Ideally, there should be a one-to-one relationship between the hosts and parasites, but the existence of several parasites per host or several hosts per parasite does not rule out a strict pattern of coevolution generated, for example, by parasite intrahost speciation, or by host speciation not followed by parasite speciation (called "parasite inertia" by Paterson and Banks, 2001). In all cases, each H-P link must be assessed separately for its fit to the coevolution hypothesis, given the uncertainties of the estimates of the two phylogenetic trees and the H-P association matrix.

Coevolution can be tested by the method we describe here, focusing on the structure of the two trees and the matrix of host–parasite links. The global null hypothesis is that evolution of the hosts and parasites has been independent. One can also consider hypotheses about individual H-P links to estimate the contribution of each link to the overall relationship.

DESCRIPTION OF THE PARAFIT TEST STATISTICS

ParaFit Statistic for the Global Test of Host–Parasite Coevolution

Statistically assessing a hypothesis of H-P coevolution requires combining three types of information that are jointly necessary to describe the situation: the phylogeny of the parasites, the phylogeny of the hosts, and the observed H-P associations (Fig. 1). A phylogeny can be described by a matrix of patristic distances among the species along the tree (Lapointe and Legendre, 1992). These in turn can be transformed into a matrix of principal coordinates (Gower, 1966; principal coordinate analysis is also described in textbooks of biological multivariate statistics, e.g., Legendre and Legendre, 1998) having n rows and at most $n - 1$ columns. If only the tree topology is to be used, without consideration of the differences in branch lengths, one may code the tree with branch lengths of 1 before computing the patristic distance matrix. Matrices **B** and **C** in Figure 1 were so obtained and respectively represent the phylogeny of the parasites and that of the hosts. Matrix **A** represents the H-P associations: With the

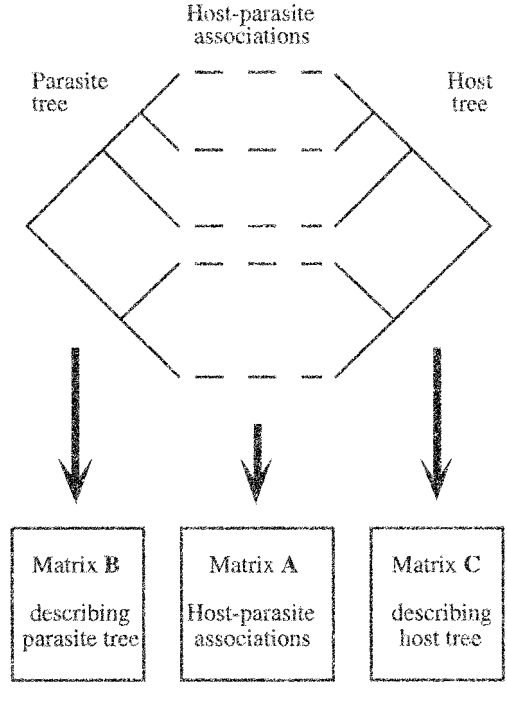

FIGURE 1. The three elements of the H-P coevolution problem can be translated into rectangular data matrices **A**, **B**, and **C**. See text.

parasites in rows and the hosts in columns, a 1 is written where a parasite has been empirically found to be associated to a host and 0's are used elsewhere.

If the reconstructed phylogeny for either the hosts or the parasites, or both, is uncertain (e.g., poorly resolved phylogeny or uncertainty among several almost equivalent trees), one can use, instead of patristic distances, a matrix of phylogenetic distances computed directly from the raw data (morphology, DNA sequences, and so forth). Matrix **B** or **C** can be derived from that matrix by using principal coordinate analysis. Because distances do not necessarily obey the four-point condition (Buneman, 1974), they are more remote from the "true tree" than an estimated phylogeny would be and thus may be considered to represent a more noisy form of information; however, this does not invalidate the use of the corresponding principal coordinates for testing a hypothesis of H-P coevolution.

Matrices **A**, **B**, and **C** can be combined in a meaningful way if they are positioned as in Figure 2. Note that matrix **C** is transposed (principal coordinates are now use for the rows of **C**) to ensure that its columns correspond to the hosts, as in the columns of **A**. Figure 2 suggests that the H-P association can be described by a matrix **D**, which crosses **B** and **C** and depicts the H-P association between the two phylogenies. **D** can be obtained by the matrix operation

$$\mathbf{D} = \mathbf{C}\,\mathbf{A}'\,\mathbf{B} \tag{1}$$

described by Legendre et al. (1997; see also Legendre and Legendre, 1998: Section 10.6), who coined the expression "fourth-corner statistics" to describe the individual values in matrix **D**, as well as some global statistic synthesizing the information in **D**. Those authors showed that when the variables in the columns of **B** and the rows of **C** are quantitative, the individual parameters in **D** are cross-products weighted by the presence–absence (1–0) values found in **A**.

Trying to interpret the individual parameter estimates found in **D** is pointless because those estimates cross principal coordinates not meant to be individually interpretable in the present context. Instead, we will derive a global H-P statistic to test the hypothesis of coevolution. For this global statistic we will use the sum of squares of the values d_{ij} found in matrix **D**:

$$\text{ParaFitGlobal} = \text{trace}(\mathbf{D}'\mathbf{D}) = \sum (d_{ij}{}^2) \tag{2}$$

This is analogous to a trace statistic computed on the matrix of sum of squares and cross products (SSCP) among either the rows or the columns of **D**—except for the centering, either by rows or by columns, that would be required to obtain a real SSCP matrix and its trace. In the present case, we have no preference between the SSCP matrix among the rows of **D** (which would require centering by rows) or that among the columns of **D** (which would require centering by columns), so we use no centering at all. For justification, we provide numerical simulations showing that the global ParaFitGlobal statistic has correct type I error.

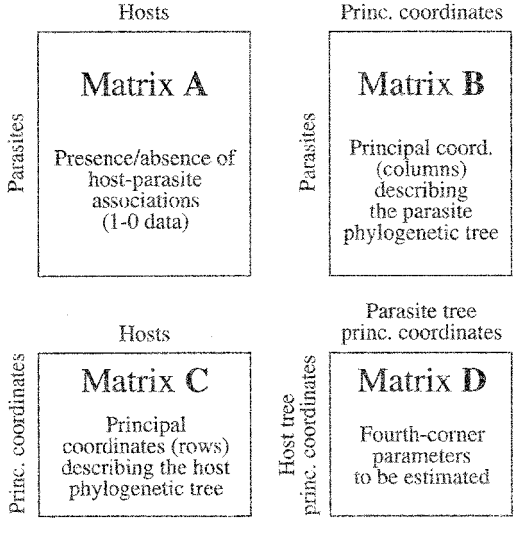

FIGURE 2. Given the information in matrices **A**, **B**, and **C**, the problem is to estimate the parameters in the fourth-corner matrix **D** that crosses the principal coordinates of the hosts with those of the parasites.

ParaFit Statistics for Tests of Individual Host–Parasite Association Links

The procedure described below will allow us to test the individual links written in matrix **A**, which represent observed H-P associations. Two statistics are developed to do that. They are based on the idea that the global statistic should decrease in value if we remove from **A** a link that represents an important contribution to the H-P relationship.

For the test, consider the H-P link k from matrix **A**. If we replace the value 1 that represents this link in matrix **A** with a 0 (no link), we obtain a new matrix **A**(k). We now compute

$$\text{trace}(k) = \sum \left(d_{ij}{}^2\right) \quad (3)$$

where the d_{ij} values now are those resulting from the product $\mathbf{D} = \mathbf{C}\,\mathbf{A}(k)'\mathbf{B}$. Using the trace statistic from Eq. 2, that we call "trace," we obtain with the following equation a first test statistic for an individual link k:

$$\text{ParaFitLink1}(k) = \text{trace} - \text{trace}(k) \quad (4)$$

This formula measures the contribution of link k to the global trace statistic.

The second statistic is constructed similar to a partial F-statistic (see, for instance, Sokal and Rohlf, 1995: Eq. 16.14) that would have lost its degrees of freedom in the numerator and denominator. This loss is of no consequence in permutation testing because the reference and permuted values of the statistic are affected in the same way by a constant multiplicative term, so that the ordering of the reference and permuted values (>, =, or <) is not changed by eliminating the degrees of freedom. The numerator of the statistic is [trace – trace(k)] from Eq. 4. The denominator is analogous to a residual sum of squares. To construct the denominator, we need a measure of the maximum trace value that can occur in **D**. The maximum possible value occurs when the hosts and parasite phylogenetic trees are fully congruent—a relevant reference situation to test a hypothesis of coevolution. In that case, it can be shown that the trace of **D** is equal to the sum of squares of the eigenvalues of the principal coordinates found in matrix **B** or matrix **C**. In principal coordinate analysis, the eigenvalues measure the variances of the principal coordinates (Gower, 1966; Legendre and Legendre, 1998). In most empirical situations, the estimates of the two phylogenetic trees will not be exactly alike, so we will actually use:

$$\begin{aligned}&\text{TraceMax}\\ &\quad = \max(\text{sum of squared eigenvalues of } \mathbf{B},\\ &\quad \text{sum of squared eigenvalues of } \mathbf{C}) \quad (5)\end{aligned}$$

as our estimate of maximum trace value.

The second test statistic that we are proposing for an individual link k is thus:

$$\text{ParaFitLink2}(k) = (\text{trace} - \text{trace}(k))/(\text{TraceMax} - \text{trace}) \quad (6)$$

This statistic cannot be used when the host and parasite phylogenetic trees are identical because, in that situation, the denominator of Eq. 6 is 0. This situation may happen in simulation work but should rarely occur when analyzing empirical data.

On occasions, trace(k) may happen to be slightly larger than trace. This is of no consequence because it indicates a situation where the H-P relationship is stronger without link k than with it. In other words, link k does not increase the global H-P relationship and thus is not indicative of H-P coevolution. The test of significance (next section) will never find such a link significant—which is the correct answer.

Testing Procedures

Novel statistics, such as described in Eqs. 2, 4, and 6, may be tested for significance by using the method of permutations, also called randomization, which is now widely used in biological work. The method is described in several texts (e.g., Sokal and Rohlf, 1995; Manly, 1997; Legendre and Legendre, 1998).

Global Test of Host–Parasite Coevolution

Consider matrices **A**, **B**, and **C** described above. The two phylogenetic trees are given as fixed because they have been formed through evolutionary time; we are not directly interested in testing their similarity. The random component is clearly the set of association links found in matrix **A**, which, under the null hypothesis, may change through ecological time. Siddall (1996) provides other arguments in favor of randomizing the H-P associations. The test will involve random permutations of the hosts associated with each parasite because the parasites parasitize the hosts, not the opposite. Accordingly, if there is no coevolutionary H-P association, each parasite species should parasitize hosts selected at random on the host phylogenetic tree. This is the null hypothesis of the test (H_0). The alternative hypothesis (H_1) is that the positions of the individual

H-P links are not random but associate corresponding branches of the two evolutionary trees; in that sense, they are critical to the overall H-P association. The testing procedure is as follows:

1. Compute matrix **D** by using Eq. 1. Compute statistic ParaFitGlobal by using Eq. 2, which provides the reference value ($\text{ParaFitGlobal}_{\text{ref}}$) for the test. Save this value as $\text{trace}_{\text{ref}}$ for use with the tests of individual H-P links, described below.
2. To obtain a realization of the null hypothesis, permute at random the values within each row of matrix **A**, and do this independently from row to row. Recompute matrix **D** by Eq. 1 and the statistic ParaFitGlobal by Eq. 2. This provides a value (ParaFitGlobal*) of the statistic under permutation. Save each value as trace* = ParaFitGlobal* for use with the tests of individual H-P links (below).
3. Repeat step 2 a large number of times to obtain an estimate of the distribution of the statistic under permutation. Add the reference value $\text{ParaFitGlobal}_{\text{ref}}$ to the distribution, following Hope (1968).
4. Calculate the one-tailed probability (*P*-value) of the data under the null hypothesis as the proportion of values in the ParaFitGlobal* distribution that are larger than or equal to $\text{ParaFitGlobal}_{\text{ref}}$. The test indicates that the data are unlikely to correspond to the null hypothesis if $\text{ParaFitGlobal}_{\text{ref}}$ is larger than or equal to most (say, 95% for $\alpha = 0.05$) of the ParaFitGlobal* values obtained under permutation.

Tests of Individual Host–Parasite Association Links

Tests of individual H-P association links can also be computed to determine the probability that individual H-P links conform to the null hypothesis. This is done as follows:

1. Compute TraceMax from matrices **B** and **C** by using Eq. 5.
2. Choose a H-P link k and remove it from matrix **A**.
3. Compute matrix **D** for the modified matrix **A** by using Eq. 1. Compute trace(k) by Eq. 3. Calculate the values of the two reference statistics for the tests, $\text{ParaFitLink1}(k)_{\text{ref}}$ (Eq. 4) and $\text{ParaFitLink2}(k)_{\text{ref}}$ (Eq. 6), by using the value $\text{trace}_{\text{ref}}$ saved from the global test of coevolution.
4. Permute at random the values within each row of matrix **A**, independently from row to row, using the same sequence of random permutations as were used during the global test of H-P coevolution (above). Recompute matrix **D** by using Eq. 1 and trace(k)* by Eq. 3, and then recompute the values of the two statistics under permutation: ParaFitLink1(k)* (Eq. 4) and ParaFitLink2(k)* (Eq. 6), using the value trace* saved from the global test of coevolution.
5. Repeat step 2 a large number of times to obtain an estimate of the distribution of the two statistics under permutation. Add the reference values $\text{ParaFitLink1}(k)_{\text{ref}}$ and $\text{ParaFitLink2}(k)_{\text{ref}}$ to the distributions, following Hope (1968).
6. Calculate the one-tailed probabilities (*P*-values) of the data under the null hypothesis as the proportions of values in the ParaFitLink1(k)* and ParaFitLink2(k)* distributions that are larger than or equal to $\text{ParaFitLink1}(k)_{\text{ref}}$ or $\text{ParaFitLink2}(k)_{\text{ref}}$, respectively. The tests indicate that link k is unlikely to be random with respect to the coevolutionary structure if $\text{ParaFitLink1}(k)_{\text{ref}}$ or $\text{ParaFitLink2}(k)_{\text{ref}}$ is larger than or equal to most (say, 95% for $\alpha = 0.05$) of the ParaFitLink1(k)* or ParaFitLink2(k)* values, respectively, obtained under permutation.

NUMERICAL SIMULATIONS

Type I Error

Simulations have been performed to check the type I error and power of the test of H-P coevolution. Type I error occurs when the null hypothesis is rejected although the data conform to H_0. To be valid, a test of significance should have a rate of rejection of the null hypothesis no larger than the nominal (α) significance level of the test when H_0 is true (Edgington, 1995).

To study type I error, we used simulated data from random phylogenetic trees for the hosts and parasites and a matrix **A** containing links sampled at random (without replication) from among all possible H-P links. Unrooted random additive (i.e., phylogenetic) trees were generated according to the method of Pruzansky et al. (1982). The two

random trees were transformed into matrices **B** (for the parasites) and **C** (for the hosts) by using principal coordinate analysis. The statistics for the global test and for the individual H-P links were calculated and tested by using random permutations.

A simulation study can never explore all parameter combinations. The simulation effort reported here for type I error involves the following parameter combinations, which we considered represented commonly encountered situations in H-P studies:

- 10 hosts, 10 parasites. Number of random H-P links: {5, 10, 15, 20, 25}.
- 10 hosts, 15 parasites. Number of random H-P links: {5, 10, 15, 20, 25}.
- 15 hosts, 10 parasites. Number of random H-P links: {5, 10, 15, 20, 25}.

For each combination of parameters, 10,000 random simulations were produced; 99 random permutations were used in each test of significance. The statistic of interest is the rate of rejection of the null hypothesis, the type I error rate, which was computed for various significance levels, $\alpha = \{0.01, 0.02, 0.03, 0.04, 0.05\}$, together with 95% confidence intervals based on the results of the 10,000 independent simulations. Optimally, the rate of rejection of the null hypothesis will be approximately equal to α, the value of which should lie within the 95% confidence interval of the rejection rate.

Power

A test of significance should be able to reject the null hypothesis in most instances when H_0 is false. The ability to reject H_0 in these circumstances is referred to as the power of the test. In the present study, power is defined as the rate of rejection of the null hypothesis when H_0 is false by construct. Power was studied by using the same type of simulations as described above, except that this time the alternative hypothesis (H_1) was made to be true. Three types of simulations were done:

1. A random additive tree was generated, as described above, for each simulation, but here the same tree was used for the parasites (matrix **B**) and the hosts (matrix **C**); this is a condition for coevolution. A H-P link was created between each host and the parasite that was found in the corresponding position on the tree. Next, a certain percentage of randomly located links were added to matrix **A** (without replication of existing links) by the same procedure as in the type I error study. The simulation parameters were as follows: 10 hosts, 10 parasites, 10 basic H-P links. The number of random H-P links added to the basic links was {0%, 10%, 20%, . . . , 100%}, or {0, 1, 2, . . . , 10} supplementary random links, for a total of {10, 11, 12, . . . , 20} links.
2. In the second series of power simulations, a random additive tree was generated for each simulation, and the same tree was used for the parasites (matrix **B**) and hosts (matrix **C**). H-P links were created between each host and the parasite in the corresponding position on the common tree. Next, a certain percentage of randomly-located links were removed from the list and replaced (without replication of existing links) with randomly located links. The simulation parameters were as follows: 10 hosts, 10 parasites, 10 basic H-P links. The number of random H-P links replacing basic links was {0%, 10%, 20%, . . . , 100%}, or {0, 1, 2, . . . , 10} links.
3. In the third type of power simulations, the host and parasite phylogenetic trees were made to have a common portion, whereas the remainder of each tree was random. The number of coevolving species was the main parameter in these simulations. (In the previous types of simulations, the trees were the same, but some of the links were random. In the present simulations, the two trees were only partially similar; coevolutionary links were created only in the similar portions of the trees, and the remaining links were random.) The simulation parameters were as follows: 10 hosts, 10 parasites, 10 H-P links. The number of species that shared the same tree (coevolving species) was: {0, 1, 2, 3, 4, 5, 6, 7, 8, 9}; hence the number of independent species (not coevolving) was {10, 9, 8, 7, 6, 5, 4, 3, 2, 1}.

In all power simulations, there were 10,000 independent simulations per combination of parameters, and 99 permutations per test.

Power was computed for various significance levels $\alpha = \{0.01, 0.02, 0.03, 0.04, 0.05\}$. The 95% confidence intervals were not plotted because they would have made

the graphs difficult to read. We want to identify the simulated situations in which power was low, medium, or high, to provide guidance for interpretation of real-case studies. We also want to see if one of the statistics (ParaFitLink1 and ParaFitLink2) created for the tests of individual H-P links had higher power than the other, at least in some situations.

All the simulations described above were repeated with random rooted ultrametric trees instead of unrooted additive trees. The generation of random ultrametric trees for n species involved two steps, described by Lapointe and Legendre (1991):

1. Random node positions were obtained by drawing $n - 1$ numbers at random from a uniform (0, 1] distribution and placing these values in the subdiagonal of an $n \times n$ matrix. The remainder of the matrix was filled by using the procedure FillMat of Lapointe and Legendre (1991).
2. The species (1 to n) were attributed at random to the leaves of the tree.

This procedure yields a random ultrametric tree, which is a special type of phylogenetic tree. The simulation results were very similar to those obtained when using random unrooted additive trees. Only the latter are discussed in detail below.

Simulation Results

Global Test of Host–Parasite Coevolution

Type I error was correct in all cases, so we can conclude that the global test is valid. Figure 3a shows the results for random and independent phylogenies for the hosts and the parasites, for equal numbers of hosts (10) and parasites (10) and various numbers (5 to 25) of random H-P links; the significance level used in the permutation tests ($\alpha = 0.01$, 0.02, 0.03, 0.04, 0.05) was always within the 95% confidence interval of the type I error rate. Repeating the simulations for type I error but using unequal numbers of hosts and parasites (10, 15 and 15, 10) gave results similar to those presented in Figure 3a: The significance levels of the permutation tests were always within the confidence intervals of the type I error rates computed from the 10,000 simulations. Similar results were obtained again when the two phylogenies were the same but the H-P links were positioned at random.

The power study is based on simulations in which the null hypothesis is false by construct; we are thus certain that the data used in these simulations represent simulated cases of coevolution. When the data exactly conformed the hypothesis of coevolution, power was maximum, the null hypothesis being rejected in nearly all cases (Figs. 3b and 3c: number of random links or supplementary random links = 0). Power decreased as the number of supplementary random links added to the coevolutionary structure increased (Fig. 3b). For $\alpha = 0.05$, power was ~55% when there were as many supplementary random links (10) as coevolutionary links (10).

Power also decreased as coevolutionary links were replaced by random links (Fig. 3c). When all coevolutionary links were replaced by random links (10 random links in Fig. 3c), the rejection level, which measures type I error in that case, was equal to the α significance level of the test. This is not a surprising result; it simply indicates that the simulation method used to generate replacement links was correct.

When the host and parasite phylogenetic trees contained a common portion, the remainder of each tree being random, power was affected by the proportion of coevolving species (Fig. 3d): With no coevolving species (left side of the graph), we are in a situation where H_0 is true, and the simulation results are identical to those of Figure 3a. Power increased with the number of coevolving species, to reach a maximum when all species were coevolving; the rejection rate was then 100% or very nearly so, as in Figures 3b and 3c (left sides). As the number of coevolving species increased, the global test was more likely to identify as significant the coevolutionary relationship present in portions of the trees, but power was low when <70% of the species were coevolving. The effect on the global test is the same as when the host and parasite trees were the same but some of the coevolutionary links had been replaced by random links (Fig. 3c, abscissa reversed). We conclude that coevolution may not be detected by the global test when a good portion of the hosts and parasites are not coevolving.

The third series of power simulations was repeated with larger numbers of hosts and

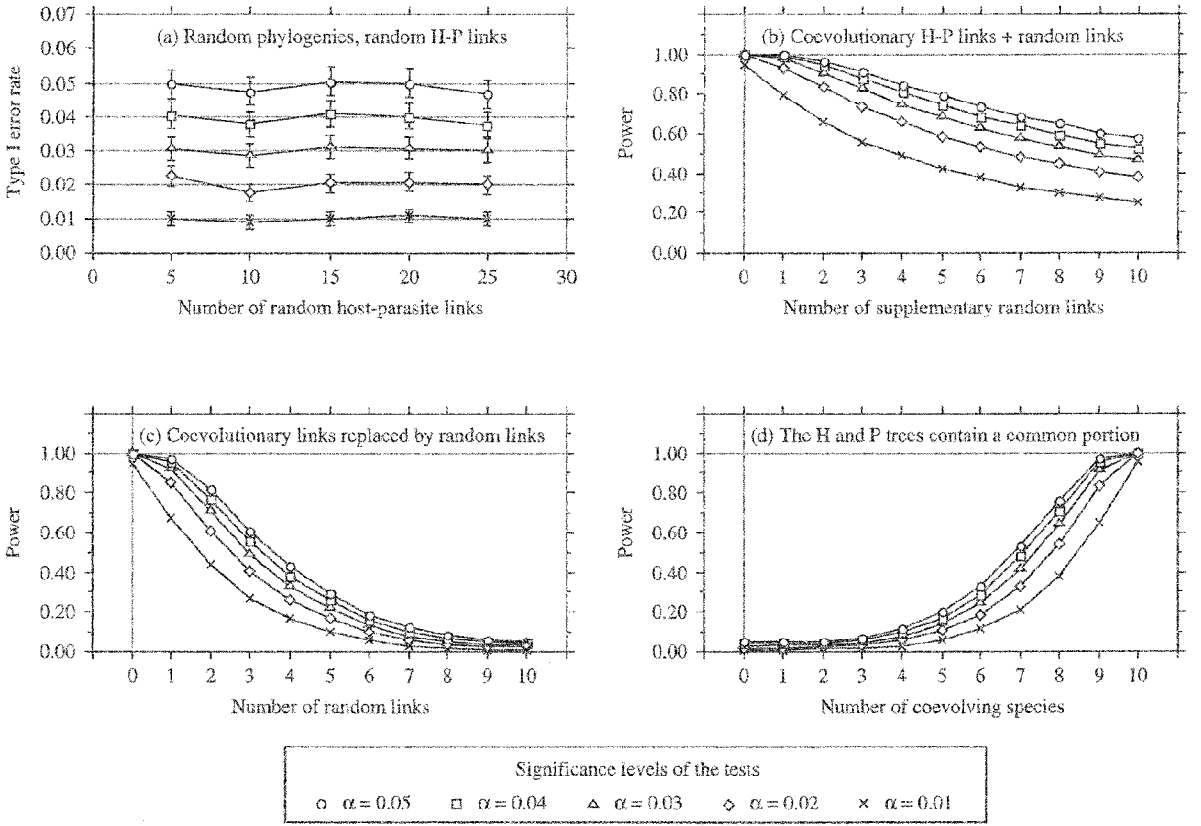

FIGURE 3. Global test of H-P coevolution (ParaFitGlobal statistic). (a) Type I error rates and 95% confidence intervals, at $\alpha = 0.01$ to 0.05, of the test for 5 to 25 H-P association links (abscissa). The phylogenies were generated independently for the hosts and the parasites. The confidence intervals are based on 10,000 replicate simulations for each combination of parameters. (b) Power of the global test of H-P coevolution in simulations in which a number of supplementary random H-P links (abscissa) were added to the coevolutionary links. (c) Same as (b), but for simulations in which some (abscissa) of the coevolutionary links were removed and replaced by random H-P links. (d) Same as (b), but for simulations in which the phylogenetic trees of the hosts and parasites contained a common portion, the remainder of each tree being random. Abscissa: number of coevolving species. All these simulations were for 10 hosts and 10 parasites.

parasites and more links. The power of the global test increased with host and parasite sample sizes, for given proportions of coevolutionary links.

Tests of Individual Host–Parasite Association Links

In tests of individual H-P association links, type I error was always correct for both statistics, ParaFitLink1 and ParaFitLink2, whether the numbers of hosts and parasites were equal (10, 10) or unequal (15, 10 or 10, 15). Therefore, the two statistics provide valid tests of significance. As an example, Figures 4a and 4b present the type I error rates obtained for the first H-P link of each simulation, in a 10,000-simulation run involving 10 hosts and 10 parasites.

For pure coevolutionary structures, the test of an individual link using statistic ParaFitLink1 had good power, but not as good as that of the global test: Compare Figure 3b with Figure 5a (left, where the number of supplementary links is 0), Figure 3c with Figure 6a (left, where the number of random links is 0), and Figure 3d with Figure 7a (right, where all 10 species were coevolving). Tests using ParaFitLink2 also have fairly good power, but this cannot be shown for pure coevolutionary structures for reasons given above. The power of individual tests increases with the number of hosts and parasites for given proportions of coevolutionary links (simulation results not shown in detail).

In the presence of supplementary random links (supplementary to a saturated coevolutionary model containing a coevolutionary link for every H-P pair), the statistic ParaFitLink1 generally had greater power than ParaFitLink2 for detecting the links that significantly contributed to H-P coevolution (Fig. 5). As noted above, statistic ParaFitLink2 cannot be computed for a perfect coevolutionary structure in the absence of random links—which is why no rejection rate is reported for the situation of no supplementary link (Fig. 5b).

The power of the test of individual coevolutionary H-P links decreased as the proportion of supplementary random links increased. For ParaFitLink1, for instance (Fig. 5a), the presence of as many supplementary links (10) as coevolutionary links (10) reduced power by about half, compared with the power for simulations without supplementary random links. With ParaFitLink1, the probability of correctly detecting a coevolutionary link (Fig. 5a) was 1.5 to 2.5 times greater than that of wrongly declaring a random link significant (Fig. 5c). The difference is not as great for statistic ParaFitLink2 (compare Figs. 5b and 5d). Accordingly, the ParaFitLink1 statistic is preferable in this situation. (In Fig. 5d, a single supplementary link added to a perfect coevolutionary structure cannot be tested because the test of significance requires that the added link be removed, which brings us back to the perfect coevolutionary case, where the ParaFitLink2 statistic cannot be computed.) Figures 5c and 5d also show that in the presence of coevolutionary links, type I error for the tests of the random links was greater than α; as a consequence, test results in this situation have to be interpreted conservatively.

When coevolutionary links were replaced by random links, the coevolutionary model was no longer saturated, meaning that it did not contain a coevolutionary link for each H-P pair. In that case, statistic ParaFitLink2 had greater power than ParaFitLink1 for detecting the links that significantly contributed to H-P coevolution (Figs. 6a, 6b). With both statistics, the probability of correctly detecting a coevolutionary link (Figs. 6a, 6b) was 1.5 to 2 times greater than that of wrongly declaring a random link significant (Figs. 6c, 6d) in the presence of 20% random links. Note that the graphs in Figures 6c and 6d converge towards the alpha significance level when all 10 coevolutionary links have been replaced by random links. Figures 6c and 6d also show that in the presence of coevolutionary links, the type I error on the random links exceeded α; consequently, test results in this situation have to be interpreted conservatively.

When the host and parasite phylogenetic trees contained a common portion and the remainder of each tree was random, both power and type I error were affected by the proportion of coevolving species (Fig. 7). The power to detect significant coevolutionary links increased with the proportion of coevolving species (Figs. 7a, 7b) in the same way as the global test of significance did (Fig. 3d). On the other hand, type I error for the test of the random links increased well above the α significance level when the number of coevolutionary species reached

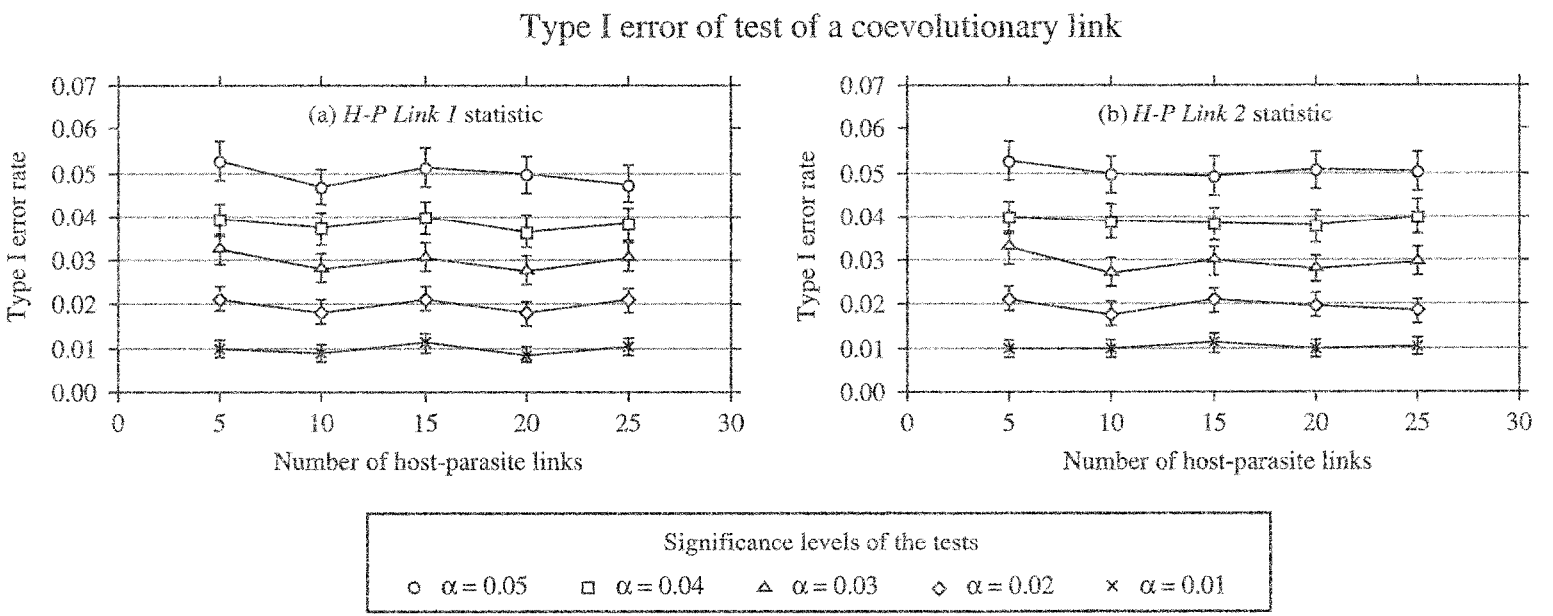

FIGURE 4. Type I error rates and 95% confidence intervals, at $\alpha = 0.01$ to 0.05, of the test of individual H-P association for the first H-P coevolutionary link generated during each simulation, for 5 to 25 H-P association links. (a) ParaFitLink1 statistic. (b) ParaFitLink2 statistic. Simulations were for 10 hosts and 10 parasites; similar results were obtained with 10 hosts and 15 parasites and with 15 hosts and 10 parasites.

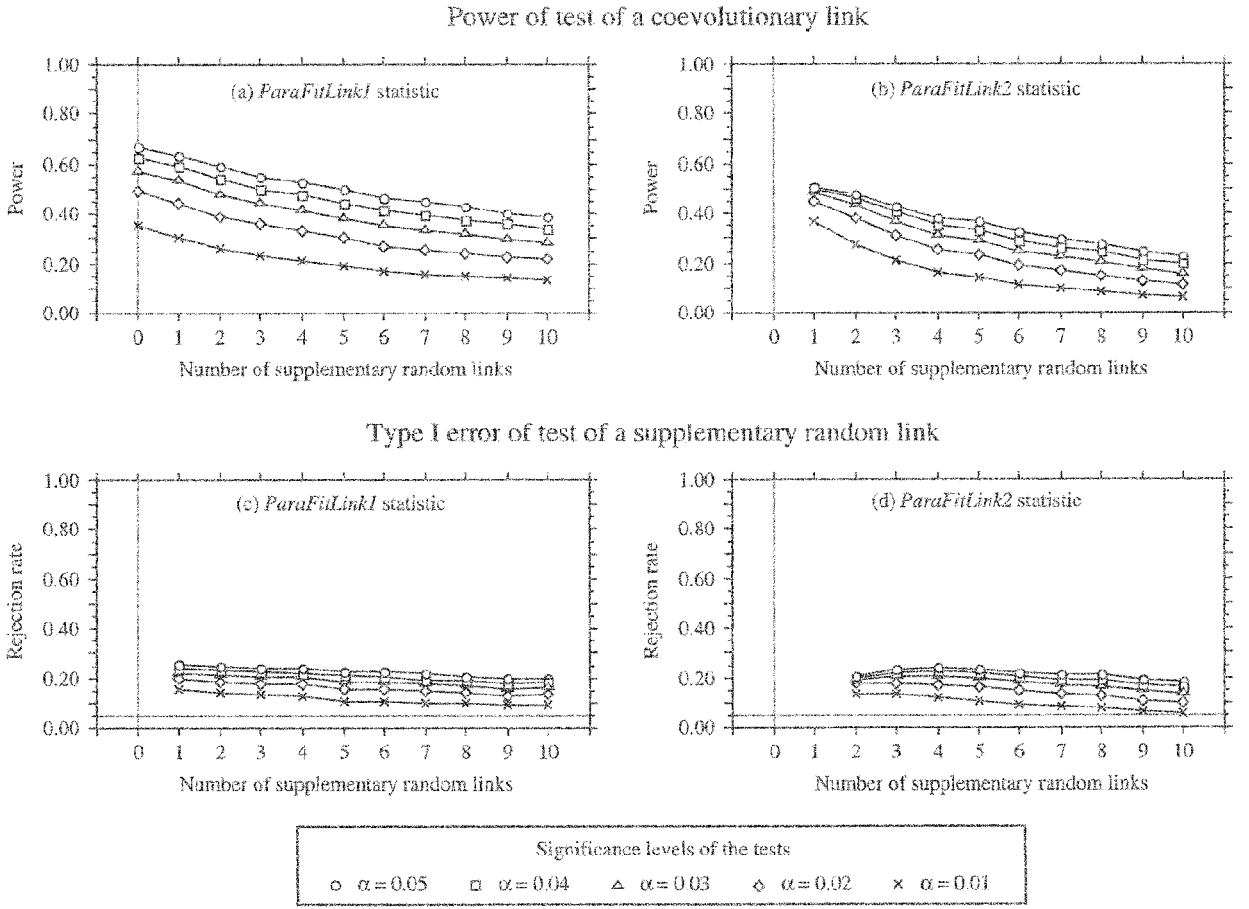

FIGURE 5. Power of the test of individual H-P association in simulations in which randomly located links were added to matrix **A**. Simulations were for 10 hosts and 10 parasites; there were also 10 coevolutionary links and 0 to 10 supplementary random links. (a) Simulation results for one of the coevolutionary H-P links, ParaFitLink1 statistic. (b) ParaFitLink2 statistic. (c) Simulation results for one of the supplementary random H-P links, ParaFitLink1 statistic; (d) ParaFitLink2 statistic.

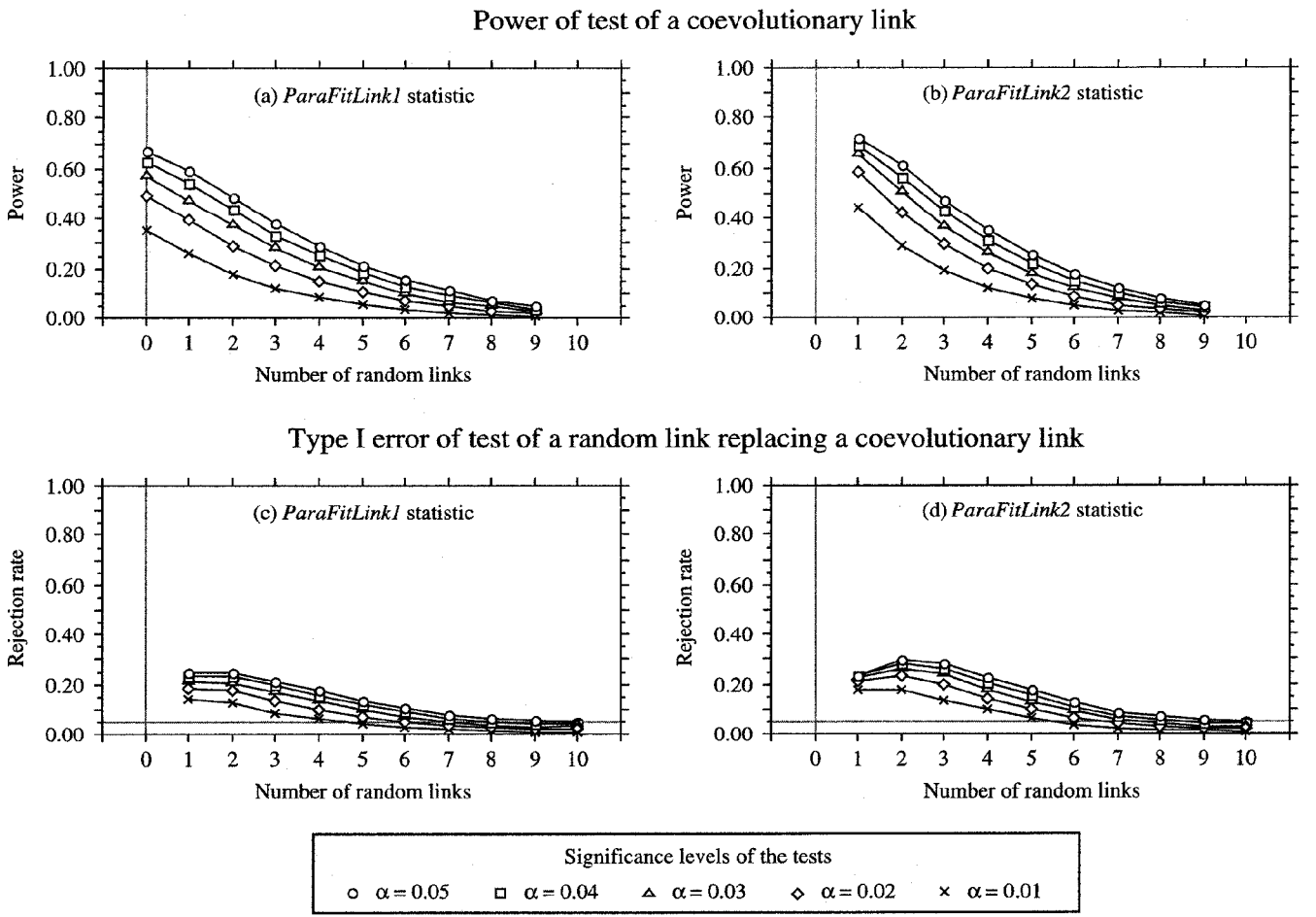

FIGURE 6. Power of the test of individual H-P association in simulations in which a certain percentage of randomly-located links were removed from matrix **A** and replaced with randomly located links. The simulations were for 10 hosts and 10 parasites. There were also 10 coevolutionary links; 0 to 10 of them were replaced with random links. (a) Simulation results for one of the coevolutionary H-P links, ParaFitLink1 statistic. (b) ParaFitLink2 statistic. (c) Simulation results for a random H-P link replacing a coevolutionary link, ParaFitLink1 statistic; (d) ParaFitLink2 statistic.

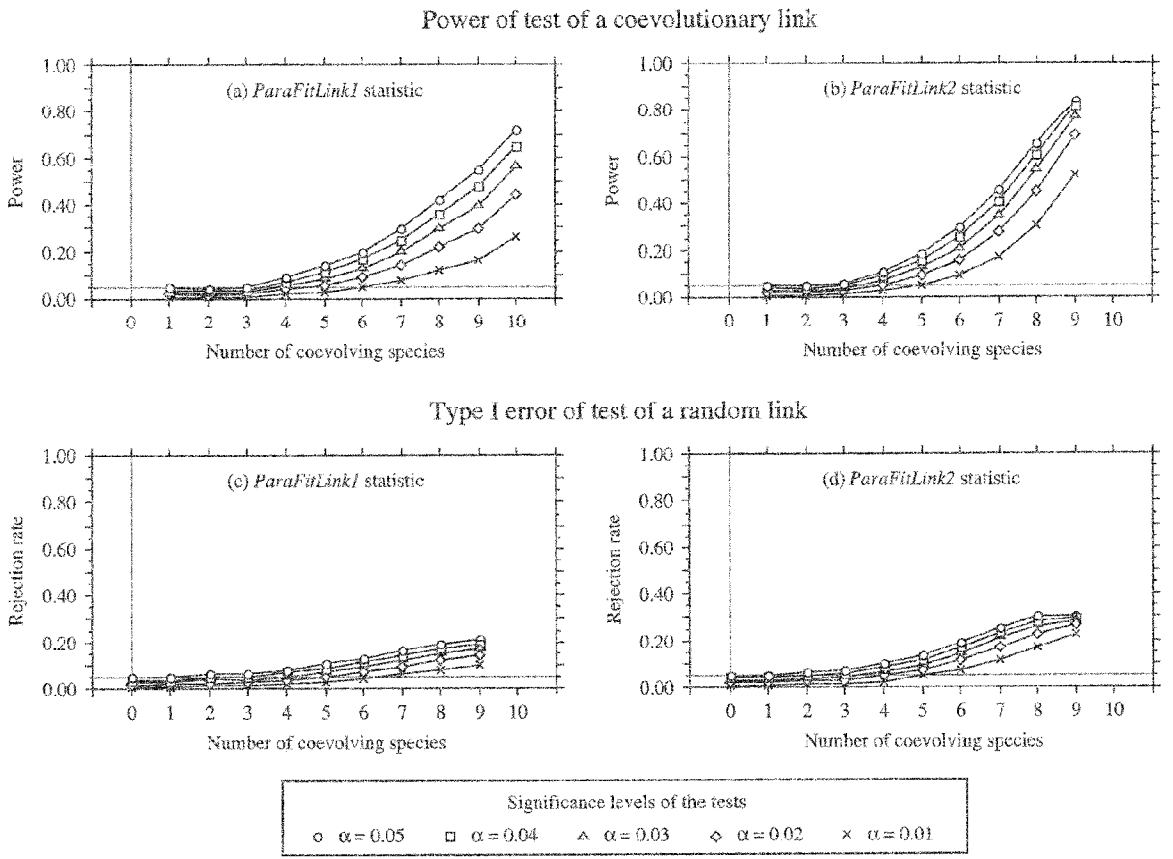

FIGURE 7. Power of the test of individual H-P association in simulations in which the phylogenetic tree of the hosts and that of the parasites were partly similar and partly random. There were 10 hosts, 10 parasites, and 10 links in this particular set of simulations. Abscissa: number of species of hosts and parasites (among 10) that shared the same tree and were linked. (a) Simulation results for one of the coevolutionary H-P link, ParaFitLink1 statistic. (b) ParaFitLink2 statistic. (c) Simulation results for one of the random H-P links, ParaFitLink1 statistic; (d) ParaFitLink2 statistic.

about one-half the total number of species; the effect on statistic ParaFitLink1 was less important than on statistic ParaFitLink2 (Figs. 7c, 7d), making ParaFitLink1 seem preferable. The probability of correctly detecting a coevolutionary link was greater than that of wrongly declaring a random link significant by about the same factor with both statistics (compare Figs. 7a with 7c and Figs. 7b with 7d); that is, the two statistics are equivalent from that point of view. Considering the smaller degree of inflation of type I error displayed by ParaFitLink1, compared with ParaFitLink2, ParaFitLink1 seems preferable in this situation. In any case, test results in this situation have to be interpreted cautiously.

Further simulations conducted with larger numbers of hosts and parasites and more links showed that power of the ParaFitLink1 and ParaFitLink2 tests increased with increasing host and parasite sample sizes for given proportions of coevolutionary links.

Interpretation of the Test Results

The null hypothesis of the global test of significance for H-P coevolution is that the evolution of the two groups, as revealed by the two phylogenetic trees and the set of H-P association links, has occurred independently. Test results are interpreted as follows:

- A significant global test result indicates that the test has detected a significant H-P association, with a probability of type I error equal to α.
- A nonsignificant global test result indicates either that there is no H-P association, or that the H-P association is masked by supplementary random H-P links (see the results of the first series of power simulations in which randomly located links were added to matrix **A**), or the structure is a mixed one with parts of the two trees coevolving whereas other portions are not coevolving (see the results of the second and third series of power simulations). The test has good power only when most of the host and parasite species are coevolving and the number of random links is not too great. Tests of individual H-P links are still possible but results must be interpreted conservatively, as discussed below.

In the tests of individual H-P association links, the null hypothesis is that the link being tested is random. Results of tests of individual links are interpreted as follows:

- When results of the global test and the test of an individual link are both significant, the test has detected a significant H-P link, with a probability of type I error equal to α.
- When the global test gives a significant result and the test of an individual link does not, the data do not support the hypothesis that the link represents a coevolutionary link. Because the test of individual links has less power than the global test, the test of a H-P link based on a small number of hosts, parasites, and coevolutionary links may turn out not to be significant because of lack of power.
- When the result of the global test is not significant but that of the test of an individual link is, then presumably we are dealing with a mixed structure containing perhaps a coevolutionary structure with some random links added (as in the first series of power simulations), or perhaps the structure has a coevolutionary portion and a random one (as in the second and third series of power simulations). Only links that are very highly significant should be considered, to compensate for the fact that the tests of individual links have inflated type I error in this situation.
- When neither the global test nor the test of an individual link gives significant results, the link is unlikely to represent a coevolutionary H-P association.

In summary, the ParaFitLink1 statistic generally behaved better in simulations and should be preferred to ParaFitLink2.

Application of the Test: Gophers and Lice

We tested our method by using phylogenetic trees for pocket gophers and their chewing lice (Hafner and Nadler, 1988, 1990; Hafner et al., 1994) reconstructed from the mtDNA cytochrome oxidase I sequences used in Hafner et al. (1994). This data set has become a test case for coevolutionary studies and has been reexamined several times with various new methods (e.g., Ronquist,

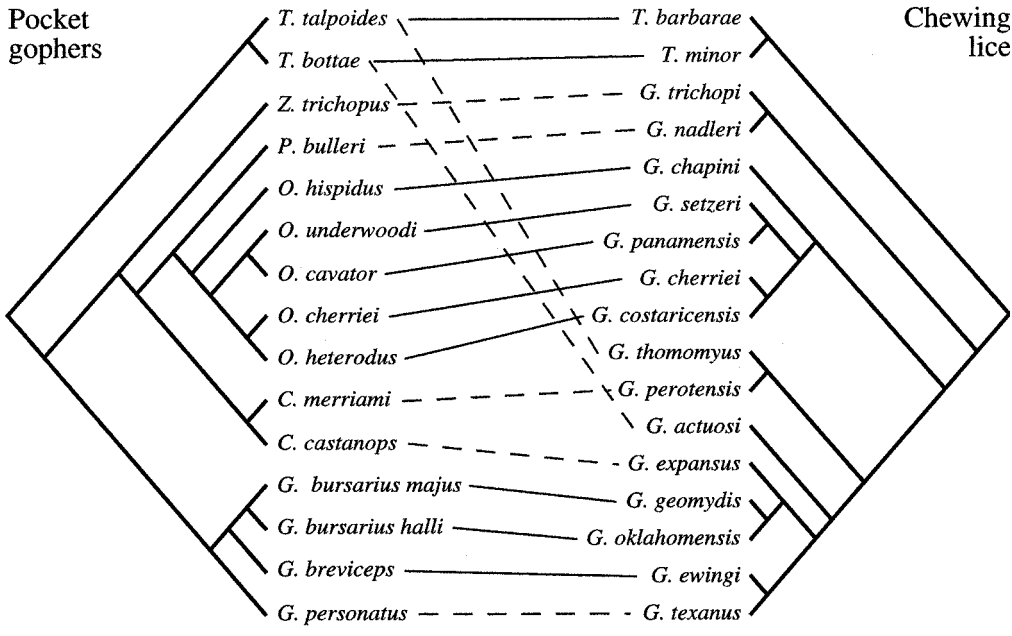

FIGURE 8. Pocket gophers and chewing lice phylogenetic trees and H-P links. Significant H-P links are represented by full lines, nonsignificant links by dashed lines.

1995; Page, 1996; Charleston, 1998). We used Hasegawa–Kishino–Yano (HKY85) corrected distances because of an observed heterogeneity of nucleotide frequencies and the occurrence of a transition bias. Trees were reconstructed using the neighbor-joining method (Saitou and Nei, 1987) available in the program PAUP* (Swofford, 2001). We used this method to quickly obtain an estimate of the phylogeny of the hosts and parasites; however, this should not be taken as an endorsement of neighbor-joining as the best method for exploring tree space. The trees that we obtained differ slightly from those published by Hafner et al. (1994) because we used distances corrected under a different evolutionary model. This produced a topology slightly different from that published in Hafner et al. (1994). The exact tree topology, however, is of little importance for illustrating our method in the present study.

Coevolution was first tested by using the global ParaFit statistic (H_0: evolution of the hosts and parasites has occurred independently). The result, that the null hypothesis must be rejected (permutational $P = 0.001$ after 999 permutations), supports the alternative hypothesis (H_1) of coevolution, revealed by the similarity of the two phylogenetic trees and the matrix of H-P association links. This result is in agreement with previous studies on this H-P model, which suggest an important amount of cospeciation between pocket gophers and their chewing lice. When we assessed the significance of each H-P association, using the ParaFitLink1 statistic (tested at $\alpha = 0.05$), we found that 7 of the 17 H-P links were not significant (Fig. 8).

The significant links display extensive coevolution between the hosts and parasites. The discrepancies between the two phylogenies are associated with the nonsignificant H-P links. If we remove those from Figure 8, and remove also the species that have no significant H-P link, the result is two identical phylogenetic trees displaying perfect coevolution for a subset of the gophers and lice (Fig. 9).

The ParaFit method does not intend to infer an evolutionary scenario explaining the observed historical association between gophers and their lice—in contrast to the TreeMap method, for example. It did allow us, however, to identify (1) the hosts and parasites that have probably undergone cospeciation and (2) the species most likely to have been subjected to host-switching or sorting events (parasite extinction, or primary absence on daughter host lineage)

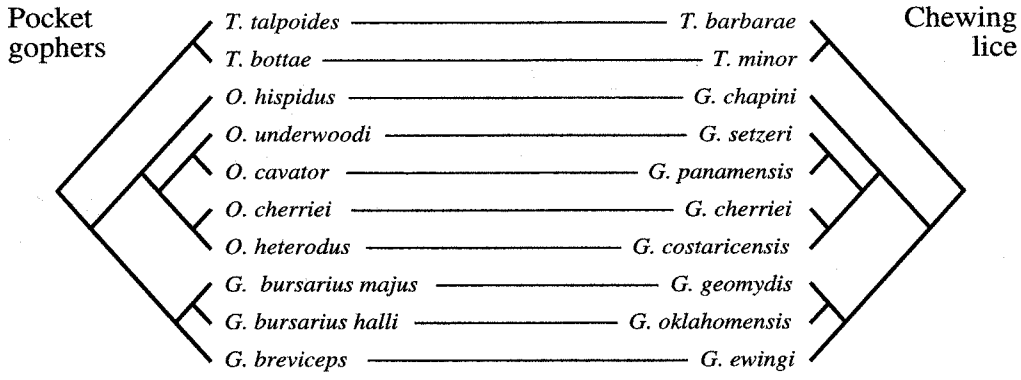

FIGURE 9. Pruned trees: The trees are now identical and display perfect coevolution for a subset of the animals.

DISCUSSION

We have described a new method designed to answer the biological question: Do the parasites tend to use hosts that occupy corresponding positions in the phylogenetic tree? The ParaFit method allows one to perform a statistical test of this particular global hypothesis of coevolution and also to test the significance of each H-P link contributing to the relationship, thereby leading to the identification of the species involved in cospeciation. Through this process, the incongruent H-P links are also identified—links that are often worth examining. Previously described methods either did not provide statistical tests (such as BPA), tested only a global fit (such as TreeMap), or tested the contributions of different kinds of events to the overall relationship (such as TreeFitter). ParaFit permits one to deal with any kind of H-P association in a reasonable amount of computing time.

Johnson et al. (2001) have recently proposed a method, involving several calculation steps, whose objective is very similar to that of ParaFit: to identify the incongruencies in the list of H-P links so as to produce a joint scenario of the cospeciation and incongruent events. They used cospeciation as the null model for statistical testing, and the test admittedly overestimated the number of congruent (i.e., cospeciation) events. In ParaFit, in contrast, the null model is the independence of the H-P associations. In hypothesis testing, identifying cospeciation events by rejecting a null hypothesis with a known (and small) type I error is better than failing to reject a null hypothesis of cospeciation with an unknown (and usually large, especially with a small sample size) type II error. The Johnson et al. method certainly takes much longer to compute than ParaFit and would be difficult to compute for moderate to large data sets.

Another advantage of ParaFit is that if, for some reason, one or the other phylogenetic tree is not available, phylogenetic distances can be used instead of trees. This is a useful feature because distances can be directly computed from raw data (morphology, sequences, and so on), without having to reconstruct a tree. The distance matrix is transformed into a rectangular matrix by principal coordinate analysis before being used in the ParaFit program. This can be interesting in the case of poorly resolved phylogenies or multiple trees, which often present a problem in this kind of studies, or if one does not want or need to estimate the phylogeny. On the other hand, when phylogenetic trees are available, they can be expressed as distance matrices by calculating patristic distances among the species (i.e., the leaves of the tree). The patristic distance matrices, transformed by principal coordinate analysis, are then used in the ParaFit tests.

The statistics ParaFitLink1 and ParaFitLink2 both have their usefulness. ParaFitLink1 has greater power for correctly detecting coevolutionary links in saturated coevolutionary models that contain additional random links, whereas ParaFitLink2 has greater power for correctly detecting coevolutionary links in unsaturated coevolutionary models, in which only a fraction of the links are coevolutionary and the other

links are random. ParaFitLink2 cannot be used in perfect coevolutionary situations because its denominator is then zero.

A FORTRAN program (PARAFIT: source code, compiled versions for Macintosh and DOS, and program documentation) to carry out the H-P coevolution test described in this paper is available on the website <http://www.fas.umontreal.ca/biol/legendre/> as well as the site of the Society for Systematic Biologists <http://www.systbiol.org>. The user's manual contains matrices **A**, **B**, and **C** used to compute the gopher–lice example discussed in this paper, as well as the patristic distance matrices that led to **B** and **C**.

ACKNOWLEDGMENTS

We are grateful to Vladimir Makarenkov, who courteously gave us a C version of the subroutine generating the unrooted random additive trees (following Pruzansky et al., 1982) that we used in our simulations, and to Mark Hafner for providing us with the DNA sequence data for pocket gophers and their lice. Interesting comments on the manuscript were provided by Roderic Page, Mark Siddall, and Adrian Paterson. This research was supported by NSERC grant no. OGP0007738 to P.L. Y.D. was supported by a scholarship from *Le Fonds pour la Formation de Chercheurs et l'Aide à la Recherche (Fonds FCAR)*, Gouvernement du Québec. E.B. was supported by a scholarship from *Le Conseil régional de Bretagne*.

REFERENCES

BARKER, S. C. 1994. Phylogeny and classification, origins, and evolution of host associations of lice. Int. J. Parasitol. 24:1285–1291.

BARRETT, J. A. 1986. Host–parasite interactions and systematics. Pages 1–17 *in* Coevolution and systematics (A. R. Stone and D. L. Hawksworth, eds.). Clarendon Press, Oxford, U.K.

BOEGER, W. A., AND D. C. KRITSKY. 1997. Coevolution of the Monogenoidea (Platyhelminthes) based on a revised hypothesis of parasite phylogeny. Int. J. Parasitol. 27:1495–1511.

BROOKS, D. R. 1977. Evolutionary history of some plagiorchioid trematodes of anurans. Syst. Zool. 26: 277–279.

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. 1985. Historical ecology: A new approach to studying the evolution of ecological associations. Ann. Mo. Bot. Gard. 72:660–680.

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. of Chicago Press, Chicago.

BUNEMAN, P. 1974. A note on the metric properties of trees. J. Comb. Theory B 17:48–50.

CHARLESTON, M. A. 1998. Jungles: A new solution to the host/parasite phylogeny reconciliation problem. Math. Biosci. 149:191–223.

DEMASTES, J. W., AND M. S. HAFNER. 1993. Cospeciation of pocket gophers (*Geomys*) and their chewing lice (*Geomydoecus*). J. Mammal. 74:521–530.

EDGINGTON, E. S. 1995. Randomization tests, 3rd edition. Marcel Dekker, New York.

EHRLICH, P. R., AND P. H. RAVEN. 1964. Butterflies and plants: A study in coevolution. Evolution 18:586–608.

FARENHOLZ, H. 1913. Ectoparasiten und Abstammungslehre. Zool. Anz. 41:371–374.

FUTUYMA, D. J., AND M. SLATKIN. 1983. Coevolution. Sinauer Associates. Sunderland, Massachusetts.

GOWER, J. C. 1966. Some distance properties of latent root and vector methods used in multivariate analysis. Biometrika 53:325–338.

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.

HAFNER, M. S, P. D. SUDMAN, F. X. VILLABLANCA, T. A. SPRADLING, J. W. DEMASTES, AND S. A. NADLER. 1994. Disparate rates of molecular evolution in cospeciating hosts and parasites. Science 265: 1087–1090.

HOPE, A. C. A. 1968. A simplified Monte Carlo test procedure. J. R. Stat. Soc. B 50:35–45.

HUELSENBECK, J. P., B. RANNALA, AND Z. YANG. 1997. Statistical tests of host–parasite cospeciation. Evolution 51:410–419.

JOHNSON, K. P., D. M. DROWN, AND D. H. CLAYTON. 2001. A data based parsimony method of cophylogenetic analysis. Zool. Scr. 30:79–87.

KLASSEN, G. J. 1992. Coevolution: A history of the macroevolutionary approach to studying host–parasite associations. J. Parasitol. 78: 573–587.

KLASSEN, G. J., AND M. BEVERLEY-BURTON. 1988. North American fresh water ancyrocephalids (Monogenea) with articulating haptoral bars: Host–parasite coevolution. Syst. Zool. 37:179–189.

LAPOINTE, F.-J., AND P. LEGENDRE. 1991. The generation of random ultrametric matrices representing dendrograms. J. Classif. 8:177–200.

LAPOINTE, F.-J., AND P. LEGENDRE. 1992. A statistical framework to test the consensus among additive trees (cladograms). Syst. Biol. 41:158–171.

LEGENDRE, P., R. GALZIN, AND M. L. HARMELIN-VIVIEN. 1997. Relating behavior to habitat: Solutions to the fourth-corner problem. Ecology 78:547–562.

LEGENDRE, P., AND L. LEGENDRE. 1998. Numerical ecology, 2nd English edition. Elsevier Science BV, Amsterdam.

MANLY, B. J. F. 1997. Randomization, bootstrap and Monte Carlo methods in biology, 2nd edition. Chapman and Hall, London.

PAGE, R. D. M. 1993a. COMPONENT 2.0. Tree comparison software for the use with Microsoft Windows. The Natural History Museum, London.

PAGE, R. D. M. 1993b. Parasites, phylogeny and cospeciation. Int. J. Parasitol. 23:499–506.

PAGE, R. D. M. 1994. Parallel phylogenies: Reconstructing the history of host–parasite assemblages. Cladistics 10:155–173.

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

PAGE, R. D. M., AND M. A. CHARLESTON. 1998. Trees within trees—Phylogeny and historical associations. Trends Ecol. Evol. 13:356–359.

PAGE, R. D. M., AND M. S. HAFNER. 1996. Molecular phylogenies and host–parasite cospeciation: Gophers and lice as a model system. Pages 255–270 *in* New uses for new phylogenies (P. H. Harvey, A. J. Leigh Brown, J. Maynard Smith, and S. Nee, eds.). Oxford Univ. Press, New York.

PAGE, R. D. M., AND E. H. HOLMES. 1998. Molecular evolution: A phylogenetic approach. Blackwell Science, Oxford, U.K.

PATERSON, A. M., AND J. B. BANKS. 2001. Analytical approaches to measuring cospeciation of host and parasites: Through a glass, darkly. Int. J. Parasitol. 31: 1012–1022.

PATERSON, A. M., R. D. GRAY, AND G. P. WALLIS. 1993. Parasites, petrels and penguins: Does louse presence reflect seabird phylogeny? Int. J. Parasitol. 23:515–526.

PRUZANSKY, S., A. TVERSKY, AND J. D. CARROLL. 1982. Spatial versus tree representations of proximity data. Psychometrika 47:3–19.

RONQUIST, F. 1995. Reconstructing the history of host–parasite associations using generalised parsimony. Cladistics 11:73–89.

RONQUIST, F. 1997. Phylogenetic approaches in coevolution and biogeography. Zool. Scr. 26:313–322.

ROY, B. A. 2001. Patterns of association between crucifers and their flower-mimic pathogens: Host jumps are more common than coevolution or cospeciation. Evolution 55:41–53.

SAITOU, N., AND M. NEI. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4:406–425.

SIDALL, M. E. 1996. Phylogenetic covariance probability: Confidence and historical associations. Syst. Biol. 45:48–66.

SOKAL, R. R., AND F. J. ROHLF. 1995. Biometry—The principles and practice of statistics in biological research, 3rd edition. W. H. Freeman, New York.

SWOFFORD, D. L. 2001. PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4b8. Sinauer Associates, Sunderland, Massachusetts.

SZIDAT, L. 1940. Beiträge zum Aubfau eines natürlichen Systems der Trematoden. I. Die Entwicklung von Echinocercaria choanophila U. Szidat zu Cathaemasia hians und die Ableitung der Fasciolidae von den Echinostomidae. Z. Parasitenk. 11:239–283.

THOMPSON, J. N. 1994. The coevolutionary process. Univ. of Chicago Press, Chicago.

Received 10 August 2001; accepted 15 November 2001
Associate Editor: Roderic D. M. Page