DE GRUYTER Stat. Appl. Genet. Mol. Biol. 2017; 16(5–6): 349–365

Colleen Nooney, Stuart Barber*, Arief Gusnanto and Walter R. Gilks

A statistical method for analysing cospeciation in tritrophic ecology using electrical circuit theory

https://doi.org/10.1515/sagmb-2016-0049

Abstract: We introduce a new method to test efficiently for cospeciation in tritrophic systems. Our method utilises an analogy with electrical circuit theory to reduce higher order systems into bitrophic data sets that retain the information of the original system. We use a sophisticated permutation scheme that weights interactions between two trophic layers based on their connection to the third layer in the system. Our method has several advantages compared to the method of Mramba et al. [Mramba, L. K., S. Barber, K. Hommola, L. A. Dyer, J. S. Wilson, M. L. Forister and W. R. Gilks (2013): "Permutation tests for analyzing cospeciation in multiple phylogenies: applications in tri-trophic ecology," Stat. Appl. Genet. Mol. Biol., 12, 679–701.]. We do not require triangular interactions to connect the three phylogenetic trees and an easily interpreted *p*-value is obtained in one step. Another advantage of our method is the scope for generalisation to higher order systems and phylogenetic networks. The performance of our method is compared to the methods of Hommola et al. [Hommola, K., J. E. Smith, Y. Qiu and W. R. Gilks (2009): "A permutation test of host–parasite cospeciation," Mol. Biol. Evol., 26, 1457–1468.] and Mramba et al. [Mramba, L. K., S. Barber, K. Hommola, L. A. Dyer, J. S. Wilson, M. L. Forister and W. R. Gilks (2013): "Permutation tests for analyzing cospeciation in multiple phylogenies: applications in tri-trophic ecology," Stat. Appl. Genet. Mol. Biol., 12, 679–701.] at the bitrophic and tritrophic level, respectively. This was achieved by evaluating type I error and statistical power. The results show that our method produces unbiased *p*-values and has comparable power overall at both trophic levels. Our method was successfully applied to a dataset of leaf-mining moths, parasitoid wasps and host plants [Lopez-Vaamonde, C., H. Godfray, S. West, C. Hansson and J. Cook (2005): "The evolution of host use and unusual reproductive strategies in achrysocharoides parasitoid wasps," J. Evol. Biol., 18, 1029–1041.], at both the bitrophic and tritrophic levels.

Keywords: cospeciation; electrical circuit; Kirchhoff matrix, tritrophic ecology; weighted permutation test.

1 Introduction

The study of host-parasite coevolution originated with the work of Von Ihering, who was the first to recognise predictable associations among hosts and their parasites (Klassen, 1992). Parasites and their hosts generally form tight ecological associations and as such it has long been assumed that the speciation of parasites is largely dependent on the speciation of their hosts (Legendre et al., 2002). However, cospeciation is not the only process that occurs, and thus host-parasite phylogenies are rarely exact mirror images. The parasite may switch lineages, speciate independently, go extinct, fail to colonise all descendants of a speciating host lineage, or fail to speciate when the host does (Page, 2003).

Figure 1 displays two simple example bitrophic systems, each consisting of Tree X, Tree Y and the interactions between their leaf nodes. We mainly focus on parasitic interactions, however other types of ecological interaction exist. These interactions may have arisen through symbiosis, mutualism, habitat or feeding relationships.

***Corresponding author: Stuart Barber,** Department of Statistics, School of Mathematics, University of Leeds, Leeds LS2 9JT, UK, e-mail: S.Barber@leeds.ac.uk
Colleen Nooney, Arief Gusnanto and Walter R. Gilks: Department of Statistics, School of Mathematics, University of Leeds, Leeds LS2 9JT, UK

There has been extensive exploration into the bitrophic interactions observed between hosts and their parasites, and between plants and specialised herbivorous insects (Forister and Feldman, 2011). As a result, many statistical tests have been developed to assess cospeciation in these systems (Page, 1996; Legendre et al., 2002; Hommola et al., 2009; Huelsenbeck et al., 2000; Mantel, 1967). However, shared evolutionary histories have been observed across more than two trophic levels (Forister and Feldman, 2011). For example, tritrophic interactions were observed between hosts, parasites and host plants (Micha et al., 2000; Ahmad et al., 2004). Recently, it was discovered that tritrophic coevolution exists between flies and parasitic nematodes on Mytaceae host plants (Nelson et al., 2014).

Mramba et al. (2013) developed the only statistical method we are aware of to test cospeciation in tritrophic systems. However, the test of Mramba et al. (2013) requires the interactions between three phylogenies to form triangles to be able to compare patristic distances on the three trees. This is often not the case in naturally occurring tritrophic systems, and thus interactions that do not form triangles are discarded along with the information they provide. We propose an improved method which can accommodate any type of interaction. Many bitrophic tests (Legendre et al., 2002; Hommola et al., 2009; Mantel, 1967) and Mramba et al.'s tritrophic test are limited to systems consisting of phylogenetic trees. Our method has the scope for generalisation to higher order systems and more generally to phylogenetic networks.

When considering the possibility of cospeciation in a tritrophic system, some care needs to be taken by exactly what one mean by 'tritrophic cospeciation'. We clarify the hypotheses being tested in Section 2.1. To draw conclusions about where cospeciation occurs within a tritrophic system, Mramba et al.'s method necessitates the permutation of every pairwise combination of three trees; that is, seven randomisations and, correspondingly, seven p-values. By contrast, our more efficient method requires the use of one sophisticated permutation scheme, which we describe in Section 2.3, resulting in one easily interpreted p-value.

Our method is a development of the correlation statistic proposed by Hommola et al. (2009). We propose a test statistic to analyse higher-order systems in Section 2.2. Our statistic, based on an analogy between phylogenetic trees and electrical circuits, applies methods from electrical circuit theory (Curtis and Morrow, 2000). We use these methods to reduce higher order systems into two vectors of "phylogenetically equivalent" distances whose correlation can be computed. These distances are "equivalent" in our electrical analogy in the sense that replacing all the tree branches and interactions by direct connections between tree tips would produce an electrical circuit that behaves in the same way as the original interacting phylogenies; Section 2.2 gives a more precise definition of this concept. These distances take into account the information contained in the connection to the third phylogenetic tree.

Our method is distinct from event-based cophylogenetic methods. These methods aim to infer evolutionary events such as cospeciation, duplication, lineage-sorting and host switching, from the phylogeny of host and parasite trees and their interaction graph. See Martínez-Aquino (2016) for a recent review.

The performance of our method is compared to the approaches proposed by Hommola et al. (2009) at the bitrophic level and Mramba et al. (2013) at the tritrophic level by means of numerical simulations in Section 3 and illustrated in the context of a real data set in Section 4.

2 Methodology

We draw an analogy between interacting phylogenetic trees and electrical circuits. We envisage an electrical circuit with junctions; correspondingly phylogenetic trees have branches, interactions, and nodes. We utilise electrical circuit theory to develop a method that can be generalised to test cospeciation hypotheses in both bitrophic and tritrophic systems.

2.1 Hypotheses

In the bitrophic case we consider two phylogenetic trees, X and Y, and the interactions between their tips. We are interested in the following hypotheses

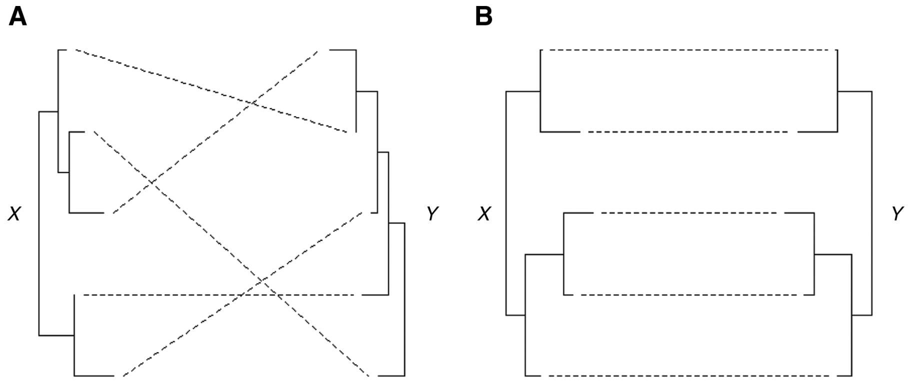

Figure 1: Randomly generated systems consistent with the bitrophic hypotheses.
The dashed lines represent the interactions between the leaf nodes of the two phylogenetic trees. (A) System generated consistent with the null hypothesis. Both trees and the interactions between them have been independently randomly generated. (B) System generated consistent with the alternative hypothesis. The trees are identical and interactions are placed at corresponding positions on the two trees.

H_0: The interaction matrix between Trees X and Y is unrelated to any cospeciation between the two trees.
H_1: The interaction matrix between Trees X and Y is related to cospeciation between the two trees.

Figure 1 displays systems generated under the extremes of the above hypotheses. The system in Figure 1A is comprised of randomly generated trees with random interactions consistent with the null hypothesis of no cospeciation. In contrast, the system in Figure 1B consists of identical trees with corresponding interactions, demonstrating the extreme of perfect cospeciation. We note that trees need not be identical to reflect a history of coevolution. Trees which are more topologically similar (congruent) than would be expected by chance also provide evidence of cospeciation. In this work, we take the trees as given and focus on the evidence of cospeciation which can be inferred from the interactions between the trees.

In the tritrophic case we consider three phylogenetic trees, X, Y and Z, and the interactions between each pair of trees. We do not simply want to know whether cospeciation exists somewhere within a tritrophic system. Rather, we are interested in how the cospeciation is driven. In particular, whether there is any "direct" cospeciation between two trees, say X and Y; that is, cospeciation which cannot be explained by the influence of the third tree Z. Therefore, we are interested in the following hypotheses

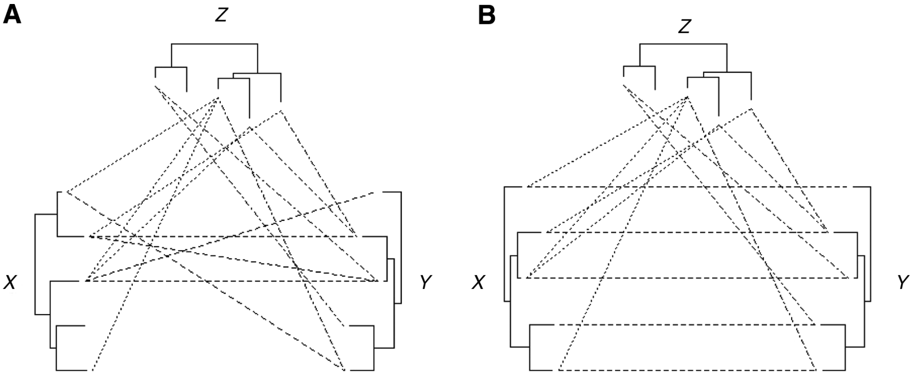

Figure 2: Randomly generated systems consistent with the tritrophic hypotheses.
The dashed lines represent the interactions between the leaf nodes of the three phylogenetic trees. (A) System generated consistent with the null hypothesis. All three trees and the interactions between them have been independently randomly generated. (B) System generated consistent with the alternative hypothesis. Two of the trees are identical with interactions placed at corresponding positions on the two trees. The third tree is independently generated and has random interactions with the other two trees.

H_0: The interaction matrix between Trees X and Y is unrelated to any direct cospeciation between Trees X and Y.

H_1: The interaction matrix between Trees X and Y is related to direct cospeciation between Trees X and Y.

Figure 2 displays systems generated under the extremes of the tritrophic hypotheses. The system in Figure 2A is comprised of three randomly generated trees with random interactions between them. Clearly, there is no cospeciation between Trees X and Y; none of the trees appear to be cospeciating on a pairwise level. Systems where Tree Z is driving the cospeciation between Trees X and Y would also be consistent with the null hypothesis. The system in Figure 2B consists of identical Trees X and Y with corresponding interactions. There is no cospeciation between these trees and Tree Z, so Tree Z does not drive the cospeciation between Trees X and Y.

2.2 Correlation statistic based on resolved distances

The methods of Hommola et al. (2009) and Mramba et al. (2013) calculate the patristic distance on each tree between each pair of interactions. Patristic distances describe the amount of genetic change that has occurred in a tree and are calculated by adding together the branch lengths that connect two leaf nodes of a phylogenetic tree (Fourment and Gibbs, 2006). Branch lengths represent an amount of evolutionary divergence and are typically a measure of distance between sequences, or obtained from a model of substitution of residues over the course of evolution (Durbin, 1998). In a bitrophic system the calculation of patristic distances is trivial. However, in a tritrophic system, there is no simple way to correlate patristic distances across all three trees. Patristic distances on the three trees can only be compared by finding pairs of interaction triangles in the system. Another situation in which patristic distances are difficult to calculate is when the system involves a phylogenetic network, as there may be more than one path between two leaf nodes.

To overcome these problems we consider electrical networks as an analogy for the network of phylogenetic trees. We apply the so-called forward problem in electrical networks, described below, to the system of phylogenetic trees to obtain phylogenetically equivalent distances between a set of carefully placed nodes. Nodes are defined to be points where two or more elements meet. In a circuit the elements are wires and in the case of a phylogenetic tree, the elements are the branches and interactions. The key analogy is that phylogenetic distance, as represented by length of branches, corresponds to electrical resistance, which is proportional to the length of a conducting wire. Electrical conductance is then the inverse of resistance and hence corresponds to evolutionary “similarity”.

Suppose we have an electrical circuit where the conductance and topology of the connections is known, as displayed in Figure 3A. Suppose we take a subset of nodes, nodes 1–4 in Figure 3A, and define these nodes to be external; all other nodes are internal. If we impose a voltage on the external nodes, we can

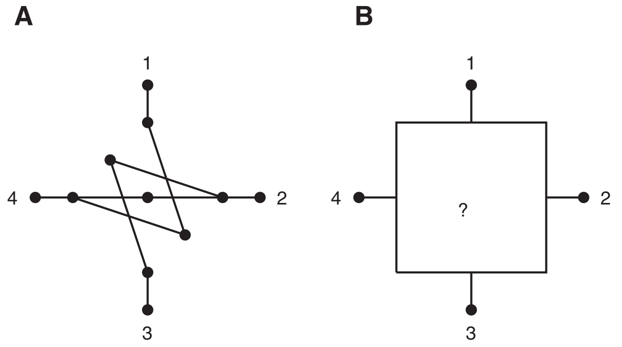

Figure 3: Schematic diagram of the forward problem in electrical networks.
(A) Example electrical circuit with nodes displayed as black circles that are connected by wires. (B) Black box containing the circuit in (A) with four nodes exposed, the wiring of the circuit inside the black box is unknown.

calculate the resulting current at these nodes. We now suppose that the circuit, excluding the external nodes is inside a black box, as displayed in Figure 3B. We no longer know how the internal nodes inside the box are connected, or the conductance on the original connections; we only have the conductances on direct connections between the external nodes. The *forward problem* assumes that we know how the circuit is connected, and the conductance on each connection. The conductance on the direct connections between the external nodes is then calculated using this information. The *inverse problem* is to obtain the full circuit from the circuit in the black box where only conductances on direct connections between the external nodes are known. The conductances of each connection in the full circuit is calculated from measurements of voltages and currents at the external nodes in the black box circuit (Curtis and Morrow, 2000).

We use the forward problem in electrical networks to calculate the conductance on direct connections between each pair of interactions between Tree *X* and Tree *Y*. In the tritrophic case these conductances will take into account how Tree *X* and Tree *Y* are connected to Tree *Z*. These conductances can then be used to calculate distances. In the bitrophic case, for each pair of interactions we will have a distance that corresponds to Tree *X*, and a distance that corresponds to Tree *Y*. In the tritrophic case these distances will take into account the connections between Trees *X* and *Y* with Tree *Z*. For a cospeciated system we expect there to be a correlation between the distance on Tree *X* and the distance on Tree *Y* associated with each pair of interactions.

To obtain direct connections between the interactions for Tree *X* and Tree *Y*, we need an external node at each end of every interaction. We introduce two artificial nodes on each interaction, dividing the interactions into three connections as displayed in Figure 4A. The artificial nodes are the external nodes and every other

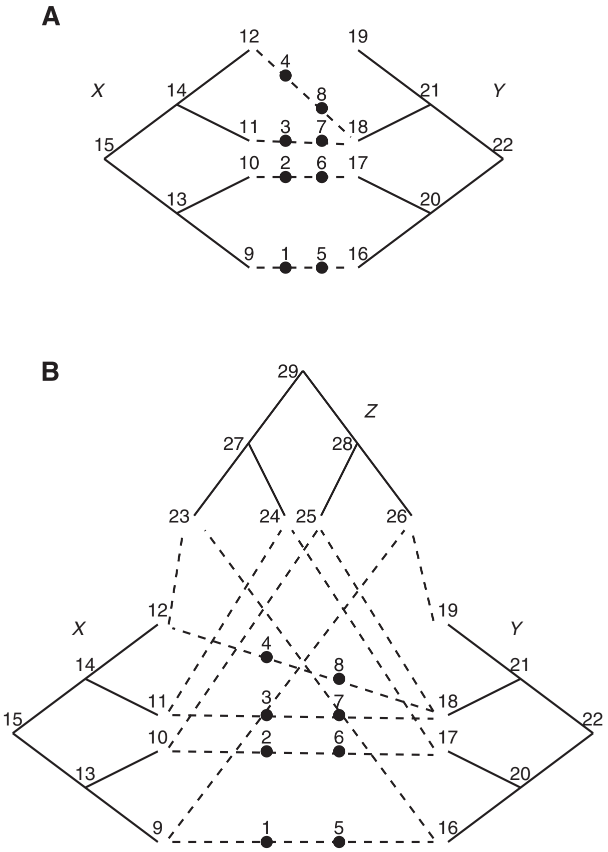

Figure 4: External node placement in (A) bitrophic and (B) tritrophic systems.
External nodes are represented by black dots. Each node in the system has been numbered. The Trees *X*, *Y* and *Z* correspond to Trees *X*, *Y* and *Z* as described in the tritrophic hypothesis.

node in the system is internal. In a tritrophic system the artificial external nodes are introduced on the interactions between Trees X and Y, as shown in Figure 4B.

Our test statistic is derived by treating the phylogenetic distances on the branches and interactions of the phylogenetic trees as electrical resistance, converting these into conductances and calculating a response matrix for the system. The conductance between nodes i and j is calculated as

$$\gamma_{i,j} = \frac{1}{d_{i,j}}, \tag{1}$$

where $d_{i,j}$ is the phylogenetic distance between nodes i and j and $\gamma_{i,j} = 0$ if nodes i and j are not directly connected by a single branch. The interactions between the phylogenetic trees do not typically have distances, therefore we assign each of the three connections that make up an interaction a constant distance, ϵ. In our analysis we chose ϵ such that the branches of the phylogenetic trees and the interactions are weighted equally. However, it may be of interest to give the branches more or less weight than the interactions. Alternatively, the interactions may be given different weights based on how strong the association is between the species in nature. The interactions can also be weighted differently to represent how likely they are to exist.

Given an interacting system of phylogenetic trees consisting of m nodes in total, we employ electrical circuit theory via the construction of an $m \times m$ Kirchhoff matrix, K, which has the following interpretation. If u is defined to be a vector of voltages applied to each node of the network, then $\phi = Ku$ is the resulting vector of current flowing through the network at each node. If a voltage of one unit is applied to node j and a voltage of zero is applied to every other node, then $K_{i,j}$ is the current in the network at each node i. Thus column j of K gives the values of the currents in the network at nodes $i = 1, \ldots, m$. The Kirchhoff matrix is a Laplacian matrix, assembled using the conductances between nodes connected by a single branch. The $(i, j)^{\text{th}}$ element of K is

$$k_{i,j} = \begin{cases} -\gamma_{i,j} & \text{if } i \neq j \\ \sum_{j \neq i} \gamma_{i,j} & \text{if } i = j. \end{cases} \tag{2}$$

Rearranging the Kirchhoff matrix in terms of the internal and external nodes of the system, where the external nodes are the nodes on the interactions and all of the tree nodes are internal, partitions the matrix as

$$K = \begin{array}{cc} & \begin{array}{cc} E & I \end{array} \\ \begin{array}{c} E \\ I \end{array} & \begin{pmatrix} A & B \\ B^T & D \end{pmatrix} \end{array},$$

where E and I correspond to the external and internal nodes, respectively.

A response matrix is simply a Kirchhoff matrix calculated for an equivalent system without internal nodes, and only direct connections between the external nodes. Therefore the response matrix contains the negative conductance on each pairwise connection between the external nodes. The response matrix, Λ_γ, is obtained by calculating the Schur complement in K of the square submatrix D, corresponding to the internal nodes of the network:

$$\Lambda_\gamma = A - BD^{-1}B^T.$$

The distances between the external nodes in the collapsed system are obtained by reversing Equations (1) and (2). We define $D^\star$ to be the resulting distance matrix, with $(i, j)^{\text{th}}$ element given by:

$$d^\star_{i,j} = \begin{cases} -\frac{1}{(\Lambda_\gamma)_{i,j}} & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases}$$

where $(\Lambda_\gamma)_{i,j}$ is the $(i, j)^{\text{th}}$ element of Λ_γ.

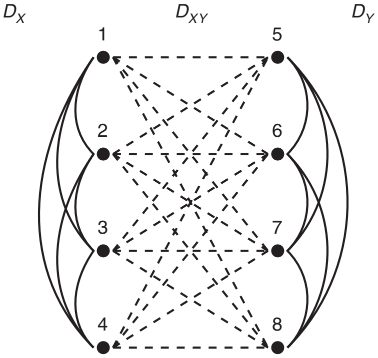

Figure 5: Connections contained in D^* for the systems displayed in Figure 4.
The external nodes are represented by black dots and numbered consistently with Figure 4. The internal nodes have been integrated out by the response matrix calculations.

The distance matrix can be partitioned in terms of the external nodes corresponding to Tree X; E_X, and the external nodes corresponding to Tree Y; E_Y, as follows:

$$D^* = \begin{array}{cc} & \begin{array}{cc} E_X & E_Y \end{array} \\ \begin{array}{c} E_X \\ E_Y \end{array} & \begin{pmatrix} D_X & D_{XY} \\ D_{XY}^T & D_Y \end{pmatrix} \end{array},$$

where D_X and D_Y are submatrices containing the distances between each pair of external nodes corresponding to Tree X and Tree Y, respectively. D_{XY} is a submatrix containing the distances between Tree X and Tree Y. In the tritrophic case, these distances will also take into account the connection with Tree Z; in higher-order systems, the distances in D_{XY} will take into account the connections with and between all trees other than X and Y. Figure 5 displays the connections corresponding to the distances contained in D^* for the systems in Figure 4.

Our statistic is obtained by calculating the Spearman's correlation coefficient, r_{obs}, between the upper triangle of D_X and D_Y. We use a rank correlation because the response matrix calculations produce large distances when there are extreme interactions between the trees.

We propose a permutation approach to determine whether the value of r_{obs} is statistically significant. A p-value, p, is obtained for r_{obs} by simulating N systems under H_0 as described in Section 2.3, then calculating

$$p = \frac{1}{N}\sum_{i=1}^{N} I(r_i \geq r_{\text{obs}}),$$

where r_i is the test statistic calculated for the i-th randomisation and $I(r_i \geq r_{\text{obs}})$ is an indicator function taking the value 1 if r_i is greater than or equal to r_{obs} and 0 otherwise. If $p \leq \alpha$ we reject H_0 at the $100\alpha\%$ significance level.

2.3 Permutations

To determine whether a value of r_{obs} is statistically significant, we require a permutation scheme that simulates compatible systems under our null hypotheses.

In a bitrophic system the connections between the external nodes are sampled with equal probability. Permutations of the connection between the external nodes that result in overlapping interactions are rejected. This is equivalent to simply randomising the existing connections between the external nodes.

Randomising in this way preserves the many to one nature of the interactions, however not all of the interactions between the two trees are possible due to the placement of the external nodes on the interactions. That is, nodes on the trees without interactions are essentially removed.

The response matrix for the system of phylogenetic trees is simply a Kirchhoff matrix calculated only for the external nodes of the phylogenetically equivalent system with the internal nodes integrated out. Therefore the response matrix infers a connection between each pair of external nodes with different conductivities based on the original connections between the trees. The conductance (analogous to evolutionary similarity) on these connections are used as weights to sample the connections between the external nodes that connect Trees X and Y. Connections consistent with H_0 have a greater probability of being sampled. To obtain these weights we recalculate the response matrix for the system with the middle connections between the external nodes removed. To randomise the tritrophic system consistent with the null hypothesis, we sample connections between the external nodes with probability proportional to their conductance in the recalculated response matrix. Note that as we do not alter the internal nodes of the trees during randomisation, our test is possibly sensitive to their shape.

Two considerations must be taken into account when sampling the connections. Firstly, the connections must be sampled such that many to one interactions between two external nodes are avoided; this would correspond to a system where there are interactions between the interactions. Secondly, permutations involving overlapping interactions are rejected, as in the bitrophic case.

3 Results

The performance of our method, at the bitrophic and tritrophic level, is analysed by investigating Type I error and assessing statistical power (see below). We compared the performance of our method to those proposed by Hommola et al. (2009) and Mramba et al. (2013) at the relevant trophic level. In every simulation we set $\epsilon = 0.5$, the average branch length of the simulated trees.

3.1 Type I error

Type I error arises as a result of incorrectly rejecting the null hypothesis when it is true. The probability of this is called the significance level, α, of the test. Type I error is estimated by simulating data under the null hypothesis. The rate of rejection of the null hypothesis for data simulated under it should be equal to α. We expect the p-values of data generated under H_0 to be uniformly distributed if the statistic is reliable. Therefore we expect a plot of the empirical cumulative distribution function (CDF) to be a straight diagonal line.

For both the bitrophic and tritrophic hypothesis, this corresponds to independently generating random phylogenetic trees with randomly assigned interactions (see Section 2.1 for the bitrophic hypothesis). The trees were generated using the `rtree` function of the R (R Core Team, 2013) package *ape* (Paradis et al., 2004). In the bitrophic case we used the same parameter combinations as Hommola et al. (2009) and Legendre et al. (2002):

- 10 tips on Tree X, 10 tips on Tree Y and 10, 15, 20, and 25 interactions;
- 10 tips on Tree X, 15 tips on Tree Y and 10, 15, 20, and 25 interactions.

For each parameter combination, 1000 systems were generated. We calculated p-values with $N = 10{,}000$ randomisations for each system using our method and the correlation method proposed by Hommola et al. (2009). The results for the first parameter combination, with 10 and 15 interactions, are displayed in Figure 6. The remaining plots for the first parameter combination are in the Supplementary Material (Figure S1), results for the second parameter combination are not shown.

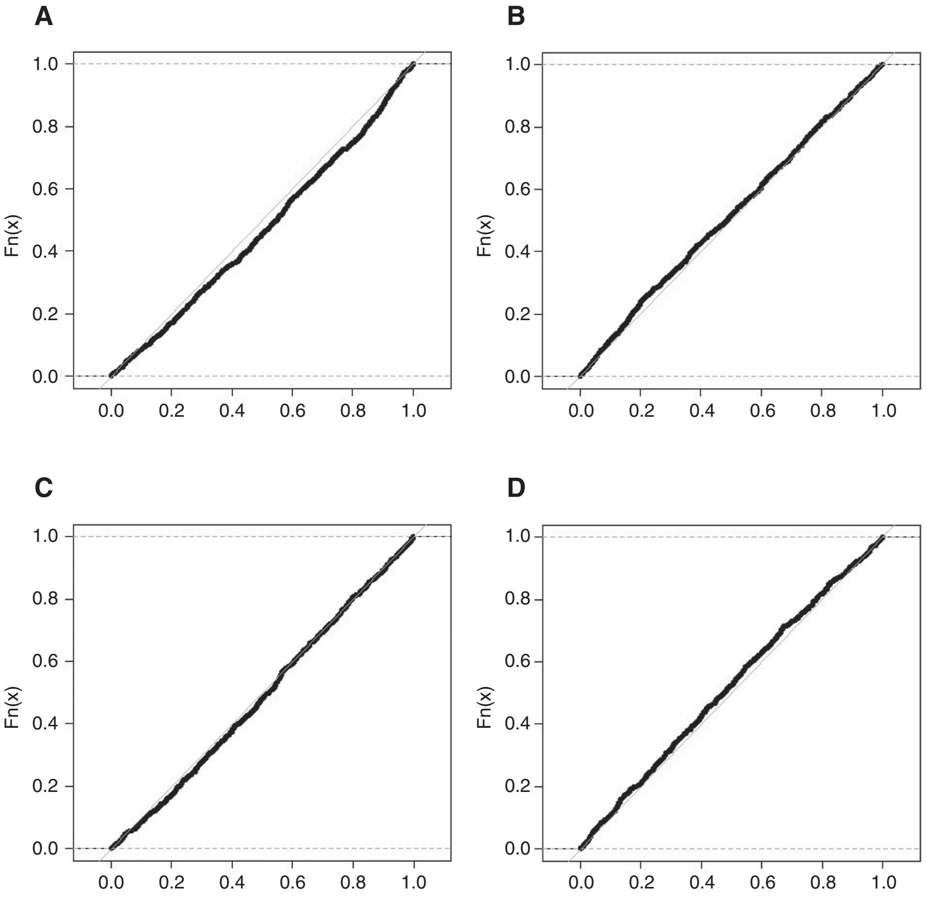

Figure 6: Empirical cumulative distribution functions for our *p*-values and Hommola et al. (2009).
Each plot corresponds to simulations with 10 tips on each tree. The first column corresponds to 10 interactions simulated and the second column corresponds to 15 interactions simulated. The top row contains the *p*-values for our method, and the bottom row contains the *p*-values for the method of Hommola et al. (2009).

For the tritrophic case we used the same parameter combinations as Mramba et al. (2013), with and without triangular interaction constraints:

- 10 tips on Tree X, 10 tips on Tree Y, 10 tips on Tree Z and 10, 15, 20, and 25 interactions between each pair of trees;
- 10 tips on Tree X, 10 tips on Tree Y, 15 tips on Tree Z and 10, 15, 20, and 25 interactions between each pair of trees.

For each parameter combination, 1000 systems were generated. We calculated *p*-values with $N = 1000$ randomisations for each system using our method and the method of Mramba et al. (2013). The results of our method, for the first parameter combination, with triangular interactions, are displayed in Figure 7, the results for the second parameter combination, with triangular interactions, are in the Supplementary Material (Figure S2).

The empirical CDF for our *p*-values lies close to the desired diagonal line for all parameter combinations in the bitrophic and tritrophic cases. The same is true of the methods of Hommola et al. (2009) and Mramba et al. (2013). However, when applied to datasets where there are no constraints on the interactions, Mramba et al. (2013) *p*-values are biased for systems where there are fewer interactions. For the parameter combinations with 10 interactions, 95% and 97% of the simulated systems could not be used to calculate *p*-values as their interactions did not form triangles, as required by that method. In the case of the parameter combinations with 15 interactions, 43% and 65% of the systems could not be used.

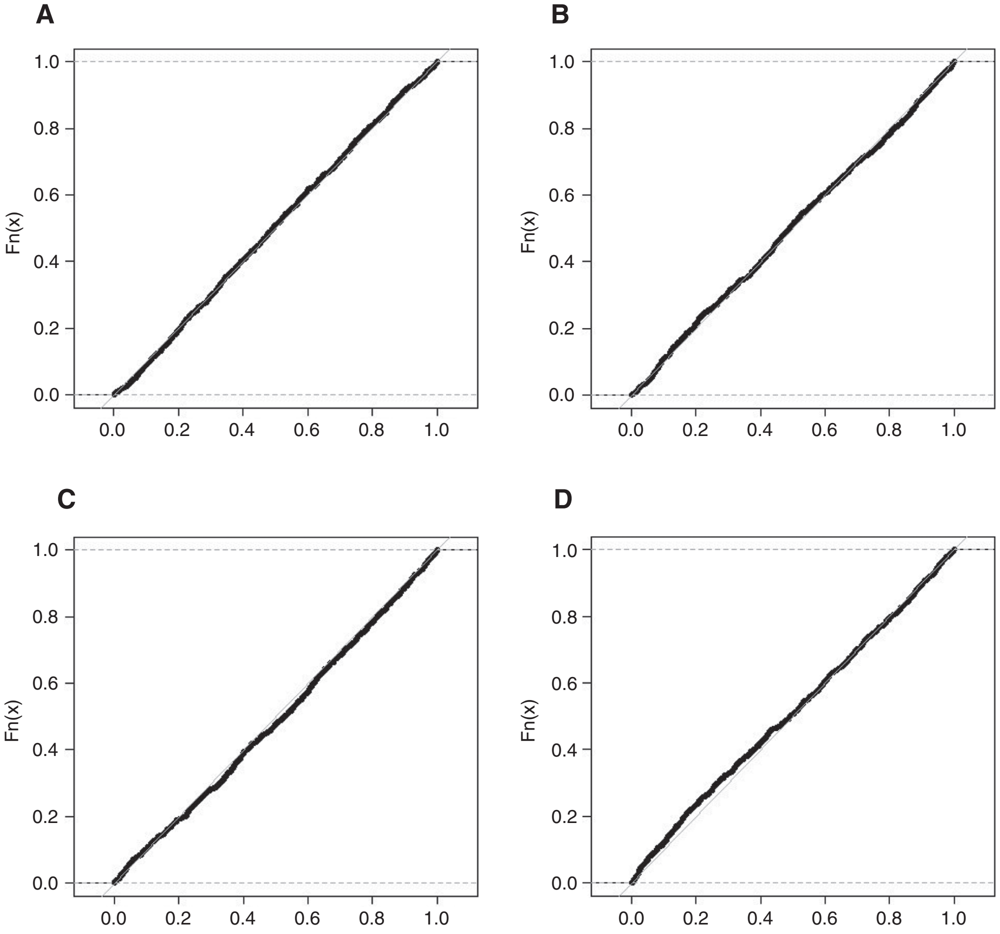

Figure 7: Empirical cumulative distribution functions for our tritrophic p-values.
Each plot corresponds to simulations with 10 tips on each tree. Each plot represents a different number of interactions simulated. From top left to bottom right, 10, 15, 20 and 25 interactions.

3.2 Bitrophic power simulations

Statistical power is the probability that the null hypothesis is correctly rejected when it is false. Statistical power has been assessed for our method as well as the correlation statistic proposed by Hommola et al. (2009) for the bitrophic case. We followed the simulation approaches adapted by Hommola et al. (2009) and Legendre et al. (2002) to generate data consistent with H_1. Noise is gradually added using the following three methods, and the proportion of correct rejections of the null hypothesis calculated in each case. In every simulation approach 1000 systems were generated. We calculated p-values with $N = 10{,}000$ randomisations for each system.

3.2.1 Simulation method 1: replacing interactions

For each simulation, Tree X and Tree Y were assigned the same randomly generated phylogenetic tree with interactions initially assigned at corresponding positions on the tree. The interactions connect each leaf node on Tree X with the same leaf node on the identical Tree Y, such that they exhibit perfect cospeciation. A percentage, 10%–50%, of these interactions are then replaced with random, non-corresponding, interactions. We used the following parameter combinations:

- 10 tips on Tree X, 10 tips on Tree Y, 10 corresponding interactions, replacing 1, 2, 3, 4, and 5 random interactions
- 20 tips on Tree X, 20 tips on Tree Y, 20 corresponding interactions, replacing 2, 4, 6, 8, and 10 random interactions

3.2.2 Simulation method 2: adding interactions

As for simulation method 1, Tree X and Tree Y were assigned the same phylogenetic tree and interactions assigned at corresponding positions on the tree. A number of random interactions were then added. This simulation approach was performed for the same parameter combinations as for simulation method 1.

3.2.3 Simulation method 3: randomise clade branch lengths

We now consider the branch lengths of the phylogenies as well as the interactions. A random base tree was generated and the branch lengths randomised to produce Tree X and Tree Y. The branch lengths were randomised by randomly sampling new branch lengths from the standard uniform distribution. The clades were chosen for randomisation based on their distance from the root node; the clades furthest from the root nodes were randomised first.

- 10 tips on Tree X, 10 tips on Tree Y, and branch lengths randomised in 1, 2, 3, 4, and 5 clades.
- 20 tips on Tree X, 20 tips on Tree Y, and branch lengths randomised in 2, 3, 4, 5 and 6 clades.

For each simulation approach, we calculated the rejection rate of the null hypothesis at the $\alpha = 0.05$ and $\alpha = 0.01$ significance levels. The rejection rate is calculated as the proportion of times that we reject the null hypothesis. Selected rejection rate plots are displayed in Figure 8. Rejection rate plots for simulation method 3 are in the Supplementary Material (Figure S3). The rejection rates increase as the systems become more cospeciated. For each of the simulation approaches, rejection rates are higher for systems with 20 tips than systems with 10 tip trees. It is also clear that the rejection rates are higher for simulation method 2 than the other simulation approaches. For each simulation method and tree size considered, our rejection rate is comparable to that of Hommola et al. (2009). We obtain similar results at the $\alpha = 0.01$ significance level.

3.3 Tritrophic power simulations

Statistical power has been assessed for our method at the tritrophic level and we have also compared our method to the permutation test proposed by Mramba et al. (2013). We followed the simulation approaches adapted by Mramba et al. (2013), and repeated these without forcing the interactions to form triangles between the three trees. In every simulation approach 100 systems were generated. We calculated p-values with $N = 10{,}000$ randomisations for each system.

3.3.1 Simulation method 1: replacing interactions

Trees X and Y were assigned the same randomly generated phylogenetic tree. To avoid computational issues with Mramba et al. (2013) method independent $N(0, 0.01^2)$ noise was added to the branch lengths, as described in Mramba et al. (2013). Interactions were initially assigned at corresponding positions between the trees, such that Tree X and Tree Y exhibit perfect cospeciation. Tree Z is unrelated to Trees X and Y, and is therefore independently generated with randomly assigned interactions between itself and Trees X and Y. The interactions between each pair of trees are then replaced with random interactions. We used the following parameter combinations:

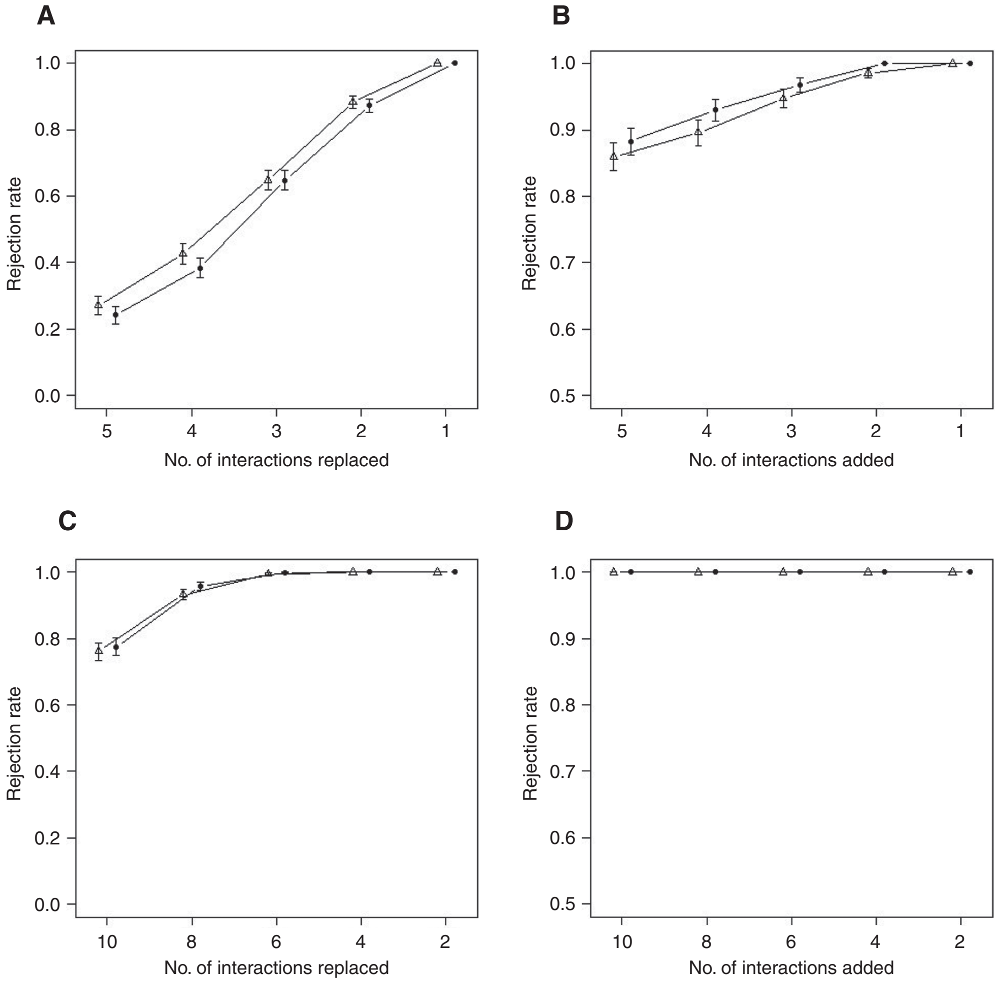

Figure 8: Rejection rates for the *p*-values generated using our method and the method of Hommola et al. (2009) at the $\alpha = 0.05$ significance level, under different simulation approaches.

Black dots are the rates obtained using our method and triangles are the rates calculated for Hommola et al. (2009) *p*-values. The points are offset on the horizontal axis to prevent overlapping. Each column corresponds to a different simulation approach. The first column corresponds to simulation method 1 and the second column corresponds to simulation method 2. The top row contains the 10 tip simulations for each approach. The bottom row contains the 20 tip simulations for each approach.

- 10 tips on Trees X, Y and Z, 10 interactions between each pair of trees, and 1, 2, …, 10 interactions replaced between each pair of trees.
- 20 tips on Trees X, Y and Z, 20 interactions between each pair of trees, and 2, 4, …, 20 interactions replaced between each pair of trees.

3.3.2 Simulation method 2: adding interactions

Again, Trees X and Y have the same phylogenetic tree with interactions assigned at corresponding positions. Tree Z is independently generated with random interactions between itself and Trees X and Y. In this approach, interactions were randomly added between each pair of trees. The same parameter combinations were used as in the previous simulation approach.

Our method can only be compared to Mramba et al. (2013) when the interactions between the three trees are forced to form triangles. The above simulation approaches are performed with and without triangular interaction constraints. Selected plots of the rejection rates are displayed in Figures 9 and 10.

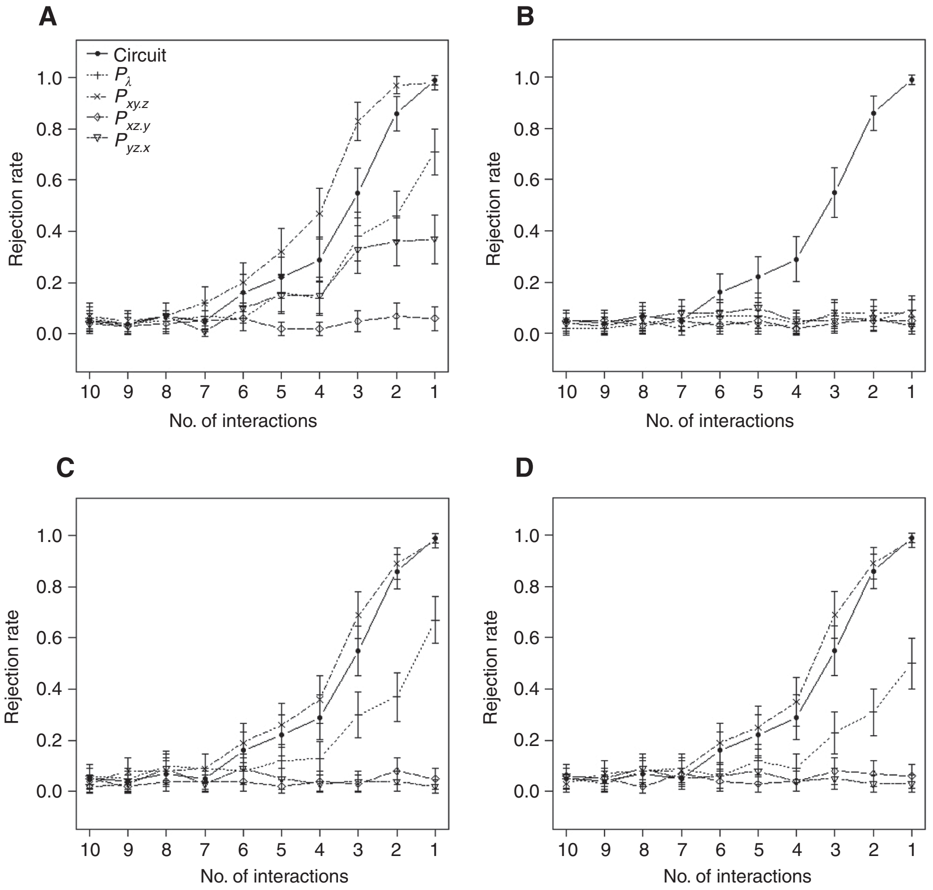

Figure 9: Rejection rates for *p*-values generated using our method and the method of Mramba et al. (2013) at the $\alpha = 0.05$ significance level, under the simulation approach where triangular interactions are replaced between three 10 tip trees. The interactions between the three trees are forced to form triangles. The horizontal axis corresponds to the number of interactions replaced between each pair of trees. Black dots are the rates obtained using our method, labelled "Circuit", and the other lines correspond to the rates calculated for the different *p*-values obtained under Mramba et al. (2013) method; P_λ, $P_{xy.z}$, $P_{xz.y}$ and $P_{yz.x}$. (A) Only *X* randomized; (B) Only *Z* randomized; (C) *X* and *Y* randomized; and (D) *X*, *Y* and *Z* randomized.

The method of Mramba et al. (2013) requires the permutation of every combination of trees, and four different *p*-values to make conclusions about cospeciation in a tritrophic system. A simple interpretation guide for the relationship between the possible permutations and the *p*-values is given in Table 1. Figure 9 displays the rejection rates for our *p*-values and Mramba et al.'s method four different *p*-values for the simulation approach where we replace triangles of interactions with random triangles of interactions. The rejection rates are calculated at the $\alpha = 0.05$ significance level. Each plot corresponds to a different randomisation in Mramba et al.'s method. The power curve for our method is repeated in each plot. Figure 9A–D correspond to the cases where only Tree *X* is randomised, only Tree *Z* is randomised, both Trees *X* and *Y* are randomised, and all three trees are randomised, respectively.

By construction, Tree *Z* is not involved in the cospeciation between Trees *X* and *Y*, thus permuting Tree *Z* reveals no effect of cospeciation. This can be seen in Figure 9B where, as expected, the rejection rates for Mramba et al.'s method are all very low. From Table 1, a significant value of $P_{xy.z}$ when Trees *X* and *Y* are involved in the randomisation indicates that there is cospeciation between Trees *X* and *Y*. This can clearly be seen in Figure 9A, C and D where the statistic corresponding to $P_{xy.z}$ is the most powerful. The statistics corresponding to $P_{xz.y}$ and $P_{yz.x}$ are less powerful because Trees *X* and *Y* are not cospeciating with Tree *Z*, and randomising Tree *X* tells us nothing about the cospeciation between Trees *Y* and *Z*. Our statistic has less power than $P_{xy.z}$ under randomisations involving Tree *X*, most clearly where only Tree *X* is randomised.

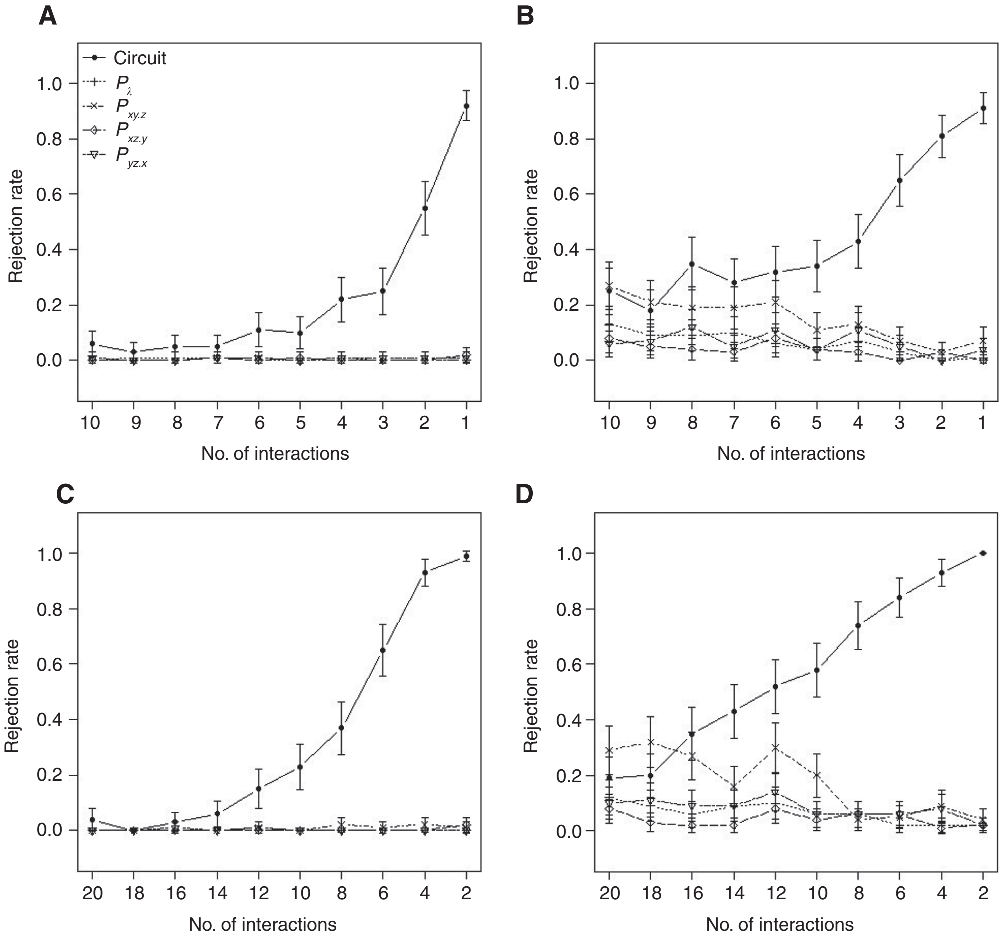

Figure 10: Rejection rates for p-values generated using our method and the method of Mramba et al. (2013) at the $\alpha = 0.05$ significance level, under different simulation approaches.
Each column corresponds to a different simulation approach; replacing and adding interactions between the three trees, respectively. The horizontal axis corresponds to the number of interactions replaced or added between each pair of trees.
In each simulation the interactions are not forced to form triangles. The rows correspond to the tree sizes. The first row contains the 10 tip simulations for each approach. The second row contains the 20 tip simulations for each approach. Each plot corresponds to the case where only Tree X is randomised for Mramba et al. (2013) method. Black dots are the rates obtained using our method, labelled “Circuit”, and the other lines correspond to the rates calculated for the different p-values obtained under Mramba et al. (2013) method; P_λ, $P_{xy.z}$, $P_{xz.y}$ and $P_{yz.x}$.

Table 1: Basic interpretation of the interaction between the possible permutations of the tritrophic system and the p-values of the method of Mramba et al. (2013).

Permutation	**P_λ significant**	**$P_{xy.z}$ significant**	**$P_{xz.y}$ significant**	**$P_{yz.x}$ significant**
X	*X* involved in cospeciation	*X* and *Y* cospeciate	*X* and *Z* cospeciate	–
Y	*Y* involved in cospeciation	*X* and *Y* cospeciate	–	*Y* and *Z* cospeciate
Z	*Z* involved in cospeciation	–	*X* and *Z* cospeciate	*Y* and *Z* cospeciate
XY	Cospeciation occurs somewhere in the system			
XZ				
YZ				
XYZ				

However, in natural systems there is no restriction that the interactions form triangles between the three phylogenetic trees. Figure 10 displays the rejection rates, calculated at the $\alpha = 0.05$ significance level, for our method and Mramba et al. (2013) for simulations with interactions that are not constrained to form triangles.

We show only one of Mramba et al. (2013) randomisations, the case where only Tree X is randomised; other plots display very similar results. Clearly our statistic is more powerful than the method of Mramba et al. (2013). Similar results were obtained at the $\alpha = 0.01$ significance level.

To calculate their p-values, the method of Mramba et al. (2013) must discard any interactions that do not form triangles. On average at least 60% of the interactions were discarded in every simulation approach; in most of these simulations over 80% of the interactions were discarded on average. Mramba et al. (2013) p-values cannot be calculated unless there are at least three triangles. Any p-values that cannot be calculated are not included in the calculation of the rejection rate. Therefore many of the rejection rates calculated for the method of Mramba et al. (2013) are calculated based on only a fraction of the systems simulated. If none of the p-values can be calculated then the rejection rate is zero.

4 Application to real data

We applied our method to a tritrophic dataset consisting of host plants, leaf-mining moths and parasitoid wasps (Lopez-Vaamonde et al., 2005). The statistic does not depend on the centre of the interactions between Trees X and Y. It only depends on the pieces of the interactions that are connected to Trees X and Y. We set the value of ϵ to be the average of all the branch distances on the tree it is connected to. We used the reconstructed phylogenetic trees calculated by Mramba et al. (2013). The three phylogenies and their interactions are displayed in Figure 11; there are 16 extant taxa in the host plant phylogeny, 28 for the moths, and 15 for the wasps. These phylogenies have 29 moth-host plant interactions, 37 moth-wasp interactions, and 24 host plant-wasp interactions, forming a total of 32 triangular interactions. The interactions do not all form the triangles that are necessary for Mramba et al. (2013) method; in fact 12 interactions had to be discarded.

Lopez-Vaamonde et al. (2005) found no evidence that the host plant, leaf-mining moth or parasitoid wasp exhibit cospeciation at a pairwise level. By contrast, Mramba et al. (2013) found mixed evidence for cospeciation but conclude that the parasitoid wasp has been central in the cospeciation of the tritrophic system. p-Values for Mramba et al. (2013) method are given in the Supplementary Material (Table S1). Our results coincide with those of Mramba et al. (2013). We obtain a p-value of 0.441, indicating that there is evidence that any cospeciation between the moth and host plant is due to the parasitoid wasp driving the cospeciation in the tritrophic system.

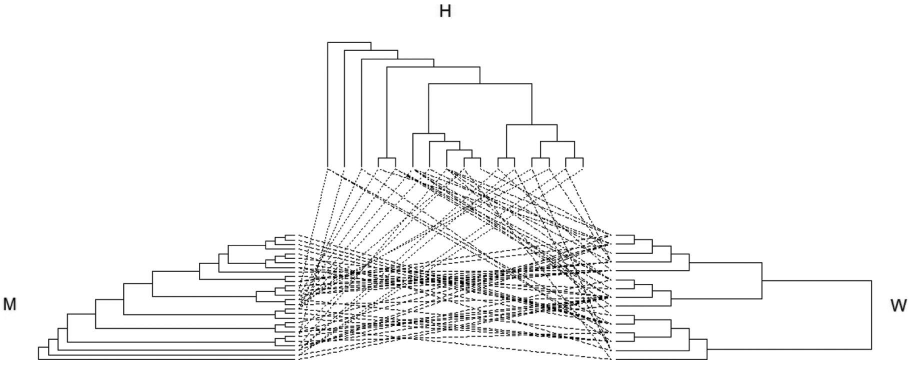

Figure 11: Tritrophic system consisting of host plants (H), leaf-mining moths (M) and parasitoid wasps (W) (Lopez-Vaamonde et al., 2005).
The phylogenetic trees were reconstructed by Mramba et al. (2013). Branch lengths have not been used to plot the trees. Plots of the phylogenetic trees using the branch lengths are given in the Supplementary Material (Figure S4). The dashed lines display the interactions between the leaf nodes of the three trees.

5 Discussion

We have introduced a method that efficiently tests cospeciation hypotheses in interacting phylogenetic systems. This method conditions on the phylogenetic trees and measures the evidence of cospeciation which is encoded in the interactions between the trees. Our method is effective at the bitrophic level. We observe unbiased *p*-values when assessing type I error has similar power to that of Hommola et al. (2009).

We use one sophisticated permutation scheme based on weighted interactions to test our hypothesis. This is an improvement on the multiple permutation scheme required by the tritrophic method proposed by Mramba et al. (2013), which requires the interactions to form sets of triangles, we do not require specific interaction patterns to be formed between the three phylogenies to calculate our statistic or to perform the randomisations. As a result no information is discarded with our method, and we obtain unbiased *p*-values. Discarding interactions results in biased *p*-values for the method of Mramba et al. (2013).

Statistical power for each method was evaluated by simulating data under the alternative hypothesis. Our method out performed Mramba et al. (2013) in all cases where the interactions were not constrained to form triangles, even when noise was introduced to the data.

We successfully applied our method to a tritrophic dataset of host plants, leaf-mining moths and parasitoid wasps. Our conclusions support those of Mramba et al. (2013).

Due to the calculation of the direct distances between the external nodes, our method is not restricted to phylogenetic trees; it can still be applied when the system involves phylogenetic networks. It is also easily generalised to higher order systems. In each case, the added complexity of the system is resolved by the reduction to the phylogenetically equivalent set of distances in the response matrix; this will work regardless of how large a set of internal nodes is being reduced. A larger set of internal nodes could accommodate more trophic levels or more general phylogenetic networks in place of trees.

Existing methods use a binary system to determine whether or not an association exists between two species. Our method allows the interactions to be weighted according to the user's criteria. For example, there may a degree of uncertainty surrounding the likelihood of an association existing.

Our method has been implemented using R (R Core Team, 2013) and the source code is available from: http://www.maths.leeds.ac.uk/~stuart/research

Acknowledgment: CN was supported by an EPSRC Doctoral Training Grant at the School of Mathematics, University of Leeds.

References

Ahmad, F., M. Aslam and M. Razaq (2004): "Chemical ecology of insects and tritrophic interactions," J. Research (Science), 15, 181–190.

Curtis, E. B. and J. A. Morrow (2000): Inverse problems for electrical networks, World Scientific, Singapore.

Durbin, R. (1998): Biological sequence analysis: probabilistic models of proteins and nucleic acids, Cambridge University Press, Cambridge.

Forister, M. L. and C. R. Feldman (2011): "Phylogenetic cascades and the origins of tropical diversity," Biotropica, 43, 270–278.

Fourment, M. and M. J. Gibbs (2006): "Patristic: a program for calculating patristic distances and graphically comparing the components of genetic change," BMC Evol. Biol., 6, 1.

Hommola, K., J. E. Smith, Y. Qiu and W. R. Gilks (2009): "A permutation test of host–parasite cospeciation," Mol. Biol. Evol., 26, 1457–1468.

Huelsenbeck, J. P., B. Rannala and B. Larget (2000): "A bayesian framework for the analysis of cospeciation," Evolution, 54, 352–364.

Klassen, G. J. (1992): "Coevolution: a history of the macroevolutionary approach to studying host-parasite associations," J. Parasitol., 78, 573–587.

Legendre, P., Y. Desdevises and E. Bazin (2002): "A statistical test for host–parasite coevolution," Syst. Biol., 51, 217–234.

Lopez-Vaamonde, C., H. Godfray, S. West, C. Hansson and J. Cook (2005): "The evolution of host use and unusual reproductive strategies in achrysocharoides parasitoid wasps," J. Evol. Biol., 18, 1029–1041.

Mantel, N. (1967): "The detection of disease clustering and a generalized regression approach," Cancer Res., 27(2 Part 1), 209–220.

Martínez-Aquino, A. (2016): “Phylogenetic framework for coevolutionary studies: a compass for exploring jungles of tangled trees,” Curr. Zoo., 62, 393–403.

Micha, S. G., S. Kistenmacher, G. Mölck and U. Wyss (2000): “Tritrophic interactions between cereals, aphids and parasitoids: discrimination of different plant-host complexes by *Aphidius rhopalosiphi* (Hymenoptera: Aphidiidae),” Eur. J. Entomol., 97, 539–544.

Mramba, L. K., S. Barber, K. Hommola, L. A. Dyer, J. S. Wilson, M. L. Forister and W. R. Gilks (2013): “Permutation tests for analyzing cospeciation in multiple phylogenies: applications in tri-trophic ecology,” Stat. Appl. Genet. Mol. Biol., 12, 679–701.

Nelson, L. A., K. A. Davies, S. J. Scheffer, G. S. Taylor, M. F. Purcell, R. M. Giblin-Davis, A. H. Thornhill and D. K. Yeates (2014): “An emerging example of tritrophic coevolution between flies (diptera: Fergusoninidae) and nematodes (nematoda: Neotylenchidae) on myrtaceae host plants,” Biol. J. Linnean Soc., 111, 699–718.

Page, R. D. (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. (2003): Tangled trees: Phylogeny, cospeciation, and coevolution, University of Chicago Press, Chicago.

Paradis, E., J. Claude and K. Strimmer (2004): “Ape: analyses of phylogenetics and evolution in r language,” Bioinformatics, 20, 289–290.

R Core Team (2013): R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Supplemental Material: The online version of this article offers supplementary material (https://doi.org/10.1515/sagmb-2016-0049).