Version dated: September 19, 2017

RH: PHYLOGENETIC COMPARATIVE METHODS ON NETWORKS

Phylogenetic comparative methods on phylogenetic networks with reticulations

Paul Bastide[1,2], Claudia Solís-Lemus[3], Ricardo Kriebel[4], K. William Sparks[4], Cécile Ané [3,4]

[1] *UMR MIA-Paris, AgroParisTech, INRA, Université Paris-Saclay, 75005, Paris, France*

[2] *MaIAGE, INRA, Université Paris-Saclay, 78352 Jouy-en-Josas, France*

[3] *Department of Statistics, University of Wisconsin-Madison, WI, 53706, USA*

[4] *Department of Botany, University of Wisconsin-Madison, WI, 53706, USA*

Corresponding author: Paul Bastide, UMR MIA-Paris, AgroParisTech, INRA, Université Paris-Saclay, 16 rue Claude Bernard, 75005, Paris, France; E-mail: paul.bastide@m4x.org.

Abstract.— The goal of Phylogenetic Comparative Methods (PCMs) is to study the distribution of quantitative traits among related species. The observed traits are often seen as the result of a Brownian Motion (BM) running along a phylogenetic tree. Reticulation events such as hybridization, gene flow or horizontal gene transfer, can substantially affect a species' traits, but are not modeled by a tree. *Phylogenetic networks* have been designed to represent reticulate evolution. As they become available for downstream analyses, new models of trait evolution are needed, applicable to networks. One natural extension of the BM is to use a weighted average model for the trait of a hybrid, at a reticulation point. We develop here an efficient recursive algorithm to compute the phylogenetic variance matrix of a trait on a network, in only one preorder traversal of the network. We then extend the standard PCM tools to this new framework, including phylogenetic regression with covariates (or phylogenetic ANOVA), ancestral trait reconstruction, and Pagel's λ test of phylogenetic signal. The trait of a hybrid is sometimes outside of the range of its two parents'. Hybrid vigor and hybrid depression is indeed a rather common phenomenon observed in present-day hybrids. Transgressive evolution can be modeled as a shift in the trait value following a reticulation point. We develop a general framework to handle such shifts, and take advantage of the phylogenetic regression view of the problem to design statistical tests for ancestral transgressive evolution in the evolutionary history of a group of species. We study the power of these tests in several scenarios, and show that recent events have indeed the strongest impact on the trait distribution of present-day taxa. We apply those methods to a dataset of *Xiphophorus* fishes, to confirm and complete previous analysis in this group. All the methods developed here are available in the user-friendly julia package PhyloNetworks.

(Keywords: Phylogenetic Networks, Phylogenetic Comparative Methods, Transgressive Evolution, Heterosis, PhyloNetworks)

40

41 The evolutionary history of species is known to shape the present-day distribution
42 of observed characters (Felsenstein 1985). Phylogenetic Comparative Methods (PCMs)
43 have been developed to account for correlations induced by a shared history in the analysis
44 of a quantitative dataset (Pennell and Harmon 2013). They usually rely on two main
45 ingredients: a time-calibrated phylogenetic tree, and a dynamical model of trait evolution,
46 that should be chosen to capture the features of the trait evolution over time. Much work
47 has been made on the second ingredient, with more and more sophisticated models of trait
48 evolution, with numerous variations around the original Brownian Motion (BM), see for
49 instance Felsenstein (1985); Hansen and Martins (1996); Hansen (1997); Blomberg et al.
50 (2003); Butler and King (2004); Beaulieu et al. (2012); Landis et al. (2013); Blomberg
51 (2016), to cite only but a few.

52 In contrast, the first assumption has not been questioned until now (Jhwueng and
53 O'Meara 2017). However, phylogenetic trees are not always adapted to capture
54 relationships between species, and *phylogenetic networks* are sometimes needed.
55 Phylogenetic networks differ from trees by added reticulation points, where two distinct
56 branches come together to create a new species. Such reticulations can represent various
57 biological mechanisms, like hybridization, gene flow or horizontal gene transfer, that are
58 known to be common in some groups of organisms (Mallet 2005, 2007). Ignoring those
59 events can lead to misleading tree inference (Kubatko 2009; Solís-Lemus et al. 2016; Long
60 and Kubatko 2017). Thanks to recent methodological developments, the statistical

inference of reliable phylogenetic networks has become possible (Maddison 1997; Degnan and Salter 2005; Kubatko 2009; Yu et al. 2012, 2014; Yu and Nakhleh 2015; Solís-Lemus and Ané 2016). Although these state-of-the-art methods are still limited by their computational burden, we believe that the use of these networks will increase in the future. The goal of this work is to propose an adaptation of standard PCMs to a group of species with reticulate evolution, related by a network instead of a tree.

We describe an extension of the BM model of trait evolution to a network. The main modeling choice is about the fate of hybrid species. How should these species inherit their trait from their two parents? In this work, we first choose a weighted-average merging rule: the trait of a hybrid is a mixture of its two parents', weighted by their relative genetic contributions. This rule can be seen as a reasonable null model. But in some cases, the trait of a hybrid is observed to be outside of the range of its two parents. This phenomenon can be modeled by a *shift* in the trait value occurring right after the reticulation point: the hybrid trait value being the weighted average of the two parents, plus an extra term specific to the hybridization event at hand. Such a shift can model several biological mechanisms, such as transgressive segregation (Rieseberg et al. 1999) or heterosis (Fiévet et al. 2010; Chen 2013), with hybrid vigor (when the hybrid species is particularly fit to its environment) or depression (when the hybrid is ill-fit). In the following, we will refer to this class of phenomena using the generic term transgressive evolution. Here, this term only refers to the hybrid trait being different from the weighted average of its parents. This model allows for an explicit mathematical derivation of the trait distribution at the tips of the network and extends to networks the use of standard PCM tools such as phylogenetic regression (Grafen 1989, 1992), ancestral state reconstruction (Felsenstein 1985; Schluter et al. 1997) or tests of phylogenetic signal (Pagel 1999).

In the following, we first describe this BM model of trait evolution and show how it fits into the standard PCM framework. We then show how to add shifts in the trait values

to model transgressive evolution. We propose a statistical test for transgressive evolution. These methods are validated with a simulation study, and with the theoretical study of the power of the tests in a range of scenarios. Finally, we revisit the analysis of a *Xiphophorus* dataset about sword index and female preference made by Cui et al. (2013), in the light of our new network methods.

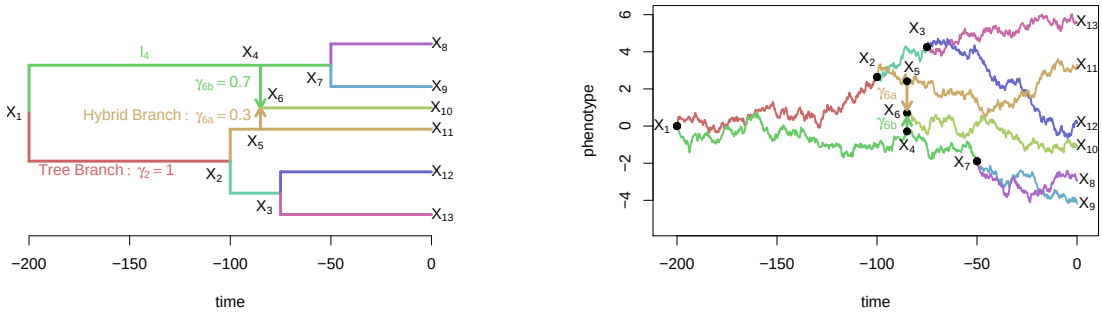
Model

In our model for trait evolution on a phylogenetic network, the novel aspect is the merging rule at reticulation events, compared to standard PCMs on trees. Our model is very similar to that defined in Jhwueng and O'Meara (2017), but we adopt a different statistical view point, based on the phylogenetic linear regression representation.

Trait Evolution on Networks

Phylogenetic Network.— In this work, we assume that we have access to a *rooted, calibrated and weighted phylogenetic network* that describes the relationships between a set of observed species (Huson et al. 2010). In such a network, reticulations, or *hybrids*, are nodes that have two parent nodes. They receive a given proportion of their genetic material from each parent. This proportion is controlled by a weight γ_e that represents the *inheritance probability* associated with each branch e of the network. A branch that is *tree-like*, i.e. that ends at a non-hybrid node, has a weight $\gamma_e = 1$. We further assume that the length ℓ_e of a branch e represents evolutionary time. In the example in Figure 1a, the two hybrid edges have length zero, but this need not to be the case, see Jhwueng and O'Meara (2017); Degnan (2017).

Brownian Motion.— Since the seminal article of Felsenstein (1985), the Brownian Motion (BM) has been intensively used to model trait evolution on phylogenetic trees. It is well



(a) A time-calibrated phylogenetic network

(b) BM on the branches of the network

Figure 1: Realization of a BM (with $\mu = 0$ and $\sigma^2 = 0.04$) on a calibrated network. The color of each branch (left) matches the color of the corresponding process (right). Only tip values are observed (here at time $t = 0$). For simplicity, the two hybrid branches were chosen to have a length of 0, but this need not be the case. Inheritance probabilities at the hybridization event are γ_{6a} and γ_{6b}, with $\gamma_{6a} + \gamma_{6b} = 1$.

adapted to model several biological processes, from random genetic drift, to rapid adaptation to a fluctuating environment (see e.g. Felsenstein 2004, Chap. 24). In order to adapt this process to a network instead of a tree, we define a weighted average merging rule at hybrids, as defined below. This rule expresses the idea that a hybrid inherits its trait from both its parents, with a relative weight determined by the proportion of genetic material received from each: if it inherits 90% of its genes from parent A, then 90% of its trait value should be determined by the trait of A. Because the BM usually models the evolution of a polygenic character, that is the additive result of the expression of numerous genes, this rule is a natural null hypothesis.

Definition 1 (BM on a Network)**.** Consider a rooted phylogenetic network with branch lengths and inheritance probabilities. Let X_v be the random variable describing the trait value of node (or vertex) v.

- At the root node ρ, we assume that $X_\rho = \mu$ is fixed.

- For a tree node v with parent node a, we assume that X_v is normally distributed with mean $X_a + \Delta_e$ and with variance $\sigma^2 \ell_e$, with σ^2 the variance rate of the BM, and ℓ_e the length of the parent edge e from a to v. Δ_e is a (fixed) shift value associated with branch e, possibly equal to 0.

- For a hybrid node v with parent nodes a and b, we assume that X_v is normally distributed with mean $\gamma_{e_a} X_a + \gamma_{e_b} X_b$, where e_a and e_b are the hybrid edges from a (and b) to v. If these edges have length 0, meaning that a, b and their hybrid v are contemporary, then X_v is assumed to have variance 0, conditional on the parent traits X_a and X_b. In general, the conditional variance of X_v is $\gamma_{e_a} \sigma^2 \ell_{e_a} + \gamma_{e_b} \sigma^2 \ell_{e_b}$. For the sake of identifiability, shifts are not allowed on hybrid branches (see Section on Transgressive Evolution for further details).

An example of such a process (without shift) is presented Figure 1b. This process is the same as in Jhwueng and O'Meara (2017), except that the shifts are treated differently. See Section on Transgressive Evolution and Discussion for more information on the links and differences between the two models. For the sake of generality, shifts are allowed on any tree edge. We will see in the next section how they are used to model transgressive evolution. In the rest of this section, we take all shifts to be zero, and only consider the un-shifted BM ($\Delta_e = 0$ for all edges e).

Note that the state at the root, μ, could also be drawn from a Gaussian distribution, instead of being fixed. This would not change the derivations below, and would simply add a constant value to all terms in the variance matrix.

Variance Matrix

From a Tree to a Network.— The distribution of trait values at all nodes, $\mathbf{X}$, can be fully characterized as a multivariate Gaussian with mean $\mu \mathbf{1}_{m+n}$ and variance matrix $\sigma^2 \mathbf{V}$,

where $\mathbf{1}_{m+n}$ is the vector of ones, n is the number of tips and m the number of internal nodes. The variance matrix $\mathbf{V}$, which depends on the topology of the network, encodes the correlations induced by the phylogenetic relationships between taxa. When the network reduces to a tree (if there are no hybrids), then $\mathbf{V}$ is the well-known BM covariance (Felsenstein 1985): $\mathbf{V}_{ij} = t_{ij}$ is the time of shared evolution between nodes i and j, i.e. the time elapsed between the root and the most recent common ancestor (mrca) of i and j.

When the network contains hybrids, this formula is not valid anymore. To see this, let's re-write t_{ij} as:

$$t_{ij} = \sum_{e \in p_i \cap p_j} \ell_e$$

where p_i is the path going from the root to node i. This formula just literally expresses that t_{ij} is the length of the shared path between the two nodes, that ends at their mrca. On a network, the difficulty is that there is *not a unique path* going from the root to a given node. Indeed, if there is a hybrid among the ancestors of node i, then a path might go "right" of "left" of the hybrid loop to go from the root to i.

Under the BM model in Definition 1 (with a fixed root), it turns out that we need to *sum* over all the possible paths going from the root to a given node, weighting paths by the inheritance probabilities γ_e of all the traversed edges:

$$V_{ij} = \sum_{\substack{p_i \in \mathcal{P}_i \\ p_j \in \mathcal{P}_j}} \left(\prod_{e \in p_i} \gamma_e \right) \left(\prod_{e \in p_j} \gamma_e \right) \sum_{e \in p_i \cap p_j} \ell_e \tag{1}$$

where $\mathcal{P}_i$ denotes the set of all the paths going from the root to node i.

This general formula for $\mathbf{V}$ was first presented in Pickrell and Pritchard (2012) in the context of population genomics. A formal proof is provided here (Appendix).

Remark 1 (Variance reduction). From the expression above, we can show that the variance of any tip i decreases with each hybridization ancestral to i. Consider time-consistent

network, in the sense that *all* paths from the root to a given hybrid node have the same length, as expected if branch lengths measure calendar time. Note that this is the opposite of the "NELP" property (No Equally Long Paths) defined by Pardi and Scornavacca (2015). For tip i, let t_i be the length of any path from the root to i. If the network is a tree, then $V_{ii} = t_i$. If the history of tip i involves one or more reticulations, then we show (Appendix), that:

$$V_{ii} < t_i \ . \tag{2}$$

This shows that hybridization events, that imply taking a weighted means of two traits, cause the trait variance to decrease.

Algorithm.— The formula above, although general, is not practical to compute. Using the recursive characterization of the process given in Definition 1, we can derive an efficient way to compute this covariance matrix across all nodes in the network (tips and internal nodes), in a single traversal of the network. This traversal needs to be in "preorder", from the root to the tips, such that any given node is listed after all of its parent(s): for any two nodes numbered i and j, if there is a directed path from i to j, then $i \leq j$. Such an ordering (also called topological sorting) can be obtained in linear time in the number of nodes and edges (Kahn 1962). On Figure 1a, nodes are numbered from 1 to 13 in preorder. The result below, proved in the Appendix, provides an efficient algorithm to compute the phylogenetic variance matrix $\mathbf{V}$ in a time linear in the number of nodes of the network, in a single preorder traversal.

Proposition 1 (Iterative computation of the phylogenetic variance)**.** *Assume that the nodes of a network are numbered in preorder. Then* $\mathbf{V}$ *can be calculated using the following step for each node* i*, from* $i = 1$ *to* $i = n + m$*:*

- *If* $i = 1$ *then* i *is the root, and* $V_{ii} = 0$*.*

- *If i is a tree node, denote by a the parent of i, and by ℓ_{e_a} the length of the branch e_a going from a to i. Then:*

$$\begin{cases} V_{ij} = V_{aj} & \text{for all } 1 \leq j \leq i-1 \\ V_{ii} = V_{aa} + \ell_{e_a} \ . \end{cases} \tag{3}$$

- *If i is a hybrid node, denote by a and b the parents of i, by ℓ_{e_a} and ℓ_{e_b} the lengths of the branches e_a and e_b going from a or b to i, and by γ_{e_a} and γ_{e_b} the associated inheritances probabilities. Then:*

$$\begin{cases} V_{ij} = \gamma_{e_a} V_{aj} + \gamma_{e_b} V_{bj} & \text{for all } 1 \leq j \leq i-1 \\ V_{ii} = \gamma_{e_a}^2 (V_{aa} + \ell_{e_a}) + \gamma_{e_b}^2 (V_{bb} + \ell_{e_b}) + 2\gamma_{e_a}\gamma_{e_b} V_{ab} \ . \end{cases} \tag{4}$$

Phylogenetic Regression

We can now define a *phylogenetic regression* on networks, the same way it is defined for phylogenetic trees (Grafen 1989, 1992).

Linear Regression Framework.— Define $\mathbf{Y}$ as the vector of trait values observed at the tips of the network. This is a sub-vector of the larger vector of trait values at all nodes. Let $\mathbf{V}^{\text{tip}}$ be the sub-matrix of $\mathbf{V}$, with covariances between the observed taxa (tips). The phylogenetic linear regression can be written as:

$$\mathbf{Y} = \mathbf{R}\boldsymbol{\theta} + \sigma^2 \mathbf{E} \quad \text{with} \quad \mathbf{E} \sim \mathcal{N}(\mathbf{0}_n, \mathbf{V}^{\text{tip}}) \tag{5}$$

where $\mathbf{R}$ is a $n \times q$ matrix of regressors, and $\boldsymbol{\theta}$ a vector of q coefficients. We can recover the distribution of $\mathbf{Y}$ under a simple BM with a fixed root value equal to μ (and no shift) by

taking $\mathbf{R} = \mathbf{1}_n$ and $\boldsymbol{\theta} = \mu$ (with $q = 1$). Regression matrix $\mathbf{R}$ can also contain some explanatory trait variables of interest. In this phylogenetic regression, the BM model applies to the residual variation not explained by predictors, $\mathbf{E}$.

This formulation is very powerful, as it recasts the problem into the well-known linear regression framework. The variance matrix $\mathbf{V}^{\text{tip}}$ is known (it is entirely characterized by the network used) so that, through a Cholesky factorization, we can reduce this regression to the canonical case of independent sampling units. This problem hence inherits all the features of the standard linear regression, such as confidence intervals for coefficients or data prediction, as explained in the next paragraph.

Ancestral State Reconstruction and Missing Data.— The phylogenetic variance matrix can also be used to do ancestral state reconstruction, or missing data imputation. Both tasks are equivalent from a mathematical point of view, rely on the Best Linear Unbiased Predictor (BLUP, see e.g. Robinson 1991) and are well known in the standard PCM toolbox. They have been implemented in many R packages, such as ape (Paradis et al. 2004, function ace), phytools (Revell 2012, function fastAnc) or more recently Rphylopars (Goolsby et al. 2017, function phylopars). In our Julia package PhyloNetworks, it is available as function ancestralStateReconstruction.

Pagel's λ.— The variance structure induced by the BM can be made more flexible using standard transformations of the network branch lengths, such as Pagel's λ (Pagel 1999). Because the network is calibrated with node ages, it is time-consistent: the time t_i elapsed between the root and a given node i is well defined, and does not depend on the path taken. Hence, the lambda transform used on a tree can be extended to networks, as shown below.

Definition 2 (Pagel's λ transform)**.** First, for any hybrid tip in the network, add a child edge of length 0 to change this tip into an internal (hybrid) node, and transfer the data from the former hybrid tip to the new tip. Next, let e be a branch of the network, with

child node i, parent node $\mathrm{pa}(i)$, and length ℓ_e. Then its transformed length is given by:

$$\ell_e(\lambda) = \begin{cases} \lambda\ell_e & \text{if } i \text{ is an internal node} \\ \ell_e + (1-\lambda)t_{\mathrm{pa}(i)} = \lambda\ell_e + (1-\lambda)t_i & \text{if } i \text{ is a tip,} \end{cases} \tag{6}$$

where t_i and $t_{\mathrm{pa}(i)}$ are the times elapsed from the root to node i and to its parent.

The interpretation of this transformation in term of phylogenetic signal is as usual: when λ decreases to zero, the phylogenetic structure is less and less important, and traits at the tips are completely independent for $\lambda = 0$. The first step of resolving hybrid tips is similar to a common technique to resolve polytomies in trees, using extra branches of length 0. This transformation does not change the interpretation of the network or the age of the hybrid. The added external edge does allow extra variation specific to the hybrid species, however, immediately after the hybridization, by Pagel's λ transformation. The second part of (6) applies to the new external tree edge, and hybrid edges are only affected by the first part of (6). The transformation's impact on the matrix $\mathbf{V}^{\mathrm{tip}}$ is not exactly the same as on trees. It still involves a simple multiplication of the off-diagonal terms by λ, but the diagonal terms are also modified. The following proposition is proved in the Appendix.

Proposition 2 (Pagel's λ effect on the variance)**.** *The phylogenetic variance of a BM running on a network transformed by a parameter λ, $\mathbf{V}(\lambda)$ is given by:*

$$\begin{cases} V(\lambda)_{ij} = \lambda V_{ij} & \text{for } i \text{ and } j \text{ such that } i \text{ or } j \text{ is an internal node, or } i \neq j \\ V(\lambda)_{ii} = \lambda V_{ii} + (1-\lambda)t_i & \text{for any tree tip } i \end{cases}$$

where $\mathbf{V} = \mathbf{V}(1)$ is the variance of the BM process on the non-transformed network.

On a tree, we have $V(\lambda)_{ii} = t_i$ for any tip i and any λ, so that the diagonal terms remain unchanged. This is not true on a network, however, as the Pagel transformation

erases the variance-reduction effect of ancestral hybridizations.

Other transformations, for instance based on Pagel's κ or δ (Pagel 1999), could be adapted to the phylogenetic network setting. Although these are not implemented for the moment, they would be straightforward to add in our linear regression framework.

Shifted BM and Transgressive Evolution

In our BM model, we allowed for shifts on non-hybrid edges. In this section, we show how those shifts can be inferred from the linear regression framework, and how they can be used to test for ancestral transgressive evolution events. When considering shifts, we again require that all tips are tree nodes. If a tip is a hybrid node, then the network is first resolved by adding a child edge of length 0 to the hybrid, making this node an internal node. This network resolution does not affect the interpretation of the network or the variance of the BM model. It adds more flexibility to the mean vector of the BM process, because the extra edge is a tree edge on which a shift can be placed.

Shift Vector.— We first describe an efficient way to represent the shifts on the network branches in a vector format. In Definition 1, we forbade shifts on hybrid branches. This does not lose generality, and is just for the sake of identifiability. Indeed, a hybrid node connects to three branches, two incoming and one outgoing. A shift on any of these three branches would impact the same set of nodes (apart from the hybrid itself), and hence would produce the same data distribution at the tips. The position of a shift on these three branches is consequently not identifiable. By restricting shifts to tree branches, the combined effect of branches with the same set of descendants is identified by a shift on a

single (tree) edge. We can combine all shift values in a vector $\mathbf{\Delta}$ indexed by nodes:

$$\Delta_i = \begin{cases} \mu & \text{if } i = \rho \text{ is the root node} \\ \Delta_e & \text{if } i \text{ is a tree node with parent edge } e \\ 0 & \text{if } i \text{ is a hybrid node.} \end{cases}$$

Note that any tree edge e is associated to its child node i in this definition. In the following, when there is no ambiguity, we will refer indifferently to one or the other.

Descendence Matrix.— For a rooted tree, a matrix of 0/1 values where each column corresponds to a clade can fully represent the tree topology. In column j, entries are equal to 1 for descendants of node number j, and 0 otherwise (Ho and Ané 2014; Bastide et al. 2017b). On a network, a node i can be a "partial" descendant of j, with the proportion of inherited genetic material represented by the inheritance probabilities γ_e. Hence, the descendence matrix of a network can be defined with non-binary entries between 0 and 1 as follows.

Definition 3 (Descendence Matrix)**.** The descendence matrix $\mathbf{U}$ of a network, given some ordering of its n tips and m internal nodes, is defined as an $(n+m) \times (n+m)$ matrix by:

$$U_{ij} = \sum_{p \in \mathcal{P}_{j \to i}} \prod_{e \in p} \gamma_e$$

where $\mathcal{P}_{j \to i}$ is the set of all the paths going from node j to node i (respecting the direction of edges). Note that, if i is not a descendant of j, then $\mathcal{P}_{j \to i}$ is empty and $U_{ij} = 0$. By convention, if $i = j$, we take $U_{ii} = 1$ (that is, a node is considered to be a descendant of itself). If the network is a tree, we recover the usual definition (all the γ_e are equal to 1). In general, the set of nodes i for which $U_{ij} > 0$ is the hardwired cluster of i, or the clade below i if the network is a tree.

Further define $\mathbf{T}$ as the (non-square) submatrix of $\mathbf{U}$ made of the rows that correspond to tip nodes (see example below).

Example 1 (Descendence Matrix and Shift Vector). The descendence matrices $\mathbf{U}$ and $\mathbf{T}$ associated with the network presented in Figure 2 are shown below, with zeros replaced by dots to improve readability:

$$\mathbf{U} = \begin{array}{c} \\ X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \\ X_6 \\ X_7 \\ \hline X_8 \\ X_9 \\ X_{10} \\ X_{11} \\ X_{12} \\ X_{13} \end{array} \begin{array}{c} \begin{array}{ccccccccccccc} X_1 & X_2 & X_3 & X_4 & X_5 & X_6 & X_7 & X_8 & X_9 & X_{10} & X_{11} & X_{12} & X_{13} \end{array} \\ \left(\begin{array}{ccccccccccccc} 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & 1 & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & \cdot & \cdot & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & 1 & \cdot & \cdot & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & \gamma_{6a} & \cdot & \gamma_{6b} & \gamma_{6a} & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & \cdot & \cdot & 1 & \cdot & \cdot & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \hline 1 & \cdot & \cdot & 1 & \cdot & \cdot & 1 & 1 & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & \cdot & \cdot & 1 & \cdot & \cdot & 1 & \cdot & 1 & \cdot & \cdot & \cdot & \cdot \\ 1 & \gamma_{6a} & \cdot & \gamma_{6b} & \gamma_{6a} & 1 & \cdot & \cdot & \cdot & 1 & \cdot & \cdot & \cdot \\ 1 & 1 & \cdot & \cdot & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & 1 & \cdot & \cdot \\ 1 & 1 & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 & \cdot \\ 1 & 1 & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 \end{array} \right) \end{array} \left. \begin{array}{c} \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \end{array} \right\} \mathbf{T}$$

The associated shift vector and associated trait means at the tips are shown below, where the only non-zero shift is assumed to correspond to transgressive evolution at the

hybridization event, captured by Δ_{10} on edge 10:

$$\boldsymbol{\Delta} = \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \\ 10 \\ 11 \\ 12 \\ 13 \end{array} \begin{pmatrix} \mu \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \Delta_{10} \\ \cdot \\ \cdot \\ \cdot \end{pmatrix} \qquad \mathbf{T}\boldsymbol{\Delta} = \begin{array}{c} 8 \\ 9 \\ 10 \\ 11 \\ 12 \\ 13 \end{array} \begin{pmatrix} \mu \\ \mu \\ \mu + \Delta_{10} \\ \mu \\ \mu \\ \mu \end{pmatrix}$$

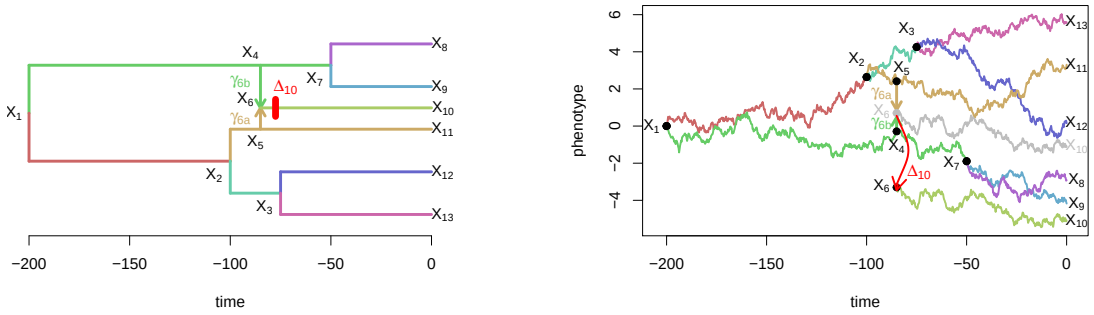
Note that rapid trait evolution or jumps in the trait value in other parts of the phylogeny could be also be modeled, by letting Δ_i be non-zero for other tree edges i.

Linear Model.— The shifted BM model in Definition 1 can be expressed by:

$$\mathbf{Y} = \mathbf{T}\boldsymbol{\Delta} + \sigma^2\mathbf{E} \quad \text{with} \quad \mathbf{E} \sim \mathcal{N}(\mathbf{0}_n, \mathbf{V}^{\text{tip}}) \tag{7}$$

where $\mathbf{Y}$ is the trait vector at the tips, and $\boldsymbol{\Delta}$ and $\mathbf{T}$ are the shift vector and the descendence matrix as defined above (see the Appendix for the proof).

Transgressive Evolution.— Even though the linear formulation above makes it easier to



(a) A phylogenetic network with transgressive evolution

(b) BM on the branches of the network

Figure 2: Realization of a univariate BM process (with $\mu = 0$ and $\sigma^2 = 0.04$) on a calibrated network, with transgressive evolution. The shift occurs right after the hybridization event, and changes the trajectory of the BM from the grey one to the colored one.

study, the problem of locating the non-zero shifts on the branches of a phylogenetic tree is difficult, and is still an active research area (see e.g. Uyeda and Harmon 2014; Bastide et al. 2017b; Khabbazian et al. 2016; Bastide et al. 2017a).

On networks as on trees, a shift can represent various biological processes. In the present work, we limit our study to shifts occurring on branches that are outgoing from a hybrid node (see Figure 2 for an example). Such shifts might represent a *transgressive evolution* effect, as defined in the introduction, and as a component of hybridization: the new species inherits its trait as a weighted average of the traits of its two parents, plus a shift representing extra variation, perhaps as a result of rapid selection.

Limiting shifts to being right after reticulations avoids the difficult exploration of all the possible locations of an unknown number of shifts on all the tree branches.

Statistical Tests for Transgressive Evolution

As there are typically only a few hybridization events in a phylogenetic network, we can test for transgressive evolution on each one individually. Thanks to the linear framework

described above, this amounts to a well-known test of fixed effects.

Statistical Model.— Denote by $\mathbf{N}$ the $n \times h$ sub-matrix of $\mathbf{T}$ containing only the columns corresponding to tree branches outgoing from hybrid nodes. We assume that $\mathbf{N}$ has full rank, that is, that the transgressive evolution shifts are identifiable. This is likely to be the case in networks that can be inferred by current methods, which typically have a small number of reticulations. We further denote by $\bar{\mathbf{N}}$ the vector of size n containing the row sums of $\mathbf{N}$: for tip i, $\bar{N}_i = \sum_{k=1}^{h} N_{ik}$. Then the phylogenetic linear regression extending (5) with transgressive evolution can be written as:

$$\mathbf{Y} = \mathbf{R}\boldsymbol{\beta} + \bar{\mathbf{N}}b + \mathbf{N}\mathbf{d} + \mathbf{E}\ , \quad \mathbf{d} \text{ such that } \sum_{k=1}^{h} d_k = 0\ , \quad \mathbf{E} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{V}^{\text{tip}}) \tag{8}$$

where $\mathbf{R}$ is a given matrix of regressors, with associated coefficients $\boldsymbol{\beta}$. These are included for the sake of generality, but usually only represent the ancestral state of the BM: $\mathbf{R} = \mathbf{1}_n$ and $\boldsymbol{\beta} = \mu$. The coefficient b represents a *common* transgressive evolution effect, that would affect all the hybridization events uniformly, while the vector $\mathbf{d}$ has h entries with a specific deviation from this common effect for each event, and represents *heterogeneity*.

Fisher Test.— When written this way, the problem of testing for transgressive evolution just amounts to testing the fixed effects b and $\mathbf{d}$. Some hypotheses that can be tested are summarized in the next table. $\mathcal{H}_0$ corresponds to the null model where the hybrids inherit their parents' weighted average. $\mathcal{H}_1$ is a model where all hybridization events share the same transgressive evolution effect, the trait being shifted by a common coefficient b. Finally, $\mathcal{H}_2$ is a model where each hybridization event k has its own transgressive evolution effect, with a shift $b + d_k$.

	Hypotheses	Linear Model
$\mathcal{H}_0$	No transgressive evolution	$b = 0$ and $\mathbf{d} = \mathbf{0}$
$\mathcal{H}_1$	Single effect transgressive evolution	$b \neq 0$ and $\mathbf{d} = \mathbf{0}$
$\mathcal{H}_2$	Multi effect transgressive evolution	$b \neq 0$ and $\mathbf{d} \neq \mathbf{0}$

Tests of fixed effects are very classic in the statistics literature (see e.g. Lehman 1986; Searle 1987). Compared to a likelihood ratio test, an F-test is exact and is more powerful, when available. Here we can define two F (Fisher) statistics F_{10} and F_{21} (see the Appendix). To see if $\mathcal{H}_2$ fits the data significantly better than $\mathcal{H}_1$, we compare F_{21} to an F distribution with degrees of freedom $r_{[\mathbf{R}\ \mathbf{N}]} - r_{[\mathbf{R}\ \bar{\mathbf{N}}]}$ and $n - r_{[\mathbf{R}\ \mathbf{N}]}$, where r is the matrix rank, and $[\mathbf{R}\ \mathbf{N}]$ is the matrix obtained by pasting the columns of $\mathbf{R}$ and $\mathbf{N}$ together. To test $\mathcal{H}_1$ versus the null model $\mathcal{H}_0$, we compare F_{10} to an F distribution with degrees of freedom $r_{[\mathbf{R}\ \bar{\mathbf{N}}]} - r_{\mathbf{R}}$ and $n - r_{[\mathbf{R}\ \bar{\mathbf{N}}]}$. We study these tests for several symmetric networks in the following section.

Simulation and Power Study

In this section, we first analyse the performance of the PCM tools described above, and then provide a theoretical power study of our statistical tests for transgressive evolution.

Implementation of the Network PCMs

All the tools described above, as well as simulation tools, were implemented in the julia package PhyloNetworks (Solís-Lemus et al. 2017). To perform a phylogenetic regression, the main function is phyloNetworklm. It relies on functions preorder! and sharedPathMatrix to efficiently compute the variance matrix using the algorithm in Proposition 1, and on julia package GLM (Bates 2016) for the linear regression. All the analysis and extraction tools provided by this GLM package can hence be used, including the ftest function to

perform the Fisher statistical tests for transgressive evolution. For the *Xiphophorus* fishes study (see below), we developed function calibrateFromPairwiseDistances! to calibrate a network topology based on pairwise genetic distances.

Simulation Study

Setting.— We considered 4 network topologies, all based on the same symmetric backbone tree with unit height and 32 tips, to which we added several hybridization events (Fig. 3, top). Those events were either taken very recent and numerous ($h = 8$ events each affecting 1 taxon) or quite ancient and scarce ($h = 2$ events each affecting 4 taxa). All networks had 8 tips with a hybrid ancestry. All the hybridization events had inheritance probability $\gamma = 0.3$. We then simulated datasets on these networks with $\mu = 0$, $\sigma^2 = 1$, and Pagel's λ transformation with λ in $\{0, 0.25, 0.5, 0.75, 1\}$. Recall that $\lambda = 0$ corresponds to all tips being independent, and $\lambda = 1$ is the simple BM on the original network. Each simulation scenario was replicated 500 times. To study the scalability of the implementation, we then reproduced these analysis on networks with 32 to 256 tips, and 1 to 8 hybridization events, each affecting 8 tips.

We analysed each dataset assuming either a BM or a λ model of evolution. When $\lambda \neq 1$, we could study the effect of wrongly using the BM. All the analyses were conducted on a laptop computer, with four Intel Core i7-6600U, and a 2.60GHz CPU speed.

Results.— When the vanilla BM model is used for both the simulation and the inference, the two parameters μ and σ^2 are well estimated, with no bias, for all the network topologies tested (Fig. 3, last two rows, red boxes for $\lambda = 1$). The estimation of μ is quite robust to the misspecification of the model, while σ^2 tends to be over-estimated (Fig. 3, last two rows, red boxes for $\lambda \neq 1$). This is expected, as in this case, the BM model wrongly tries to impose a strong correlation phylogenetic structure on the data, and can

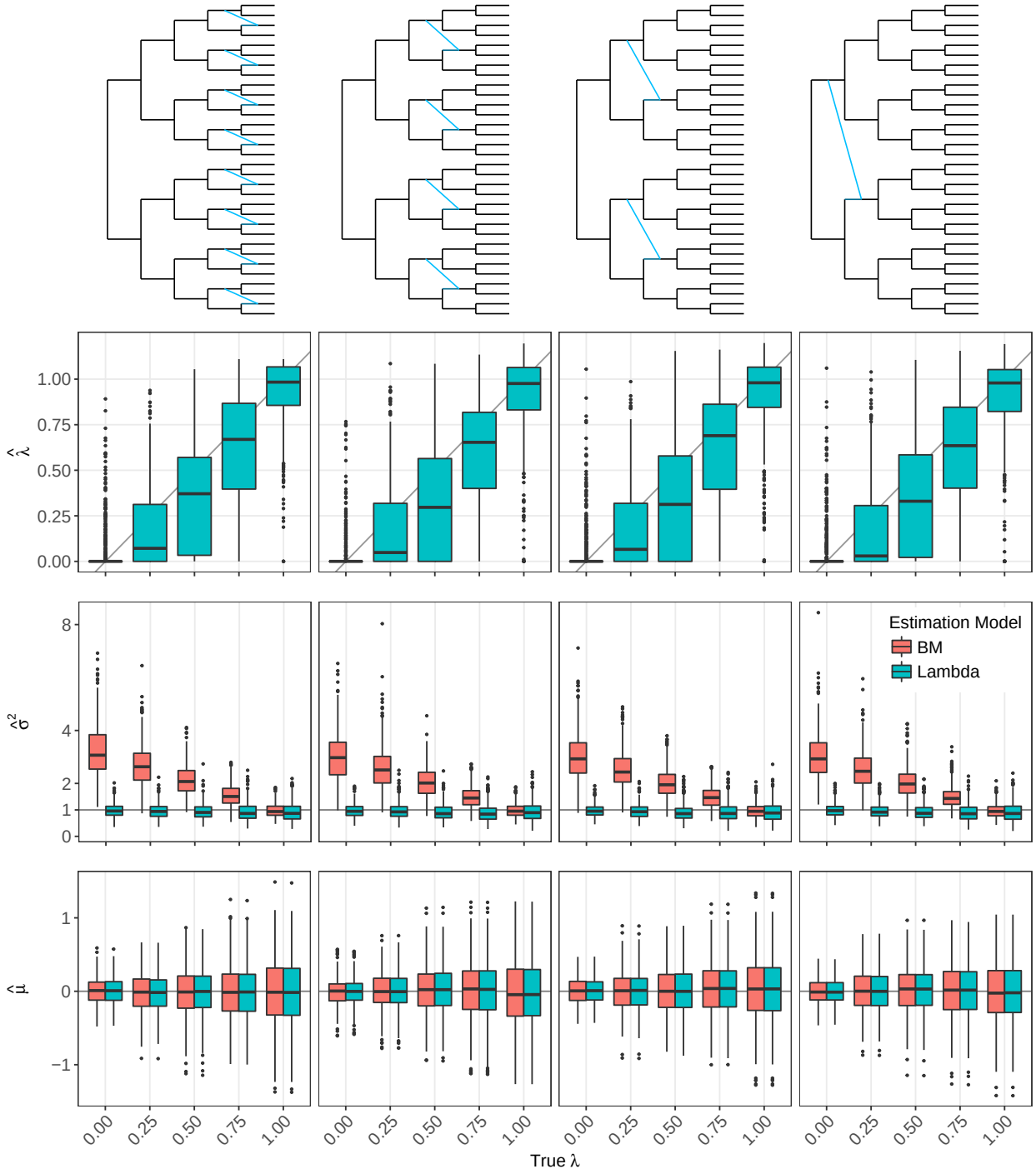


Figure 3: Estimated λ, σ^2 and μ values for several network topologies, with $\gamma = 0.3$, when the data are simulated according to a BM process with Pagel's λ transformation. Data were analyzed either with a straight BM model, which corresponds to $\lambda = 1$ (red), or with Pagel's λ transformed model (blue). True values are marked by a grey line. Boxplots show variation across 500 replicates.

only account for the observed diversity by raising the estimated variance, to accommodate both phylogenetic variance and independent intra-specific variation. When we use the true λ model for the inference, this bias is corrected, and both μ and σ^2 are correctly estimated (Fig. 3, last two rows, blue boxes). Furthermore, the λ estimate has a small bias but rather high variance (Fig. 3, second row). As expected, when the number of taxa increases, this variance decreases (data not shown). Finally, our implementation is quite fast (Fig. 4), with computing times ranging between 1 and 10 ms for a BM fit, and between 10 ms and 1 s for a Pagel's λ fit.

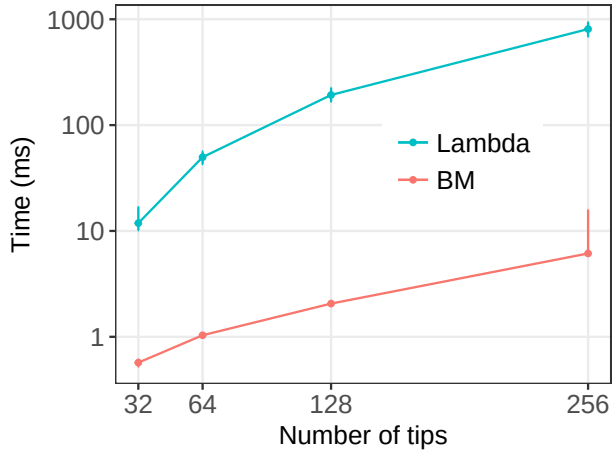


Figure 4: Computing time needed for fitting a continuous trait evolution model in PhyloNetworks. Median and confidence interval for 6000 repetitions in various conditions for each number of taxa. A log scale is used for the computing time.

Power Study of the Statistical Tests for Transgressive Evolution

We determined that our test statistics have the following noncentral Fisher distributions:

$$\text{Under } \mathcal{H}_1,\ F_{10} \sim \mathcal{F}\left(r_{[\mathbf{R}\ \bar{\mathbf{N}}]} - r_{\mathbf{R}},\ n - r_{[\mathbf{R}\ \bar{\mathbf{N}}]},\ \frac{b^2}{2\sigma^2}\Delta^2_{10}(\mathbf{R}, \bar{\mathbf{N}}, \mathbf{V}^{\text{tip}})\right) \tag{9}$$

$$\text{Under } \mathcal{H}_2,\ F_{21} \sim \mathcal{F}\left(r_{[\mathbf{R}\ \mathbf{N}]} - r_{[\mathbf{R}\ \bar{\mathbf{N}}]}, n - r_{[\mathbf{R}\ \mathbf{N}]}, \frac{1}{2\sigma^2}\Delta^2_{21}(\mathbf{d}, \mathbf{R}, \bar{\mathbf{N}}, \mathbf{N}, \mathbf{V}^{\text{tip}})\right) \tag{10}$$

The noncentral coefficient are determined by Δ_{10} and Δ_{21}, detailed in the Appendix. These Δ terms are zero under the null hypothesis ($\mathcal{H}_0$ for Δ_{10} and $\mathcal{H}_1$ for Δ_{21}), and depend on the network topology through the metric defined by $\mathbf{V}^{\text{tip}}$, and through the regression matrix $\mathbf{N}$.

Because we know the exact distribution of our F statistics, we do not need to resort to simulations to assess the power of these tests. In the following, we present a theoretical power study.

Test $\mathcal{H}_0$ vs $\mathcal{H}_1$.— We first studied the theoretical power to detect a single transgressive evolution effect, depending on the size b of this effect, and on the position of the hybridization event on the network. We considered 4 network topologies, using the same backbone tree than in the simulation study above, but adding only one hybridization event, occurring at various depths, from a very recent event affecting a single taxon to a very ancient event affecting 8 taxa (Fig. 5, top). The inheritance probability of this added hybrid branch was fixed to $\gamma = 0.4$. This parameter proved to have little influence to detect transgressive evolution (data not shown), for all the values tested, between 0 and 0.5. The underlying BM process had fixed ancestral value $\mu = 0$, and variance rate $\sigma^2 = 1$. Finally, for each network topology, we varied the transgressive evolution effect from 0 to 5, and computed the power of the test $\mathcal{H}_0$ vs $\mathcal{H}_1$ for three fixed standard levels (α in $\{0.01, 0.05, 0.1\}$).

As expected, the power improves with the size of the effect, reaching approximately 1 for $b = 5$ in all scenarios (Fig. 5, bottom). In addition, the transgressive evolution effect seems easier to detect for *recent* hybridization events, even if they affect fewer tips. One intuition for that is that ancient hybridization effects are "diluted" by the variance of the BM, and are hence harder to detect, even if they affect more tips. This may be similar to the difficulty of detecting ancient hybridization compared to recent hybridizations.

Test $\mathcal{H}_1$ vs $\mathcal{H}_2$.— We used a similar framework to study the power of the test to detect

heterogeneity in the transgressive evolution effects. We used here the same 4 networks than in the simulation study, with 32 tips and 2 to 8 hybridization events (Fig. 6, top), but with inheritance probabilities fixed to $\gamma = 0.4$. Transgressive evolution effects were set to $\mathbf{d} = d\mathbf{d}^u$, with $\mathbf{d}^u$ fixed to $d^u_i = 1$ for $i \leq h/2$ and $d^u_i = -1$ for $i > h/2$, h being the number of hybrids, which was even in all the scenarios we considered. Note that the average transgressive evolution effect was 0, because the d^u_i values sum up to 0. This allowed us to reduce the "strength of heterogeneity" to a single parameter d, which we varied between 0 and 5 (see appendices for the reduced expression of the noncentral coefficient). Like before, we computed the power of the test $\mathcal{H}_1$ vs $\mathcal{H}_2$ for three fixed standard levels (α in $\{0.01, 0.05, 0.1\}$).

Figure 6 (bottom) shows a similar pattern: the test is more powerful for a high heterogeneity coefficient, and for recent hybridization events. For variation of about 3.5 in transgressive evolution, the power is close to one in all the scenarios considered here.

Xiphophorus FISHES

Methods

Network inference.— We revisited the example in Solís-Lemus and Ané (2016) and re-analyzed transcriptome data from Cui et al. (2013) to reconstruct the evolutionary history of 23 swordtails and platyfishes (Xiphophorus: Poeciliidae). The original work included 24 taxa, but we eliminated *X. nezahualcoyotl*, because the individual sequenced in Cui et al. (2013) was found to be a lab hybrid not representative of the wild species *X. nezahualcoyotl* (personal communication). We re-analyzed their first set of 1183 transcripts, and BUCKy (Larget et al. 2010) was performed on each of the 8,855 4-taxon

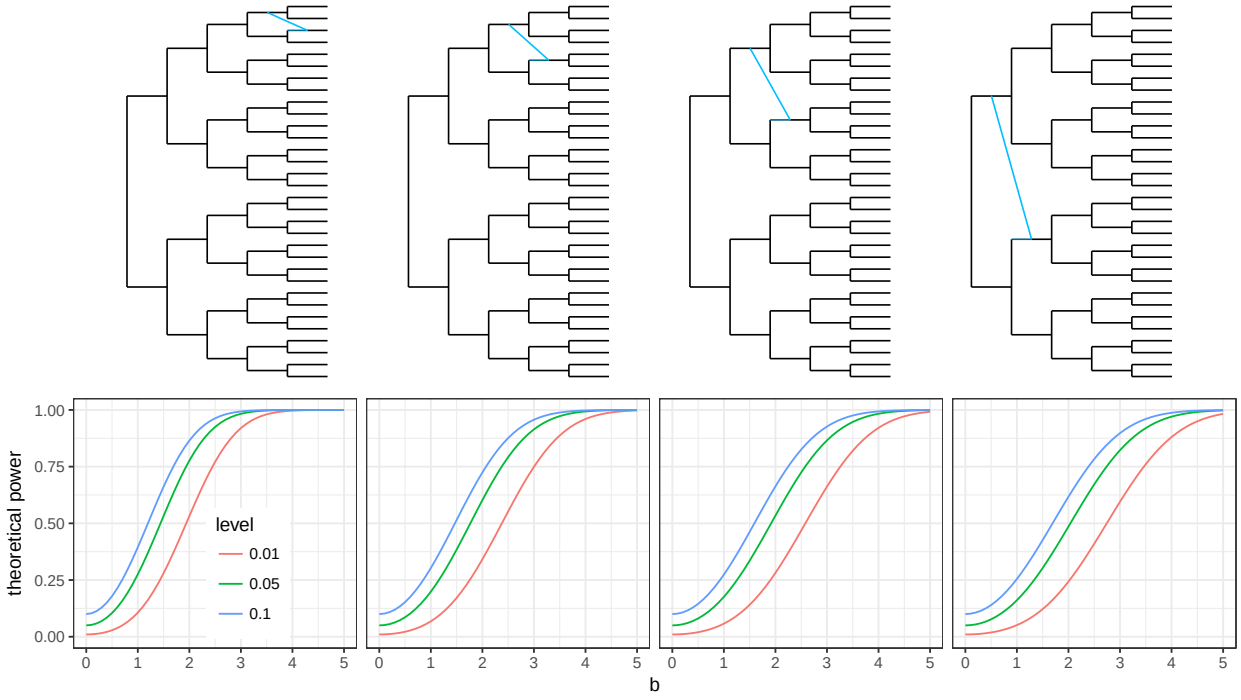


Figure 5: Theoretical power of the shared transgressive evolution test $\mathcal{H}_0$ vs $\mathcal{H}_1$, for four different networks topologies with inheritance probability $\gamma = 0.4$ (top), and a BM with ancestral value $\mu = 0$ and variance rate $\sigma^2 = 1$. The power of the test increases with the transgressive evolution effect b (bottom).

sets. The resulting quartet CFs were used in SNaQ (Solís-Lemus and Ané 2016), using $h = 0$ to $h = 5$ and 10 runs each. The network scores (negative log-pseudolikelihood) decreased very sharply from $h = 0$ to 1, strongly between $h = 1$ to 3, then decreased only slightly and somewhat linearly beyond $h = 3$ (Fig. 7, top left). Using a broken stick heuristic, one might suggest that $h = 1$ or perhaps $h = 3$ best fits the data. Given our focus on PCMs, we used both networks ($h = 1$ and 3) as well as the tree ($h = 0$) to study trait evolution, and to compare results across the different choices of reticulation numbers.

Network calibration.— SNaQ estimates branch lengths in coalescent units, which are not expected to be proportional to time, and are also not estimable for some edges (like

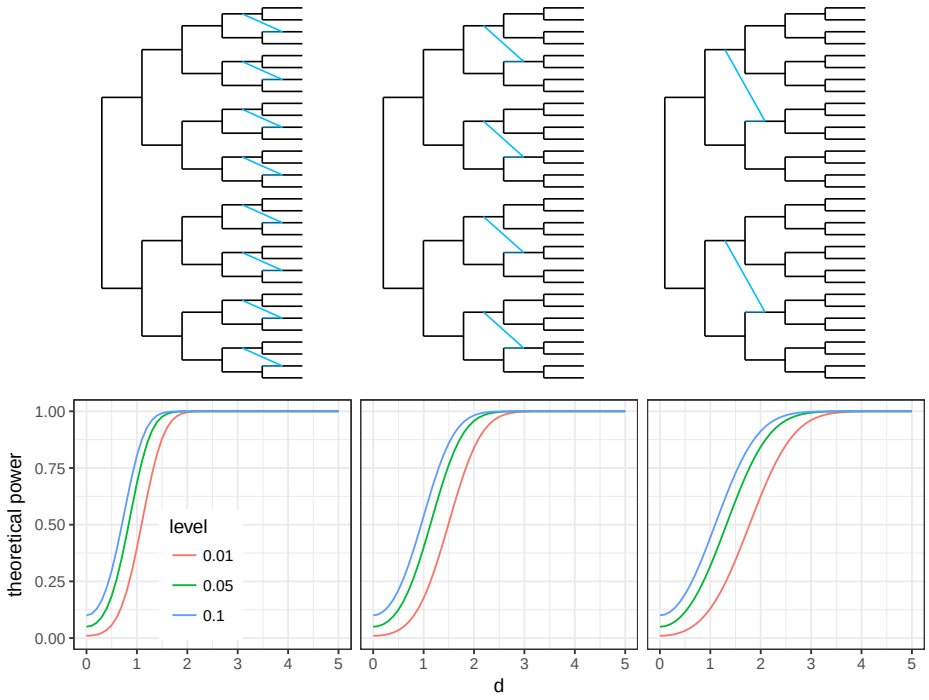


Figure 6: Theoretical power of the test for heterogeneous transgressive evolution $\mathcal{H}_1$ vs $\mathcal{H}_2$, for three different networks topologies with inheritance probability $\gamma = 0.4$ (top), and a BM with ancestral value $\mu = 0$ and variance $\sigma^2 = 1$. The power of the test increases with the heterogeneity coefficient d (bottom).

external branches to taxa represented by a single individual). To calibrate the network, we estimated pairwise genetic distances between taxa, and then optimized node divergence times using a least-square criterion, as detailed below.

To estimate pairwise distances, individual gene trees were estimated with RAxML, using the HKY model and gamma-distributed rate variation among sites. For each locus, branch lengths were rescaled to a median of 1 to reduce rate variation across loci, before obtaining a pairwise distance matrix from each rescaled gene tree. Loci with one or more missing taxon were then excluded (leaving 1019 loci), and pairwise distance matrices were averaged across loci.

This average pairwise distance matrix was used to estimate node ages on each

network ($h = 0, 1, 3$). The network pairwise distance between taxa i and j was taken as the weighted average distance between i and j on the trees displayed by the network, where the weight of a displayed tree is the product of the inheritance probabilities γ_e for all edges e retained in the tree. We estimated node ages that minimized the ordinary least-squares mismatch between the genetic pairwise distances and the network pairwise distances.

Traits.— With data presented in Cui et al. (2013) and following their study on sword evolution, we revisited the hypotheses that females have a preference for males with longer swords, and that the common ancestor of the genus *Xiphophorus* likely had a sword. Rather than using the methods of parsimony character mapping and independent contrasts as in Cui et al. (2013), we tested the effect of hybridization on the ancestral state reconstructions and the correlation between both traits using networks with zero, one or three hybridization events, using phyloNetworklm. For each network, the topology and branch lengths were assumed to be perfectly estimated, and fixed. We also tested for phylogenetic signal in both traits on all networks using Pagel's λ, as well as for transgressive evolution, using the F statistics defined above. For the phylogenetic regression, more than half of the species were excluded because they lack information on female preference.

Along with the datasets used, two executables julia markdown (.jmd) files are provided in the online supplementary material, allowing the interested reader to reproduce all the analyses described here.

Results

The *Xiphophorus* fish topologies with zero, one, and three hybridization events were calibrated using pairwise genetic distances (Fig. 7, bottom, for $h = 0$ and 3). With $h = 1$, the reticulation event did not necessarily imply the existence of unsampled or extinct taxa,

so we constrained this reticulation to occur between contemporary populations (with an edge length of 0). For the network with $h = 3$, two reticulation events implied the existence of unsampled taxa, so we calibrated this network without constraint, to allow minor reticulation edges of positive lengths. Optimized branch lengths were similar between networks. Branch lengths were estimated to be 0 for some tree edges and some unconstrained hybrid edges, creating polytomies.

Using networks with 0, 1 or 3 hybridization events, we found a positive correlation between female preference and longer swords in males, but this relationship was not statistically significant ($h = 0$: $p = 0.096$; $h = 1$: $p = 0.110$; $h = 3$: $p = 0.106$). Ancestral state reconstruction of sword index shows the presence of a sword at the MRCA of each network because unsworded species were assigned a value of 0.275 in Cui et al. (2013) and the ancestral state in all networks was reconstructed to be 0.46. Phylogenetic signal was high for both traits with estimated $\lambda = 1.0$ on all networks (or above 1.0 with unconstrained maximum likelihood).

We also applied our tests for transgressive evolution on both traits, using the network with 3 hybridization events (Fig. 7, lower right). For the sword index, we found no evidence of transgressive evolution ($p = 0.55$ and $p = 0.28$, respectively, for homogeneous or heterogeneous transgressive evolution). However, we did find some evidence for an heterogeneous transgressive evolution effect for female preference. Testing $\mathcal{H}_2$ against $\mathcal{H}_1$ gives $p = 0.0087$. Testing $\mathcal{H}_2$ against $\mathcal{H}_0$ directly, we get $p = 0.0064$ (see the Appendix for a description of this third test, also based on a Fisher statistic). However, transgressive evolution effects were in opposite directions (one positive and two negative), such that the common effect was not significantly different from 0: $\mathcal{H}_1$ vs $\mathcal{H}_0$ gave $p = 0.11$.

Discussion

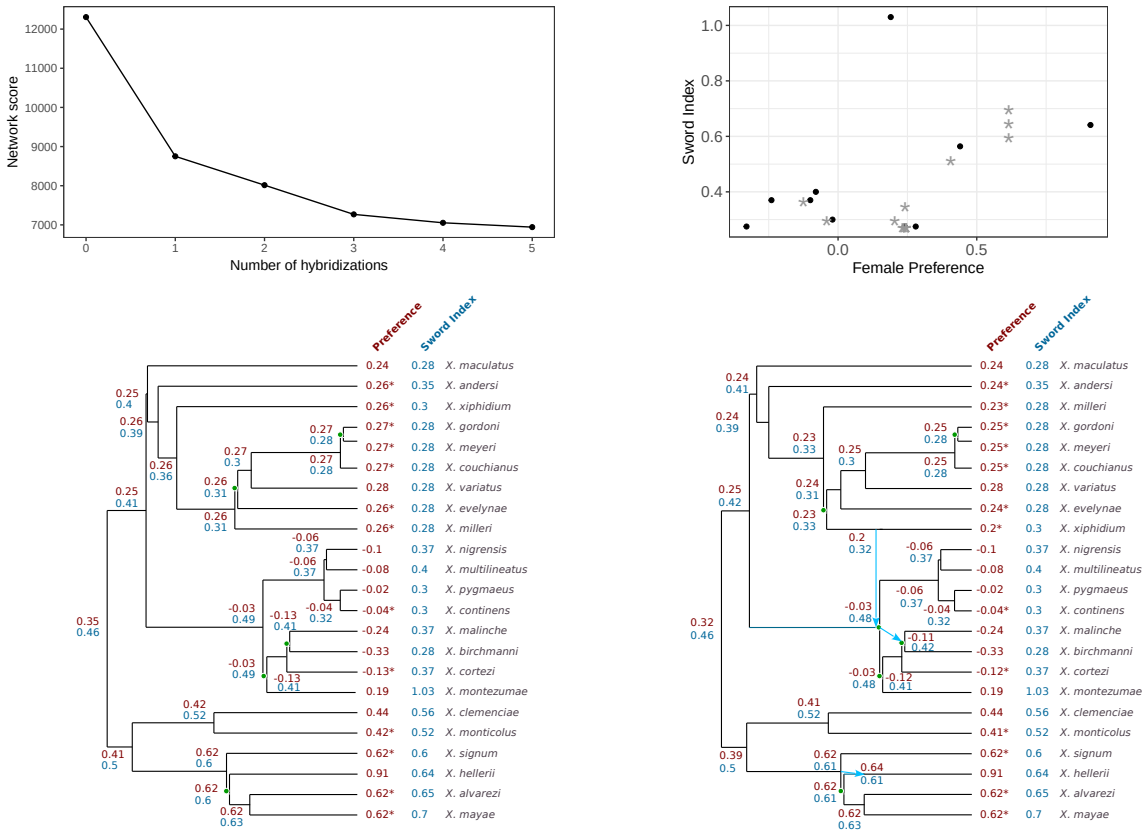


Figure 7: Results of the analysis on the fish dataset. Top left: negative pseudo log-likelihood score of the estimated networks with various numbers of hybridizations. Top right: scatter plot of sword index and female preference. Gray stars are taxa missing female preference data, for which female preference was predicted using ancestral state reconstruction of the trait on the network (independent of sword index). Bottom: ancestral state reconstruction of both traits, independently, using a BM model on the tree ($h = 0$, left) or on the network with $h = 3$ (right). Starred values indicate taxa with missing preference data, and imputed female preference values. Branches with an estimated length zero are indicated by a green dot, to show the network topologies.

Impact of the Network.— The results from the fish dataset analysis using a tree ($h = 0$) or a network ($h = 1$ or $h = 3$) show that taking the hybridization events into account has a small impact on the ancestral state reconstruction and on the estimation of parameters, both for the regression analysis and for the test for phylogenetic signal. This finding was corroborated by simulations: when we ignored hybridization events, using a tree while the true underlying model was a network, we found that the estimation of parameters μ and σ^2 was only slightly affected (data not shown). These results may indicate that major previous findings, that used a phylogenetic tree where a phylogenetic network might have been more appropriate, are likely to be robust to a violation of the tree-like ancestry assumption. Our new model may simply refine previous estimates in many cases.

However, the structure of the network has a strong impact on the study of transgressive evolution. This is expected, as the model allows for shifts below each inferred hybrid. If one reticulation is undetected, or if one was incorrectly located on the network, then the model will be ill-fitted, leading to potentially misleading conclusions. As an example, we reproduced the analysis of transgressive evolution for female preference on the network with three hybridization events, but this time pruning the network, to keep only the taxa with a measured trait. Preference data were missing for species *X. signum*, *X. alvarezi* and *X. mayae*, such that *X. helleri* became the only species impacted by one of the reticulation event, which became a simple loop in the network. In other words, *X. helleri* was the only descendant of the reticulation, and also the closest relative of the hybrid's parent among the remaining taxa. The reticulation could be dropped from the pruned network. This new and simplified network only retained the two hybridization events associated with negative shifts. As a consequence, and contrary to the conclusion we found in the main text, we found support for *homogeneous* transgressive evolution ($p = 0.0071$ for $\mathcal{H}_1$ vs $\mathcal{H}_0$), and no support for heterogeneous effects ($p = 0.88$ for $\mathcal{H}_1$ vs $\mathcal{H}_0$). This illustrates that caution is needed for the interpretation of tests of transgressive

evolution, as those highly depend on the quality of the input network inference, which is a recognized hard problem.

Network Calibration.— To conduct PCMs, we developed a distance-based method to calibrate a network topology into a time-consistent network. This is a basic method that makes a molecular clock assumption on the input pairwise distance matrix. Important improvements could be made to account for rate variation across lineages, and to use calibration dates from fossil data, like in relaxed clock calibration methods for phylogenetic trees such as r8s (Sanderson 2003) or BEAST (Drummond et al. 2006). In our fish example, we averaged pairwise distances across loci, to mitigate a violation of the molecular clock that might be specific to each locus.

Our method estimated some branch lengths to be 0, thereby creating polytomies. This behavior is shared by other well-tested distance-based methods like Neighbor-Joining (Saitou and Nei 1987), which can also estimate 0 or even negative branch lengths.

We also noticed that several calibrations could fit the same matrix of genetic pairwise distances equally well, pointing to a lack of identifiability of some node ages. This issue occurred for the age of hybrid nodes and of their parent nodes. Branch lengths and node ages around reticulation points were also found to be non-identifiable by Pardi and Scornavacca (2015), when the input data consist of the full set of trees displayed by the network, and when these trees are calibrated. This information on gene trees can only identify the "unzipped" version of the network, where unzipping a reticulation means moving the hybrid point as close as possible to its child node (see Pardi and Scornavacca 2015, for a rigorous description of "canonical" networks). This unzipping operation creates a polytomy after the reticulation point. We observed such polytomies for two events in our calibrated network (Fig. 7, bottom right). Pardi and Scornavacca (2015) proved that the lack of identifiability is worse for time-consistent networks, which violates their "NELP"

property (no equally-long paths). Lack of identifiable branch lengths around reticulations is thus observed from different sources of input data, and requires more study. Methods utilizing multiple sources of data might be able to resolve the issue. For instance, gene tree discordance is informative about branch lengths in coalescent units around reticulation nodes, and could rescue the lack of information from other input data like pairwise distances or calibrated displayed trees. More work is also needed to study the robustness of transgressive evolution tests to errors in estimated branch lengths.

Comparison with Jhwueng and O'Meara (2017).— In their model, Jhwueng and O'Meara (2017) include hybridization events as *random* shifts. Using their notations, each hybrid k shifts by a coefficient $\log\beta + \delta_k$, with δ_k a random Gaussian with variance ν_H: $\delta_k \sim \mathcal{N}(0, \nu_H)$. This formulation provides a *mixed effects* linear model, with shifts appearing as random effects. In this framework, the test of heterogeneity ($\mathcal{H}_2$ vs $\mathcal{H}_1$) amounts to a test of null variance, $\nu_H = 0$. In the context of mixed effects linear models, such tests are also well studied, but are known to be more difficult than tests of fixed effects (Lehman 1986; Khuri et al. 1998). Assuming that the variance ν_H is 0, our test for a common transgressive evolution effect ($\mathcal{H}_1$ vs $\mathcal{H}_0$) is then similar to the likelihood-based test for $\log\beta = 0$ in Jhwueng and O'Meara (2017). A mixed-effect model is legitimate, although it might be more difficult to study theoretically, and its inference can be more tricky. Jhwueng and O'Meara (2017) indeed report some numerical problems, and rather large sampling error for both $\log\beta$ and ν_H. Current state-of-the-art methods to infer phylogenetic networks cannot handle more than 30 taxa and no more than a handful of reticulation events (Hejase and Liu 2016). Hence, it might not be surprising that estimating a variance ν_H for an event that is only observed two or three times is indeed difficult. On data sets with few reticulations, we believe that our fixed effect approach can be better suited. However, our approach adds a parameter for each hybridization event,

whereas the random-effect approach of Jhwueng and O'Meara (2017) maintains only two parameters (mean and variance). As the available networks are likely to grow over the next few decades, this later approach might be preferable in the future.

Perspectives.— As stated in the introduction, PCMs rely on two fundamental components: the species relationship model (tree or network), and the model of trait evolution. Here, we showed how a network could be used instead of a tree, but we used the most simple model of trait evolution (BM). Future developments could adapt some of the more refined models to the network framework, in order to capture the diverse tempo and modes of evolution. In doing so, the salient point to be careful about is the merging rule one might adopt for all these processes.

For instance, the Ornstein-Uhlenbeck (OU) process is popular to model trait evolution (Hansen 1997). It has extra parameters compared to the BM: a primary optimum for the trait, and α, a rubber band parameter that controls how the trait is pulled toward its optimum. Either one might vary across lineages. What behavior would be biologically realistic at reticulation points? For an OU with one single optimum value over the whole tree, the weighted average merging rule could be adopted. But how should transgressive evolution be modeled? With the OU process, shifts have been traditionally considered on the optimal value rather that directly on the process' value, as we did for the BM (Butler and King 2004; Beaulieu et al. 2012). If a transgressive evolution shift is allowed on the optimum value, this would result in several optima on different regions of the network, which might not capture biological realism. A related problem is to find a realistic merging rule for reticulations between two species evolving in two different phylogenetic groups with different optima.

More generally, the numerous improvements that have been developed for PCMs on trees should be adapted to phylogenetic networks, such as support for measurement error

or intra-specific variation (as in, e.g. Lynch 1991; Ives et al. 2007; Felsenstein 2008; Goolsby et al. 2017); distinct regimes of evolution on different regions of the network (see e.g. Beaulieu et al. 2012); and multivariate processes (Felsenstein 1985; Bartoszek et al. 2012; Clavel et al. 2015).

Sticking with the vanilla BM, it could also be interesting to look into other merging rules at reticulation points. For instance, instead of taking a weighted average, one could draw either one of the two parents' trait for the hybrid, with probabilities defined by the weights γ_a and γ_b of the parents. If such a rule could be justified from a modelling point of view, further work would be needed to derive the induced distribution of the trait at the tips of the network.

Acknowledgments

PB would like to thank Mahendra Mariadassou and Stéphane Robin for enlightening discussions, and useful comments on an early version of this work. PB also thanks Tristan Mary-Huard for sharing his extensive knowledge on the linear mixed model. The authors thank Mohammad Khabbazian for insights on the topological sorting algorithm.

Funding

The visit of PB to the University of Wisconsin-Madison during the fall of 2015 was funded by a grant from the Franco-American Fulbright Commission. This work was funded in part by the National Science Foundation (DEB 1354793) and by a Vilas Associate award to CA from the University of Wisconsin-Madison.

*

625 References

626 Bartoszek K, Pienaar J, Mostad P, Andersson S, Hansen TF. 2012. A phylogenetic
627 comparative method for studying multivariate adaptation. *Journal of Theoretical*
628 *Biology*. 314:204–215.

629 Bastide P, Ané C, Robin S, Mariadassou M. 2017a. Inference of Adaptive Shifts for
630 Multivariate Correlated Traits. *bioRxiv*. .

631 Bastide P, Mariadassou M, Robin S. 2017b. Detection of adaptive shifts on phylogenies by
632 using shifted stochastic processes on a tree. *Journal of the Royal Statistical Society:*
633 *Series B (Statistical Methodology)*. 79:1067–1093.

634 Bates D. 2016. Generalized linear models in Julia.
635 `https://github.com/JuliaStats/GLM.jl`.

636 Beaulieu JM, Jhwueng DC, Boettiger C, O'Meara BC. 2012. Modeling stabilizing
637 selection: Expanding the Ornstein-Uhlenbeck model of adaptive evolution. *Evolution*.
638 66:2369–2383.

639 Blomberg SP. 2016. Beyond Brownian motion and the Ornstein-Uhlenbeck process:
640 Stochastic diffusion models for the evolution of quantitative characters. *bioRxiv e-print*. .

641 Blomberg SP, Garland T, Ives AR. 2003. Testing for Phylogenetic Signal in Comparative
642 Data: Behavioral Traits Are More Labile. *Evolution*. 57:717–745.

643 Butler MA, King AA. 2004. Phylogenetic Comparative Analysis: A Modeling Approach for
644 Adaptive Evolution. *The American Naturalist*. 164:683–695.

645 Chen ZJ. 2013. Genomic and epigenetic insights into the molecular bases of heterosis.
646 *Nature Reviews Genetics*. 14:471–482.

647 Clavel J, Escarguel G, Merceron G. 2015. mvMORPH: an R package for fitting
648 multivariate evolutionary models to morphometric data. *Methods in Ecology and*
649 *Evolution*. 6:1311–1319.

650 Cui R, Schumer M, Kruesi K, Walter R, Andolfatto P, Rosenthal GG. 2013.
651 Phylogenomics reveals extensive reticulate evolution in *xiphophorus* fishes. *Evolution;*
652 *international journal of organic evolution*. 67:2166–79.

653 Degnan J. 2017. Modeling hybridization with incomplete lineage sorting. *Systematic*
654 *Biology*. p. submitted for Symposium issue.

655 Degnan JH, Salter LA. 2005. Gene tree distributions under the coalescent process.
656 *Evolution; international journal of organic evolution*. 59:24–37.

657 Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. 2006. Relaxed phylogenetics and
658 dating with confidence. *PLOS Biology*. 4:e88.

659 Felsenstein J. 1985. Phylogenies and the Comparative Method. *The American Naturalist*.
660 125:1–15.

661 Felsenstein J. 2004. Inferring Phylogenies. Sinauer Associates, Sunderland, Mass.

662 Felsenstein J. 2008. Comparative Methods with Sampling Error and WithinSpecies
663 Variation: Contrasts Revisited and Revised. *The American Naturalist*. 171:713–725.

664 Fiévet JB, Dillmann C, de Vienne D. 2010. Systemic properties of metabolic networks lead
665 to an epistasis-based model for heterosis. *Theoretical and Applied Genetics*. 120:463–473.

666 Goolsby EW, Bruggeman J, Ané C. 2017. Rphylopars: fast multivariate phylogenetic
667 comparative methods for missing data and within-species variation. *Methods in Ecology*
668 *and Evolution*. 8:22–27.

669 Grafen A. 1989. The Phylogenetic Regression. *Philosophical Transactions of the Royal*
670 *Society B: Biological Sciences*. 326:119–157.

671 Grafen A. 1992. The uniqueness of the phylogenetic regression. *Journal of Theoretical*
672 *Biology*. 156:405–423.

673 Hansen TF. 1997. Stabilizing selection and the comparative analysis of adaptation.
674 *Evolution*. 51:1341.

675 Hansen TF, Martins EP. 1996. Translating between microevolutionary process and
676 macroevolutionary patterns: The correlation structure of interspecific data. *Evolution*.
677 50:1404.

678 Hejase HA, Liu KJ. 2016. A scalability study of phylogenetic network inference methods
679 using empirical datasets and simulations involving a single reticulation. *BMC*
680 *Bioinformatics*. 17:422.

681 Ho LST, Ané C. 2014. Intrinsic inference difficulties for trait evolution with
682 Ornstein-Uhlenbeck models. *Methods in Ecology and Evolution*. 5:1133–1146.

683 Huson DH, Rupp R, Scornavacca C. 2010. Phylogenetic Networks. Cambridge: Cambridge
684 University Press.

685 Ives AR, Midford PE, Garland T, Oakley T. 2007. Within-species variation and
686 measurement error in phylogenetic comparative methods. *Systematic Biology*.
687 56:252–270.

688 Jhwueng DC, O'Meara BC. 2017. Trait evolution on phylogenetic networks. *Systematic*
689 *Biology*. p. submitted for Symposium issue.

690 Kahn AB. 1962. Topological sorting of large networks. *Communications of the ACM*.
691 5:558–562.

692 Khabbazian M, Kriebel R, Rohe K, Ané C. 2016. Fast and accurate detection of
693 evolutionary shifts in Ornstein-Uhlenbeck models. *Methods in Ecology and Evolution*.
694 7:811–824.

695 Khuri AI, Mathew T, Sinha BK. 1998. Statistical Tests for Mixed Linear Models. Wiley
696 series in Probabilities and Statistics.

697 Kubatko LS. 2009. Identifying hybridization events in the presence of coalescence via
698 model selection. *Systematic Biology*. 58:478–488.

699 Landis MJ, Schraiber JG, Liang M. 2013. Phylogenetic analysis using Lévy processes:
700 Finding jumps in the evolution of continuous traits. *Systematic Biology*. 62:193–204.

701 Larget B, Kotha S, Dewey C, Ané C. 2010. BUCKy: Gene tree / species tree reconciliation
702 with Bayesian concordance analysis. *Bioinformatics*. 26:2910–2911.

703 Lehman EL. 1986. Testing Statistical Hypotheses. Springer Texts in Statistics. New York,
704 NY: Springer New York.

705 Long C, Kubatko L. 2017. The effect of gene flow on coalescent-based species-tree
706 inference. *Systematic Biology*. p. submitted for Symposium issue.

707 Lynch M. 1991. Methods for the Analysis of Comparative Data in Evolutionary Biology.
708 *Evolution*. 45:1065–1080.

709 Maddison WP. 1997. Gene trees in species trees. *Systematic Biology*. 46:523.

710 Mallet J. 2005. Hybridization as an invasion of the genome. *Trends in Ecology &*
711 *Evolution*. 20:229–237.

712 Mallet J. 2007. Hybrid speciation. *Nature*. 446:279–283.

713 Pagel M. 1999. Inferring the historical patterns of biological evolution. *Nature*.
714 401:877–884.

715 Paradis E, Claude J, Strimmer K. 2004. APE: Analyses of Phylogenetics and Evolution in
716 R language. *Bioinformatics*. 20:289–290.

717 Pardi F, Scornavacca C. 2015. Reconstructible phylogenetic networks: Do not distinguish
718 the indistinguishable. *PLOS Computational Biology*. 11:e1004135.

719 Pennell MW, Harmon LJ. 2013. An integrative view of phylogenetic comparative methods:
720 connections to population genetics, community ecology, and paleobiology. *Annals of the*
721 *New York Academy of Sciences*. 1289:90–105.

722 Pickrell JK, Pritchard JK. 2012. Inference of population splits and mixtures from
723 genome-wide allele frequency data. *PLoS Genetics*. 8:e1002967.

724 Revell LJ. 2012. phytools: An R package for phylogenetic comparative biology (and other
725 things). *Methods in Ecology and Evolution*. 3:217–223.

726 Rieseberg LH, Archer Ma, Wayne RK. 1999. Transgressive segregation, adaptation and
727 speciation. *Heredity*. 83:363–372.

728 Robinson GK. 1991. That BLUP is a good thing: The estimation of random effects.
729 *Statistical Science*. 6:15–32.

730 Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing
731 phylogenetic trees. *Molecular Biology and Evolution*. 4:406–425.

732 Sanderson MJ. 2003. r8s: inferring absolute rates of molecular evolution and divergence
733 times in the absence of a molecular clock. *Bioinformatics*. 19:301–302.

734 Schluter D, Price T, Mooers AØ, Ludwig D. 1997. Likelihood of ancestor states in
735 adaptive radiation. *Evolution*. 51:1699–1711.

736 Searle SR. 1987. Linear Models for Unbalanced Data. Wiley Series in Probability and
737 Statistics. Wiley.

738 Solís-Lemus C, Ané C. 2016. Inferring phylogenetic networks with maximum
739 pseudolikelihood under incomplete lineage sorting. *PLoS Genetics*. 12:e1005896.

740 Solís-Lemus C, Bastide P, Ané C. 2017. PhyloNetworks: a package for phylogenetic
741 networks. *Molecular Biology and Evolution*. .

742 Solís-Lemus C, Yang M, Ané C. 2016. Inconsistency of species tree methods under gene
743 flow. *Systematic Biology*. 65:843–851.

744 Uyeda JC, Harmon LJ. 2014. A novel Bayesian method for inferring and interpreting the
745 dynamics of adaptive landscapes from phylogenetic comparative data. *Systematic*
746 *Biology*. 63:902–918.

747 Yu Y, Degnan JH, Nakhleh L. 2012. The probability of a gene tree topology within a
748 phylogenetic network with applications to hybridization detection. *PLoS Genetics*.
749 8:e1002660.

750 Yu Y, Dong J, Liu KJ, Nakhleh L. 2014. Maximum likelihood inference of reticulate
751 evolutionary histories. *PNAS*. 111:16448–16453.

752 Yu Y, Nakhleh L. 2015. A maximum pseudo-likelihood approach for phylogenetic
753 networks. *BMC Genomics*. 16:S10.

754 Proof of the Variance Formula and Algorithm

We prove here both formula (1) for the BM variance matrix and Proposition 1 giving an efficient algorithm to calculate this matrix. We do so by induction on the number of nodes in the network: $N = n + m$. When the network is made of a single node $i = 1$, equation (1) and Proposition 1 are obviously correct. We now assume that these results are correct for any phylogenetic network with up to $N - 1$ nodes, and we consider a network with N nodes. When these nodes are sorted in preorder, the last node $i = N$ is necessarily a tip (with no descendants), so removing it and its parent edges from the original network gives a valid phylogenetic network with $N - 1$ nodes. Using the same notations as in the main text, we can focus on the case $i = N$. Because of the preorder, there is no directed path from i to j for any $j < i$. We use here the formulas of Definition 1, and assume $\sigma^2 = 1$ without loss of generality.

- If i is a tree node, then $X_i = X_a + \epsilon$, with $\epsilon \sim \mathcal{N}(0, \ell_{e_a})$, ϵ independent of the values X_j in the subnetwork ($j < i$). Moreover, $a < i$ because of the preorder. Then:

$$\mathbb{C}\mathrm{ov}\left[X_i; X_j\right] = \begin{cases} \mathbb{C}\mathrm{ov}\left[X_a; X_j\right] & \text{if } j < i \\ \mathbb{C}\mathrm{ov}\left[X_a; X_a\right] + \ell_{e_a} & \text{if } j = i \end{cases}$$

  and all the needed quantities on the right-hand side have already been computed because $a < i$. This proves (3) in Proposition 1. Next, we seek to prove (1). Note that it is valid by induction for all nodes in the subnetwork, and we just need to prove it for $i = N$ and any $j \leq i$. By induction, we have that, for any $j < i$,

$$\mathbb{C}\mathrm{ov}\left[X_a; X_j\right] = \sum_{\substack{p_a \in \mathcal{P}_a \\ p_j \in \mathcal{P}_j}} \left(\prod_{e \in p_a} \gamma_e\right)\left(\prod_{e \in p_j} \gamma_e\right) \sum_{e \in p_a \cap p_j} \ell_e \,.$$

  Because a is the only parent of node $i = N$, any paths from the root to i must start as a path from the root to a, and then follow e_a between a and i. In other words, any

774 path from the root to a corresponds to a unique path from the root to i:

$$\mathcal{P}_i = \{p_i = (p_a, e_a) : p_a \in \mathcal{P}_a\} \ .$$

775 Moreover, the inheritance weight of path p_a and $p_i = (p_a, e_a)$ are the same, because e_a
776 is a tree edge with $\gamma_{e_a} = 1$: $\prod_{e \in p_i} \gamma_e = \prod_{e \in p_a} \gamma_e$. Now take $j < i$. Any path p_j from
777 the root to j cannot go through i (because of the preorder), therefore it cannot go
778 through e_a, and the edges shared by p_i and p_j are exactly the same as the edges
779 shared by p_a and p_j. Putting these considerations together, we get:

$$\begin{aligned}
\mathbb{C}\mathrm{ov}\left[X_i; X_j\right] = \mathbb{C}\mathrm{ov}\left[X_a; X_j\right] &= \sum_{\substack{p_a \in \mathcal{P}_a \\ p_j \in \mathcal{P}_j}} \left(\prod_{e \in p_a} \gamma_e\right)\left(\prod_{e \in p_j} \gamma_e\right) \sum_{e \in p_a \cap p_j} \ell_e \\
&= \sum_{\substack{p_i \in \mathcal{P}_i \\ p_j \in \mathcal{P}_j}} \left(\prod_{e \in p_i} \gamma_e\right)\left(\prod_{e \in p_j} \gamma_e\right) \sum_{e \in p_a \cap p_j} \ell_e \ ,
\end{aligned}$$

780 which proves (1) for $i = N$ and $j < i$. For $j = i$, any path p_j from the root to $j = i$
781 must go through a and e_a, so that the shared edges between p_i and p_j are the edges
782 shared by p_a and p_j, plus edge e_a. Therefore, we get that

$$\begin{aligned}
\mathbb{C}\mathrm{ov}\left[X_i; X_i\right] &= \mathbb{C}\mathrm{ov}\left[X_a; X_a\right] + \ell_{e_a} \\
&= \sum_{\substack{p_1 \in \mathcal{P}_a \\ p_2 \in \mathcal{P}_a}} \left(\prod_{e \in p_1} \gamma_e\right)\left(\prod_{e \in p_2} \gamma_e\right) \sum_{e \in p_1 \cap p_2} \ell_e \ + \ \ell_{e_a} \\
&= \sum_{\substack{p_1 \in \mathcal{P}_i \\ p_2 \in \mathcal{P}_i}} \left(\prod_{e \in p_1} \gamma_e\right)\left(\prod_{e \in p_2} \gamma_e\right)\left(\left(\sum_{e \in p_1 \cap p_2} \ell_e\right) - \ell_{e_a}\right) \ + \ \ell_{e_a} \\
&= \sum_{\substack{p_1 \in \mathcal{P}_i \\ p_2 \in \mathcal{P}_i}} \left(\prod_{e \in p_1} \gamma_e\right)\left(\prod_{e \in p_2} \gamma_e\right) \sum_{e \in p_1 \cap p_2} \ell_e
\end{aligned}$$

where the last equality follows from $\sum_{p\in\mathcal{P}_i}\left(\prod_{e\in p}\right)=1$. This completes the proof of (1), for $i=j$.

- If i is a hybrid node, then $X_i=(\gamma_{e_a}X_a+\gamma_{e_b}X_b)+(\gamma_{e_a}\epsilon_a+\gamma_{e_b}\epsilon_b)$, with $\epsilon_k\sim\mathcal{N}(0,\ell_{e_k})$, and ϵ_k independent of the all values X_j in the subnetwork ($j<i$) for $k=a$ and $k=b$. Again, $a<i$ and $b<i$ because of the preorder. Then:

$$\mathbb{C}\mathrm{ov}\left[X_i;X_j\right]=\begin{cases}\gamma_{e_a}\mathbb{C}\mathrm{ov}\left[X_a;X_j\right]+\gamma_{e_b}\mathbb{C}\mathrm{ov}\left[X_b;X_j\right] & \text{if } j<i\\ \gamma_{e_a}^2\left(\mathbb{C}\mathrm{ov}\left[X_a;X_a\right]+\ell_{e_a}\right)+\gamma_{e_b}^2\left(\mathbb{C}\mathrm{ov}\left[X_b;X_b\right]+\ell_{e_b}\right) & \\ \quad+2\gamma_{e_a}\gamma_{e_b}\mathbb{C}\mathrm{ov}\left[X_a;X_b\right] & \text{if } j=i\,.\end{cases}$$

This proves (4) in Proposition 1. Next, we focus on proving (1). Again, it is valid by induction for all nodes in the subnetwork, and we need to prove it for $i=N$ and any $j\leq i$. By induction, (1) holds for a, b, and any $j<i$. Then, because a and b are the only parents of i, any path p_i from the root to i must go through a and e_a, or through b and e_b (and not both). In other words:

$$\mathcal{P}_i=\{(p_a,e_a):p_a\in\mathcal{P}_a\}\cup\{(p_b,e_b):p_b\in\mathcal{P}_b\}\,.$$

Now considering node $j<i$ and a path p_j from the root to j, p_j cannot go through i so it cannot go through e_a or e_b. Therefore, the shared edges between p_j and $p_i=(p_a,e_a)$ are exactly the same edges as those shared between p_j and p_a, and the shared edges between p_j and $p_i=(p_b,e_b)$ are also the same as those shared between

797 p_j and p_b. For $j < i$, we get:

$$\begin{aligned}
&\sum_{\substack{p_i\in\mathcal{P}_i\\ p_j\in\mathcal{P}_j}} \left(\prod_{e\in p_i}\gamma_e\right)\left(\prod_{e\in p_j}\gamma_e\right)\sum_{e\in p_i\cap p_j}\ell_e \\
&\quad= \sum_{\substack{p_a\in\mathcal{P}_a\\ p_j\in\mathcal{P}_j}} \left(\prod_{e\in p_a}\gamma_e\right)\gamma_{e_a}\left(\prod_{e\in p_j}\gamma_e\right)\sum_{e\in p_a\cap p_j}\ell_e + \sum_{\substack{p_b\in\mathcal{P}_b\\ p_j\in\mathcal{P}_j}} \left(\prod_{e\in p_b}\gamma_e\right)\gamma_{e_b}\left(\prod_{e\in p_j}\gamma_e\right)\sum_{e\in p_b\cap p_j}\ell_e \\
&\quad= \gamma_{e_a}\mathbb{C}\mathrm{ov}\left[X_a; X_j\right] + \gamma_{e_b}\mathbb{C}\mathrm{ov}\left[X_b; X_j\right] && \text{by induction} \\
&\quad= \mathbb{C}\mathrm{ov}\left[X_i; X_j\right] && \text{from above,}
\end{aligned}$$

798 proving (1) for $i = N$ and $j < i$. For $j = i = N$, we similarly decompose the set of

799 paths $\mathcal{P}_i$ into two sets, either going through a or through b:

$$\begin{aligned}
&\sum_{\substack{p_1\in\mathcal{P}_i\\ p_2\in\mathcal{P}_i}} \left(\prod_{e\in p_1}\gamma_e\right)\left(\prod_{e\in p_2}\gamma_e\right)\sum_{e\in p_1\cap p_2}\ell_e \\
&\quad= \sum_{\substack{p_1\in\mathcal{P}_a\\ p_2\in\mathcal{P}_a}} \left(\prod_{e\in p_1}\gamma_e\right)\gamma_{e_a}\left(\prod_{e\in p_2}\gamma_e\right)\gamma_{e_a}\left(\left(\sum_{e\in p_1\cap p_2}\ell_e\right)+\ell_{e_a}\right) \\
&\qquad+ 2\times\sum_{\substack{p_1\in\mathcal{P}_a\\ p_2\in\mathcal{P}_b}} \left(\prod_{e\in p_1}\gamma_e\right)\gamma_{e_a}\left(\prod_{e\in p_2}\gamma_e\right)\gamma_{e_b}\sum_{e\in p_1\cap p_2}\ell_e \\
&\qquad+ \sum_{\substack{p_1\in\mathcal{P}_b\\ p_2\in\mathcal{P}_b}} \left(\prod_{e\in p_1}\gamma_e\right)\gamma_{e_b}\left(\prod_{e\in p_2}\gamma_e\right)\gamma_{e_b}\left(\left(\sum_{e\in p_1\cap p_2}\ell_e\right)+\ell_{e_b}\right) \\
&\quad= \gamma_{e_a}^2\left(\mathbb{C}\mathrm{ov}\left[X_a; X_a\right]+\ell_a\right) + 2\gamma_{e_a}\gamma_{e_b}\mathbb{C}\mathrm{ov}\left[X_a; X_b\right] + \gamma_{e_b}^2\left(\mathbb{C}\mathrm{ov}\left[X_b; X_b\right]+\ell_b\right) \\
&\quad= \mathbb{C}\mathrm{ov}\left[X_i; X_i\right] && \text{by induction, and from above.}
\end{aligned}$$

800 This completes the proof of (1), for $i = j$, and for the last case when i is a hybrid

801 node.

Variance Reduction

Here, we prove Formula (2). As in the main text, consider a time-consistent network. For tip i, let t_i be the length of any path from the root to i. If the history of tip i involves one or more reticulations then take any two paths p_i and q_i in $\mathcal{P}_i$. We have: $\sum_{e\in p_i\cap q_i}\ell_e < \sum_{e\in p_i}\ell_e = t_i$, with a strict inequality if p_i and q_i are different paths. Seeing $\pi_{p_i} = \prod_{e\in p_i}\gamma_e$ as the probability associated with the path p_i (with $\sum_{p_i\in\mathcal{P}_i}\pi_{p_i}=1$), we get from Equation (1):

$$V_{ii} < \sum_{p_i,q_i\in\mathcal{P}_i}\pi_{p_i}\pi_{q_i}t_i \leq t_i,$$

with the equality fulfilled if there is a unique path from the root to taxon i, i.e. if i has no hybrid ancestry.

Pagel's λ Variance

Proof of Proposition 2. In Equation 1, the first equation is straightforward, because all the edges shared by the paths to i and to j are internal edges, whose lengths are multiplied by λ. Now take a tip node i. The first step of the transformation ensures that i is a tree node. Let a be its parent node, and parent branch e_a. From the recursive formula given in Proposition 1, the variance at node i is proportional to:

$$V_{ii}(\lambda) = V(\lambda)_{aa} + \ell_{e_a}(\lambda) = \lambda V_{aa} + \lambda\ell_{e_a} + (1-\lambda)t_i = \lambda V_{ii} + (1-\lambda)t_i\ ,$$

hence the announced formulas. □

Shifted BM model with the Descendence Matrix

Proof of Formula (7). The shifts are fixed, so they do not impact the variance structure of the traits, and we only need to show that $\mathbb{E}[\mathbf{Y}] = \mathbf{T}\boldsymbol{\Delta}$. Here, we prove a slightly more general formula on the complete vector of trait values at all the nodes, that is: $\mathbb{E}[\mathbf{X}] = \mathbf{U}\boldsymbol{\Delta}$. The original equality is easily derived from this one by keeping the tip values only.

We show this equality recursively. Assume that the nodes are numbered in preorder. Denote by $\mathbf{U}^i$ the i^{th} row-vector of $\mathbf{U}$. Node $i = 1$ is the root, which is the descendant of no other node than itself, so

$$\mathbb{E}[X_1] = \mu = \Delta_1 = \mathbf{U}^1\boldsymbol{\Delta} \ .$$

We now assume that $\mathbb{E}[X_j] = \mathbf{U}^j\boldsymbol{\Delta}$ for all nodes $j < i$, and we seek to prove that this property is also true for node i.

- If i is a tree node, then denote by a its unique parent and by e_a the edge from a to i. For any node $k \neq i$, $\mathcal{P}_{k\to i} = \{(p_a, e_a) : p_a \in \mathcal{P}_{k\to a}\}$. Since e_a is a tree edge with $\gamma_{e_a} = 1$, we get from definition 3 that:

$$U_{ik} = \begin{cases} U_{ak} & \forall\ k \neq i \\ 1 & \text{if } k = i \ , \end{cases}$$

  hence

$$\mathbb{E}[X_i] = \mathbb{E}[X_a] + \Delta_i = \mathbf{U}^a\boldsymbol{\Delta} + \Delta_i = \mathbf{U}^i\boldsymbol{\Delta} \ .$$

- If i is a hybrid, then denote by a and b its two parents, by e_a and e_b the corresponding edges, with coefficients γ_{e_a} and γ_{e_b}. Then for any node $k \neq i$, we have:

$\mathcal{P}_{k\to i} = \{(p_a, e_a) : p_a \in \mathcal{P}_{k\to a}\} \cup \{(p_b, e_b) : p_b \in \mathcal{P}_{k\to b}\}$, and using definition 3:

$$U_{ik} = \begin{cases} \gamma_{e_a} U_{ak} + \gamma_{e_b} U_{bk} & \forall\ k \neq i \\ 1 & \text{if } k = i\ . \end{cases}$$

Since no shift can occur on the hybrid branches, $\Delta_i = 0$ by convention and:

$$\mathbb{E}\left[X_i\right] = \gamma_{e_a}\mathbb{E}\left[X_a\right] + \gamma_{e_b}\mathbb{E}\left[X_b\right] = \gamma_{e_a}\mathbf{U}^a\mathbf{\Delta} + \gamma_{e_a}\mathbf{U}^b\mathbf{\Delta} = \mathbf{U}^i\mathbf{\Delta}\ .$$

This ends the recursion, and the proof of (7). □

Note that this proof also gives an efficient recursive way to compute the descendence matrix $\mathbf{U}$.

Fisher Test for Transgressive Evolution

The Fisher statistics used in Section Transgressive Evolution have the following expression:

$$F_{10} = \frac{\left\|\mathbf{Y} - \mathrm{Proj}_{\mathbf{R}}\,\mathbf{Y}\right\|^2_{(\mathbf{V}^{\mathrm{tip}})^{-1}} - \left\|\mathbf{Y} - \mathrm{Proj}_{[\mathbf{R}\ \bar{\mathbf{N}}]}\,\mathbf{Y}\right\|^2_{(\mathbf{V}^{\mathrm{tip}})^{-1}}}{\left\|\mathbf{Y} - \mathrm{Proj}_{[\mathbf{R}\ \bar{\mathbf{N}}]}\,\mathbf{Y}\right\|^2_{(\mathbf{V}^{\mathrm{tip}})^{-1}}} \frac{n - r_{[\mathbf{R}\ \bar{\mathbf{N}}]}}{r_{[\mathbf{R}\ \bar{\mathbf{N}}]} - r_{\mathbf{R}}}$$

$$F_{21} = \frac{\left\|\mathbf{Y} - \mathrm{Proj}_{[\mathbf{R}\ \bar{\mathbf{N}}]}\,\mathbf{Y}\right\|^2_{(\mathbf{V}^{\mathrm{tip}})^{-1}} - \left\|\mathbf{Y} - \mathrm{Proj}_{[\mathbf{R}\ \mathbf{N}]}\,\mathbf{Y}\right\|^2_{(\mathbf{V}^{\mathrm{tip}})^{-1}}}{\left\|\mathbf{Y} - \mathrm{Proj}_{[\mathbf{R}\ \mathbf{N}]}\,\mathbf{Y}\right\|^2_{(\mathbf{V}^{\mathrm{tip}})^{-1}}} \frac{n - r_{[\mathbf{R}\ \mathbf{N}]}}{r_{[\mathbf{R}\ \mathbf{N}]} - r_{[\mathbf{R}\ \bar{\mathbf{N}}]}}$$

where $\mathrm{Proj}_{\mathbf{M}}$ denotes the projection onto the linear space spanned by the columns of matrix $\mathbf{M}$, with respect to the metric defined by $\mathbf{V}^{\mathrm{tip}}$: $\|\mathbf{X}\|^2_{(\mathbf{V}^{\mathrm{tip}})^{-1}} = \mathbf{X}^T(\mathbf{V}^{\mathrm{tip}})^{-1}\mathbf{X}$. In other words, for any vector $\mathbf{X}$:

$$\mathrm{Proj}_{\mathbf{M}}\,\mathbf{X} = \underset{\mathbf{U}\in\mathrm{Span}(\mathbf{M})}{\mathrm{argmin}}\ \|\mathbf{X} - \mathbf{U}\|^2_{(\mathbf{V}^{\mathrm{tip}})^{-1}} = \mathbf{M}(\mathbf{M}^T(\mathbf{V}^{\mathrm{tip}})^{-1}\mathbf{M})^{-1}\mathbf{M}^T(\mathbf{V}^{\mathrm{tip}})^{-1}\mathbf{X}\ .$$

These statistics follow a noncentral Fisher distribution as given in (9) and (10) of the main text, where

$$\begin{cases} \Delta^2_{10}(\mathbf{R}, \bar{\mathbf{N}}, \mathbf{V}^{\text{tip}}) = \left\|(\mathbf{I} - \text{Proj}_{\mathbf{R}})\bar{\mathbf{N}}\right\|^2_{(\mathbf{V}^{\text{tip}})^{-1}} \\ \Delta^2_{21}(\mathbf{d}, \mathbf{R}, \bar{\mathbf{N}}, \mathbf{N}, \mathbf{V}^{\text{tip}}) = \left\|(\mathbf{I} - \text{Proj}_{[\mathbf{R}\ \bar{\mathbf{N}}]})\mathbf{N}\mathbf{d}\right\|^2_{(\mathbf{V}^{\text{tip}})^{-1}} \end{cases} .$$

When studying the power of the test $\mathcal{H}_1$ vs $\mathcal{H}_2$, we took $\mathbf{d} = d\mathbf{d}^u$, so that the noncentral coefficient is:

$$\frac{1}{2\sigma^2}\Delta^2_{21}(\mathbf{d}, \mathbf{R}, \bar{\mathbf{N}}, \mathbf{N}, \mathbf{V}^{\text{tip}}) = \frac{d^2}{2\sigma^2}\left\|(\mathbf{I} - \text{Proj}_{[\mathbf{R}\ \bar{\mathbf{N}}]})\mathbf{N}\mathbf{d}_u\right\|^2_{(\mathbf{V}^{\text{tip}})^{-1}}$$

and, as the networks are fixed, it only varies with the heterogeneity coefficient d.

Note that a third statistic, F_{20}, can be defined in a similar way to test $\mathcal{H}_2$ vs $\mathcal{H}_0$ directly. We first re-write the linear model as:

$$\mathbf{Y} = \mathbf{R}\boldsymbol{\beta} + \mathbf{N}\boldsymbol{\delta} + \mathbf{E}\ , \quad \mathbf{E} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{V}^{\text{tip}})\ ,$$

where there are no constraints on coefficients $\boldsymbol{\delta}$. Then the F statistic can be written as:

$$F_{20} = \frac{\|\mathbf{Y} - \text{Proj}_{\mathbf{R}}\,\mathbf{Y}\|^2_{(\mathbf{V}^{\text{tip}})^{-1}} - \left\|\mathbf{Y} - \text{Proj}_{[\mathbf{R}\ \mathbf{N}]}\,\mathbf{Y}\right\|^2_{(\mathbf{V}^{\text{tip}})^{-1}}}{\left\|\mathbf{Y} - \text{Proj}_{[\mathbf{R}\ \mathbf{N}]}\,\mathbf{Y}\right\|^2_{(\mathbf{V}^{\text{tip}})^{-1}}} \frac{n - r_{[\mathbf{R}\ \mathbf{N}]}}{r_{[\mathbf{R}\ \mathbf{N}]} - r_{\mathbf{R}}}\ .$$

In the same way, it follows under $\mathcal{H}_2$ a noncentral Fisher distribution:

$$F_{20} \sim \mathcal{F}\left(r_{[\mathbf{R}\ \mathbf{N}]} - r_{\mathbf{R}}, n - r_{[\mathbf{R}\ \mathbf{N}]}, \frac{1}{2\sigma^2}\Delta^2_{20}(\mathbf{d}, \mathbf{R}, \mathbf{N}, \mathbf{V}^{\text{tip}})\right)\ ,$$

with

$$\Delta^2_{20}(\boldsymbol{\delta}, \mathbf{R}, \mathbf{N}, \mathbf{V}^{\text{tip}}) = \|(\mathbf{I} - \text{Proj}_{\mathbf{R}})\mathbf{N}\boldsymbol{\delta}\|^2_{(\mathbf{V}^{\text{tip}})^{-1}}\ .$$

Thank to the flexible framework provided by the GLM ftest function, all these tests are

857 readily implemented, as long as one can fit the three models ($\mathcal{H}_0$, $\mathcal{H}_1$, and $\mathcal{H}_2$).