

# An Assessment of Phylogenetic Tools for Analyzing the Interplay Between Interspecific Interactions and Phenotypic Evolution

J. P. DRURY<sup>1,2,\*</sup>, G. F. GRETHER<sup>2</sup>, T. GARLAND Jr.<sup>3</sup>, AND H. MORLON<sup>1</sup>

<sup>1</sup>Département de Biologie, Institut de Biologie, École Normale Supérieure, CNRS UMR 8197, 46 rue d'Ulm 75005 Paris, France; <sup>2</sup>Department of Ecology & Evolutionary Biology, University of California, 621 Charles E. Young Dr. S, Los Angeles, CA 90095-1606, USA; and <sup>3</sup>Department of Evolution, Ecology, and Organismal Biology, University of California, 900 University Ave., Riverside, CA 92521, USA

\*Correspondence to be sent to: Department of Ecology & Evolutionary Biology, University of California, 621 Charles E. Young Dr. S, Los Angeles, CA 90095-1606, USA; E-mail: druryj@ucla.edu.

Received 25 October 2016; reviews returned 20 September 2017; accepted 21 September 2017

Associate Editor: Luke Harmon

**Abstract.**—Much ecological and evolutionary theory predicts that interspecific interactions often drive phenotypic diversification and that species phenotypes in turn influence species interactions. Several phylogenetic comparative methods have been developed to assess the importance of such processes in nature; however, the statistical properties of these methods have gone largely untested. Focusing mainly on scenarios of competition between closely-related species, we assess the performance of available comparative approaches for analyzing the interplay between interspecific interactions and species phenotypes. We find that many currently used statistical methods often fail to detect the impact of interspecific interactions on trait evolution, that sister-taxon analyses are particularly unreliable in general, and that recently developed process-based models have more satisfactory statistical properties. Methods for detecting predictors of species interactions are generally more reliable than methods for detecting character displacement. In weighing the strengths and weaknesses of different approaches, we hope to provide a clear guide for empiricists testing hypotheses about the reciprocal effect of interspecific interactions and species phenotypes and to inspire further development of process-based models. [Character displacement; competition; interspecific interactions; phylogenetic comparative methods; trait evolution.]

Interactions between species are a fundamental aspect of life on earth, and understanding the evolutionary and ecological consequences of such interactions are a central goal of many classical theoretical frameworks in ecology and evolutionary biology. Identifying both the predictors of interspecific interactions and the consequences of such interactions for diversification and coexistence is thus an important contemporary research area (Weber et al. 2017), with strong implications for conservation biology.

Several phylogenetic comparative methods have been deployed with the goal of elucidating how interspecific interactions drive (or are driven by) character evolution, but the reliability and efficacy of these methods remain largely untested. Here, we focus on methods used to study interactions between closely related species (e.g., members of the same family) that arise from similarity in morphology, signaling traits, or habitat (Brown and Wilson 1956; Schluter 2000; Pfennig and Pfennig 2009), rather than on community-wide interactions and interaction networks (Webb et al. 2002; Rezende et al. 2007; Cavender-Bares et al. 2009; Cadotte et al. 2013).

Classical character displacement theory (Brown and Wilson 1956; Grether et al. 2009; Pfennig and Pfennig 2009) predicts that, where heterospecifics compete, selection should favor divergence in the traits responsible for competition, until lineages in sympatry no longer compete intensely. In a seminal example, selection resulting from exploitative competition between medium and large ground finches (*Geospiza fortis* and *G. magnirostris*) has driven bill size divergence on Daphne Major in the Galápagos (Grant and Grant 2006). Investigators who conduct comparative studies of divergent character

displacement often test the prediction that coexisting species will be more phenotypically divergent than noncoexisting ones by looking for a relationship between biogeographic overlap and trait dissimilarity. Recent studies on *Bicyclus* butterflies and *Euglossa* bees, for example, show that male chemical cues are more distinct between sympatric species than allopatric species, suggesting that reproductive character displacement has driven signal divergence in these taxa (Bacquet et al. 2015; Weber et al. 2016).

Interspecific interactions can also lead to convergent, rather than divergent, character displacement (Cody 1969, 1973; Grant 1972; Grether et al. 2013). Agonistic character displacement theory (Grether et al. 2013) predicts convergence in traits mediating interspecific aggression when species compete strongly for the same resources. In other words, between-species similarity in resource use may make interspecific territoriality adaptive, resulting in subsequent convergence in signaling traits involved in mediating territorial interactions (e.g., song in ovenbirds, Tobias et al. 2014). Therefore, tests of convergent character displacement typically test the prediction that sympatric lineages are more phenotypically similar than allopatric ones. Because sympatric similarity can also result from convergence to local conditions (e.g., habitat, climate), it is important for empiricists to account for abiotic factors in tests of character convergence.

In some instances, rather than identifying the effect of species interactions on trait evolution, empiricists aim to identify traits that mediate particular pairwise interactions, such as hybridization or interspecific aggression. In this case, investigators test for a

relationship between the measured interactions and trait similarity. Recent studies on New World warblers (Parulidae), for example, show that hybridization occurs more often between species with similar songs and that interspecific territoriality occurs more often between species that share similar plumage and territorial song phenotypes (Willis et al. 2014; Losin et al. 2016).

Although the examples presented here largely represent scenarios where interactions between species are competitive, empiricists apply methods discussed here to other noncompetitive interactions as well (e.g., predicting links in plant/pollinator networks and identifying Müllerian mimicry rings: Elias et al. 2008; Eklöf and Stouffer 2016). Regardless of the biological question, a particularity of comparative tests aimed at understanding the interplay between interspecific interactions and species phenotypes is that they largely involve testing correlations between pairwise data (e.g., range overlap, phenotypic similarity, and frequency of hybridization). In contrast, most phylogenetic comparative methods have been developed and tested on tip data (e.g., range size and morphological trait values), and the statistical properties of methods adapted to handle pairwise data (Appendix 1) have gone untested (but see Harmon and Glor 2010). Furthermore, species interactions are inherently affected by the biogeographic history of dispersal and speciation in an evolving clade and the resulting patterns of range overlap. Patterns of trait dissimilarity between sympatric lineages—the classic test of character displacement—may actually be the null expectation if allopatric speciation is the norm, because then sympatric species pairs will tend to be more distantly related than allopatric species pairs (Weir and Price 2011; Tobias et al. 2014).

Here, we apply the main phylogenetic comparative methods that investigators use to test hypotheses about interactions between closely related lineages and phenotypes (Appendix 1, Fig. 1) to data sets simulated under different evolutionary histories of speciation, dispersal, species interactions, and trait evolution. We then compare the efficacy of these methods, discuss the relative merits of each, and outline directions for future research.

## METHODS

We compared the performance of different phylogenetic comparative methods by measuring their statistical power (e.g., probability of detecting divergence when divergence is simulated) and Type I error rate (e.g., probability of detecting an effect of species interactions when such an effect is not simulated) across three scenarios.

### *Phylogeny and Range Simulations*

We jointly simulated trees (# spp. = 20, 50, 100, 150, 200, 250) and biogeographies under the dispersal-extinction-cladogenesis model of biogeographical evolution (i.e.,

DEC+J, with the inclusion of founder event speciation) in BioGeoBEARS (Ree and Smith 2008; Matzke 2014). Briefly, the DEC+J model is a model of range evolution in which species ranges change along the branches of a phylogeny as a function of dispersal and local extinction and are inherited by daughter taxa at speciation according to several possible cladogenetic scenarios (see more details in Supplementary Methods available on Dryad at <http://dx.doi.org/10.5061/dryad.ch0vn>). For each tree, we started with a single ancestral species occupying one of ten equidistant regions, and simulated trees with constant rates of speciation and local extinction. We considered different biogeographic scenarios by varying the rate of dispersal events between ranges (“high” and “low” dispersal; see details in Supplementary Methods available on Dryad) and the probability that speciation events occur in sympatry versus allopatry (“high” and “low” sympatric speciation; Supplementary Methods available on Dryad). Each of these simulations resulted in a phylogeny (the tree of extant species) and its associated biogeography (the set of regions in which each lineage occurred throughout the history of the clade). Lineages were identified as sympatric if they co-occurred in at least one of the ten geographic regions, and allopatric if they did not co-occur in any.

We simulated four biogeographic scenarios (combinations of low or high dispersal and low or high sympatric speciation) for each tree size. The resulting biogeographies span scenarios where sympatric speciation is common and dispersal is low (e.g., lizards on islands) to scenarios where allopatric speciation is the main mode of speciation and dispersal between regions is high (e.g., birds on continents). These parameter combinations produced a range of realistic proportions of sister taxa that are sympatric (Supplementary Fig. S1A available on Dryad) and a range of realistic differences in age between sympatric and allopatric sister taxa (Pigot and Tobias 2014; Supplementary Fig. S1B available on Dryad), at least for animal taxa (but see Anacker and Strauss 2014 for different patterns in plants). In defining sympatry as any overlap, the mean magnitude of range overlap fell between 33% and 42% across all tree sizes and simulation parameters (Supplementary Fig. S1C,D available on Dryad), which falls well within the range of overlap of sympatric taxa defined under commonly used minimum threshold values applied to continuous indices of range overlap (e.g., Pigot and Tobias 2014; Tobias et al. 2014).

For each combination of tree sizes and DEC parameter combinations ( $n=24$ ), we performed 100 simulations, resulting in a bank of 2400 trees with associated biogeographies.

### *Character Displacement*

*The model.*—To simulate both divergent and convergent character displacement, we simulated a continuous trait  $z$  under a model in which trait values of sympatric species in an evolving clade are repelled from (or

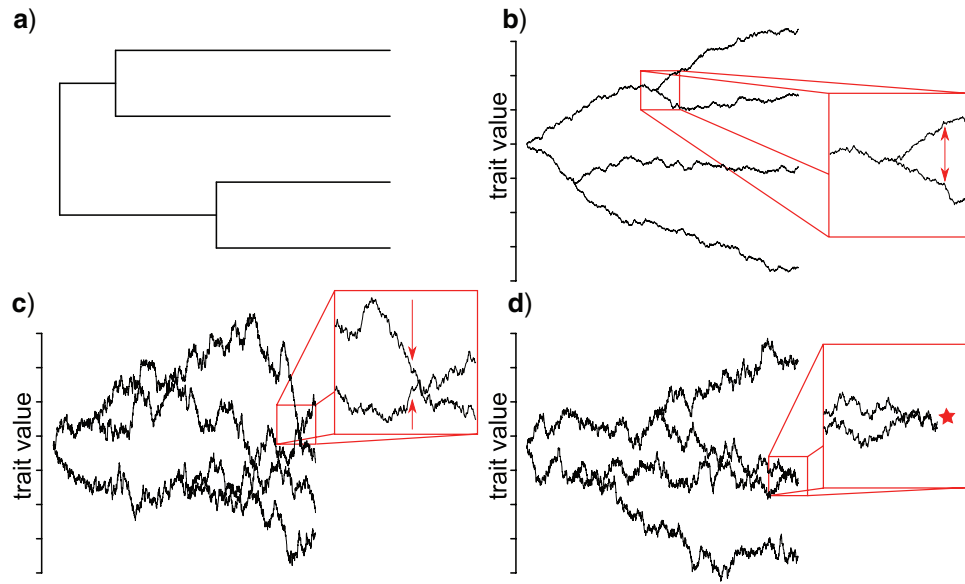


FIGURE 1. Schematic examples of the processes examined in our simulation study. a) Phylogeny along which the trait evolves, b) a trait evolving via divergent character displacement, c) a trait evolving via convergent character displacement, and d) a species interaction that exists at present due to pairwise trait similarity. For simulation details, see the main text and Supplementary Methods of the Supplementary Material available on Dryad.

drawn toward) one another. In divergent character displacement, trait divergence is driven by pairwise similarity in that same trait  $z$ ; in convergent character displacement however, convergence in trait  $z$  (e.g., a signaling trait) is driven by pairwise similarity in another trait  $y$  (e.g., a resource use trait). To create a generic model of character displacement, we thus modified the matching competition (MC) model (Nuismer and Harmon 2015; Drury et al. 2016) by describing the mean value for trait  $z$  in lineage  $i$  after an infinitesimally small time step  $dt$  by:

$$z_i(t+dt) = z_i(t) + \psi(\theta - z_i(t))dt + m \left[ \sum_{j \neq i}^n \mathbf{A}_{i,j} \times \text{sign}(z_i(t) - z_j(t)) \times e^{-\alpha(y_i(t) - y_j(t))^2} \right] dt + \delta \quad (1)$$

where  $y=z$  in the case of divergent character displacement and  $y \neq z$  in the case of convergent character displacement,  $\psi(\theta - z_i(t))$  describes attraction to a single stationary peak [i.e., the Ornstein–Uhlenbeck (OU) process, Felsenstein 1988; Garland et al. 1993; Hansen and Martins 1996],  $n$  is the number of species,  $\delta$  is a random variable with mean 0, and variance  $=\sigma^2 dt$  [the Brownian motion (BM) rate parameter, describing the stochastic component of trait evolution], and  $\mathbf{A}$  is a piecewise-constant matrix representing biogeographical overlap such that  $\mathbf{A}_{i,j}$  equals 1 if species  $i$  and  $j$  are sympatric at time  $t$ , and 0 otherwise. The “sign” portion determines the relative position of each species in trait space (i.e., it equals +1 if  $z_i$  is larger than  $z_j$ , and –1 otherwise). The  $\alpha$  value ( $\alpha > 0$ ) determines the effect of

pairwise similarity in trait  $y$  on competition: if  $\alpha$  is close to zero, all lineages sympatric with lineage  $i$  have the same competitive effect on  $i$ , regardless of their similarity in trait  $y$ ; conversely, if  $\alpha$  is large, sympatric lineages similar to  $i$  in terms of the  $y$  trait will have a much stronger competitive effect on  $i$  than sympatric lineages dissimilar to  $i$  in terms of the  $y$  trait. The parameter  $m$  represents the magnitude of the effect of competition when two lineages have identical  $y$  values (i.e., it provides an upper bound for the deterministic effect of competition). When  $m=0$ , this equation reduces to an OU model, whereas positive  $m$  values result in pairwise divergence and negative values result in pairwise convergence. When both  $m$  and  $\psi=0$ , this model reduces to BM. For additional simulation details, see Supplementary Methods available on Dryad.

We use a lineage-based “phenomenological” model for our simulations rather than an individual-based model to have the computational ability to produce data sets of a size comparable to the maximum sometimes reached in empirical comparative phylogenetic studies (i.e., often reaching several hundreds of species). Models derived from microevolutionary first principles (e.g., Grether et al. 2009; Nuismer and Harmon 2015) generate similar patterns of sympatric shifts resulting from character displacement, and using such a model here would be much more computationally intensive, therefore restricting the range of parameter values that can be studied. For simplicity, this model also omits the effect of a species’ geographic structure and the effect of gene flow between distinct populations on the evolution of the mean species phenotype. This simplification is reasonable in the context of our study because there is no reason to expect that it will systematically bias the

patterns generated in such a way as to yield different conclusions regarding the performance of the various analytical approaches that we use here. Finally, in all of our simulations, we considered sympatry to be a binomial variable, so  $A_{i,j}$  equaled either 1 (if species  $i$  and  $j$  are sympatric) or 0 (if species  $i$  and  $j$  are allopatric). This index of sympatry is similar to commonly used indices (Pigot and Tobias 2014; Tobias et al. 2014), but other formulations of sympatry, such as continuous measurements of range overlap (Bothwell et al. 2015; Martin et al. 2015) are also possible. We did not explore continuous measurements of range overlap here, but have uploaded our simulation scripts to RPANDA (Morlon et al. 2016; <https://github.com/hmorlon/PANDA>), which could easily be modified to do so.

*Divergent character displacement.*—We simulated data sets with divergent character displacement by setting  $y = z$  in Eq. 1 such that trait divergence is driven by pairwise similarity in that trait. Biologically, this could represent a feeding trait that covaries with resource use (e.g., bill shape in Galápagos finches, Grant and Grant 2011) and which would directly affect interspecific competition. To assess whether each method could detect divergent character displacement when it occurred and did not erroneously detect character displacement when it was absent, we simulated data sets both with repulsion ( $m = 2$ ) and without repulsion ( $m = 0$ ) (see Supplementary Methods available on Dryad). We also simulated data sets with  $\{\psi = 2\}$  and without  $\{\psi = 0\}$  the OU process. In all simulations, we held  $\sigma^2$  constant at 0.5,  $\alpha$  constant at 1, and both the state at the root ( $z_0$ ) and the OU optimum ( $\theta$ ) constant at 0.

In additional simulations run only on 100-species trees, we analyzed the effect of both the maximum strength of repulsion ( $m = 0, 1, 2, 10$ ) and, to understand how the opposing forces of repulsion and attraction to an optimum influence analyses, the ratio of attraction to the maximum effect of competition ( $\psi:m = 0, 0.2, 0.5, 1$ ). To achieve these ratios of  $\psi:m$ , we varied  $\psi$  while holding  $m$  constant (e.g., for the case where  $m = 2$ , we simulated data sets where  $\psi = 0, 0.4, 1$ , and 2, respectively). As above, these values were arbitrarily chosen based on visual inspection of realized simulations.

For each parameter combination, we simulated 10 data sets for each tree, resulting in 1000 simulations for each tree size/biogeographic scenario combination.

*Convergent character displacement.*—We simulated data sets with convergent character displacement under Eq. 1, where the term  $y$  represents a trait determining resource use or niche occupation evolving via BM or OU. A species' trait  $z$  in this model—a trait used as a territorial signal—is thus attracted most strongly to the signal trait values of sympatric lineages with the most similar resource-use traits. Biologically, this represents a scenario where selection favors interspecific territoriality—mediated by similarity in territorial signals—because the benefits of excluding heterospecifics are similar to the benefits of excluding conspecifics

(Grether et al. 2009). As a species' resource-use trait becomes less similar to that of sympatric species, the strength of attraction decreases to zero.

We simulated resource-use traits under both BM ( $\sigma_{\text{resource}}^2 = 0.5$ ,  $\psi_{\text{resource}} = 0$ ) and OU ( $\sigma_{\text{resource}}^2 = 0.5$ ,  $\psi_{\text{resource}} = 2$ ,  $\theta_{\text{resource}} = 0$ ) models. For the signal trait, we simulated data sets both with convergence ( $m = -0.25$ ) and without convergence ( $m = 0$ ). We did not include attraction toward a stable peak for the signal trait (i.e.,  $\psi$  was held constant at 0). As above, we held  $\sigma^2 = 0.5$  and  $z_0 = 0$ , though we held  $\alpha$  constant at 10, since smaller values result in rapid, cladewise convergence in traits. To analyze the effect of the maximum strength of convergence, we ran another set of simulations on 100-species trees varying  $m$  ( $m = 0, -0.1, -0.25, -0.5$ ) (see Supplementary Methods available on Dryad). The resource trait ( $y$ ) and signal trait ( $z$ ) were modeled as unlinked and genetically uncorrelated.

As above, we simulated 10 data sets for each tree, resulting in 1000 simulations for each tree size/biogeographic scenario combination.

#### Predictors of Interspecific Interactions

In some cases, investigators wish to identify which factors explain the occurrence of particular interspecific interactions. For example, investigators may want to understand which traits cause species to hybridize (e.g., Willis et al. 2014). In this scenario, species interactions vary according to phenotypic similarity between sympatric species pairs (i.e., species pairs that could potentially interact). Additionally, and unlike character displacement analyses, predicting the occurrence of interspecific interactions requires treating trait similarity as a predictor variable rather than a response variable. Thus, we generated data sets where the presence of interactions between sympatric taxa depends on pairwise similarity in traits.

First, we simulated the independent evolution of two traits along the phylogeny. One of these traits (Trait 1) represents the measured, focal trait: the investigator wants to know if this trait (e.g., plumage color) affects interactions (e.g., hybridization). The other trait (Trait 2) represents an uncorrelated trait (e.g., song) that potentially also affects interactions but is not the focal trait, and is not necessarily measured. We simulated this second trait in order to check whether the effect of a nonfocal trait on interactions could be misinterpreted as the effect of the focal trait (a sort of Type I error) and also to determine how the effect of an unmeasured trait on interactions affects the ability to identify an effect due to the measured trait. We evolved Trait 1 under a BM ( $\sigma^2 = 0.5$ ,  $\psi = 0$ ) or OU ( $\sigma^2 = 0.5$ ,  $\psi = 2$ ,  $\theta = 0$ ) model and Trait 2 under a BM model ( $\sigma_{\text{unmeasured}}^2 = 1$ ,  $\psi_{\text{unmeasured}} = 0$ ).

Next, we generated data sets where the probability  $P$  for two species to interact depends on similarity in trait

space at the present:

$$P = \frac{e^{(b_1 D + b_2 D_2)}}{1 + e^{(b_1 D_1 + b_2 D_2)}} \quad (2)$$

(e.g., Hilbe 2009) where  $D_n$  is the trait distance between species (i.e., distance between tip values) in simulated trait  $n$  (simulated using fastBM in phytools, Revell 2012), and  $b_n$  is the coefficient determining the magnitude of the relationship between the species interaction and similarity in trait  $n$ . As the effect of  $b_n$  on species interaction depends on  $D_n$ , which in turn depends on the total height of the tree, we scaled the trees to a height of one prior to simulating data sets to facilitate comparison of results across trees and parameter space.

To determine the power to identify an effect of trait similarity on interactions, we generated species interactions based on similarity in the focal trait ( $b_1 = -4, b_2 = 0$ ). To assess the Type I error rate, we simulated species interactions based on similarity in the nonfocal trait ( $b_1 = 0, b_2 = -4$ ). It is also possible that both the focal trait and an unmeasured trait influence species interactions. To determine how the effect of an unmeasured trait on interactions affects the ability to identify an effect due to the measured trait, we ran another set of simulations on 100-species trees varying  $b_1$  ( $b_1 = 0, -2, -4, -6, -8$ ) and holding  $b_2 = -4$ . As above, we ran 1000 simulations for each tree size/biogeographic scenario combination.

#### Phylogenetic Tests

Among our tests of character displacement (both divergent and convergent), the “correlation” tests involved assessing the significance of the relationship between phenotypic similarity and coexistence, using either the “full” data set (all species pairs) or the “sister taxa” subset obtained by culling sister taxa from trees with  $\geq 150$  tips (Appendix 1, Diagram S1). To the full data sets, we applied standard nonphylogenetic regression analyses that ignore phylogenetic nonindependence (Appendix 1), the raw and phylogenetically permuted partial Mantel tests (Appendix 1), phylogenetic linear mixed models (PLMMs, Appendix 1), and the simulation approach (Appendix 1, Supplementary Methods available on Dryad). To the sister-taxa data sets, we applied nonphylogenetic regression analyses (Appendix 1), PLMMs (Appendix 1), the simulation approach (Appendix 1), sister-taxa GLMs (Appendix 1), and fit process based models in EvoRAG (Appendix 1, Supplementary Methods available on Dryad). We did not perform Mantel tests on the sister-taxa data because such tests require complete matrices and distance matrices with data for only sister taxa would mostly contain empty cells (i.e., all those cells that correspond to nonsister taxa species pairs). We compared the fit of process-based phenotypic models with and without species interactions (BM, OU, diversity dependent, and MC models; see Appendix 1 and Supplementary Methods available on Dryad) to the full data sets from divergence scenarios using the R packages geiger (Pennell et al.

2014) and RPANDA (Morlon et al. 2016). We acknowledge that diversity-dependent models were not designed to analyze character displacement *per se*, but because they incorporate interspecific interactions, we hypothesized that (and wanted to test if) they could be useful in doing so. We did not apply process-based models to convergence scenarios because the necessary model fitting tools have yet to be developed (see Discussion).

Our tests of predictors of species interactions involved assessing the significance of the relationship between phenotypic similarity and species interactions (i.e., whether the species interact where they occur in sympatry). Since the response variable is binary, we fit nonphylogenetic logistic regressions, logistic PLMMs, and employed the simulation approach (see Supplementary Methods available on Dryad). We did not perform Mantel tests or sister-taxa analyses because the species pair matrix was incomplete (species that do not coexist cannot interact) and typically too few sister taxa occurred in sympatry for regression analysis.

## RESULTS

### Divergent Character Displacement

When all possible pairwise comparisons are included in analyses, the ability of most methods to detect divergent character displacement in simulated data sets depends on the presence of the OU process. As expected, nonphylogenetic regression analyses have high Type I error rates [Fig. 2Ai,iv, Supplementary Fig. S2Ai,iv available on Dryad (NB: throughout, results for low sympatric speciation biogeographies are plotted in the main text and high sympatric speciation biogeographies in the supplement)]. When the OU process is present ( $\psi = 2$ ), all phylogenetic methods generally have low Type I error rates and high power (Fig. 2Aiv–vi, Supplementary Fig. S2iv–vi and Tables available on Dryad). However, when there is no pull toward a peak ( $\psi = 0$ ), the Type I error rate is higher for Mantel tests (Fig. 2Ai,ii, Supplementary Fig. S2i,ii available on Dryad), and the power is much lower for all methods, though the pppMantel and raw Mantel perform better than the simulation and PLMM methods (Fig. 2Aii, Supplementary Fig. S2iii available on Dryad). Repulsion is easier to detect against an OU background of traits converging toward a common optimum than against a background of traits diverging under BM, likely because the repulsion process is more active when species occupy similar trait space (Supplementary Figs. S3 and S4 available on Dryad). High rates of sympatric speciation and dispersal tend to slightly decrease the power of all methods (Supplementary Fig. S2iii,vi and Tables available on Dryad).

The ability to detect divergence was relatively similar for  $m = 1$  and  $m = 2$ , but declined for  $m = 10$  (Supplementary Fig. S5 available on Dryad). This is due to a positive relationship between the ability to detect character displacement and the ratio of  $\psi:m$  (Supplementary Fig. S6 available on Dryad), resulting from a higher absolute

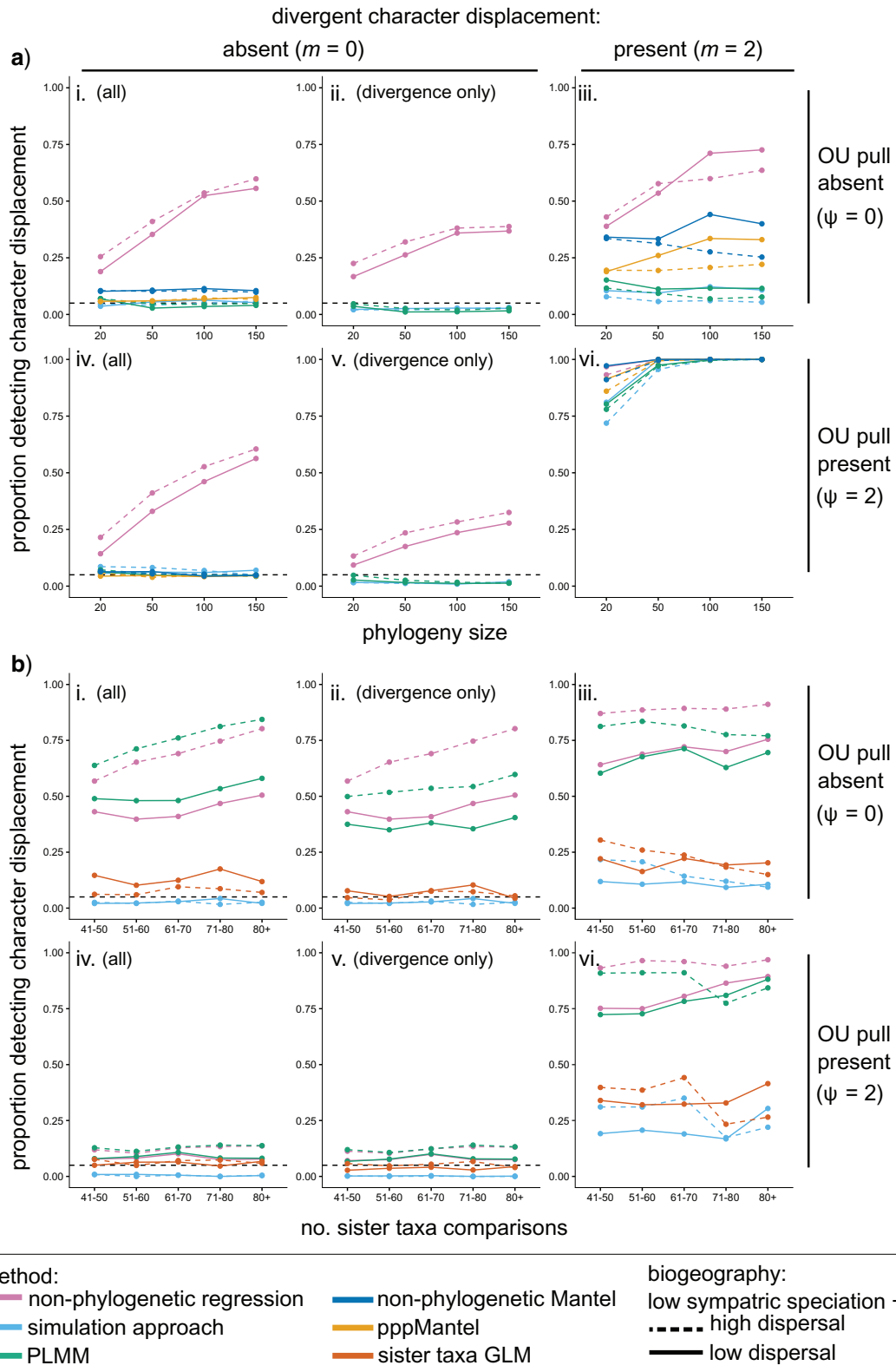


FIGURE 2. Proportion of statistically significant analyses in data sets simulated under divergent character displacement in biogeographic scenarios with low sympatric speciation rates. a) Results from approaches using data from all pairwise comparisons in a clade, plotted as a function of the phylogeny size and dispersal rate when (i–ii)  $m=0$  and  $\psi=0$  [(i) all analyses and (ii) only analyses returning divergence in sympatry], (iii)  $m=2$  and  $\psi=0$ , (iv–v)  $m=0$  and  $\psi=2$  [(iv) all analyses and (v) only analyses returning divergence in sympatry], and (vi)  $m=2$  and  $\psi=2$ . b) Results from analyses of sister taxa culled from complete phylogenies binned by the number of resulting species pairs, plotted as a function of the number of sister taxa comparisons and dispersal rate when (i–ii)  $m=0$  and  $\psi=0$  [(i) all analyses and (ii) only analyses returning divergence in sympatry], (iii)  $m=2$  and  $\psi=0$ , (iv–v)  $m=0$  and  $\psi=2$  [(iv) all analyses and (v) only analyses returning divergence in sympatry], and (vi)  $m=2$  and  $\psi=2$ . For scenarios where  $m=2$ , only the proportion of significant results showing divergence are plotted. Dashed horizontal lines represent a Type I error rate of 5%.

magnitude of repulsion when both processes are present (Supplementary Figs. S4 and S6 available on Dryad), indicating that this ratio impacts the ability to detect divergence more than the raw value of  $m$ .

For several analyses using only sister-taxa comparisons, there is a high probability of falsely concluding that character displacement occurred in data sets simulated under BM and, to a lesser extent, OU, when data are analyzed with simple linear regressions or PLMMs (Fig. 2Bi,iv, Supplementary Fig. S2Bi,iv available on Dryad). As with the whole-tree approach, the power tends to increase and Type I error rate tends to decrease in data sets with attraction toward a single-stationary peak (Fig. 2Biv–vi, Supplementary Fig. S2Biv–vi available on Dryad). However, the overall power to infer the presence of divergence was low ( $<0.8$ ) with analyses conducted on sister taxa (Fig. 2Biii,vi, Supplementary Fig. S2Biii,vi available on Dryad, Table 2), regardless of the analytical approach used. Inferences were generally better when dispersal was high, which may reflect the elevated observed divergence in high-dispersal scenarios (Supplementary Fig. S3 available on Dryad). Allopatric speciation increased the probability of Type I error (e.g., Fig. 2Bi–ii).

For the phylogenetic trait model-fitting analyses, BM and OU were generally correctly chosen when they were the generating models (i.e., when  $m=0$  and when  $\psi=0$  or 2, respectively, Fig. 3, Supplementary Fig. S8 available on Dryad). When  $\psi=0$  and  $m>0$ , the MC model with biogeography is consistently the best-fit model (Fig. 3A, Supplementary Fig. S8A available on Dryad). When  $m>0$  and  $\psi=2$ , the diversity dependent exponential ( $DD_{exp}$ ) model with biogeography was favored over other models in most scenarios (Fig. 3B, Supplementary Fig. S8B available on Dryad), with positive rate parameters estimated in the maximum likelihood solution (Supplementary Fig. S9 available on Dryad). The biogeographic scenario did not greatly affect the outcome of model fitting, though correct models were slightly more supported when dispersal was high (Supplementary Fig. S10 available on Dryad), in agreement with the observed magnitude of repulsion (Supplementary Fig. S3 available on Dryad). Although the models are less identifiable when  $m=10$  and  $\psi=2$  (Fig. 3, Supplementary Fig. S8 available on Dryad), this results from variation in the  $\psi:m$  ratio—there is a ratio of  $\psi:m$  around which these models cannot be distinguished (Supplementary Fig. S11 available on Dryad).

Process-based models fit to sister-taxa data sets in EvoRAG did not mistakenly identify an effect of species interactions when they were absent (Supplementary Fig. S4A available on Dryad, C, Table 2), but they were unable to identify the effect of competition when  $\psi=0$  (Supplementary Fig. S4B available on Dryad, Table 2). However, as with process-based models fit to the whole phylogeny, when data were simulated with both repulsion and a pull toward a stable peak, a model where evolutionary rates vary linearly with the number of sympatric taxa is often the best-fit model, though generally with only a marginally lower AICc value (i.e.,

$\Delta AICc < 2$ ) than BM (Supplementary Fig. S4 available on Dryad, Table 2).

### *Convergent Character Displacement*

As with divergent character displacement, with all pairwise species combinations, the ability of most methods to detect convergent character displacement depends on the presence of the OU process on the resource-use trait: data sets simulated with convergent character displacement and an OU pull on resource-use traits were more likely to be statistically significant (Fig. 4A.vi, Supplementary Fig. S12A.vi available on Dryad) across all methods than those simulated with convergent character displacement and no OU pull on resource-use traits (Fig. 4A.iii, Supplementary Fig. S12A.iii available on Dryad). Again, this is likely because the presence of the OU process in the resource-use trait amplifies the magnitude of convergence (Supplementary Figs. S13 and S14 available on Dryad). Overall, however, only the simulation approach had substantial power ( $>0.80$ ) to detect convergent character displacement (Table 1), and only in trees with 100 or more tips and data sets with the OU process in the simulated resource-use trait. Indeed, the nonphylogenetic regressions often (spuriously) detected divergence rather than the simulated convergence, especially in smaller trees (Fig. 4Ai vs. Aii, Supplementary Fig. S12Aiv,ii and Tables available on Dryad). Both types of Mantel tests were unable to detect convergence, in fact having a higher Type I error rate (detecting divergence in BM simulated data sets, Supplementary Tables available on Dryad) than power. As with divergent character displacement, there was a tendency for higher power in lower dispersal scenarios.

The power to detect convergence generally increased with increasingly negative values of  $m$ , the maximum strength of attraction in the signal trait when species are identical in the resource-use trait (Supplementary Fig. S15 available on Dryad), though as  $m$  gets large, the probability that all species converge on the same trait value increases, especially when  $\psi_{resource}=2$  (see Supplementary Methods available on Dryad).

Regardless of whether resource-use traits are simulated under OU or BM, when there is no convergence, nonphylogenetic regressions and PLMMs used for analyses of sister taxa data sets tend to have high Type I error rates, though these analyses return an erroneous inference of divergence, rather than convergence, between sister taxa (Fig. 4Bi,ii,iv,v, Supplementary Fig. S12Bi,ii,iv,v available on Dryad, Table 2, Supplementary Tables available on Dryad). Sister-taxa analyses had overall very low power ( $<0.6$ ) to detect convergence when it did exist, and nonphylogenetic regressions often detected divergence, rather than convergence (Table 2, Supplementary Tables available on Dryad). As with divergent character displacement simulations, the allopatric speciation biogeographic scenarios were more likely to lead to higher Type I error rates (Fig. 4Bi,iv). Process-based models fit to sister-taxa data sets

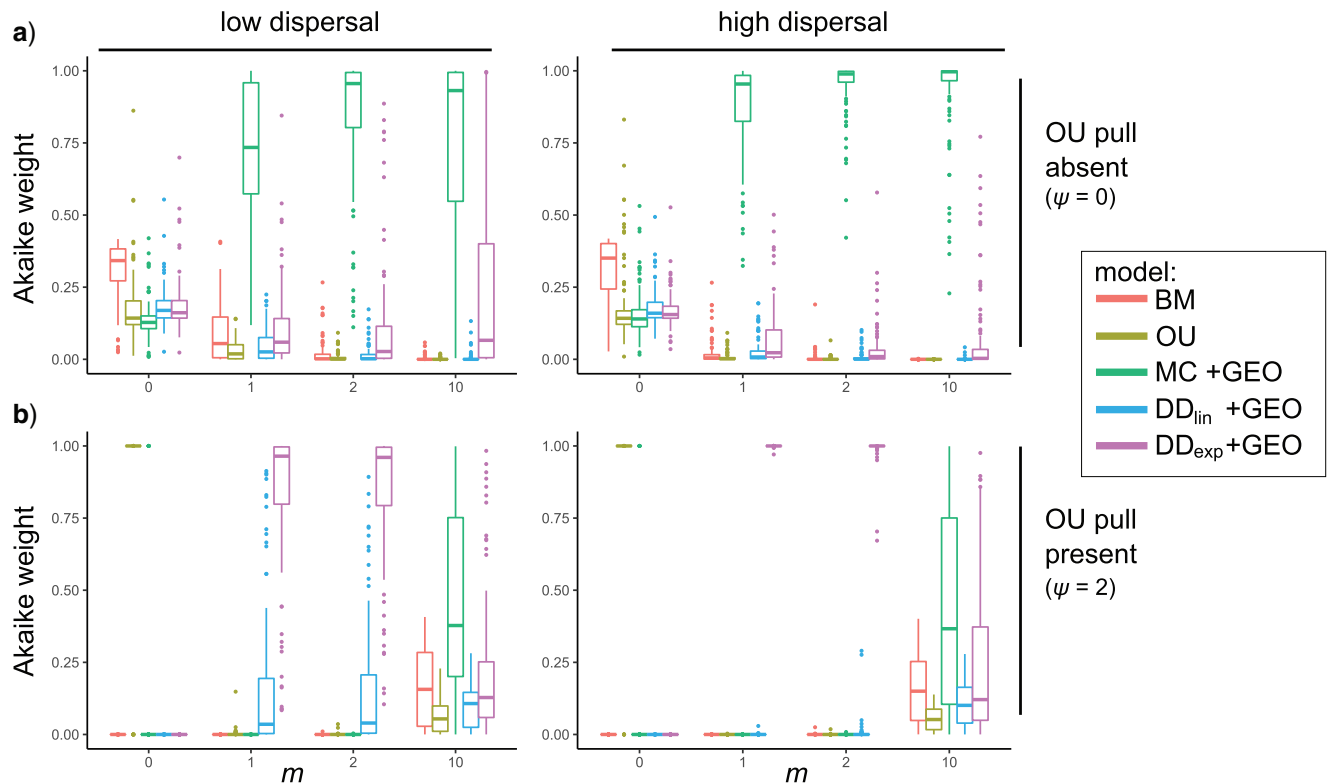


FIGURE 3. Boxplots of Akaike weights for each trait model fit to simulated data sets in biogeographic scenarios with low sympatric speciation rates as a function of  $m$  in trees with 100 species. a) When OU is absent, BM is the best-fit model when  $m=0$ , and the MC model with biogeography is the best model when competitive divergence is present. b) When OU is present, OU is the best-fit model when  $m=0$ , and the diversity-dependent exponential model with biogeography is the best model when competitive divergence is present and  $\psi:m$  is relatively high.

in EvoRAG did not erroneously detect divergence or convergence (i.e., BM was the best-fit model when  $m=0$ , Supplementary Fig. S14A,C available on Dryad, Table 2), but they could not detect an effect of species interactions when convergence was present, at least for the number of sister taxa in this study, as OU was the best-fit model when  $m=-0.25$  (Supplementary Fig. S14B,C available on Dryad, Table 2).

#### Predicting Interspecific Interactions

Although all three methods used to identify traits that are causally related to interspecific interactions had high power ( $\gg 0.8$ , Table 1, Supplementary Tables available on Dryad) to do so in the parameter space explored here (Fig. 5ii,iv, Supplementary Fig. S16ii,iv available on Dryad), only the simulation approach had both high power and a low Type I error rate (Table 1), whereas nonphylogenetic regressions and PLMMs had fairly high Type I error rates (Table 1) when interactions were simulated based on similarity in a trait other than the measured one (Fig. 5i,iii, Supplementary Fig. S16i,iii available on Dryad). The power to detect an effect of trait similarity on species interactions was not greatly affected by the presence of an additional, unmeasured trait that also affected the interaction (Supplementary Fig. S17 available on Dryad). Biogeography did not have a large impact on analyses, though there were

slightly higher Type I error in low-dispersal scenarios (Fig. 5i,iii).

#### DISCUSSION

As open-access databases with species range, trait, and phylogenetic data rapidly expand, investigators are able to test hypotheses about the relationships between interspecific interactions and phenotypic evolution at an unprecedented scale. Understanding the relative strengths and weaknesses of phylogenetic comparative methods available for testing such hypotheses is thus paramount. We found that many currently used methods for detecting causal relationships between interspecific interactions and species phenotypes suffer from severe limitations (Tables 1,2).

Overall, standard methods are better at detecting divergent character displacement when divergence does not drive unbounded trait evolution (i.e., when selection acts against extreme phenotypes, as can be modeled by the OU process). Consistent with previous reports (Harmon and Glor 2010; Guillot and Rousset 2013), Mantel tests had high Type I error rates and both standard and pppMantel tests have low power (Table 1, Fig. 2Ai, Supplementary Fig. S2Ai available on Dryad). We found that several analytical tools used on sister-taxa data sets have high Type I error rates (Table 2, Figs. 2Bi,iv, 4Bi,iv, Supplementary Figs. S2Bi,iv, S12Bi,iv, Tables available on



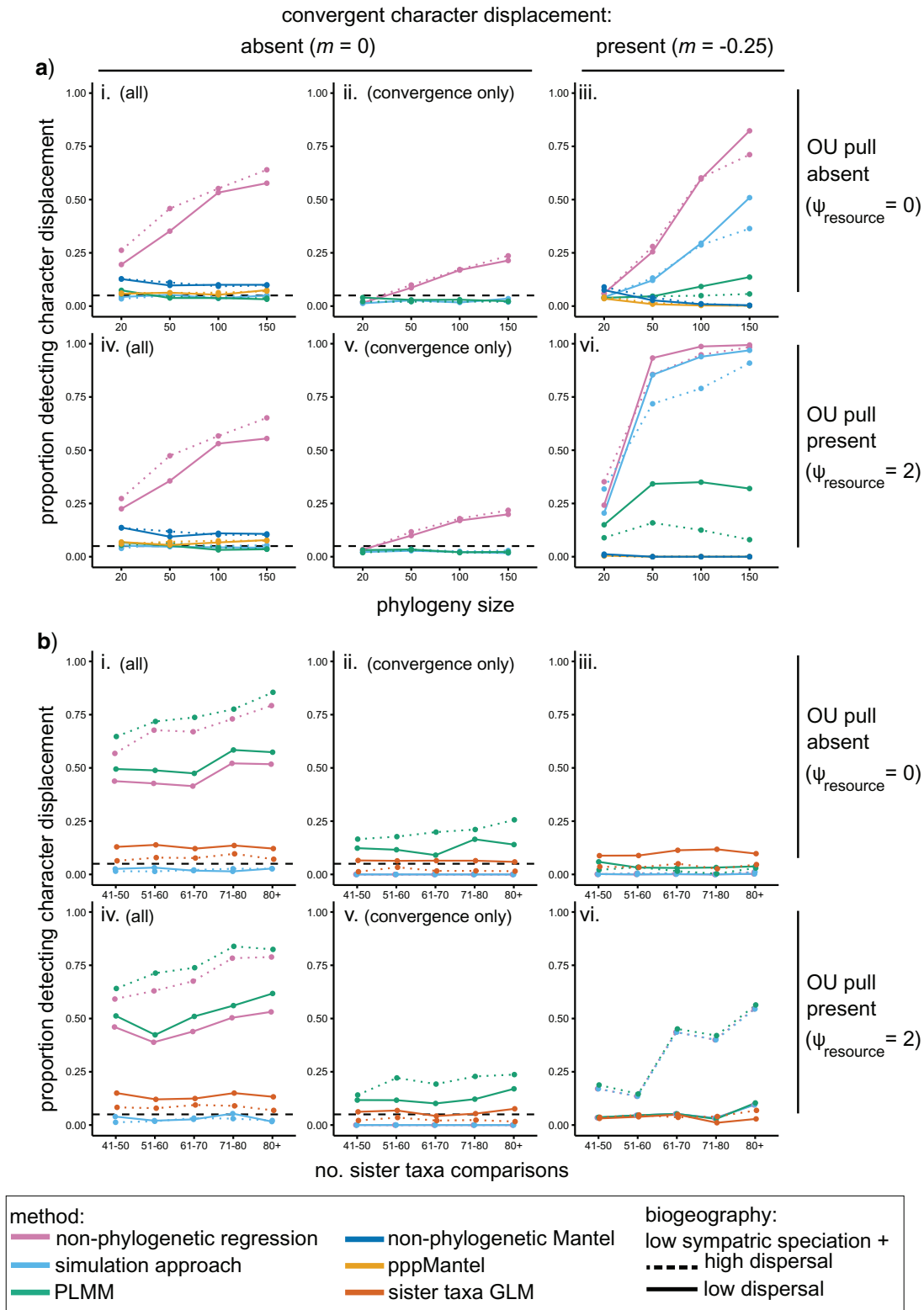


FIGURE 4. Proportion of statistically significant analyses in data sets simulated under convergent character displacement in biogeographic scenarios with low sympatric speciation rates. a) Results from approaches using data from all pairwise comparisons in a clade, plotted as a function of the phylogeny size and dispersal rate when (i–ii)  $m = 0$  and  $\psi_{\text{resource}} = 0$  [(i) all analyses and (ii) only analyses returning convergence in sympatry], (iii)  $m = -0.25$  and  $\psi_{\text{resource}} = 0$ , (iv–v)  $m = 0$  and  $\psi_{\text{resource}} = 2$  [(iv) all analyses and (v) only analyses returning convergence in sympatry], and (vi)  $m = -0.25$  and  $\psi_{\text{resource}} = 2$ . b) Results from analyses of sister-taxa culled from complete phylogenies binned by the number of resulting species pairs, plotted as a function of the number of sister taxa comparisons and dispersal rate when (i–ii)  $m = 0$  and  $\psi_{\text{resource}} = 0$  [(i) all analyses and (ii) only analyses returning convergence in sympatry], (iii)  $m = -0.25$  and  $\psi_{\text{resource}} = 0$ , (iv–v)  $m = 0$  and  $\psi_{\text{resource}} = 2$  [(iv) all analyses and (v) only analyses returning convergence in sympatry], and (vi)  $m = -0.25$  and  $\psi_{\text{resource}} = 2$ . For scenarios where  $m = -0.25$ , only the proportion of significant results showing convergence are plotted. Dashed horizontal lines represent a Type I error rate of 5%.

TABLE 1. Summary of the statistical properties of the analytical approaches tested under scenarios using data from all tips (i.e., with sister-taxa analyses excluded)

Analysis	Nonphylogenetic regression		Mantel test		pppMantel test		PLMM		Simulation test		Process-based models	
	Type I	Power	Type I	Power	Type I	Power	Type I	Power	Type I	Power	Type I*	Power <sup>†</sup>
Divergent char. displacement	0.37–0.61	0.51–1	0.05–0.10	0.28–1	0.04–0.06	0.20–1	0.05–0.06	0.12–1	0.05–0.07	0.07–1	0.01–0.04	0.92–0.93
Convergent char. displacement	0.40–0.60	0.31–0.99	0.08–0.09	0–0.02	0.05–0.06	0–0.01	0.05–0.07	0.07–0.26	0.04–0.05	0.12–0.91	—	—
Predicting spp. interactions	0.08–0.3	1	—	—	—	—	0.07–0.18	1	0.03–0.04	1	—	—

Notes: Values refer to the range of average type I error rates and power levels for each tree size  $\geq 50$  across biogeographic scenarios and scenarios where  $\psi$  or  $\psi_{\text{resource}} = 0$  or 2. Power refers to only those statistically significant tests in the appropriate tail (i.e., in the lower tail for divergent character displacement and upper tail for convergent character displacement). For each analytical scenario, the cell with the method with the best trade-off between Type I error and power is shaded.

\*Type I error rate calculated as the proportion of data sets simulated without divergent character displacement for which a model that includes species interactions— $DD_{\text{exp}}$ ,  $DD_{\text{lin}}$ , or MC—was chosen by model selection (i.e., for which  $\Delta\text{AICc} = 0$  and  $\Delta\text{AICc}$  for all other models  $> 2$ ).

<sup>†</sup>Power calculated as the proportion of data sets simulated with divergent character displacement for which either  $DD_{\text{exp}}$ ,  $DD_{\text{lin}}$ , or MC was chosen by model selection.

TABLE 2. Summary of the statistical properties of the analytical approaches tested under scenarios using sister-taxa analyses

Analysis	Non-phylogenetic regression		Sister-taxa GLM		PLMM		Simulation test		Process-based models in EvoRAG	
	Type I	Power	Type I	Power	Type I	Power	Type I	Power	Type I*	Power <sup>†</sup>
Divergent char. displacement	0.07–0.42	0.69–0.75	0.05–0.07	0.19–0.32	0.08–0.50	0.69–0.78	0.01–0.03	0.18–0.30	0.04–0.07	0.03–0.37
Convergent char. displacement	0.33–0.43	0.01–0.2	0.07	0.04–0.06	0.41–0.5	0.02–0.21	0.03	0.01–0.2	0.04	0.09–0.55

Notes: Values refer to the range of type I error rates and power levels, averaged across biogeographic scenarios, and scenarios where  $\psi$  or  $\psi_{\text{resource}} = 0$  or 2. Power refers to only those statistically significant tests in the appropriate tail (i.e., in the upper tail for divergent character displacement and lower tail for convergent character displacement). Since no method has both low Type-I error rates and high power, we caution against using sister-taxa approaches to test for character displacement.

\*Type I error rate calculated as the proportion of data sets simulated without divergent character displacement for which a model that includes a linear dependency on the number of sympatric lineages— $BM_{\text{linear}}$  or  $OU_{\text{linear\_beta}}$ —was chosen by model selection (i.e., for which  $\Delta\text{AICc} = 0$  and  $\Delta\text{AICc}$  for all other models  $> 2$ ).

<sup>†</sup>Power calculated as the proportion of data sets simulated with divergent character displacement for which either  $BM_{\text{linear}}$  or  $OU_{\text{linear\_beta}}$  was chosen by model selection.

Dryad), which would lead investigators to conclude that divergent character displacement had occurred when, in fact, it had not, and no statistical approaches for sister-taxa analyses have a reasonable combination of Type I error and power. Given the lack of a method that has reasonable Type I error rate and power, we discourage empiricists from using sister-taxa approaches to study character displacement. If no other data are available for testing for character displacement on the whole tree, then we recommend phylogenetic simulations or sister-taxa GLMs, as they are the only methods with generally low type I error rates, even though they suffer from low power (Tables 1 and 2). Moreover, for analyses conducted with phylogenetic simulations or sister-taxa GLMs, though at risk of falsely rejecting the hypothesis of character displacement owing to low power, empiricists can be fairly confident that positive signals of character displacement are trustworthy.

Fitting process-based phylogenetic trait models to data sets simulated with divergent character displacement

yielded more consistent patterns (Fig. 3). Without attraction toward a single stationary peak to bound trait evolution, the MC model with biogeography was predominantly the best-fit model. For data sets simulated with the OU process, the diversity-dependent exponential ( $DD_{\text{exp}}$ , see Appendix 1) model with biogeography was the best-fit model, and similarly a model with a linear relationship between evolutionary rates and the number of sympatric taxa often fit sister-taxa data sets, though with much lower power overall (Supplementary Fig S4 available on Dryad, Table 2). In the  $DD_{\text{exp}}$  model, rates of trait evolution vary exponentially with the number of sympatric lineages through time, thereby incorporating the effect of interspecific interactions on the rate of trait of evolution but not explicitly modeling the process of character displacement acting on the mean trait values. It may nonetheless provide a useful proxy for detecting patterns that are similar to those left by character displacement, in the absence of a process-based model that incorporates both attraction

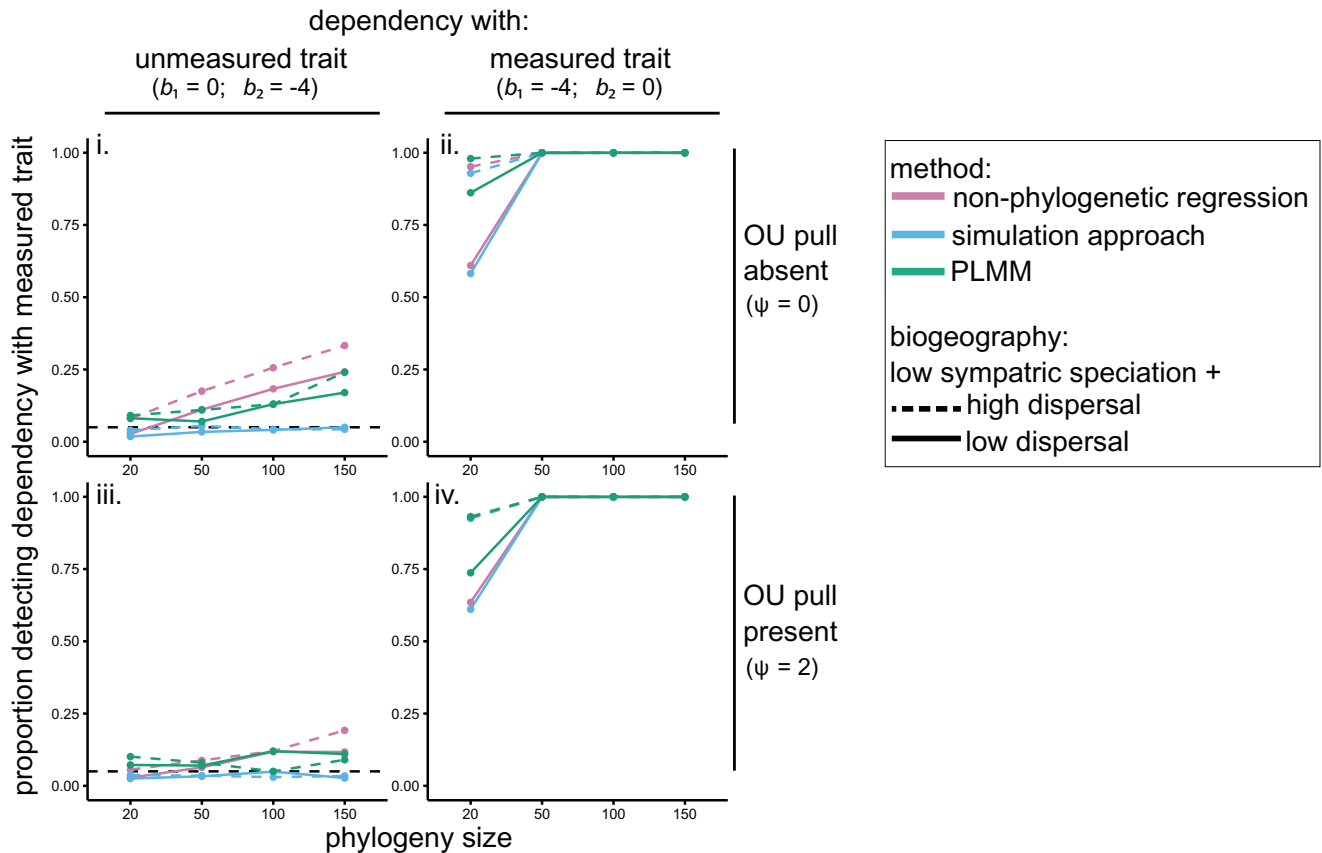


FIGURE 5. Proportion of statistically significant analyses in data sets with interactions simulated under a simple phenotype matching process in biogeographic scenarios with low sympatric speciation rates. Results from analyses where the measured trait was simulated under BM (i, ii) or OU (iii, iv), plotted as a function of the phylogeny size and dispersal rate when i.  $b_1$  (the simulation coefficient determining the relationship between the interaction and the measured trait) = 0,  $b_2$  (the simulation coefficient for an unmeasured trait) = -4, and  $\psi = 2$ , ii.  $b_1 = -4$ ,  $b_2 = 0$ , and  $\psi = 2$ , iii.  $b_1 = 0$ ,  $b_2 = -4$ , and  $\psi = 0$ , and (iv)  $b_1 = -4$ ,  $b_2 = 0$ , and  $\psi = 0$ .

toward an optimum trait value and divergent character displacement. We emphasize, however, that statistical support for phylogenetic process-based trait models incorporating interspecific interactions does not in itself constitute decisive evidence that character displacement has occurred, as other processes may generate similar patterns (e.g., increasing evolutionary rates with increasing lineage diversity). Given that the  $DD_{exp}$  model is the best-fit model in parameter space where other methods also perform well, combined evidence from model-fitting and other, nonprocess based methods would constitute a strong case for the presence of character displacement. In the absence of tip data (e.g., due to incomplete sampling or traits that are inherently measured as pairwise properties), process-based models are unsuitable and we recommend using data from as many species pairs as possible—not just sister taxa—and using simulation approaches or PLMMs. In other words, to detect divergent character displacement, we recommend that empiricists fit the MC model to their data set when possible. High support for the MC model would constitute evidence that character displacement has acted on a trait. If the MC model does not provide a good fit for the data, this could be because character

displacement proceeds in the presence of bounded trait evolution, in which case a signature of the  $DD_{exp}$  model with a positive rate parameter and/or a signature of sympatric divergence in phylogenetic simulations or PLMMs would constitute evidence consistent with divergent character displacement.

Interestingly, even though most previous investigators have used the  $DD_{exp}$  model to represent a decline in ecological opportunity with increasing species richness (Mahler et al. 2010; Weir and Mursleen 2013), the maximum likelihood estimates of the rate parameters for this model were positive, rather than negative, when both divergence and the OU process were present (Supplementary Fig. S9 available on Dryad). This is consistent with our finding of increasing evolutionary rates with increasing species richness (Supplementary Figs. S3, S4 and S7 available on Dryad) in this scenario. An increase in the rate of evolutionary changes in trait values toward the present likely results from selection not only restricting species to certain trait space but also partitioning that space. The resulting adaptive landscape is therefore changing rapidly, causing accelerating evolutionary rates as lineages fill this increasingly constrained space.

The MC model (Appendix 1) is similar to the model used to simulate data (Eq. 1), with the assumption that  $\alpha$  is very small ( $\ll 1$ ) and consequently, competitive interactions are affected by the mean trait values of all sympatric species, rather than by pairwise similarity (Nuismer and Harmon 2015; Drury et al. 2016). Biologists, however, generally assume that competition is stronger between phenotypically similar species (Brown and Wilson 1956). Our results show that the assumption of a small  $\alpha$  does not render the MC model useless for studying character displacement, as the MC model is the best-fit model for many data sets simulated under the character displacement model used here. Nevertheless, the finding that the  $DD_{\text{exp}}$  model is the best-fit model in data sets simulated under character displacement including OU indicates that the MC model is not a perfect model of character displacement. Recently, approximate Bayesian computational (ABC) tools have been published to fit a model of character displacement in which, like in our simulation model, the strength of competition depends on similarity in trait space (Clarke et al. 2017). This model provides an alternative tool for detecting character displacement in comparative data sets, and we hope that further development of methods such as this ABC method will help ameliorate the statistical issues shown here.

For data sets simulated including the OU process, the ratio of the pull-parameter in the OU portion of the model to the maximum amount of repulsion ( $\psi:m$ ) had a consistent impact across all methods, which results from the overall larger magnitude of evolutionary changes in traits in scenarios with a high  $\psi:m$  ratio (Supplementary Figs. S3, S4 and S7 available on Dryad). As  $\psi:m$  approached 1, all methods were better at detecting character displacement. Currently, there are no analytical approaches that can disentangle the simultaneous impact of attraction toward a peak and divergence due to competition, though we hope our results will inspire development of such tools. We also note that the ratio of the BM rate parameter  $\sigma^2$  and  $m$  will also likely impact the ability to detect character displacement, though we have not explored this here.

Unlike for divergent character displacement, available statistical methods for detecting convergence in comparative data sets generally do a poor job of detecting convergence, with the simulation method outperforming others (Table 1). With whole-data set approaches, Type I error rates are acceptable for phylogenetic analyses (~5%), however, so although detecting convergence is difficult, the risk of mistakenly detecting convergence is low. In sister-taxa analyses, although Type I error rates are high for PLMMs (Table 2), these largely return erroneous divergence results, rather than erroneous convergence (Figs. 4Bii,v, Supplementary Fig. S12Bii,v available on Dryad). In short, if an empiricist detects convergence in their data set, they can be fairly confident in the result. Yet if empiricists do not detect convergence, this could simply be a result of lower power of the available analytical tools. Currently, there are no tools to fit phylogenetic trait models of convergence between

species (e.g., Nuismer and Harmon 2015); such tools might more successfully identify convergent character displacement in comparative data sets than the available statistical methods.

For both divergent and convergent character displacement scenarios, we found that sister-taxa GLMs and the simulation approach applied to sister-taxa data sets had a mean Type I error rate near 5% (Table 2). However, in some scenarios, the Type I error for sister-taxa GLMs was slightly higher than for the simulation approach (Figs. 2Bi and 4Bi, Supplementary Tables available on Dryad), which suggests that including a model-based estimate of the rate of trait evolution more properly accounts for the effect of divergence than simply including the branch lengths separating sister taxa as a covariate in analyses to control for variation in the amount of time sister taxa have had to diverge from one another (but see Appendix 1 for other extensions of sister-taxa GLMs). The high overall Type I error rate for analyses conducted on sister-taxa data sets may also result from the unrealistic assumption, common to all sister-taxa analyses, that transitions between allopatry and sympatry are uncommon along branches connecting sister taxa (Weir and Price 2011; Tobias et al. 2014). Supporting this explanation, we found that biogeographic scenarios with high levels of sympatric speciation and low dispersal tended to have overall lower Type I error rates (cf. Figs. 2, 4, Supplementary Figs. S2 and S12 available on Dryad).

The statistical properties of analyses used for identifying which traits drive species interactions are less variable than for character displacement scenarios. The statistical methods available to test for causal relationships between phenotypic similarity and interactions between species have very high power. The simulation approach has a low Type I error rate when causal relationships are simulated based on an unmeasured trait, although non-phylogenetic regressions and PLMMs suffer from relatively high Type I error rates (Table 1). Thus, we recommend that empiricists interested in predicting pairwise species interactions based on trait data use phylogenetic simulations. While we did not simulate interactions between clades, our results are likely applicable to other empirical questions, such as identifying traits that predict links in ecological networks (Rafferty and Ives 2013; Hadfield et al. 2014; Eklöf and Stouffer 2016).

By simulating data sets with various types of interactions between species across different modes of speciation and dispersal rates, we have shown that many of the methods that investigators use to analyze empirical data sets have low power to detect such patterns (Table 1). In particular, widely-used sister taxa analyses, including standard regressions and, in some scenarios, sister-taxa GLMs, often detected character displacement in data sets that were simulated under a simple BM model (Figs. 2Bi-ii and 4Bi-ii). We therefore urge investigators to use caution when interpreting the results of such analyses, even in cases where sympatry is delineated using other criteria than the one considered here. When

process-based models could be fit to these data sets, they tended to correctly identify patterns of divergence (i.e., either the MC model or a diversity-dependent model is the best fit model >92% of the time). Thus, when possible, empiricists should employ such methods. Statistical tools to fit process-based models of phenotypic evolution including species interactions are in their infancy (Drury et al. 2016; Manceau et al. 2016) and many possible models are not yet available (e.g., convergent character displacement, character divergence in the presence of an adaptive pull towards a peak). We hope that our results encourage the continued development of such tools.

In closing, we note that divergent character displacement is erroneously detected with many statistical approaches, indicating that there may be an overrepresentation of empirical studies that imply that divergence has occurred. In particular, studies that have used sister-taxa methods to document character displacement using standard regressions or PLMMs may have falsely interpreted a null expectation—larger trait differences between sympatric lineages owing to allopatric speciation—as evidence for divergent character displacement. Conversely, convergent character displacement is often hard to detect with existing methods, suggesting that convergence in signal traits (e.g., Cody 1969, Cody 1973; Pigot and Tobias 2014; Losin et al. 2016) might be more prevalent than previously thought.

#### ACKNOWLEDGEMENTS

We thank M. Manceau, J. Clavel, E. Lewitus, O. Maliet, O. Missa, L. Harmon, and several anonymous reviewers for feedback, N. Matzke for assistance simulating trees with biogeography, and F. Hartig for assistance streamlining our simulation script.

#### FUNDING

This research was funded by the European Research Council (grant 616419-PANDA to H.M.) and the National Science Foundation (grant DEB-1457844 to G.F.G.).

#### SUPPLEMENTARY MATERIAL

Data available from the Dryad Digital Repository: <http://dx.doi.org/10.5061/dryad.ch0vn>.

#### APPENDIX 1. METHODS FOR ASSESSING THE INTERPLAY BETWEEN INTERSPECIFIC INTERACTIONS AND SPECIES PHENOTYPES

Comparative analyses of the interplay between interspecific interactions and species phenotypes can either be conducted on entire clades, or, commonly, on sister taxa—species pairs that share a most recent common ancestor—that are culled from larger phylogenies. Such analyses generally consist of testing the statistical significance of correlations between either phenotypic similarity and geographic overlap (to test for divergent or convergent character displacement) or

species interactions and phenotypic similarity (to find predictors of species interactions). As we are looking for correlations between pairwise comparisons (e.g., trait similarity, biogeographical overlap, hybridization, magnitude of pre-zygotic isolation), rather than “tip values” belonging to a single species, phylogenetically independent contrasts and extensions of PGLS analyses (Felsenstein 1985; Rezende and Diniz-Filho 2012) cannot be used, and alternative tests have been developed. For a guide to which analytical tools can be applied to each empirical question, see Supplementary Diagram 1 available on Dryad.

#### 1. Non-phylogenetic regressions

“Non-phylogenetic regressions” refers to Generalized Linear Models (GLMs) that ignore phylogenetic structure. Though less commonly applied to whole-clade analyses, investigators sometimes use non-phylogenetic regressions for sister-taxa analyses, on the basis that branches connecting sister taxa represent independent evolutionary histories (Felsenstein 1985). Non-phylogenetic regressions can be used in tests for character displacement or in analyses of predictors of interspecific interactions.

#### 2. Mantel tests

Several previous investigators have implemented Mantel tests (Mantel 1967) to test for character displacement between species pairs (e.g., Roncal et al. 2012). These tests are designed to assess correlations between matrices, which here comprise interspecific trait distances or differences. Existing accounts of Mantel tests describe procedures only for complete matrices, so they cannot be used in many cases, including sister-taxa analyses (for which most off-diagonal elements of distances matrices are by definition excluded) and in identifying predictors of species interactions (e.g., hybridization), as only sympatric lineages can interact and setting values for allopatric comparisons to zero would not make biological sense.

#### 3. Phylogenetically permuted partial Mantel tests

Phylogenetically permuted partial Mantel (pppMantel) tests account for phylogenetic non-independence (e.g., see Lapointe and Garland 2001) by permuting null data sets that are structured phylogenetically, and are popular among investigators testing for character displacement (e.g., Allen et al. 2014; Willis et al. 2014; Medina-García et al. 2015). Like Mantel tests, pppMantel tests also require complete interaction matrices.

#### 4. Phylogenetic linear mixed models

In recent years, researchers have adapted animal models from quantitative genetics to incorporate phylogenies as random effects in mixed-effect regressions on comparative data sets (Hadfield and Nakagawa 2010; de Villemereuil and Nakagawa 2014). Such PLMMs have been modified to accommodate pairwise species data

(Tobias et al. 2014), wherein the identity of the species being compared and the node connecting them in the phylogeny are included as random effects. PLMMs are promising new tools, as they are not limited to sister-taxa data and model predictions can be generated and plotted.

### 5. Phylogenetic simulations

Simulation approaches are widely used to control for phylogenetic nonindependence in tip data (Martins and Garland 1991; Garland et al. 1993), and have been applied to pairwise species comparisons (Elias et al. 2008; Drury et al. 2015; Losin et al. 2016). In these approaches, trait evolution is simulated along phylogenies, often scaled such that the simulated tip data resemble real data. Pairwise comparisons are then calculated on many simulated data sets and used to generate a phylogenetically informed null distribution of test statistics against which to compare test statistics calculated from nonphylogenetic regressions on the real data.

### 6. Process-based models of phenotypic evolution

In the statistical approaches outlined thus far, the data analyzed are measurements of pairwise differences between species, and the statistical tests for the effect of species interactions on trait evolution consist of testing for significant correlations between either phenotypic similarity and geographic overlap or species interactions and trait similarity. However, it is also possible to detect a signature of interspecific competition in the distributions of continuous trait values across the tips of a phylogeny by fitting process-based models of phenotypic evolution to the data. These models allow testing hypotheses about which processes are most likely to have generated the observed distribution of traits in a clade (Hansen and Martins 1996; Harmon et al. 2010).

Interspecific interactions have recently been incorporated into such models in two ways. First, in diversity-dependent (DD) models, evolutionary rates change as a function [either linear (DD<sub>lin</sub>) or exponential (DD<sub>exp</sub>)] of the number of extant lineages through time (e.g., Weir and Mursleen 2013). Secondly, in the “matching competition” (MC) model, trait evolution in an evolving lineage varies as a function of the values of traits in other evolving lineages (Nuismer and Harmon 2015, Drury et al. 2016). Comparing the fit of these models to other models that exclude interspecific interactions (e.g., BM and OU models) tests whether there is evidence that interspecific interactions have influenced the trajectory of trait evolution in a clade.

### 7. Sister-taxa GLMs

If allopatric speciation is common, then sympatry occurs after a period of initial isolation, resulting in a pattern where sympatric sister taxa are older than allopatric sister taxa. Thus, even random genetic drift can generate a pattern in which sympatric lineages have more divergent traits compared to allopatric lineages,

simply because divergence has had more time to evolve (Weir and Price 2011; Tobias et al. 2014). To control for variation in the evolutionary distance between sister taxa in tests for character displacement, “sister-taxa GLMs” include patristic distance as a predictor in non-phylogenetic regressions (e.g., Davies et al. 2007; Martin et al. 2010). Extensions to sister-taxa GLMs include (i) non-linear transformations of patristic distances (Weber et al. 2016) and (ii) comparisons of the divergence of sister taxa relative to a third taxon, with one sister allopatric to and the other sympatric with that third taxon (Noor 1997).

### 8. Sister-taxa model fitting

Recently, tools have been described for fitting process-based models to sister taxa data sets using maximum likelihood (Weir and Wheatcroft 2011; Weir and Lawson 2015). With these tools, it is possible to test whether models that allow evolutionary rates to vary as a linear function of a gradient (e.g., whether male plumage coloration varies as a function of the strength of sexual selection, Seddon et al. 2013) better fit sister-taxa data sets than constant rates models. When the gradient is the number of sympatric lineages, these models are conceptually similar to the linear diversity dependent models described above.

## REFERENCES

- Allen W.L., Stevens M., Higham J.P. 2014. Character displacement of Cercopithecini primate visual signals. *Nat. Commun.* 5:4266.
- Anacker B.L., Strauss S.Y. 2014. The geography and ecology of plant speciation: range overlap and niche divergence in sister species. *Proc. Roy. Soc. B* 281:20132980.
- Bacquet P.M.B., Brattström O., Wang H.L., Allen C.E., Löfstedt C., Brakefield P.M., Nieberding C.M. 2015. Selection on male sex pheromone composition contributes to butterfly reproductive isolation. *Proc. R. Soc. London B Biol. Sci.* 282:20142734.
- Bothwell E., Montgomerie R., Loughheed S.C., Martin P.R. 2015. Closely related species of birds differ more in body size when their ranges overlap—in warm, but not cool, climates. *Evolution* 69:1701–1712.
- Brown W.L.L., Wilson E.O. 1956. Character displacement. *Syst. Zool.* 5:49–64.
- Cadotte M., Albert C.H., Walker S.C. 2013. The ecology of differences: assessing community assembly with trait and evolutionary distances. *Ecol. Lett.* 16:1234–1244.
- Cavender-Bares J., Kozak K.H., Fine P.V.A., Kembel S.W. 2009. The merging of community ecology and phylogenetic biology. *Ecol. Lett.* 12:693–715.
- Clarke M., Thomas G.H., Freckleton R.P. 2017. Trait evolution in adaptive radiations: modelling and measuring interspecific competition on phylogenies. *Am. Nat.* 189:121–137. doi: 10.1086/689819.
- Cody M.L. 1969. Convergent characteristics in sympatric species: a possible relation to interspecific competition and aggression. *Condor* 71:223–239.
- Cody M.L. 1973. Character convergence. *Annu. Rev. Ecol. Syst.* 4:189–211.
- Davies J.T., Meiri S., Barraclough T.G., Gittleman J.L. 2007. Species co-existence and character divergence across carnivores. *Ecol. Lett.* 10:146–152.
- Drury J., Clavel J., Manceau M., Morlon H. 2016. Estimating the effect of competition on trait evolution using maximum likelihood inference. *Syst. Biol.* 65:700–710.
- Drury J.P., Okamoto K.W., Anderson C.N., Grether G.F. 2015. Reproductive interference explains persistence of aggression between species. *Proc. R. Soc. B* 282:20142256.

- Eklöf A., Stouffer D.B. 2016. The phylogenetic component of food web structure and intervality. *Theor. Ecol.* 9:107–115.
- Elias M., Gompert Z., Jiggins C., Willmott K. 2008. Mutualistic interactions drive ecological niche convergence in a diverse butterfly community. *PLoS Biol.* 6:e300.
- Felsenstein J. 1985. Phylogenies and the comparative method. *Am. Nat.* 125:1–15.
- Felsenstein J. 1988. Phylogenies and quantitative characters. *Annu. Rev. Ecol. Syst.* 19:445–471.
- Garland T., Dickerman A.W., Janis C.M., Jones J.A. 1993. Phylogenetic analysis of covariance by computer simulation. *Syst. Biol.* 42:265–292.
- Grant P.R. 1972. Convergent and divergent character displacement. *Biol. J. Linn. Soc.* 4:39–68.
- Grant P.R., Grant B.R. 2006. Evolution of character displacement in Darwin's finches. *Science* 313:224–226.
- Grant P.R., Grant B.R. 2011. How and why species multiply: the radiation of Darwin's finches. Princeton (NJ): Princeton University Press.
- Grether G.F., Anderson C.N., Drury J.P., Kirschel A.N.G., Losin N., Okamoto K., Peiman K.S. 2013. The evolutionary consequences of interspecific aggression. *Ann. N. Y. Acad. Sci.* 1289:48–68.
- Grether G.F., Losin N., Anderson C.N., Okamoto K. 2009. The role of interspecific interference competition in character displacement and the evolution of competitor recognition. *Biol. Rev.* 84:617–635.
- Guillot G., Rousset F. 2013. Dismantling the Mantel tests. *Methods Ecol. Evol.* 4:336–344.
- Hadfield J.D., Krasnov B.R., Poulin R., Nakagawa S. 2014. A tale of two phylogenies: comparative analyses of ecological interactions. *Am. Nat.* 183:174–187.
- Hadfield J.D., Nakagawa S. 2010. General quantitative genetic methods for comparative biology: phylogenies, taxonomies and multi-trait models for continuous and categorical characters. *J. Evol. Biol.* 23:494–508.
- Hansen T.F., Martins E.P. 1996. Translating between microevolutionary process and macroevolutionary patterns: the correlation structure of interspecific data. *Evolution* 50:1404–1417.
- Harmon L.J., Glor R.E. 2010. Poor statistical performance of the Mantel test in phylogenetic comparative analyses. *Evolution* 64:2173–2178.
- Harmon L.J., Losos J.B., Jonathan Davies T., Gillespie R.G., Gittleman J.L., Bryan Jennings W., Kozak K.H., McPeck M.A., Moreno-Roark F., Near T.J., Purvis A., Ricklefs R.E., Schluter D., Schulte J.A., Seehausen O., Sidlauskas B.L., Torres-Carvajal O., Weir J.T., Mooers A.T. 2010. Early bursts of body size and shape evolution are rare in comparative data. *Evolution* 64:2385–2396.
- Hilbe J.M. 2009. Logistic regression models. Boca Raton, FL: CRC press.
- Lapointe F.J., Garland T. 2001. A generalized permutation model for the analysis of cross-species data. *J. Classif.* 18:109–127.
- Losin N., Drury J.P., Peiman K.S., Storch C., Grether G.F. 2016. The ecological and evolutionary stability of interspecific territoriality. *Ecol. Lett.* 19:260–267.
- Mahler D.L., Revell L.J., Glor R.E., Losos J.B. 2010. Ecological opportunity and the rate of morphological evolution in the diversification of greater Antillean anoles. *Evolution* 64:2731–2745.
- Manceau M., Lambert A., Morlon H. 2016. A unifying comparative phylogenetic framework including traits coevolving across interacting lineages. *Syst. Biol.* 66:551–568. doi: 10.1093/sysbio/syw115.
- Mantel N. 1967. The detection of disease clustering and a generalized regression approach. *Cancer Res.* 27:209–220.
- Martin P.R., Montgomerie R., Loughheed S.C. 2010. Rapid sympatry explains greater color pattern divergence in high latitude birds. *Evolution* 64:336–347.
- Martin P.R., Montgomerie R., Loughheed S.C. 2015. Color patterns of closely related bird species are more divergent at intermediate levels of breeding-range sympatry. *Am. Nat.* 185:443–451.
- Martins E.P., Garland T. Jr. 1991. Phylogenetic analyses of the correlated evolution of continuous characters: a simulation study. *Evolution* 45:534–557.
- Matzke N.J. 2014. Model selection in historical biogeography reveals that founder-event speciation is a crucial process in island clades. *Syst. Biol.* 63:951–970.
- Medina-García A., Araya-Salas M., Wright T.F. 2015. Does vocal learning accelerate acoustic diversification? Evolution of contact calls in Neotropical parrots. *J. Evol. Biol.* 28:1782–1792.
- Morlon H., Lewitus E., Condamine F.L., Manceau M., Clavel J., Drury J. 2016. RPANDA: an R package for macroevolutionary analyses on phylogenetic trees. *Methods Ecol. Evol.* 7:589–597.
- Nuismer S.L., Harmon L.J. 2015. Predicting rates of interspecific interaction from phylogenetic trees. *Ecol. Lett.* 18:17–27.
- Pennell M.W., Eastman J.M., Slater G.J., Brown J.W., Uyeda J.C., FitzJohn R.G., Alfaro M.E., Harmon L.J. 2014. geiger v2.0: an expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. *Bioinformatics* 30:2216–2218.
- Pfennig K.S., Pfennig D.W. 2009. Character displacement: ecological and reproductive responses to a common evolutionary problem. *Q. Rev. Biol.* 84:253–276.
- Pigot A.L., Tobias J.A. 2014. Dispersal and the transition to sympatry in vertebrates. *Proc. R. Soc. B Biol. Sci.* 282:20141929.
- Rafferty N.E., Ives A.R. 2013. Phylogenetic trait-based analyses of ecological networks. *Ecology* 94:2321–2333.
- Ree R.H., Smith S.A. 2008. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Syst. Biol.* 57:4–14.
- Revell, Liam J. 2012. phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* 3.2: 217–223.
- Rezende E.L., Diniz-Filho J.A.F. 2012. Phylogenetic analyses: comparing species to infer adaptations and physiological mechanisms. *Compr. Physiol.* 2:639–674.
- Rezende E.L., Lavabre J.E., Guimaraes P.R. Jr., Jordano P., Bascompte J. 2007. Non-random coextinctions in phylogenetically structured mutualistic networks. *Nature* 448:925–929.
- Roncal J., Henderson A., Borchsenius F., Cardoso S.R.S., Balslev H. 2012. Can phylogenetic signal, character displacement, or random phenotypic drift explain the morphological variation in the genus *Geonoma* (Arecaceae)? *Biol. J. Linn. Soc.* 106:528–539.
- Schluter D. 2000. The ecology of adaptive radiation. Oxford (UK): Oxford University Press.
- Seddon N., Botero C.A., Tobias J.A., Dunn P.O., Macgregor H.E., Rubenstein D.R., Uy J.A., Weir J.T., Whittingham L.A., Safran R.J. 2013. Sexual selection accelerates signal evolution during speciation in birds. *Proc. R. Soc. B Biol. Sci.* 280:20131065.
- Tobias J.A., Cornwallis C.K., Derryberry E.P., Claramunt S., Brumfield R.T., Seddon N. 2014. Species coexistence and the dynamics of phenotypic evolution in adaptive radiation. *Nature* 506: 359–363.
- de Villemereuil P., Nakagawa S. 2014. General quantitative genetic methods for comparative biology. In: Garamszegi L., editor. *Modern phylogenetic comparative methods and their application in evolutionary biology*. New York, NY: Springer. p. 287–303.
- Webb C.O., Ackerly D.D., McPeck M.A., Donoghue M.J. 2002. Phylogenies and community ecology. *Annu. Rev. Ecol. Syst.* 33: 475–505.
- Weber M.G., Mitko L., Eltz T., Ramírez S.R. 2016. Macroevolution of perfume signalling in orchid bees. *Ecol. Lett.* 19:1314–1323.
- Weber M., Wagner C., Best R., Harmon L., Matthews B. 2017. Evolution in a community context: on integrating ecological interactions and macroevolution. *Trends Ecol. Evol.* 32:291–304.
- Weir J.T., Lawson A. 2015. Evolutionary rates across gradients. *Methods Ecol. Evol.* 6:1278–1286.
- Weir J.T., Mursleen S. 2013. Diversity-dependent cladogenesis and trait evolution in the adaptive radiation of the auks (Aves: Alcidae). *Evolution* 67:403–416.
- Weir J.T., Price T.D. 2011. Limits to speciation inferred from times to secondary sympatry and ages of hybridizing species along a latitudinal gradient. *Am. Nat.* 177:462–469.
- Weir J.T., Wheatcroft D. 2011. A latitudinal gradient in rates of evolution of avian syllable diversity and song length. *Proc. R. Soc. B Biol. Sci.* 278:1713–20.
- Willis P.M., Symula R.E., Lovette I.J. 2014. Ecology, song similarity and phylogeny predict natural hybridization in an avian family. *Evol. Ecol.* 28:299–322.