A potential solution to the unresolved challenge of false correlation between discrete characters

James D. Boyko[1,*] and Jeremy M. Beaulieu[1]

[1]*Department of Biological Sciences, University of Arkansas, Fayetteville, Arkansas, 72701 USA*

∗Correspondence to be sent to: *Department of Biological Sciences, University of Arkansas, Fayetteville, Arkansas, 72701 USA, E*-mail: jboyko@uark.edu

Abstract

The correlation between two characters is often interpreted as evidence that there exists a significant and biologically important relationship between them. However, Maddison and FitzJohn (2015) recently pointed out that in certain situations find evidence of correlated evolution between two categorical characters is often spurious, particularly, when the dependent relationship stems from a single replicate deep in time. Here we will show that there may, in fact, be a statistical solution to the problem posed by Maddison and FitzJohn (2015) naturally embedded within the expanded model space afforded by the hidden Markov model (HMM) framework. We demonstrate that the problem of single unreplicated evolutionary events manifests itself as rate heterogeneity within our models and that this is the source of the false correlation. Therefore, we argue that this problem is better understood as model misspecification rather than a failure of comparative methods to account for phylogenetic pseudoreplication. We utilize HMMs to develop a multi-rate independent model which, when implemented, drastically reduces support for correlation. The problem itself extends beyond categorical character evolution, but we believe that the practical solution presented here may lend itself to future extensions in other areas of comparative biology.

Key words: Macroevolution, phylogenetic comparative methods, hidden Markov models, rate heterogeneity, model adequacy.

Correlated or dependent evolution on a macroevolutionary scale is defined as a change in a character state (e.g., plumage color) that is linked to the presence of a particular state in a separate character (e.g., beak color). In other words, the evolution of character X can be said to be dependent on character Y if in the presence of a particular state of Y (e.g., Y_0), shifts within character X occur in a different way from when the lineage is in an alternative state of Y (e.g., Y_1). For example, a shift from X_0 to X_1 may occur more quickly when paired with Y_1 than with Y_0 resulting in a distribution with many character pairs X_1Y_1. It is often the case that these sorts of dependent relationships between characters seem obvious, especially if the observations of many individuals are consistent.

However, what happens when all observations of the pair come from, for example, one biogeographic region? In other words, there may have been many individual pairs of X_1Y_1 observed, but they all came from one population. Since the strength of the relationship is related to the number of individual observations, the non-independence of them raises concerns about the validity of the proposed correlation. This problem extends to interspecific comparisons too, but rather than observations being linked to one of two populations, they are associated with particular taxonomic groupings and shared histories. This fact was well understood as early as Darwin (1859) and the tools for dealing with the resulting statistical non-independence have been available to comparative biologists since the foundational work of Felsenstein (1985). Nevertheless, this issue of “phylogenetic pseudoreplication”, where species are non-independent due to their shared ancestry, served as the basis for the concerns raised by Maddison and FitzJohn (2015) regarding tests of dependent character evolution.

Maddison and FitzJohn (2015) demonstrated that the most widely used phylogenetic method for detecting correlated evolution between categorical characters (Pagel 1994), almost always indicates strong evidence of correlation when singular events deep in time can account for the co-distribution of two characters. To demonstrate their point, they fit correlated models to datasets generated under their so-called, “Darwin's” and the “Unreplicated Burst” scenarios (Fig. 1). Darwin’s scenario results in the perfect co-distribution of two characters, which in practice, might occur when testing for correlations between two synapomorphies (e.g., presence/absence of middle ear bones and fur). Under the Unreplicated Burst scenario, only one of the two characters has phylogenetically replicated change. This scenario occurs when one of the characters is a synapomorphy for the clade, with the other character undergoing several changes within the focal clade. The issue is that, when applied to either Darwin’s or the Unreplicated Burst scenario, commonly used comparative methods (Pagel 1994) will almost always indicate

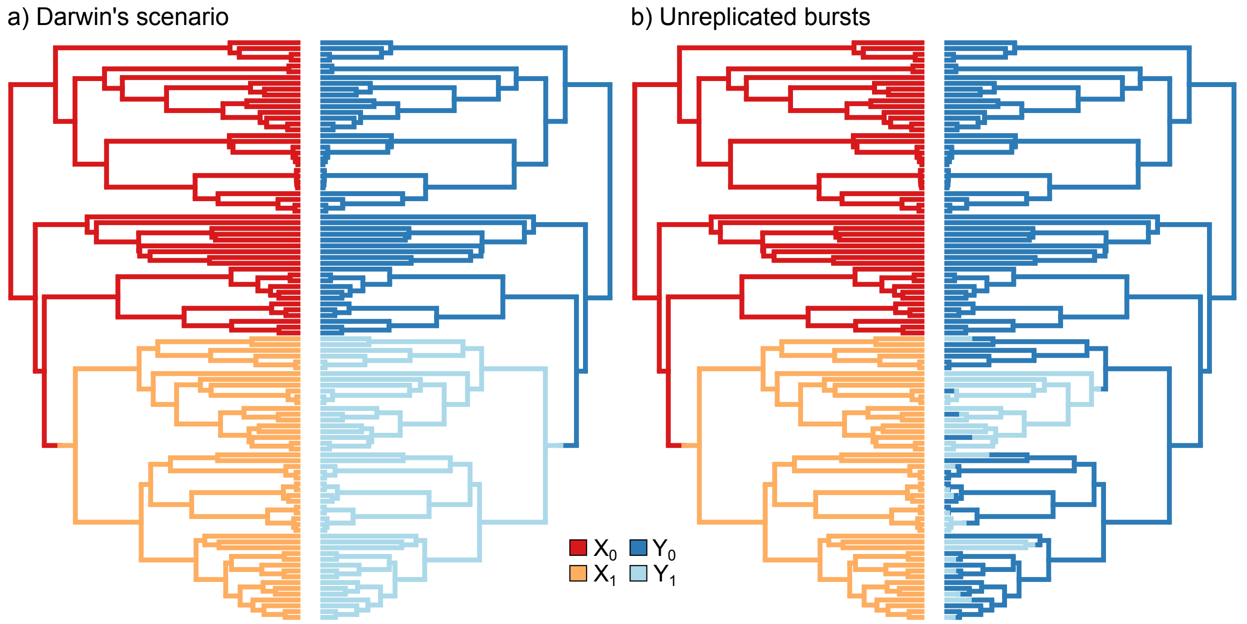


Figure 1. The two problematic scenarios from Maddison and FitzJohn (2015) for the evolution of characters X and Y. Character X is painted on the left phylogeny using red and orange for state X_0 and X_1, whereas character Y is painted on the right phylogeny using dark blue and light blue for state Y_0 and Y_1. a) Darwin’s scenario is depicted as a single event deep in time that has led to the co-distribution of X_0Y_0 outside of the focal clade and X_1Y_1 within the focal clade. b) Unreplicated bursts scenario is where a single event deep in time has led to the co-distribution of X_0Y_0 outside of the focal clade and X_1Y_0 and X_1Y_1 within the focal clade.

strong evidence of correlation despite the dependent relationship arising from little more than a single event deep in time.

There is considerable interest in understanding and, ultimately, finding a resolution to the problem posed by Maddison and FitzJohn (2015). Recently, Uyeda et al. (2018) suggested that for Darwin's scenario, the relatively long periods of stasis between the two characters (i.e., minimal trait change) is the primary cause for their significant dependent relationship. In fact, they showed that the probability of selecting a character-dependent model (i.e., a model of correlated evolution between the two characters) over a character-independent model (i.e., a model where the two characters are explicitly not correlated) was proportional to the ratio between the length of the branch where the shift occurred and the total length of the tree. The nature of this ratio ensured that a correlated model would always be supported in cases where singular evolutionary events led to a co-distribution of characters. Another study, by Gardner and Organ (2021), tested a variety of correlated models beyond Markov models and examined the structure of datasets which are susceptible to the problem of false dependence. They found that all the tested comparative methods produced erroneous correlations when datasets were phylogenetically pseudoreplicated.

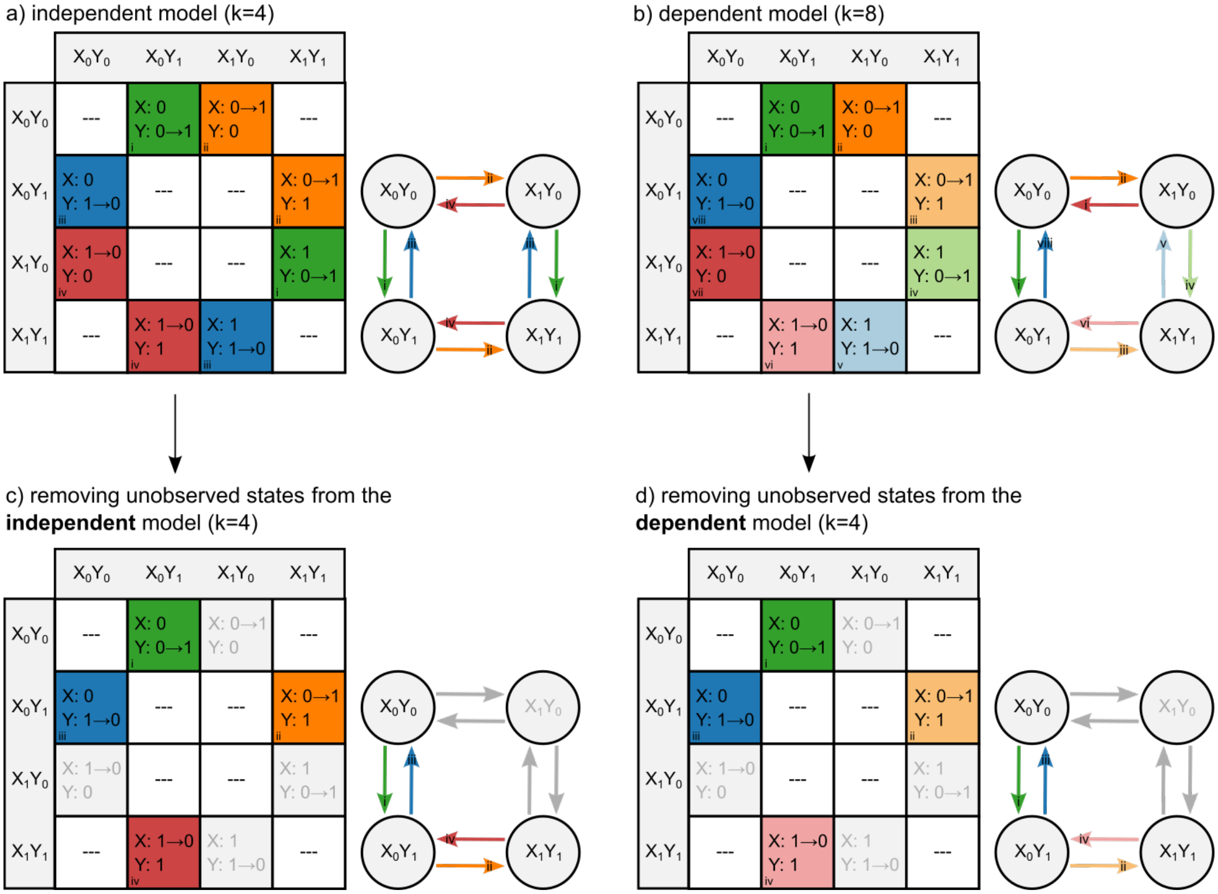
In both of these studies, the authors have addressed the problem by encouraging scientists to think critically about their models. While this recommendation is certainly admirable and correct, it is not a direct and satisfying solution to the statistical problems presented so far, as no amount of methodological vigilance will ever prevent analyses from being marred by phylogenetic pseudoreplication. However, prior analyses have limited model comparisons to only a few models, and have overlooked the very large set of alternative Markov models which can also be consistent with correlation or independence depending on the model's structure.

These alternative models have been briefly discussed previously (Pagel 1994; Pagel and Meade 2006) and, as we will show, the inclusion of a few examples within the model set can play a crucial role in ensuring a fair test of correlation. These underrepresented models, in addition to enormous model space provided by hidden Markov models (HMMs) for addressing rate heterogeneity across the tree (Beaulieu et al. 2013; Boyko and Beaulieu 2021), form the basis of our putative statistical solution to the problem posed by Maddison and FitzJohn (2015). We acknowledge that the problem itself extends beyond categorical character evolution, but we believe that the practical solution presented here may lend itself to future extensions in other areas.

We draw on two important insights as they relate to models of categorical character evolution. The first is that model space is severely underexplored and that the inclusion of more complex, character-independent models within our modeling set helps reduce evidence of false correlation. We note that estimates of transition rates to and from unobserved character states are not statistically identifiable, revealing that the canonical character-dependent model is over-parameterizationed in phylogenetically pseudoreplicated datasets like Darwin's scenario (Fig. 1a). When only two or three of the four possible character state combinations are observed, we produce models nested within the correlated and independent model that are overwhelmingly favored over both. Second, the issue of false dependent relationships is not one of stasis *per se*, but rather, a failure to account for rate heterogeneity. We demonstrate that an explicit character-independent hidden Markov model (HMM) provides significant evidence for models of independent evolution in cases where a correlated model would have previously been supported. This is because under the classic Pagel (1994) framework, support for correlation comes from both a dependent relationship between characters and a strong signal of rate heterogeneity. By amending the Pagel framework with a model which allows for rate heterogeneity independent of a focal character, we correct the bias towards correlation. We also reiterate that the relative support of each model should be considered when interpreting biologically sound results rather than examining tests of character dependence against “trivial” nulls (Beaulieu and O’Meara 2016; Caetano et al. 2018; O’Meara and Beaulieu 2021).

Correlated models depend on observations of intermediate states

While much has been written about the specifics of Pagel’s model, we briefly review aspects of it in order to better illustrate our point — namely, that certain transition rates are not estimatable and that their inclusion may be an additional cause of false correlations uncovered by Maddison and FitzJohn (2015). The correlated or dependent model of discrete character evolution, introduced by Pagel (1994), uses a continuous-time Markov process to estimate the rate of transitions between character states (Fig. 2ab). With a single binary character, *X*, the transition rate matrix, denoted as **Q**, is a simple 2x2 matrix, which contains all the information necessary to estimate the probability of a transition occurring between two states of character *X* over a given period of time. At its most complex, **Q** would contain two transition rates: from state X_0 to state X_1, and from state X_1 to state X_0. If we introduce a second binary character, *Y*, the number of possible observed state combinations is expanded — that is, the possible observed state combinations become X_0Y_0, X_0Y_1, X_1Y_0, and X_1Y_1. Consequently, this requires an expansion of **Q** to a 4x4 matrix, to account for all the possible transitions between state combinations. This model is considerably more complex, as the number of transitions goes from a maximum of two to a maximum of 12. However, the model introduced by Pagel (1994) is constrained specifically for the purpose of detecting correlations between characters by examining whether the state of one variable affects the probability of change in the other. To do this, dual transitions (i.e., changes in both *X* and *Y* occurring in a single time step) are removed. As noted by Pagel (1994), setting dual transition rates to zero does not rule out dual transitions over long periods of time.



a) independent model (k=4)

	X_0Y_0	X_0Y_1	X_1Y_0	X_1Y_1
X_0Y_0	---	X: 0 Y: 0→1 i	X: 0→1 Y: 0 ii	---
X_0Y_1	X: 0 Y: 1→0 iii	---	---	X: 0→1 Y: 1 ii
X_1Y_0	X: 1→0 Y: 0 iv	---	---	X: 1 Y: 0→1 i
X_1Y_1	---	X: 1→0 Y: 1 iv	X: 1 Y: 1→0 iii	---

b) dependent model (k=8)

	X_0Y_0	X_0Y_1	X_1Y_0	X_1Y_1
X_0Y_0	---	X: 0 Y: 0→1 i	X: 0→1 Y: 0 ii	---
X_0Y_1	X: 0 Y: 1→0 viii	---	---	X: 0→1 Y: 1 iii
X_1Y_0	X: 1→0 Y: 0 vii	---	---	X: 1 Y: 0→1 iv
X_1Y_1	---	X: 1→0 Y: 1 vi	X: 1 Y: 1→0 v	---

c) removing unobserved states from the **independent** model (k=4)

	X_0Y_0	X_0Y_1	X_1Y_0	X_1Y_1
X_0Y_0	---	X: 0 Y: 0→1 i	X: 0→1 Y: 0	---
X_0Y_1	X: 0 Y: 1→0 iii	---	---	X: 0→1 Y: 1 ii
X_1Y_0	X: 1→0 Y: 0	---	---	X: 1 Y: 0→1
X_1Y_1	---	X: 1→0 Y: 1 iv	X: 1 Y: 1→0	---

d) removing unobserved states from the **dependent** model (k=4)

	X_0Y_0	X_0Y_1	X_1Y_0	X_1Y_1
X_0Y_0	---	X: 0 Y: 0→1 i	X: 0→1 Y: 0	---
X_0Y_1	X: 0 Y: 1→0 iii	---	---	X: 0→1 Y: 1 ii
X_1Y_0	X: 1→0 Y: 0	---	---	X: 1 Y: 0→1
X_1Y_1	---	X: 1→0 Y: 1 iv	X: 1 Y: 1→0	---

Figure 2. Representations of the different transition rate matrices, **Q**, with *k* number of parameters associated with each. Where transitions are fixed to occur at the same rate, the squares are colored to be the same. Unique parameters are also indicated with a roman numeral in the bottom left corner of the square. To the right of each matrix, a ball and stick representation of the model is presented with colors and parameter numbers matching the transitions indicated in the matrix, **Q**. The ball and stick representation is organized such that internal arrows represent transitions from 1 to 0, and external arrows represent transitions from 0 to 1. Additionally, arrows which cross the vertical midpoint indicate transitions in character *X*, whereas transitions across the horizontal midpoint indicate transitions in character *Y*. a) An independent model with four unique parameters, which fixes transitions within a character such that changes in *X* or *Y* do not depend on the state of the other character. b) A dependent model with eight unique parameters, whichs model allows transitions within a character to depend on the state of the other character. c) A model which removes transitions to and from an unobserved state from the independent model (a). d) A model that removes transitions to and from an unobserved state from the dependent model (b). In (c) and (d) the unobserved state is based on the Unreplicated Burst scenario where X_0Y_1 is not observed.

Rather, a dual transition from X_0Y_0 must first pass through state X_0Y_1 or X_1Y_0, before finally transitioning to X_1Y_1. Equating the rates of transitions between particular pathways allows for the construction and testing of an independent model (Pagel and Meade 2006). A model of independent evolution is nested within the correlated model but assumes that the transition rates between states of a character are equal to one another regardless of the state of the other character (e.g., [X_0 to X_1 | Y_0] = [X_0 to X_1 | Y_1]; Fig. 2ab). In other words, if these two characters, X and Y, are independent, the presence of one character will have no influence on the change of the other and thus model selection criteria should choose the simpler model.

Using this specific nested framework, we were able to replicate the results of Maddison and Fitzjohn (2015). Specifically, we generated 100 datasets for Darwin's scenario and the Unreplicated Bursts scenario. Phylogenies were simulated under a λ=1 and μ=0.5 until 100 extant taxa were reached, and each resulting tree was then evaluated for a focal monophyletic group between 40 and 60 taxa. For Darwin's scenario, extant species within the focal clade were assigned X_1Y_1 and species outside the clade were assigned X_0Y_0. We simulated Unreplicated Bursts by assigning all species outside the focal clade X_0, and all species within the clade X_1. Next, character Y was simulated at a rate of 100 transitions per million years. Outside of the focal clade, species were assigned Y_0 whereas within the focal clade, the simulated data resulted in both Y_0 and Y_1. We used *corHMM* (Beaulieu et al. 2013; Boyko and Beaulieu 2021) to fit and compare the four-state independent model (Fig. 2a) against the four-state correlated model (Fig. 2b) using Akaike Information Criterion (AIC). In all cases, we found overwhelming support for the correlated model for both Unreplicated Bursts and Darwin's scenario datasets (See Supplemental Materials). The mean AIC weight for the correlated model under Darwin's scenario was 92.52% and under Unreplicated Bursts it was 99.96%. As expected, an independent model was never favored over a correlated model in either scenario.

For Darwin's scenario, setting aside the critical analytical issues regarding phylogenetic pseudoreplication, we had additional concerns with the structure of the data and how this might impact estimates of transition rates. Under any continuous-time Markov process, the estimates of the transition rates among all possible character combinations are reflective of the observed state frequencies and distribution at the tips. But, what if two of the four character combinations are not observed at all? Here we are referring to the two combinations, X_0Y_1 and X_1Y_0, not observed in any of the tips under Darwin's scenario. There may be biological reasons for not observing intermediate state combinations. For example, these combinations may be at some selective disadvantage, resulting in rapid transitions to another, more viable character combination (e.g., X_0Y_0 or X_1Y_1). Alternatively, it could be that one or both combinations are never possible due to some underlying genetic or developmental reasons (e.g., certain fruit character combinations, see Beaulieu and Donoghue 2013). However, whatever biological meaning is attributed to the lack of intermediate character state observations, in this case, is beside the point. There seems to be obvious, and yet unrecognized, identifiability issues with including transitions to and from these unobserved state combinations in the model, calling into question fitting the correlated model to these types of data. That is to say, if we never see intermediate state combinations at the tips, how can the model ever favor one pathway over the other?

To illustrate this point, we examined the likelihood surface of one of the datasets simulated under Darwin's scenario and fit under Pagel's correlated model (Fig. 3). Whether starting from X_0Y_0 or X_1Y_1, transition rate estimates to either of the unobserved character combinations fall along a ridge of equal likelihood, where changing the rate of transition to one unobserved state determines the rate for the transitions to the other unobserved state. When a lineage transitions into one of the states, the likelihood surface for transitions out of these states to either state X_0Y_0 or X_1Y_1 are completely flat, with all rates ranging from 0.1 to 100 transitions per unit time all having nearly identical likelihoods. Taken together, the preferred model estimates for various transition rates arise simply by chance of the optimization procedure, but more importantly, there are parameters which are clearly unneeded to explain the data.

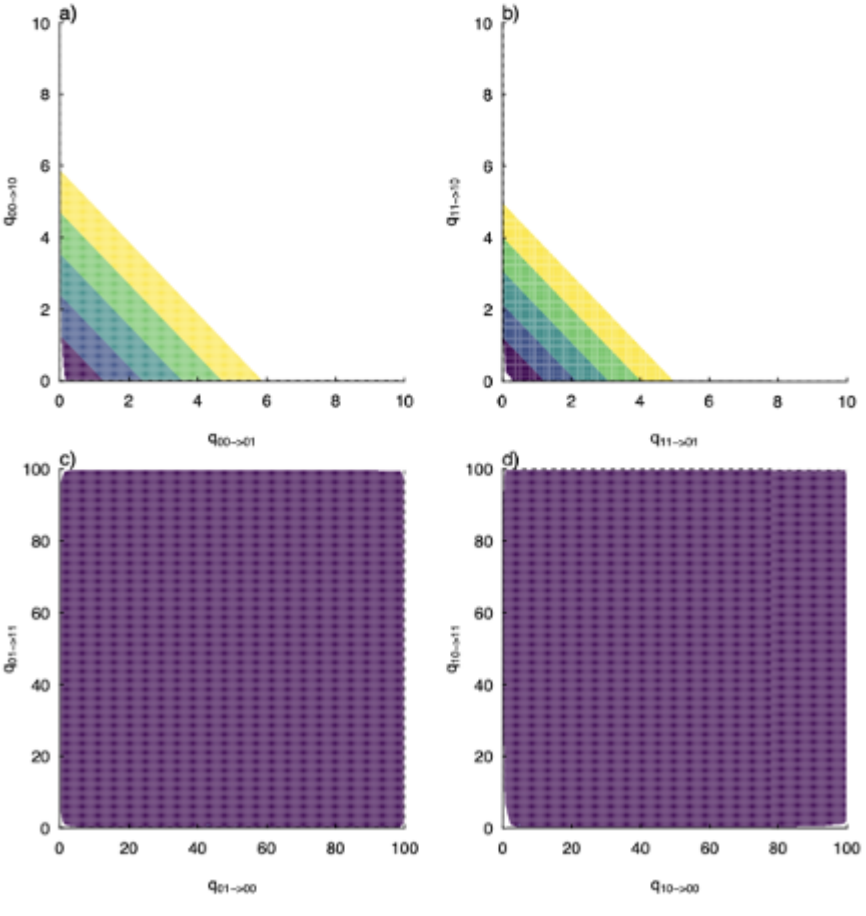


Figure 3. An example likelihood surface of a correlated model when applied to one of the 100 Darwin's scenario datasets. The color of the plot indicates the likelihood of a particular pair of parameters when the remaining transition rates are optimized. Thus, each point represents the maximum likelihood estimate when the transition rates indicated by the axes are fixed. a) Transitions from X_0Y_0 to an intermediate state result in several likelihood ridges. b) Transitions from X_1Y_1 to an intermediate state result in several likelihood ridges. c) Transitions from X_0Y_1 to either X_0Y_0 or X_1Y_1 result in a completely flat likelihood surface. d) Transitions from X_1Y_0 to either X_0Y_0 or X_1Y_1 result in a completely flat likelihood surface.

One obvious solution is to simply remove the unobserved character combinations from the model completely. From a modeling perspective, removing unobserved states removes the parameters that fall along the likelihood ridge and should lead to a model that ends up being well estimated. Consequently, the question of whether independent or dependent models better explain the data becomes irrelevant as the two models collapse into one another when unobserved states are removed (Fig. 2c,d). This is clearly seen when the collapsed model is applied to an Unreplicated Burst scenario. Whether one starts with an independent model (Fig. 2a) or a correlated model (Fig. 2b), once unobserved states are removed, comparing alternative transition pathways between X_0Y_0 and X_1Y_1 are no longer possible. For example, take transitions between states of character X. Both the correlated and independent models estimate transitions from X_0 to X_1 as depending only on Y_0, since X_0Y_1 is not observed in the dataset. Since it is not possible to compare the likelihood of alternative scenarios of dependence a comparison of correlation and independence becomes irrelevant.

Including a collapsed model as part of our model set drastically changes the results. We found complete support for a collapsed state model for both Darwin's scenario and Unreplicated Bursts (see Supplemental Materials). The average AIC weight for the collapsed model is 99.7% under Darwin's scenario and 100.0% under an Unreplicated burst scenario. This suggests that the support for the correlated models over simpler independent models is a result of an intuitive, but necessary parameter constraint. Specifically, in an independent model, transitions between observed states are constrained to be identical to transitions between unobserved states (e.g., X_0Y_0 to X_0Y_1 must be identical to X_0Y_1 to X_1Y_1, even if X_0Y_1 is never observed). In contrast, the correlated model is not subject to these constraints. This is, of course, the important distinction between the two models and what allows us to test for correlated evolution. In this case, the support for dependence occurs because, in a sense, the correlated model is free to “throw away” the inestimatable transition rates which describe movement to and from intermediate states, while the independent model is forced to evaluate them. However, this issue becomes moot when exclusively modeling observed state combinations because the dependent and independent models become equivalent descriptions of the evolutionary process and are, therefore, indistinguishable for the given data.

Rate heterogeneity is necessary when testing for correlation between categorical variables

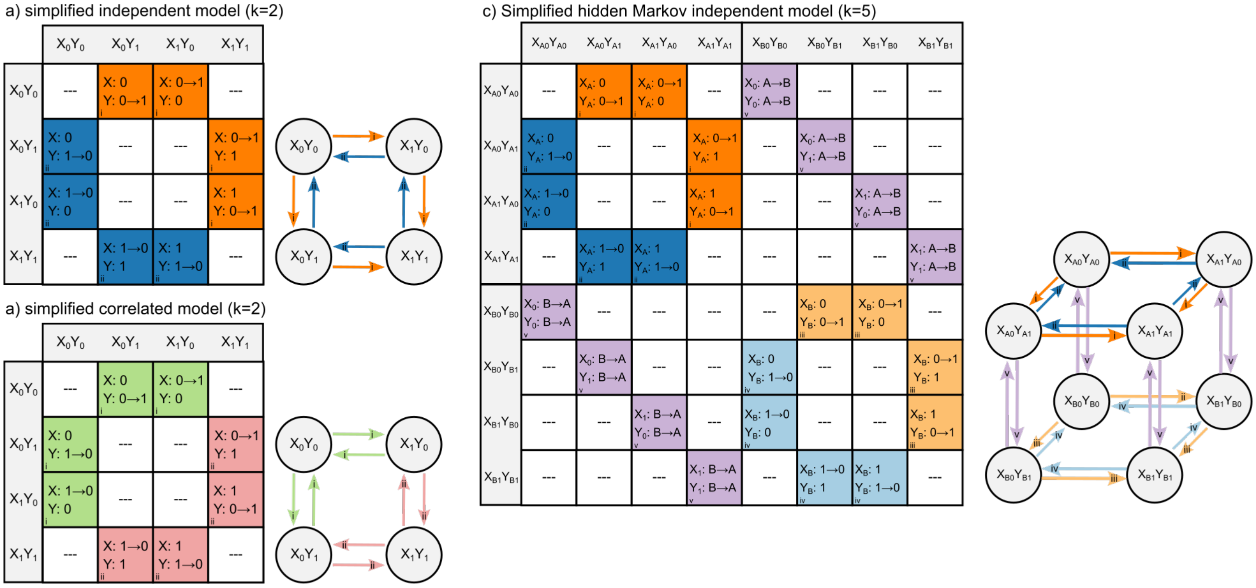
A major issue for the collapsed model described above is that in Darwin’s scenario, a single observation of X_0Y_1 and X_1Y_0 removes the possibility of collapsing the model structure. This suggests that modeling only observed state combinations is not a generalizable solution to the phylogenetic pseudoreplication of categorical characters. As we will show, with only a single observation of intermediate character combinations, support for the correlated model over an independent model remains substantial. Even so, the results above highlight information limitations and that the strong evidence for dependent models may be due to a lack of viable alternative independent models rather than being irrefutable evidence of correlation.

It is worth considering again the possible explanations of the data under Darwin’s scenario. One possibility is that the characters X and Y evolve slowly and that their co-distribution is the result of two independent events deep in time. The probability of this scenario has been explored in-depth and its implausibility is a major contributor to the recurrent issues of false correlation when comparing dependent and independent models (Uyeda et al. 2018). We propose a complementary explanation for the correlated model’s support: the independent model structure fixes the transition X_0 to X_1 to always be the same rate in the context of the state of Y

(Fig. 2a), whereas a dependent model structure allows transitions from X_0 to X_1 to vary depending on the state of Y (Fig. 2b). Support for the correlated model, therefore, comes from the fact that the best explanation of the data is not one that has a single slow transition rate for the characters. Instead, the most likely description of the process is one in which transitions between X_0 and X_1 or Y_0 and Y_1 are allowed to occur rapidly within the focal clade and occur slowly outside of the focal clade. The relative stasis of X_0 outside the focal clade and the rapid accumulation of X_1 within the clade suggests that changes in X are not consistent throughout the tree.

Hidden Markov models (HMMs) are a natural way to deal with this kind of rate heterogeneity across the tree. The underlying mathematical framework of an HMM is no different than a typical Markov model. They utilize a rate matrix, $\mathbf{Q}$, to estimate the probabilities of transitioning between discrete states and arrive at the likelihood of the model given the observed dataset (Felsenstein and Churchill 1996). However, HMMs introduce a so-called “hidden-state”, which can represent any number of unobserved factors, biological or otherwise. Based on the presence or absence of this hidden-state, changes between observed states are allowed to vary. In the most extreme cases, the absence of the hidden state may halt the evolutionary process and result in periods of stasis. For example, Marazzi et al. (2012) conceptualized the hidden-state as a “precursor” trait and only in its presence could extrafloral nectaries (EFNs) emerge. It is important to emphasize that the precursor state was never directly observed and that the information for its presence or absence of the hidden state came from the rate heterogeneity of EFNs transitions. In some parts of the tree, the model EFNs emerged rapidly and in others there were periods of stasis. Of course, HMMs are more general than either halting or actuating the evolutionary process and are used to quantify rate heterogeneity without the necessity of stasis (e.g., comparing fast, slow, or intermediate rates as in Beaulieu et al. 2013). The key point here is that they allow for rate heterogeneity that is unlinked to another observed character.

We developed and tested a hidden Markov independent model (HMIM) which accounts for rate heterogeneity while maintaining the independence of the observed focal characters *X* and *Y* (Fig. 4). In our view, the inclusion of our model within the evaluated set better levels the playing field between correlated and independent models. For example, if we focus on character *X*, our proposed model utilizes hidden states to vary transition rates between X_0 and X_1 based on an unobserved character. This is similar to the way that the correlated model allows transition rates between X_0 and X_1 to differ based on the observed state of *Y*. If the cause of false correlation was, as we suspect, not accounting for rate heterogeneity, then both the hidden state



a) simplified independent model (k=2)

	X_0Y_0	X_0Y_1	X_1Y_0	X_1Y_1
X_0Y_0	---	X: 0 Y: 0→1 i	X: 0→1 Y: 0 i	---
X_0Y_1	X: 0 Y: 1→0 ii	---	---	X: 0→1 Y: 1 i
X_1Y_0	X: 1→0 Y: 0 ii	---	---	X: 1 Y: 0→1 i
X_1Y_1	---	X: 1→0 Y: 1 ii	X: 1 Y: 1→0 ii	---

a) simplified correlated model (k=2)

	X_0Y_0	X_0Y_1	X_1Y_0	X_1Y_1
X_0Y_0	---	X: 0 Y: 0→1 i	X: 0→1 Y: 0 i	---
X_0Y_1	X: 0 Y: 1→0 i	---	---	X: 0→1 Y: 1 ii
X_1Y_0	X: 1→0 Y: 0 i	---	---	X: 1 Y: 0→1 ii
X_1Y_1	---	X: 1→0 Y: 1 ii	X: 1 Y: 1→0 ii	---

c) Simplified hidden Markov independent model (k=5)

	$X_{A0}Y_{A0}$	$X_{A0}Y_{A1}$	$X_{A1}Y_{A0}$	$X_{A1}Y_{A1}$	$X_{B0}Y_{B0}$	$X_{B0}Y_{B1}$	$X_{B1}Y_{B0}$	$X_{B1}Y_{B1}$
$X_{A0}Y_{A0}$	---	X_A: 0 Y_A: 0→1 i	X_A: 0→1 Y_A: 0 i	---	X_0: A→B Y_0: A→B v	---	---	---
$X_{A0}Y_{A1}$	X_A: 0 Y_A: 1→0 ii	---	---	X_A: 0→1 Y_A: 1 i	---	X_0: A→B Y_1: A→B v	---	---
$X_{A1}Y_{A0}$	X_A: 1→0 Y_A: 0 ii	---	---	X_A: 1 Y_A: 0→1 i	---	---	X_1: A→B Y_0: A→B v	---
$X_{A1}Y_{A1}$	---	X_A: 1→0 Y_A: 1 ii	X_A: 1 Y_A: 1→0 ii	---	---	---	---	X_1: A→B Y_1: A→B v
$X_{B0}Y_{B0}$	X_0: B→A Y_0: B→A v	---	---	---	---	X_B: 0 Y_B: 0→1 iii	X_B: 0→1 Y_B: 0 iii	---
$X_{B0}Y_{B1}$	---	X_0: B→A Y_1: B→A v	---	---	X_B: 0 Y_B: 1→0 iv	---	---	X_B: 0→1 Y_B: 1 iii
$X_{B1}Y_{B0}$	---	---	X_1: B→A Y_0: B→A v	---	X_B: 1→0 Y_B: 0 iv	---	---	X_B: 1 Y_B: 0→1 iii
$X_{B1}Y_{B1}$	---	---	---	X_1: B→A Y_1: B→A v	---	X_B: 1→0 Y_B: 1 iv	X_B: 1 Y_B: 1→0 iv	---

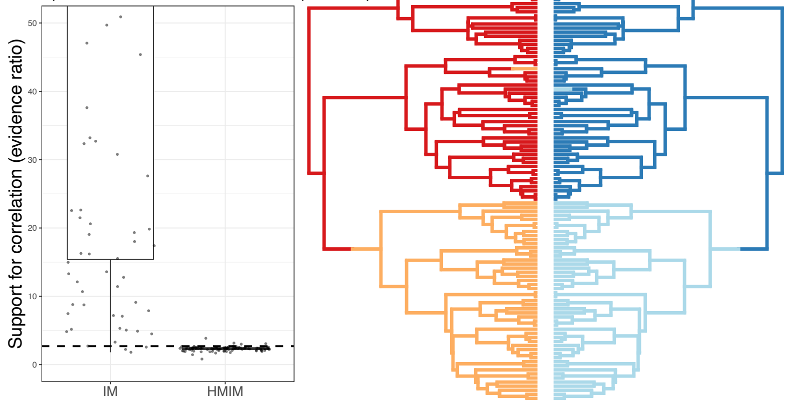
Figure 4. The hidden Markov independent model (HMIM), which allows transitions within a character to have rate heterogeneity without it necessarily being linked to an observed character. This matrix can be read as a block matrix, with 4x4 blocks representing transitions between observed characters following an independent model (top left and bottom right) and transitions between hidden rate classes *A* and *B* (top right and bottom left). The independent model is essentially duplicated in the top left (blue and green) and bottom right (red and orange) of the block matrix with transitions occurring between these different types of independent models (purple). Here, transition rates between the hidden states are fixed to be the same (parameter ix), but it is straightforward to allow the transition between rate class *A* and *B* to differ.

independent and correlated model should be preferable to the simple independent model and evidence of correlation between X and Y should be greatly reduced.

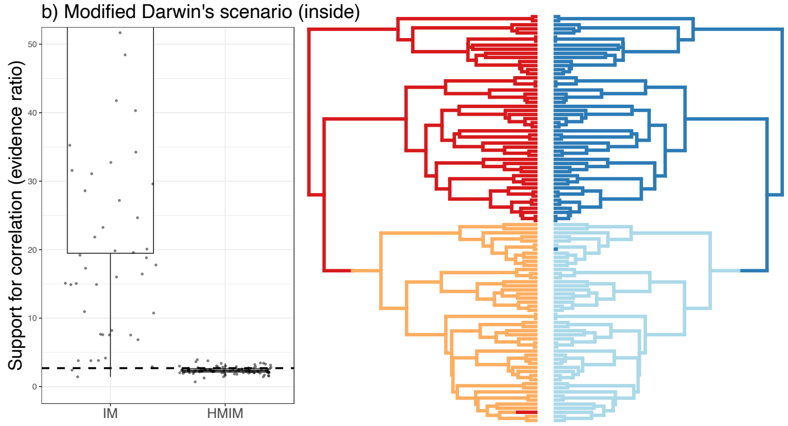
We first removed the possibility of collapsing the Markov model by modifying Darwin's scenario. We defined the focal clade as being the monophyletic group where all observations of X_1Y_1 occur and randomly add the intermediate state observations of X_0Y_1 and X_1Y_0 within the focal clade (which refer to as "inside" hereafter), outside of the focal clade (which we refer to as "outside" hereafter), and both within and outside the focal clade (which refer to as "both" hereafter) (Fig. 5). Next, we verified that this modified Darwin's scenario still suffers from the problems of the original Darwin's scenario by comparing the independent and correlated models *sensu* Pagel (1994). We then added the hidden Markov independent model to the model set and evaluated two questions: (1) when comparing independent models to one another, is there evidence of rate heterogeneity? and (2) is support for the correlated model reduced when compared to an independent model with rate heterogeneity? In addition to AIC weight, we utilized evidence ratios (*ER*) to explore the relative likelihood of our models. Evidence ratios are a simple extension of AIC weights, but as a means of evaluation, are important here since they allow us to focus on evaluating the relative evidence of pairs of models irrespective of other models in the set (Burnham and Anderson 2002). The evidence for model *i* over model *j* is the ratio between their AIC weights: $ER = w_i / w_j$ and it can help quantify whether the best model in our comparison is convincingly best. With alternative samples, a convincingly best model is likely to be chosen again sample to sample. However, if evidence for a model is low, we expect model selection uncertainty to be high. Following Burnham and Anderson (2002), an evidence ratio of greater than 2.7 is used as a guide to justify judging support for one model being better than another. This also neatly corresponds to a ΔAIC = 2. We emphasize that this value should *not* be misconstrued as a significant test in a frequentist sense since we are not evaluating the probability of rejecting a null hypothesis.

For all modified Darwin's scenarios, we found substantial evidence (ER > 2.7) for a correlated model over a single rate class independent model (Fig. 5). The geometric mean evidence ratio for the correlated model over the single rate independent model was $ER_{outside}$= 59.51, ER_{inside}= 78.16, ER_{both}= 11.44 (Fig. 5), thus we, again, successfully recreated the conditions of Maddison and FitzJohn (2015) under a modified Darwin's scenario. Next, we examined the evidence for rate heterogeneity by comparing a single rate independent model to the hidden Markov independent model. We found substantial evidence for rate heterogeneity across all scenarios, with all mean evidence ratios of the HMIM over the standard independent model well over 20, indicating substantial support for rate heterogeneity ($ER_{outside}$= 24.45, ER_{inside}= 24.33, ER_{both}= 50.45). Finally, we tested whether there is still conclusive evidence of correlation between characters if we include the hidden state independent model within our modeling set. We found that the evidence for a correlated model over the hidden Markov independent model was greatly reduced when compared to the single rate class independent model (Fig. 5; $ER_{outside}$= 2.43, ER_{inside}= 3.21, ER_{both}= 0.22; Fig. 5). In fact, with only two observations of each intermediate state combination (X_0Y_1 and X_1Y_0), support for the hidden Markov independent model over the correlated model was substantial (evidence for HMIM over a correlated model: $\mathrm{ER}_{\mathrm{both}}$= 4.41). Taken together, these findings suggest that 1) there is indeed substantial evidence of rate heterogeneity, and that this is causing the signal of false correlation; and 2) including a hidden Markov independent model can, at least, muddle evidence for correlation.

a) Modified Darwin's scenario (outside)



b) Modified Darwin's scenario (inside)



c) Modified Darwin's scenario (both)

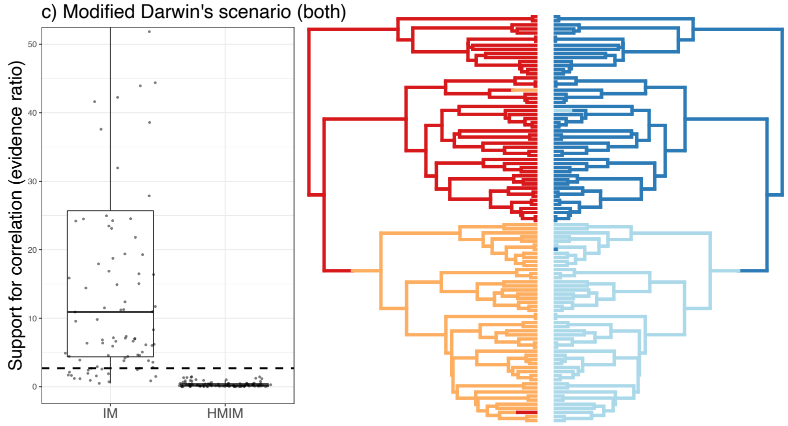


Figure 5. The amount of evidence for correlation when comparing a correlated model to ether an independent model (IM) or hidden Markov independent model (HMIM). The models are fit to data of the modified version of Darwin's scenario where a single observation of X_0Y_1 and X_1Y_0 is added outside of the focal clade (a), inside of the focal clade (b), and both within and outside of the focal clade (c). Evidence ratios for each m`odel comparison are plotted as boxplots to the left of the simulation scenario. In all cases, the evidence ratio of the correlated model over the independent model is substantially greater than 2.7 (left boxplot) but, the correlated model receives much less support over the hidden Markov independent model (right boxplot).

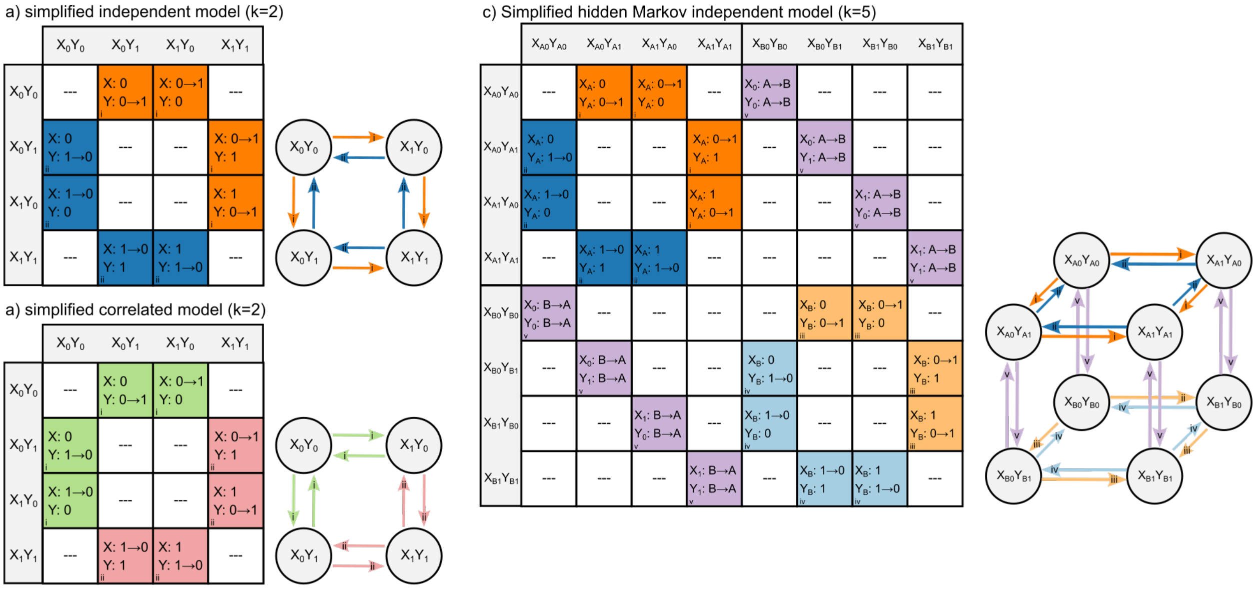
A (potentially) complete solution to biased correlation between synapomorphies

It was still concerning to us that for the original and two of the modified Darwin's scenarios (specifically the "outside" and "inside" sets; see Fig. 5), support for the correlated model was still often greater than the hidden state independent model. Although the addition of character independent rate heterogeneity muddles support for the correlated model, in the most extreme cases the best model remained the dependent model. To deal with this issue, we applied what we learned thus far, with regards to the over-parameterization of models and the necessity of rate heterogeneity and added a new set of simpler and nested models within the set presented thus far to specifically address the issues of Darwin's scenario.

It is critical to emphasize that model space has been underexplored and that there are many nested model structures that are consistent with either independence or correlation depending on their constraints (see also Pagel and Meade 2006). Here we describe two constrained versions of the independent and correlated models that achieve the most efficient description of the data. One simplified version of the correlated model suggests that when either character X or Y is in state 0, rates of change are slower or faster than when either character is in state 1 (Fig. 6b). We refer to this as the "simplified correlated" model and it represents the simplest way to model a dependent relationship between two binary characters. Next, we created a "simplified independent" model of equal parameterization to the simplified correlated model, which equates all changes from 0 to 1 regardless of the character and the same is done for changes from 1 to 0 (Pagel and Meade 2006; Fig. 6a).

The structures of these simplified models have certain qualities that may make them apt descriptions of data like Darwin's scenario. Primarily, these models suggest that changes between states 0 and 1 do not necessarily depend on the specific identity of character X or Y

since they are constrained to be equal. When we consider the redundancy of a dataset composed of two synapomorphies, it is obvious that there is little to no information that distinguishes the two characters– that is, it makes no difference whether one analyzes character X or character Y since their distributions are identical. The simplified models make that assumption explicit. It is also important to note that the simplified independent model and simplified correlated model maintain independence and dependence *sensu* Pagel (1994). The background state of the unchanging character does not influence changes in the case of the simplified independent model, whereas the background state of the unchanging character will influence rates of change in the case of the simplified dependent model (Pagel and Meade 2006). Finally, we can introduce rate heterogeneity by modeling the simplified independent and correlated models as two rate class hidden Markov models (Fig. 6c).



a) simplified independent model (k=2)

	X_0Y_0	X_0Y_1	X_1Y_0	X_1Y_1
X_0Y_0	---	X: 0 Y: 0→1 i	X: 0→1 Y: 0 i	---
X_0Y_1	X: 0 Y: 1→0 ii	---	---	X: 0→1 Y: 1 i
X_1Y_0	X: 1→0 Y: 0 ii	---	---	X: 1 Y: 0→1 i
X_1Y_1	---	X: 1→0 Y: 1 ii	X: 1 Y: 1→0 ii	---

a) simplified correlated model (k=2)

	X_0Y_0	X_0Y_1	X_1Y_0	X_1Y_1
X_0Y_0	---	X: 0 Y: 0→1 i	X: 0→1 Y: 0 i	---
X_0Y_1	X: 0 Y: 1→0 i	---	---	X: 0→1 Y: 1 ii
X_1Y_0	X: 1→0 Y: 0 i	---	---	X: 1 Y: 0→1 ii
X_1Y_1	---	X: 1→0 Y: 1 ii	X: 1 Y: 1→0 ii	---

c) Simplified hidden Markov independent model (k=5)

	$X_{A0}Y_{A0}$	$X_{A0}Y_{A1}$	$X_{A1}Y_{A0}$	$X_{A1}Y_{A1}$	$X_{B0}Y_{B0}$	$X_{B0}Y_{B1}$	$X_{B1}Y_{B0}$	$X_{B1}Y_{B1}$
$X_{A0}Y_{A0}$	---	X_A: 0 Y_A: 0→1 i	X_A: 0→1 Y_A: 0 i	---	X_0: A→B Y_0: A→B v	---	---	---
$X_{A0}Y_{A1}$	X_A: 0 Y_A: 1→0 ii	---	---	X_A: 0→1 Y_A: 1 i	---	X_0: A→B Y_1: A→B v	---	---
$X_{A1}Y_{A0}$	X_A: 1→0 Y_A: 0 ii	---	---	X_A: 1 Y_A: 0→1 i	---	---	X_1: A→B Y_0: A→B v	---
$X_{A1}Y_{A1}$	---	X_A: 1→0 Y_A: 1 ii	X_A: 1 Y_A: 1→0 ii	---	---	---	---	X_1: A→B Y_1: A→B v
$X_{B0}Y_{B0}$	X_0: B→A Y_0: B→A v	---	---	---	---	X_B: 0 Y_B: 0→1 iii	X_B: 0→1 Y_B: 0 iii	---
$X_{B0}Y_{B1}$	---	X_0: B→A Y_1: B→A v	---	---	X_B: 0 Y_B: 1→0 iv	---	---	X_B: 0→1 Y_B: 1 iii
$X_{B1}Y_{B0}$	---	---	X_1: B→A Y_0: B→A v	---	X_B: 1→0 Y_B: 0 iv	---	---	X_B: 1 Y_B: 0→1 iii
$X_{B1}Y_{B1}$	---	---	---	X_1: B→A Y_1: B→A v	---	X_B: 1→0 Y_B: 1 iv	X_B: 1 Y_B: 1→0 iv	---

Figure 6. a) A simplified independent model. In this model, transitions from 0 to 1 all occur at the same rate and transitions from 1 to 0 all occur at the same rate. b) A simplified correlated model. Under this model, transitions between states of character X and Y depend on the background state of the other character. c) A simplified hidden Markov independent model, where the simple independent model of (a) is used in the hidden Markov framework which allows for rate heterogeneity independent of focal characters. The same can be done for the simple correlated model (not shown).

Returning to the modified Darwin’s scenario datasets, we found consistent and overwhelming support for the simplified hidden Markov independent model across all scenarios (Table 1). The average AIC weight of the simplified HMIM when fit to modified Darwin’s scenarios are $w_{outisde}$= 89.6%, w_{inside}= 90.2%, and w_{both}= 93.5%. The set of models applied to this data included all models discussed thus far as well as more complicated versions of those previously described (such as a standard correlated model with multiple rate classes). Additionally, to ensure that these models are not biased towards being favored across all datasets, we simulated data under a simplified correlated, simplified independent, and simplified hidden Markov independent models. We then fit each model to these datasets and found that the generating model is consistently chosen as the best fitting model (see Supplemental Materials). In summary, our findings suggest that when a complete model set is considered, the bias towards a correlation noted by Maddison and Fitzjohn (2015) disappears. The model which best describes data under a strict Darwin’s scenario is not one of correlation, but a simplified independent model with character independent rate heterogeneity.

Broadly applicable solutions

The issue discussed herein is recognized as being broadly applicable to several comparative methods that test for associations between variables (FitzJohn 2010; Rabosky and Goldberg 2015; Uyeda et al. 2018; Nakov et al. 2019; Gardner and Organ 2021). It is concerning that such a significant issue has seemingly gone unresolved for so long given comparative methods are of critical importance for understanding macroevolutionary patterns. However, in our view, the prevalence of the problems identified over the past few years is due to a singular overarching cause, namely, model misspecification, which occurs when a model, or set of models, is incomplete. Within the context of their model sets, authors of previous studies have correctly portrayed and analyzed the correlation bias of modeling dependence between discrete characters (Maddison and FitzJohn 2015; Uyeda et al. 2018; Gardner and Organ 2021). However, the danger of model misspecification is that the inferences drawn from an incomplete set are highly susceptible to unforeseen biases – a fact will hold true in both theoretical and empirical contexts. Here, we are arguing that the model set is incomplete without the inclusion of models that allow for rate heterogeneity that is independent of the focal characters. The canonical character independent model of Pagel (1994) has no way to account for multiple rates of evolution, whereas support for a correlated model can come from both evidence of correlation and evidence of rate heterogeneity. The additional support from explaining rate heterogeneity is not a feature exclusive to correlated characters, and thus accounting for independent rate heterogeneity is necessary to resolve the model set misspecification. This misspecified model set has led to consistently biased evidence towards correlation, and it is the same issue addressed by the inclusion of the character independent models within state-dependent speciation extinction models (Beaulieu and O'Meara 2016). In that case, the biased association was between diversification rates and phenotype (Rabosky and Goldberg 2015), but the cause is the same. Models in which there are no differences in diversification are compared to models which tested for the presence of a correlation between character and diversification rate (which necessarily allow for multiple rates of diversification).

One difference between the problem of false correlation in SSE models and the problems within simpler Markov models is the narrative surrounding them. In the case of SSE models, the problem was viewed as a high false positive rate (Rabosky and Goldberg 2015), whereas in the case of discrete character evolution we are led towards viewing rate heterogeneity through the lens of single unreplicated evolutionary events (Maddison and FitzJohn 2015). However, both points contribute to the same problem and if we view single evolutionary events as examples of where evolution has changed in tempo or mode, then the inclusion of hidden Markov models as a solution arises naturally from the problem.

Since we as comparative biologists are involved in a historical science, we will inevitably encounter single evolutionary events of large importance. However, it must be recognized that datasets which are susceptible to biases from singular events are not amenable to most phylogenetic comparative tests. Although here we have resolved the statistical biases associated with false correlations, there is no amount of methodological massaging that will allow for a satisfying test of macroevolutionary correlation between two synapomorphies. This is because comparative methods rely on several independent replicates of correlation such that the associations found between the variables may be considered robust even when extended beyond the dataset used for the analysis. If there is only one example of the correlation arising in the entire dataset, we should not have confidence in extending our inferences beyond the clade and should be wary of the correlation even within the focal clade. However, that is not to say there is no mechanistic reason for an association between synapomorphies. It is entirely possible that two characters which share identical evolutionary histories have an underlying biological link. Nonetheless, conclusions about the potential links between these characters cannot come from studies conducted at a macroevolutionary scale, and they should instead be investigated at a smaller scale (Beaulieu and O'Meara 2018, 2019; Donoghue and Edwards 2019). Additional lines of evidence and a more mechanistic explanation will be necessary in order for a conclusion of correlation to be satisfying (Gardner and Organ 2021). In a sense, the hidden rate classes of our proposed framework may represent lineage-specific factors that, once present, readily allow for a shift in the tempo and mode of a lineage's evolution (Maddison and FitzJohn 2015; Ogburn and Edwards 2015).

A broader methodological conclusion that can be drawn from our results, which have been echoed elsewhere (Beaulieu and O’Meara 2016; Caetano et al. 2018; O’Meara and Beaulieu 2021), is that testing against simple null hypotheses is usually not a productive way to do science. Rather than testing for a binary outcome of whether or not correlation is present, it is often beneficial to examine what these models suggest about the evolutionary process. Utilizing model comparison and finding that correlation exists is certainly interesting, but the real utility of modeling macroevolutionary processes is interpreting parameters that could not have been identified from the pattern alone. Within reason, it is often possible to look at the distribution of two discrete characters and be able to say whether the two are correlated before doing any modeling. However, it is more difficult to specify numerical values for the rates at which these characters evolve. For example, neither a glance at the dataset nor summary statistics will be consistently informative as to how many orders of magnitude faster a lineage in state Y_0 evolves character X than a lineage in state Y_1. Additionally, transition rates which are measured in changes per million years (more specifically, changes per time unit of the phylogenetic tree) are directly comparable across any comparative study. For instance, changes in flower color in one study can be compared directly to changes in mammalian diet in another, because the parameters of transition have the same unit (event per unit of time). With these parameter estimates we may more robustly test hypotheses based on a well-defined model of macroevolution (Pennell and Harmon 2013). Furthermore, an examination of parameter estimates applies to most commonly used macroevolutionary models. For example, Vasconcelos et al. (2021) tested a set of three hypotheses related to how the mode of seed dispersal related to climatic niche evolution using

Ornstein-Uhlenbeck models. This is not new for these types of models, but a key point from this study is that the model support was not as important as the relative value of the parameters. Instead of examining whether model A was more supported than model B, they looked at how specific hypotheses (i.e., that abiotically dispersed seeds tend to have a more arid climatic optima) were differentially supported across a diverse set of models. A focus on parameter estimates rather than relative model support underscores that we are uncertain about the best model, but we wish to estimate parameters which reflect that uncertainty and robustly relate them to our hypotheses. This insight spurred the adoption of model-averaging by comparative biologists, which is now recognized as vital for macroevolutionary studies (see Caetano et al. 2018).

Concluding Remarks

Sparked by an appreciation of the limitations of PCMs, several commonly used phylogenetic comparative methods have seen critical challenges recently, which have led to advancements useful for both developers and users (Boettiger et al. 2012; Maddison and FitzJohn 2015; Rabosky and Goldberg 2015; Louca and Pennell 2020). Here, too, the critiques of classic tests of correlation (Pagel 1994) are not wrong, and the recommendations of past studies remain useful (Maddison and FitzJohn 2015; Uyeda et al. 2018; Gardner and Organ 2021). Instead, what we have demonstrated is that the statistical bias towards correlation is primarily due to a misspecification of the model set and a failure to account for character independent rate heterogeneity. We have highlighted that the inclusion of non-standard Markov models in the model set can be critical for the quality of the inferences being made. We acknowledge that choosing a diverse set of models *a priori* is not always straightforward, but both likelihood and

Bayesian methods will only be as effective as the plausibility of the models set being analyzed (Burnham and Anderson 2002). We know that a homogeneous process over millions of years and across thousands of lineages is incorrect (Eldredge and Gould 1972) and that the individual parts of an organism do not evolve independently (Levins and Lewontin 1985). While we may not be able to always specify each of these individual processes, we must try to incorporate them in our modeling. Accounting for rate heterogeneity through HMMs is a simplified way that we can bring realism to our modeling while also making statistically consistent and unbiased estimates of evolutionary parameters. From there, undoubtedly more work will be necessary (e.g., Goldberg and Foo 2020). But comparative analyses must at the very least attempt to account for what we know about macroevolution while making us aware of the wonderful idiosyncrasies of evolutionary history.

Data Availability

Data and a user guide are available from the following github repository: https://github.com/jboyko/2022_unsolved-challenge

Acknowledgements

We thank Brian O’Meara for his comments and discussion of the ideas presented here. We would also like to thank Thais Vasconcelos and Eric Hagen for their helpful comments and edits of this work at various stages.

Tables

Table 1. Average ΔAIC values for 100 datasets with standard deviations shown in brackets. Each column represents a scenario described in the main text and each row represents a different Markov model structure which may be consistent with independence or correlation. For each scenario, 8 or 9 models were fit to the datasets. The collapsed model is fit only when not all potential state combinations are directly observed and therefore are not fit in modified scenarios. A ΔAIC of 0 indicates the best model and models within 2 AIC units of each other are generally considered good fits to the data (Burnham and Anderson 2002).

Scenario	*Darwin's*	*Unreplicated bursts*	*Modified Darwin's (outside)*	*Modified Darwin's (inside)*	*Modified Darwin's (both)*
Collapsed	**0.0 (± 0.0)**	**0.0 (± 0.0)**	NA	NA	NA
Independent	17.9 (± 12.3)	36.8 (±9.0)	14.3 (±3.6)	14.8 (±4.0)	15.6 (±4.7)
Simplified independent	13.9 (±2.3)	67.3 (±15.8)	10.3 (±3.6)	10.8 (±4.0)	11.6 (±4.7)
Correlated	12.0 (±0.2)	8.0 (±0.1)	6.1 (±0.5)	6.1 (±0.7)	10.8 (±2.6)
Simplified correlated	13.9 (±2.3)	30.0 (±8.2)	9.8 (±3.6)	10.4 (±4.1)	11.6 (±4.7)
Hidden Markov independent	20.8 (±6.8)	9.2 (±0.4)	7.9 (±1.2)	8.4 (±3.3)	7.8 (±2.2)
Simplified hidden Markov independent	5.5 (±0.1)	36.3 (±9.1)	**0.0 (±0.0)**	**0.0 (±0.0)**	**0.0 (±0.0)**
Correlated hidden Markov	29.7 (±0.3)	24.9 (±0.8)	22.9 (±0.7)	23.5 (±0.8)	23.2 (±1.4)
Simplified correlated hidden Markov	18.8 (±2.1)	34.3 (±7.7)	14.2 (±3.3)	14.3 (±2.8)	15.7 (±3.5)

Figures

Figure 1. The two problematic scenarios from Maddison and FitzJohn (2015) for the evolution of characters *X* and *Y*. Character *X* is painted on the left phylogeny using red and orange for state X_0 and X_1, whereas character *Y* is painted on the right phylogeny using dark blue and light blue for state Y_0 and Y_1. a) Darwin's scenario is depicted as a single event deep in time that has led to the co-distribution of X_0Y_0 outside of the focal clade and X_1Y_1 within the focal clade. b) Unreplicated bursts scenario is where a single event deep in time has led to the co-distribution of X_0Y_0 outside of the focal clade and X_1Y_0 and X_1Y_1 within the focal clade.

Figure 2. Representations of the different transition rate matrices, **Q**, with *k* number of parameters associated with each. Where transitions are fixed to occur at the same rate, the squares are colored to be the same. Unique parameters are also indicated with a roman numeral in the bottom left corner of the square. To the right of each matrix, a ball and stick representation of the model is presented with colors and parameter numbers matching the transitions indicated in the matrix, **Q**. The ball and stick representation is organized such that internal arrows represent transitions from 1 to 0, and external arrows represent transitions from 0 to 1. Additionally, arrows which cross the vertical midpoint indicate transitions in character X, whereas transitions across the horizontal midpoint indicate transitions in character Y. a) An independent model with four unique parameters, which fixes transitions within a character such that changes in X or Y do not depend on the state of the other character. b) A dependent model with eight unique parameters, whichs model allows transitions within a character to depend on the state of the other character. c) A model which removes transitions to and from an unobserved state from the independent model (a). d) A model that removes transitions to and from an unobserved state from the dependent model (b). In (c) and (d) the unobserved state is based on the Unreplicated Burst scenario where X_0Y_1 is not observed.

Figure 3. An example likelihood surface of a correlated model when applied to one of the 100 Darwin's scenario datasets. The color of the plot indicates the likelihood of a particular pair of parameters when the remaining transition rates are optimized. Thus, each point represents the maximum likelihood estimate when the transition rates indicated by the axes are fixed. a) Transitions from X_0Y_0 to an intermediate state result in several likelihood ridges. b) Transitions

from X_1Y_1 to an intermediate state result in several likelihood ridges. c) Transitions from X_0Y_1 to either X_0Y_0 or X_1Y_1 result in a completely flat likelihood surface. d) Transitions from X_1Y_0 to either X_0Y_0 or X_1Y_1 result in a completely flat likelihood surface.

Figure 4. The hidden Markov independent model (HMIM), which allows transitions within a character to have rate heterogeneity without it necessarily being linked to an observed character. This matrix can be read as a block matrix, with 4x4 blocks representing transitions between observed characters following an independent model (top left and bottom right) and transitions between hidden rate classes *A* and *B* (top right and bottom left). The independent model is essentially duplicated in the top left (blue and green) and bottom right (red and orange) of the block matrix with transitions occurring between these different types of independent models (purple). Here, transition rates between the hidden states are fixed to be the same (parameter ix), but it is straightforward to allow the transition between rate class *A* and *B* to differ.

Figure 5. The amount of evidence for correlation when comparing a correlated model to ether an independent model (IM) or hidden Markov independent model (HMIM). The models are fit to data of the modified version of Darwin's scenario where a single observation of X_0Y_1 and X_1Y_0 is added outside of the focal clade (a), inside of the focal clade (b), and both within and outside of the focal clade (c). Evidence ratios for each model comparison are plotted as boxplots to the left of the simulation scenario. In all cases, the evidence ratio of the correlated model over the independent model is substantially greater than 2.7 (left boxplot) but, the correlated model receives much less support over the hidden Markov independent model (right boxplot).

Figure 6. a) A simplified independent model. In this model, transitions from 0 to 1 all occur at the same rate and transitions from 1 to 0 all occur at the same rate. b) A simplified correlated model. Under this model, transitions between states of character *X* and *Y* depend on the background state of the other character. c) A simplified hidden Markov independent model, where the simple independent model of (a) is used in the hidden Markov framework which allows for rate heterogeneity independent of focal characters. The same can be done for the simple correlated model (not shown).

References

Beaulieu J.M., Donoghue M.J. 2013. Fruit Evolution and Diversification in Campanulid Angiosperms. Evolution. 67:3132–3144.

Beaulieu J.M., O'Meara B.C. 2016. Detecting Hidden Diversification Shifts in Models of Trait-Dependent Speciation and Extinction. Syst Biol. 65:583–601.

Beaulieu J.M., O'Meara B.C. 2018. Can we build it? Yes we can, but should we use it? Assessing the quality and value of a very large phylogeny of campanulid angiosperms. Am. J. Bot. 105:417–432.

Beaulieu J.M., O'Meara B.C. 2019. Diversity and skepticism are vital for comparative biology: a response to Donoghue and Edwards (2019). Am. J. Bot. 106:613–617.

Beaulieu J.M., O'Meara B.C., Donoghue M.J. 2013. Identifying Hidden Rate Changes in the Evolution of a Binary Morphological Character: The Evolution of Plant Habit in Campanulid Angiosperms. Syst Biol. 62:725–737.

Boettiger C., Coop G., Ralph P. 2012. Is Your Phylogeny Informative? Measuring the Power of Comparative Methods. Evolution. 66:2240–2251.

Boyko J.D., Beaulieu J.M. 2021. Generalized hidden Markov models for phylogenetic comparative datasets. Methods Ecol Evol. 12:468–478.

Burnham K.P., Anderson D.R. 2002. Model selection and multimodel inference: a practical information-theoretic approach. New York: Springer.

Caetano D.S., O'Meara B.C., Beaulieu J.M. 2018. Hidden state models improve state-dependent diversification approaches, including biogeographical models: HMM and the adequacy of SSE models. Evolution. 72:2308–2324.

Darwin C. 1859. On the Origin of Species, 1859. Routledge.

Donoghue M.J., Edwards E.J. 2019. Model clades are vital for comparative biology, and ascertainment bias is not a problem in practice: a response to Beaulieu and O'Meara (2018). Am. J. Bot. 106:327–330.

Eldredge N., Gould S.J. 1972. Punctuated equilibria: an alternative to phyletic gradualism. Models in paleobiology. 1972:82–115.

Felsenstein J. 1985. Phylogenies and the Comparative Method. Am. Nat. 125:1–15.

Felsenstein J., Churchill G.A. 1996. A Hidden Markov Model approach to variation among sites in rate of evolution. Mol. Biol. Evol. 13:93–104.

FitzJohn R.G. 2010. Quantitative Traits and Diversification. Syst Biol. 59:619–633.

Gardner J.D., Organ C.L. 2021. Evolutionary Sample Size and Consilience in Phylogenetic Comparative Analysis. Syst Biol. 70:1061–1075.

Goldberg E.E., Foo J. 2020. Memory in trait macroevolution. Am. Nat. 195:300–314.

Levins R., Lewontin R. 1985. The dialectical biologist. Harvard University Press.

Louca S., Pennell M.W. 2020. Extant timetrees are consistent with a myriad of diversification histories. Nature. 580:502–505.

Maddison W.P., FitzJohn R.G. 2015. The Unsolved Challenge to Phylogenetic Correlation Tests for Categorical Characters. Syst Biol. 64:127–136.

Marazzi B., Ané C., Simon M.F., Delgado-Salinas A., Luckow M., Sanderson M.J. 2012. Locating Evolutionary Precursors on a Phylogenetic Tree. Evolution. 66:3918–3930.

Nakov T., Beaulieu J.M., Alverson A.J. 2019. Diatoms diversify and turn over faster in freshwater than marine environments. Evolution. 73:2497–2511.

Ogburn M.R., Edwards E.J. 2015. Life history lability underlies rapid climate niche evolution in the angiosperm clade Montiaceae. Mol. Phylogenet. Evol. 92:181–192.

O'Meara B., Beaulieu J. 2021. Potential survival of some, but not all, diversification methods. Available from ecoevorxiv.org/w5nvd.

Pagel M. 1994. Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. Proc. R. Soc. B: Biol. Sci. 255:37–45.

Pagel M., Meade A. 2006. Bayesian Analysis of Correlated Evolution of Discrete Characters by Reversible-Jump Markov Chain Monte Carlo. Am. Nat. 167:808–825.

Pennell M.W., Harmon L.J. 2013. An integrative view of phylogenetic comparative methods: connections to population genetics, community ecology, and paleobiology. Annals of the New York Academy of Sciences. 1289:90–105.

Rabosky D.L., Goldberg E.E. 2015. Model Inadequacy and Mistaken Inferences of Trait-Dependent Speciation. Syst Biol. 64:340–355.

Uyeda J.C., Zenil-Ferguson R., Pennell M.W. 2018. Rethinking phylogenetic comparative methods. Syst Biol. 67:1091–1109.

Vasconcelos T., Boyko J.D., Beaulieu J.M. 2021. Linking mode of seed dispersal and climatic niche evolution in flowering plants. J. Biogeogr. In press.

Supplemental Figures

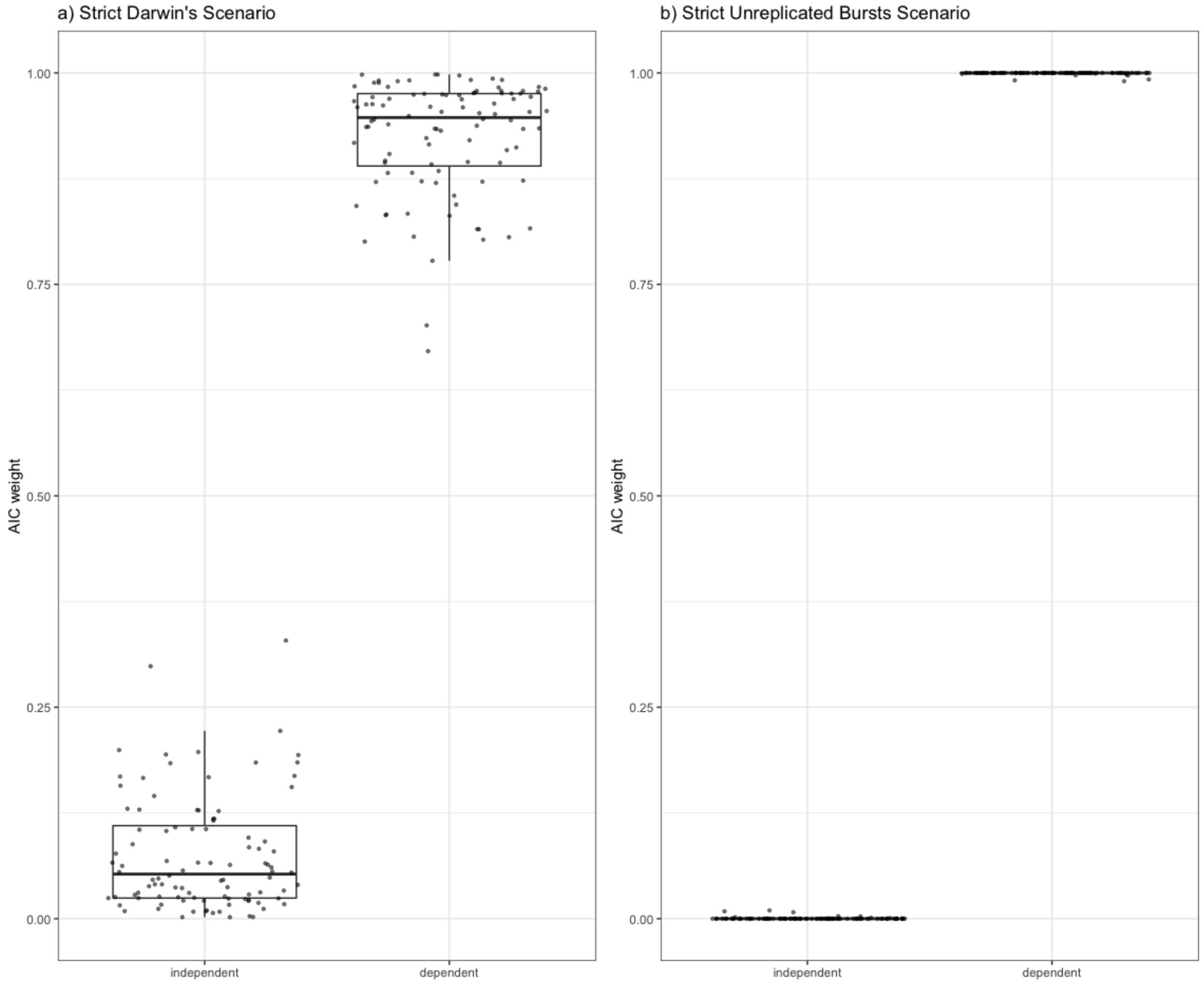


Figure S1. Replicated the Maddison and Fitzjohn (2015) result with our simulation and model fitting framework. Support for a dependent/ correlated model is consistently greater than an independent model.

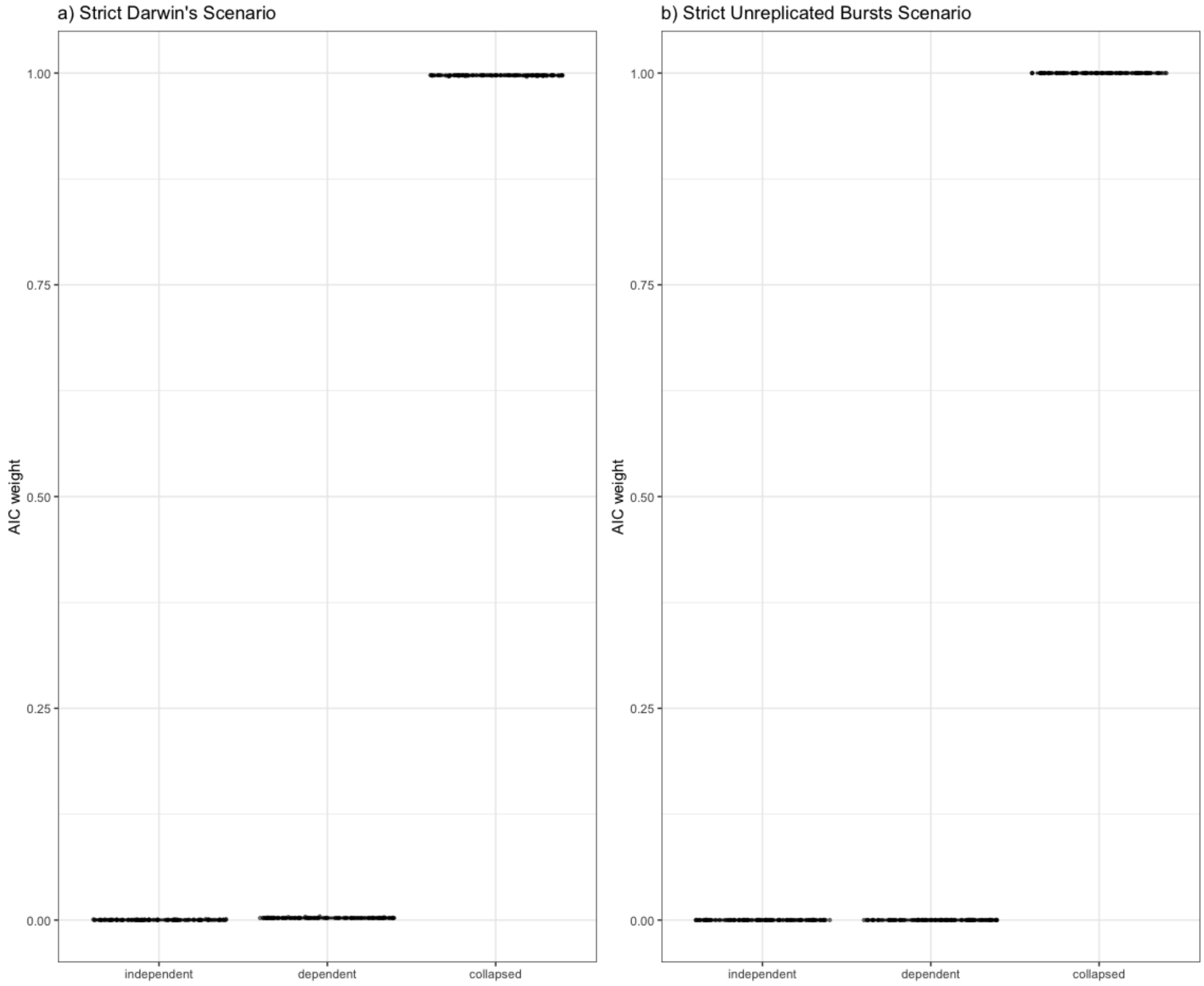


Figure S2. The same model set used by Maddison and Fitzjohn (2015), but with the inclusion of a collapsed model. SUpport for the collapsed model is overwhelming.

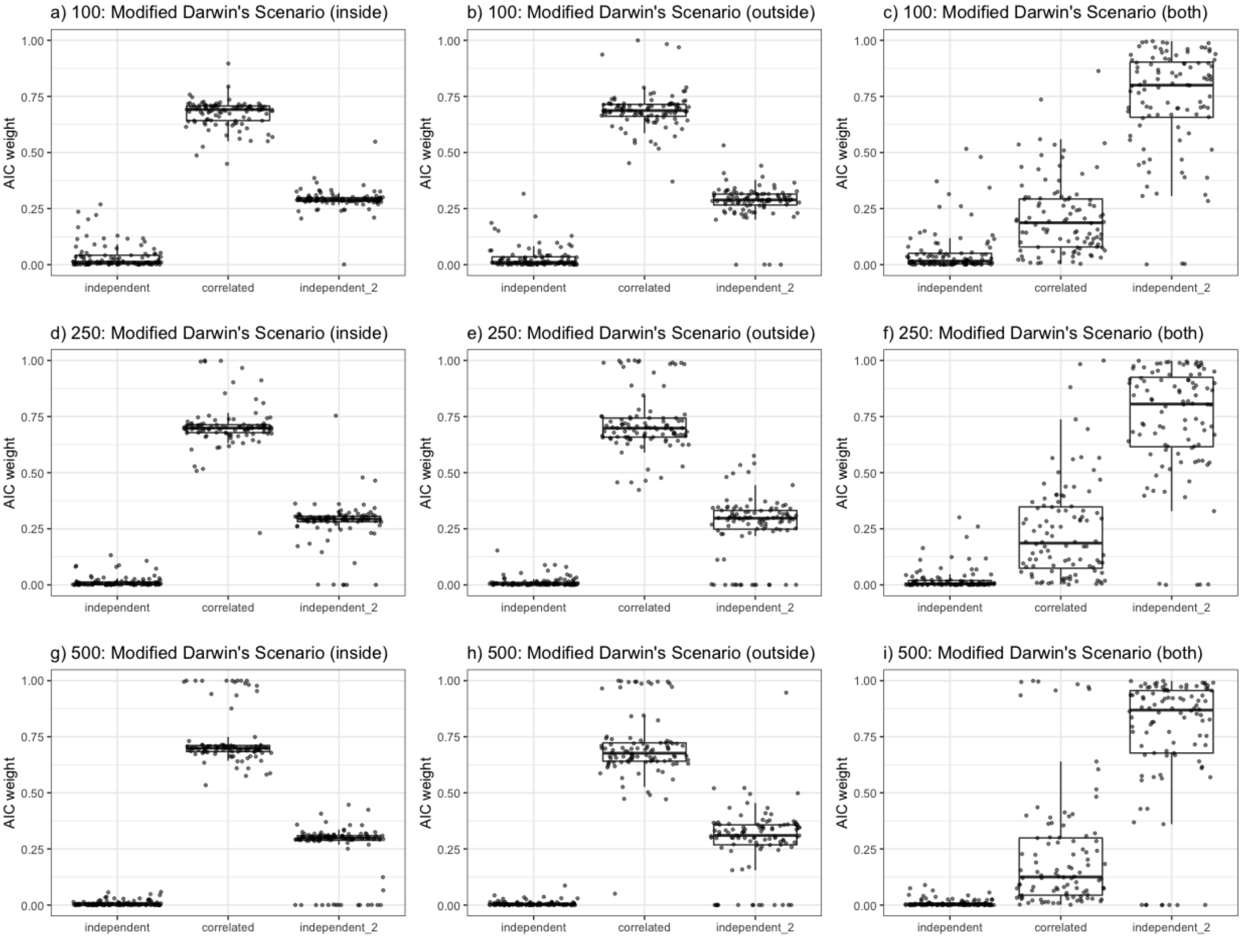


Figure S3. The effect of increasing the number of taxa on model support. Shown here are the two standard Pagel (1994) models (independent and correlated) as well as the unsimplified hidden state independent model. Support for the models is consistent across 100, 250, 500 taxa.

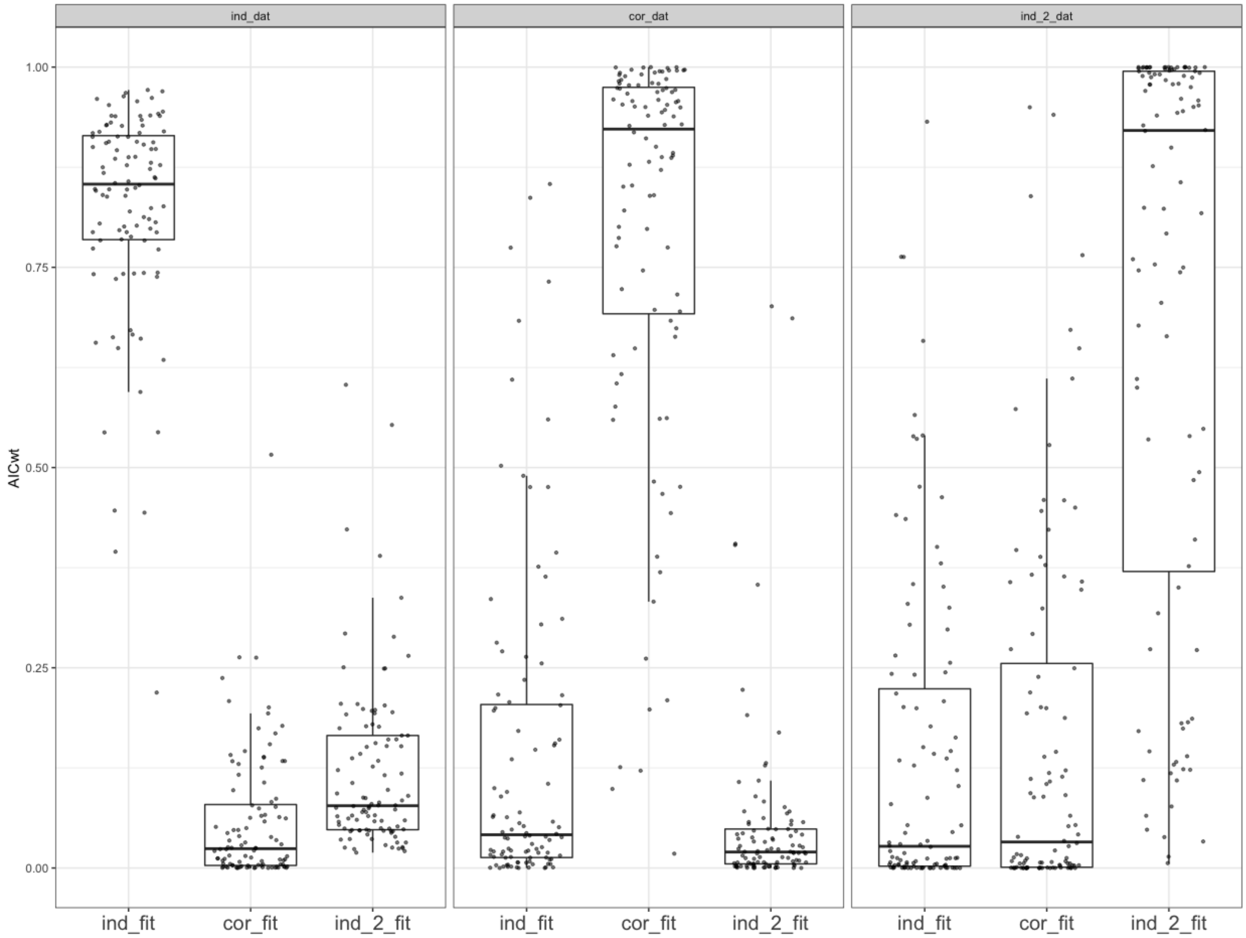


Figure S4. Akaike model weights are shown for data simulated under a simplified independent model (ind_dat), simplified correlated model (cor_dat) and simplified hidden Markov independent model (ind_2) for 100 unique datasets (See Figure 6 for model structure). For the simple independent and dependent models, the rates of evolution were 1 and 5 changes per million years. With the addition of the hidden states, we added rates of 2 and 10 for the second rate category as well as a transition rate of 4 between rate classes. Phylogenetic trees of 100 taxa were simulated with a birth rate of 1 and death rate of 0.75. Total branching time in the tree was rescaled to a total of 10 MY.