Assessing the Impact of Continuous Traits on the Evolution of Discrete Traits: The Ancestral Condition Test.

Running title: AncCond

Keywords: comparative method, phylogenetics, trait correlation

Nathan Anderson1, Richard H. Adams2, Jeffery P. Demuth3, and Heath Blackmon4,5

1 Department of Integrative Biology, University of Wisconsin – Madison, Madison, WI, USA

2 Department of Computer and Electrical Engineering and Computer Science, Florida Atlantic University, FL, USA

3 Department of Biology, University of Texas at Arlington, TX, USA

4 Department of Biology, Texas A&M, College Station, TX USA

5 coleoguy@gmail.com

**ABSTRACT**

1. Analyses of the co-evolution of multiple traits has the potential to reveal the drivers and limits to biological evolution. A variety of methods are available to study the interaction between either two continuous traits or a discrete trait that impacts the evolution of a continuous trait. However, few methods are available to study the impact of a continuous trait on the evolution of a discrete trait.
2. Here we present the ancestral condition test, a new comparative method that evaluates whether a binary trait tends to transition when a continuous trait has values more extreme than expected if both traits were evolving independently. This approach leverages ancestral state estimates of both the continuous and the binary trait to test whether extreme values of the continuous trait are associated with transitions in the binary trait, and to assess statistical significance.
3. We explore the robustness of our approach under a range of parameter values and patterns of trait evolution. We find that either a relatively strong contingency between the two traits or a large number of taxa is required to detect the underlying relationships reliably. Statistical power of the test is highest when the binary trait evolves unidirectionally, and we find that the false-positive rate remains acceptable for a bidirectionally evolving binary trait. In comparison to existing methods that might be employed, we show that the ancestral condition test has both higher power and a lower false-positive rate.
4. The types of questions that this approach allows us to test are common in evolutionary biology and, unlike existing methods, the ancestral condition test incorporates the temporal order of transitions – moving a step closer to inferring causality rather than merely identifying correlation. An implementation of this test is distributed in the r package evobiR.

**INTRODUCTION**

One of the central goals of evolutionary biology is to understand the evolution of traits among species. With the advent of statistically robust methods of phylogenetic tree inference, we can now fit probabilistic models to the evolution of both binary traits (Lewis 2001; Felsenstein 2012) and continuous traits (Felsenstein 1985; Butler & King 2004; O'Meara *et al.* 2006; Landis, Schraiber & Liang 2012). A variety of methods that test for correlations between multiple continuous traits have been developed (reviewed in Felsenstein 2004). Robust methods are also available to study how a binary character impacts the evolution of a continuous trait (O'Meara *et al.* 2006). However, available methods to identify and analyze a co-evolutionary relationship where a continuous trait impacts the evolution of a binary trait are less well developed (but see Hadfield 2010; Felsenstein 2012; Revell 2014).

Common hypotheses suggest that extreme values of a continuous trait are associated with transitions in a binary trait(Ross *et al.* 2012; Blackmon & Demuth 2015; Blackmon, Hardy & Ross 2015). In a recent study of sex chromosome evolution, for example, we were tested whether loss of the Y chromosome (a binary trait) occurred in taxa with a significantly higher total number of chromosomes (a continuous trait) than would be expected if the two were unrelated (Blackmon & Demuth 2015). Investigators could employ several existing approaches in such a situation. First, by discretizing chromosome number into a low and high category, we could apply Pagel’s test for the correlation of two binary characters (Pagel 1994). This approach is problematic, however, as defining threshold values for high and low classes involves arbitrary decisions that may not be biologically realistic. Furthermore, if several threshold values are tested, p-values must be corrected for multiple comparisons. Additionally, by discretizing a continuous variable, we effectively reduce the amount of information being applied to the problem and decrease our power to detect a significant relationship (Uusitalo 2007). Another approach would be to assume that our binary state (presence or absence of a Y chromosome) could be modeled as if it was produced by an underlying continuous trait that is unobserved. If this assumption is justified, we could employ the threshold model (Wright 1934; Felsenstein 2012), which would allow estimation of the correlation between chromosome number and the presence or absence of the Y chromosome. A third potential approach would be to use a phylogenetic mixed model that treats chromosome number as a predictor variable and presence or absence of the Y chromosome as a response variable (Hadfield 2010). However, each of these approaches suffer from the same fundamental problem: correlation does not informthe direction of causality. For example, does high chromosome number lead to Y chromosome loss, or does Y chromosome loss lead to high chromosome number (Blackmon & Demuth 2015)? Furthermore, many of these current approaches suffer from excessively high false-positive rates (Maddison & FitzJohn 2014).

In light of these issues, we have developed a flexible, conservative statistical test to infer significant relationships between the value of a continuous trait and transitions in a binary trait. Briefly, our approach estimates the mean value of the continuous trait when the binary character undergoes transitions. We then evaluate the significance of this mean value by comparison to a null distribution that is generated by simulating mean ancestral states of the continuous trait under the hypothesis that there is no relationship. We evaluate the ancestral condition test using both simulated and empirical datasets. We find that our method is conservative, with the highest power when the number of taxa and/or the degree of contingency between traits is large. In comparison to existing methods (Pagel’s and Threshold), the test we develop offers higher power and lower false-positive rates. In conjunction with improved performance, our approach also grants insight into the temporal order of changes and allows us to get closer to the possible causal relationships between traits. Our ancestral condition test represents a novel addition to the phylogenetic comparative method toolkit that is useful for studying the interplay between the state of a continuous trait and the evolution of a binary character.

Furthermore, the approach we have developed offers a large degree of flexibility that is missing in many existing methods. For instance, in many cases we may know that the derived state of the binary character is associated with lower values of the continuous trait, but is this because low values of the continuous trait lead to changes in the binary character, or does the derived state of the binary character lead to reductions in the continuous trait? When this is a concern, we can restrict our estimate of the continuous trait to only use data from species that have retained the ancestral condition of the binary trait. This approach provides a much more rigorous test than can be implemented with existing methods. In practice, there is not always a clear ancestral and derived state, particularly when a binary character evolves quickly and lacks fossil data. This uncertainty does not pose a problem for the ancestral condition test, but in such cases, the power of our approach is reduced.

**METHODS**

*The Ancestral Condition Test*

Our approach determines whether there is evidence that nodes in the phylogenetic tree subtending transitions of a binary character tend to exhibit extreme values of the continuous character under study. We refer to this method as the Ancestral Condition (AncCond) test, which involves the following four steps:

1) Estimate the ancestral values of the continuous character while optionally pruning data from species that exhibit the derived state of the binary character. By pruning data from taxa exhibiting the derived state of the binary character, we create a more conservative test that will be applicable in cases where the derived state of the binary character is expected to lead to changes in the selective forces acting on the continuous character (addressing the potential for feedback mentioned above). Ancestral state estimation is accomplished under a Brownian motion model using the existing function *anc.ML* in phytools (Figure 1A) (Revell 2012).

2) Ancestral state estimation of the binary character by way of stochastic character mapping (Figure 1B) (Huelsenbeck, Nielsen & Bollback 2003; Bollback 2006). A user specified number of stochastic mappings (default = 100) to incorporate uncertainty in the ancestral state estimations. We may choose to fix a single state at the root of the tree in cases where this is appropriate. For instance, when the ancestral state is informed by fossil records or when prior biological knowledge suggests that transitions are expected to occur in only one direction (Blackmon, Hardy & Ross 2015), users may also restrict the simulated evolution of the binary character by allowing only transitions from the ancestral to the derived state. Stochastic mapping is performed using the *make.simmap* function provided in phytools (Figure 1B) (Revell 2012). For simplicity, we will refer to the two states as one and two; state one will denote the ancestral state, and state two is the derived state.

3) Classify nodes. The stochastic maps are then processed to classify all nodes in the tree into one of three groups: A) Nodes exhibiting state one (i.e., ancestral state) of the binary character and no transitions in the binary character in immediate daughter branches (Figure 1C, blue nodes), B) Nodes exhibiting state two (i.e., derived state) of the binary character and no transitions in the binary character in immediate daughter branches (Figure 1C, green nodes), or C) Producing nodes exhibiting either state of the binary character but with one of the immediate daughter branches showing a transition into the other state of the binary character (Figure 1C, red nodes). In the bidirectional case, we additionally parse producing nodes to select those that subtend branches with transitions from state one to state two or from state two to state one. In the unidirectional case, this parsing is not necessary. Next, we calculate the mean value of the continuous trait at the producing nodes in each group for each stochastic mapping. producing nodes (type C for a stochastic mapping This process is repeated across stochastic mappings incorporating ancestral state estimation uncertainty. Finally, the mean of the observed statistics is calculated. This value is referred to as the “estimated ancestral condition” for each of the states of the binary character (Figure 1C). In the case where transitions are restricted to a single direction, there will be a single estimated ancestral condition for transitions from the ancestral to the derived state. In the more general case, there will be two estimated ancestral conditions for transitions from state one to state two and for transitions from state two to state one.

4)

We construct a null distribution by simulating a user specified number (default = 100, meaning a default of 100 rate matrices each with 100 simulations totaling 10,000 simulations) of evolutionary histories of the discrete trait under each rate matrix inferred in step 2 using the *sim.history* function in phytools (Revell 2012). We then process each of the simulated null histories in the manner described in step 3, yielding a null observed statistic for each of these histories. The null ancestral conditions are calculated by taking the mean of one of the **ordered** observed statistics for each of the rate matrices. --- I.e. we are averaging over the number of different rate matrices, not over the number of evolutionary histories simulated for each rate matrix. Furthermore, we are averaging the ordered observed statistic calculated for each rate matrix; the smallest observed statistic for each rate matrix are averaged, the second smallest observed statistic for each rate matrix are averaged etc. This ordering preserves the variance in the observed statistics by ensuring that, for example, the largest and smallest observed statistics are not averaged together. If this were allowed to happen, the null ancestral conditions would regress towards the mean of the null observed statistics, effectively narrowing our null distribution. --- These null ancestral conditions define a null distribution, and we compare our test statistic (the estimated ancestral condition) to this distribution to obtain an empirical p-value (Figure 1D).

***Simulated data***

*Scenario one–unidirectional change in a binary character*

We simulated 100 trees using a birth-death model with a birth rate of 3.0 and death rate of 1.0, allowing trees to grow until 200 extant species were reached. Phylogenies were rescaled to unit length. Trees were simulated using the function *trees* in Diversitree (FitzJohn 2012). Next, we simulated a continuous character evolving via Brownian motion with a rate parameter of 0.2 and a starting mean of 0 on each tree using the function *sim.char* in Geiger (Harmon *et al.* 2008). We then used the function *anc.ML* in Phytools (Revell 2012) to estimate the mean value of the continuous trait for each node in the phylogeny.

To simulate a binary trait whose evolution was contingent on the continuous trait, we used a branch scaling approach. We first identified those branches that had a mean value of the continuous trait in the upper or lower quartiles. Branches whose mean was in the lower quartile were scaled by a multiplicative factor of 1/*x*, while branches in the upper quartile were scaled by a multiplicative factor of *x*. We repeated this process for five values (–between 1 and 10) on each of our 100 trees. We then simulated the evolution of the binary character on these scaled trees. We set the root to state one (ancestral state) and allowed the trait to evolve under an Mk2 model where transitions to state two (derived state) were allowed, but not transitions from state two to state one (i.e., modeling unidirectional evolution), as we expected our approach to be most powerful under this mode of evolution. This simulation was done using the *sim.char* function in Geiger (Harmon *et al.* 2008). Various transition rates from state one to state two were evaluated, and we found that a rate of 0.2 was sufficient to ensure that multiple transitions occurred in all simulated datasets. If a dataset did not contain at least ten extant taxa exhibiting each state of the binary character, it was discarded (between 7% and 55% of simulations were discarded for a given scaling factor, with a higher percent being discarded for smaller scaling factors). Thus, our total of 500 simulated datasets consisted of five degrees of contingency each applied to 100 trees.

By simulating the evolution of the binary trait on these scaled trees, we generated a contingency between the binary and continuous traits. There exists more opportunity for transitions on the branches where the continuous trait value is in the upper quartile and the branch length was increased. Alternatively, there exists less opportunity for transitions on the branches where the continuous trait value is in the lower quartile and the branch length was decreased. This branch scaling technique effectively altered the rate of evolution of the binary trait wherever the continuous trait took extreme values, increasing the density of transitions on branches where the continuous trait had high values and decreasing the density of transitions for branches with low values.

To assess the relationship between the number of taxa included in an analysis and the performance of our approach, we simulated birth-death trees and continuous characters using the methods described above but with tree sizes that ranged from 20 to 200 taxa. If a dataset did not include at least five extant taxa in each state of the binary character, it was discarded, and a new dataset was generated (between 14% and 91% of simulations were discarded for different sized trees, with small trees leading to a higher proportion of discarded simulations). We then simulated binary datasets with scaling factors *1 or 5* as described above, allowing us to measure both power (scaling factor = 5) and false-positive rate (scaling factor = 1; i.e. how often we detect a relationship under conditions where the data were simulated with no contingency between traits). A total of 200 data sets (scaling factors of one and five) were generated for each of the five tree sizes for a total of 1000 simulated datasets.

*Scenario two – bidirectional change in a binary character*

We also used simulated data sets to evaluate the performance of the AncCond test in scenarios for which a binary character has transitions in both directions (i.e., bidirectional), and when we are unable to confidently assign an ancestral - derived relationship. We repeated the above procedure with two key differences. First, we did not fix the root of the tree to any one state. Instead, we allowed the root to take either state with equal probability. Second, we allowed for transitions in either direction, from state one to state two and from state two to state one, with equal rates. A transition rate of 0.6 was used to ensure there were several forward and reverse transitions in every dataset. This process generated an additional 1500 simulated datasets.

*Scenario three – empirical phylogeny*

To test whether our approach is impacted when more complex processes generate trees and when continuous traits have variable rates of evolution, we used a phylogeny of living cetaceans (Steeman *et al.* 2009) and body size data for 74 species (Slater *et al.* 2010). We then simulated the unidirectional evolution of a binary character, from a known ancestral state, whose evolution is entirely independent of the continuous character using the sim.char function from Geiger (Harmon *et al.* 2008). This process generated 100 simulated datsets.

# *Scenario four –* Ornstein–Uhlenbeck model

We also sought to test the performance of our new method in determining the direction of causality in any correlation between the discrete and continuous trait evolution. To this end, we also developed a scenario where the evolution of the continuous trait is altered by the discrete trait. This contrasts to the above analyses where the evolutionary history of the discrete trait was affected by extreme values of the continuous trait. We first simulated the evolutionary history of the discrete trait on an unaltered tree using the same method as scenario 1 (??????). We then used the *OUwie.sim* function provided in OUwie (cite) to simulate the evolution of the continuous trait using a two optima OU model where the optima changes with the state of the discrete trait. In our simulations, the two optima (thetas) were 1 and 9, the adaptation parameter (alpha) was 1, the drift parameter (sigma^2) was 0.75 and the starting phenotype was 1, the optimum. On each simulated dataset, we performed three analyses using AncCond. First we pruned the continuous trait data on tips with the derived state of the discrete trait, this pruning will remove the “feedback” signal mentioned above. Second, we pruned the continuous trait data on branches exhibiting the ancestral state. We do this to explore the performance of AncCond whenever the data pruning is incorrectly applied. Third we apply AncCond using all available data, with no pruning.

*Assessing the performance of comparative tests*

We analyzed these datasets using our AncCond function implemented in the R package EvobiR (Blackmon H 2015). We supplied this R function with the unaltered phylogenetic trees and the corresponding tip data of the binary and continuous traits. For the case of unidirectional evolution (i.e., scenario one), we set the rate of evolution from state two to state one to zero and fixed the root of the tree to state one. We also chose to prune continuous data from taxa exhibiting the derived state of the binary character. This pruning is not entirely applicable to this simulated data, as the continuous character was simulated completely independent of the binary character. However, we chose to use this option in our analysis because, in practice, it is not always clear when the continuous character’s evolution is independent of the binary character’s evolution, and because it is important to test our method under the most restrictive conditions. When assessing data generated under scenario two, we followed the same approach but allowed two separate rates of transitions in the binary trait (transitions from one to two and two to one), made no assumptions on the original state at the root of the tree, and pruned no continuous data. In assessing the data from scenario three, we followed the same procedure as scenario one.

We compared the ancestral condition test to two existing approaches. For these comparisons, we used a subset of data simulated under scenario one (trees with 200 taxa and scaling factors of one and five). First, we applied the threshold model as implemented in phytools (Revell 2012). This approach allowed us to test for a correlation between our continuous character and the binary character by inferring an underlying continuous value that we observe as a discretely expressed character. Second, we discretized our continuous character at the median value and applied Pagel’s test for correlation of binary characters also implemented in phytools (Revell 2012).

Below we show the performance of our approach with the simulated data, reporting both false-positive rate and power. Tests were considered statistically significant at a p-value ≤ 0.05 for data analyzed under scenario one and three but at a p-value of ≤ 0.025 for scenario two. The reduction in p-value accounts for the fact that we are testing two hypotheses 1) transitions from state one to two are associated with extreme values of the continuous trait, and 2) that transitions from two to one are associated with extreme values of the continuous trait. All analyses were completed with RStudio version 1.1.463 running R version 3.5.2 on a windows computer with 16GB of 1200 MHz RAM and a 3.2 GHz processor (R Development Core Team 2013; Team 2015). All code used in simulation and testing of the ancestral condition test are available online via GitHub (<https://github.com/coleoguy/anc.cond.analyses>). We have implemented the ancestral condition test in the R package EvobiR which is publicly available and can be installed directly from GitHub (<https://github.com/coleoguy/evobir>). This package includes a variety of other useful functions for evolutionary biology (table 1).

**RESULTS**

Scenario one - Unidirectional change in the binary character

Analysis of the 100 simulated datasets with a scaling factor of one (no contingency between the two traits) showed a reasonable false-positive rate of 6% (Figure 2, black “x”). Analysis of the datasets with the scaling factor ranging from two to ten allowed us to assess the power of this approach for increasing degrees of contingency between the binary and continuous characters. We found that our ability to detect contingency in unidirectional evolution ranged from 15% when the scaling factor was two to a high of 78% when the scaling factor was eight (Figure 2, blue line). In our initial analyses, scaling factors of nine and ten had power slightly less than scaling factor eight. We performed subsequent analyses and found that scaling factors of 8-10 all produce power in the mid-seventies and that the slight reduction in power on scaling factors of 9-10 is simply due to the finite sample size of our analyses. The power of 78% likely indicates the maximum power of our approach to detect increasingly strong contingency in 200 species phylogenies. We also calculated the performance of existing methods to detect correlation using data sets simulated with scaling factors of one and five. Pagel’s test and the threshold test had powers of 22% and 12%, respectively, and false-positive rates of 7% and 18%. Under these same conditions, the ancestral condition test has a power of 50% and a false-positive rate of 6% (figure 3).

The number of taxa included in comparative analyses often varies by orders of magnitude. We found that our ancestral condition test requires moderately large sample sizes to detect a significant relationship reliably. Figure 4 shows that when our simulated phylogenetic trees contained 20 extant taxa, we were only able to detect a significant relationship 4% of the time (Figure 4, red line filled circles), a reduction in power of approximately 93% relative to the maximum power found for a scaling factor of 5. The maximum power we measured was for the simulations with 180 taxa, where we detected the underlying relationship in 58% of the simulated datasets. The power dipped below 50% with 200 taxa, although additional analyses showed that power with 180 and 200 taxa both range between 50% and 60%. Regardless of the number of taxa included in the analysis, the false-positive rate of our approach was approximately 5% (Figure 4, red line open circles).

Scenario two - Bidirectional change in the binary character

To test the performance of our method when the evolutionary history of the binary trait is more complex, we analyzed datasets where the binary state at the root of the tree was set to either one or two with equal probability, and transitions in either direction were allowed. We evaluated performance with a scaling factor of one to measure false-positive rate, and scaling factors from 2-10 to measure the power of our method under increasingly strong relationships between the continuous and binary characters. We measured a false-positive rate of 6% (Figure 2, black square). Power for this scenario ranged from 11% for a scaling factor of two to 54% for a scaling factor of eight (Figure 2 blue line). These power estimates were calculated with a level of significance alpha = 0.025 to correct for testing for contingency in two directions. We confirmed the ancestral condition test requires a large sample size in the bidirectional case as well. For small phylogenies, the ancestral condition test has power in the single digits and only ever reaches a power higher than 40% when there are 200 taxa present (Figure 4, blue line filled circles), the highest value tested. Regardless of the number of taxa included in the analysis, the false-positive rate of our approach was approximately 5% (Figure 4, blue line open circles).

Scenario 3 - Empirical Phylogeny

Our approach exhibited similar performance and when applied to an empirical phylogeny. Of the 100 simulated datasets, whose evolution is independent of body size, the Ancestral Condition test detected contingency (alpha = 0.05) in seven of these cases. This result indicates that our method has a false-positive rate of 7% when applied to this empirical phylogeny. This value is not considerably different than the false-positive rate measured using simulated phylogenies, indicating that the ancestral condition test performs similarly with simulated and empirical phylogenies.

**DISCUSSION**

Our simulation study demonstrates the conservative nature of our ancestral condition test, which has acceptable false-positive rates, but relatively low power when the sample size and/or contingency between the traits is small. In situations where no true relationship exists between the origin of a derived state of a binary trait and the value of a continuous trait (scaling factor of one), the test will incorrectly identify a significant relationship ~6% of the time under our simulated conditions. With a scaling factor of five (i.e., stronger contingency), branches with a high value of the continuous trait effectively have a transition rate into the derived state of the binary trait that is five times higher than branches with intermediate values. Likewise branches with a low value of the continuous trait effectively have a transition rate into the derived state of the binary trait that is five times lower than branches with intermediate values. This is arguably a very strong relationship, but even in this case, we detect the relationship in only ~50% of simulated datasets.

The ancestral condition test becomes less powerful when the evolution of the binary trait is bidirectional. For the highest scaling factor (i.e., *x* = 10, the strongest contingency) evaluated in our simulations, the power reduced from 74% to 53% when we allowed for bidirectional evolution. However, the false-positive rate remains a moderate 6% in both cases. Thus, these results indicate that the test can identify contingency in the evolution of traits with a high degree of certainty. However, the power of the ancestral condition test to detect even strong contingency in bidirectional evolution is more limited. In other words, under the conditions we have simulated, whenever the ancestral condition test detects contingency, there is a 94% chance it is correct, but when it does not detect contingency, then it is correct three out of four times, at best.

In several of our analyses, we found that as scaling factors increased (i.e., higher contingency between traits) we reached a plateau where no further increase in power was achieved, despite ever-stronger contingency. In retrospect this is not surprising. at the very highest contingency levels, every branch in the tree that is classified into the upper or lower quartiles is scaled to such an extent that those in the lower quartiles almost never exhibit changes in the binary character and branches in the upper quartile almost always exhibit transitions in the binary character. Based on our tree sizes and rates of evolution, we can infer this occurred when our scaling factor reached a value of eight. This plateau is clear in the right-hand side of Figure 2, where power plateaus at a scaling factor of eight for both unidirectional and bidirectional evolution scenarios. This point at which power plateaus should be a function of tree size where larger trees plateau at higher levels of contingency.

While current approaches for jointly analyzing continuous and binary traits can be effective for identifying correlations between traits in some scenarios, our AncCond test provides an important and useful extension to current comparative methods as it explicitly provides information about the order in which traits evolved across a phylogenetic tree. Furthermore, our approach is robust to possible confounding effects that one of the states of binary trait might exert on the evolution of the continuous trait. For instance, if we believe that state two of a binary trait may lead to selection for lower values in a corresponding continuous trait, we can prune the continuous data for all species in state two. By doing this we effectively ask if species still in state one yield an ancestral state estimation for the continuous trait that suggests that state two of the binary character is arising in parts of the phylogeny that have lower values than expected if the traits are truly independent. This approach removes a typical caveat from the interpretation of our results and moves us closer to inferring causation rather than simple correlation. Like many comparative methods, the power of the ancestral condition test is dependent on having a sufficiently large tree with transitions in the binary character. However, the false-positive rate remains acceptable even for the smallest phylogenies tested.

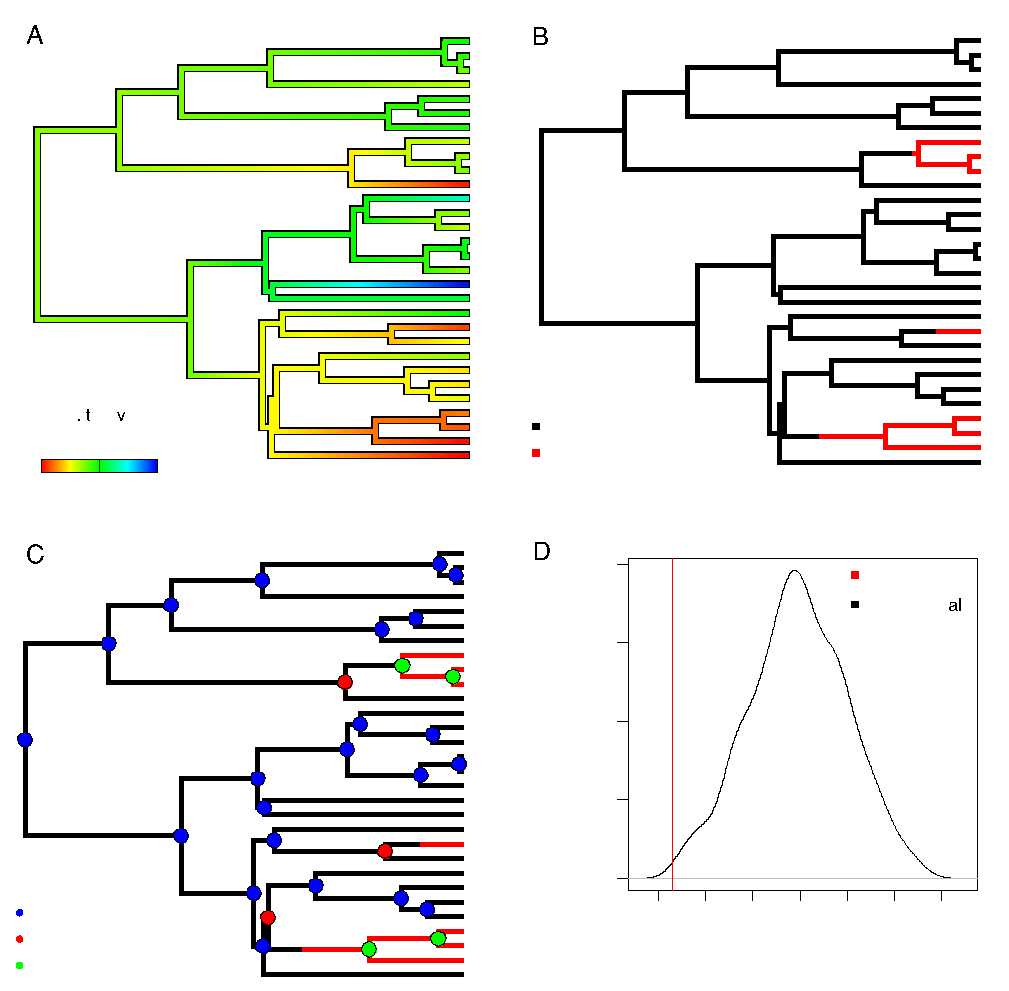
Our approach assumes an Mk2 model for the evolution of the binary trait and a Brownian motion model for the evolution of the continuous trait. The adequacy of these underlying models should be evaluated before using the ancestral condition test. Model adequacy of the continuous trait can be accomplished in several ways (Garland, Harvey & Ives 1992; Pennell *et al.* 2015). Perhaps the most robust of these methods involves first calculating test statistics on the phylogenetic independent contrasts from the observed data. Simulations are then performed, and these same test statistics are calculated for each simulation - generating null distributions. The benefit of this approach is that deviation of observed test statistics can determine if a model is adequate while providing information about what type of alternative model might perform better for a given dataset (Pennell *et al.* 2015). Evaluation of model adequacy for binary traits is currently less developed. One potential option is to use a Monte Carlo approach to determine if the model and parameter estimates are able to generate data with similar proportions of taxa in each discrete state (Price *et al.* 2012; Blackmon & Demuth 2014). An additional approach is to examine the number of state changes expected under parsimony for the observed data and compare this to the number of state changes expected under parsimony for simulated datasets (Beaulieu, O'Meara & Donoghue 2013). One advantage of the AncCond test is that while our implementation assumes a simple Brownian model for the continuous trait and an Mk2 model for the discrete trait, the same approach could be used with alternative ancestral state estimates generated under models found to be more appropriate for a practitioner’s particular dataset. Although we only implement the Brownian motion and Mk models in the ancestral condition test the process described in figure 1 could be repeated substituting a different model during the ancestral state reconstructions step A and B.

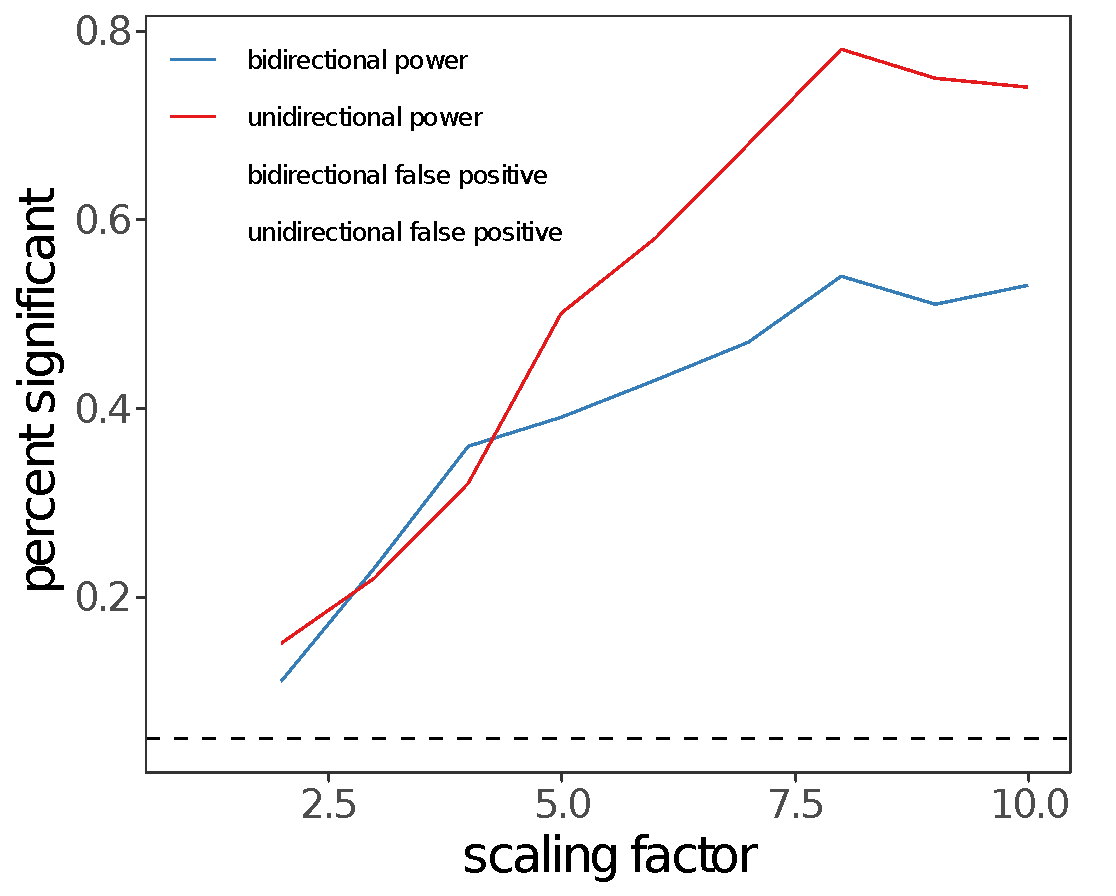
Often users may not have a single best tree, but rather a sample of a posterior distribution of trees. In these cases, the ancestral condition test can be run independently on each tree in the posterior. Users could then report the proportion of trees that support a significant relationship. Alternatively, our function returns the full null distribution generated for a tree, and these null distributions could be combined to create a null that incorporates phylogenetic uncertainty.

Many comparative methods can suffer from pseudoreplication (Maddison & FitzJohn 2014), returning significant results when some would argue that they should not (i.e., high false-positive rates), and our ancestral condition test could suffer from such shortcomings. For example, one could imagine a situation where the most recent common ancestor of a clade evolved a higher value for the continuous trait, followed by the evolution of an additional trait that increased the probability of transitions to the derived state of the binary state. Such a pattern could conceivably produce a significant result despite having only a single origin for the high continuous trait value. We believe the best approach to avoiding the errors associated with pseudoreplication is thorough data exploration before applying nearly any comparative method, including our AncCond test. There are many tools now available that allow simultaneous visualization of trees and data that can be useful in the early stages of analysis to alert researchers of possible issues. An alternative approach to solving the pseudoreplication problem is to perform what is effectively a meta-analysis. With this approach, the original dataset is divided into several independent data sets and the results of all of the independent tests can be evaluated (Mayrose *et al.* 2011).

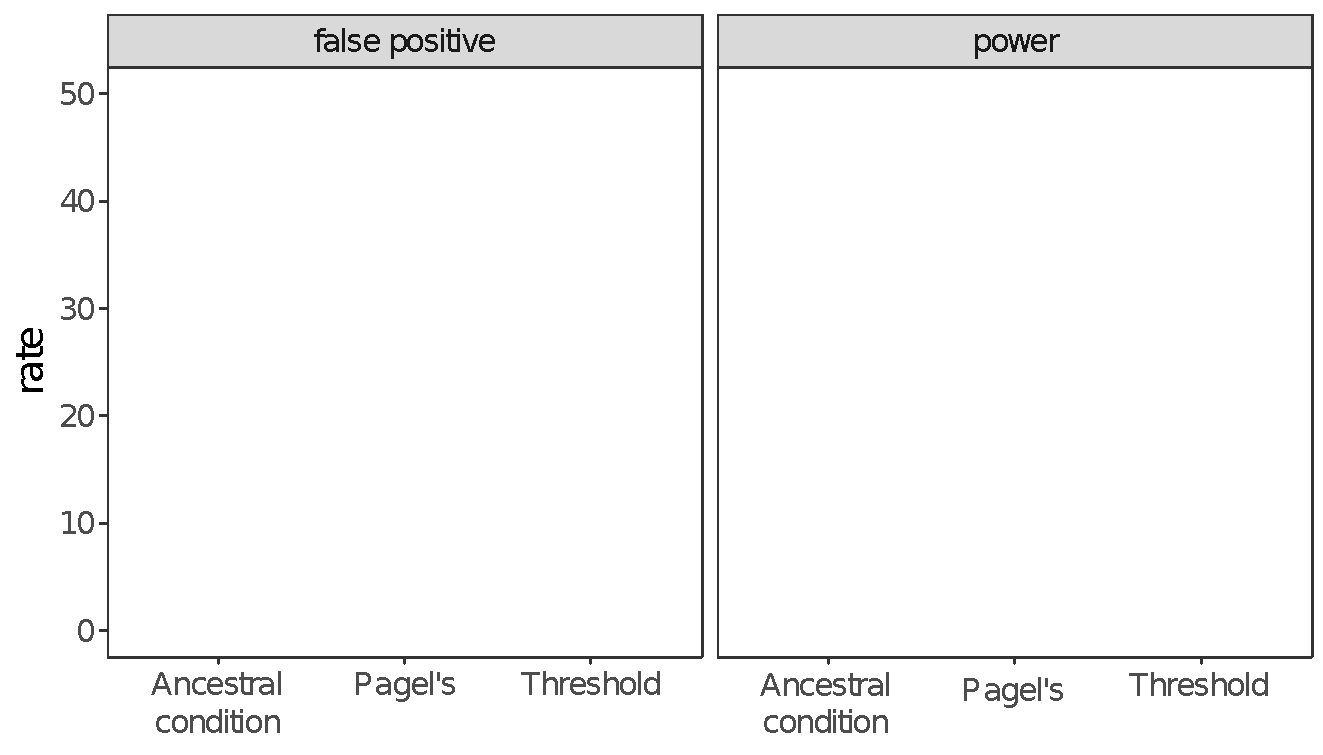
|  |  |
| --- | --- |
| **Function** | **Description** |
| **Applied Phylogenetics** |  |
| AncCond | Tests whether a derived state of a binary character originates when a continuous character has extreme values |
| **Population Genetics** |  |
| CalcD | Test of introgression algorithm described in (Durand et al. 2011) |
| WinCalcD | Sliding window version of CalcD |
| CalcPopD | Test of introgression algorithm described in (Durand et al. 2011) |
| **Miscellaneous** |  |
| ResSel | For use in selection experiments identifies those individuals for high or low selection lines after regression of one trait on another. |
| SampleTrees | Processes large nexus files, removing burn-in, randomly sampling, and saving in various formats. |
| FuzzyMatch | Identifies records in trees and trait dataset that may be lost due to differences in spelling. |
| SlidingWindow | Applies any function that operates on a vector to a sliding window across a vector or matrix |
| SuperMatrix | Combines multiple alignments with varying taxa sets into a single supermatrix. |

Table 1 Summary of functions available in EvobiR.

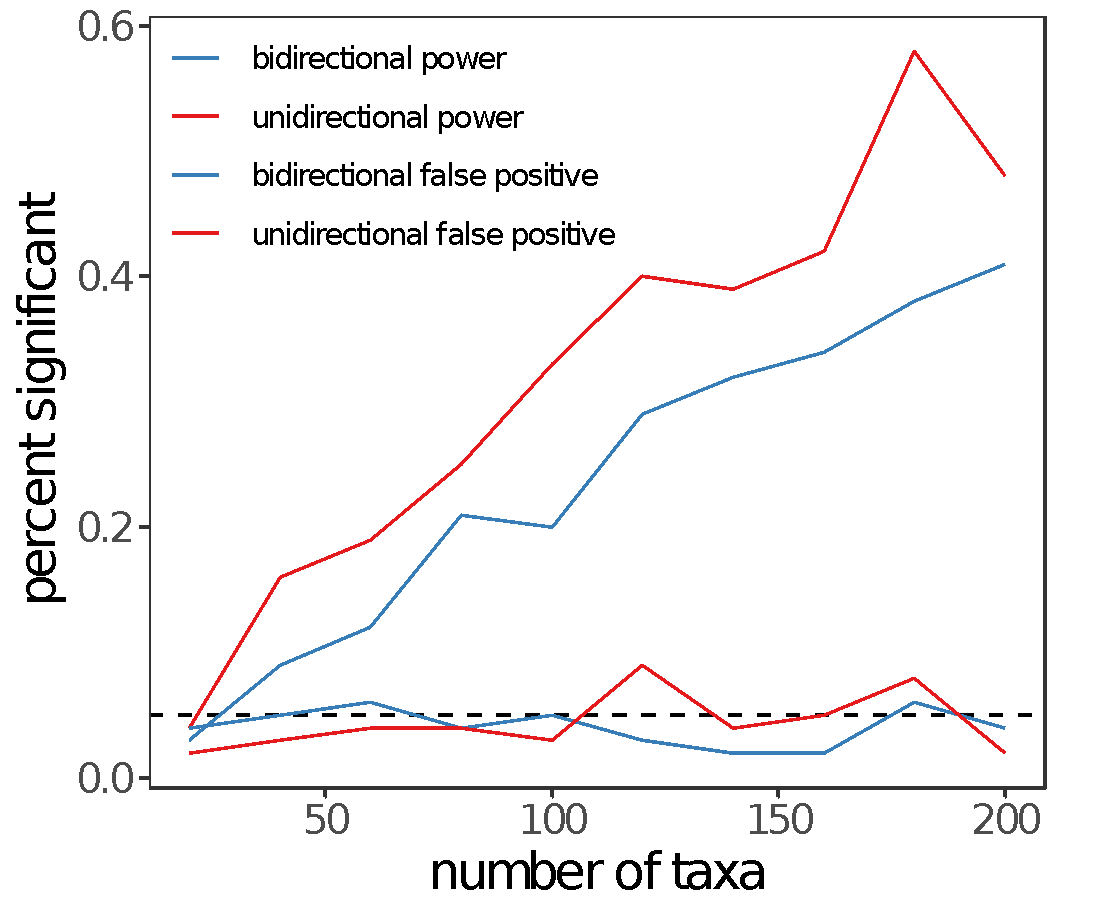
Figure 1. The four steps in the ancestral condition test. A) Ancestral states are estimated for the continuous character assuming a Brownian motion model of evolution. B) Possible evolutionary histories for the binary trait are generated through stochastic mapping C) Nodes are categorized as either ancestral or derived, and ancestral nodes that subtend an origin of the derived state of the binary character are annotated. D) Depiction of the null distribution and the observed mean of producing nodes estimated from the data. In this example the producing nodes have a lower continuous value than expected if there is no relationship between the traits.



**Figure 2. False-positive rate and power of the AncCond test as a function of the branch scaling factor.** Scaling factor used to generate binary traits (where higher values indicate greater contingency between traits) shown on the horizontal axis plotted against the proportion of simulations for which the ancestral condition test detected significant contingency on the vertical axis. For a scaling factor of one, this proportion represents the false-positive rate of our method, indicated for scenario one (unidirectional evolution) with an X and indicated for scenario two (bidirectional evolution) with an open box. For scaling factors greater than one, the significant results are correct and represent the power of the ancestral condition test to detect the true relationship with different strengths of contingency in evolution. The blue line denotes the performance of our test in scenario one, the red line scenario two. The dashed black line marks 5%.



**Figure 3. Comparison to existing methods.** In the left panel false-positive rate for the ancestral condition test is compared to Pagel’s method with discretization of the continuous trait and to the threshold method where the binary trait is assumed to be an expression of an underlying continuous trait. In the right panel the power of the ancestral condition is compared to Pagel’s method and to the threshold method.



**Figure 4. The impact of sample size on test performance.** The number of extant taxa included in the analysis plotted on the horizontal axis against the proportion of simulations for which AncCond detected significant contingency plotted on the vertical axis. Lines with open circles indicate the false-positive rate, and lines with filled circles represent power. In both cases red lines indicate scenario two (bidirectional evolution) and blue lines indicate scenario one (unidirectional evolution), Scaling factors is one for false-positive analysis and five for power analysis. The dashed black line marks 5%.

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. *Systematic Biology,* **62,** 725-737.

Blackmon, H. & Demuth, J.P. (2014) Estimating tempo and mode of Y chromosome turnover: explaining Y chromosome loss with the fragile Y hypothesis. *Genetics,* **197,** 561-572.

Blackmon, H. & Demuth, J.P. (2015) The fragile Y hypothesis: Y chromosome aneuploidy as a selective pressure in sex chromosome and meiotic mechanism evolution. *Bioessays,* **37,** 942-950.

Blackmon, H., Hardy, N.B. & Ross, L. (2015) The evolutionary dynamics of haplodiploidy: Genome architecture and haploid viability. *Evolution,* **69,** 2971-2978.

Blackmon H, R.A.A. (2015) EvobiR: tools for comparative analyses and teaching evolutionary biology.

Bollback, J.P. (2006) SIMMAP: stochastic character mapping of discrete traits on phylogenies. *BMC Bioinformatics,* **7,** 88.

Butler, M.A. & King, A.A. (2004) Phylogenetic comparative analysis: a modeling approach for adaptive evolution. *The American Naturalist,* **164,** 683-695.

Felsenstein, J. (1985) Phylogenies and the comparative method. *American Naturalist***,** 1-15.

Felsenstein, J. (2004) *Inferring phylogenies*. Sinauer Associates, Sunderland, MA.

Felsenstein, J. (2012) A comparative method for both discrete and continuous characters using the threshold model. *The American Naturalist,* **179,** 145-156.

FitzJohn, R.G. (2012) Diversitree: comparative phylogenetic analyses of diversification in R. *Methods in Ecology and Evolution,* **3,** 1084-1092.

Garland, T., Harvey, P.H. & Ives, A.R. (1992) Procedures for the analysis of comparative data using phylogenetically independent contrasts. *Systematic Biology,* **41,** 18-32.

Hadfield, J.D. (2010) MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. *Journal of Statistical Software,* **33,** 1-22.

Harmon, L.J., Weir, J.T., Brock, C.D., Glor, R.E. & Challenger, W. (2008) GEIGER: investigating evolutionary radiations. *Bioinformatics,* **24,** 129-131.

Huelsenbeck, J.P., Nielsen, R. & Bollback, J.P. (2003) Stochastic mapping of morphological characters. *Systematic Biology,* **52,** 131-158.

Landis, M.J., Schraiber, J.G. & Liang, M. (2012) Phylogenetic analysis using Lévy processes: finding jumps in the evolution of continuous traits. *Systematic Biology,* **62,** 193-204.

Lewis, P.O. (2001) A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic Biology,* **50,** 913-925.

Maddison, W.P. & FitzJohn, R.G. (2014) The unsolved challenge to phylogenetic correlation tests for categorical characters. *Systematic Biology***,** syu070.

Mayrose, I., Zhan, S.H., Rothfels, C.J., Magnuson-Ford, K., Barker, M.S., Rieseberg, L.H. & Otto, S.P. (2011) Recently formed polyploid plants diversify at lower rates. *Science,* **333,** 1257-1257.

O'Meara, B.C., Ane, C., Sanderson, M.J. & Wainwright, P.C. (2006) Testing for different rates of continuous trait evolution using likelihood. *Evolution,* **60,** 922-933.

Pagel, M. (1994) Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. *Proceedings of the Royal Society of London B: Biological Sciences,* **255,** 37-45.

Pennell, M.W., FitzJohn, R.G., Cornwell, W.K. & Harmon, L.J. (2015) Model adequacy and the macroevolution of angiosperm functional traits. *The American Naturalist,* **186,** E33-E50.

Price, S.A., Hopkins, S.S., Smith, K.K. & Roth, V.L. (2012) Tempo of trophic evolution and its impact on mammalian diversification. *Proceedings of the National Academy of Sciences,* **109,** 7008-7012.

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

Revell, L.J. (2012) phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution,* **3,** 217-223.

Revell, L.J. (2014) Ancestral character estimation under the threshold model from quantitative genetics. *Evolution,* **68,** 743-759.

Ross, L., Hardy, N.B., Okusu, A. & Normark, B.B. (2012) Large population size predicts the distribution of asexuality in scale insects. *Evolution,* **67,** 196-206.

Slater, G.J., Price, S.A., Santini, F. & Alfaro, M.E. (2010) Diversity versus disparity and the radiation of modern cetaceans. *Proceedings of the Royal Society B: Biological Sciences,* **277,** 3097-3104.

Steeman, M.E., Hebsgaard, M.B., Fordyce, R.E., Ho, S.Y., Rabosky, D.L., Nielsen, R., Rahbek, C., Glenner, H., Sørensen, M.V. & Willerslev, E. (2009) Radiation of extant cetaceans driven by restructuring of the oceans. *Systematic Biology,* **58,** 573-585.

Team, R. (2015) RStudio: Integrated Development for R.RStudio, Inc., Boston, MA.

Uusitalo, L. (2007) Advantages and challenges of Bayesian networks in environmental modelling. *Ecological modelling,* **203,** 312-318.

Wright, S. (1934) An analysis of variability in number of digits in an inbred strain of guinea pigs. *Genetics,* **19,** 506.