1 Assessing the Impact of Continuous Traits on the Evolution of Discrete Traits: The Ancestral
2 Condition Test.
3
4 Running title: ACT
5
6 Keywords: comparative method, phylogenetics, trait correlation
7
8 Nathan Anderson[1], Jeffery P. Demuth[2], Richard H. Adams[3], and Heath Blackmon[4,5]
9
10 [1] Department of Math, Texas A&M, College Station, TX USA
11 [2] Department of Biology, University of Texas at Arlington, TX, USA
12 [2] Department of Computer and Electrical Engineering and Computer Science, Florida Atlantic University, FL, USA
13 [3] Department of Biology, Texas A&M, College Station, TX USA
14 [4] coleoguy@gmail.com
15
16 **ABSTRACT**
17
18 Co-evolutionary relationships whereby the value of a continuous trait affects the evolution of a

19 discrete trait remain challenging to study. Here we present the ancestral condition test, a new

20 comparative method that evaluates whether a discrete trait tends to transition when a continuous

21 trait has values more extreme than expected if both traits were evolving independently. This

22 approach leverages ancestral state estimates of both the continuous and the discrete trait to test

23 whether extreme values of the continuous trait appear to be associated with transitions in the

24 discrete trait, and to assess statistical significance. We explore the robustness of our approach

25 under a range of parameter values and patterns of trait evolution. We find that either a relatively

26 strong contingency between the two traits or a large number of taxa is required to detect the

27 underlying relationships reliably. Statistical power of the test is highest when the discrete trait

28 evolves unidirectionally, and we find that the false-positive rate remains acceptable for a

29 bidirectionally evolving discrete trait in many scenarios. In comparison to existing methods that

30 can be used to explore similar hypotheses, we show that the ancestral condition test has both

31 higher power and a lower false-positive rate. The types of questions that this approach allows us

32 to test are common in evolutionary biology and, unlike existing methods, the ancestral condition

33 test can be applied to account for the temporal order of transitions – moving a step closer to

34 inferring causality rather than merely identifying correlation.

35

36

37

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 discrete traits (Felsenstein, 2012, Lewis, 2001) and continuous traits (Butler & King, 2004, Felsenstein, 1985, Landis et al., 2012, O'Meara et al., 2006). 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 the way a discrete 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 affects the evolution of a discrete trait are less well developed (but see Felsenstein, 2012, Hadfield, 2010, Revell, 2014).

Hypotheses that suggest that extreme values of a continuous trait are associated with transitions in a discrete trait are common in evolutionary biology (Blackmon & Demuth, 2015, Blackmon et al., 2015, Ross et al., 2012). In a recent study of sex chromosome evolution, for example, we were interested in testing whether loss of the Y chromosome (a discrete transition) 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 biological 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. Another approach

3

64   would be to assume that our discrete state (presence or absence of a Y chromosome) could be

65   modeled as if it was produced by an underlying continuous trait that is unobserved. If this

66   assumption is justified, we could employ the threshold model (Felsenstein, 2012, Wright, 1934),

67   which would allow us to estimate the correlation between chromosome number and the presence

68   or absence of the Y chromosome. A third potential approach would be to use a phylogenetic

69   mixed model that treats chromosome number as a predictor variable and presence or absence of

70   the Y chromosome as a response variable (Hadfield, 2010). However, each of these approaches,

71   as commonly applied, suffers from the same fundamental problem: correlation does not provide

72   information about the direction of causality. For example, does high chromosome number lead to

73   Y chromosome loss, or does Y chromosome loss lead to high chromosome number (Blackmon &

74   Demuth, 2015)? Furthermore, many of these current approaches suffer from excessively high

75   false-positive rates (Maddison & FitzJohn, 2014).

76

77   In light of these issues, we have developed a flexible, conservative statistical test to infer

78   significant relationships between the value of a continuous trait and transitions in a discrete trait.

79   Briefly, our approach estimates the mean value of the continuous trait when the discrete character

80   undergoes transitions. We then evaluate the significance of this mean value by comparison to a

81   null distribution that is generated by sampling estimated mean ancestral states of the continuous

82   trait under the hypothesis that there is no relationship. We evaluate the ancestral condition test

83   using both simulations and empirical datasets. We find that our method is conservative, with the

84   highest power when the number of taxa and/or the degree of contingency between traits is large.

85   In comparison to existing methods (Pagel's and Threshold), the test we develop offers higher

86   power and lower false-positive rates. In conjunction with improved performance, our approach

87   also grants insight into the temporal order of changes and allows us to get closer to the possible

88   causal relationships between traits. Our ancestral condition test represents a novel addition to the

89   phylogenetic comparative method toolkit that is useful for studying the interplay between the state

90   of a continuous trait and the evolution of a discrete character.

91

92   Furthermore, the approach we have developed offers a large degree of flexibility that is missing in

93   many existing methods. For instance, in many cases we may know that the derived state of the

94   discrete character is associated with lower values of the continuous trait, but is this because low

95   values of the continuous trait lead to changes in the discrete character, or does the derived state

96   of the discrete character lead to reductions in the continuous trait? When this is a concern, we can

97   restrict our estimate of the continuous trait to only use data from species that have retained the

98   ancestral condition of the discrete trait. This approach provides a much more conservative test

99   than can be implemented with existing methods. In practice, there is not always a clear ancestral

100 and derived state, particularly when a discrete character evolves quickly and lacks fossil data.

101 This uncertainty does not pose a problem for the ancestral condition test, but in such cases, the

102 power of our approach is reduced.

103

104 **METHODS**

105

106 *The Ancestral Condition Test*

107 Our approach determines whether there is evidence that nodes in the phylogenetic tree

108 subtending transitions of a discrete character tend to exhibit extreme values of the continuous

109 character under study. We refer to this method as the Ancestral Condition (AncCond) test, which

110 involves the following four steps:

111

112   1) Estimate the ancestral values of the continuous character while optionally pruning data

113   from species that exhibit the derived state of the discrete character. By pruning data from

114   taxa exhibiting the derived state of the discrete character, we create a more conservative

5

115   test that will be applicable in cases where the derived state of the discrete character is

116   expected to lead to changes in the selective forces acting on the continuous character

117   (addressing the potential for feedback mentioned above based on the users selection).

118   Ancestral state estimation is accomplished under a Brownian motion model using the

119   existing function *anc.ML* in phytools (Figure 1A) (Revell, 2012).

120

121   2) Next, we perform ancestral state estimation of the discrete character by way of

122   stochastic character mapping (Figure 1B) (Bollback, 2006, Huelsenbeck et al., 2003). We

123   may choose to fix a single state at the root of the tree in cases where this is appropriate.

124   For instance, when the ancestral state is informed by fossil records or when prior biological

125   knowledge suggests that transitions are expected to occur in one direction (Blackmon et

126   al., 2015). In the latter case, users may also restrict the simulated evolution of the discrete

127   character by allowing only transitions from the ancestral to the derived state. Stochastic

128   mapping is performed using the *make.simmap* function provided in phytools (Figure 1B).

129   For simplicity, we will refer to the two states as one and two; state one will denote the

130   ancestral state, and state two is the derived state.

131

132   3) We then process the stochastic maps to classify all nodes in the tree into one of three

133   groups: a) Nodes exhibiting state one (i.e., ancestral state) of the discrete character and

134   no transitions in discrete character in immediate daughter branches (Figure 1C, blue

135   nodes), b) Nodes exhibiting state two (i.e., derived state) of the discrete character and no

136   transitions in discrete character in immediate daughter branches (Figure 1C, green nodes),

137   c) Producing nodes exhibiting either state of the discrete character but with one of the

138   immediate daughter branches showing a transition into the other state of the discrete

139   character (Figure 1C, red nodes). In the bidirectional case, we then parse producing nodes

140   to select those that subtend branches with transitions from state one to state two or from

6

141   state two to state one. In the unidirectional case, this parsing is not necessary. Next, we

142   calculate the mean value of the continuous trait at the producing nodes in each group. This

143   process is repeated across stochastic mappings incorporating ancestral state estimation

144   uncertainty. The mean of type c (producing nodes) is our observed statistic. We refer to

145   this value as the "estimated ancestral condition" for each of the states of the discrete

146   character (Figure 1C). In the case where transitions are restricted to a single direction,

147   there will be a single estimated ancestral condition for transitions from the ancestral to the

148   derived state. In the more general case, there will be two estimated ancestral conditions

149   for transitions from state one to state two and for transitions from state two to state one.

150

151   4) We construct a null distribution by repeatedly sampling a random subset of the

152   continuous trait estimates from all nodes in the correct discrete trait state (i.e. sampling

153   irrespective of whether transitions happen in immediate daughter branches). The number

154   of nodes in the subset is equal to the number of producing nodes sampled for the mean

155   ancestral condition statistic (e.g. 3 in the Figure 1 example). This sampling process is

156   repeated for a user-specified number of replicates (default = 10^3). These sampled null

157   ancestral condition values define a null distribution and we compare our test statistic (the

158   estimated ancestral condition) to this distribution to obtain an empirical p-value (Figure

159   1D).

160

161   **Simulated data**

162

163   *Scenario one–unidirectional change in a discrete character*

164   We simulated 100 trees using a birth-death model with a birth rate of 3.0 and death rate of 1.0,

165   allowing trees to grow until 200 extant species were reached. Phylogenies were rescaled to unit

166   length. Trees were simulated using the function *trees* in Diversitree (FitzJohn, 2012). Next, we

7

167 simulated a continuous character evolving via Brownian motion with a rate parameter of 0.2 and a

168 starting mean of 0 on each tree using the function *sim.char* in Geiger (Harmon et al., 2008). We

169 then used the function *anc.ML* in Phytools (Revell, 2012) to estimate the mean value of the

170 continuous trait for each node in the phylogeny.

171

172 To simulate a discrete trait whose evolution was contingent on the continuous trait, we used a

173 branch scaling approach. We first identified those branches that had a mean value of the

174 continuous trait in the extreme upper or lower quartiles. Branches whose mean was in the lower

175 quartile were scaled by a multiplicative factor of $1/x$, while branches in the upper quartile were

176 scaled by a multiplicative factor of x. We repeated this process for ten values ($x = 1 - 10$) on each

177 of our 100 trees. We then simulated the evolution of the discrete character on these scaled trees.

178 We set the root to state one (ancestral state) and allowed the trait to evolve under an Mk2 model

179 where transitions to state two (derived state) were allowed, but not transitions from state two to

180 state one (i.e., modeling unidirectional evolution), as we expected our approach to be most

181 powerful under this mode of evolution. This simulation was done using the *sim.char* function in

182 Geiger. Various transition rates from state one to state two were evaluated, and we found that a

183 rate of 0.1 was sufficient to ensure that multiple transitions occurred in all simulated datasets. If a

184 dataset did not contain at least ten extant taxa exhibiting each state of the discrete character, it

185 was discarded (Between 7% and 55% of simulations were discarded for a given scaling factor,

186 with a higher percent being discarded for smaller scaling factors). Thus, our total of 10^3 simulated

187 datasets consisted of ten degrees of contingency each applied to 100 trees.

188

189 By simulating the evolution of the discrete trait on these scaled trees, we generated a contingency

190 between the discrete and continuous traits. There exists more opportunity for transitions on the

191 branches where the continuous trait value is in the upper quartile and the branch length was

192 increased. Alternatively, there exists less opportunity for transitions on the branches where the

8

193   continuous trait value is in the lower quartile and the branch length was decreased. This branch

194   scaling technique effectively altered the rate of evolution of the discrete trait wherever the

195   continuous trait took extreme values, increasing the density of transitions on branches where the

196   continuous trait had high values and decreasing the density for branches with low values.

197

198   To assess the relationship between the number of taxa included in an analysis and the

199   performance of our approach, we simulated birth-death trees and continuous characters using the

200   methods described above but with tree sizes that ranged from 20 to 200 taxa in increments of 20.

201   If a dataset did not include at least five extant taxa in each state of the discrete character, it was

202   discarded, and a new dataset was generated (between 14% and 91% for different sized trees,

203   with small trees leading to a higher proportion of discarded simulations). We then simulated

204   discrete datasets with scaling factors $x = 1$ or $x = 5$ as described above, allowing us to measure

205   both power ($x = 5$) and false-positive rate ($x = 1$; i.e. how often we detect a relationship under

206   conditions where the data were simulated with no contingency between traits). A total of 200 data

207   sets (scaling factors of one and five) were generated for each of the ten tree sizes for a total of

208   2000 simulated datasets.

209

210   *Scenario two – bidirectional change in a discrete character*

211   We also used simulated data sets to evaluate the performance of the AncCond test in scenarios

212   for which a discrete character has transitions in both directions (i.e., bidirectional), and when we

213   are unable to confidently assign an ancestral - derived relationship. We repeated the above

214   procedure with two key differences. First, we did not fix the root of the tree to any one state.

215   Instead, we allowed the root to take either state with equal probability. Second, we allowed for

216   transitions in either direction, from state one to state two and from state two to state one, with

217   equal rates. A transition rate of 0.3 was used to ensure there were several forward and reverse

218   transitions in every dataset.

9

219

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 discrete 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).

227

Assessing the performance of comparative tests

229

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 discrete 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 discrete character. This pruning is not entirely applicable to this simulated

data, as the continuous character was simulated completely independent of the discrete

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 discrete 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 discrete 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 taxa. In

assessing the data from scenario three, we followed the same procedure as scenario one.

244

10

245   We compared the ancestral condition test to two existing approaches. For these comparisons, we

246   used a subset of data simulated under scenario one (trees with 200 taxa and scaling factors of

247   one and five). First, we applied the threshold model as implemented in phytools. This approach

248   allowed us to test for a correlation between our continuous character and the discrete character

249   by inferring an underlying continuous value that we observe as a discretely expressed character.

250   Second, we discretized our continuous character at the median value and applied Pagel's test for

251   correlation of discrete characters also implemented in phytools (Revell, 2012).

252

253   Below we show the performance of our approach with the simulated data, reporting both false-

254   positive rate and power. Tests were considered statistically significant at a p-value ≤ 0.05 for data

255   analyzed under scenario one and three but at a p-value of ≤ 0.025 for scenario two. The reduction

256   in p-value accounts for the fact that we are testing two hypotheses 1) transitions from state one to

257   two are associated with extreme values of the continuous trait, and 2) that transitions from two to

258   one are associated with extreme values of the continuous trait. All analyses were completed with

259   RStudio version 1.1.463 running R version 3.5.2 on a windows computer with 16GB of 1200 MHz

260   RAM and a 3.2 GHz processor (R Development Core Team, 2013, Team, 2015)

261

262   **RESULTS**

263

264   Scenario one - Unidirectional change in the discrete character

265

266   Analysis of the 100 simulated datasets with a scaling factor of one (no contingency between the

267   two traits) showed a reasonable false-positive rate of 6% (Figure 2, black "x"). Analysis of the

268   datasets with the scaling factor ranging from two to ten allowed us to assess the power of this

269   approach for increasing degrees of contingency between the discrete and continuous characters.

270   We found that our ability to detect contingency in unidirectional evolution ranged from 15% when

11

271   the scaling factor was two to a high of 78% when the scaling factor was eight (Figure 2, blue line).

272   In our initial analyses, scaling factors of nine and ten had power slightly less than scaling factor

273   eight. We performed subsequent analyses and found that scaling factors of 8-10 all produce

274   power in the mid-seventies and that the slight reduction in power on scaling factors of 9-10 is

275   simply due to the finite sample size of our analyses. The power of 78% likely indicates the

276   maximum power of our approach to detect increasingly strong contingency in 200 species

277   phylogenies. We also calculated the performance of existing methods to detect correlation using

278   data sets simulated with scaling factors of one and five. Pagel's test and the threshold test had

279   powers of 22% and 12%, respectively, and false-positive rates of 7% and 18%. Under these same

280   conditions, the ancestral condition test has a power of 50% and a false-positive rate of 6% (figure

281   4).

282

283   The number of taxa included in comparative analyses often varies by orders of magnitude. We

284   found that our ancestral condition test requires moderately large sample sizes to detect a

285   significant relationship reliably. Figure 3 shows that when our simulated phylogenetic trees

286   contained 20 extant taxa, we were only able to detect a significant relationship 4% of the time

287   (Figure 3, red line filled circles), a reduction in power of approximately 93% relative to the

288   maximum power found in this analysis. The maximum power we measured was for the

289   simulations with 180 taxa, where we detected the underlying relationship in 58% of the simulated

290   datasets. The power dipped below 50% with 200 taxa, although additional analyses showed that

291   power with 180 and 200 taxa both range between 50% and 60%. Regardless of the number of

292   taxa included in the analysis, the false-positive rate of our approach was approximately 5%

293   (Figure 3, red line open circles).

294

295   Scenario two - Bidirectional change in the discrete character

296

12

297   To test the performance of our method when the evolutionary history of the discrete trait is more

298   complex, we analyzed datasets where the binary state at the root of the tree was set to either one

299   or two with equal probability, and transitions in either direction were allowed. We evaluated

300   performance with a scaling factor of one to test false-positive rate, and scaling factors from 2-10 to

301   analyze the power of our method under increasingly strong relationships between the continuous

302   and discrete characters. We measured a false-positive rate of 6% (Figure 2, black square). Power

303   for this scenario ranged from 11% for a scaling factor of two to 54% for a scaling factor of eight

304   (Figure 2 blue line). These power estimates were calculated with a level of significance alpha =

305   0.025 to correct for testing for correlation in two directions. We confirmed the ancestral condition

306   test requires a large sample size in the bidirectional case as well. For small phylogenies, the

307   ancestral condition test has power in the single digits and only ever reaches a power higher than

308   40% when there are 200 taxa present (Figure 3, blue line filled circles), the highest value tested.

309   Regardless of the number of taxa included in the analysis, the false-positive rate of our approach

310   was approximately 5% (Figure 3, blue line open circles).

311

312   Scenario 3 - Empirical Phylogeny

313

314   Our approach exhibited similar performance and when applied to empirical phylogenies. Of the

315   100 discrete characters, whose evolution is independent of body size, the Ancestral Condition test

316   detected correlation (alpha = 0.05) in seven of these cases. This result indicates that our method

317   has a false-positive rate of 7% when applied to an empirical phylogeny. This value is not

318   considerably different than the false-positive rate measured using simulated data, indicating that

319   the ancestral condition test performs similarly with simulated and empirical phylogenies.

320

321   **DISCUSSION**

322

13

323 Our simulation study demonstrates the conservative nature of our ancestral condition test, which

324 has acceptable false-positive rates, but relatively low power when the sample size and/or

325 contingency between the traits is small. In situations where no true relationship exists between the

326 origin of a derived state of a discrete trait and the value of a continuous trait (scaling factor of

327 one), the test will incorrectly identify a significant relationship ~6% of the time under our simulated

328 conditions. With a scaling factor of five (i.e., stronger contingency), branches with a high value of

329 the continuous trait effectively have a transition rate into the derived state of the discrete trait that

330 is five times higher than branches with intermediate values and 25 times higher than branches

331 with low values. This is arguably a very strong relationship, but even in this case, we detect the

332 relationship in only ~50% of simulated datasets.

333

334 The ancestral condition test becomes more conservative when the evolution of the binary trait is

335 bidirectional. For the highest scaling factor (i.e., $x = 10$, the strongest contingency) evaluated in

336 our simulations, the power reduced from 74% to 53% when we allowed for bidirectional evolution.

337 However, the false-positive rate remains a moderate 6% in both cases. Thus, these results

338 indicate that the test can identify correlation in the evolution of traits with a high degree of

339 certainty. However, the power of the ancestral condition test to detect even strong correlation in

340 bidirectional evolution limits the test's applications in those scenarios. In other words, under the

341 conditions we have simulated, whenever the ancestral condition test detects correlation, there is a

342 94% chance it is correct, but when it does not detect correlation, then it is correct three out of four

343 times, at best.

344

345 In several of our analyses, we found that as scaling factors increased (i.e., higher contingency

346 between traits) we reached a plateau where no further increase in power was achieved, despite

347 ever-stronger contingency. At the very highest contingency levels, every branch in the tree that is

348 classified into the upper or lower quartiles is scaled to such an extent that those in the lower

14

349   quartiles never exhibit changes in the discrete character and branches in the upper quartile all

350   exhibit transitions in the discrete character. Based on our tree sizes and rates of evolution, we can

351   infer this occurred when our scaling factor reached a value of eight. This plateau is clear in the

352   right hand side of Figure 2, where power plateaus at a scaling factor of eight for both

353   unidirectional and bidirectional evolution scenarios.

354

355   While current approaches for jointly analyzing continuous and discrete traits can be effective for

356   identifying correlations between traits in some scenarios, our novel AncCond test provides an

357   important and useful extension to current comparative methods as it explicitly provides information

358   about the order in which traits evolved across a phylogenetic tree. Furthermore, our approach is

359   robust to possible confounding effects that one of the states of discrete trait might exert on the

360   evolution of the continuous trait. For instance, if we believe that state two of a binary trait may lead

361   to selection for lower values in a corresponding continuous trait, we can prune the continuous

362   data for all species in state two. By doing this we effectively ask if species still in state one yield an

363   ancestral state estimation for the continuous trait that suggests that state two of the discrete

364   character is arising in parts of the phylogeny that have lower values than expected if the traits are

365   truly independent. This approach removes a typical caveat from the interpretation of our results

366   and moves us closer to inferring causation rather than simple correlation. Like many comparative

367   methods, the power of the ancestral condition test is dependent on having a sufficiently large tree

368   with transitions in the discrete character. However, the false-positive rate remains acceptable

369   even for the smallest phylogenies tested.

370

371   Our approach assumes an Mk2 model for the evolution of the discrete trait and a Brownian motion

372   model for the evolution of the continuous trait. The Mk2 model implemented assumes a discrete

373   character that has an ancestral and derived state and does not experience reversion. It would be

374   straightforward to extend our approach to allow for other models for the discrete trait; however,

15

375   more complex models will often lead to greater uncertainty in ancestral state estimates, and thus

376   may led to lower power for detecting significant relationships. The adequacy of these underlying

377   models should be evaluated before using the ancestral condition test. Model adequacy of the

378   continuous trait can be accomplished in several ways (Garland et al., 1992, Pennell et al., 2015).

379   Perhaps the most robust of these methods involves first calculating test statistics on the

380   phylogenetic independent contrasts from the observed data. Simulations are then performed, and

381   these same test statistics are calculated for each simulation - generating null distributions. The

382   benefit of this approach is that deviation of observed test statistics can determine if a model is

383   adequate while providing information about what type of alternative model might perform better for

384   a given dataset (Pennell et al., 2015). Evaluation of model adequacy for discrete traits is currently

385   less developed. One potential option is to use a Monte Carlo approach to determine if the model

386   and parameter estimates are able to generate data with similar proportions of taxa in each

387   discrete state (Blackmon & Demuth, 2014, Price et al., 2012). An additional approach is to

388   examine the number of state changes expected under parsimony for the observed data and

389   compare this to the number of state changes expected under parsimony for simulated datasets

390   (Beaulieu et al., 2013). One advantage of the AncCond test is that while our implementation

391   assumes a simple Brownian model for the continuous trait and an Mk2 model for the discrete trait,

392   the same approach could be used with alternative ancestral state estimates generated under

393   models found to be more appropriate for a practitioner's particular dataset. Often users may not

394   have a single best tree, but rather a sample of a posterior distribution of trees. In these cases, the

395   ancestral condition test can be run independently on each tree in the posterior. Users could then

396   report the proportion of trees that support a significant relationship. Alternatively, our function

397   returns the full null distribution generated for a tree, and these null distributions could be combined

398   to create a null that incorporates phylogenetic uncertainty.

399

400   Many comparative methods can suffer from pseudoreplication (Maddison & FitzJohn, 2014),

401   returning significant results when some would argue that they should not (i.e., high false-positive

402   rates), and our ancestral condition test could suffer from such shortcomings. For example, one

403   could imagine a situation where the most recent common ancestor of a clade evolved a higher

404   value for the continuous trait, followed by the evolution of an additional trait that increased the

405   probability of transitions to the derived state of the discrete state. Such a pattern could

406   conceivably produce a significant result despite having only a single origin for the high continuous

407   trait value. We believe the best approach to avoiding the errors associated with pseudoreplication

408   is thorough data exploration before applying nearly any comparative method, including our

409   AncCond test. There are many tools now available that allow simultaneous visualization of trees

410   and data that can be useful in the early stages of analysis to alert researchers of possible issues.

411   An alternative approach to solving the pseudoreplication problem is to perform what is effectively

412   a meta-analysis. With this approach, the original dataset is divided into several independent data

413   sets and the results of all of the independent tests can be evaluated (Mayrose et al., 2011).

414

415   We have implemented ancestral condition test in the R package EvobiR which is publicly available

416   and can be installed directly from GitHub (https://github.com/coleoguy/evobir). This package

417   includes a variety of other useful functions for evolutionary biology (table 1).

418

419

420

Function	Description
Applied Phylogenetics	
AncCond	Tests whether a derived state of a binary character originates when a continuous character has extreme values
PPSDiscrete	Produce posterior predictive datasets based on log files from programs such as BayesTraits or diversitree.
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
SuperMatrix	Combines multiple alignments with varying taxa sets into a single supermatrix.

421 Table 1 Summary of functions available in EvobiR.
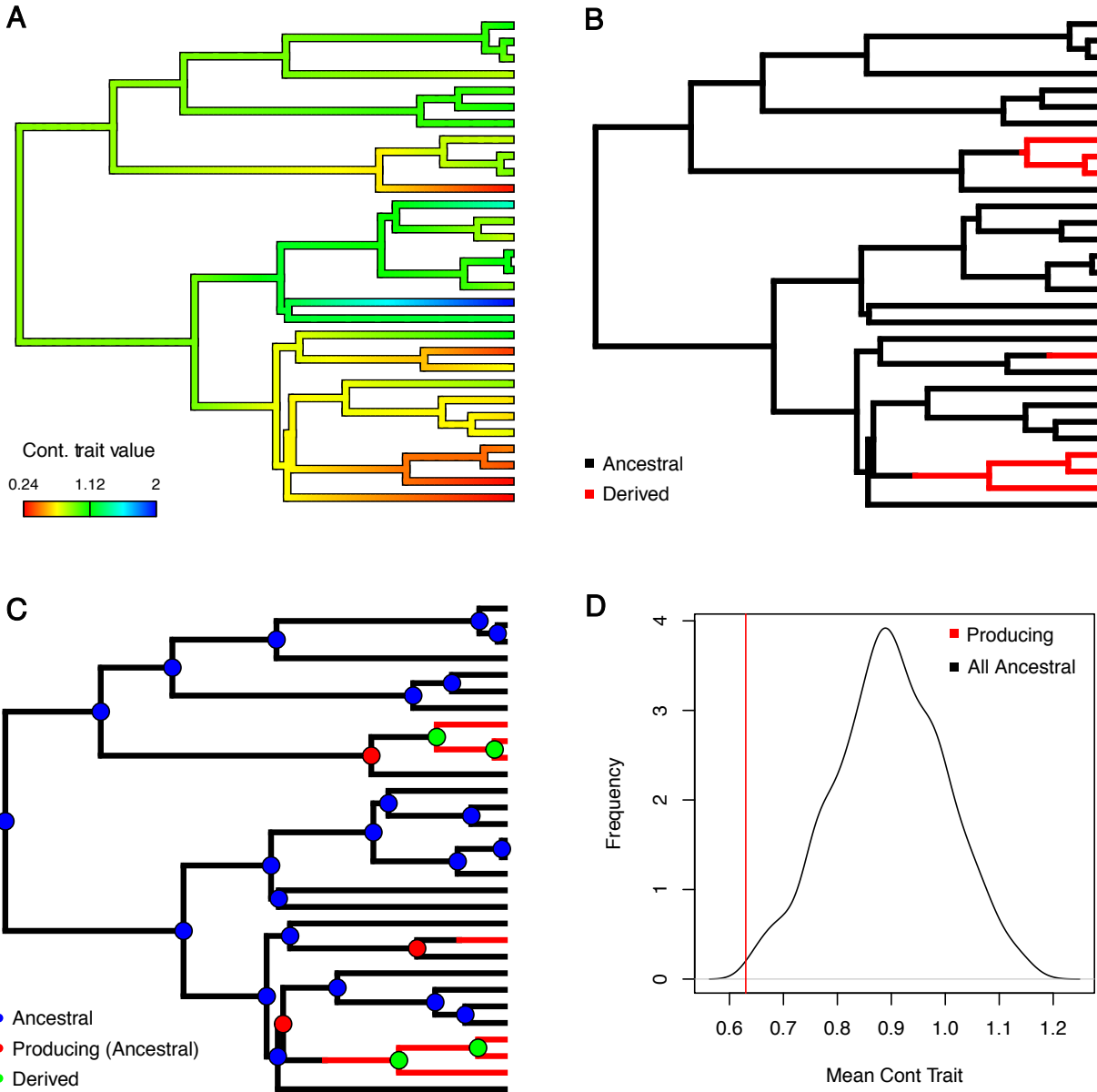422

18

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 discrete 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 discrete 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.
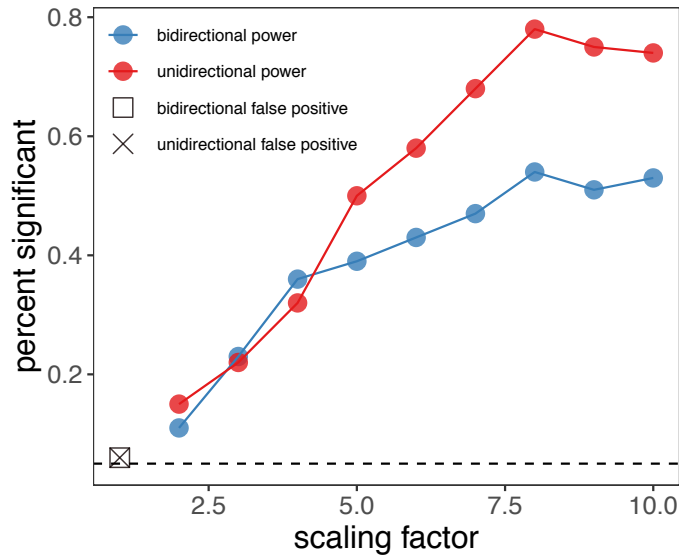
Figure 2. False-positive rate and power of the AncCond test as a function of the branch scaling factor. Scaling factor used to generate discrete traits (where higher values indicate greater contingency between traits) shown on the horizontal axis plotted against the proportion of simulations for which 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%.

20
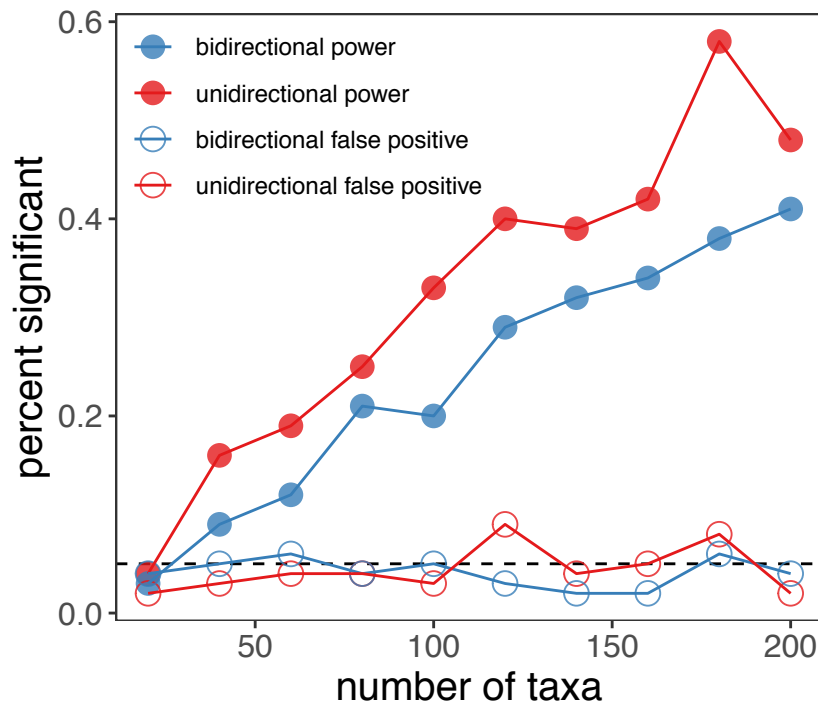
Figure 3. The impact of sample size on test performance. The number of extant taxa included in the analysis plotted on the horizontal axis against the proportion of simulations for which AncCond detected significant contingency plotted on the vertical axis. Lines with open circles indicate the false-positive 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%.

458 Figure 4. Comparison to existing methods. In the left panel false-positive rate for the ancestral
459 condition test is compared to Pagel's method with discretization of the continuous trait and to the
460 threshold method where the discrete trait is assumed to be an expression of an underlying
461 continuous trait. In the right panel the power of the ancestral condition is compared to Pagel's
462 method and to the threshold method.
463
464
465
466
467

468

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. pp.

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**: 129-31.

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.

515 Price, S. A., Hopkins, S. S., Smith, K. K. & Roth, V. L. 2012. Tempo of trophic evolution and its
516   impact on mammalian diversification. *Proceedings of the National Academy of Sciences*
517   **109**: 7008-7012.
518 R Development Core Team (2013) R: A Language and Environment for Statistical Computing. pp.,
519   Vienna, Austria.
520 Revell, L. J. 2012. phytools: an R package for phylogenetic comparative biology (and other
521   things). *Methods in Ecology and Evolution* **3**: 217-223.
522 Revell, L. J. 2014. Ancestral character estimation under the threshold model from quantitative
523   genetics. *Evolution* **68**: 743-759.
524 Ross, L., Hardy, N. B., Okusu, A. & Normark, B. B. 2012. Large population size predicts the
525   distribution of asexuality in scale insects. *Evolution* **67**: 196-206.
526 Slater, G. J., Price, S. A., Santini, F. & Alfaro, M. E. 2010. Diversity versus disparity and the
527   radiation of modern cetaceans. *Proceedings of the Royal Society B: Biological Sciences*
528   **277**: 3097-3104.
529 Steeman, M. E., Hebsgaard, M. B., Fordyce, R. E., Ho, S. Y., Rabosky, D. L., Nielsen, R., Rahbek,
530   C., Glenner, H., Sørensen, M. V. & Willerslev, E. 2009. Radiation of extant cetaceans
531   driven by restructuring of the oceans. *Systematic biology* **58**: 573-585.
532 Team, R. (2015) RStudio: Integrated Development for R. pp. RStudio, Inc., Boston, MA.
533 Wright, S. 1934. An analysis of variability in number of digits in an inbred strain of guinea pigs.
534   *Genetics* **19**: 506.

535