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 Anderson[1], Richard H. Adams[2], Jeffery P. Demuth[3], and Heath Blackmon[4,5]

[1] Department of Math, Texas A&M, College Station, TX 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 the way 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).

Hypotheses that suggest that extreme values of a continuous trait are associated with transitions in a binary trait are common in evolutionary biology (Ross *et al.* 2012; Blackmon & Demuth 2015; Blackmon, Hardy & Ross 2015). In a recent study of sex chromosome evolution, for example, we were interested in testing 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).

2

69   Another approach would be to assume that our binary state (presence or absence of a Y

70   chromosome) could be modeled as if it was produced by an underlying continuous trait that is

71   unobserved. If this assumption is justified, we could employ the threshold model (Wright 1934;

72   Felsenstein 2012), which would allow us to estimate the correlation between chromosome number

73   and the presence or absence of the Y chromosome. A third potential approach would be to use a

74   phylogenetic mixed model that treats chromosome number as a predictor variable and presence

75   or absence of the Y chromosome as a response variable (Hadfield 2010). However, each of these

76   approaches suffers from the same fundamental problem: correlation does not provide information

77   about the direction of causality. For example, does high chromosome number lead to Y

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

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

80   false-positive rates (Maddison & FitzJohn 2014).

81

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

83   significant relationships between the value of a continuous trait and transitions in a binary trait.

84   Briefly, our approach estimates the mean value of the continuous trait when the binary character

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

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

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

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

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

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

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

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

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

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

95   of a continuous trait and the evolution of a binary character.

96

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

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

99   binary character is associated with lower values of the continuous trait, but is this because low

100   values of the continuous trait lead to changes in the binary character, or does the derived state of

101   the binary character lead to reductions in the continuous trait? When this is a concern, we can

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

103   ancestral condition of the binary trait. This approach provides a much more rigorous test than can

104   be implemented with existing methods. In practice, there is not always a clear ancestral and

105   derived state, particularly when a binary character evolves quickly and lacks fossil data. This

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

107   of our approach is reduced.

108

109   **METHODS**

110

111   *The Ancestral Condition Test*

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

113   subtending transitions of a binary character tend to exhibit extreme values of the continuous

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

115   involves the following four steps:

116

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

118   from species that exhibit the derived state of the binary character. By pruning data from

119   taxa exhibiting the derived state of the binary character, we create a more conservative

4

120 test that will be applicable in cases where the derived state of the binary character is

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

122 (addressing the potential for feedback mentioned above). Ancestral state estimation is

123 accomplished under a Brownian motion model using the existing function *anc.ML* in

124 phytools (Figure 1A) (Revell 2012).

125

126 2) Next, we perform ancestral state estimation of the binary character by way of stochastic

127 character mapping (Figure 1B) (Huelsenbeck, Nielsen & Bollback 2003; Bollback 2006).

128 We may choose to fix a single state at the root of the tree in cases where this is

129 appropriate. For instance, when the ancestral state is informed by fossil records or when

130 prior biological knowledge suggests that transitions are expected to occur in only one

131 direction (Blackmon, Hardy & Ross 2015). In the latter case, users may also restrict the

132 simulated evolution of the binary character by allowing only transitions from the ancestral

133 to the derived state. Stochastic mapping is performed using the *make.simmap* function

134 provided in phytools (Figure 1B) (Revell 2012). For simplicity, we will refer to the two states

135 as one and two; state one will denote the ancestral state, and state two is the derived

136 state.

137

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

139 groups: a) Nodes exhibiting state one (i.e., ancestral state) of the binary character and no

140 transitions in the binary character in immediate daughter branches (Figure 1C, blue

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

142 transitions in the binary character in immediate daughter branches (Figure 1C, green

143 nodes), c) Producing nodes exhibiting either state of the binary character but with one of

144 the immediate daughter branches showing a transition into the other state of the binary

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

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

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

148   we calculate the mean value of the continuous trait at the producing nodes in each group.

149   This process is repeated across stochastic mappings incorporating ancestral state

150   estimation uncertainty. The mean of type c (producing nodes) is our observed statistic. We

151   refer to this value as the "estimated ancestral condition" for each of the states of the binary

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

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

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

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

156

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

158   continuous trait estimates from all nodes in the correct binary trait state (i.e. sampling

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

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

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

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

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

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

165   1D).

166

167   **Simulated data**

168

169   *Scenario one–unidirectional change in a binary character*

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

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

6

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

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

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

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

176 continuous trait for each node in the phylogeny.

177

178 To simulate a binary trait whose evolution was contingent on the continuous trait, we used a

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

180 continuous trait in the upper or lower quartiles. Branches whose mean was in the lower quartile

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

182 multiplicative factor of x. We repeated this process for ten values ($x = 1 - 10$) on each of our 100

183 trees. We then simulated the evolution of the binary character on these scaled trees. We set the

184 root to state one (ancestral state) and allowed the trait to evolve under an Mk2 model where

185 transitions to state two (derived state) were allowed, but not transitions from state two to state one

186 (i.e., modeling unidirectional evolution), as we expected our approach to be most powerful under

187 this mode of evolution. This simulation was done using the *sim.char* function in Geiger (Harmon *et*

188 *al.* 2008). Various transition rates from state one to state two were evaluated, and we found that a

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

190 dataset did not contain at least ten extant taxa exhibiting each state of the binary character, it was

191 discarded (between 7% and 55% of simulations were discarded for a given scaling factor, with a

192 higher percent being discarded for smaller scaling factors). Thus, our total of 1000 simulated

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

194

195 By simulating the evolution of the binary trait on these scaled trees, we generated a contingency

196 between the binary and continuous traits. There exists more opportunity for transitions on the

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

7

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

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

200   scaling technique effectively altered the rate of evolution of the binary trait wherever the

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

202   continuous trait had high values and decreasing the density of transitions for branches with low

203   values.

204

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

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

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

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

209   discarded, and a new dataset was generated (between 14% and 91% of simulations were

210   discarded for different sized trees, with small trees leading to a higher proportion of discarded

211   simulations). We then simulated binary datasets with scaling factors $x = 1$ or $x = 5$ as described

212   above, allowing us to measure both power ($x = 5$) and false-positive rate ($x = 1$; i.e. how often we

213   detect a relationship under conditions where the data were simulated with no contingency

214   between traits). A total of 200 data sets (scaling factors of one and five) were generated for each

215   of the ten tree sizes for a total of 2000 simulated datasets.

216

217   *Scenario two – bidirectional change in a binary character*

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

219   for which a binary character has transitions in both directions (i.e., bidirectional), and when we are

220   unable to confidently assign an ancestral - derived relationship. We repeated the above procedure

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

222   allowed the root to take either state with equal probability. Second, we allowed for transitions in

223   either direction, from state one to state two and from state two to state one, with equal rates. A

8

224   transition rate of 0.3 was used to ensure there were several forward and reverse transitions in

225   every dataset. This process generated an additional 3000 simulated datasets.

226

227   *Scenario three – empirical phylogeny*

228   To test whether our approach is impacted when more complex processes generate trees and

229   when continuous traits have variable rates of evolution, we used a phylogeny of living cetaceans

230   (Steeman *et al.* 2009) and body size data for 74 species (Slater *et al.* 2010). We then simulated

231   the unidirectional evolution of a binary character, from a known ancestral state, whose evolution is

232   entirely independent of the continuous character using the sim.char function from Geiger (Harmon

233   *et al.* 2008). This process generated 100 simulated datsets.

234

235   *Assessing the performance of comparative tests*

236

237   We analyzed these datasets using our AncCond function implemented in the R package EvobiR

238   (Blackmon H 2015). We supplied this R function with the unaltered phylogenetic trees and the

239   corresponding tip data of the binary and continuous traits. For the case of unidirectional evolution

240   (i.e., scenario one), we set the rate of evolution from state two to state one to zero and fixed the

241   root of the tree to state one. We also chose to prune continuous data from taxa exhibiting the

242   derived state of the binary character. This pruning is not entirely applicable to this simulated data,

243   as the continuous character was simulated completely independent of the binary character.

244   However, we chose to use this option in our analysis because, in practice, it is not always clear

245   when the continuous character's evolution is independent of the binary character's evolution, and

246   because it is important to test our method under the most restrictive conditions. When assessing

247   data generated under scenario two, we followed the same approach but allowed two separate

248   rates of transitions in the binary trait (transitions from one to two and two to one), made no

9

249   assumptions on the original state at the root of the tree, and pruned no continuous data. In

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

251

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

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

254   one and five). First, we applied the threshold model as implemented in phytools (Revell 2012).

255   This approach allowed us to test for a correlation between our continuous character and the

256   binary character by inferring an underlying continuous value that we observe as a discretely

257   expressed character. Second, we discretized our continuous character at the median value and

258   applied Pagel's test for correlation of binary characters also implemented in phytools (Revell

259   2012).

260

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

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

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

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

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

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

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

268   RAM and a 3.2 GHz processor (R Development Core Team 2013; Team 2015). All code used in

269   simulation and testing of the ancestral condition test are available online via GitHub

270   (https://github.com/coleoguy/anc.cond.analyses). We have implemented the ancestral condition

271   test in the R package EvobiR which is publicly available and can be installed directly from GitHub

272   (https://github.com/coleoguy/evobir). This package includes a variety of other useful functions for

273   evolutionary biology (table 1).

274

10

275

RESULTS

276

277

278   Scenario one - Unidirectional change in the binary character

279

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

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

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

283   approach for increasing degrees of contingency between the binary and continuous characters.

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

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

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

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

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

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

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

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

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

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

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

295   3).

296

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

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

299   significant relationship reliably. Figure 4 shows that when our simulated phylogenetic trees

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

11

301 (Figure 4, red line filled circles), a reduction in power of approximately 93% relative to the

302 maximum power found for a scaling factor of 5. The maximum power we measured was for the

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

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

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

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

307 (Figure 4, red line open circles).

308

309 Scenario two - Bidirectional change in the binary character

310

311 To test the performance of our method when the evolutionary history of the binary trait is more

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

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

314 performance with a scaling factor of one to measure false-positive rate, and scaling factors from

315 2-10 to measure the power of our method under increasingly strong relationships between the

316 continuous and binary characters. We measured a false-positive rate of 6% (Figure 2, black

317 square). Power for this scenario ranged from 11% for a scaling factor of two to 54% for a scaling

318 factor of eight (Figure 2 blue line). These power estimates were calculated with a level of

319 significance alpha = 0.025 to correct for testing for contingency in two directions. We confirmed

320 the ancestral condition test requires a large sample size in the bidirectional case as well. For small

321 phylogenies, the ancestral condition test has power in the single digits and only ever reaches a

322 power higher than 40% when there are 200 taxa present (Figure 4, blue line filled circles), the

323 highest value tested. Regardless of the number of taxa included in the analysis, the false-positive

324 rate of our approach was approximately 5% (Figure 4, blue line open circles).

325

326 Scenario 3 - Empirical Phylogeny

12

327

328 Our approach exhibited similar performance and when applied to an empirical phylogeny. Of the

329 100 simulated datasets, whose evolution is independent of body size, the Ancestral Condition test

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

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

332 considerably different than the false-positive rate measured using simulated phylogenies,

333 indicating that the ancestral condition test performs similarly with simulated and empirical

334 phylogenies.

335

336 **DISCUSSION**

337

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

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

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

341 origin of a derived state of a binary trait and the value of a continuous trait (scaling factor of one),

342 the test will incorrectly identify a significant relationship ~6% of the time under our simulated

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

344 the continuous trait effectively have a transition rate into the derived state of the binary trait that is

345 five times higher than branches with intermediate values. Likewise branches with a low value of

346 the continuous trait effectively have a transition rate into the derived state of the binary trait that is

347 five times lower than branches with intermediate values. This is arguably a very strong

348 relationship, but even in this case, we detect the relationship in only ~50% of simulated datasets.

349

350 The ancestral condition test becomes less powerful when the evolution of the binary trait is

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

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

13

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

354   indicate that the test can identify contingency in the evolution of traits with a high degree of

355   certainty. However, the power of the ancestral condition test to detect even strong contingency in

356   bidirectional evolution is more limited. In other words, under the conditions we have simulated,

357   whenever the ancestral condition test detects contingency, there is a 94% chance it is correct, but

358   when it does not detect contingency, then it is correct three out of four times, at best.

359

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

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

362   ever-stronger contingency. In retrospect this is not surprising. at the very highest contingency

363   levels, every branch in the tree that is classified into the upper or lower quartiles is scaled to such

364   an extent that those in the lower quartiles almost never exhibit changes in the binary character

365   and branches in the upper quartile almost always exhibit transitions in the binary character. Based

366   on our tree sizes and rates of evolution, we can infer this occurred when our scaling factor

367   reached a value of eight. This plateau is clear in the right-hand side of Figure 2, where power

368   plateaus at a scaling factor of eight for both unidirectional and bidirectional evolution scenarios.

369   This point at which power plateaus should be a function of tree size where larger trees plateau at

370   higher levels of contingency.

371

372   While current approaches for jointly analyzing continuous and binary traits can be effective for

373   identifying correlations between traits in some scenarios, our AncCond test provides an important

374   and useful extension to current comparative methods as it explicitly provides information about the

375   order in which traits evolved across a phylogenetic tree. Furthermore, our approach is robust to

376   possible confounding effects that one of the states of binary trait might exert on the evolution of

377   the continuous trait. For instance, if we believe that state two of a binary trait may lead to selection

378   for lower values in a corresponding continuous trait, we can prune the continuous data for all

14

379   species in state two. By doing this we effectively ask if species still in state one yield an ancestral

380   state estimation for the continuous trait that suggests that state two of the binary character is

381   arising in parts of the phylogeny that have lower values than expected if the traits are truly

382   independent. This approach removes a typical caveat from the interpretation of our results and

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

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

385   with transitions in the binary character. However, the false-positive rate remains acceptable even

386   for the smallest phylogenies tested.

387

388   Our approach assumes an Mk2 model for the evolution of the binary trait and a Brownian motion

389   model for the evolution of the continuous trait. The adequacy of these underlying models should

390   be evaluated before using the ancestral condition test. Model adequacy of the continuous trait can

391   be accomplished in several ways (Garland, Harvey & Ives 1992; Pennell *et al.* 2015). Perhaps the

392   most robust of these methods involves first calculating test statistics on the phylogenetic

393   independent contrasts from the observed data. Simulations are then performed, and these same

394   test statistics are calculated for each simulation - generating null distributions. The benefit of this

395   approach is that deviation of observed test statistics can determine if a model is adequate while

396   providing information about what type of alternative model might perform better for a given dataset

397   (Pennell *et al.* 2015). Evaluation of model adequacy for binary traits is currently less developed.

398   One potential option is to use a Monte Carlo approach to determine if the model and parameter

399   estimates are able to generate data with similar proportions of taxa in each discrete state (Price *et*

400   *al.* 2012; Blackmon & Demuth 2014). An additional approach is to examine the number of state

401   changes expected under parsimony for the observed data and compare this to the number of

402   state changes expected under parsimony for simulated datasets (Beaulieu, O'Meara & Donoghue

403   2013). One advantage of the AncCond test is that while our implementation assumes a simple

404   Brownian model for the continuous trait and an Mk2 model for the discrete trait, the same

15

405 approach could be used with alternative ancestral state estimates generated under models found

406 to be more appropriate for a practitioner's particular dataset. Although we only implement the

407 Brownian motion and Mk models in the ancestral condition test the process described in figure 1

408 could be repeated substituting a different model during the ancestral state reconstructions step A

409 and B.

410

411 Often users may not have a single best tree, but rather a sample of a posterior distribution of

412 trees. In these cases, the ancestral condition test can be run independently on each tree in the

413 posterior. Users could then report the proportion of trees that support a significant relationship.

414 Alternatively, our function returns the full null distribution generated for a tree, and these null

415 distributions could be combined to create a null that incorporates phylogenetic uncertainty.

416

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

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

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

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

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

422 probability of transitions to the derived state of the binary state. Such a pattern could conceivably

423 produce a significant result despite having only a single origin for the high continuous trait value.

424 We believe the best approach to avoiding the errors associated with pseudoreplication is thorough

425 data exploration before applying nearly any comparative method, including our AncCond test.

426 There are many tools now available that allow simultaneous visualization of trees and data that

427 can be useful in the early stages of analysis to alert researchers of possible issues. An alternative

428 approach to solving the pseudoreplication problem is to perform what is effectively a meta-

429 analysis. With this approach, the original dataset is divided into several independent data sets and

430 the results of all of the independent tests can be evaluated (Mayrose *et al.* 2011).

431

16

432

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.

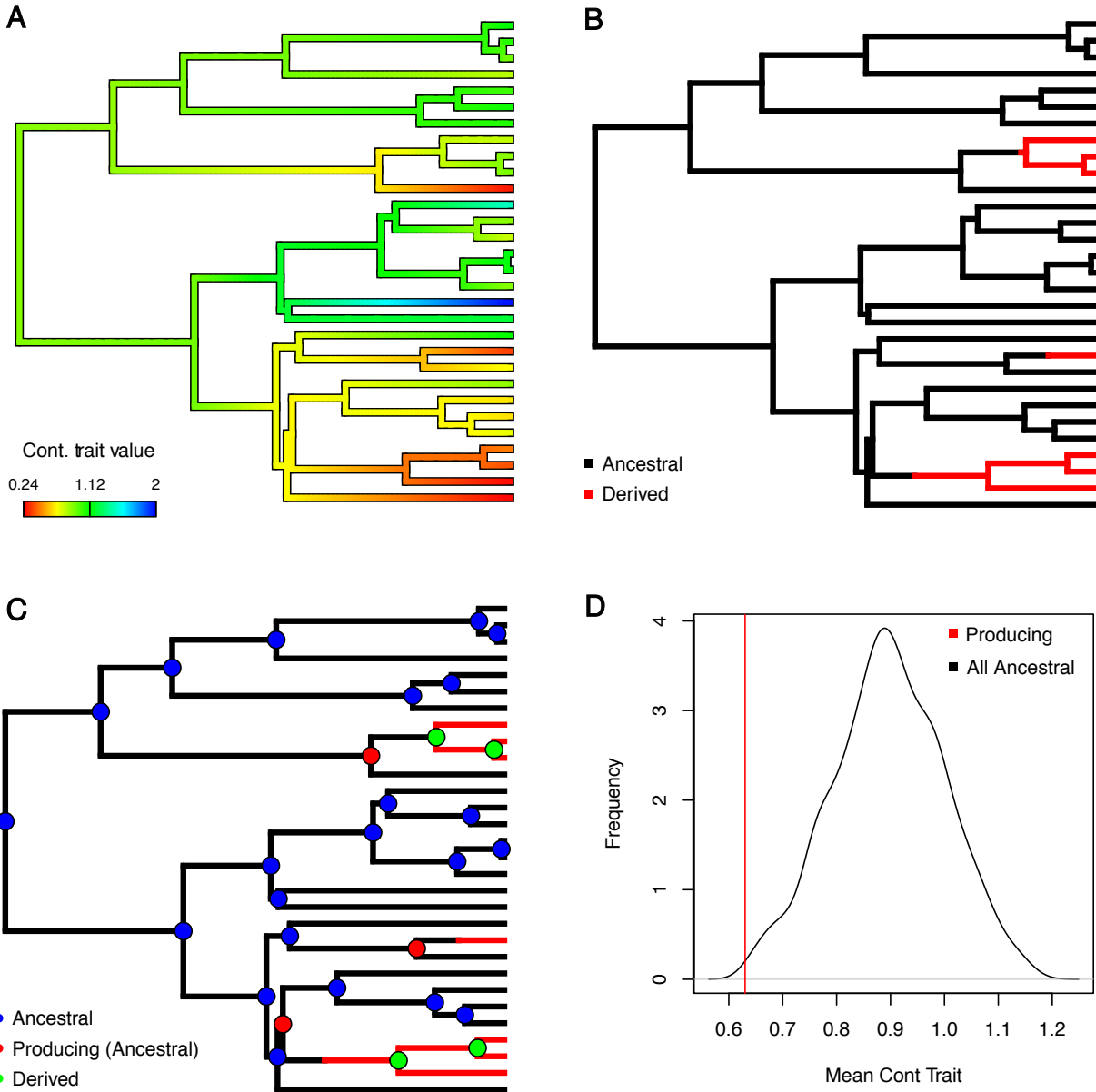433   Table 1 Summary of functions available in EvobiR.
434

17

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.
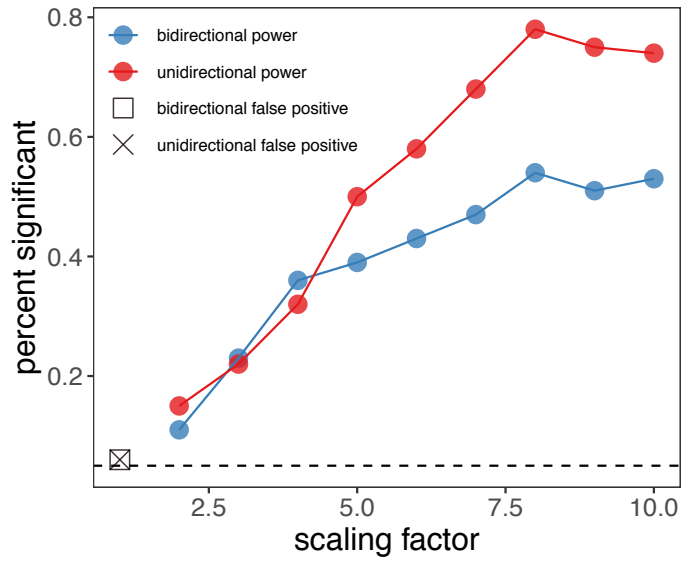
Figure 2. False-positive rate and power of the AncCond test as a function of the branch scaling factor. Scaling factor used to generate binary traits (where higher values indicate greater contingency between traits) shown on the horizontal axis plotted against the proportion of simulations for which 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%.
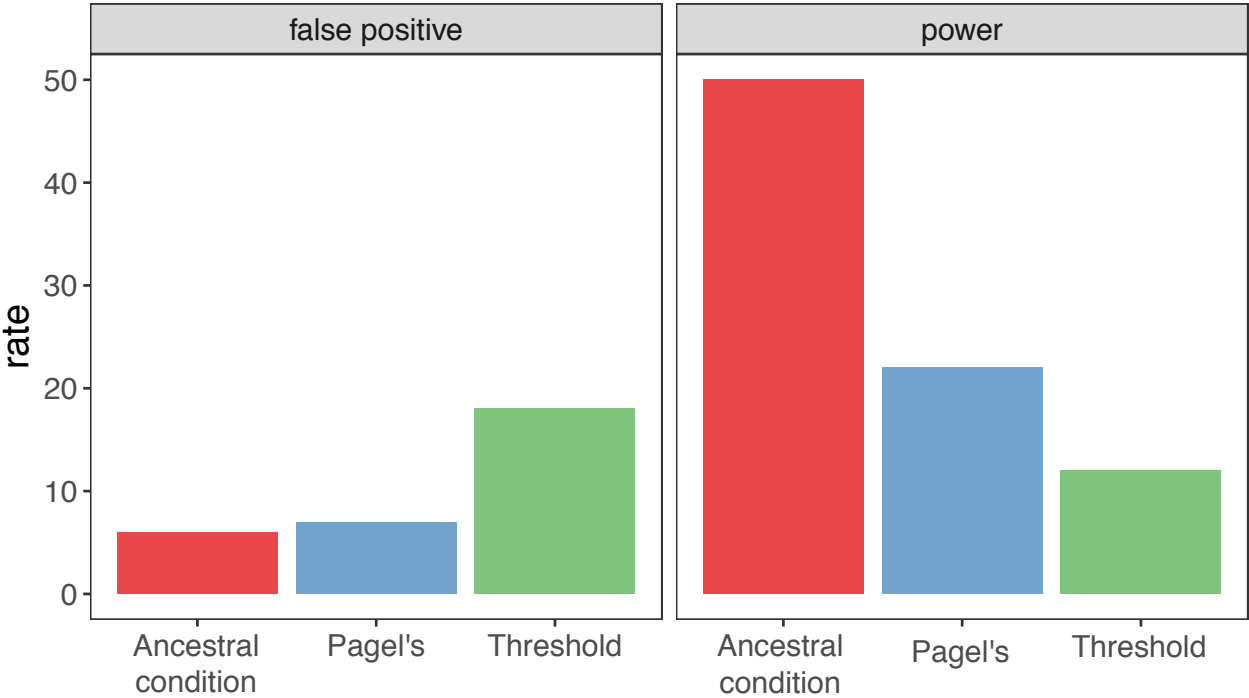
19

457
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
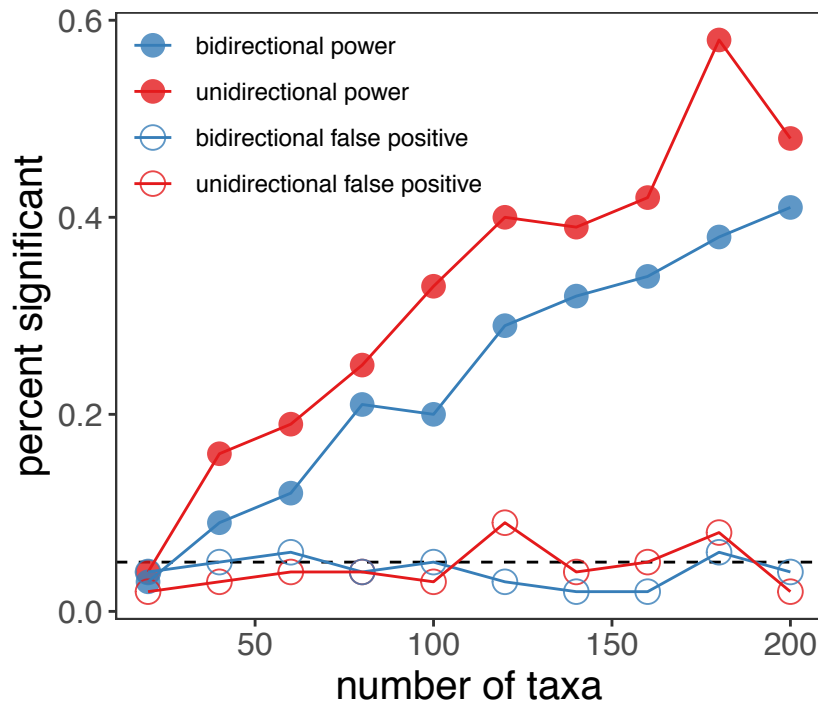method and to the threshold method.

467
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%.

477
478 Beaulieu, J.M., O'Meara, B.C. & Donoghue, M.J. (2013) Identifying hidden rate changes in the
479   evolution of a binary morphological character: the evolution of plant habit in campanulid
480   angiosperms. *Systematic Biology,* **62,** 725-737.
481 Blackmon, H. & Demuth, J.P. (2014) Estimating tempo and mode of Y chromosome turnover:
482   explaining Y chromosome loss with the fragile Y hypothesis. *Genetics,* **197,** 561-572.
483 Blackmon, H. & Demuth, J.P. (2015) The fragile Y hypothesis: Y chromosome aneuploidy as a
484   selective pressure in sex chromosome and meiotic mechanism evolution. *Bioessays,* **37,**
485   942-950.
486 Blackmon, H., Hardy, N.B. & Ross, L. (2015) The evolutionary dynamics of haplodiploidy:
487   Genome architecture and haploid viability. *Evolution,* **69,** 2971-2978.
488 Blackmon H, R.A.A. (2015) EvobiR: tools for comparative analyses and teaching evolutionary
489   biology.
490 Bollback, J.P. (2006) SIMMAP: stochastic character mapping of discrete traits on phylogenies.
491   *BMC Bioinformatics,* **7,** 88.
492 Butler, M.A. & King, A.A. (2004) Phylogenetic comparative analysis: a modeling approach for
493   adaptive evolution. *The American Naturalist,* **164,** 683-695.
494 Felsenstein, J. (1985) Phylogenies and the comparative method. *American Naturalist,* 1-15.
495 Felsenstein, J. (2004) *Inferring phylogenies.* Sinauer Associates, Sunderland, MA.
496 Felsenstein, J. (2012) A comparative method for both discrete and continuous characters using
497   the threshold model. *The American Naturalist,* **179,** 145-156.
498 FitzJohn, R.G. (2012) Diversitree: comparative phylogenetic analyses of diversification in R.
499   *Methods in Ecology and Evolution,* **3,** 1084-1092.
500 Garland, T., Harvey, P.H. & Ives, A.R. (1992) Procedures for the analysis of comparative data
501   using phylogenetically independent contrasts. *Systematic Biology,* **41,** 18-32.
502 Hadfield, J.D. (2010) MCMC methods for multi-response generalized linear mixed models: the
503   MCMCglmm R package. *Journal of Statistical Software,* **33,** 1-22.
504 Harmon, L.J., Weir, J.T., Brock, C.D., Glor, R.E. & Challenger, W. (2008) GEIGER: investigating
505   evolutionary radiations. *Bioinformatics,* **24,** 129-131.
506 Huelsenbeck, J.P., Nielsen, R. & Bollback, J.P. (2003) Stochastic mapping of morphological
507   characters. *Systematic Biology,* **52,** 131-158.
508 Landis, M.J., Schraiber, J.G. & Liang, M. (2012) Phylogenetic analysis using Lévy processes:
509   finding jumps in the evolution of continuous traits. *Systematic Biology,* **62,** 193-204.
510 Lewis, P.O. (2001) A likelihood approach to estimating phylogeny from discrete morphological
511   character data. *Systematic Biology,* **50,** 913-925.
512 Maddison, W.P. & FitzJohn, R.G. (2014) The unsolved challenge to phylogenetic correlation tests
513   for categorical characters. *Systematic Biology,* syu070.
514 Mayrose, I., Zhan, S.H., Rothfels, C.J., Magnuson-Ford, K., Barker, M.S., Rieseberg, L.H. & Otto,
515   S.P. (2011) Recently formed polyploid plants diversify at lower rates. *Science,* **333,** 1257-
516   1257.
517 O'Meara, B.C., Ane, C., Sanderson, M.J. & Wainwright, P.C. (2006) Testing for different rates of
518   continuous trait evolution using likelihood. *Evolution,* **60,** 922-933.
519 Pagel, M. (1994) Detecting correlated evolution on phylogenies: a general method for the
520   comparative analysis of discrete characters. *Proceedings of the Royal Society of London
521   B: Biological Sciences,* **255,** 37-45.
522 Pennell, M.W., FitzJohn, R.G., Cornwell, W.K. & Harmon, L.J. (2015) Model adequacy and the
523   macroevolution of angiosperm functional traits. *The American Naturalist,* **186,** E33-E50.
524 Price, S.A., Hopkins, S.S., Smith, K.K. & Roth, V.L. (2012) Tempo of trophic evolution and its
525   impact on mammalian diversification. *Proceedings of the National Academy of Sciences,*
526   **109,** 7008-7012.
527 R Development Core Team (2013) R: A Language and Environment for Statistical Computing.
528   Vienna, Austria.

22

529   Revell, L.J. (2012) phytools: an R package for phylogenetic comparative biology (and other
530   things). *Methods in Ecology and Evolution,* **3,** 217-223.
531   Revell, L.J. (2014) Ancestral character estimation under the threshold model from quantitative
532   genetics. *Evolution,* **68,** 743-759.
533   Ross, L., Hardy, N.B., Okusu, A. & Normark, B.B. (2012) Large population size predicts the
534   distribution of asexuality in scale insects. *Evolution,* **67,** 196-206.
535   Slater, G.J., Price, S.A., Santini, F. & Alfaro, M.E. (2010) Diversity versus disparity and the
536   radiation of modern cetaceans. *Proceedings of the Royal Society B: Biological Sciences,*
537   **277,** 3097-3104.
538   Steeman, M.E., Hebsgaard, M.B., Fordyce, R.E., Ho, S.Y., Rabosky, D.L., Nielsen, R., Rahbek,
539   C., Glenner, H., Sørensen, M.V. & Willerslev, E. (2009) Radiation of extant cetaceans
540   driven by restructuring of the oceans. *Systematic Biology,* **58,** 573-585.
541   Team, R. (2015) RStudio: Integrated Development for R. RStudio, Inc., Boston, MA.
542   Uusitalo, L. (2007) Advantages and challenges of Bayesian networks in environmental modelling.
543   *Ecological modelling,* **203,** 312-318.
544   Wright, S. (1934) An analysis of variability in number of digits in an inbred strain of guinea pigs.
545   *Genetics,* **19,** 506.

546