1   *Abstract*

2   Whether modeling the evolution of a discrete or continuous character, the focal trait of interest

3   does not evolve in isolation and require comparative methods that model multivariate evolution.

4   Progress along these lines has involved modeling multivariate evolution of the same class of

5   character and there are fewer options when jointly modeling traits when one character is discrete

6   and the other is continuous. Here we develop such a framework to explicitly estimate the joint

7   likelihood for discrete and continuous characters. Specifically, our model combines the

8   probability of observing the continuous character under a generalized OU process with the

9   probability of the discrete character under a hidden Markov model, linked by a shared underlying

10   regime. We use simulation studies to demonstrate that this approach, *hOUwie*, can accurately

11   evaluate parameter values across a broad set of models. We then apply our model to test whether

12   fleshy and dry fruits of Ericaceae lineages are correlated with their climatic niche evolution as

13   represented by the aridity index. Consistent with expectations, we find that dry fruits have higher

14   rates of climatic niche evolution, that the climatic niche of fleshy fruits is more conserved, and

15   dry fruits have a more humid climatic optimum.

16

17   A common theme in comparative biology is the detection of causal, or least mechanistic,

18   factors that affect the evolution of quantitative characters. Questions of how plant life habit

19   influence genome size evolution (Beaulieu et al. 2012), how substrate use alters limb length

20   evolution (Mahler et al. 2013), or how tooth morphology slowly changes in response to habitat

21   and diet (Toljagić et al. 2018) are all examples of testing whether evolutionary changes in a

22   discrete variable may have altered evolutionary trajectories of a continuously varying trait. One

23   very common phylogenetic comparative approach for these types of questions is to employ an

24   Ornstein-Uhlenbeck (OU) model, which assumes distinct regimes, described by the evolution of

25   a discrete character, are known completely *a priori* (e.g., Butler and King 2004; Hansen et al.

26   2008; Beaulieu et al. 2012), or assumes that "shifts" in regimes can be inferred directly from the

27   distribution of the continuous trait (e.g. Ingram and Mahler 2013; Uyeda and Harmon 2014;

28   Khabbazian et al. 2016). While these approaches are practical, the discrete trait is assumed the

29   driving force underlying the evolution of the continuous character. However, dependence rarely

30   flows just one way in evolution, and we suspect that as often as a discrete character causes

31   change in the continuous character, continuous characters also influence discrete character

32   evolution, or at the very minimum, can provide information about how they may be evolving in

33   tandem.

34   Progress along these lines has mostly involved acknowledging uncertainty in the

35   evolution of the discrete character by fitting models over a large set of stochastically generated

36   character mappings. That is, a large set of alternative reconstructions of the discrete character are

37   obtained completely uninformed by the continuous trait's evolution, then the likelihood of the

38   continuous character becomes the average of the likelihoods across these maps (e.g., Revell

39   2012). The advantage of this approach is that there is an explicit model for how regimes change

40   through time, but the evolution of these regimes remains entirely independent of the continuous

41   trait, and the probability of these regimes is not explicitly considered. For example, it is possible

42   that the model that best fits the discrete data generates stochastic maps that does not provide a

43   good fit to the continuous data.

44   A promising approach was recently described for detecting adaptive codon evolution

45   (Jones et al 2020), where a set of maps obtained for a discrete phenotype under a standard

46   Markov process is optimized along with parameters associated with genotype properties, thus

47   forcing an emergent dependency between the two. Similarly, May and Moore (2020) developed

48   a joint model for discrete and continuous characters under a state-dependent Brownian motion

49   model. Their approach takes advantage of prior probabilities within a Bayesian framework to

50   accommodate variation in the "background" rate of evolution in the continuous trait (i.e., rate

51   variation across lineages that is independent of the discrete character under consideration). The

52   novel Bayesian pipeline recently developed by Tribble et al. (2021) is the first attempt that we

53   are aware of for jointly modeling discrete and continuous traits under an OU framework. Their

54   approach samples discrete stochastic mappings informed by the discrete trait along with regime

55   mappings which were informed by the continuous trait while accounting for the potential of

56   hidden variation. While a more effective test of correlation between discrete and continuous

57   characters, one drawback is that they do not explicitly account for the joint probability of the

58   discrete and continuous parameter estimates together. They assume that the combination of

59   independently estimated discrete and continuous models produces a joint estimate.

60   Here we develop and implement a framework that provides an explicitly joint estimate of

61   the likelihood for a discrete and continuous character. Specifically, our model combines the

62   probability of the continuous character given a particular regime evolving under a generalized

63   OU process, and the probability of that discrete regime painting obtained from an expanded set

64   of Markov models, integrated over many regime paintings. We demonstrate how our framework,

65   which we call *hOUwie*, can be used to test hypotheses of correlated evolution between discrete

66   and continuous characters while also accounting for hidden character states and unobserved

67   variation. Finally, we apply several *hOUwie* models to test the correlated dynamics of the mode

68   of seed dispersal and climatic niche evolution and compare our results to those that did not

69   account for the potential joint evolution of discrete and continuous variables.

70

71   **Materials and Methods**

72   *The hOUwie model*

73   Our model is composed of two processes: one describing the evolution of a discrete

74   character and the other describing the evolution of a continuous character. To model the

75   evolution of a single continuous character we use an Ornstein-Uhlenbeck (OU) model (Hansen

76   1997; Butler and King 2004; Hansen et al. 2008; Beaulieu et al. 2012; Ho and Ané 2014a).

77   Formally, the OU process is an Itô diffusion satisfying:

78                    $$dX(t) = \alpha(\theta(t) - X(t)) + \sigma dB(t).$$

79   Conceptually, this model combines the stochastic evolution of a trait through time with a

80   deterministic component that models the tendency for a trait to evolve towards an "optimum." In

81   this model, the value of a trait, $X(t)$, is pulled towards an optimum, $\theta(t)$, at a rate scaled by the

82   parameter α. The optimum, $\theta(t)$, is a piecewise constant on intervals and takes values in a finite

83   set $\{\theta_i\}$. This can represent the set of "selective regimes", "regimes", or Simpson's "adaptive

84   zones" (Cressler et al. 2015), though it is consistent with a variety of true underlying

85   microevolutionary models (Hansen 2014). Additionally, random deviations are introduced by

86   Gaussian white noise $dB(t)$, which is distributed as a normal random variable with mean zero

87   and variance equal to $\sigma^2 dt$. Thus, σ^2 is a constant describing the rate of stochastic evolution

88   away from the optimum. We use the set of extensions introduced by Beaulieu et al. (2012) and

89   implemented in the R package *OUwie*, which allows for multiple primary optima $\theta(t)$ in which

90   both the pull strength (α) and the rate of stochastic evolution (σ^2) can vary across the phylogeny.

91   However, the algorithm used to calculate the likelihood described in Beaulieu et al. (2012)

92   involves a computationally costly matrix inversion procedure. Here we implement a linear-time

93   computation of the likelihood of Gaussian trait models following (Ho and Ané 2014a). To do

94   this, we first transform the phylogeny such that its variance covariance matrix, V, is 3-point

95   structured. We can write the variance covariance matrix of the untransformed phylogeny as $V =$

96      $D_u \tilde{V} D_u$, where following Beaulieu et al. (2012) and Ho and Ané (2014),

$$\tilde{V}_{ij} = \sum_{\gamma=1}^{\kappa(i,j)} \frac{\sigma^2_{ij,\gamma}}{2\alpha_{ij,\gamma}} (e^{2\alpha_{ij,\gamma}s_{ij,\gamma}} - e^{2\alpha_{ij,\gamma}s_{ij,\gamma-1}}),$$

97

$$\text{and, } D_u = e^{\sum_{\gamma=1}^{\kappa(i)} \alpha_{i,\gamma}(s_{i,\gamma}-s_{i,\gamma-1})},$$

98

99   where, s_γ is the distance from the root to the beginning of the selective regime (γ) for the κ

100   number of selective regimes along the path from the root to the last common ancestor of i and j,

101    $\kappa(i,j)$, or from the root to the terminal tip i, $\kappa(i)$. Our transformed phylogeny now has a variance

102   covariance matrix $\tilde{V}_{ij}$ and diagonal matrix D_u. We can then calculate the quadratic quantities and

103   determinant of V (Ho and Ané 2014a). The probability of our continuous trait is given by

$$\log(P(X|D,z,\vartheta,\psi)) = n\log(2\pi) + \log(det(V)) + \frac{P'V^{-1}P - 2P'V^{-1}Q + Q'V^{-1}Q}{2},$$

104

105   where n is the number of tips in the phylogeny (ψ), P is the continuous trait value of each

106   species, and Q is the expected value of each species given the continuous trait model calculated

107   following equation (11) of Beaulieu et al. (2012), D is the discrete character data, z is a particular

108   regime mapping, and ϑ are the parameters of the *hOUwie* model.

109   Next, we describe the calculation of the probability of the underlying regime structure, γ,

110   that is the joint probability of discrete characters (D) and stochastic mapping (z). This calculation

111   is analogous to the pathway likelihood of Steel and Penny (2000). To calculate the probability of

112   discrete characters (D) and stochastic mapping (z) we instead use an approximation. Our

113   approximation relies on a finite number of degree-2 internodes and uses the standard Chapman-

114   Kolmgorov equation to calculate the probabilities of beginning in a particular state i and ending

115   in state j (Pagel 1994) and is identical to a joint probability of a set of state reconstructions

116   (Yang 2006). As the number of internodes increase, the amount of time between nodes decreases

117   and the approximation improves (Rao and Teh 2013). The joint probability of a regime structure

118   and the discrete character i

119
$$P(D, z | Q, \psi) \;=\; P(x_0 | Q, \psi) \prod_{\ell=1}^{n-1} P(z_\ell | Q, T_\ell),$$

120   where $\mathbf{Q}$ the instantaneous rate matrix ($\mathbf{Q} \in \vartheta$), ψ is the phylogeny, $P(x_0 | \mathbf{Q}, \psi)$ is the root state

121   probability (Pagel 1994; Yang 2006; Maddison et al. 2007), n is the number of external nodes

122   (tips), internal nodes, and internodes (degree-2 nodes) summed, ℓ indicates a particular branch,

123    $P(z_\ell | \mathbf{Q}, T_\ell) = e^{\mathbf{Q} T_\ell} \mathbb{1}_\gamma$, where $\mathbb{1}_\gamma$ is an indicator function which ensures that we only use the

124   probability of states indicated by the specific the regime mapping instead of summing over all

125   possible state combinations. The continuous character probability requires the discrete state(s) to

126   be defined along the entire branch, thus we place transitions halfway between any two nodes.

127   For each set of parameters evaluated during the maximum likelihood search, a set of

128   possible mappings of discrete states and continuous regimes are generated to evaluate the

129   discrete and continuous likelihoods. Ideally, we would calculate the likelihood by summing

130   across all possible reconstructions (note that we want the sum across the reconstructions, not the

131   single reconstruction with highest likelihood). The number of such reconstructions is very large,

132                                $n_{\text{states}}^{(2*\text{number of taxa-2})(1+\text{number of degree two internodes per edge})}$,

133   which is particularly daunting as the sum must be calculated anew for every unique examined set

134   of parameter values as part of search. We found in early work where we did look at this

135   exhaustively that a few mappings made up the vast majority of the total likelihood, so we set up

136   the analysis to focus on calculating total likelihood given the highest probability mappings.

137   To do this, we first approximate the conditional state probabilities at nodes. The

138   conditional state probability, unlike the more common marginal reconstruction or joint state

139   reconstruction (Pupko et al. 2000; Felsenstein 2004; Yang 2006), calculates the probability that a

140   node has a particular state value conditioned only on the observations of its descendants. For a

141   particular focal node, we calculate the probability of the observing all pairwise descendant values

142   given the OU model parameters, integrated over all possible rootward node states, and observed

143   tipward discrete states (Fig. 1). Although this is only an approximation of the conditional state

144   probabilities, it proves to be an essential improvement over the typical procedure of sampling

145   many stochastic maps based solely on the discrete process. Next, the conditional probabilities of

146   states at nodes are sampled starting with the root. Once the root is sampled, descendent states are

147   sampled based on both the conditional ancestral values and the sampled ancestral state. This is

148   achieved by multiplying the conditional probability of the node states by the probability of

149   starting in the sampled rootward ancestral value and ending in any of the tipward states (the latter

150   is calculated using familiar matrix exponentiation methods; e.g., Pagel 1994). Finally, under

151   usual stochastic mapping procedures we would use rejection sampling (Nielsen 2002; Rao and

152 Teh 2013) to simulate a path between the sampled rootward and tipward nodes. However, for

153 increased computational efficiency, we opt to place transitions at pre-defined internodes. After

154 nodes and internodes are sampled in step two, mappings are evaluated to ensure consistency with

155 the discrete model (i.e., impossible transitions do not occur) and branches are painted based on

156 the sampled nodes with transitions occurring half-way between nodes (and remember that a

157 single edge may have multiple internodes placed on it).
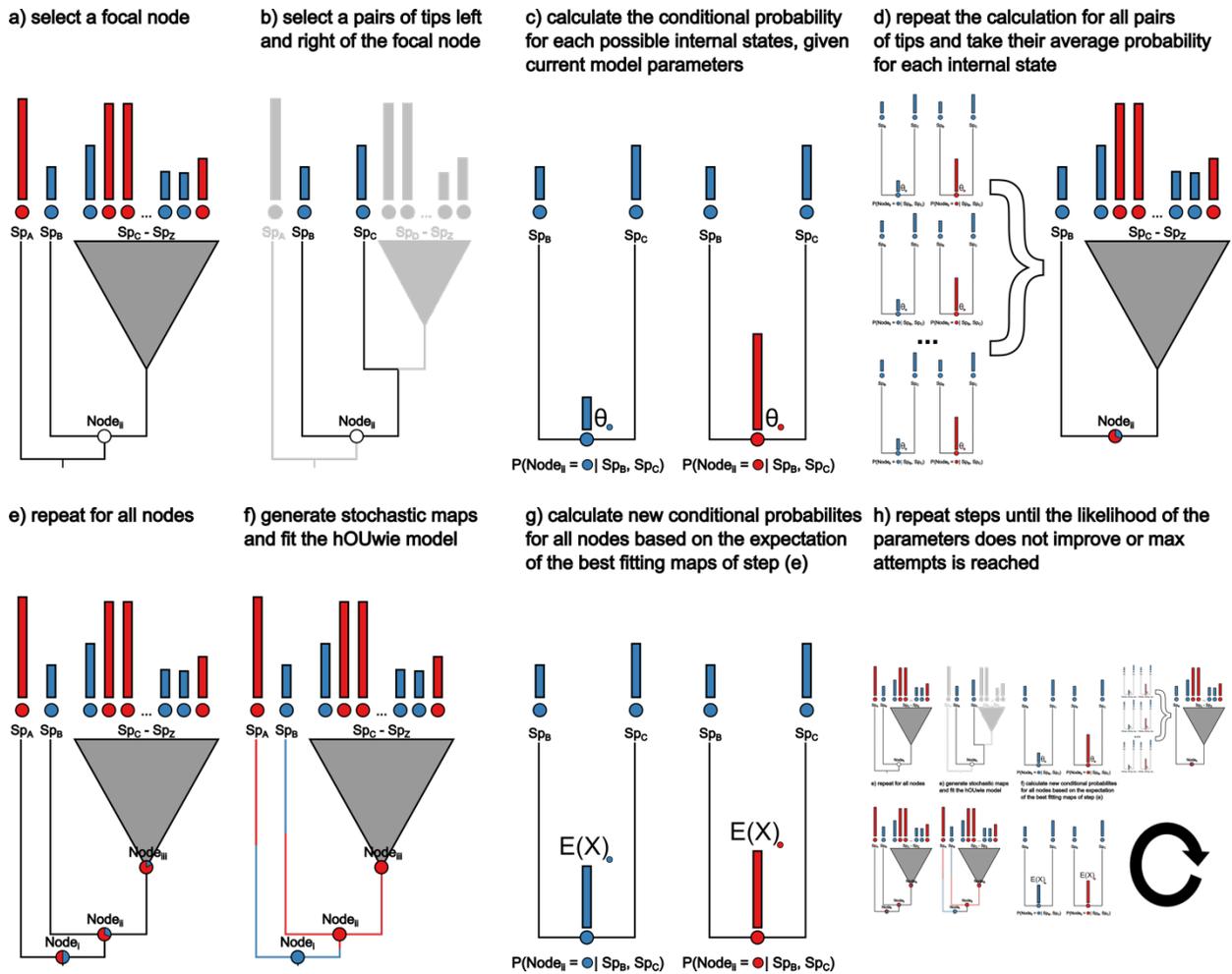
Figure 1. A visual representation of the algorithm underlying the calculation of conditional node probabilities and the adaptive sampling procedure. The goal of the procedure is to produce underlying regime paintings well suited to both the discrete and continuous character. a) select the focal node for which we will be calculating the joint conditional probabilities of the discrete and continuous characters. b) on each side of the node we select a pair of tips. c) the conditional probability of the observed discrete and continuous character is calculated for each discrete regime state with an ancestral continuous value equal to θ of that regime state. d) the conditional probability of the focal node is calculated as the average probability of each regime state for all pairs of observed tips. e) the conditional probabilities are calculated for all internal nodes. This can be turned off within hOUwie by setting the sample_nodes argument to false. f) A stochastic map is generating using forward simulation rejection sampling. g) adaptive sampling uses the highest joint probability of previously generated underling regimes to generate a set of ancestral continuous character values. This differs from previous ancestral values because instead of assuming the value θ for each regime state, it calculates the expected value given the root state and regime mapping for that particular node. h) we repeat steps d) through g) until the joint likelihood of the set of underlying regimes does not improve.

159   Our function for the joint probability of a continuous and a discrete character is,

160
$$P(X, D|\vartheta, \psi) = \sum_z P(X| D, z, \vartheta, \psi)P(D, z|\vartheta, \psi),$$
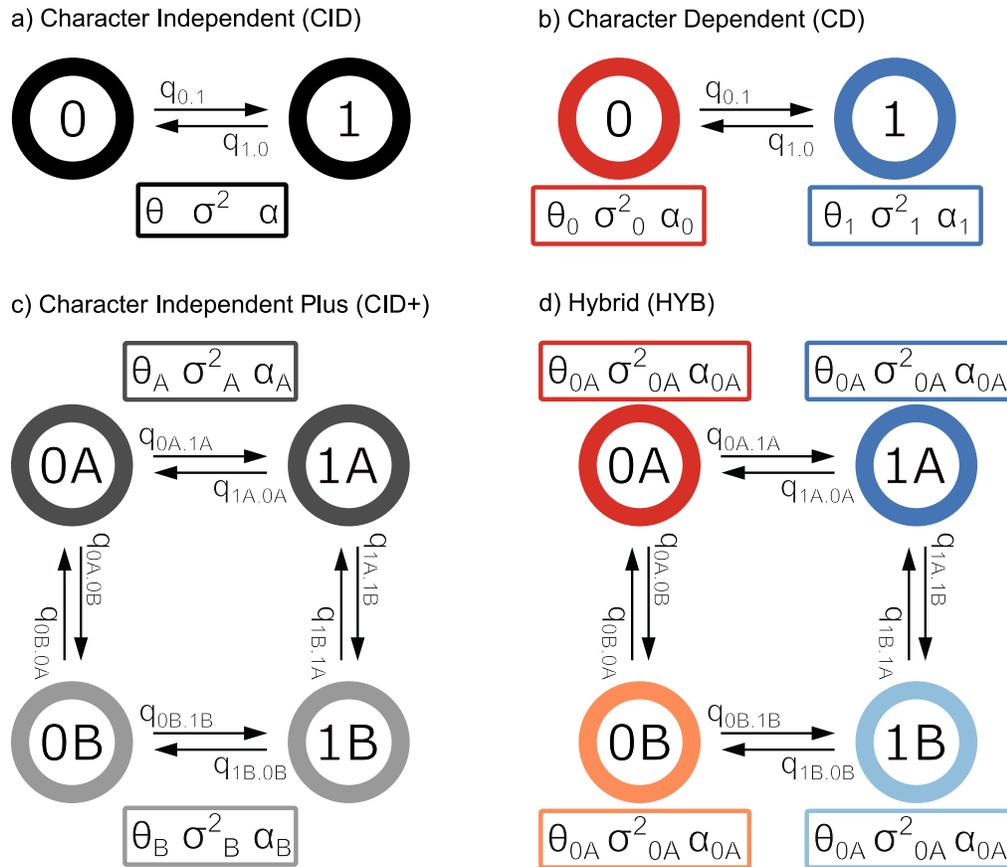
161 where summing over all generated maps (z), $P(X| D, z, \vartheta, \psi)$ is the probability of the continuous

162 character (X) given the discrete character data (D), mapping (z), hOUwie parameters (ϑ), and

163 phylogeny (ψ). $P(D, z|\vartheta, \psi)$ is the joint probability of the discrete character data (D) and

164 stochastic mapping (z) given the *hOUwie* parameters (ϑ) and phylogeny (ψ).

165

166   *The hOUwie model space*

167   Our simulation studies examined 22 possible *hOUwie* model structures for a binary

168 discrete character, although the possible number of models is significantly higher because any

169 number of discrete characters and states can be modeled together. For the discrete component of

170 the model, we assumed that transitions between the observed characters were equal. We

171 constrained transitions between hidden states to be the same for observed states, but this

172 constraint can be relaxed if desired. The continuous model structures allowable in *hOUwie* are a

173 generalized form of those allowed in *OUwie* and now include models in which only α varies

174 (OUA), only σ^2 varies (OUV), and combinations of an OU and BM process (OUBM1 and

175 OUBMV). We note that the OUBM1 model within *hOUwie* differs from The Ornstein–

176 Uhlenbeck Brownian-motion (OUBM) model presented in Hansen et al. (2008) and Bartoszek et

177 al. (2012) since the latter models are of multiple continuous characters, rather than different

178 processes describing the same continuous character.

179   The potential model structures range from completely character-dependent to character-

180 independent. Character-dependent (CD) models are models in which *any* continuous OU

181 parameter differs between observed discrete state, whereas character-independent models (CID)

Figure 2. A state-transition diagram describing the model classes allowable in hOUwie. Each panel is comprised of observed discrete states 0 and 1 with possible hidden states A and B. Transitions between states are described with the q parameter. Continuous model parameters appear in a box below the states they describe, and their association is displayed with a subscript specific to that state. a) A simple character independent model in which the two observed states do not influence the continuous character which will have the same θ, σ^2, α throughout the phylogeny. b) A character dependent model in which the continuous character depends on the discrete character by virtue of θ, σ^2, α being associated with a particular observed discrete state. c) A character independent model with rate heterogeneity. The two observed states (0 and 1) are not directly linked to the continuous character. However, the continuous character is still allowed to have multiple θ, σ^2, α describing its evolution, but these parameters are associated with hidden states A and B. d) A hybrid model in which each combined observed and hidden state is allowed to have its own θ, σ^2, α. Under this model, the continuous character is linked to both character dependent differences (parameters associated with 0 and 1) and character independent differences (A and B). Though this diagram shows a binary observed and hidden character, either can have more states (up to 26 states for each in theory, though few datasets will have enough power to estimate the necessary number of parameters).

182   test whether observed discrete states can be described by the same OU parameters. There are two

183   types of character-independent model (Fig. 2). First, character-independent models include

184   structures where there are no differences between any OU parameters. Under this model the

185   entire evolutionary history of the clade can be described by a single α, σ^2, and θ (Fig. 2a). To

186   combat this unrealistic assumption we introduce a character-independent model which allows for

187   differences in the OU parameters to depend upon an unobserved hidden state (CID+) and has

188   been shown to correct for the bias towards detecting correlation (Boyko and Beaulieu 2022).

189   This addition allows for heterogeneity within the evolutionary process without the necessity of it

190   being linked to a focal trait (Fig. 2c). In total we examine 22 unique model structures (2 CID, 10

191   CD, and 10CID+).

192

193   *Simulation study*

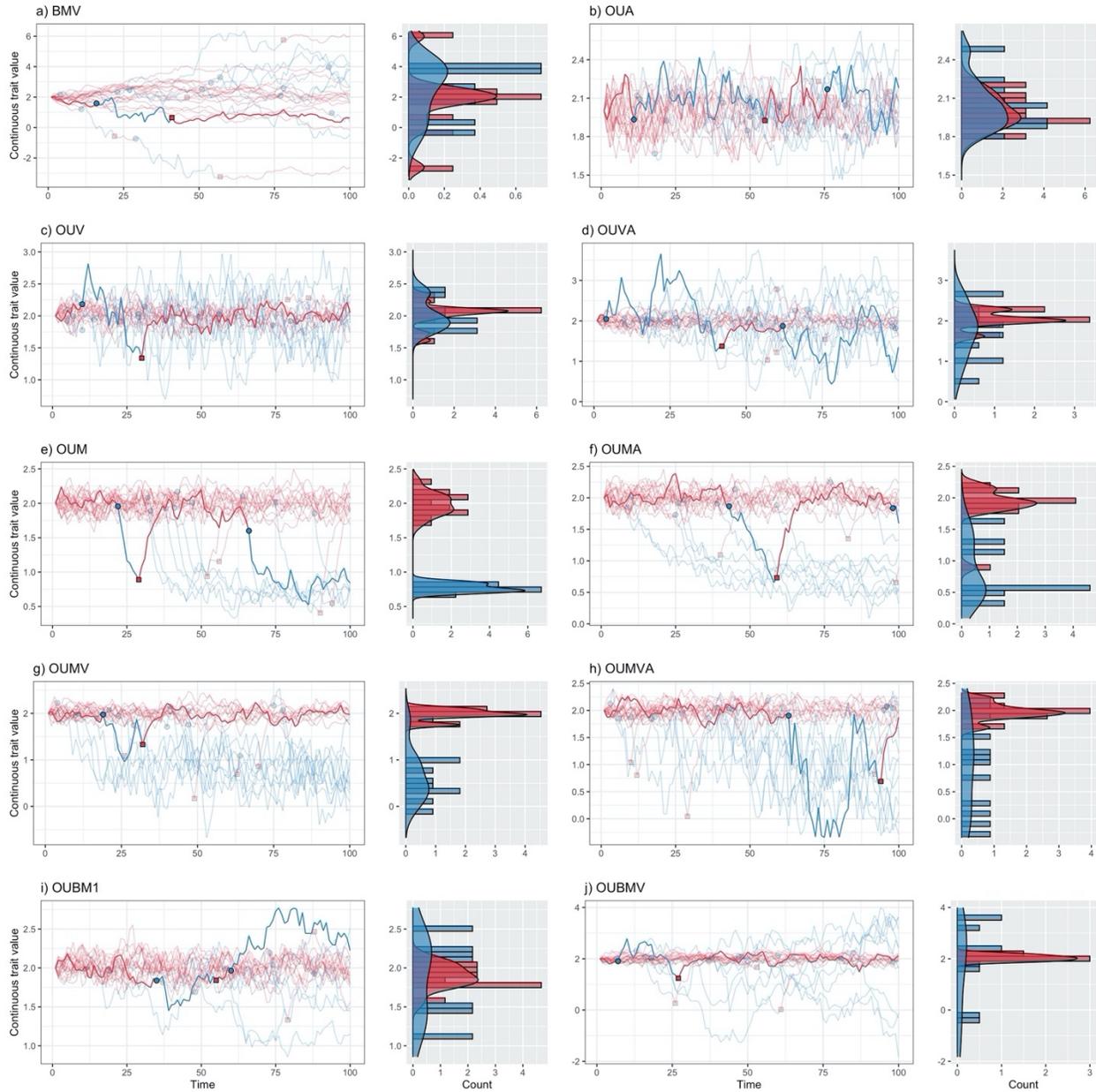194   For each of the 22 *hOUwie* model structures, we simulated 50 datasets for phylogenies of

195   25, 100, and 250 taxa for a total of 3300 unique datasets. Phylogenies were pure birth

196   phylogenetic trees with $\lambda = 1$, rescaled tree height to 1, and the root state was fixed to state 1.

197   The parameters used to generate a phenotypic dataset depend on the structure of the generating

198   model. For example, an OUM model and OU1 model can have identical q_{ij}, α, and σ^2, but they

199   must differ in θ or else OUM will collapse into OU1. The simulating parameters were chosen to

200   match Beaulieu et al. (2012) with $q_{ij} = 0.1, \alpha_1 = 3, \alpha_2 = 1.5, \sigma_1^2 = 0.35, \sigma_2^2 = 1, \theta_1 =$

201   2, and $\theta_2 = 0.75$ (Fig. 3). Once a phylogeny and phenotypic dataset were simulated, we fit our

202   models to assess parameter estimation accuracy and model selection power. Although this

203   represents a small subset of the potentially vast parameter space available to OU models, the

204   behavior of these models has been thoroughly characterized and thus we chose parameters within

205   the range of typical identifiability (Beaulieu et al. 2012; Ho and Ané 2014*a*; Cressler et al. 2015).

206   Additionally, because *hOUwie* uses a variable number of mappings, we evaluate changing the

Figure 3. A visual representation of binary discrete character *hOUwie* model types. Discrete time forward simulations are conducted starting in the red state and the distribution of the continuous character is plotted on the right as a histogram and density plot. Each line represents a continuous character value at some time. Transitions occur at colored points and each line is colored by the current discrete state. 100 time-steps are simulated with the same parameters as our simulation study ($q_{ij} = 0.1, \alpha_1 = 3, \alpha_2 = 1.5, \sigma_1^2 = 0.35, \sigma_2^2 = 1, \theta_1 = 2$, and $\theta_2 = 0.75$). The highlighted line was randomly chosen from the set in which at least one discrete state transition occurred.

207   number of stochastic maps. We fit each model using 25, 100, and 250 stochastic mappings per

208   likelihood evaluation. Each dataset was evaluated using the true generating model, a BM1, an

209　　OU1, and either the character-dependent or character-independent counterpart to the generating

210　　model. For example, if the data were simulated under a character-dependent OUM model where

211　　the value of θ_1 and θ_2 depend on the observed character, a character-independent OUM model

212　　would also be fit as part of the model set. Under the CID+ OUM model, a variable θ is still

213　　allowed, but it is unlinked to the focal character and thus should provide a more reliable

214　　character independent null hypothesis than BM1 or OU1 (Beaulieu and O'Meara 2016; Uyeda et

215　　al. 2018; May and Moore 2020; Boyko and Beaulieu 2022).

216

217　　　　　　　　　　　　　*The impact of climatic variables on seed dispersal*

218　　　　　For sedentary organisms, such as plants, dispersal is mainly limited to a brief stage of

219　　their life cycle and mediated mainly through the movement of seeds (Levin et al. 2003).

220　　Generally, the expectation is that seeds dispersed by frugivores are going to be dispersed to

221　　environments more like their parents' environment, whereas abiotically dispersed seeds are

222　　likely to be more erratic in their dispersal patterns (Schupp 1993; Westoby et al. 1996).

223　　Furthermore, it has been proposed that adaptations for frugivorous dispersal is linked to tropical

224　　and subtropical biomes, because in these warmer and wetter habitats, large trees create shady

225　　environments where competition for light is more important. A shadier habitat then imposes a

226　　selective pressure for larger seeds because more nutrients are needed for germination and initial

227　　survival (Foster and Janson 1985). However, the evolution of larger seeds comes with a tradeoff

228　　as they have a significantly lower dispersal potential (Howe and Smallwood 1982). Thus, we

229　　might expect that the climatic variables of a habitat influence the probability of transitioning

230　　between abiotic and biotic modes of dispersal, with transition rates from abiotic to biotic being

231　　greater in less arid environments.

232   Here we use dry or fleshy fruit morphology as a proxy for abiotic or biotic seed dispersal

233   (Lorts et al. 2008) to evaluate three predictions outlined in Vasconcelos et al. (2021), but

234   specifically measuring the aridity index. First, we expect that the climatic optima for fleshy fruits

235   will be more humid compared to dry fruits ($\theta_{dry} < \theta_{fleshy}$). Second, we expect that dry fruits

236   will have faster rates of climatic niche evolution ($\sigma^2_{dry} > \sigma^2_{fleshy}$). Finally, we expect that the

237   climatic niches of fleshy fruits will be more conserved through time ($\alpha_{dry} < \alpha_{fleshy}$). We apply

238   several *hOUwie* models to test these hypotheses and compare our results to those discussed in

239   Vasconcelos et al. (2021). We expect that any differences found between this study and

240   Vasconcelos et al. (2021) are because we can explicitly account for the joint probability of the

241   discrete and continuous characters. We focus our attention on Ericaceae specifically because

242   Vasconcelos et al. (2021) found two counter-intuitive results. Namely, they found that the

243   phenotypic optima of dry fruits were more humid than fleshy fruited lineages, and that the rate of

244   climatic evolution was greater in fleshy fruits than dry fruits.

245   We included 25 *hOUwie* models within our model set: 2 CID, 10 CD, 10 CID+, and 3

246   HYB. *Gaultheria* is technically a dry-fruited genus within Ericaceae but has a persistent fleshy

247   calyx that attracts frugivores (Stevens et al. 2004). However, since we are interested in the

248   association between dispersal and fruit type, we code this as fleshy fruited within our dataset.

249   Models are evaluated using the sample size corrected Akaike Information Criterion (AICc) and

250   model averaging is conducted when discussing how our results relate to our hypotheses

251   (Burnham and Anderson 2002). Measurement error is included for each model fit as within

252   species variance (the sample-sized weighted average of the individual species variances

253   following Labra et al. (2009) and Vasconcelos et al. (2021)). We evaluate then model averaged

254   parameter estimates of θ, σ^2, and α for fleshy and dry fruited lineages, as they relate to our

255   hypotheses and compare our results to Vasconcelos et al. (2021). Finally, we conduct a

256   parametric bootstrap of 100 simulated datasets to evaluate the standard error of our model

257   averaged parameter estimates.

258

259   **Results**

260   *Simulation study*

261   For character-independent (CID) models, our heuristic adaptive sampling algorithm,

262   which uses information from the discrete and continuous characters to guess at mappings,

263   consistently produced more probable mappings than using purely discrete mappings for all

264   models examined. On average, adaptive sampling produced mappings which were roughly 38

265   log likelihood units better than purely discrete sampling when examining joint probabilities. This

266   was driven primarily by the improved continuous probabilities which were on average 38.4 log

267   likelihood units better. In contrast, the discrete probability of each mapping was similar with

268   discrete-only simulations producing maps that were on average 0.39 log likelihood units better

269   (Table 1; Fig. S1). For character-dependent models, the difference was negligible (not shown).

270   This is because when the discrete and continuous character are strongly linked, discrete-only

271   mappings will match the continuous character's distribution quite well.

272   Most character-dependent models (CD) had lower overall deviations from the generating model

273   across all model types. The RMSE was largest for alpha at 1.76 and 1.65 (if variable alpha) and

274   errors were generally higher for more complex models. All other parameters had relatively

275   similar RMSE, ranging from 0.1 for discrete the rate to 0.75 for σ_2^2. The BMV (BM with

276   variable σ), OUV (OU with variable σ), OUA (OU with variable α), and OUM (OU with

277   variable θ) models generally had the lowest errors, but there were some biases present (Table 2).

278 Most notably, alpha was biased upwards for OUM and OUV models and under variable alpha

279 models (OUA, OUMA, OUVA, OUMVA), the difference between the alpha estimates tended to

280 be larger than the generating parameter difference. The more complex models had larger error

281 variances but showed similar biases as the simple models. Finally, OUBM models showed a

282 significantly downward biased α, suggesting BM like processes (Fig. S2).

283 **Table 1.** A comparison of the effectiveness of the adaptive sampling procedure and standard
284 discrete only sampling of maps. Regardless of the sampling procedure, all probabilities are
285 calculated in the same way and so any differences in probabilities reflects each procedure's
286 ability to generate appropriate mappings. 50 regime mappings are used to calculate the likelihood
287 of the parameters. A higher log_e likelihood is better (that is, -16.43 is better than -16.48; 10.54 is
288 better than 9.19)For each model type, data are simulated following our methods with $q_{ij} =$
289 $0.1, \alpha_1 = 3, \alpha_2 = 1.5, \sigma_1^2 = 0.35, \sigma_2^2 = 1, \theta_1 = 2,$ and $\theta_2 = 0.75$. The generating parameters are
290 used to evaluate probability of each dataset and thus the probabilities represented here are not
291 necessarily the same as those derived from the MLE. Generally, adaptive sampling improves the
292 joint estimate by improving the probability of the continuous character and is most effective for
293 variable θ models. As expected, discrete only sampling produces regime paintings which better
294 reflect the discrete character than adaptive sampling, but the difference is minor.

Model class	Model type	Sampling procedure	Discrete marginal log_e likelihood	Continuous marginal log_e likelihood	Joint log_e likelihood
CID+	BMV	adaptive sampling	-16.48	10.54	**-10.59**
		discrete only	-16.43	9.19	-10.59
	OUA	adaptive sampling	-15.46	44.34	**25.14**
		discrete only	-15.53	43.11	24.96
	OUV	adaptive sampling	-30.89	47.86	**12.17**
		discrete only	-30.14	46.00	12.11
	OUVA	adaptive sampling	-11.88	36.91	**21.14**
		discrete only	-11.17	36.27	21.08
	OUM	adaptive sampling	-11.94	57.57	**39.08**
		discrete only	-11.19	53.56	32.21
	OUMA	adaptive sampling	-9.94	35.01	**17.39**
		discrete only	-9.38	2.19	-20.48
	OUMV	adaptive sampling	-19.96	20.77	**-15.64**
		discrete only	-14.76	-2.92	-25.83
	OUMVA	adaptive sampling	-13.91	25.47	**7.48**
		discrete only	-13.23	26.36	4.48
	OUBM1	adaptive sampling	-14.26	42.20	**24.39**
		discrete only	-14.88	40.89	24.22
	OUBMV	adaptive sampling	-19.17	49.10	**18.84**
		discrete only	-19.01	33.45	7.71

295 **Table 2.** The average accuracy of *hOUwie* parameter estimates across several model classes and
296 types as measured by root-mean-square error (RMSE). RMSE is calculated for each model type
297 by taking the square root of the mean squared error (MSE), where MSE is the average squared
298 difference between the MLE and the simulating parameters. Data is generated with $q_{ij} =$
299 $0.1, \alpha_1 = 3, \alpha_2 = 1.5, \sigma_1^2 = 0.35, \sigma_2^2 = 1, \theta_1 = 2,$ and $\theta_2 = 0.75,$ and for phylogenies with 25,
300 100, and 250 taxa. Finally, model fits use either 25, 100, or 250 stochastic maps per likelihood
301 iteration. The table shown here calculates RMSE integrating over all phylogenetic tree sizes and
302 number of stochastic maps (n=8217). Dashes indicate a parameter that is not estimated for a
303 given model type. Generally, character independent (CID+) models had higher errors than
304 character dependent (CD) models. The greatest errors occurred when estimating alpha in variable
305 alpha models for both CD and CID+ model classes. Estimates of the optimum and transition
306 rates generally had the lowest errors.

Model class	Model type	RMSE q	RMSE α_1	RMSE α_2	RMSE σ_1^2	RMSE σ_2^2	RMSE θ_1	RMSE θ_2
CD	BMV	0.12	-	-	0.10	0.28	0.22	-
	OUV	0.11	1.27	-	0.15	0.33	0.05	-
	OUA	0.12	1.55	1.63	0.11	-	0.06	-
	OUM	0.13	1.49	-	0.10	-	0.07	0.13
	OUVA	0.09	1.44	1.11	0.14	0.98	0.06	-
	OUMV	0.16	1.82	-	0.16	0.32	0.07	0.17
	OUMA	0.15	2.11	2.48	0.28	-	0.12	0.50
	OUMVA	0.18	1.62	1.12	0.12	1.07	0.76	1.06
	OUBM1	0.1	2.64	-	0.08	-	0.08	-
	OUBMV	0.09	2.29	-	0.13	2.37	0.08	-
CID+	BMV	0.05	-	-	0.27	10.11	0.24	-
	OUV	0.04	1.13	-	0.32	1.83	0.05	-
	OUA	0.05	2.93	1.34	0.33	-	0.07	-
	OUM	0.09	2.53	-	0.15	-	0.44	0.20
	OUVA	0.05	1.26	1.11	0.27	13.44	0.07	-
	OUMV	0.1	2.50	-	0.16	2.12	1.30	0.68
	OUMA	0.05	8.28	1.27	0.23	-	5.88	0.8
	OUMVA	0.07	5.54	1.24	0.20	9.37	8.76	1.35
	OUBM1	0.05	3.33	-	0.32	-	0.14	-
	OUBMV	0.05	3.50	-	0.27	8.79	0.14	-

307

308 Character-independent models with rate heterogeneity models generally performed well

309 in terms of parameter estimates, but as expected, due to their inherit uncertainty, CID+ models

310 had larger errors than CD models. The largest error was estimates of σ_2^2 which had an RMSE of

311 8.5, although the median error value was only 0.03, suggesting that the large RMSE is driven by

312 a long rightward tail of the estimates. Like CD models, α_1 and α_2 consistently showed the

313 largest RMSE at 3.6 and 1.2. In general, α was underestimated with medians of -0.4 and -1.4

314 below the simulating values of 3 and 1.5. This means that models for CID+ models tended to be

315 more BM like even under an OU generated data (Fig. S2). Increasing the number of taxa

316 examined improved both CD and CID+ performance. The RMSE for α was nearly cut in half

317 between when moving from 25 tips to 250 tips from 5.2 to 2.8 under CID+ models (Table 3).

318 Nonetheless, some parameters continued to be estimated poorly, such as σ_2^2. Interestingly,

319 increasing the number of stochastic maps improved CID+ performance, but did not substantially

320 improve estimation under CD models (Fig. S2c).

321 **Table 3.** Average AIC weight as the number of taxa increases for each model class. Gray cells
322 indicate the AIC weight of the generating model class. In general, as the number of taxa
323 increases the average support for the generating model class increases.

Generating model class	nTaxa	AICwt BM1	AICwt OU1	AICwt CD	AICwt CID+
CD	25	0.12	0.22	0.51	0.15
	100	0.06	0.22	0.70	0.02
	250	0.02	0.14	0.82	0.02
CID+	25	0.28	0.35	0.24	0.14
	100	0.21	0.4	0.23	0.15
	250	0.11	0.34	0.32	0.22

324

325   Generally, evidence of CD when it was the generating model was consistent across all

326 model types. The lowest support for the OUA and OUBM1 models at an average AICwt of 0.31

327 and 0.13. For complex models, such as OUMVA, model support for was 0.81 and highest for

328 OUMV at 0.97. CID+ models fared worse in terms of generating consistent support even when

329 they were the generating model. Models which were difficult to estimate under character

330 dependence were difficult to find consistent support for under character independence. The most

331 extreme case was OUA model for which CID+ model was never chosen as the best supported

332 model. However, models which performed well for CD tended to perform well under CID+. For

333 example, OUM models garnered consistent support when with an average AICwt of 0.733

334   (Table S1; Fig. S3). While the best model under AICc need not be the generating model (for

335   example, for a small dataset a simpler model may lose less information than the generating

336   model) given the size of the simulated trees and distinctness of the models we expect the

337   generating model to generally be the best.

338   For both CD and CID+ models, support improved when increasing the number of tips

339   analyzed. Support for a CD model when CD was the generating model increased from $w_{CD} =$

340   0.5 to $w_{CD} = 0.67$ to $w_{CD} = 0.79$ for 25, 100, 250 tips and support for a CID+ model when it

341   was the generating model increased from $w_{CID+} = 0.11$ to $w_{CID+} = 0.15$ to $w_{CID+} = 0.22$

342   (Table 3). Similarly, increasing the number of regime maps generally improved the fit, but not as

343   much as increasing the number of tips. We found that the false evidence of correlation (as

344   measured by the average AICwt of a character-dependent model when character-independence

345   was the generating model) was generally not an issue for variable θ models (OUM*). Variable θ

346   models had average AICwts for false character-dependence ranging from 0.03 to 0.23 and for

347   none of our simulations models was a CD model best supported. Under a simple OUM model,

348   CID+ models helped correct any potential bias with an average AICwt of 0.68. However, false

349   evidence of correlation was an issue for variable σ_i^2 and α_i models. False support for CD as

350   measured by AIC weight ranged from 0.34 to 0.44 when θ was fixed and α_i and/or σ_i^2 varied.

351   Although CID+ models did not garner much support when these models were fit, OU1 and BM1

352   models served as reasonable null hypotheses in these cases. In general, we found that when CID

353   models were the generating model, evidence of CID was strongest and when CD models were

354   the generating model, evidence of character dependence was strongest. This suggests that the

355   effect of rate heterogeneity causing false correlations is not as pronounced as other comparative

356   methods (Maddison and FitzJohn 2015; Rabosky and Goldberg 2015).

358   We found evidence of a character-dependent model over either a simple or hidden state

359   character-independent model, suggesting a link between the climatic niche of Ericaceae lineages

360   and their fruit type (Table S2). The best supported models were OUMVA and OUVA with AIC

361   weights of 0.41 and 0.32 respectively. This suggests that there were character dependent

362   differences in phenotypic optima, rates of evolution, and overall phylogenetic signal. To evaluate

363   support for our hypotheses we examined the model averaged parameter estimates (Table 4). The

364   estimated optimum 0.81 $ln(AI)$ ($\pm$ 0.28) for fleshy fruits suggests a more arid environment for

365   their optimal habitat, and the 0.97 $ln(AI)$ ($\pm$ 0.011) of dry fruits corresponds to a more humid

366   environment (Middleton and Thomas 1997), where AI is measured as mean annual precipitation

367   (P) dived by average annual potential evapotranspiration (PET). However, both optima

368   correspond to non-dryland humid environments. Both σ^2 and α interact to create tip variance, so

369   in addition to σ^2, we measured the stationary variance $V = \frac{\sigma^2}{2\alpha}$. As predicted, we found that

370   Ericaceae lineages with dry fruits were more variable in their climatic niche evolution ($\sigma^2_{dry} =$

371   0.011 $ln(AI)^2 MY^{-1}$, $V_{dry} = 0.37\ ln(AI)^2$) compared to fleshy fruits ($\sigma^2_{fleshy} =$

372   0.007 $ln(AI)^2 MY^{-1}$, $V_{fleshy} = 0.15\ ln(AI)^2$). Additionally, the strength of pull of fleshy fruited

373   lineages was greater than dry fruited lineages ($\alpha_{fleshy} = 0.022 MY^{-1} > \alpha_{dry} = 0.014 MY^{-1}$).

374   This corresponds to phylogenetic half-lives of $t_{1/2,dry} = 46.4$ MY and $t_{1/2,fleshy} = 30.3$ MY

375   which are 38% and 25% of the total tree height respectively. Transitions to fleshy fruit occurred

376   at 0.0015 transitions per million years which is more than 4.3 times faster than transitions to dry

377   fruits (0.0004 transitions per million years). The waiting time $\left(\frac{1}{q}\right)$ of fleshy fruits (2,500 MY)

378   was substantially longer than that of dry fruits (667 MY). Given that the total branch length in

379   the tree is $10,120 \, MY$, we expect that lineages were typically under the fleshy fruit regime and

380   evolving towards a preference for more humid environments. Perhaps for this reason we found

381   that, on average, lineages were in more arid environments than predicted by the model (average

382   difference of 0.19 AI), with some species expected to be in much more humid environments

383   (difference between current AI and optimal AI ranged from -4.4 to 0.85).
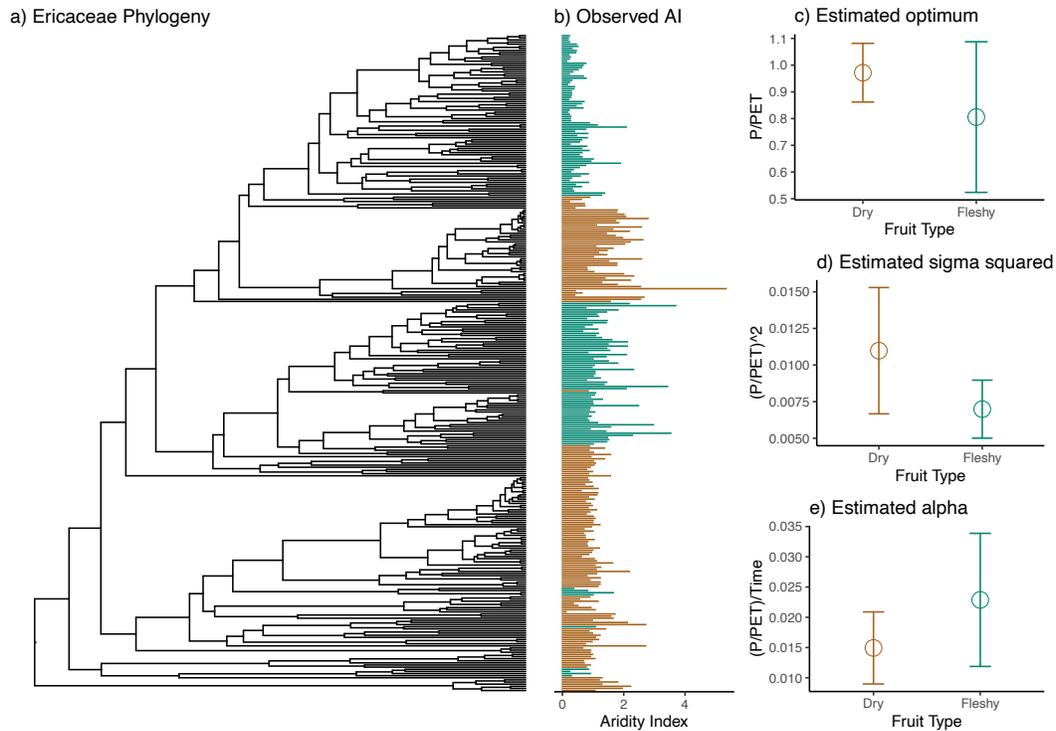
Figure 4. a) Ericaceae phylogeny for which we had data (n=309). b) Ln aridity index dataset where each bar is colored by dry (brown) and fleshy (green) fruit type. c) Model averaged parameter estimates with standard error calculated from 100 parametric bootstraps.

384   **Table 4.** Model averaged parameter estimates and standard errors for Ericaceae aridity index and
385   fruit type data. Models with higher AIC weights contribute more overall to the parameter values.
386   The units for α, σ^2, and θ are $\frac{P}{PET} \div time$, $\left(\frac{P}{PET}\right)^2$, and $\frac{P}{PET}$ respectively. P is the average annual
387   precipitation and PET is average annual potential evapotranspiration. Rates of q are measured in
388   transitions per million years.

Continuous parameter estimates				Discrete parameter estimates	
	α	σ^2	θ		
Dry	0.015 (±0.0059)	0.011 (±0.0043)	0.97 (±0.011)	$q_{dry\,to\,fleshy}$	0.0015 (±0.00058)
Fleshy	0.023 (±0.011)	0.007 (±0.002)	0.81 (±0.28)	$q_{fleshy\,to\,dry}$	00036 (±0.000086)

Discussion

Phylogenetic comparative methods have been widely applied to study discrete and continuous characters separately. Due primarily to computational limitations there are few options which jointly evaluate both classes of character. The *hOUwie* framework proposed here overcomes these limitations, and we demonstrate how it is used to test hypotheses of correlated evolution between discrete and continuous characters while accounting for hidden character states and unobserved variation. Our model jointly models discrete and continuous characters by linking both via a common regime painting. However, unlike other similar methods, our likelihood formula explicitly calculates the probability of the underlying regimes. This has the advantage of describing the discrete character evolution probabilistically and allows information from the discrete and continuous characters to jointly contribute to the overall likelihood.

Relationship to existing methods

Considerable progress has been made towards more realistic models of continuous character evolution within the last two decades. Continuous character models which initially relied on either single rate Brownian motion or simple Ornstein-Uhlenbeck models (Felsenstein 1985; Hansen 1997) have seen several extensions to allow for heterogeneity in the evolutionary process as well as the deterministic influence of underlying independent variables. Generally, these models can be classified as either being "hypothesis driven" or "data driven" (Martin et al. 2022). Hypothesis driven models are those which require *a priori* hypotheses regarding where evolutionary rates may differ throughout the phylogeny. These include models which have extended simple single-rate BM to incorporate rate variation based on discrete regime mappings (e.g., O'Meara et al. 2006; Thomas et al. 2006; Revell and Collar 2009; Caetano and Harmon

412   2017) or more generalized Ornstein-Uhlenbeck models where parameters are allowed to vary

413   based on an underlying regime mapping (e.g., Butler and King 2004; Bartoszek et al. 2012;

414   Beaulieu et al. 2012). In contrast, several methods have focused on the development of data

415   driven, shift-detection methods (which may indeed be used in testing hypotheses, but these

416   hypotheses are not directly used in creating the regime map). These methods utilize an Ornstein-

417   Uhlenbeck process to automatically detect where in the phylogeny evolutionary rates and

418   phenotypic optima shift (Ingram and Mahler 2013b; Uyeda and Harmon 2014; Khabbazian et al.

419   2016; Bastide et al. 2017). Furthermore, some recently developed methods have allowed for rate

420   variation without the assumption of constant regimes at all. Instead, these models assume the

421   rates themselves evolve and change throughout the phylogeny under various Brownian motion-

422   like processes (Lemey et al. 2010; Eastman et al. 2013; Revell 2021; Martin et al. 2022) or single

423   optima Ornstein-Uhlenbeck processes (Hansen et al. 2008; Mitov et al. 2019). The method

424   presented here is most like the latter group. *hOUwie* attempts to explicitly model the evolution of

425   rate shifts according to regimes which jointly influence discrete and continuous character

426   evolution. The regimes themselves are never fixed a priori and each is evaluated as a partial

427   contribution to the overall probability of the data. The advantage of this approach is that it

428   acknowledges the uncertainty in the underlying regime paintings and allows them to change

429   through time.

430   Additionally, unlike *hOUwie*, the "hypothesis driven" or "data driven" models do not

431   explicitly account for the joint modeling of the discrete and continuous characters. Most progress

432   in this area has, until recently, been made via phylogenetic logistic regressions (Ives and Garland

433   2010) or threshold models in which the discrete character is modeled by a continuously varying

434   unobserved lability (Felsenstein 2012; Revell 2014; Cybis et al. 2015). However, these models

435   rely on more simplistic evolutionary models without character independent rate heterogeneity

436   (such as single rate Brownian motion). This lack of character independent rate heterogeneity has

437   recently been recognized as a potential source of inflated correlation between discrete and

438   continuous characters. Such was the reasoning for the MuSSCRat model (May and Moore 2020).

439   Like *hOUwie*, MuSSCRat allows for character-independent rate heterogeneity following a

440   multiple rate Brownian motion model to be directly contrasted against character correlation to

441   correct for potential biases towards correlation. However, the way the underlying discrete

442   character is calculated in *hOUwie*, as well as how rate heterogeneity is modeled, differs

443   substantially from May and Moore (2020). Finally, Tribble et al. (2021) has recently developed a

444   method which is similar to the one presented here. One of the primary differences between

445   *hOUwie* and the Bayesian pipeline discussed in Tribble et al. (2021) is how discrete character

446   evolution is treated. Specifically, Tribble et al. (2021) assumed that character-independent

447   mappings are generated under the same parameters which best fit their focal discrete character.

448   In contrast, *hOUwie* allows the free estimation of character-independent discrete rates which best

449   fit both discrete and continuous data. This difference may lead to biases against null models in

450   the Tribble et al. (2021) approach since the character-independent regimes are forced to follow a

451   character-dependent discrete model.

452

453   *Character-independent models and null hypotheses*

454   There is a growing appreciation that comparing constant-rate null models to variable-rate

455   alternative models will consistently favor rate heterogeneity, regardless of whether there is a

456   genuine association with a focal variable (Maddison and FitzJohn 2015; Rabosky and Goldberg

457   2015; Beaulieu and O'Meara 2016; Uyeda et al. 2018; O'Meara and Beaulieu 2021; Boyko and

458   Beaulieu 2022). This problem, termed the "straw-man effect" by May and Moore (2020), has

459   been demonstrated to lead to nearly 100% error rates for evidence of discrete character

460   correlation (Maddison and FitzJohn 2015; Boyko and Beaulieu 2022), and has severely biased

461   evidence towards state-dependent speciation and extinction (Rabosky and Goldberg 2015;

462   Beaulieu and O'Meara 2016). Given these often-overwhelming error rates in other comparative

463   methods, we expected to find a similarly consistent bias towards correlation between discrete and

464   continuous characters. However, we found that support for single rate character-independent null

465   models was greater than character-dependent models even when simulated under character-

466   independent models with rate heterogeneity. Although the inclusion of explicit multi-rate

467   character independent models (CID+) models did help reduce evidence of false correlation in

468   some cases, by and large, simplistic null models performed admirably. This is not to say that the

469   error rates for discrete and continuous character correlation should be dismissed outright. If our

470   simulations correctly assess that nearly one-third of results find false evidence of a correlation

471   between continuous character rates of evolution and discrete characters, then better null models

472   are certainly needed. But, in comparison to the profound effect that model misspecification has

473   had in other comparative analyses (Beaulieu and O'Meara 2016; Boyko and Beaulieu 2022), the

474   joint models tested here have substantially lower error rates.

475   We suspect that part of the reason that the correlation between discrete and continuous

476   characters is less susceptible to "straw-man" effects than other PCMs is related to the

477   inefficiency of sampling potential maps from the univariate stochastic mapping model. A

478   common approach to fitting OU models involves simulating many stochastic maps to represent

479   underlying regimes from parameters estimated only from the discrete character (Revell 2013).

480   The resulting distribution of underlying regimes will therefore reflect a distribution appropriate

481   for the discrete character, but not necessarily suitable for the continuous character. This is

482   especially true if the continuous character is unlinked to the focal discrete character. Indeed, we

483   found that if the discrete and continuous characters are unlinked, most stochastic maps, even

484   though good descriptions of the discrete characters, were completely inadequate representations

485   of continuous regimes. Thus, any joint model with these maps contributed little to the overall

486   likelihood. Under our simulation protocol, for a typical run, 90% of the total likelihood for the

487   best set of parameters came from just 2% of the attempted maps.

488   In some ways the substantial contributions of only a few underlying regimes to the

489   overall likelihood is good. First, it makes spurious links between a randomly distributed discrete

490   character and a continuous character more unlikely since associations between regimes and

491   continuous variables tend to be specific. This ultimately reduces the potential "straw-man"

492   effect. Second, the continuous characters can inform the placement of shared regimes and

493   therefore shift detection methods, where the continuous data are all that provides information

494   about regimes shifts (Ingram and Mahler 2013; Uyeda and Harmon 2014; Khabbazian et al.

495   2016; Bastide et al. 2017), may be appropriate across a broad range of scenarios. However, this

496   property also makes sampling a good set of regimes to get an accurate estimate of the likelihood

497   difficult and is why the development of our adaptive sampling heuristics was necessary.

498   Adaptive sampling, in combination with our approximation of the joint conditional distributions,

499   helped make parameter estimation more accurate. Increasing the amount of sampled regime

500   mappings is useful in improving precision (Fig. S1), at the cost of longer run time.

501

502

503

505   In many studies that deal with the correlation of discrete and continuous traits, it is often

506   assumed that the discrete trait functions as the independent trait and the continuous trait as the

507   dependent trait. This assumption is baked into methods that map the discrete trait first and then

508   analyze the continuous trait given these mappings, but it would be easy to fall into this form of

509   thinking even with *hOUwie*, which does not have this assumption. Instead, *hOUwie* can help

510   understand whether and how traits are correlated. For example, one could see if mammal body

511   size correlates with trophic level: are hypercarnivores larger on average than herbivores? It could

512   be that an herbivorous (discrete character) beaver evolves a taste for meat and then grows bigger

513   (continuous character) so it can take down bigger prey; it could be that once things get to be the

514   size of a bison (continuous character) they start adding more and more rodents to their diet,

515   eventually becoming carnivores (discrete character). Causality can go both directions, and of

516   course both traits may be evolving based on some other third trait and not functionally related to

517   each other.

518   *hOUwie* is part of a series of hidden state models developed by our research groups (i.e.,

519   Beaulieu et al. 2013; Beaulieu and O'Meara 2016; Caetano et al. 2018; Boyko and Beaulieu

520   2021, 2022; Vasconcelos et al. 2022). One misconception we have noted in use of these methods

521   is the thought that there is a single, discrete, hidden character in the biology. These models do

522   model a single hidden character (with potentially many states), but this could be reflecting

523   multiple characters evolving together or other factors that change in a heritable manner through

524   time. It is a way to allow heterogeneity, especially by factors that vary by clades. With *hOUwie*,

525   this heterogeneity can affect the discrete trait, the continuous trait, both, or neither.

526

528		Here we revaluated three hypotheses related to climatic niche evolution and seed

529	dispersal and found that: (1) the climatic optima of dry fruits was more humid than fleshy fruits

530	($\theta_{fleshy} < \theta_{dry}$), (2) lineages with dry fruits had faster rates of climatic niche evolution ($\sigma^2_{dry} >$

531	σ^2_{fleshy}), and (3) climatic niches of fleshy fruits are more conserved through time ($\alpha_{dry} <$

532	α_{fleshy}). In contrast to previous findings, the higher rate and stationary variance of climatic niche

533	evolution for dry seeds matched our original hypothesis (Vasconcelos et al. 2021). This is to be

534	expected because abiotically dispersed seeds are likely to be more erratic in their dispersal

535	patterns (Schupp 1993; Westoby et al. 1996). Additionally, that our results differ from previous

536	findings (Vasconcelos et al. 2021) suggests that jointly modeling climatic niche evolution

537	alongside fruit type changed our parameter estimation in a meaningful way.

538		Our final hypothesis, which stated that fleshy, biotically dispersed, seeds are more likely

539	to be associated with humid environments, was not supported. However, it has been suggested

540	that a trade-off between seed persistence, seed size, and dispersal strategies can be also common

541	in arid environments (Venable and Brown 1988; Nunes et al. 2017). Specifically, large seed size

542	may occasionally help withstand unfavorable conditions associated with increased aridity (Nunes

543	et al. 2017). With an increased seed size, biotic seed dispersal and fleshy fruits, may become

544	necessary for seed dispersal. This may be the case for Styphelieae, which is distributed in the

545	arid Australian heathland and, of all predominately fleshy-fruited groups, lies the furthest from

546	the inferred aridity optima. Additionally, it has been found that the proportion of abiotically

547	dispersed seeds increases as elevation increases, due to the decreasing availability of frugivores

548	(Chapman et al. 2016). Given that several radiations of Ericaceae lineages are associated with

549	montane habitats (Schwery et al. 2015), it may be that the distribution of dry and fleshy fruits are

550   a consequence of elevation rather than being directly linked to climatic niche evolution. Finally,

551   it has been noted Ericaceae lineages are often found in well-leached soils and epiphytic habitats

552   (Schwery et al. 2015). If associations with soil type are more important than links to climatic

553   optima, we may expect that fruit-dependent climatic optima are consequence of unmodeled

554   factors. Although our modeling explicitly considers hidden variables that may lead to rate

555   heterogeny, if the proposed hidden variable (soil condition) is closely linked to our modeled

556   variable (aridity), then we may not be able to detect the presence of hidden variation. This may

557   be the case between soil condition and aridity (Moreno-Jiménez et al. 2019).

558

559   *Caveats and possible extensions*

560   There are three important caveats to our proposed modeling framework. First, our

561   discrete mapping probability, $P(D, z|\vartheta, \psi)$, is only an approximation. What we calculate is the

562   probability of starting in a particular state i and ending a particular state j, summed over all

563   possible paths. However, the continuous model probability is based off a particular pathway

564   history that is defined throughout the entire branch (Hansen 1997). Ultimately, this means that

565   the underlying regimes are not treated identically for the continuous and discrete characters. The

566   second caveat is that we do not force *hOUwie* to sum over all possible mappings z. This is

567   because the number of mappings will grow exponentially as the number of nodes and internodes

568   increases and the computation will quickly become infeasible (see Jones et al. 2020). Although

569   this may not be entirely necessary since we have shown that only a small percentage of possible

570   mappings contribute to the overall joint probability. Nonetheless, an ideal solution could be the

571   use Markov-Modulated Ornstein-Uhlenbeck models (Huang et al. 2016) since this would remove

572   the need for a regime mapping approach, but these have yet to be applied in phylogenetic

573 comparative biology. *hOUwie* currently only deals with one discrete and one continuous trait at a

574 time – a set of discrete traits can be handled by converting them to a single multistate character,

575 but incorporating multiple continuous traits requires adding correlations between them. Finally,

576 it is possible to extend *hOUwie* to include state-dependent speciation and extinction dynamics

577 which have been shown to influence the distribution of discrete characters (Maddison 2006) and

578 would therefore influence continuous characters if the two were linked. However, this extension

579 would require a different calculation of the underlying regime mapping probability. Approaches

580 for stochastically mapping SSE models already exist (Freyman and Höhna 2019), so the largest

581 remaining challenge of this extension would be generating high joint probability mappings.

582 *Concluding remarks*

583 The use of pre-defined discrete character mappings can be useful for testing hypotheses

584 which rely on distinct, well-defined differences in the evolutionary histories of lineages.

585 However, this approach assumes that the underlying mapping is known with complete accuracy

586 and ignores the probabilistic nature of discrete regimes. *hOUwie*'s methodology integrates over

587 the uncertainty of high probability character mappings and relies on the interpretation of

588 parameter estimates from contrasting model structures to find evidence for hypotheses. Rather

589 than assuming an *a priori* mapping, *hOUwie* can utilize the mutual information about the

590 discrete and continuous characters to learn something about the underlying regimes evolution.

591

Tables

593   **Table 1.** A comparison of the effectiveness of the adaptive sampling procedure and standard
594   discrete only sampling of maps. Regardless of the sampling procedure, all probabilities are
595   calculated in the same way and so any differences in probabilities reflects each procedure's
596   ability to generate appropriate mappings. 50 regime mappings are used to calculate the likelihood
597   of the parameters. A higher log_e likelihood is better (that is, -16.43 is better than -16.48; 10.54 is
598   better than 9.19)For each model type, data are simulated following our methods with $q_{ij} =$
599   $0.1, \alpha_1 = 3, \alpha_2 = 1.5, \sigma_1^2 = 0.35, \sigma_2^2 = 1, \theta_1 = 2,$ and $\theta_2 = 0.75$. The generating parameters are
600   used to evaluate probability of each dataset and thus the probabilities represented here are not
601   necessarily the same as those derived from the MLE. Generally, adaptive sampling improves the
602   joint estimate by improving the probability of the continuous character and is most effective for
603   variable θ models. As expected, discrete only sampling produces regime paintings which better
604   reflect the discrete character than adaptive sampling, but the difference is minor.
605

Model class	Model type	Sampling procedure	Discrete marginal log_e likelihood	Continuous marginal log_e likelihood	Joint log_e likelihood
CID+	BMV	adaptive sampling	-16.48	10.54	**-10.59**
		discrete only	-16.43	9.19	-10.59
	OUA	adaptive sampling	-15.46	44.34	**25.14**
		discrete only	-15.53	43.11	24.96
	OUV	adaptive sampling	-30.89	47.86	**12.17**
		discrete only	-30.14	46.00	12.11
	OUVA	adaptive sampling	-11.88	36.91	**21.14**
		discrete only	-11.17	36.27	21.08
	OUM	adaptive sampling	-11.94	57.57	**39.08**
		discrete only	-11.19	53.56	32.21
	OUMA	adaptive sampling	-9.94	35.01	**17.39**
		discrete only	-9.38	2.19	-20.48
	OUMV	adaptive sampling	-19.96	20.77	**-15.64**
		discrete only	-14.76	-2.92	-25.83
	OUMVA	adaptive sampling	-13.91	25.47	**7.48**
		discrete only	-13.23	26.36	4.48
	OUBM1	adaptive sampling	-14.26	42.20	**24.39**
		discrete only	-14.88	40.89	24.22
	OUBMV	adaptive sampling	-19.17	49.10	**18.84**
		discrete only	-19.01	33.45	7.71

606
607

608

609

610 **Table 2.** The average accuracy of *hOUwie* parameter estimates across several model classes and
611 types as measured by root-mean-square error (RMSE). RMSE is calculated for each model type
612 by taking the square root of the mean squared error (MSE), where MSE is the average squared
613 difference between the MLE and the simulating parameters. Data is generated with $q_{ij} =$
614 $0.1, \alpha_1 = 3, \alpha_2 = 1.5, \sigma_1^2 = 0.35, \sigma_2^2 = 1, \theta_1 = 2,$ and $\theta_2 = 0.75$, and for phylogenies with 25,
615 100, and 250 taxa. Finally, model fits use either 25, 100, or 250 stochastic maps per likelihood
616 iteration. The table shown here calculates RMSE integrating over all phylogenetic tree sizes and
617 number of stochastic maps (n=8217). Dashes indicate a parameter that is not estimated for a
618 given model type. Generally, character independent (CID+) models had higher errors than
619 character dependent (CD) models. The greatest errors occurred when estimating alpha in variable
620 alpha models for both CD and CID+ model classes. Estimates of the optimum and transition
621 rates generally had the lowest errors.

Model class	Model type	RMSE q	RMSE α_1	RMSE α_2	RMSE σ_1^2	RMSE σ_2^2	RMSE θ_1	RMSE θ_2
CD	BMV	0.12	-	-	0.10	0.28	0.22	-
	OUV	0.11	1.27	-	0.15	0.33	0.05	-
	OUA	0.12	1.55	1.63	0.11	-	0.06	-
	OUM	0.13	1.49	-	0.10	-	0.07	0.13
	OUVA	0.09	1.44	1.11	0.14	0.98	0.06	-
	OUMV	0.16	1.82	-	0.16	0.32	0.07	0.17
	OUMA	0.15	2.11	2.48	0.28	-	0.12	0.50
	OUMVA	0.18	1.62	1.12	0.12	1.07	0.76	1.06
	OUBM1	0.1	2.64	-	0.08	-	0.08	-
	OUBMV	0.09	2.29	-	0.13	2.37	0.08	-
CID+	BMV	0.05	-	-	0.27	10.11	0.24	-
	OUV	0.04	1.13	-	0.32	1.83	0.05	-
	OUA	0.05	2.93	1.34	0.33	-	0.07	-
	OUM	0.09	2.53	-	0.15	-	0.44	0.20
	OUVA	0.05	1.26	1.11	0.27	13.44	0.07	-
	OUMV	0.1	2.50	-	0.16	2.12	1.30	0.68
	OUMA	0.05	8.28	1.27	0.23	-	5.88	0.8
	OUMVA	0.07	5.54	1.24	0.20	9.37	8.76	1.35
	OUBM1	0.05	3.33	-	0.32	-	0.14	-
	OUBMV	0.05	3.50	-	0.27	8.79	0.14	-

622

623

624

625

626

627 **Table 3.** Average AIC weight as the number of taxa increases for each model class. Gray cells
628 indicate the AIC weight of the generating model class. In general, as the number of taxa
629 increases the average support for the generating model class increases.

Generating model class	nTaxa	AICwt BM1	AICwt OU1	AICwt CD	AICwt CID+
CD	25	0.12	0.22	0.51	0.15
	100	0.06	0.22	0.70	0.02
	250	0.02	0.14	0.82	0.02
CID+	25	0.28	0.35	0.24	0.14
	100	0.21	0.4	0.23	0.15
	250	0.11	0.34	0.32	0.22

630

631

632 **Table 4.** Model averaged parameter estimates and standard errors for Ericaceae aridity index and
633 fruit type data. Models with higher AIC weights contribute more overall to the parameter values.
634 The units for α, σ^2, and θ are $\frac{P}{PET} \div time$, $\left(\frac{P}{PET}\right)^2$, and $\frac{P}{PET}$ respectively. P is the average annual
635 precipitation and PET is average annual potential evapotranspiration. Rates of q are measured in
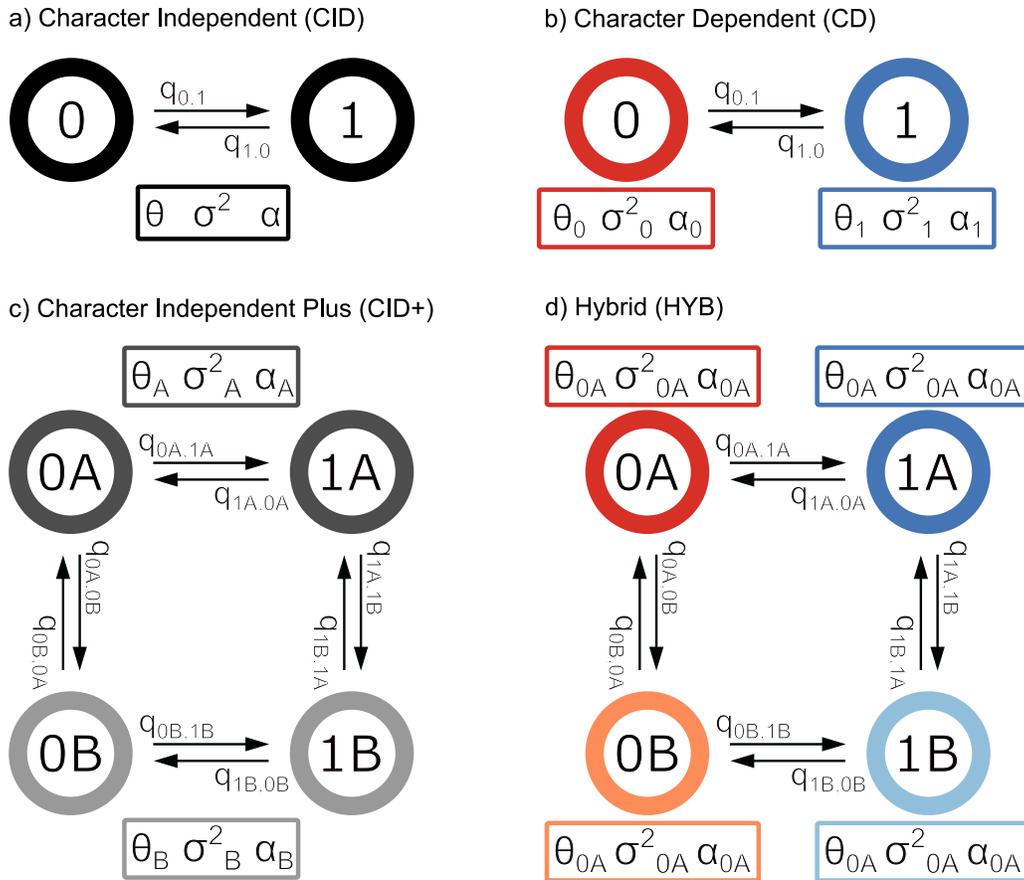636 transitions per million years.

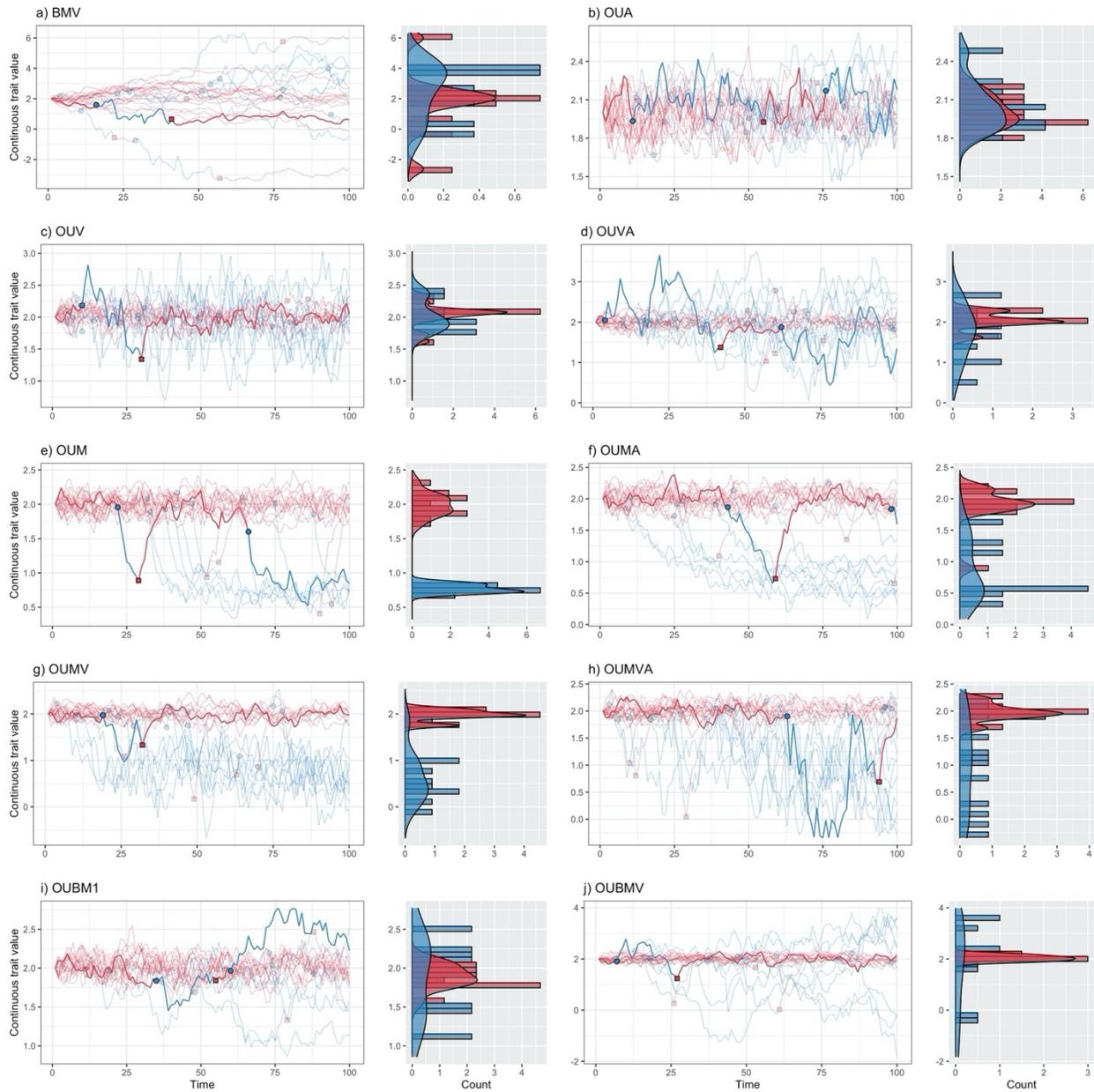| Continuous parameter estimates				Discrete parameter estimates	
	α	σ^2	θ		
Dry	0.015 (±0.0059)	0.011 (±0.0043)	0.97 (±0.011)	$q_{dry\ to\ fleshy}$	0.0015 (±0.00058)
Fleshy	0.023 (±0.011)	0.007 (±0.002)	0.81 (±0.28)	$q_{fleshy\ to\ dry}$	00036 (±0.000086)

637

Figures

Figure 1. A visual representation of the algorithm underlying the calculation of conditional node probabilities and the adaptive sampling procedure. The goal of the procedure is to produce underlying regime paintings well suited to both the discrete and continuous character. a) select the focal node for which we will be calculating the joint conditional probabilities of the discrete and continuous characters. b) on each side of the node we select a pair of tips. c) the conditional probability of the observed discrete and continuous character is calculated for each discrete regime state with an ancestral continuous value equal to θ of that regime state. d) the conditional probability of the focal node is calculated as the average probability of each regime state for all pairs of observed tips. e) the conditional probabilities are calculated for all internal nodes. This can be turned off within *hOUwie* by setting the sample_nodes argument to false. f) A stochastic map is generating using forward simulation rejection sampling. g) adaptive sampling uses the highest joint probability of previously generated underling regimes to generate a set of ancestral continuous character values. This differs from previous ancestral values because instead of assuming the value θ for each regime state, it calculates the expected value given the root state and regime mapping for that particular node. h) we repeat steps d) through g) until the joint likelihood of the set of underlying regimes does not improve.

a) Character Independent (CID)

b) Character Dependent (CD)

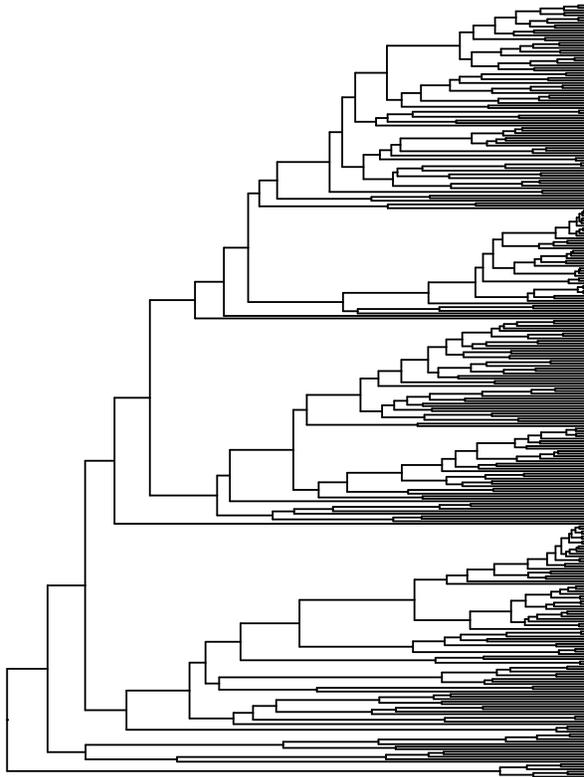**c) Character Independent Plus (CID+)**

d) Hybrid (HYB)

657

Figure 2. A state-transition diagram describing the model classes allowable in hOUwie. Each panel is comprised of observed discrete states 0 and 1 with possible hidden states A and B. Transitions between states are described with the q parameter. Continuous model parameters appear in a box below the states they describe, and their association is displayed with a subscript specific to that state. a) A simple character independent model in which the two observed states do not influence the continuous character which will have the same θ, σ^2, α throughout the phylogeny. b) A character dependent model in which the continuous character depends on the discrete character by virtue of θ, σ^2, α being associated with a particular observed discrete state. c) A character independent model with rate heterogeneity. The two observed states (0 and 1) are not directly linked to the continuous character. However, the continuous character is still allowed to have multiple θ, σ^2, α describing its evolution, but these parameters are associated with hidden states A and B. d) A hybrid model in which each combined observed and hidden state is allowed to have its own θ, σ^2, α. Under this model, the continuous character is linked to both character dependent differences (parameters associated with 0 and 1) and character independent differences (A and B). Though this diagram shows a binary observed and hidden character, either can have more states (up to 26 states for each in theory, though few datasets will have enough power to estimate the necessary number of parameters).
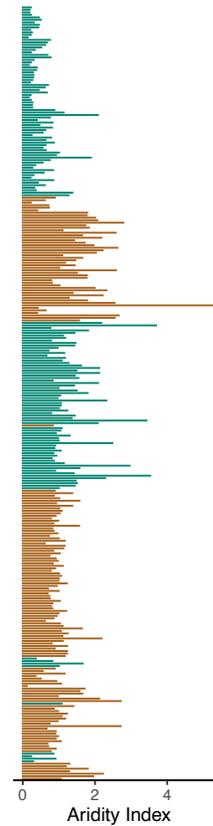
676

Figure 3. A visual representation of binary discrete character *hOUwie* model types. Discrete time forward simulations are conducted starting in the red state and the distribution of the continuous character is plotted on the right as a histogram and density plot. Each line represents a continuous character value at some time. Transitions occur at colored points and each line is colored by the current discrete state. 100 time-steps are simulated with the same parameters as our simulation study ($q_{ij} = 0.1, \alpha_1 = 3, \alpha_2 = 1.5, \sigma_1^2 = 0.35, \sigma_2^2 = 1, \theta_1 = 2,$ and $\theta_2 = 0.75$). The highlighted line was randomly chosen from the set in which at least one discrete state transition occurred.
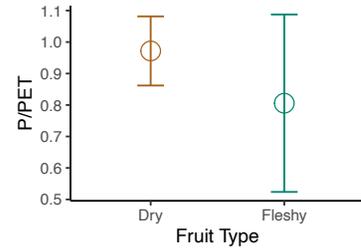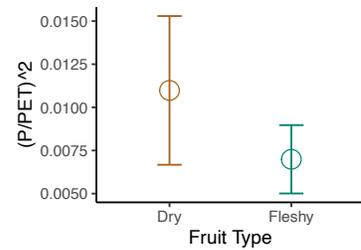
685

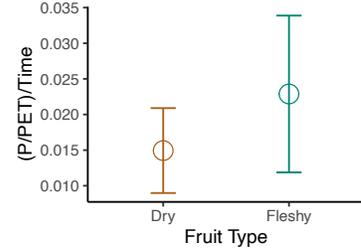**Figure 4.** a) Ericaceae phylogeny for which we had data (n=309). b) Ln aridity index dataset where each bar is colored by dry (brown) and fleshy (green) fruit type. c) Model averaged parameter estimates with standard error calculated from 100 parametric bootstraps.

689

Literature cited

690   **Literature cited**

691   Bartoszek K., Pienaar J., Mostad P., Andersson S., Hansen T.F. 2012. A phylogenetic
692   comparative method for studying multivariate adaptation. Journal of Theoretical Biology.
693   314:204–215.

694   Bastide P., Mariadassou M., Robin S. 2017. Detection of adaptive shifts on phylogenies by using
695   shifted stochastic processes on a tree. Journal of the Royal Statistical Society: Series B
696   (Statistical Methodology). 79:1067–1093.

697   Beaulieu J.M., Jhwueng D.-C., Boettiger C., O'Meara B.C. 2012. Modeling Stabilizing
698   Selection: Expanding the Ornstein–Uhlenbeck Model of Adaptive Evolution. Evolution.
699   66:2369–2383.

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

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

705   Boyko, J., & Beaulieu, J. M. (2022, April 1). Reducing the biases in false correlations between
706   discrete characters. https://doi.org/10.32942/osf.io/e2kj8

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

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

711   Butler M.A., King A.A. 2004. Phylogenetic Comparative Analysis: A Modeling Approach for
712   Adaptive Evolution. The American Naturalist. 164:683–695.

713   Caetano D.S., Harmon L.J. 2017. ratematrix: An R package for studying evolutionary integration
714   among several traits on phylogenetic trees. Methods in Ecology and Evolution. 8:1920–
715   1927.

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

719   Chapman H., Cordeiro N.J., Dutton P., Wenny D., Kitamura S., Kaplin B., Melo F.P.L., Lawes
720   M.J. 2016. Seed-dispersal ecology of tropical montane forests. Journal of Tropical
721   Ecology. 32:437–454.

722   Cressler C.E., Butler M.A., King A.A. 2015. Detecting Adaptive Evolution in Phylogenetic
723   Comparative Analysis Using the Ornstein–Uhlenbeck Model. Systematic Biology.
724   64:953–968.

725   Cybis G.B., Sinsheimer J.S., Bedford T., Mather A.E., Lemey P., Suchard M.A. 2015.
726   ASSESSING PHENOTYPIC CORRELATION THROUGH THE MULTIVARIATE
727   PHYLOGENETIC LATENT LIABILITY MODEL. Ann Appl Stat. 9:969–991.

728   Eastman J.M., Wegmann D., Leuenberger C., Harmon L.J. 2013. Simpsonian "Evolution by
729   Jumps" in an Adaptive Radiation of Anolis Lizards. arXiv:1305.4216 [q-bio].

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

731   Felsenstein J. 2004. Inferring phylogenies. Sinauer associates Sunderland, MA.

732   Felsenstein J. 2012. A Comparative Method for Both Discrete and Continuous Characters Using
733   the Threshold Model. The American Naturalist. 179:145–156.

734   Foster S., Janson C.H. 1985. The relationship between seed size and establishment conditions in
735   tropical woody plants. Ecology. 66:773–780.

736   Freyman W.A., Höhna S. 2019. Stochastic Character Mapping of State-Dependent
737   Diversification Reveals the Tempo of Evolutionary Decline in Self-Compatible
738   Onagraceae Lineages. Systematic Biology. 68:505–519.

739   Hansen T.F. 1997. Stabilizing Selection and the Comparative Analysis of Adaptation. Evolution.
740   51:1341–1351.

741   Hansen T.F. 2014. Use and Misuse of Comparative Methods in the Study of Adaptation. In:
742   Garamszegi L.Z., editor. Modern Phylogenetic Comparative Methods and Their
743   Application in Evolutionary Biology. Berlin, Heidelberg: Springer Berlin Heidelberg. p.
744   351–379.

745   Hansen T.F., Pienaar J., Orzack S.H. 2008. A Comparative Method for Studying Adaptation to a
746   Randomly Evolving Environment. Evolution. 62:1965–1977.

747   Ho L. si, Ané C. 2014a. A Linear-Time Algorithm for Gaussian and Non-Gaussian Trait
748   Evolution Models. Syst Biol. 63:397–408.

749   Ho L.S.T., Ané C. 2014b. Intrinsic inference difficulties for trait evolution with Ornstein-
750   Uhlenbeck models. Methods in Ecology and Evolution. 5:1133–1146.

751   Howe H.F., Smallwood J. 1982. Ecology of Seed Dispersal. Annual Review of Ecology and
752   Systematics. 13:201–228.

753   Huang G., Jansen H.M., Mandjes M., Spreij P., Turck K.D. 2016. Markov-modulated Ornstein-
754   Uhlenbeck processes. Advances in Applied Probability. 48:235–254.

755   Ingram T., Mahler D.L. 2013a. SURFACE: detecting convergent evolution from comparative
756   data by fitting Ornstein-Uhlenbeck models with stepwise Akaike Information Criterion.
757   Methods in ecology and evolution. 4:416–425.

758   Ingram T., Mahler D.L. 2013b. SURFACE: detecting convergent evolution from comparative
759   data by fitting Ornstein-Uhlenbeck models with stepwise Akaike Information Criterion.
760   Methods in Ecology and Evolution. 4:416–425.

761   Ives A.R., Garland T. 2010. Phylogenetic Logistic Regression for Binary Dependent Variables.
762   Syst Biol. 59:9–26.

763   Jones C.T., Youssef N., Susko E., Bielawski J.P. 2020. A Phenotype–Genotype Codon Model
764   for Detecting Adaptive Evolution. Systematic Biology. 69:722–738.

765   Khabbazian M., Kriebel R., Rohe K., Ané C. 2016. Fast and accurate detection of evolutionary
766   shifts in Ornstein–Uhlenbeck models. Methods in Ecology and Evolution. 7:811–824.

767   Labra A., Pienaar J., Hansen T.F. 2009. Evolution of Thermal Physiology in Liolaemus Lizards:
768   Adaptation, Phylogenetic Inertia, and Niche Tracking. The American Naturalist.
769   174:204–220.

770   Lemey P., Rambaut A., Welch J.J., Suchard M.A. 2010. Phylogeography Takes a Relaxed
771   Random Walk in Continuous Space and Time. Molecular Biology and Evolution.
772   27:1877–1885.

773   Levin S.A., Muller-Landau *Helene C., Nathan *Ran, Chave *Jérôme. 2003. The Ecology and
774   Evolution of Seed Dispersal: A Theoretical Perspective. Annual Review of Ecology,
775   Evolution, and Systematics. 34:575–604.

776   Lorts C.M., Briggeman T., Sang T. 2008. Evolution of fruit types and seed dispersal:A
777   phylogenetic and ecological snapshot. Journal of Systematics and Evolution. 46:396.

778   Maddison W.P. 2006. Confounding Asymmetries in Evolutionary Diversification and Character
779   Change. Evolution. 60:1743–1746.

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

782   Maddison W.P., Midford P.E., Otto S.P., Oakley T. 2007. Estimating a Binary Character's Effect
783   on Speciation and Extinction. Syst Biol. 56:701–710.

784   Mahler D.L., Ingram T., Revell L.J., Losos J.B. 2013. Exceptional convergence on the
785   macroevolutionary landscape in island lizard radiations. Science. 341:292–295.

786   Martin B.S., Bradburd G.S., Harmon L.J., Weber M.G. 2022. Modeling the Evolution of Rates of
787   Continuous Trait Evolution. :2022.03.18.484930.

788   May M.R., Moore B.R. 2020. A Bayesian Approach for Inferring the Impact of a Discrete
789   Character on Rates of Continuous-Character Evolution in the Presence of Background-
790   Rate Variation. Syst Biol. 69:530–544.

791   Middleton N., Thomas D. 1997. World atlas of desertification.. ed. 2. .

792   Mitov V., Bartoszek K., Stadler T. 2019. Automatic generation of evolutionary hypotheses using
793   mixed Gaussian phylogenetic models. Proceedings of the National Academy of Sciences.
794   116:16921–16926.

795   Moreno-Jiménez E., Plaza C., Saiz H., Manzano R., Flagmeier M., Maestre F.T. 2019. Aridity
796   and reduced soil micronutrient availability in global drylands. Nat Sustain. 2:371–377.

797   Nielsen R. 2002. Mapping Mutations on Phylogenies. Systematic Biology. 51:729–739.

798   Nunes A., Köbel M., Pinho P., Matos P., Bello F. de, Correia O., Branquinho C. 2017. Which
799   plant traits respond to aridity? A critical step to assess functional diversity in
800   Mediterranean drylands. Agricultural and Forest Meteorology. 239:176–184.

801   O'Meara B. 2008. Using Trees: Myrmecocystus Phylogeny and Character Evolution and New
802   Methods for Investigating Trait Evolution and Species Delimitation (PhD Dissertation).
803   Nat Prec.:1–1.

804   O' Meara, B., & Beaulieu, J. M. (2021, November 6). Potential survival of some, but not all,
805   diversification methods. https://doi.org/10.32942/osf.io/w5nvd

806   O'Meara B.C., Ané C., Sanderson M.J., Wainwright P.C. 2006. Testing for Different Rates of
807   Continuous Trait Evolution Using Likelihood. Evolution. 60:922–933.

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

810   Pupko T., Pe I., Shamir R., Graur D. 2000. A Fast Algorithm for Joint Reconstruction of
811   Ancestral Amino Acid Sequences. Mol Biol Evol. 17:890–896.

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

814   Rao, V., & Teh, Y. W. (2013). Fast MCMC Sampling for Markov Jump Processes and
815   Extensions. Journal of Machine Learning Research, 14(11).

816   Revell L.J. 2013. A Comment on the Use of Stochastic Character Maps to Estimate Evolutionary
817   Rate Variation in a Continuously Valued Trait. Systematic Biology. 62:339–345.

818   Revell L.J. 2014. Ancestral Character Estimation Under the Threshold Model from Quantitative
819   Genetics. Evolution. 68:743–759.

820   Revell L.J. 2021. A variable-rate quantitative trait evolution model using penalized-likelihood.
821   PeerJ. 9:e11997.

822   Revell L.J., Collar D.C. 2009. Phylogenetic Analysis of the Evolutionary Correlation Using
823   Likelihood. Evolution. 63:1090–1100.

824   Schupp E.W. 1993. Quantity, quality and the effectiveness of seed dispersal by animals.
825   Vegetatio. 107:15–29.

826   Schwery O., Onstein R.E., Bouchenak-Khelladi Y., Xing Y., Carter R.J., Linder H.P. 2015. As
827   old as the mountains: the radiations of the Ericaceae. New Phytologist. 207:355–367.

828   Steel M., Penny D. 2000. Parsimony, Likelihood, and the Role of Models in Molecular
829   Phylogenetics. Mol Biol Evol. 17:839–850.

830   Stevens P.F., Luteyn J., Oliver E.G.H., Bell T.L., Brown E.A., Crowden R.K., George A.S.,
831   Jordan G.J., Ladd P., Lemson K., Mclean C.B., Menadue Y., Pate J.S., Stace H.M.,
832   Weiller C.M. 2004. Ericaceae. In: Kubitzki K., editor. Flowering Plants · Dicotyledons:
833   Celastrales, Oxalidales, Rosales, Cornales, Ericales. Berlin, Heidelberg: Springer. p. 145–
834   194.

835   Thomas G.H., Freckleton R.P., Székely T. 2006. Comparative analyses of the influence of
836   developmental mode on phenotypic diversification rates in shorebirds. Proceedings of the
837   Royal Society B: Biological Sciences. 273:1619–1624.

838   Toljagić O., Voje K.L., Matschiner M., Liow L.H., Hansen T.F. 2018. Millions of Years Behind:
839   Slow Adaptation of Ruminants to Grasslands. Syst Biol. 67:145–157.

840   Tribble C.M., May M.R., Jackson-Gain A., Zenil-Ferguson R., Specht C.D., Rothfels C.J. 2022.
841   Unearthing modes of climatic adaptation in underground storage organs across Liliales.
842   bioRxiv.:2021.09.03.458928.

843   Uyeda J.C., Harmon L.J. 2014. A novel Bayesian method for inferring and interpreting the
844   dynamics of adaptive landscapes from phylogenetic comparative data. Systematic
845   biology. 63:902–918.

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

848   Vasconcelos, T., Boyko, J. D., & Beaulieu, J. M. (2021). Linking mode of seed dispersal and
849   climatic niche evolution in flowering plants. Journal of Biogeography.

850   Vasconcelos T., O'Meara B.C., Beaulieu J.M. 2022. A flexible method for estimating tip
851   diversification rates across a range of speciation and extinction scenarios. Evolution.

852   Venable D.L., Brown J.S. 1988. The Selective Interactions of Dispersal, Dormancy, and Seed
853   Size as Adaptations for Reducing Risk in Variable Environments. The American
854   Naturalist. 131:360–384.

855   Westoby M., Leishman M., Lord J. 1996. Comparative ecology of seed size and dispersal.
856   Philosophical Transactions of the Royal Society of London. Series B: Biological
857   Sciences. 351:1309–1318.

858   Yang Z. 2006. Computational molecular evolution. Oxford University Press Oxford.

859

860   **Supplementary Tables**

861   **Table S1.** AIC weights summarizing the average support for each model class when they are the
862   generating model. Data is generated with $q_{ij} = 0.1, \alpha_1 = 3, \alpha_2 = 1.5, \sigma_1^2 = 0.35, \sigma_2^2 = 1, \theta_1 = $
863   2, and $\theta_2 = 0.75$ for phylogenies with 25, 100, and 250 taxa and model fits using either 25, 100,
864   or 250 stochastic maps per likelihood iteration. When the generating model class is character
865   dependent (CD) or character independent (CID+) we expect that the AICwt will be highest for
866   that model when fit. Character dependent models generally show that pattern, however CID+
867   models generally perform poorly. An additional concern is datasets simulated by a character
868   independent model with rate heterogeneity (datasets generated by a CID+ model) are best fit by
869   CD models – which would be a spurious correlation. Although there was often some signal of
870   character dependence in these models (AICwt of CD when CID+ is generating), most of the AIC
871   weight was for simple character independent models (BM1 or OU1).

| Generating model class	Generating model type	AICwt of BM1	AICwt of OU1	AICwt of CD	AICwt of CID+	Proportion generating model chosen as best
CD	BMV	0.18	0.17	**0.64**	0.02	0.62
	OUV	0.03	0.22	**0.74**	0.02	0.73
	OUA	0.07	**0.56**	0.31	0.06	0.15
	OUM	0.04	0.02	**0.9**	0.04	0.92
	OUVA	0.04	0.21	**0.7**	0.06	0.7
	OUMV	0.02	0.02	**0.93**	0.03	0.95
	OUMA	0.12	0.15	**0.64**	0.09	0.66
	OUMVA	0.05	0.13	**0.76**	0.06	0.76
	OUBM1	0.19	**0.58**	0.13	0.10	0.08
	OUBMV	0.07	0.20	**0.71**	0.02	0.73
CID+	BMV	**0.36**	0.28	0.33	0.03	0.01
	OUV	0.04	**0.49**	0.43	0.04	0.01
	OUA	0.06	**0.56**	0.37	0.02	0
	OUM	0.21	0.09	0.03	**0.67**	0.71
	OUVA	0.07	**0.55**	0.35	0.04	0.03
	OUMV	0.24	0.19	0.14	**0.44**	0.44
	OUMA	**0.41**	0.40	0.13	0.06	0.06
	OUMVA	0.24	**0.39**	0.21	0.16	0.15
	OUBM1	0.24	**0.55**	0.16	0.05	0.01
	OUBMV	0.23	**0.37**	0.30	0.10	0.08

872
873
874
875
876
877
878

879

880 **Table S2:** Modeling results from the 25 models fit to Ericaceae aridity index and fruit type data.
881 Model classes are character independent without rate heterogeneity (CID), character dependence
882 (CD), character independence with rate heterogeneity (CID+), and mixed character dependent
883 and character independence (HYB). Character dependent models suggest that climatic niche
884 evolution will be linked to the fruit type. We found substantial support for OUVA (variable σ^2
885 and α) and OUMVA (variable σ^2, α, and θ) models. np is the number of freely estimated
886 parameters. lnLik is the joint likelihood of the MLE. DiscLik and ContLik are the marginal
887 likelihood of the discrete and continuous datasets respectively, given the maximum joint
888 likelihood estimate of the parameters. AIC is the Akaike information criterion, ΔAIC is the
889 difference from the best fit model measured as the difference between each model's AIC, and
890 AICwt is the relative support for each model.

| Model class	Model type	np	lnLik	DiscLik	ContLik	AIC	ΔAIC	AICwt
CID	BM1	4	-243.89	-32.62	-206.67	495.78	39.07	0
	OU1	5	-225.5	-32.62	-188.28	461.01	4.30	0.05
CD	BMV	5	-243.78	-32.62	-207.08	497.56	40.85	0
	OUV	6	-225.49	-32.62	-188.47	462.98	6.27	0.02
	OUA	6	-224.95	-32.58	-189.48	461.9	5.19	0.03
	OUM	6	-224.12	-32.57	-187.79	460.24	3.53	0.07
	OUVA	*7*	*-221.62*	*-32.58*	*-184.44*	*457.24*	*0.53*	*0.32*
	OUMV	7	-224.05	-32.62	-188.15	462.10	5.39	0.03
	OUMA	7	-223.21	-32.58	-187.97	460.42	3.71	0.06
	OUMVA	*8*	*-220.35*	*-32.60*	*-183.27*	*456.71*	*0*	*0.41*
	OUBM1	5	-243.84	-32.57	-206.67	497.68	40.97	0
	OUBMV	6	-243.79	-32.61	-206.99	499.57	42.87	0
CID+	BMV	7	-244.80	-33.11	-205.78	503.59	46.89	0
	OUV	8	-228.77	-32.98	-190.16	473.55	16.84	0
	OUA	8	-226.42	-33.17	-188.53	468.84	12.13	0
	OUM	8	-226.43	-33.32	-189.07	468.87	12.16	0
	OUVA	9	-244.38	-33.43	-202.12	506.76	50.05	0
	OUMV	9	-225.20	-33.39	-182.88	468.39	11.68	0
	OUMA	9	-225.57	-32.68	-189.92	469.14	12.43	0
	OUMVA	10	-227.39	-33.13	-185.15	474.79	18.08	0
	OUBM1	7	-244.44	-33.16	-206.67	502.88	46.17	0
	OUBMV	8	-225.58	-32.71	-186.58	467.17	10.46	0
HYB	BMS	9	-244.46	-33.08	-204.83	506.93	50.22	0
	OUM	10	-224.12	-32.67	-188.99	468.23	11.52	0
	OUMVA	16	-226.56	-33.03	-179.11	485.13	28.42	0

891

Supplementary Figures
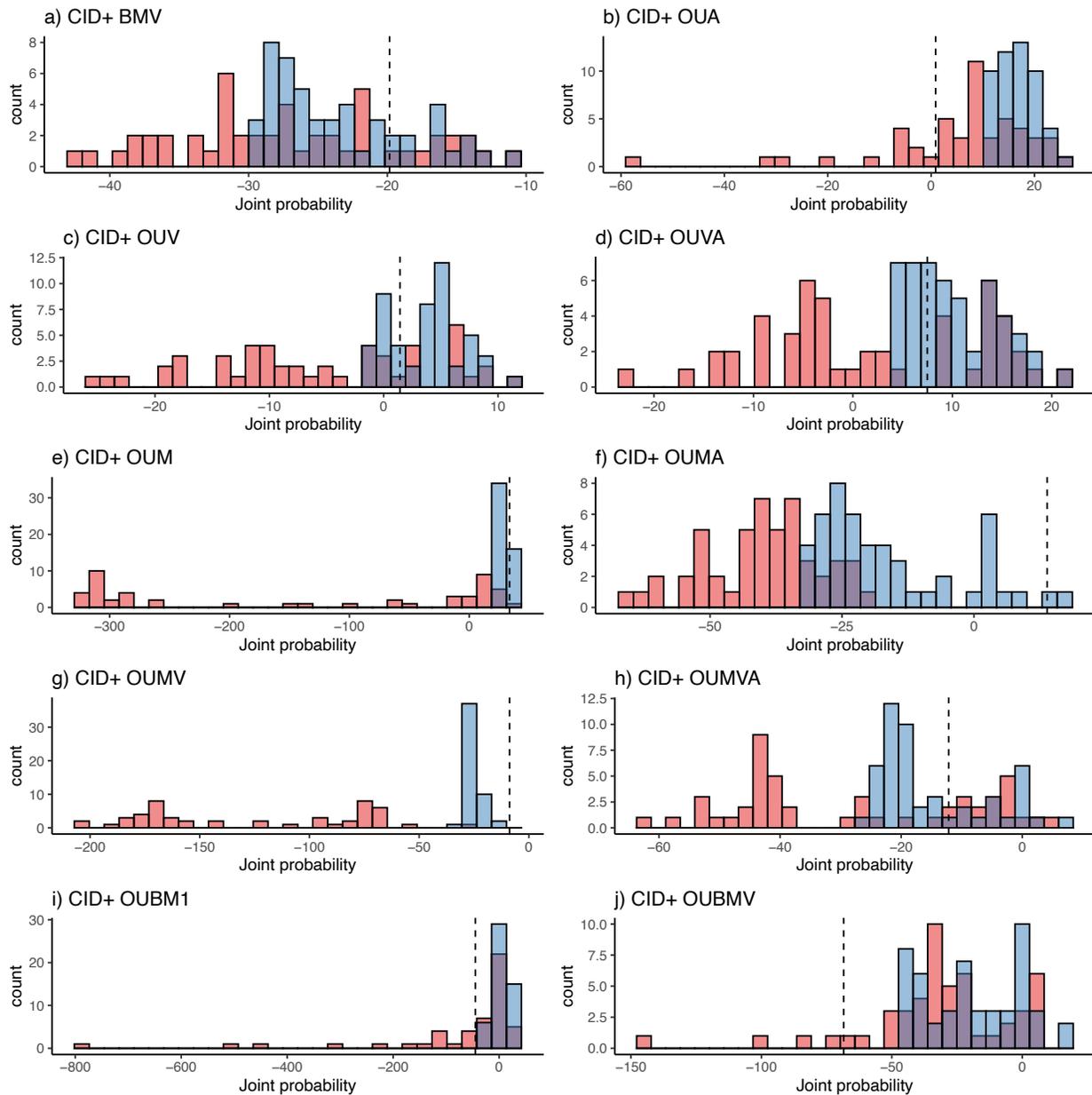
Figure S1. Overlapping histograms comparing the effectiveness of the adaptive sampling procedure (blue) and standard discrete only sampling (red) of maps. Regardless of the sampling procedure, all probabilities are calculated in the same way and so any differences in probabilities reflects each procedure's ability to generate appropriate mappings. 50 stochastic mappings are used to calculate the likelihood of the parameters. For each model type, data are simulated following our methods with $q_{ij} = 0.1, \alpha_1 = 3, \alpha_2 = 1.5, \sigma_1^2 = 0.35, \sigma_2^2 = 1, \theta_1 = 2,$ and $\theta_2 = 0.75$. Dashed line likelihood under generating map.

902
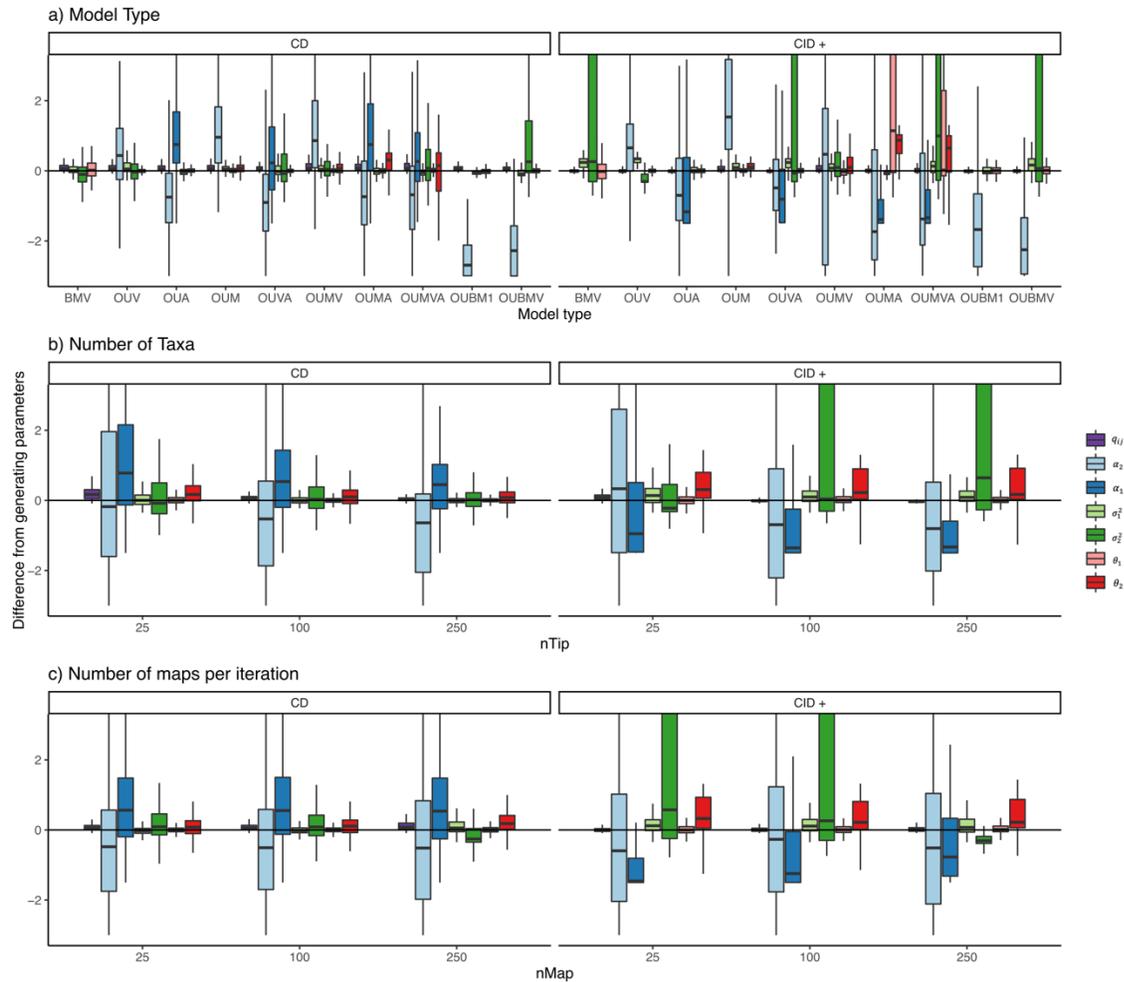
Figure S2. The raw difference of the maximum likelihood parameter estimates and the
generating values depending on the a) model type, b) number of taxa in the dataset, and c)
number of stochastic maps per iteration of the likelihood search. Generally, variable alpha
models had the highest biases with alpha being consistently underestimated. As the number of
taxa increased, estimation of CD model parameters was estimated with less error. The number of
maps per iteration had the greatest effect on character independent models with rate
heterogeneity.
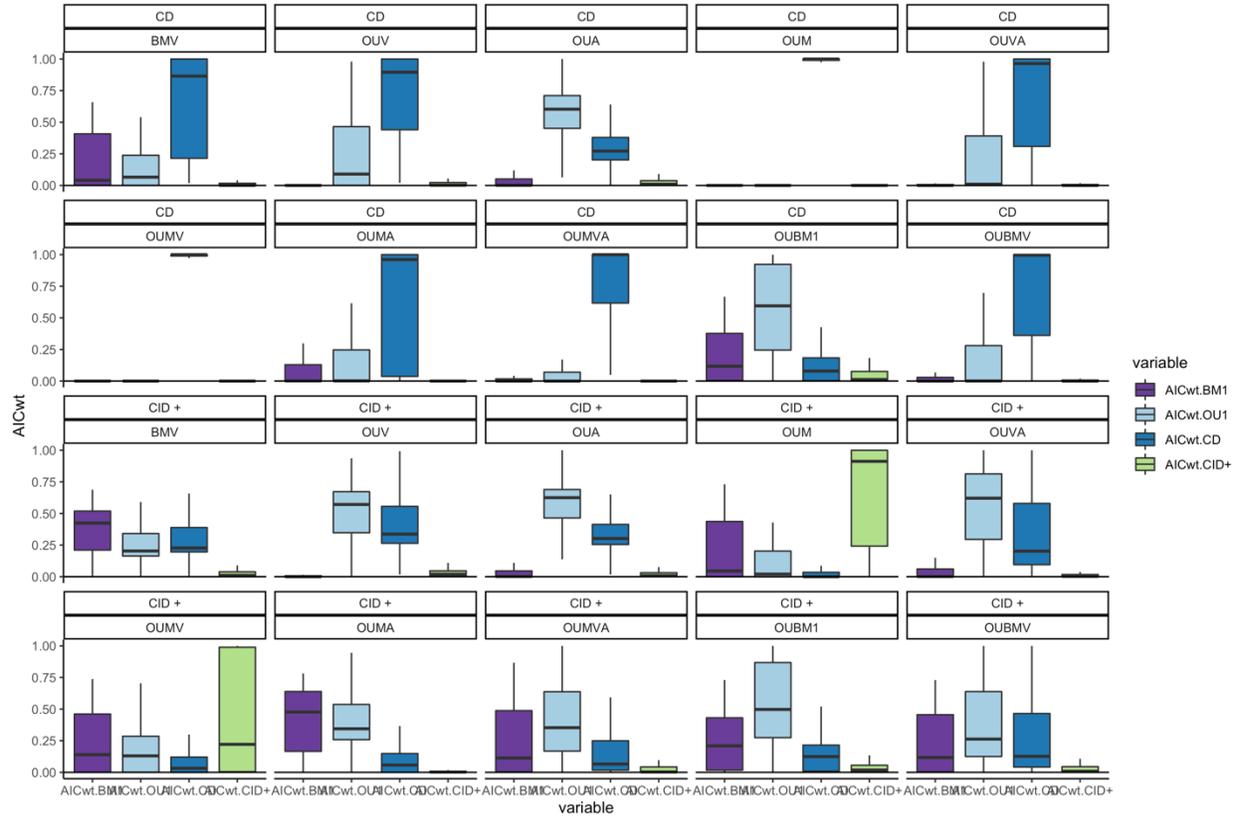
911
912 **Figure S3.** AIC weights summarizing the average support for particular model classes and model
913 type when they are the generating model. Headings indicate the generating model type and
914 model class. Data was generated with $q_{ij} = 0.1, \alpha_1 = 3, \alpha_2 = 1.5, \sigma_1^2 = 0.35, \sigma_2^2 = 1, \theta_1 =$
915 $2,$ and $\theta_2 = 0.75$ for phylogenies with 25, 100, and 250 taxa and model fits using either 25, 100,
916 or 250 stochastic maps per likelihood iteration. When the generating model class is character
917 dependent (CD) or character independent (CID+) we expect that the AICwt will be highest for
918 that model when fit.
919