Jointly Modeling the Evolution of Discrete and Continuous Traits

James D. Boyko1\*, Brian O’Meara2, Jeremy M. Beaulieu1

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

*2Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, TN, USA*

*\*To whom correspondence should be addressed (jboyko@uark.edu).*

Running head: Jointly Modeling Discrete and Continuous Traits

Key words: rate variation

*Abstract*

Rates of evolution have changed throughout the history of life and produced the rich diversity of morphology, behaviour, and ecology that characterizes the biodiversity we see today.

Questions of how plant life habit influence genome size evolution (Beaulieu et al. 2012), how substrate use alters limb length evolution (Mahler et al. 2013), or how tooth morphology slowly changes in response to habitat and diet (Toljagić et al. 2018) are evidence of a recurring interest in testing whether evolutionary changes in one variable is linked to change in another. To study questions such as these, one very common phylogenetic comparative approach is to employ an Ornstein-Uhlenbeck (OU) model, which assumes distinct regimes, described by the evolution of a discrete character, are known completely *a priori* (e.g., Butler and King 2004; Hansen et al. 2008; Beaulieu et al. 2012), or assumes that “shifts” in regimes can be inferred directly from the distribution of the continuous trait (e.g. Ingram and Mahler 2013; Uyeda and Harmon 2014; Khabbazian et al. 2016). While these approaches are practical, they assume that the combination of independently estimated discrete and continuous models will produce a joint model, but with the discrete trait being the driving force underlying the evolution of the continuous character. However, evolution is rarely as simple as cause and effect, and we suspect that as often as a discrete character causes change in the continuous character, continuous characters also influence discrete character evolution, or at the very minimum, can provide information about this joint evolution.

Progress along these lines has mostly involved acknowledging uncertainty in the evolution of the discrete character by way of stochastic maps. That is, a large set of alternative reconstructions of the discrete character are obtained completely uninformed by the continuous trait’s evolution, then the likelihood of the continuous character becomes the average of the likelihoods across these maps (e.g., Revell 2012). The advantage of this approach is that there is an explicit model for how regimes change through time, but the evolution of these regimes remains entirely independent of the continuous trait, and the probability of these regimes is not explicitly considered. For example, it may be possible that the model which best fits the discrete data generates stochastic maps which do not fit the continuous data. A promising approach was recently described for detecting adaptive codon evolution (Jones et al 2020), where a set of maps obtained for a discrete phenotype under a standard Markov process is optimized along with parameters associated with genotype properties, thus forcing an emergent dependency between the two. Similarly, May and Moore (2020) developed a joint model for discrete and continuous characters under a state-dependent Brownian motion model. Their approach takes advantage of a Bayesian framework through the use of priors in order to accommodate variation in the “background” rate of evolution in the continuous trait (i.e., rate variation across lineages that is independent of the discrete character under consideration). The novel Bayesian pipeline recently developed by Tribble et al. (2021) is the first attempt that we are aware of for jointly modeling discrete and continuous traits under an OU framework. Their approach samples discrete stochastic mappings informed by the discrete trait along with regime mappings which were informed by the continuous trait while accounting for the potential of hidden variation. This method allows for a more effective test of correlation between the discrete and continuous characters. One drawback, however, is that they do not explicitly account for the joint probability of the discrete and continuous parameter estimates together. They assume that the combination of independently estimated discrete and continuous models produces a joint estimate.

Here we propose an explicitly joint estimate of the likelihood for a discrete and continuous character by combining the probability of the continuous character given a particular regime and the probability of that discrete regime painting, integrated over many regime paintings. Specifically, we combine hidden Markov models of discrete character evolution (Beaulieu et al. 2013; Boyko and Beaulieu 2021) with generalized Ornstein-Uhlenbeck models (Hansen 1997; Butler and King 2004; Hansen et al. 2008; Beaulieu et al. 2012; Ho and Ané 2014a). We demonstrate how our framework, which we call “hOUwie”, can be used to test hypotheses of correlated evolution between discrete and continuous characters while also accounting for hidden character states and unobserved variation. Finally, we apply several hOUwie models to test the correlated dynamics of the mode of seed dispersal and climatic niche evolution and compare our results to those that did not account for the potential joint evolution of discrete and continuous variables.

**Materials and Methods**

*The hOUwie model*

Our model is composed of two processes: one that describes the evolution of a discrete character and the other describes the evolution of a continuous character. To model the evolution of a single continuous character we use an Ornstein-Uhlenbeck (OU) model (Hansen 1997; Butler and King 2004; Hansen et al. 2008; Beaulieu et al. 2012; Ho and Ané 2014a). Formally, the OU process is an Itô diffusion satisfying:

.

This model combines the stochastic evolution of a trait through time with a deterministic component which models the tendency for a trait to evolve towards an optimum. In this model, the value of a trait, , is pulled towards an optimum, , at a rate scaled by the parameter . The optimum, , is a piecewise constant on intervals and takes values in a finite set. This can represent the set of “selective regimes”, “regimes”, or Simpson’s “adaptive zones” (Cressler et al. 2015), though it is consistent with a variety of true underlying microevolutionary models (Hansen 2014). Additionally, random deviations are introduced by Gaussian white noise , which is distributed as a normal random variable with mean zero and variance equal to . Thus, is a constant describing the rate of stochastic evolution away from the optimum. We use the set of extensions introduced by Beaulieu et al. (2012) and implemented in the R package OUwie. This allows for multiple primary optima in which both the pull strength () and the rate of stochastic evolution () can vary across the phylogeny. However, the algorithm used to calculate the likelihood described in Beaulieu et al. (2012) involves matrix inversion -- a computationally costly procedure. Therefore, we implement a linear-time computation of the likelihood of Gaussian trait models following (Ho and Ané 2014a). To do this, we first transform the phylogeny such that its variance covariance matrix, , is 3-point structured. We can write the variance covariance matrix of the untransformed phylogeny as , where following Beaulieu et al. (2012) and Ho and Ané (2014),

and,

where, is the distance from the root to the beginning of the selective regime () for the number of selective regimes along the path from the root to the last common ancestor of and , or from the root to the terminal tip. Our transformed phylogeny now has a variance covariance matrix and diagonal matrix . We can then calculate the quadratic quantities and determinant of (Ho and Ané 2014a). The probability of our continuous trait is given by

where is the number of tips in the phylogeny (, is the continuous trait value of each species, and is the expected value of each species given the continuous trait model – calculated following equation (11) of Beaulieu et al. (2012), *D* is the discrete character data, is a particular regime mapping, and are the parameters of the hOUwie model.

Next, we describe the calculation of the probability of the underlying regime structure, , which is the joint probability of discret­e characters () and stochastic mapping (). This calculation is analogous to the pathway likelihood of Steel and Penny (2000). Recently, May and Moore (2020) suggested that the joint probability of a regime structure and the discrete character is the product of the probabilities of exponentially distributed waiting times. By this definition, branch lengths are the sum of waiting times. However, under this definition, when we calculate the probability of starting and ending a branch in state *i*, the likelihood of a regime structure was unaffected by the number of transitions (Supplemental Materials), even though the maximum likelihood estimate should be zero transitions (O’Meara 2008). We suspect this is an issue of treating the probability density functions independently, when their convolution is required, because the time of the second transition depends on the timing of the first as well as the total branch length available. Ultimately, the effect of this inaccuracy is that the number of transitions has no influence on the probability of a branch’s state reconstruction and that the sum of the joint probabilities will be greater than the marginal probability. For further details, and R code necessary to demonstrate these issues, we refer readers to the Supplemental Materials.

To calculate the probability of discret­e characters () and stochastic mapping () we instead use an approximation. Our approximation relies on a finite number of degree-2 internodes and uses the standard Chapman-Kolmgorov equation to calculate the probabilities of beginning in a particular state and ending in state (Pagel 1994) and is identical to a marginal probability of a set of state reconstructions (Yang 2006). As the number of internodes increase, the amount of time between nodes decreases and the approximation improves (Rao and Teh 2013). The joint probability of a regime structure and the discrete character i

where the instantaneous rate matrix (, is the phylogeny, is the root state probability (Pagel 1994; Yang 2006; Maddison et al. 2007), is the number of external nodes (tips), internal nodes, and internodes (degree-2 nodes) summed, indicates a particular branch, , where is an indicator function which ensures that we only use the probability of states indicated by the specific the regime mapping instead of summing over all possible state combinations. The continuous character probability requires the discrete state(s) to be defined along the entire branch, thus we place transitions halfway between any two nodes.

For each set of parameters evaluated during the maximum likelihood search, a set of stochastic mappings are generated to evaluate the discrete and continuous likelihoods (Fig. 3). To do this, we first approximate of the conditional state probabilities at nodes. The conditional state probability, unlike the more common marginal reconstruction or joint state reconstruction (Pupko et al. 2000; Felsenstein 2004; Yang 2006), calculates the probability that a node has a particular state value conditioned only on the observations of its descendants. For a particular focal node, we calculate the probability of the observing all pairwise descendant values given the OU model parameters, integrated over all possible rootward node states, and observed tipward discrete states (Fig 3d, see Supplemental for more detail). Although this is only an approximation of the conditional state probabilities, it proves to be an essential improvement over the typical procedure of sampling many stochastic maps based solely on the discrete process (Fig. 4). Next, the conditional probabilities of states at nodes are sampled starting with the root. Once the root is sampled, descendent states are sampled based on both the conditional ancestral values and the sampled ancestral state. This is achieved by multiplying the conditional probability of the node states by the probability of starting in the sampled rootward ancestral value and ending in any of the tipward states (the latter is calculated using familiar matrix exponentiation methods (e.g., Pagel 1994). Finally, under usual stochastic mapping procedures we would use rejection sampling (Nielsen 2002; Rao and Teh 2013) to simulate a path between the sampled rootward and tip ward nodes. However, for increased computational efficiency, we opt to place transitions at pre-defined internodes. Thus, under our procedure, after nodes and internodes are sampled in step two, mappings are evaluated to ensure consistency with the discrete model (i.e., impossible transitions do not occur) and branches are painted based on the sampled nodes with transitions occurring half-way between nodes.

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

where summing over all generated maps (), is the probability of the continuous character () given the discrete character data (), mapping (), hOUwie parameters (), and phylogeny (). is the joint probability of the discrete character data () and stochastic mapping () given the hOUwie parameters () and phylogeny ().

*The hOUwie model space*

Our simulation studies examined 22 possible hOUwie model structures for a binary discrete character, although the possible number of models is significantly higher because any number of discrete characters and states can be modeled together. For the discrete component of the model, we assumed that transitions between the observed characters were equal. We constrained transitions between hidden states to be the same for observed states, but this constraint can be relaxed if desired. The continuous model structures allowable in hOUwie are a generalized form of those allowed in OUwie and now include models in which only alpha varies (OUA), only sigma varies (OUV), and combinations of an OU and BM process (OUBM1 and OUBMV). We note that the OUBM1 model within hOUwie, differs from The Ornstein–Uhlenbeck Brownian-motion (OUBM) model presented in Hansen et al. (2008) and Bartoszek et al. (2012) since latter models are of multiple continuous characters, rather than different processes describing the same continuous character.

The potential model structures range from completely character-dependent to character-independent. Character-dependent (CD) models are models in which any continuous OU parameter differs between observed discrete state, whereas character-independent models (CID) test whether observed discrete states can be described by the same OU parameters. There are two types of character-independent model (Fig. 1). First, character-independent models include structures where there are no differences between any OU parameters. Under this model the entire evolutionary history of the clade can be described by a single alpha, sigma, and optimum value (Fig. 1a). To combat this unrealistic assumption we introduce a character-independent model which allows for differences in the OU parameters to depend upon an unobserved hidden state (CID+) and has been shown to correct for the bias towards detecting correlation (Boyko and Beaulieu 2022). This addition allows for heterogeneity within the evolutionary process without the necessity of it being linked to a focal trait (Fig. 1c). In total we examine 22 unique model structures (2 CID, 10 CD, and 10CID+).

*Simulation study*

For each of the 22 hOUwie model structures, we simulated 50 datasets for phylogenies of 25, 100, and 250 taxa for a total of 3300 unique datasets. Phylogenies were pure birth phylogenetic tree with , and rescaled tree height to 1. The root state was fixed to state 1. The parameters used to generate a phenotypic dataset depend on the structure of the generating model. For example, an OUM model and OU1 model can have identical , but they must differ in or else OUM will collapse into OU1 (model structures associated with model name are shown in Table 1). The simulating parameters were chosen to match Beaulieu et al. (2012) with . Once a phylogeny and phenotypic dataset were simulated, we fit our models to assess parameter estimation accuracy and model selection power. Although this represents a small subset of the potentially vast parameter space available to OU models, the behavior of these models has been thoroughly characterized and thus we chose parameters within the range of typical identifiability (Beaulieu et al. 2012; Ho and Ané 2014*a*; Cressler et al. 2015). Additionally, because hOUwie uses a variable number of mappings, we evaluate changing the number of stochastic maps. We fit each model using 25, 100, and 250 stochastic mappings per likelihood evaluation. Each dataset was evaluated using the true generating model, a BM1, an OU1, and either the character-dependent or character-independent counterpart to the generating model. For example, if the data was simulated under a character-dependent OUM model where the value of depend on the observed character, a character-independent OUM model would also be fit as part of the model set. Under the CID+ OUM model, a variable is still allowed, but it is unlinked to the focal character and thus should provide a more reliable character independent null hypothesis than BM1 or OU1 (Beaulieu and O’Meara 2016; Uyeda et al. 2018; May and Moore 2020; Boyko and Beaulieu 2022).

*The impact of climatic variables on seed dispersal*

For sedentary organisms, such as plants, dispersal is mainly limited to a brief stage of their life cycle and mediated mainly through the movement of seeds (Levin et al. 2003). Generally, the expectation is that seeds dispersed by frugivores are going to be dispersed to environments more like their parents’ environment, whereas abiotically dispersed seeds are likely to be more erratic in their dispersal patterns (Schupp 1993; Westoby et al. 1996). Furthermore, it has been proposed that adaptations for frugivorous dispersal is linked to tropical and subtropical biomes. This is because in these warmer and wetter habitats, large trees create shady environments where competition for light is more important. A shadier habitat then imposes a selective pressure for larger seeds because more nutrients are needed for germination (Foster and Janson 1985). However, the evolution of larger seeds comes with a tradeoff as they have a significantly lower dispersal potential (Howe and Smallwood 1982). Thus, we might expect that the climatic variables of a habitat influence the probability of transitioning between abiotic and biotic modes of dispersal, with transition rates from abiotic to biotic being greater in less arid environments.

Here we use dry or fleshy fruit morphology as a proxy for abiotic or biotic seed dispersal (Lorts et al. 2008) to evaluate three predictions outlined in Vasconcelos et al. (2021), but specifically measuring the aridity index. First, we expect that the climatic optima for fleshy fruits will be more humid compared to dry fruits (). Second, we expect that dry fruits will have faster rates of climatic niche evolution (). Finally, we expect that the climatic niches of fleshy fruits will be more conserved through time (). We apply several hOUwie models to test these hypotheses and compare our results to those discussed in Vasconcelos et al. (2021). We expect that any differences found between this study and Vasconcelos et al. (2021) are because we can explicitly account for the joint probability of the discrete and continuous characters. We choose Ericaceae specifically because Vasconcelos et al. (2021) found two counter-intuitive results. First, they found that the phenotypic optima of dry fruits were more humid than fleshy fruited lineages. Second, they found that the rate of climatic evolution was greater in fleshy fruits than dry fruits.

We include 25 hOUwie models within our model set: 2 CID, 10 CD, 10 CID+, and 3 HYB. *Gaultheria* is technically a dry-fruited genus within Ericaceae but has a persistent fleshy calyx that attracts frugivores (Stevens et al. 2004). However, since we are interested in the association between dispersal and fruit type, we code this as fleshy fruited within our dataset. Models are evaluated using the sample size corrected Akaike Information Criterion (AICc) and model averaging is conducted when discussing how our results relate to our hypotheses (Burnham and Anderson 2002). Next, we conduct a parametric bootstrap of 100 simulated datasets to evaluate the standard error of our model averaged parameter estimates. Finally, we evaluate the expected values, the optima, as they relate to our hypotheses and compare our results to Vasconcelos et al. (2021).

**Results**

*Simulation study*

For character independent models, our heuristic adaptive sampling algorithm consistently produced more probable mappings than using purely discrete mappings for all models examined. On average, adaptive sampling produced mappings which were 38 log likelihood units better than purely discrete sampling when examining joint probabilities. This was driven primarily by the improved continuous probabilities which were on average 38.4 log likelihood units better. In contrast, the discrete probability of each mapping was similar with discrete-only simulations producing maps that were on average 0.39 log likelihood units better (Table 1; Figure 3). For character-dependent models, the difference was negligible (not shown). This is because when the discrete and continuous character are linked, discrete-only mappings will match the continuous character’s distribution well.

Most character dependent models showed low parameter estimation errors across all model types. Like previous studies, the RMSE was largest for alpha at 1.76 and 1.65 (if variable alpha) and errors were generally higher for more complex models. All other parameters had relatively similar RMSE, ranging from 0.1 for discrete the rate to 0.75 for sigma^2\_2. The BMV, OUV, OUA, and OUM models generally had the lowest errors, but there were some biases present. Most notably, alpha was biased upwards for OUM and OUV models and under variable alpha models (OUA, OUMA, OUVA, OUMVA), the difference between the alpha estimates tended to be larger than the generating parameter difference. The more complex models had larger error variances but showed similar biases as the simple models. Finally, OUBM models showed a significantly downward biased alpha, suggesting BM like processes (Figure 5; Table 2).

Character independent models with rate heterogeneity models generally performed well in terms of parameter estimates, but as expected, due to their inherit uncertainty, CID+ models had larger errors than CD models. The largest error was estimates of sigma^2\_2 which had an RMSE of 8.5, although the median error value was only 0.03, suggesting that the large RMSE is driven by a long rightward tail of the estimates. Similar to CD models, alpha\_1 and alpha\_2 consistently showed the largest RMSE at 3.6 and 1.2. In general, alpha was underestimated with medians of -0.4 and -1.4 below the simulating values of 3 and 1.5. This means that models for CID+ models tended to be more BM like even under an OU generated data (Figure 5; Table 2).

Increasing the number of taxa examined improved both CD and CID+ performance. Alpha RMSE was nearly cut in half between when moving from 25 tips to 250 tips from 5.2 to 2.8 under CID+ models. Nonetheless, some parameters continued to be estimated poorly, such as sigma2\_2. Interestingly, increasing the number of stochastic maps improved CID+ performance, but did not substantially improve estimation under CD models (Fig. 5bc).

Generally, evidence of CD when it was the generating model was consistent across all model types. The lowest support for the OUA and OUBM1 models at an average AICwt of 0.31 and 0.13. For complex models, such as OUMVA, model support for was 0.81 and highest for OUMV at 0.97. CID+ models fared worse in terms of generating consistent support even when they were the generating model. Models which were difficult to estimate under character dependence were difficult to find consistent support for under character independence. The most extreme case was OUA model for which CID+ model was never chosen as the best supported model. However, models which performed well for CD tended to perform well under CID+. For example, OUM models garnered consistent support when with an average AICwt of 0.733 (Table 3; Figure 6).

For both CD and CID+ models, support improved when increasing the number of tips analyzed. Support for a CD model when CD was the generating model increased from to to for 25, 100, 250 tips and support for a CID+ model when it was the generating model increased from to to . Similarly, increasing the number of stochastic maps generally improved the fit, but not as much as increasing the number of tips. We found that false evidence of correlation (as measured by the average AICwt of a character-dependent model when character-independence was the generating model) was generally not an issue for variable theta models. Variable theta models had average AICwts for false character-dependence ranging from 0.03 to 0.23 and for none of our simulations models was a CD model best supported. Under a simple OUM model, CID+ models helped correct any potential bias with an average AICwt of 0.68. However, false evidence of correlation was an issue for variable sigma and alpha models. False support for CD ranged from 0.34 to 0.44 when theta was fixed and alpha and/or sigma varied. Although CID+ models did not garner much support when these models were fit, OU1 and BM1 models served as reasonable null hypotheses in these cases. In general, we found that when CID models were the generating model, evidence of CID was strongest and when CD models were the generating model, evidence of character dependence was strongest. This suggests that the effect of rate heterogeneity causing false correlations is not as pronounced as other comparative methods (Maddison and FitzJohn 2015; Rabosky and Goldberg 2015).

*Seed dispersal and climatic evolution*

We found evidence of a character dependent model over either a simple or hidden state character independent model, suggesting a link between the climatic niche of Ericaceae lineages and their fruit type (Table 6). The best supported models were OUMVA and OUVA with AIC weights of 0.41 and 0.32 respectively. This suggests that there were character dependent differences in phenotypic optima, rates of evolution, and overall phylogenetic signal. To evaluate support for our hypotheses we examined the model averaged parameter estimates (Table 7). The estimated optimum 0.81 for fleshy fruits suggests a more arid environment for their optimal habitat, and the 0.97 AI of dry fruits corresponds to a more humid environment (Middleton and Thomas 1997). However, both optima correspond to non-dryland humid environments. Both and interact to create tip variance, so in addition to , we measured stationary variance . As predicted, we found that Ericaceae lineages with dry fruits were more variable in their climatic niche evolution (, ) compared to fleshy fruits (, ). Additionally, the phylogenetic signal of fleshy fruited lineages was greater than dry fruited lineages (). This corresponds to phylogenetic half-lives of and which are 38% and 25% of the total tree height respectively. Transitions to fleshy fruit occurred at 0.0015 transitions per million years which is more than 4.3 times faster than transitions to dry fruits (0.00035 transitions per million years). Given the total branch length in the tree is , we would expect 15.6 transitions to fleshy fruit and 3.6 transitions to dry fruits to have occurred throughout the history of Ericaceae. Finally, on average lineages were in more arid environments than predicted by the model (average difference of 0.19 AI), with some species expected to be in much more humid environments (difference between current AI and optimal AI ranged from -4.4 to 0.85; Figure 7).

**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. Here we describe the hOUwie model and demonstrate how it can be 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 few 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 2017) or more generalized Ornstein-Uhlenbeck models where parameters are allowed to vary based on an underlying regime mapping (e.g., Butler and King 2004; Bartoszek et al. 2012; Beaulieu et al. 2012). In contrast, several methods have focused on the development of data driven, shift-detection, methods. These methods utilize an Ornstein-Uhlenbeck process to automatically detect where in the phylogeny evolutionary rates and phenotypic optima shift (Ingram and Mahler 2013b; Uyeda and Harmon 2014; Khabbazian et al. 2016; Bastide et al. 2017). Furthermore, some recently developed methods have allowed for rate variation without the assumption of constant regimes at all. Instead, these models assume the rates themselves evolve and change throughout the phylogeny under various Brownian motion-like processes (Lemey et al. 2010; Eastman et al. 2013; Revell 2021; Martin et al. 2022) or single optima Ornstein-Uhlenbeck processes (Hansen et al. 2008; Mitov et al. 2019). The method presented here is most similar to the latter group. hOUwie attempts to explicitly model the evolution of rate shifts according to regimes which jointly influence discrete and continuous character evolution. The regimes themselves are never fixed and each is evaluated as a partial contribution to the overall probability of the data. The advantage of this approach is that it acknowledges the uncertainty in the underlying regime paintings and allows them to change through time.

Additionally, unlike hOUwie, the models discussed thus far, do not explicitly account for the joint modeling of the discrete and continuous characters. Most progress in this area has, until recently, been made via phylogenetic logistic regressions (Ives and Garland 2010) or threshold models in which the discrete character is modeled by a continuously varying unobserved lability (Felsenstein 2012; Cybis et al. 2015). However, these models have yet to be extended beyond binary discrete characters and rely on more simplistic evolutionary models without character independent rate heterogeneity (such as single rate Brownian motion). This lack of character independent rate heterogeneity has recently been recognized as a potential source of inflated correlation between discrete and continuous characters. Such was the reasoning for the MuSSCRat model (May and Moore 2020). Similar to hOUwie, MuSSCRat allows for character independent rate heterogeneity following a multiple rate Brownian motion model to be directly contrasted against character correlation in an attempt to correct for potential biases towards correlation. However, as we outlined in our methods, the way the underlying discrete character is calculated as well as how rate heterogeneity is modeled, differs substantially from May and Moore (2020). Finally, Tribble et al. (2021) has recently developed a method which is similar to the one presented here. One of the primary differences between hOUwie and the Bayesian pipeline discussed in Tribble et al. (2021) is how discrete character evolution is treated. Specifically, Tribble et al. (2021) assume that character-independent mappings are generated under the same parameters which best fit their focal discrete character. In contrast, hOUwie allows the free estimation of character-independent discrete rates which best fit both discrete and continuous data. This difference may lead to be biases against null models since the character-independent regimes are forced to follow a character-dependent discrete model.

*Character-independent models and null hypotheses*

There is a growing appreciation that comparing constant-rate null models to variable-rate alternative models will consistently favor rate heterogeneity, regardless of whether there is a genuine association with a focal variable (Maddison and FitzJohn 2015; Rabosky and Goldberg 2015; Beaulieu and O’Meara 2016; Uyeda et al. 2018; Boyko and Beaulieu 2022). This problem, termed the “straw-man effect” by May and Moore (2020), has been demonstrated to lead to nearly 100% error rates for evidence of discrete character correlation (Maddison and FitzJohn 2015; Boyko and Beaulieu 2022), and has severely biased evidence towards state-dependent speciation and extinction (Rabosky and Goldberg 2015; Beaulieu and O’Meara 2016). Given these often-overwhelming error rates in other comparative methods, we expected to find a similarly consistent bias towards correlation between discrete and continuous characters. However, we found that support for single rate character-independent null models was often greater than character-dependent models even when simulated under character-independent models with rate heterogeneity. Although, the inclusion of explicit multi-rate character independent models (CID+) models did help reduce evidence of false correlation in some cases, by and large, simplistic null models performed admirably. This is not to say that the error rates for discrete and continuous character correlation should be dismissed outright. If our simulations correctly assess that nearly one-third of results find false evidence of a correlation between continuous character rates of evolution and discrete characters, then better null models are certainly needed. But, in comparison to the profound effect that model misspecification has had in other comparative analyses (Beaulieu and O’Meara 2016; Boyko and Beaulieu 2022), the joint models tested here have substantially lower error rates.

We suspect that part of the reason that correlation between discrete and continuous characters is less susceptible to “straw-man” effects than other PCMs is related to the inefficiency of sampling potential maps from the univariate stochastic mapping model. A common approach to fitting OU models involves simulating many stochastic maps to represent underlying regimes from parameters estimated only from the discrete character (Revell 2013). The resulting distribution of underlying regimes will therefore reflect a distribution appropriate for the discrete character, but not necessarily suitable for the continuous character. This is especially true if the continuous character is unlinked to the focal discrete character. Indeed, we found that if the discrete and continuous characters are unlinked, most stochastic maps, even though fairly good for discrete characters, were completely inadequate representations of continuous regimes. Thus, any joint model with these maps contributed little to the overall likelihood. Under our simulation protocol, for a typical run, 90% of the total likelihood for the best set of parameters came from just 2% of the attempted stochastic maps.

In some ways this is good. First, it makes spurious links between a randomly distributed discrete character and a continuous character more unlikely since associations between regimes and continuous variables tend to be specific. This ultimately reduces the potential “straw-man” effect. Second, it is clear that the continuous characters can inform the placement of shared regimes and therefore shift detection methods, where the continuous data are all that provides information about regimes shifts (Ingram and Mahler 2013; Uyeda and Harmon 2014; Khabbazian et al. 2016; Bastide et al. 2017), may be appropriate across a broad range of scenarios. However, this property also makes sampling a good set of regimes to get an accurate estimate of the likelihood difficult and is why the development of our adaptive sampling heuristics was necessary. Adaptive sampling, in combination with our approximation of the joint conditional distributions, helped make parameter estimation more accurate.

*Seed dispersal and climatic niche evolution in Ericaceae*

Here we revaluated three hypotheses related to climatic niche evolution and seed dispersal and found that: (1) the climatic optima of dry fruits was more humid than fleshy fruits (), (2) lineages with dry fruits had faster rates of climatic niche evolution (), and (3) climatic niches of fleshy fruits are more conserved through time (). In contrast to previous findings, the higher rate and stationary variance of climatic niche evolution for dry seeds matched our original hypothesis (Vasconcelos et al. 2021). This is to be expected because abiotically dispersed seeds are likely to be more erratic in their dispersal patterns (Schupp 1993; Westoby et al. 1996). Additionally, that our results differ from previous findings (Vasconcelos et al. 2021) suggests that jointly modeling climatic niche evolution alongside fruit type changed our parameter estimation in a meaningful way.

Our final hypothesis which stated that fleshy, biotically dispersed, seeds are more likely to be associated with humid environments did not find support. However, it has been suggested that a trade-off between seed persistence, seed size, and dispersal strategies can be also common in arid environments (Venable and Brown 1988; Nunes et al. 2017). Specifically, large seed size may occasionally help withstand unfavorable conditions associated with increased aridity (Nunes et al. 2017). With an increased seed size, biotic seed dispersal and fleshy fruits, may become necessary for seed dispersal. This may be the case for Styphelieae, which is distributed in the arid Australian heathland and, of all predominately fleshy-fruited groups, lies the furthest from the inferred aridity optima. Additionally, it has been found that the proportion of abiotically dispersed seeds increases as elevation increases, due to the decreasing availability of frugivores (Chapman et al. 2016). Given that several radiations of Ericaceae lineages are associated with montane habitats (Schwery et al. 2015), it may be that the distribution of dry and fleshy fruits are a consequence of elevation rather than being directly linked to climatic niche evolution. Finally, it has been noted Ericaceae lineages are often found in well-leached soils and epiphytic habitats (Schwery et al. 2015). If associations with soil type are more important than links to climatic optima, we may expect that fruit-dependent climatic optima are consequence of unmodeled factors. Although our modeling explicitly considers hidden variables that may lead to rate heterogeny, if the proposed hidden variable (soil condition) is closely linked to our modeled variable (aridity), then we may not be able to detect the presence of hidden variation. This may be the case between soil condition and aridity (Moreno-Jiménez et al. 2019).

*Caveats and possible extensions*

There are three important caveats to our model. First, our discrete mapping probability ( is only an approximation. What we calculate here is the probability of starting in a particular state and ending a particular state , summed over all possible paths. However, the continuous model probability is based off a particular pathway history which is defined throughout the entire branch (Hansen 1997). Ultimately, this means regimes that the underlying regimes are not treated identically for the continuous and discrete characters. The second caveat is that we do not force hOUwie to sum over all possible mappings . This is because the number of mappings will grow exponentially as the number of nodes and internodes increases and the computation will quickly become infeasible (Jones et al. 2020). Although this may not be entirely necessary since we have shown that only a small percentage of possible mappings contribute to the overall joint probability. Nonetheless, an ideal solution could be the use Markov-Modulated Ornstein-Uhlenbeck models (Huang et al. 2016) since this would remove the need for a regime mapping approach, but these have yet to be applied in phylogenetic comparative biology. Finally, it is possible to extend hOUwie to include state-dependent speciation and extinction dynamics which have been shown to influence the distribution of discrete characters (Maddison 2006) and would therefore influence continuous characters if the two were linked. However, this extension would require a different calculation of the underlying regime mapping probability. Approaches for stochastically mapping SSE models already exist (Freyman and Höhna 2019), so the largest remaining challenge of this extension would be generating high joint probability mappings.

*Conclusion*

The use of pre-defined discrete character mappings can be useful for testing hypotheses which rely on distinct, well-defined differences in the evolutionary histories of lineages. However, this approach assumes that the underlying mapping is known with complete accuracy and ignores the probabilistic nature of discrete regimes. hOUwie’s methodology integrates over the uncertainty of high probability character mappings and relies on the interpretation of parameter estimates from contrasting model structures to find evidence for hypotheses. Rather than assuming an apriori mapping, hOUwie is able to utilize the mutual information about the discrete and continuous characters to learn something about the underlying regimes evolution.

**Tables**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| model\_type | sampling\_procedure | disc | cont | joint |
| CID+\_BMV | adaptive\_sampling | -15.17 | 22.5 | 7.33 |
| CID+\_BMV | discrete\_only | -17.52 | 13.38 | -4.13 |
| CID+\_OUA | adaptive\_sampling | -16.45 | 45.85 | 29.4 |
| CID+\_OUA | discrete\_only | -19.63 | 39.72 | 20.09 |
| CID+\_OUV | adaptive\_sampling | -26.6 | 45.44 | 18.84 |
| CID+\_OUV | discrete\_only | -29.85 | 37.16 | 7.31 |
| CID+\_OUVA | adaptive\_sampling | -13.22 | 51.52 | 38.3 |
| CID+\_OUVA | discrete\_only | -16.33 | 35.51 | 19.18 |
| CID+\_OUM | adaptive\_sampling | -42.09 | 45.97 | 3.88 |
| CID+\_OUM | discrete\_only | -37.57 | -61.17 | -98.74 |
| CID+\_OUMA | adaptive\_sampling | -18.59 | 44.11 | 25.52 |
| CID+\_OUMA | discrete\_only | -16.65 | 1.01 | -15.64 |
| CID+\_OUMV | adaptive\_sampling | -30.56 | 45.13 | 14.57 |
| CID+\_OUMV | discrete\_only | -21.11 | -9.42 | -30.53 |
| CID+\_OUMVA | adaptive\_sampling | -28.52 | 37.72 | 9.2 |
| CID+\_OUMVA | discrete\_only | -26.37 | 14.25 | -12.12 |
| CID+\_OUBM1 | adaptive\_sampling | -19.1 | 5.59 | -13.51 |
| CID+\_OUBM1 | discrete\_only | -19.8 | -59.5 | -79.31 |
| CID+\_OUBMV | adaptive\_sampling | -26.96 | 18.36 | -8.6 |
| CID+\_OUBMV | discrete\_only | -28.45 | -32.44 | -60.89 |

**Table 1:** Discrete, continuous, and joint log probabilities for 100 stochastic maps. Maps are generated based on parameters which were used to generate the dataset being fit.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| model\_type | model\_class | rate | alpha\_1 | alpha\_2 | sigma.sq\_1 | sigma.sq\_2 | theta\_1 | theta\_2 |
| BMV | CD | 0.11 | NA | NA | 0.09 | 0.27 | 0.21 | NA |
| OUV | CD | 0.1 | 1.17 | NA | 0.15 | 0.3 | 0.05 | NA |
| OUA | CD | 0.11 | 1.64 | 1.77 | 0.12 | NA | 0.06 | NA |
| OUM | CD | 0.11 | 1.33 | NA | 0.09 | NA | 0.07 | 0.12 |
| OUVA | CD | 0.09 | 1.41 | 1.1 | 0.12 | 0.96 | 0.06 | NA |
| OUMV | CD | 0.13 | 1.72 | NA | 0.14 | 0.31 | 0.06 | 0.17 |
| OUMA | CD | 0.12 | 2.28 | 2.66 | 0.32 | NA | 0.12 | 0.51 |
| OUMVA | CD | 0.12 | 1.58 | 1.04 | 0.1 | 1.17 | 0.86 | 1.13 |
| OUBM1 | CD | 0.09 | 2.71 | NA | 0.08 | NA | 0.07 | NA |
| OUBMV | CD | 0.08 | 2.26 | NA | 0.13 | 1.72 | 0.07 | NA |
| BMV | CID | 0.04 | NA | NA | 0.27 | 11.03 | 0.23 | NA |
| OUV | CID | 0.04 | 1.03 | NA | 0.32 | 2.01 | 0.05 | NA |
| OUA | CID | 0.04 | 3.26 | 1.29 | 0.39 | NA | 0.07 | NA |
| OUM | CID | 0.07 | 2.35 | NA | 0.14 | NA | 0.49 | 0.19 |
| OUVA | CID | 0.05 | 1.19 | 1.15 | 0.28 | 16.72 | 0.07 | NA |
| OUMV | CID | 0.08 | 2.46 | NA | 0.15 | 2.4 | 1.49 | 0.75 |
| OUMA | CID | 0.04 | 10.28 | 1.26 | 0.26 | NA | 7.11 | 0.82 |
| OUMVA | CID | 0.06 | 6.3 | 1.26 | 0.2 | 10.33 | 10.12 | 1.47 |
| OUBM1 | CID | 0.05 | 3.61 | NA | 0.37 | NA | 0.13 | NA |
| OUBMV | CID | 0.04 | 3.78 | NA | 0.29 | 10.63 | 0.14 | NA |

**Table 2:** Average RMSE per model type and class.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| model\_type | model\_class | AICwt.BM1 | AICwt.OU1 | AICwt.CD | AICwt.CID+ | best\_model |
| BMV | CD | 0.15 | 0.15 | 0.69 | 0.01 | 0.67 |
| OUV | CD | 0.02 | 0.2 | 0.77 | 0.01 | 0.76 |
| OUA | CD | 0.06 | 0.55 | 0.32 | 0.07 | 0.17 |
| OUM | CD | 0.03 | 0.02 | 0.91 | 0.04 | 0.92 |
| OUVA | CD | 0.03 | 0.14 | 0.76 | 0.07 | 0.77 |
| OUMV | CD | 0.01 | 0.01 | 0.96 | 0.02 | 0.97 |
| OUMA | CD | 0.1 | 0.08 | 0.71 | 0.11 | 0.73 |
| OUMVA | CD | 0.04 | 0.08 | 0.81 | 0.07 | 0.81 |
| OUBM1 | CD | 0.18 | 0.57 | 0.13 | 0.12 | 0.08 |
| OUBMV | CD | 0.05 | 0.15 | 0.78 | 0.02 | 0.8 |
| BMV | CID | 0.36 | 0.27 | 0.35 | 0.03 | 0.01 |
| OUV | CID | 0.03 | 0.49 | 0.44 | 0.04 | 0.01 |
| OUA | CID | 0.05 | 0.55 | 0.39 | 0.02 | 0 |
| OUM | CID | 0.19 | 0.08 | 0.03 | 0.69 | 0.73 |
| OUVA | CID | 0.06 | 0.51 | 0.39 | 0.04 | 0.03 |
| OUMV | CID | 0.23 | 0.18 | 0.14 | 0.45 | 0.46 |
| OUMA | CID | 0.41 | 0.39 | 0.15 | 0.05 | 0.05 |
| OUMVA | CID | 0.22 | 0.38 | 0.23 | 0.16 | 0.16 |
| OUBM1 | CID | 0.23 | 0.55 | 0.18 | 0.04 | 0.01 |
| OUBMV | CID | 0.22 | 0.33 | 0.34 | 0.11 | 0.1 |

**Table 3**: Average AIC weights summarized over 150 simulated datasets

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| nTaxa | model\_type | AICwt.BM1 | AICwt.OU1 | AICwt.CD | AICwt.CID+ |
| 25 | CD | 0.12 | 0.22 | *0.51* | 0.15 |
| 100 | CD | 0.06 | 0.22 | *0.7* | 0.02 |
| 250 | CD | 0.02 | 0.14 | *0.82* | 0.02 |
| 25 | CID | 0.28 | 0.35 | 0.24 | *0.14* |
| 100 | CID | 0.21 | 0.4 | 0.23 | *0.15* |
| 250 | CID | 0.11 | 0.34 | 0.32 | *0.22* |

**Table 4:** Average AIC weight depending on the number of taxa. Italics highlights the generating model support.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Model | np | lnLik | DiscLik | ContLik | AIC | dAIC | AICwt |
| FitSD=CD\_BMV | 5 | -243.78 | -32.62 | -207.08 | 497.56 | 40.85 | 0 |
| FitSD=CD\_OUA | 6 | -224.95 | -32.58 | -189.48 | 461.9 | 5.19 | 0.03 |
| FitSD=CD\_OUBM1 | 5 | -243.84 | -32.57 | -206.67 | 497.68 | 40.97 | 0 |
| FitSD=CD\_OUBMV | 6 | -243.79 | -32.61 | -206.99 | 499.57 | 42.87 | 0 |
| FitSD=CD\_OUM | 6 | -224.12 | -32.57 | -187.79 | 460.24 | 3.53 | 0.07 |
| FitSD=CD\_OUMA | 7 | -223.21 | -32.58 | -187.97 | 460.42 | 3.71 | 0.06 |
| FitSD=CD\_OUMV | 7 | -224.05 | -32.62 | -188.15 | 462.1 | 5.39 | 0.03 |
| *FitSD=CD\_OUMVA* | ***8*** | ***-220.35*** | ***-32.6*** | ***-183.27*** | ***456.71*** | ***0*** | ***0.41*** |
| FitSD=CD\_OUV | 6 | -225.49 | -32.62 | -188.47 | 462.98 | 6.27 | 0.02 |
| FitSD=CD\_OUVA | 7 | -221.62 | -32.58 | -184.44 | 457.24 | 0.53 | 0.32 |
| FitSD=CID\_BM1 | 4 | -243.89 | -32.62 | -206.67 | 495.78 | 39.07 | 0 |
| FitSD=CID\_OU1 | 5 | -225.5 | -32.62 | -188.28 | 461.01 | 4.3 | 0.05 |
| FitSD=CID+\_BMV | 7 | -244.8 | -33.11 | -205.78 | 503.59 | 46.89 | 0 |
| FitSD=CID+\_OUA | 8 | -226.42 | -33.17 | -188.53 | 468.84 | 12.13 | 0 |
| FitSD=CID+\_OUBM1 | 7 | -244.44 | -33.16 | -206.67 | 502.88 | 46.17 | 0 |
| FitSD=CID+\_OUBMV | 8 | -225.58 | -32.71 | -186.58 | 467.17 | 10.46 | 0 |
| FitSD=CID+\_OUM | 8 | -226.43 | -33.32 | -189.07 | 468.87 | 12.16 | 0 |
| FitSD=CID+\_OUMA | 9 | -225.57 | -32.68 | -189.92 | 469.14 | 12.43 | 0 |
| FitSD=CID+\_OUMV | 9 | -225.2 | -33.39 | -182.88 | 468.39 | 11.68 | 0 |
| FitSD=CID+\_OUMVA | 10 | -227.39 | -33.13 | -185.15 | 474.79 | 18.08 | 0 |
| FitSD=CID+\_OUV | 8 | -228.77 | -32.98 | -190.16 | 473.55 | 16.84 | 0 |
| FitSD=CID+\_OUVA | 9 | -244.38 | -33.43 | -202.12 | 506.76 | 50.05 | 0 |
| FitSD=HYB\_BMS | 9 | -244.46 | -33.08 | -204.83 | 506.93 | 50.22 | 0 |
| FitSD=HYB\_OUM | 10 | -224.12 | -32.67 | -188.99 | 468.23 | 11.52 | 0 |
| FitSD=HYB\_OUMVA | 16 | -226.56 | -33.03 | -179.11 | 485.13 | 28.42 | 0 |

**Table 5:** Empirical model fitting results.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Continuous** | **Alpha** | **Sigma.sq** | **Theta** | **Discrete** | **To dry** | **To fleshy** |
| **Dry** | **0.015** | **0.011** | **0.97** | **From dry** | **-** | **0.0015** |
| **Fleshy** | **0.023** | **0.007** | **0.81** | **From fleshy** | **00036** | **-** |

**Table 6:** Model averaged parameter estimates.

**Figures**

A picture containing diagram

Description automatically generated

Figure 1: A state-transition diagram describing the examined hOUwie model structures.

Calendar

Description automatically generated

Figure 2: Forward simulations of some of the different possible model types in hOUwie.

Diagram

Description automatically generated

Figure 3: Adaptive sampling procedure.



Figure 4: For a single simulation, the performance of adaptive sampling (blue) compared to pure discrete stochastic mappings (red). Dashed line likelihood under generating map.

Chart, box and whisker chart

Description automatically generated

**Figure 5:** Parameter estimation performance. Raw differences from generating parameters.

Timeline, box and whisker chart

Description automatically generated  
**Figure 6:** Model AIC weights by simulating model.



**Figure 7:** Ericaceae dataset and phylogeny. Standard error shown calculated from 100 parametric bootstraps.

**Works Cited**

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

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

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

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

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

Boyko J., Beaulieu J. 2022. A potential solution to the unresolved challenge of false correlation between discrete characters. .

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Rao V., Teh Y.W. 2013. Fast MCMC Sampling for Markov Jump Processes and Extensions. :26.

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

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

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

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

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

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

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

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

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

Tribble C.M., May M.R., Jackson-Gain A., Zenil-Ferguson R., Specht C.D., Rothfels C.J. 2021. Unearthing modes of climatic adaptation in underground storage organs across Liliales. .

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

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

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

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

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

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