**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**

**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. As such, models favoring variation in rates and states has fast become the rule, as opposed to the rare exception in empirical applications of comparative methods. The challenge then becomes how to model this heterogeneous process with the limitation that, for most comparative biologists, our data comes exclusively from extant species. Our model, which we call hOUwie, uses hidden Markov models as a statistically rigorous way to learn about variation in the evolutionary process. hOUwie is composed of two processes: one describes the evolution of a discrete character (such as whether a mammal is a carnivore, herbivore, or omnivore) and the other the evolution of a continuous character (such as mammalian body size).

The past is inherently uncertain… 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 causal in the change of 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 a more realistic view of evolutionary change is one which takes a dialectical approach (Levins and Lewontin 1985) – namely, 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, help provide information about this 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) with generalized Ornstein-Uhlenbeck models (Hansen 1997; Butler and King 2004; Hansen et al. 2008; Beaulieu et al. 2012; Ho and Ané 2014*b*). 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. We also consider and explore the potential for Type-I error that may arise within these joint models. When the evolution of a continuous trait is heterogeneous and unlinked to a discrete focal character, are they more likely to support a character-dependent model than simple, single rate, character independent models? 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 2014*b*). 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 (2014*b*). 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 2014). 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). As the number of internodes increase, the amount of time between nodes decreases and the approximation improves (Rao and Teh 2013). Formally, we calculate the joint probability of a regime structure and the discrete character as

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 (which can be internal, external, or internodes).

For each set of parameters evaluated during the maximum likelihood search, stochastic mappings are generated using a 3-step process (Nielsen 2002). First, the conditional state probabilities at nodes are calculated based on our discrete character and continuous character model. 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. Second, 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 calculated in step one 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 ladder is calculated using familiar matrix exponentiation methods (e.g., Pagel 1994)). The third step under usual stochastic mapping procedures would be to 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 or internodes. It should be noted that, while the conditional probability under a discrete model is well-defined, we use an approximation to calculate a conditional probability of the discrete character at nodes under the continuous model. Specifically, we calculate the probability of the observing the mean value of all descendants given the OU model parameters, integrated over all possible rootward node, and observed tip ward discrete states for both branches stemming from the focal ancestral node. Although this is only an approximation of the conditional state probabilities based on the continuous tip-values and continuous model, it proves to be an essential improvement over the typical procedure of sampling many stochastic maps based solely on the discrete process.

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

where summing over all possible 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. For the discrete component of the model, we assumed that transitions between the observed characters were equal, although when hidden states were included, rates within the hidden rate classes differed. Furthermore, due to concerns with parameter estimation, 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 (OU1A), only sigma varies (OU1S), and combinations of an OU and BM process (OUBM). The 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 potentially unrealistic assumption we introduce a character-independent model which allows for differences in the OU parameters to depend upon an unobserved hidden state (CID+). 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+). HYB models are not examined because of their increased parameter complexity.

A screenshot of a cell phone

Description automatically generated with low confidence

**Figure 1**: A state-transition diagram describing the examined hOUwie model structures. It is possible to add more complexity, such as allowing transitions between hidden states to depend on the observed character (e.g., 0A to 0B differing from 1A to 1B). Different shapes represent alternative observed states (square and circle for observed state 0 and 1). Different shades of the same color are used to describe different observed states within the same hidden state (light orange and light purple are state 0, while dark orange and dark purple are state 1). Different colors are used to represent different hidden states (orange indicates hidden state A and purple indicates hidden state B). Parameters which are shared between states are shown above the states, whereas parameters specifically associated with a particular state are shown to their left or right. Transitions between states are shown with colored arrows.

*Proof of Concept: Joint probability for all possible mappings*

To evaluate the likelihood of a set of parameters, it would be preferable to marginalize the probability of the discrete and continuous characters over all possible stochastic mappings. However, this is computationally infeasible for most empirical datasets and thus our approach samples stochastic mappings proportional to the discrete character probabilities. Nonetheless, for a small phylogenetic tree and a modest number of internodes, it is possible to decompose the discrete and continuous probabilities for all potential stochastic mappings and explore the theoretical properties of jointly modeling the two character classes. We simulated a continuous character on the regime shown in Fig. 2 with parameters on a tree with height 1. These extreme parameter values are chosen to ensure a strong link between the underlying regime and the simulated continuous data. Regimes were then treated as either an observed (Fig. 2a) or hidden state (Fig. 2b). When treated as a hidden state, one of the two binary observed states was assigned to either tip within the hidden regime. We then fit the hOUwie model by integrating over all possible stochastic mappings, rather than a subset. For a CD model there are 8 possible mappings with no internodes, and for a CID+ model there are 1024 possible mappings with no internodes.

*A picture containing qr code

Description automatically generated*

**Figure 2**:Two scenarios for continuous and discrete character evolution. a) Continuous evolution and observed discrete character evolution are linked with the same underlying regime painting. Under this scenario, the observed discrete characters (circle and square) provide information about the regime structure. b) Continuous evolution and hidden discrete state evolution (green and purple) are linked. The observed state (circle and square) is not linked to the underlying regime or the continuous trait. Information to infer the underlying continuous regime comes directly from the continuous character distribution.

*Simulation study*

For each of the 22 hOUwie models tested here, we simulated 50 datasets for phylogenies of 50, 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 50, 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 (Boyko et al. 2022, in prep).

*The impact of climatic variables on seed dispersal*

An ideal way to address the complexity of a trait syndrome would be to enumerate all the phenotypes which comprise the seed dispersal syndrome and study the dynamics in-depth. However, on macroevolutionary scales, this approach is intractable due to the vast amount of species-specific data required and the increasing number of parameters introduced into models as the number of traits increases. Thus, we propose an alternative which will use a single feature of fruit morphology as a proxy for seed dispersal, but also acknowledges that not all lineages with the same fruit morphology are going to evolve in the same way. It is for this reason that we use hidden Markov models (HMM) as the basis for describing discrete character evolution. With HMMs we allow for the fact that dynamics within broadly defined categories are likely inadequate descriptions when treated homogeneously.

Here we examine three predictions outlined in Vasconcelos et al. (2021), but specifically measuring aridity. First, we expect that the climatic optima for abiotically dispersed seeds will be drier compared to biotically dispersed seeds (). Second, we expect that abiotically dispersed seeds will have faster rates of climatic niche evolution (). Finally, we expect that the climatic niches of biotically dispersed seeds 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 abiotically dispersed seeds was more humid than biotically dispersed seeds. Second, they found that the rate of climatic evolution was greater in biotically dispersed seeds than abiotically dispersed seeds. Additionally, this previous work assumed either character dependence or simple character independence, but heterogeneous character independence may lead to a false signal of character dependence as in other macroevolutionary methods (Beaulieu and O’Meara 2016). To account for this, we include the CID+ model within our model set.

We include 27 hOUwie models within our model set: 2 CID, 10 CD, 10 CID+, and 5 HYB. *Gaultheria* is technically a dry-fruited genus within Ericaceae but has a persistent fleshy calyx that attracts frugivores (Stevens et al. 2004). For that reason, 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). Finally, we evaluate the expected values, the optima, and compare our results to Vasconcelos et al. (2021).

**Results**

*Parameter estimation given the generating model*

Parameter estimation in hOUwie did not vary greatly depending on the number of stochastic maps or number of internodes included (Fig. S2). Most variation in parameter estimation was instead related to the model complexity and the specific parameters which were allowed to vary (Table 1). Datasets which were generated with a variable alpha rate were never chosen as the best fitting model. This parameter estimation problem was amplified when alpha was allowed to vary alongside other parameters (as was the case in Beaulieu et al. 2012). Increasing the number of taxa has been shown to improve inference of OU models, but even with more taxa, if alpha is not large enough relative to sigma, estimation may remain poor. (Beaulieu et al. 2012; Ho and Ané 2014*a*; Cressler et al. 2015). CID+ parameter estimations were highly sensitive to model complexity. Although there are examples of low RMSE for each of the OU parameters, when these parameters were allowed to vary, error increased rapidly. This suggests that alternative approaches may be necessary to accurately estimate a hidden state OU model and possibly shift detection methods can supplement this gap in the interim (Uyeda and Harmon 2014; Khabbazian et al. 2016).

The sign error for variable theta was 0.05 or less, regardless of additional model complexity. Variable sigma models also performed well, with only models that also included a variable alpha (M7 and M12) having a sign error rate above 0.15 (Table 1). Finally, identifying the sign difference of alpha proved difficult for all variable alpha models. This is consistent with previous results which found the parameters of variable alpha models are generally difficult to estimate (Beaulieu et al. 2012). Sign errors cannot be evaluated for CID models because the parameter values associated with hidden states can freely switch between rate classes and have identical likelihoods. For example, if we were to simulate a dataset where states A and B are hidden states unlinked to the observed discrete character, , and , there would be no difference in likelihood if , and or , and . However, this could be distinguishable if both the observed discrete character and continuous character were linked to the hidden state.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| model | type\_1 | type\_2 | k | alpha\_free | sigma\_free | theta\_free | alpha\_sign\_error | sigma\_sign\_error | theta\_sign\_error | alpha\_rmse | sigma\_rmse | theta\_rmse | prop\_best | avg\_AICwt | prop\_CD | mean\_obs\_exp | sd\_obs\_exp |
| M1 | BM | CID | 3 | FALSE | FALSE | FALSE | NA | NA | NA | NA | 0.086 | 0.321 | 0.8 | 0.313 | 0.2 | 0.032 | 0.153 |
| M2 | BM | CD | 4 | FALSE | TRUE | FALSE | NA | 0.05 | NA | NA | 0.419 | 0.391 | 0.367 | 0.213 | 0.433 | -0.001 | 0.141 |
| M3 | BM | CID | 6 | FALSE | TRUE | FALSE | NA | NA | NA | NA | 97.669 | 0.283 | 0 | 0.007 | 0.3 | 0.026 | 0.188 |
| M4 | OU | CID | 4 | FALSE | FALSE | FALSE | NA | NA | NA | 1.88 | 0.344 | 0.096 | 0.333 | 0.179 | 0.433 | -0.011 | 0.098 |
| M5 | OU | CD | 5 | TRUE | FALSE | FALSE | 0.556 | NA | NA | 9.979 | 0.179 | 0.104 | 0.036 | 0.033 | 0.536 | 0.004 | 0.064 |
| M6 | OU | CD | 5 | FALSE | TRUE | FALSE | NA | 0.05 | NA | 1.44 | 0.644 | 0.102 | 0.167 | 0.139 | 0.533 | -0.018 | 0.105 |
| M7 | OU | CD | 6 | TRUE | TRUE | FALSE | 0.4 | 0.4 | NA | 2.49 | 4.18 | 0.101 | 0 | 0.039 | 0.5 | 0.013 | 0.064 |
| M8 | OU | CD | 5 | FALSE | FALSE | TRUE | NA | NA | 0.05 | 0.804 | 0.618 | 11.52 | 0.167 | 0.126 | 1 | -0.025 | 0.196 |
| M9 | OU | CD | 6 | TRUE | FALSE | TRUE | 0.75 | NA | 0 | 2.803 | 0.336 | 19.421 | 0.067 | 0.067 | 1 | 0.024 | 0.126 |
| M10 | OU | CD | 6 | FALSE | TRUE | TRUE | NA | 0.15 | 0.05 | 0.982 | 13.478 | 86.098 | 0.233 | 0.226 | 1 | 0.129 | 0.274 |
| M11 | OU | CD | 7 | TRUE | TRUE | TRUE | 0.5 | 0.2 | 0 | 2.667 | 20.419 | 53.158 | 0.3 | 0.271 | 1 | 0.183 | 0.267 |
| M12 | OU | CID+ | 7 | TRUE | FALSE | FALSE | NA | NA | NA | 23.696 | 1.733 | 0.75 | 0 | 0.001 | 0.345 | -0.18 | 0.067 |
| M13 | OU | CID+ | 7 | FALSE | TRUE | FALSE | NA | NA | NA | 3.942 | 89.838 | 0.122 | 0 | 0.004 | 0.433 | 0.012 | 0.102 |
| M14 | OU | CID+ | 8 | TRUE | TRUE | FALSE | NA | NA | NA | 35.524 | 59.036 | 1.277 | 0.038 | 0.039 | 0.308 | 0.488 | 0.102 |
| M15 | OU | CID+ | 7 | FALSE | FALSE | TRUE | NA | NA | NA | 1.397 | 2.091 | 31.886 | 0 | 0.001 | 0.667 | 0.001 | 0.422 |
| M16 | OU | CID+ | 8 | TRUE | FALSE | TRUE | NA | NA | NA | 2.878 | 4.048 | 57.683 | 0 | 0 | 0.667 | 0.276 | 0.483 |
| M17 | OU | CID+ | 8 | FALSE | TRUE | TRUE | NA | NA | NA | 1.472 | 80.775 | 49.334 | 0 | 0.002 | 0.467 | -0.081 | 0.395 |
| M18 | OU | CID+ | 9 | TRUE | TRUE | TRUE | NA | NA | NA | 3.227 | 150.971 | 51.689 | 0 | 0 | 0.533 | 0.349 | 0.537 |
| M19 | mixed | CD | 4 | TRUE | FALSE | FALSE | NA | NA | NA | 9.085 | 2.38 | 0.218 | 0.036 | 0.071 | 0.393 | 0.027 | 0.103 |
| M20 | mixed | CD | 5 | TRUE | TRUE | FALSE | NA | 0.05 | NA | 1.024 | 6.754 | 0.314 | 0.138 | 0.174 | 0.724 | 0.043 | 0.169 |
| M21 | mixed | CID+ | 6 | TRUE | FALSE | FALSE | NA | NA | NA | 22.298 | 2.631 | 0.199 | 0 | 0.002 | 0.267 | 0.015 | 0.11 |
| M22 | mixed | CID+ | 7 | TRUE | TRUE | FALSE | NA | NA | NA | 16.703 | 82.599 | 2.181 | 0 | 0.003 | 0.429 | -0.13 | 0.164 |

**Table 1**: Model is model name. type\_1 is the type of continuious process. type\_2 is whether the model is character dependent (CD) or independent (CID). k is the number of parameters. Alpha, sigma, and theta free are whether or not that parameter is free to vary in the model. sigma, theta, alpha sign error are the proportion of times the incorrect sign difference was inferred compared to simulated.theta, sigma, alpha rmse are the reduced mean squared errors of the best fitting model compared to the simulated values. Prop\_best is the proportion of fitted models that matched the generating model. Avg aic weight is the average weight of the generating model to its simulated data when fit and compared to all other models. prop\_CD is the proportion of models that had any CD model as the best fitting. Mean\_obs\_exp is the average difference between the expected and observed values for the best fitting model. SD\_obs\_exp is the standard deviation of the differences between observed and expected values (taken per species per simulation).

Despite variable theta models having the lowest sign errors and highest successful detection of character dependent models, theta estimates still had a high RMSE. One advantageous property of theta is that, unlike sigma and alpha, it describes a directly observable feature of the organism with easily interpretable units (the same as the continuous phenotype). Thus, we can examine the distribution of the simulated character and compare these distributions with the inferred optimal values and expected values to understand if the estimates are meaningful given the dataset. This post-hoc analysis revealed an interesting feature of our simulated dataset and hOUwie’s potential to find variation outside of the observed data (Fig. 3). We find that, although there is significant error around the optima estimate, the simulated data rarely included any species that had a phenotypic optimum near the simulated optima (Fig. 3b). This suggests that hOUwie can estimate phenotypic optimum outside of the observed phenotypic distributions. We are also able to examine the expected continuous trait values according to the fitted model and determine how well they match the observed phenotypic values (Fig. b,c; Table 1). The general match between observed and expected values suggests a reasonable approximation of the data by the model. Furthermore, although the RMSE in theta is large, it is likely a consequence of a lack of direct phenotypic observations around the optimum. This may be indicative of the potential for hOUwie to detect optima well outside the range of observed phenotypes. While the exact estimate of the phenotypic optima will not be highly accurate, its direction compared to the observed phenotypic distribution is still interesting and may be useful for hypothesis testing.

![Chart, bar chart

Description automatically generated]()

**Figure 3**: The two columns are data simulated with an initial discrete root state of 1 or 2 which will correspond to an initial continuous value of 12 or 24. a) Model averaged parameter estimates for theta. The dashed blue and red lines represent the simulating values of theta for states 1 and 2 respectively. b) the observed continuous character frequency. When data is simulated with root state 1 initially, most continuous data tends to be clustered around the initial value of 12. The same is true of root state 2, but most continuous data are clustered around the initial value of 24. c) The model expectation of the phenotypic value of the tips. The exact equations for calculating the expected value are described in Butler and King (2004) and Beaulieu et al. (2012) but are essentially the weighted average of time spent in a particular regime state given OU parameters and proportionally weighted by the nearness of the regime state to the present given an initial value. Summarized here are all non-hidden Markov models for data simulated under a variable theta CD model. HMMs are not included because model averaging hidden state associated parameter values is difficult outside the context of tip averages.

*The power to detect the generating model*

Simple character independent (CID) and character dependent (CD) models performed the best in our power analyses (Fig. 4). Of the character dependent models, variable alpha, and variable sigma models (M5, M7, M9) had low AIC weights across simulations even when they were the generating modes (Table 1). This reinforces our previous findings that the more complex variable alpha or variable sigma models are the most difficult to estimate and should be applied and interpreted with caution. Conversely, CD models which varied theta always found evidence for character dependence when the generating model included a link between discrete and continuous characters. Power analysis also suggests that hOUwie can detect character independence when that is the true model. However, when character independent models are heterogeneous (i.e., there is variation in how a continuous trait evolves unlinked to the focal character) hOUwie either favored simpler versions of character independent models or, more concerningly, character dependent models (Fig. 4c). This suggests that even in cases where our focal discrete and continuous characters are independent, if the evolutionary process is heterogeneous, we may still find false support for character dependence.

![Chart, box and whisker chart

Description automatically generated]()

**Figure 4**: Boxplots of summed AICwt for each model class for three simulating scenarios. Models are summarized into three classes (CD, CID, or CID+) although within each class models there are different model structures (e.g., variable theta vs variable alpha – see table 1). a) The simulating model is character dependent. The percent best fit (proportion from each class with dAIC = 0) for each model class was CD, CID, and CID +. b) The simulating model is character independent. The percent best fit for each model class was CD, CID, and CID +. c) The simulating model is character independent. The percent best fit for each model class was CD, CID, and CID +.

*Joint probability for all possible mappings*

Using all possible regime mappings, we were able to successfully distinguish between character dependent and character independent models (Table 2). When data was generated following a variable theta character dependent scenario, a character dependent model was preferred with a dAIC of 5.7, suggesting strong evidence for character dependence (Burnham and Anderson 2002). Interestingly, the maximum likelihood parameter estimates for the CD and CID+ models were identical in this scenario. The reason that a CD model will fit better despite having the same parameter estimates is the additional uncertainty when estimating hidden states in the CID+ model. Conversely, when continuous traits were unlinked to the underlying regime, a hidden state character independent model was preferred with strong evidence (dAIC=15.4). Taken together, this points to the theoretical possibility of estimating OU processes even when unlinked to an observed character. However, a different approach may be necessary as integrating over all possible mappings is not feasible for most empirical datasets.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| generating model | fit model | lnlik | k | aic | dAIC | discrete rate | alpha | sigma2 | theta\_1 | theta\_2 |
| CD | CD | -4.8094 | 5 | 19.61883 | 0 | 1.75 | 10 | 5 | 5 | 10.9375 |
| CD | CID+ | -7.6937 | 5 | 25.38755 | 5.76872 | 1.75 | 10 | 5 | 5 | 10.9375 |
| CID+ | CD | -11.924 | 5 | 33.84911 | 15.4376 | 2.03125 | 0.625 | 11.322632 | 6.234131 | 7.1875 |
| CID+ | CID+ | -4.2057 | 5 | 18.41151 | 0 | 48.99516 | 50 | 2.895839 | 4.991589 | 10 |

**Table 2**: Model evaluation of CD and CID+ models when all possible stochastic mappings are integrated over. The CD and CID+ models are distinguishable by AIC and likelihood. Maximum likelihood parameter estimates are also given.

*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 3). The results presented in Table 3 are summarized into four broad categories (CD, CID, CID+, and HYB), but we tested 27 different scenarios and the overall best model was M11, in which alpha, sigma2, and theta were all allowed to vary (AICwt=0.87; Supplemental Materials). Both fleshy and dry fruits had phenotypic optima which correspond to non-arid environments. However, the 0.59 AI of fleshy fruits suggests dryland as their optimal habitat, and the 1.044 AI of dry fruits corresponds to a non-dryland environment (Middleton and Thomas 1997). Both and interact to create tip variance, so in addition to , we measured stationary variance . We found that Ericaceae lineages with dry fruits were less 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 15% and 25% of the total tree height respectively. Transitions to fleshy fruit occurred at 0.0017 transitions per million years which is more than 6 times faster than transitions to dry fruits (0.0003 transitions per million years). But in both cases, the average waiting time to the next transition is much greater than the height of the tree (117.3 MY) suggesting few transitions between fruit types throughout the Ericaceae’s history. Taken together with the relatively short half-lives of each fruit type, our modeling suggests that the climatic optima of Ericaceae lineages is relatively stable. Finally, we note that the difficulty in estimating complex models is evident in this empirical application. Models with higher number of parameters should provide a more likely explanation of the data, but that was not always the case (Supplemental Materials, Table 3). This points to the failure of more complex models finding a maximum likelihood estimate. However, this is consistent with our earlier simulation study which suggests that without integrating over all possible mappings, it may not be possible to reliably find the MLE when the hidden states are associated with the OU process.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| model | np | lnLik | DiscLik | ContLik | AIC | dAIC | AICwt |
| M1 | 4 | -265.91 | -32.54 | -233.36 | 539.81 | 69.63 | 0 |
| M2 | 5 | -263.85 | -32.54 | -224.52 | 537.69 | 67.51 | 0 |
| M3 | 6 | -267.67 | -34.31 | -229.3 | 547.35 | 77.17 | 0 |
| M4 | 5 | -233.63 | -32.55 | -201.08 | 477.27 | 7.09 | 0.03 |
| M5 | 6 | -257.11 | -32.55 | -222.62 | 526.22 | 56.04 | 0 |
| M6 | 6 | -233.15 | -32.55 | -196.19 | 478.3 | 8.12 | 0.01 |
| M7 | 7 | -241.35 | -32.57 | -205.82 | 496.7 | 26.52 | 0 |
| M8 | 6 | -231.8 | -32.53 | -196.32 | 475.6 | 5.42 | 0.06 |
| M10 | 7 | -231.32 | -32.55 | -196.01 | 476.65 | 6.47 | 0.03 |
| **M11** | **8** | **-227.09** | **-32.56** | **-192.87** | **470.18** | **0** | **0.87** |
| M12 | 7 | -235.42 | -34.34 | -197.12 | 484.84 | 14.66 | 0 |
| M13 | 7 | -309.15 | -34.57 | -271.72 | 632.31 | 162.13 | 0 |
| M15 | 7 | -330.78 | -50.77 | -264.39 | 675.56 | 205.37 | 0 |
| M16 | 8 | -3931.97 | -35.42 | -2430.46 | 7879.94 | 7409.76 | 0 |
| M17 | 8 | -259.76 | -34.36 | -222.04 | 535.52 | 65.33 | 0 |
| M18 | 9 | -391.77 | -35.98 | -354.05 | 801.54 | 331.36 | 0 |
| M19 | 5 | -257.09 | -32.54 | -222.73 | 524.18 | 54 | 0 |
| M20 | 6 | -241.15 | -32.55 | -208.48 | 494.3 | 24.12 | 0 |
| M21 | 6 | -235.4 | -34.32 | -197.68 | 482.81 | 12.62 | 0 |
| M22 | 7 | -336.79 | -35.17 | -299 | 687.59 | 217.41 | 0 |
| M23 | 8 | -596.26 | -51.8 | -544.45 | 1208.51 | 738.33 | 0 |
| M24 | 9 | -307.66 | -42.7 | -260.65 | 633.32 | 163.14 | 0 |
| M25 | 12 | -296.3 | -43.01 | -247.55 | 616.6 | 146.42 | 0 |
| M26 | 12 | -362.65 | -61.05 | -281.89 | 749.3 | 279.12 | 0 |
| M27 | 15 | -354.15 | -66.97 | -268.04 | 738.3 | 268.12 | 0 |

Supplemental Table – Modeling results from Ericaceace seed dispersal and aridity.

**Discussion**

Phylogenetic comparative methods have been widely applied to study discrete and continuous characters separately, but there are surprisingly few options which simultaneously use both classes of character (Felsenstein 2012). 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.

*The significance of jointly modeling characters*

Most implementations of an OU process within PCMs assume that the regimes are known or estimate them without considering discrete character evolution (Uyeda and Harmon 2014). These approaches leave little room for inference about the regimes themselves and how they change through time, nor the possibility that the evolution of the continuous character could influence change in the regimes or vice versa. The primary difference between hOUwie and other applications of the OU process is the explicit joint modeling of the underlying regime structure and continuous trait. There are three main advantages to jointly modeling discrete and continuous traits. First, it represents a more realistic view of trait evolution. The evolution of any discrete or continuous trait is not independent of all other traits a lineage possesses. Second, the mutual information of multiple variables makes our inferences more robust (Cover and Thomas 1991). If we are certain of one variable (seed size), but uncertain about another variable (climate), we can use the knowledge of the relationship between our variables to draw more accurate inferences about the uncertain variable. Finally, using hidden Markov models allows for the discovery of hidden variation. There is little reason to believe that any relationship between discrete and continuous variables will be homogenous throughout the history of a lineage. With hOUwie, it is theoretically possible to discover regions of the phylogeny where different types of relationships emerge.

A joint model also makes the parameter estimates related to continuous trait evolution independent of pre-defined regime paintings. The use of pre-defined regime mappings can be often be useful for hypothesis testing (Butler and King 2004; Beaulieu et al. 2012). For example, testing for evidence for character displacement by mapping either a sympatric or allopatric evolutionary history (Butler and King 2004). However, when using this approach, the mapping is taken as an absolute certainty. There is no room for inference of other potential patterns nor is there the acknowledgement that the *apriori* mapping is a probabilistic sample of many possibilities. Instead of contrasting mappings, hOUwie’s methodology integrates over the uncertainty of alternative regimes and relies on the interpretation of parameter estimates from contrasting model structures to find evidence for hypotheses. To that end, modeling the joint probabilities puts our models into a likelihood framework. This allows for model comparison, model averaging, and other tools to be utilized when comparing macroevolutionary hypotheses (Burnham and Anderson 2002). This framework may be preferable to testing whether a particular hypothesis is better than a trivial null model (Beaulieu and O’Meara 2016; Caetano et al. 2018). Detecting a significant correlation between two traits is an important first step in establishing an evolutionary relationship, however a richer understanding may be achieved through examining the models’ parameterization in relation to specific hypotheses. Model averaged parameters and tip rates are a way to both test preexisting hypotheses and begin developing new hypotheses based on interpretation (misse). Additionally, because the relative explanatory power of each model is related to parameter estimates, by using model averaging we account for model structure uncertainty in addition to parameter estimation uncertainty (Burnham and Anderson 2002).

*It might be tempting to critique our map sampling maps as inefficient in comparison to popular MCMC techniques*

*"One thing that surprised us was the inefficiency of sampling potential maps from the univariate simmap model. For our simulation conditions, many potential simmaps, even though fairly good for discrete characters (thus their sampling under this process), are absolutely terrible for the continuous regimes and so the joint model with these maps contributes little to the overall likelihood. For a typical run, 90% of the total likelihood for the best set of parameters came from just <small percent> of the attempted simmaps. In some ways this is good: it is clear that the continuous characters have information about the placement of regimes (see classic OU examples (bayou, ouwie, ouch, etc.) where the continuous data are all that provides info on regimes). But it also makes sampling good regimes to get an accurate estimate of the likelihood hard -- we developed our "cherry" algorithm to help with this, as sampling discrete only simmaps was not efficient enough. Approaches that pick simmaps at complete random might work even worse."*

*Seed dispersal and climatic niche evolution in Ericaceae*

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 revaluated three hypotheses related to climatic niche evolution and seed dispersal: (1) the climatic optima for abiotically dispersed seeds will be drier compared to biotically dispersed seeds (), (2) abiotically dispersed seeds will have faster rates of climatic niche evolution (), and (3) climatic niches of biotically dispersed seeds will be more conserved through time (). Consistent with previous findings we did not find evidence of these hypotheses in Ericaceae. This may have been the result of *Dracophyllum,* a large dry-fruited genus, being distributed in particularly wet environments (Wagstaff et al. 2010) or the possibility that arid environments tend to have a higher climatic niche evolution. The latter hypothesis is consistent with our finding that lineages in arid environments (which was predominantly associated with dry fruits) transitioned to fleshy fruits at a rate 6-times faster than the alternative transition.

*Caveats*

There are three important caveats of our model. First, our discrete mapping probability ( is merely 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). Our justification for this approximation is that as the length between internodes decreases this approximation improves because the probability of more than a single transition (or no transitions) decrease. 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. Instead, we simulate node, internodes, and tip states (tip states only in the case of hidden Markov models) using the stochastic mapping procedure described in Bollback (2006) and modified to include the possibility of internodes. Our final caveat is that because we generate node mappings using the stochastic mappings procedure, they would typically be based solely on the discrete character process. This can be particularly problematic for character independent models where the discrete character and continuous character are decoupled. An imperfect solution to this problem is to weight tip probabilities when first calculating the conditional likelihoods (Bollback 2006). Typically, tip values for hidden Markov models are given a weight of 1 for each state (Beaulieu et al. 2013). However, we can modify this value to be proportional to the probability of sampling a particular tip’s continuous trait value given the models theta and sigma value for each discrete state. This is the approach we take; we assume that the conditional probability of a tip being in a particular hidden state is not equal, but rather proportional to ; the probability of sampling the continuous trait value under a normal distribution for each possible state . In cases where there are no differences between parameters, conditional probabilities have equal weights.

*Future extensions*

We demonstrate a potentially serious problem when fitting OU models without allowing for heterogeneity in the evolution process. In cases where discrete and continuous characters are unlinked, but there is heterogeneity in the dynamics continuous character, we find high false-positive support for character dependent models. Although hOUwie is designed to mitigate these concerns, due to the caveats previously discussed it is far from a complete solution. The joint estimation procedure implemented in hOUwie has high parameter estimation error when focal discrete characters are unlinked to the continuous phenotype. There are several avenues to improve the hOUwie model that are actively being explored. Statistical work on Markov-Modulated Ornstein-Uhlenbeck models represent an intriguing possibility but have yet to be applied in phylogenetic comparative biology (Huang et al. 2016). However, if implemented in a phylogenetic context, it would remove the need for a simulation approach. Finally, it is possible to extend hOUwie to include state-dependent speciation and extinction dynamics. This extension would require a different calculation of the underlying regime mapping probability but would be relatively straightforward. A challenging aspect of this extension would be generating high joint probability mappings, although approaches for stochastically mapping SSE models exist (Freyman and Höhna 2019).

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. Our simulations have shown how hidden states modeling heterogeneity in the evolutionary process can lead to false character dependent signals. Although we have shown that hOUwie can theoretically resolves this issue, in practice, additional work to improve the computational costs of integrating over all possible stochastic mappings is needed. Nonetheless, the implementation of a joint discrete and continuous model in a likelihood framework improves our inferences about the parameters related to the underlying regimes. 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.

Works Cited

Automatic citation updates are disabled. To see the bibliography, click Refresh in the Zotero tab.