Chapter 15
Modelling Stabilizing Selection: The Attraction of Ornstein–Uhlenbeck Models

Brian C. O'Meara and Jeremy M. Beaulieu

Abstract Ornstein–Uhlenbeck models are a generalization of Brownian motion models that allow trait values to evolve to follow optima. They have become broadly popular in evolutionary studies due to their ability to better fit empirical data as well as for the biological conclusions which can be drawn based on their parameter estimates, especially optimum trait values. We include a survey of available software implementing these models in phylogenetics as well as cautions regarding the use of this software.

15.1 Introduction

The mean value of a trait in a species is affected by multiple factors: physical constraints on evolution, lack of variation, change due to finite population size, and trade-offs between different optima. From one generation to the next, a trait value could change due to processes such as genetic drift, selection towards an optimum, or mutational pressure. If these movements are independent and identically distributed and have an additive effect through time, by the central limit theorem, evolution will fit a Brownian motion process (if the movements have a multiplicative effect through time, the log of the trait value will be evolving under Brownian motion). An Ornstein–Uhlenbeck (OU) process would better describe the process if these movements tended to be in the direction of a particular trait value (such that species with a trait value larger tend to evolve a smaller trait value).

B. C. O'Meara (✉)
Department of Ecology and Evolutionary Biology, Knoxville, TN, USA
e-mail: bomeara@utk.edu

J. M. Beaulieu
National Institute for Biological and Mathematical Synthesis, University of Tennessee, Knoxville, TN 37996, USA

By way of a rough example, consider the position of a toddler attached to their parent by an elastic band. On average, the toddler's position is centred on the parent's position. This mean trait position (with trait units) is often denoted θ and called the "optimum" as an analogy to models of adaptive quantitative evolution. In some phylogenetics models, θ is fixed; in others, there can be discrete shifts in optima (equivalent to attaching the parent end of the lead to a different parent), and in others, θ can move (equivalent to a parent walking down the street with toddler attached). The rate at which a toddler wiggles, σ^2 (in units of squared trait value over time), corresponds to the equivalent parameter with Brownian motion. Finally, the strength of pull by the band (α) can also vary: a strong band pulls the toddler back to the parent more rapidly than a weak band does. A different way to express this is in terms of the amount of time that is expected for a trait to move halfway to the mean value (phylogenetic half-life) which is simply $\ln(2)/\alpha$ (and is measured in time units) (Hansen 1997). For more detailed information about OU models, see Chap. 14.

15.2 Utility

In phylogenetics, we can use the OU model to describe the motion of one trait that depends on the state of another trait or the motion of a trait thought to be constrained. Note that in the former case, the independent trait need not be explicitly included in the model. OU models are often interpreted as models of adaptation, with θ thought to be an adaptive optimum and σ^2 thought to be variance due to genetic drift. This is generally an incorrect interpretation, however. As noted by Hansen (1997), on a macroevolutionary timescale, there is an almost instantaneous movement of a trait to its optimum, in contrast to the half-life of millions of years often discovered from application of OU models. Such a model, therefore, describes how a trait optimum itself moves. If a trait value for a species is not at an optimum, it is more likely that the true optimum of that species is not at the θ given in the model rather than that the trait value is itself far from the optimum.

On the whole, OU models in phylogenetics are simply phenomenological models of optimum movement rather than quantitative genetics models of adaptation within species. Nonetheless, these models have substantial utility. First, they can adjust for nonlinear accumulation of variance with time. Under Brownian motion, two diverging species have trait variance that increases linearly with time: species sharing an ancestor 50 MYA have twice the trait variance of species sharing an ancestor 25 MYA. However, on a long timescale, we might expect this rate to slow down: two sister species of flowering plants may differ in height by 10 % after just a short divergence time, but we do not expect over many millions of years to have one species microscopic and the other taller than the tallest redwood. This pattern of rate of trait divergence slowing through time (due to factors such as soft constraints on trait values) can be fit by using a single mean OU model, which has the effect of shortening especially rootward edges (in units

of expected change) in comparison with the expectation under Brownian motion. This correction can be useful for models such as independent contrasts (Felsenstein 1985) that rely on branch lengths in units of expected change.

These models can also be used to test evolutionary hypotheses. For example, Whittall and Hodges (2007) evaluated the idea that there are three distinct evolutionary optima for nectar spur length based on pollinator type using OU models. They mapped on OU regimes based on pollinator syndrome and compared the fit of an OU model with three θ parameters (one parameter per regime) with fits of models that had just one optimum or Brownian motion. Many studies use a similar strategy of comparing models with multiple pre-assigned optimum trait value parameters with models with a single optimum or a continually moving optimum (Brownian motion). Recent models (Beaulieu et al. 2012) allow for θ, σ^2, and/or α to all vary on the tree. For example, one could investigate whether the rate of evolution, σ^2, varies over the tree but use a single optimum. Figure 15.1 shows simulations of these models.

Another use of these models is in a more exploratory vein. Ingram and Mahler (2013) developed an approach that rather than a priori assignment of θ parameters to regimes on the tree allows the data and tree to drive this assignment (see also Chap. 18). This allows detection of unexpected clumping of optimality parameters. However, this can be interpreted in a hypothesis-testing framework as well. For example, Mahler et al. (2013) examined *Anolis* lizards and find support for convergence of morphological optima, consistent with earlier work on this group (e.g. Losos 1992; Jackman et al. 1997), but also recovered unique optima that could be investigated in the future.

Regardless of whether a model is being used to test a hypothesis or to investigate parameters, one must choose which model to use. Sometimes, this is the actual question: Is a model with regimes mapped based on habitat better than a model with regimes mapped based on diet? In other cases, the relevant question is the actual parameter estimate: Is the optimal body size for mammals in the temperate region greater than the optimal size for tropical mammals? For an investigation where the model chosen is of primary interest, a likelihood ratio test (comparing the fit of two nested models) or a reversible jump MCMC approach, where an algorithm can move between different models, returning the posterior probability of each, would be appropriate. Many biologists incorrectly use the Akaike information criterion as a proxy for significance, though it is appropriate for determining which model loses the least amount of information as well as relative weights for different models. Biology is complex: Ornstein–Uhlenbeck models describe a phenomenological process, and it is unlikely that all taxa in a clade have exactly the same OU parameters. Given enough power, a more complex model is likely to be chosen. For model comparison questions, therefore, it is important to consider carefully what the actual question is, as ability to reject a simple BM or OU model may come more from power than a biological process. For many questions, we recommend a parameter estimation approach instead, rather than merely demonstrating that Brownian motion is rejected in favour of a single mean OU model, which shows that under the best model, phylogenetic half-life is greater than the age of the tree, suggesting a very slow pull towards an
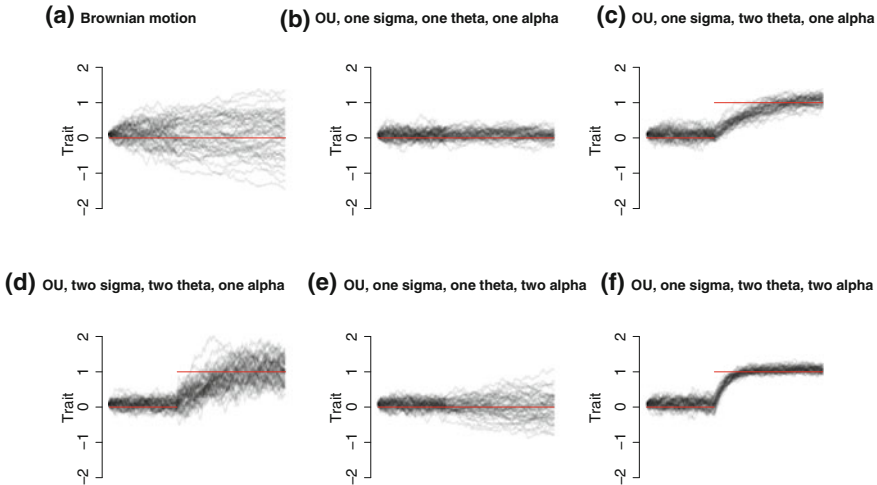
Fig. 15.1 Simulations of OU processes. On all plots, a *red line* shows the θ value for that time period, and 50 independent simulations are shown. **a** shows Brownian motion, with the same σ^2 as used in the other plots. **b** has the same parameters as **a**, except that now α is positive, making it an OU rather than BM process. **c** shows a shift in θ part-way through the simulation. Note how the approach of the simulations to θ slows as they get closer. **d** is the same as **c**, but with σ^2 increasing at the same time as the θ shifts. **e** matches **b**, except that α decreases part-way through the simulation, allowing more deviation from the optimum value. **f** matches **c**, but with α increasing at the same time as the θ shifts. **a** Brownian motion, **b** OU, one sigma, one theta, one alpha, **c** OU, one sigma,two theta, one alpha, **d** OU, two sigma,two theta, one alpha, **e** OU, one sigma,one theta, two alpha, **f** OU, one sigma,two theta, two alpha

optimal value. Biological significance of parameters should be considered as well as statistical significance. Given frequent uncertainty in choosing the best model, a multimodel inference approach (Burnham and Anderson 2004) may be appropriate where inferences about parameter values are based on multiple models weighted by their fit (often using Akaike weights) rather than just from a single model.

15.3 Historical Development

Ornstein–Uhlenbeck models have a long history in physics (Uhlenbeck and Ornstein 1930; Doob 1942) and have also been used in finance (Barndorff-Nielsen and Shephard 2001). Within phylogenetics, their adoption was proposed by Felsenstein (1988), yet their widespread use can be traced to their advocacy by Hansen (1997) and development in an information-theoretic context by Butler and King (2004). In fact, Butler and King's (2004) R package OUCH was the first to provide biologists a useful framework for analysing models of Brownian motion and Ornstein–Uhlenbeck models with one or more means. With OUCH, the means are painted as "regimes" on the tree by assigning different evolutionary parameters to different

branches of a tree. For example, one could imagine that all parts of the tree that are reconstructed as having bats as pollinators may be assigned to have θ_1, while the parts of the tree reconstructed as having insect pollinators could be assigned θ_2. One could also assign regimes based on taxonomy: assign one θ parameter to the angiosperm clade and a different θ parameter to the paraphyletic group of non-angiosperm vascular plants. The optimal values for these parameters can be found under likelihood. Models (differing in number and mapping of regimes, as well as non-OU models) may be compared using the Akaike information criterion (Akaike 1973). Chapter 12 covers model selection in more detail. In this way, a model with one mean parameter can be compared with one with multiple mean parameters. Beaulieu et al. (2012) extended this by allowing for θ, σ^2, and/or α to all vary on the tree, in their R package OUwie. This also allows painting of sets of model parameters (regimes) on different parts of the tree. In theory, regimes may change within a branch, but by default, many programs assume no more than one change per branch and assign regimes to nodes. In OUwie, if two ends of a branch differ in regime, the regime change is assumed to happen halfway along a branch; in OUCH, in this situation, the regime change is assumed to happen at the beginning of the branch. In OUwie, stochastic character mapping (Huelsenbeck et al. 2003) from the R package phytools (Revell 2012) can be used to reconstruct the state of a discrete character everywhere along a tree and then assign regime based on this reconstruction. However, this can cause issues if there is uncertainty in this mapping (Revell 2013), as some parts of branches will be misassigned to the wrong regime.

A separate trend has been the development of multivariate models. Hansen et al. (2008) developed an approach to relate a character evolving under an OU process to a mean evolving under a BM model. This allows for a natural lag between the state of the predictor variable (the trait evolving under Brownian motion) and the state of the character evolving under OU (which has an optimum value that depends on the state of the predictor variable). This was later extended by Bartoszek et al. (2012) to a case where multiple characters are coevolving, perhaps in addition to a predictor variable.

A quasi-multivariate approach was developed by Ingram and Mahler (2013) in the R package SURFACE. This wraps OUCH, so uses its model with a single σ^2 and α value for a character and one or more θ values. However, it tries different painting of θ regimes over the tree and finds the painting that minimizes the information lost in the model. Its quasi-multivariate nature is that the painting of the regimes, but not the parameter values themselves, is shared across multiple characters, allowing them to jointly inform this placement.

Finally, an OU model with a single regime everywhere on the tree has the effect of simply transforming branch lengths, in the same way that Pagel's kappa or lambda (Pagel 1997, 1999) are ways to stretch a tree so that branch lengths better represent the evolutionary process. Thus, there are several software packages that can choose a single α value that results in a set of branch lengths that best fit a model for available data, either for a single trait or for dealing with phylogenetic non-independence when doing regression or correlation between multiple traits. These include PHYSIG (Blomberg et al. 2003), ape (Paradis et al. 2004),

Table 15.1 Models implementing OU processes for trait evolution

Software name	Citation	Maximum number of regimes	Allows variation in θ parameters	Allows variation in σ^2 parameters	Allows variation in α parameters	Regression	Ancestral state estimation at multiple nodes	Software type
PHYSIG	Blomberg et al. (2003)	One per tree	N	N	N	Y	N	MATLAB scripts
OUCH	Butler and King (2004)	One per branch	Y	N	N	N	N	R package
Ape	Paradis et al. (2004)	One per tree	N	N	N	N	Y	R package
COMPARE	Martins (2004)	θ varies with other traits	Y	N	N	Y	Y	Java program
Geiger	Harmon et al. (2008)	One per tree	N	N	N	N	Y	R package
SLOUCH	Hansen et al. (2008)	θ varies with other traits	Y	N	N	Y	N	R scripts
OUwie	Beaulieu et al. (2012)	Arbitrarily high	Y	Y	Y	N	N	R package
MvSLOUCH	Bartoszek et al. (2012)	θ varies with other traits	Y	N	N	Y	N	R package
SURFACE	Ingram and Mahler (2013)	One per branch	Y	N	N	N	N	R package
Phylolm	Ho and Ané (2014)	One per tree	N	N	N	Y	N	R package

COMPARE (Martins 2004), geiger (Harmon et al. 2008), and phylolm (Ho and Ané 2014). Table 15.1 compares multiple software packages which implement OU models for phylogenetics.

15.4 Caveats

With Brownian motion, the expected value after any amount of time is the initial value. Even so, as time information about ancestral state decays, so does information about potentially different models operating deeper in the tree. With Ornstein–Uhlenbeck processes, the expected value is a weighted average of the initial value and the optimal value. Longer amounts of time, and stronger attraction parameters, mean that historical signal will begin to disappear. Even for large trees, there may be very little information about past regimes. Uncertainty in returned parameter values may be estimated by using an approximation based on the slope of the likelihood surface at its point of maximum likelihood (Beaulieu et al. 2012). A better way to estimate this is to look at the actual likelihood surface over a range of parameter values (as in SLOUCH (Hansen et al. 2008)). Parametric bootstrapping (simulating under the recovered model) is another way to estimate uncertainty: Under the assumption that a model is true, what distributions of parameter estimates are recovered if evolution were rerun under that model? High α values erase history about past processes, but low α values may also be problematic in that they make multiple θ parameters more difficult to estimate, as the final trait values could depend less on θ. In addition to issues arising with a low or high α, some of the more complex models with multiple α and σ^2 values can also be difficult for parameter estimation. The number of regimes on a tree can increase without limit, as each branch can be broken into multiple regimes, and this can rapidly exhaust any information in the data. Even assuming no more than one regime per branch and thus a branch-specific estimate of α, σ^2, and θ, on a tree with N taxa, there are $3 \times (2N-2)$ parameters to estimate but no more than $N \times$ number of characters (typically one) to provide data.

Dealing with the state at the root can also be problematic. With Brownian motion, the root state is estimated based on the parameter values and branches and does not depend on the rate of evolution. With Ornstein–Uhlenbeck processes, the root state can be estimated, but where there is little information about the past due to strong attraction, it is biased towards zero (Beaulieu et al. 2012). Some software, such as OUwie or phylolm, default to assuming the root state comes from the stationary distribution of the evolutionary process, but has an option to separately estimate the root state. However, the lack of information about the root state can mean that separately estimating this parameter can lead to inaccurate estimates for other parameters. It is important to note that while OUCH originally estimated a state at the root, more recent versions of the package assume stationarity. Bartoszek et al. (2012) fix the root state at the optimal value for the regime on the root branches.

A related issue is whether Brownian motion is a restriction of an Ornstein–Uhlenbeck model. In other words, as α approaches zero, do the model's parameter estimates and likelihood converge towards those of a Brownian motion model? There is, of course, the caveat that the uncertainty in estimates of θ should increase without bound as one approaches α of zero. Nevertheless, since the treatment of the root state differs between OU model implementations, but not between Brownian motion implementations, in some software (such as OUwie or geiger), an OU model with one regime approaches Brownian motion as α approaches zero. However, in others, such as the current version of OUCH, the implemented Brownian motion model is not nested within the implemented OU model: the likelihood of an OU model with a single θ does not converge to the likelihood of a Brownian motion model as α approaches zero. This is an active area of discussion that has yet to be resolved.

An important issue when dealing with any model is understanding what the parameters mean. The θ parameter is in the units of the trait(s) under investigation (i.e., kg for body mass), and this is true regardless of implementation. The σ^2 rate is in units of trait units squared over branch length units (i.e., kg/MY), and α is in units of reciprocal branch length units (i.e., MY^{-1}). Phylogenetic half-life has time units (i.e., MY). One unfortunate trend, in both our work and the work of others, is to fail to report these units. Users should note that some programs in this area rescale trees before analysis. One common rescaling is dividing each branch length by the total height of the tree, which makes all root to tip lengths one. This can help deal with numerical issues arising from software with finite precision, but it also means that σ^2 and α will not have the same units as in the original tree, hindering interpretation. However, these parameters can be rescaled to the original units later in the process.

Biologists have become used to computational issues involved in tree inference and so treat such inferences with caution; fitting a single model to an existing tree can seem like a trivial issue in comparison. However, estimating the likelihood of an OU model and getting good parameter estimates for it can, in practice, be difficult. Part of this stems from nearly flat likelihood surfaces: numerical optimization may terminate before reaching a peak if changes in parameter values have only a slight effect on the likelihood. Certain pairs of parameters may also tend to form a ridge in likelihood space, making them difficult to optimize. There is also the temptation to use overly complex models: a thirty-taxon dataset may simply not have enough information in one character to estimate multiple α values. While model selection approaches should not choose models that are too complex for the available data, they are not guaranteed to work all the time. The details of the mathematical steps used to calculate likelihood for OU models in many software packages can make these methods prone to encountering errors due to finite precision, though these errors may be hidden from most users (but may still affect whether optimization works properly). Many of the developers of software packages are biologists or mathematicians foremost who may lack the expertise to test and optimize every part of a program. For these various reasons, users of OU software should examine results with some skepticism. The open source nature of

software in this field will allow users to examine its interior workings and do things such as try different starting points for parameter optimization (some programs do this automatically, others do not). However, even more basic tests can be performed without knowing how to program that will give a sense of whether the parameter estimates coming back are reasonable or may reflect a problem with the software. For example, if the exact same analysis is run again, are the same likelihood scores and parameter estimates returned each time? Inconsistency may indicate that the search is sensitive to starting values and that more runs must be attempted to find the best values. While not all models compared need to be nested, in cases where one model is truly a restriction of another model, the likelihood of the restricted model (but not the AIC score) must be the same or worse than the likelihood of the general model: Does this occur? For programs that allow user specification of fixed points at which to evaluate likelihood, what is the shape of the likelihood surface: Are points near the returned maximum likelihood estimates of the parameters always worse in likelihood than what the program returned as the best values? In the case where two programs implement the same model, do they return the same parameter estimates for the same data (though note the issue about potential rescaling, above)? Is a returned parameter value at one of the preset maximum or minimum bounds of the software? If so, it may make sense to change this bound, as the maximum likelihood estimate is probably outside this region, though this may result in numerical precision problems within the program (often a reason for setting default bounds). Doing these steps takes time, but spending a few extra days to verify that the returned results are correct may be a worthwhile investment after the months to years it can take to get the phylogeny and trait data required to address a biological question using these models.

15.5 Example

Within flowering plants, there is a strong growth form-dependent distribution in genome size (i.e., the amount of DNA in any given cell), with woody species containing smaller genome sizes, on average, as well as lower overall variance, when compared to herbaceous species. It has been suggested (Beaulieu et al. 2008, 2010, 2012) that these patterns largely reflect differences in life history. Woody angiosperms generally take longer to reach reproductive maturity (Verdú 2002), leading to longer generation times and fewer opportunities for random insertion/deletions to occur on a per unit time basis. Indeed, woody lineages consistently show slower overall rates of genome size evolution when compared to herbaceous lineages (Beaulieu et al. 2010, 2008).

Here, we illustrate how the OU framework can be used to uncover rate differences in genome size evolution between woody and herbaceous growth form states within the Fabaceae (i.e., legumes). This example is further fleshed out in the Online Practical Material (hereafter OPM) available at http://www.mpcm-evolution.org. Genome size estimates were taken from the Plant DNA

C-value database (Bennett and Leitch 2010), and we arrayed onto the time-calibrated phylogeny of legumes from Beaulieu et al. (2010). We focus these analyses on the monoploid genome size, or the 1Cx value, in order to correct for the possibility that polyploidy can inflate rates of genome size evolution. The monoploid genome size represents the amount of DNA in the unreplicated monoploid chromosome set and is calculated by dividing the 2C DNA amount by ploidy. These values were $\log_{10}$-transformed prior to all analyses to ensure that these data minimally conformed to Brownian motion evolution (Oliver et al. 2007).

For this analysis, regimes are mapped on the tree based on likelihood estimation of ancestral states using the package corHMM (Beaulieu et al. 2013) though stochastic character mapping, parsimony, or other ways of assigning regimes to branches could be used. A variety of continuous trait models are then investigated. These include Brownian motion with a single rate ("BM1" in the program), Brownian motion with a different rate allowed for each discrete state regime ("BMS"), OU with a single optimum ("OU1"), OU with a different optimum for each regime, but with a constant α and rate of evolution ("OUM"), and finally a model with a different optimum and rate of evolution for each regime ("OUMV"). Other models are available, such as one that varies optimum, α, rate of evolution for each regime or various other restrictions of this model, but the set of models for the example is limited to a workable set which can run relatively quickly.

For all the models, the parameter estimates as well as the AIC with small sample correction (AICc) are stored. The best (smallest) AICc value is subtracted from the AICc values for each model to get the ΔAICc value. In this example, the best model is OUMV, a model with one α across the tree but different σ^2 and θ for each discrete state regime. The next best model has a ΔAICc much higher (42.8), which corresponds to an Akaike weight that is much tinier (<1 billionth) than that of the best model. If support were much more similar across models, model averaging would make sense as a way to deal with uncertainty in the models, but given this difference is would not be expected to help. In fact, model averaging may even be problematic if a complex model with very little support generates very bad parameter estimates, as the low weight on the model might not be enough to counter the magnitude of the poor estimates. For the best model, the optimal value for herbaceous plants was 0.254 ± 0.036 $\log_{10}$ (pg), which corresponds to 1.289 ± 1.037 pg, while for woody plants, it was 0.904 ± 1.029 pg after transformation out of log-space. The half-life of the process is 0.115 MY, while the tree is 59 MY old, suggesting a strong pull to each of these values. The rate of evolution of genome size while herbaceous is over five times greater than the rate of evolution while woody (2.968 ± 0.303 $\log_{10}$(pg)/MY vs. 0.574 ± 0.105 $\log_{10}$ (pg)/MY).

There are two lessons from this analysis. The first is that history does not matter much for genome size in this group: the short half-life indicates that the trait value of a species is quickly pulled from its ancestral state to whatever the optimal state is. However, the optimal states seem very similar, and each is within one standard error of the other, suggesting little evidence for different optimal states. In contrast, the σ^2 rates do differ, having a biologically significant fivefold rate difference as well as a statistically meaningful rate difference. This suggests that while history

does not matter much for mean value, state matters a great deal for rate of evolution. This analysis also points out a limitation of software: given the similarity in OU means across states but not OU σ^2 parameters, an even better model might be one with one α and θ across the tree but different σ^2 for each discrete state (in OUwie's jargon, this would be an OUV model) but this has yet to be implemented.

15.6 Future Directions

There are numerous potential advances in this area. One trivial advance would be inference of ancestral states under an OU process. This is possible with a single OU mean tree transform, as can be implemented in geiger (Harmon et al. 2008) or COMPARE (Martins 2004), but has not yet been implemented in software for more complex OU models. Another straightforward advance would be a wrapping of the Beaulieu et al. (2012) family of OU models in the SURFACE (Ingram and Mahler 2013) approach to check for regime shifts for σ^2 and α in the same way this is done for OUCH-type models.

Most work in this area has been in a regression or information-theoretic framework, and the utility of Bayesian approaches has yet to be explored fully. They have potential in allowing a way to bring in information from external sources about parameters as priors without having to fix this information. However, given frequent uncertainty in parameter estimates in these models, it will be essential to make sure that the results are driven in part by the data rather than only reflecting the priors.

Ornstein–Uhlenbeck models are among the most complex models of continuous trait evolution available to date. They give information about the parameters of evolutionary change on a macroevolutionary timescale but may not reflect microevolutionary processes (Hansen 1997). Models that operate at the level of population genetics mechanisms may be important in the future to allow inferences of processes rather than just fitting evolutionary patterns. Continued development of multivariate approaches remains important as well.

References

Akaike H (1973) Information theory as an extension of the maximum likelihood principle. In: Petrov BN, Csaki F (eds) Second international symposium on information theory. Akademiai Kiado, Budapest, pp 267–281

Barndorff-Nielsen OE, Shephard N (2001) Non-gaussian Ornstein–Uhlenbeck-based models and some of their uses in financial economics. J Roy Stat Soc Ser B (Stat Methodol) 63(2):167–241. doi:10.2307/2680596

Bartoszek K, Pienaar J, Mostad P, Andersson S, Hansen TF (2012) A phylogenetic comparative method for studying multivariate adaptation. J Theor Biol 314:204–215

Beaulieu JM, Jhwueng D-C, Boettiger C, O'Meara BC (2012) Modeling stabilizing selection: expanding the Ornstein–Uhlenbeck model of adaptive evolution. Evolution 66(8):2369–2383

Beaulieu JM, Leitch IJ, Patel S, Pendharkar A, Knight CA (2008) Genome size is a strong predictor of cell size and stomatal density in angiosperms. New Phytol 179(4):975–986

Beaulieu JM, O'Meara BC, Donoghue MJ (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

Beaulieu JM, Smith SA, Leitch IJ (2010) On the tempo of genome size evolution in angiosperms. J Bot. doi: 10.1155/2010/989152

Bennett MD, Leitch IJ (2010) Plant DNA C-values database (release 6.0, Dec. 2012)

Blomberg SP, Garland T, Ives AR (2003) Testing for phylogenetic signal in comparative data: behavioral traits are more labile. Evolution 57(4):717–745

Burnham KP, Anderson DR (2004) Multimodel inference—understanding AIC and BIC in model selection. Sociol Methods Res 33(2):261–304

Butler MA, King AA (2004) Phylogenetic comparative analysis: a modeling approach for adaptive evolution. Am Nat 164(6):683–695

Doob JL (1942) The Brownian movement and stochastic equations. Ann Math 43(2):351–369

Felsenstein J (1985) Phylogenies and the comparative method. Am Nat 125(1):1–15

Felsenstein J (1988) Phylogenies and quantitative characters. Annu Rev Ecol Syst 19:445–471

Hansen TF (1997) Stabilizing selection and the comparative analysis of adaptation. Evolution 51(5):1341–1351

Hansen TF, Pienaar J, Orzack SH (2008) A comparative method for studying adaptation to a randomly evolving environment. Evolution 62(8):1965–1977

Harmon LJ, Weir JT, Brock CD, Glor RE, Challenger W (2008) GEIGER: investigating evolutionary radiations. Bioinformatics 24(1):129–131

Ho LST, Ané C (2014) A linear-time algorithm for Gaussian and non-Gaussian trait evolution models. Syst Biol 63(3): 397–408

Huelsenbeck JP, Nielsen R, Bollback JP (2003) Stochastic mapping of morphological characters. Syst Biol 52(2):131–158. doi:10.1080/10635150390192780

Ingram T, Mahler DL (2013) SURFACE: detecting convergent evolution from comparative data by fitting Ornstein–Uhlenbeck models with stepwise Akaike Information Criterion. Methods Ecol Evol 4(5):416–425

Jackman T, Losos JB, Larson A, de Queiroz K (1997) Phylogenetic studies of convergent adaptive radiations in Caribbean Anolis lizards. In: Molecular evolution and adaptive radiation. pp 535–557

Losos JB (1992) The evolution of convergent structure in Caribbean Anolis communities. Syst Biol 41(4):403–420

Mahler DL, Ingram T, Revell LJ, Losos JB (2013) Exceptional convergence on the macroevolutionary landscape in island lizard radiations. Science 341(6143):292–295

Martins EP (2004) COMPARE, version 4.6 b. Computer programs for the statistical analysis of comparative data. Distributed by the author at http://compare.bio.indiana.edu/, Department of Biology, Indiana University, Bloomington, IN

Oliver MJ, Petrov D, Ackerly D, Falkowski P, Schofield OM (2007) The mode and tempo of genome size evolution in eukaryotes. Genome Res 17(5):594–601

Pagel M (1997) Inferring evolutionary processes from phylogenies. Zoolog Scr 26(4):331–348

Pagel M (1999) Inferring the historical patterns of biological evolution. Nature 401(6756):877–884

Paradis E, Claude J, Strimmer K (2004) APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20(2):289

Revell LJ (2012) Phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol Evol 3(2):217–223

Revell LJ (2013) A comment on the use of stochastic character maps to estimate evolutionary rate variation in a continuously valued trait. Syst Biol 62(2):339–345

Uhlenbeck GE, Ornstein LS (1930) On the theory of the Brownian motion. Phys Rev 36(5):823–841

Verdú M (2002) Age at maturity and diversification in woody angiosperms. Evolution 56(7):1352–1361. doi:10.1111/j.0014-3820.2002.tb01449.x

Whittall JB, Hodges SA (2007) Pollinator shifts drive increasingly long nectar spurs in columbine flowers. Nature 447(7145):706–709