

OUwie 2.0

Jeremy M. Beaulieu and Brian C. O'Meara

It's been quite a while since we've done any active development of **OUwie**. However, it was recently brought to our attention that the likelihoods between **OUwie** and **ouch** when the model assumes different OU means (what we refer to as "OUM"). We would like to thank and credit Clay Cressler for working through our code to identify these issues. So, we have taken this opportunity to release a new version of **OUwie** that corrects these issues as well as provide a number of new capabilities. Some of these new functions were implemented at the request of users, for our own research, and as part of tutorials for various workshops.

Bug fixes

There was a bug in the most recent version of the weight matrix generation code. For some reason, while looping over the different regimes, the function was not resetting the regime to the regime at the root, and the calculation for the weights for the root regime was missing a $W_{0,i} + e^{(-\alpha_{0,i})}$ term. These are now fixed, and fortunately does not cause any detectable effect on the likelihood, though it does impact the estimates of the regime optima.

We have also added modification to the way the variance-covariance matrix is constructed when assuming stationarity at the root. In the original Beaulieu et al. (2012) we expanded upon equation A5 in Butler and King (2004), which assumed that the root optima was estimated. However, by default we drop the root optima and assume that the system started at stationarity. Ho and Ane (2013) showed that the covariance calculated assuming under the stationarity there is an additional variance term that has to be added to this calculation. For example, consider two tips that diverge at the root. Under the original formulation of Butler and King (2004) the covariance between the two tips was zero because $s_{ij} = 0$, so $1 - e^{-2\alpha s_{ij}} = 0$. With the Ho and Ane (2013) method, $V_{ij} = \frac{\sigma^2}{2\alpha} e^{-2\alpha}$ is added to account for the fact that, up until $T = 0$, the lineage is assumed to have been evolving in the ancestral regime.

New Features

BIC – Ho and Ane new sample size corrections – BRIAN

Identifiability tests – BRIAN

Matrix conditioning – BRIAN

Ancestral trait reconstruction

Since **OUwie** was first released we've received a deluge of requests to allow for ancestral trait reconstruction. One such request made its way onto the public R-SIG-PHYLO discussion forum, which stimulated an important conversation about not whether or not you **could** estimate ancestral states, but, rather, **should** you. The answer is "it's complicated" and we don't recommend it. In our view, the intended use case is just to visualize what the model is saying about evolution to help intuition. For example, is the model something you can believe in? But, if you don't want to listen to us regarding the merits of ancestral trait reconstructions, here is a sample of comments from other experts in the field:

So in short, yes, you can do it, with any number of methods. But why? If you can answer your biological question with methods that do not involve estimation of a parameter that is inherently fraught with error, it might be better to go another way. Bottom line - use caution and be thoughtful! – Marguerite Butler

I would add an extra caveat to Marguerite's excellent post: Most researchers work with extant taxa only, ignoring extinction. This causes a massive ascertainment bias, and the character states of the extinct taxa can often be very different to the ancestral state reconstructions, particularly if the evolutionary model is wrong. E.g. there has been an evolutionary trend for example. Ancestral state reconstructions based only on extant taxa should be treated as hypotheses to be tested with fossil data. I wouldn't rely on them for much more. – Simone Blomberg

While I am at it, let me echo Simone and Marguerite's warnings. The predicted ancestral states will reflect the process you assumed to predict them. Hence, if you use them to make inferences about evolution, you will recover your own assumptions. I.e. if you predict from a model with no trend, you will find no trend, etc. Many comparative studies are flawed for this reason. – Thomas Hansen

Let me add more warnings to Marguerite and Thomas's excellent responses. People may be tempted to infer ancestral states and then treat those inferences as data (and also to infer ancestral environments and then treat those inferences as data). In fact, I wonder whether that is not the main use people make of these inferences. But not only are those inferences very noisy, they are correlated with each other. So if you infer the ancestral state for the clade (Old World Monkeys, Apes) and also the ancestral state for the clade (New World Monkeys, (Old World Monkeys, Apes)) the two will typically not only be error-prone, but will also typically be subject to strongly correlated errors. Using them as data for further inferences is very dubious. It is better to figure out what your hypothesis is and then test it on the data from the tips of the tree, without the intermediate step of taking ancestral state inferences as observations. The popular science press in particular demands a fly-on-the-wall account of what happened in evolution, and giving them the ancestral state inferences as if they were known precisely is a mistake. – Joe Felsenstein

The minor twist I would throw in is that it's difficult to make universal generalizations about the quality of ancestral state estimation. If one is interested in the ancestral state value at node N, it might be reasonably estimated if it is nested high up within the phylogeny, if the rates of change aren't high, etc. And (local) trends etc might well be reliably inferred. We are pretty confident that the common ancestor of humans and chimps was larger than many deeper primate ancestors, for instance. If N is the root of your available phylogeny, however, you have to be much more cautious. – Nick Matzke

I'll also add that I think there's a great deal to be skeptical of ancestral trait reconstruction even when large amounts of fossil data is available. You can try the exercise yourself: simulate pure BM on a non-ultrametric tree with lots of 'extinct' tips, and you'll still find pretty large confidence intervals on the estimates of the trait values. What does it mean to do ancestral trait reconstruction, if our calculations of uncertainty are that broad? – Dave Bapst

These people probably know better than anyone about the power and limitations of the OU model in phylogenetics. So, don't listen to us, listen to them!

Still determined? Ok, fair enough. It is straightforward to run the ancestral trait reconstruction in `OUwie`. All you need is an object of class `OUwie`, which is plugged directly into the new `OUwie.anc()` function:

Ah! If you run the code above a somewhat snarky response is printed to the screen. Since you are here reading this, we think this is sufficient, and you are at least aware that we are not huge fans of this approach. So, there is no need to read the manual (it's basically the same as what you see here), and the proper way to call the `OUwie.anc()` function call is simply add in `knowledge=TRUE`:

```
recon <- OUwie.anc(ouwiefit, knowledge=TRUE)
```

From there, you can then easily plot these reconstructions using an internal function that recognizes the `OUwie` and `OUwie.anc` class:

```
plot(recon, fsize=0.5)
```

