**Phylogenetic analyses for my convergence and disparity studies.**

*18th February 2014*

*These are notes which outline my approach and rational for the phylogenetic part of my convergence and disparity analyses. I can use them when it comes to writing the methods part of my chapter in a similar way to the notes that I created for each of my morphometric analyses*

I have a distribution of mammal super trees from ([Kuhn et al., 2011](#_ENREF_3)). They used mammalian super trees with approximately 50% of nodes as polytomies and developed a new method of resolving the polytomies to give dichotomous trees. So the result is a distribution of phylogenies which differ in how the polytomies have been resolved. I have two files; 100 and 10,000 of these resolved trees.

I will need to use a distribution of trees (I don’t know how many yet) but I worked out how to do various manipulations using a single tree first.

For each super tree, I pruned out the species which are not in my morphological data set (NB: I’m using the corrected taxonomy for my morphological data). This means that I will need separate phylogenies for each of the morphological data sets since some of the species involved are slightly different (e.g. if I didn’t have any unbroken skulls or mandibles for one particular species which is still represented in the other morphological data sets).

However, not all of the species which are in my morphological data are included in the phylogenies. Six of these missing species are Microgale tenrecs so I used the phytools add.species.to.genus function ([Revell, 2012](#_ENREF_9)) to add these missing species at random into the Microgale genus. However, I couldn’t use the same approach to add the missing golden mole species because there weren’t any other members of their genus in the phylogeny. Therefore, I identified the node in the phylogeny which is the most recent common ancestor of all of the golden moles (findMRCA function in phytools) and added the additional golden mole species as extra tips to this common ancestral node. This creates new polytomies within the phylogeny. However, for further analyses which involved calculating pairwise phylogenetic distances among species I needed a dichotomous tree. One way to deal with this issue is with the multi2d function in ape ([Paradis et al., 2004](#_ENREF_8)) which randomly transforms multichotomies into a series of dichotomies with branches of length 0.

The problem here is that there are obviously going to be many possible ways of resolving the multichotomies, which is the problem that ([Kuhn et al., 2011](#_ENREF_3)) faced when producing the resolved super trees in the first place. They used a Bayesian approach to create a distribution of resolved trees, resolution which that I destroyed by adding in the extra species. Given that I’m planning on using a distribution of these resolved trees for my analysis, I could, theoretically at least, create a second distribution for each of those trees with the multiple ways of resolving the polytomies that I created.

However, the variation with which those polytomies are resolved should only affect the phylogenetic distances within groups of species. For example, the exact placement of the added Microgale tenrecs will influence the calculated phylogenetic distances within Microgale but the overall phylogenetic distances between any of the Microgale and other species will not change. Similarly, the way the polytomies in the golden moles are resolved will affect phylogenetic distances within golden moles but not between that family and other species. For the purposes of convergence studies I’m interested in phylogenetic distances among unrelated species pairs so these issues don’t matter. There will, however, be an issue when it comes to disparity studies since those calculations rely on estimates of phylogenetic distances within family groups (tenrecs or golden moles).

Assuming that this rational for dealing with adding species into the phylogenies is acceptable the steps above give a distribution of phylogenies containing all of the species in each of my morphological data sets. The next step is to use these phylogenies for comparisons of morphological and phylogenetic distances among pairs of species and also for simulations of shape evolution which are then applied to both the convergence and disparity measurements. The issue is whether I repeat these analyses across the distribution of phylogenies (multiple comparisons of shape and phylogenetic distances and multiple simulations for each phylogeny) or do I use a single tree for the analyses. It’s almost certainly better to use a distribution of phylogenies.

In the longevity paper, Kevin and Thomas treated Kuhn’s distribution of independently resolved mammal trees as the equivalent of a Bayesian posterior distribution of trees because no such tree analysis exists for all mammals. They fitted the models to each of the trees and then combined the model outputs to give model estimates which incorporate error across the 500 trees – so you get a posterior distribution for each coefficient. In terms of adapting this for my data it means calculating pairwise phylogenetic distances among species for every tree in the distribution. So I’d get an estimated value and confidence intervals for the phylogenetic distance between two species. Is the next step then to use the estimated values for the morphological vs. phylogenetic distances plot?

I still don’t know how to deal with simulating morphlogical evolution across multiple phylogenies. The Revell 2007 paper below might have a solution so I need to go back and read that again.

**19th February 2014**

I created the SkVent\_100phylogenies script to deal with the phylogeny for my ventral skull data set. I used the smaller sample of phylogenies from the Kuhn paper (turns out to be 101 rather than 100 phylogenetic trees). I pruned each tree, added the missing Microgale species to the Microgale genus and added the missing golden mole species to the ancestral node of all golden moles. Adding the golden moles created polytomies in the tree so I used multi2di to randomly resolve the polytomies (I included some steps to support the idea that random resolution of the polytomies doesn’t make a huge difference).

I calculated pairwise phylogenetic distances for each phylogeny so now I have 101 phylogenetic difference matrices for each species pair.

**Phylogenies used in other convergence studies.**

([Mahler et al., 2013](#_ENREF_4)) SURFACE paper – single tree - maximum clade credibility from a Bayesian phylogenetic inference

([Moen et al., 2013](#_ENREF_6)) – single trees; compared results using a maximum likelihood tree from one study and a consensus tree from post-burnin sample from a Bayesian distribution

([Kawahara and Rubinoff, 2013](#_ENREF_2)) – maximum likelihood and Bayesian consensus trees were identical – they used a distribution of trees but not for pairwise contrasts analyses

([Segar et al., 2013](#_ENREF_11)) – single consensus tree from a Bayesian phylogenetic estimation

([Muschick et al., 2012](#_ENREF_7)) – single maximum likelihood phylogeny

([Revell et al., 2007](#_ENREF_10)) – maximum likelihood and Bayesian distribution trees; simulated character evolution using the variance-covariance matrix of independent contrasts for the 11 morphological characters calculated separately for each phylogeny and then pooled across phylogenies – Might be a solution for my multiple phylogenies issue so I need to come back to this paper

([Melville et al., 2006](#_ENREF_5)) – single maximum likelihood tree

([Harmon et al., 2005](#_ENREF_1)) – majority-rule consensus tree from a Bayesian posterior distribution of 180 trees and calculated support values from the distribution (the proportion of each of the trees which included the branching patterns in the consensus tree)

([Stayton, 2006](#_ENREF_12)) – compared two single phylogenies from 1988 and 1998 – presumably they’re maximum likelihood but I haven’t checked

Harmon, L. J., Kolbe, J. J., Cheverud, J. M. & Losos, J. B. 2005. Convergence and the multidimensional niche. *Evolution,* 59**,** 409-421.

Kawahara, A. & Rubinoff, D. 2013. Convergent evolution of morphology and habitat use in the explosive Hawaiian fancy case caterpillar radiation. *Journal of evolutionary biology,* 26**,** 1763-1773.

Kuhn, T. S., Mooers, A. Ø. & Thomas, G. H. 2011. A simple polytomy resolver for dated phylogenies. *Methods in Ecology and Evolution,* 2**,** 427-436.

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

Melville, J., Harmon, L. J. & Losos, J. B. 2006. Intercontinental community convergence of ecology and morphology in desert lizards. *Proceedings of the Royal Society B: Biological Sciences,* 273**,** 557-563.

Moen, D., Irschick, D. & Wiens, J. 2013. Evolutionary conservatism and convergence both lead to striking similarity in ecology, morphology and performance across continents in frogs. *Proceedings of the Royal Society B: Biological Sciences,* 280**,** 20132156.

Muschick, M., Indermaur, A. & Salzburger, W. 2012. Convergent evolution within an adaptive radiation of cichlid fishes. *Current Biology,* 22**,** 1-7.

Paradis, E., Claude, J. & Strimmer, K. 2004. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics,* 20**,** 289-290.

Revell, L. J. 2012. phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution,* 3**,** 217-223.

Revell, L. J., Johnson, M. A., Schulte, J. A., Kolbe, J. J. & Losos, J. B. 2007. A phylogenetic test for adaptive convergence in rock-dwelling lizards. *Evolution,* 61**,** 2898-2912.

Segar, S., Pereira, R., Compton, S. & Cook, J. 2013. Convergent structure of multitrophic communities over three continents. *Ecology letters,* 16**,** 1436-1445.

Stayton, C. T. 2006. Testing hypotheses of convergence with multivariate data: morphological and functional convergence among herbivorous lizards *Evolution,* 60**,** 824-841.