Reply to Englund: Molecular evolution and diversification—Counting species is better than counting nodes when the phylogeny is unknown

Recently, we demonstrated that rates of molecular evolution are related to clade size (the number of extant species per family) in birds (1). This result confirmed previous findings of a link between molecular rates and net diversification (the net result of species addition by speciation and removal by extinction) but went further by showing that it could not be an artifact of the node-density effect and by providing evidence that the relationship is caused by differences in mutation rate.

Englund (2) suggests that clade size was an inappropriate measure of net diversification because for any chosen species, the number of nodes along the path leading to that species (N) might not reflect the clade size of that family. Clade size and N are both metrics that reflect the number of nodes in a phylogenetic tree. Both metrics reflect the net result of speciation and extinction [plus any ecological limits on diversity (3)], so neither can be interpreted simply as an estimate of the number of speciation events. The two measures differ in that clade size is a property of an entire clade, whereas N is a property of a path leading to a single species. Clade size provides an accurate measure of the number of nodes in a clade even if the relationships between species are unknown, because the number of nodes in a bifurcating tree is one fewer than the number of tips. On the other hand, N will only provide an accurate measure of the number of nodes along a path leading to a single species if the relationships between all extant species are known. Thus, it would have been impossible to accurately estimate N for the species in our study, because a complete phylogeny of all bird species is unknown.

Englund suggests that when the phylogeny is unknown, net diversification should be estimated by taking the mean value of N, for a chosen species, over all possible topologies. However, when calculated in this way, the mean value of N for a chosen species is directly proportional to the logarithm of clade size (4), which is the measure of net diversification that we used. It would be impossible to average over all topologies for even modest clade sizes because of the astronomical number of possible trees (5), and estimating the mean value of N over a large sample of topologies would be a time-consuming and imprecise method of estimating a quantity that can be calculated directly from clade size (4), as demonstrated in Fig. 1.

The effect Englund describes could have added noise to our analyses, which could have obscured the signal of a link between net diversification and rates of molecular evolution, but it could not introduce systematic bias. Given that we found a significant association between clade size and rates of molecular evolution, it does not seem to have had a negative impact on our findings.

*Robert Lanfear*[a,1], *Lindell Bromham*[a], *and Simon Y. W. Ho*[b]
[a]*Centre for Macroevolution and Macroecology, Ecology Evolution and Genetics, Research School of Biology, Australian National University, Canberra ACT 0200, Australia; and* [b]*School of Biological Sciences, University of Sydney, Sydney NSW 2006, Australia*

1. Lanfear R, Ho SYW, Love D, Bromham L (2010) Mutation rate is linked to diversification in birds. *Proc Natl Acad Sci USA* 107:20423–20428.
2. Englund M (2011) Clade size is not always an appropriate measure of diversification. *Proc Natl Acad Sci USA* 108:E84.
3. Rabosky DL (2009) Ecological limits and diversification rate: alternative paradigms to explain the variation in species richness among clades and regions. *Ecol Lett* 12: 735–743.
4. Steel M, McKenzie A (2001) Properties of phylogenetic trees generated by Yule-type speciation models. *Math Biosci* 170:91–112.
5. Edwards AWF, Cavalli-Sforza LL (1964) in *Phenetic and Phylogenetic Classification*, eds Heywood VH, McNeill J (Systematics Association, London), pp 67–76.
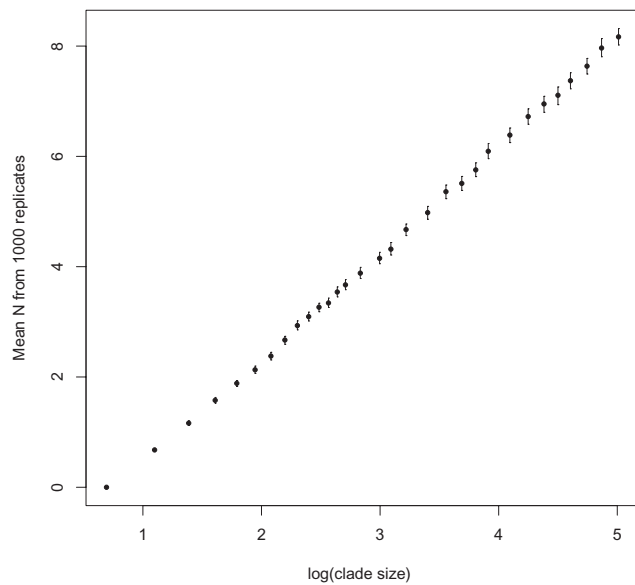
Fig. 1. Relationship between the logarithm of clade size and the mean number of nodes through which a randomly chosen root-to-tip path on a tree passes (mean *N*). Each mean *N* is calculated from 1,000 replicates in which *N* is calculated for a randomly selected tip from a randomly generated Yule tree. The error bars represent 95% confidence intervals calculated by nonparametric bootstrapping.