

## **Blue tail, red tail: towards a solution to inapplicable discrete characters in phylogenetics**

Thomas Guillerme, Martin Smith and Martin Brazeau

guillert@tcd.ie - <http://tguillerme.github.io/>

The last five years have seen a resurgence of the use of discrete morphological characters in phylogenetic studies, whether in terms of data collection, evolutionary models or inference methods. As such study becomes more popular and includes an increase number of taxa and characters, there is also a resurgence of the problem of inapplicable characters. Characters coding can sometimes not be applied to different groups of taxa as illustrated by Maddison (1993)'s blue tail and red tail problem. For example, most vertebrate characters can not be applied on non-vertebrates organisms if combining both groups into a phylogenetic matrix.

Here we propose a solution to this problem by updating ancestral states reconstruction to properly deal with inapplicable data. In other terms, we suggest only considering changes from inapplicable data to applicable data. We've implemented this in a four pass step counting algorithm (parsimony) and tested it on several published matrices. Using this algorithm, we found that the most parsimonious topologies inferred under normal Fitch step counting do greatly vary in terms of length if applying our inapplicable data step counting.

These results suggests that more care should be given in using accurate models that reflect our ways of coding discrete morphological characters. In fact, this algorithm allows actual inclusion of inapplicable data as phy-

logenetic information, a crucial step for combining discrete morphological matrices in order to include fossil data in the Tree of Life.