

Fregin, S., Haase, M., Olsson, U. & Alström, P. 2011 (in press). **Pitfalls in comparisons of genetic distances: A case study of the avian family Acrocephalidae.** *Molecular Phylogenetics and Evolution*

Abstract

Genetic distances are increasingly being used for identification and species delimitation, especially since the introduction of ‘‘barcoding’’. While for phylogenetic inferences great care is generally taken to choose the best-fit evolutionary model, this is usually neglected in calculating genetic distances. Moreover, distances obtained from others than best-fit models, different lengths of sequences, and even different loci are often freely compared. We examined the influence of different methods on calculating genetic distances using mitochondrial cytochrome b sequences for the passerine family Acrocephalidae.

We found substantial differences between: (1) corrected distances based on the best-fit model (TrN + C) vs. uncorrected p-distances; (2) distances calculated based on different parts of the same gene; and (3) distances calculated using the methods of ‘‘complete deletion’’ vs. ‘‘pairwise deletion’’ for sequences that included uncertain nucleotides. All these methodological differences affected comparisons between species and potential taxonomical conclusions.

We suggest that (1) different loci are incomparable. (2) Only perfectly homologous regions (same length, same part of locus) should be compared. (3) In the case of sequences with some uncertain nucleotides, only distances calculated by the method of ‘‘complete deletion’’ are fully comparable. (4) Only distances based on the optimal substitution model should be used. (5) Even within the same locus, corrected genetic distances are unique to the study in which they are calculated, as they are conditional on the particular dataset and model selected for that dataset.