

Bias in phylogenetic reconstruction and the structure of real and simulated sequence data

Christoph Mayer, Ruhr Universität Bochum

Model dependent phylogenetic methods have become increasingly popular in recent years since they are more accurate and robust than non-parametric methods under most circumstances. With few exceptions (covarion and mixture models), however, these methods assume that the evolution of sequences along a tree is governed by a single, time-reversible model of nucleotide substitutions, thereby taking into account only a small fraction of the true complexity of evolution. Violations of these highly restrictive assumptions are known to be able to introduce a bias in the data set, with the possible consequence that a wrong phylogenetic tree is (re)constructed and high support values are obtained for wrong branches (Lockhart 1992, Steel 1993, Lopez 2002, Kolaczkowski 2004, Philippe 2005, Wägele 2008). In this talk, I review known phylogenetic biases and present recent results that show how a simulated evolution with different substitution models on different lineages of a tree can affect the ability of various phylogenetic methods to reconstruct the correct tree. Furthermore, and in view of these biases, the information content of real data sets is compared to that of simulated data sets obtained by a parametric bootstrap approach.

Lockhart PJ, Howe CJ, Bryant DA, Beanland TJ, Larkum AW: **Substitutional bias confounds inference of cyanelle origins from sequence data.** *J Mol Evol* 1992, **34**:153-162.

Steel MA, Lockhart PJ, Penny D: **Confidence in evolutionary trees from biological sequence data.** *Nature* 1993, **364**:440-442.

Lopez P, Casane D, Philippe H: **Heterotachy, an important process of protein evolution.** *Mol Biol Evol* 2002, **19**:1-7.

Kolaczkowski B, Thornton JW: **Performance of maximum parsimony and likelihood phylogenetics when evolution is heterogeneous.** *Nature* 2004, **431**:980-984.

Philippe H, Zhou Y, Brinkmann H, Rodrigue N, Delsuc F: **Heterotachy and long-branch attraction in phylogenetics.** *BMC Evol Biol* 2005, **5**:50.

Wägele JW, Mayer C: **Visualizing differences in phylogenetic information content of alignments and distinction of three classes of long-branch effects.** *BMC Evol Biol* 2007, **7**:147.