82. JSplits: a frame-work for phylogenetic analysis using trees and networks

Daniel H. Huson [1]

Keywords: phylogeny, trees, networks, splits, quartets

1 Introduction

The goal of phylogenetic analysis is to determine the order and approximate timing of speciation events in the evolution of a given set of species. This is usually done by comparing sequence data for individual genes, and more recently, on the level of whole genomes, and then building a tree (or trees) inferred from the comparisons. However, although the stated goal is often to obtain a tree, it is sometimes desirable to apply phylogenetic reconstruction methods that do not force the given data onto any particular tree but rather are allowed to produce a more "networked" graph when this is the best representation for the data at hand. Bandelt and Dress [BD92] suggested a number of phylogenetic analysis methods, such as split-decomposition method and parsimony splits, that produce phylogenetic networks. More recently, Bryant and Moulton [BM02] described a new method Neighbor-net, that brings together both split decomposition and the well-known neighbor-joining method [SN87].

2 Phylogenetic trees and phylogenetic networks

There exist a number of packages for performing phylogenetic analysis, e.g. [Swo00, SvH96, MM02]. However, they are all use trees as the fundamental data structure. In contrast, the SplitsTree program [Hus98] is based on so-called splits and phylogenetic networks and is dedicated to methods such as split decomposition.

Our new java package `jSplits` is aimed at providing a frame-work for both tree- and network-oriented phylogenetic analysis, including median networks [BFSR95]. Fundamental data types supported by the program included unaligned- and aligned sequences, distances, splits, trees, networks and quartets. The package provides a number of the most commonly used distance-based algorithms. A main feature of the program is a "plug-and-play" design for the incorperation of new methods.

Additionally, new rank-based methods as recently proposed by Andreas Dress and Alex Grossman (personal communication) will be implemented in the package. Finally, the package will provide an implementation of the disk-covering method for trees and networks [HNW99].

`JSplits` will be released early 2003 as Open Source.

[1] Center for Bioinformatics Tübingen (ZBIT), Tübingen University, Sand 14, 72076 Tübingen, Germany. E-mail: `huson@informatik.uni-tuebingen.de`

References

[BD92] H.-J. Bandelt and A. W. M. Dress. A canonical decomposition theory for metrics on a finite set. *Advances in Mathematics*, 92:47–105, 1992.

[BFSR95] H.-J. Bandelt, P. Forster, B.C. Sykes, and M.B. Richards. Mitochondrial portraits of human population using median networks. *Genetics*, 141:743–753, 1995.

[BM02] D. Bryant and V. Moulton. Neighbornet: An agglomerative method for the construction of planar phylogenetic networks. In R. Guigó and D. Gusfield, editors, *Algorithms in Bioinformatics, WABI 2002*, volume LNCS 2452, pages 375,391, 2002.

[HNW99] D. Huson, S. Nettles, and T. Warnow. Obtaining highly accurate topology estimates of evolutionary trees from very short sequences, 1999. Proceedings"RECOMB'99".

[Hus98] D. H. Huson. SplitsTree: A program for analyzing and visualizing evolutionary data. *Bioinformatics*, 14(10):68–73, 1998.

[MM02] W. Maddison and D. Maddison. Mesquite- a modular system for evolutionary analysis. `http://mesquite.biosci.arizona.edu/mesquite/mesquite.html`, 2002.

[SN87] N. Saitou and M. Nei. The Neighbor-Joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, 4:406–425, 1987.

[SvH96] K. Strimmer and A. von Haeseler. Quartet puzzling: a quartet maximum likelihood method for reconstructing tree topologies. *Molecular Biology and Evolution*, 13:964–969, 1996.

[Swo00] D. L. Swofford. PAUP*: Phylogenetic analysis using parsimony (*: and other methods), version 4.2, 2000.