

SplitsTree 4.0 - Computation of phylogenetic trees and networks

Daniel H. Huson ¹, Tobias Kloeppe ², David Bryant ³

Keywords: phylogeny, evolution, trees, networks, graphs.

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. In the classic theory of phylogenetic analysis the species are assumed to evolve along a (bifurcating) X-tree, experiencing point mutations along the way. Recent results from genome wide comparison of gene trees seem to indicate that these models may fit well for single genes but fail to represent the complex evolution of a genome. A natural generalisation of the phylogenetic tree, that seems to fit well with the complex evolution of a genome, is the phylogenetic network. There are two main approaches to generate such phylogenetic networks. The first one is the combined analysis approach, and the second one is the individual analysis approach. In the first approach the phylogenetic network is generated directly from the given combined information. In the second approach individual trees are generated for the given information sets and all individual trees are then combined into one network. Bandelt and Dress [BD92] suggested a number of combined analysis methods, such as the split-decomposition method and the parsimony splits. 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]. Holland and Moulton [HM03] presented a method to join individual gene trees for a common set of species. In 2004 Huson et al. [HDKS04] presented the Z-Super-Network method, which merges individual *partial* gene trees.

2 The SplitsTree program

There exist a number of packages for performing phylogenetic analysis, e.g. [Swo00, SvH96, MM02]. However, they all use trees as the fundamental data structure. In contrast, the **SplitsTree** program [HB05] is based on so-called splits and phylogenetic networks. It is aimed at providing a general framework for both tree- and network-oriented phylogenetic analysis. Fundamental data types supported by the program include unaligned- and aligned sequences, distances, splits, trees, networks and quartets. The package provides commonly used distance-based algorithms. We have also implemented the methods mentioned in the introduction. The package provides numerous visualisation methods for phylogenetic trees and networks, examples are split graphs [DH04] and the equal-daylight method [Fel04]. We [HKLS05] have recently implemented methods for the reconstruction of reticulation networks and the transformation from split networks to reticulation networks.

The main features of the program are:

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

²Center for Bioinformatics Tübingen (ZBIT), Institute for Computer Science, Tübingen University, Sand 14, 72076 Tübingen, Germany. E-mail: kloeppe@informatik.uni-tuebingen.de

³McGill Centre for Bioinformatics, Lyman Duff Building, 3775 University, Montreal, Quebec, H3A 2B4, Canada. E-mail: bryant@math.mcgill.ca

- it runs on any machine with minimal installation requirements based on Java or Java Web Start.
- GUI version for interactive use, command-line version for scripting pipelines.
- Flexible frame-work for doing phylogenetic analysis.
- De-centralized plug-in concept for adding new methodology.
- Fundamental data types include splits and quartets.
- Uses Nexus file format, with one-to-one correspondence between internal data classes and external nexus blocks, and supports most other formats.
- Transformations of molecular sequences to distances.
- Combined and individual analysis methods.
- Visualisation of phylogenetic trees and networks.
- Interactive exploration of the visualisation.
- Bootstrapping

3 Availability.

The program is freely available from www-ab.informatik.uni-tuebingen.de/software

References

- [BD92] H.-J. Bandelt, and A.W.M. Dress, Split Decomposition: A new and useful approach to phylogenetic analysis of distance data. *Molecular Phylogenetics and Evolution* 1(3):242-252, 1992.
- [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.
- [DH04] A.W.M. Dress and D.H. Huson. Constructing splits graphs, *IEEE/ACM Transactions in Computational Biology and Bioinformatics*, volume 1(3), pages 109-115, 2004.
- [Fel04] J. Felsenstein. Inferring Phylogenies *Sinauer Associates, Inc.*, pages 582ff, 2004.
- [HM03] B.Holland and V.Moulton. Consensus networks: A method for visualizing incompatibilities in collections of trees. *Algorithms in Bioinformatics, WABI 2003*, volume LNBI 2812, pages 165-176, 2003
- [HB05] D. H. Huson and D.Bryant. Estimating phylogenetic trees and networks using SplitsTree 4, note=Manuscript in preparation, software available from www-ab.informatik.uni-tuebingen.de/software
- [HDKS04] D.H. Huson, T. DeZulian, T. Kloeppe and M.A. Steel. Phylogenetic Super-Networks from Partial Trees. *Algorithms in Bioinformatics, WABI 2004*, in press, 2004.
- [HKLS05] D.H. Huson, T. Kloeppe, P.J. Lockhart and M.A. Steel. Reconstruction of Reticulate Networks from Gene Trees accepted for *RECOMB 2005*.
- [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.