Introduction to Phylogenetic Networks

Daniel H. Huson

Center for Bioinformatics (ZBIT), Tübingen University, Sand 14, 72076 Tübingen, Germany
Email: husoninformatik.uni-tuebingen.de
Web: www-ab.informatik.uni-tuebingen.de

1 Biographical data

Daniel H. Huson studied mathematics at Bielefeld University and received the PhD degree there in 1990. From 1990-9, he held a variety of different research positions at Bielefeld University and was supported during this time by a two year DFG research scholarship, and received his "Habilitation" in mathematics there in 1997. He then spent two years, 1997-99, as a postdoctorate working with Tandy Warnow at Princeton University. He then joined Celera Genomics as a Senior Research Scientist working in Gene Myers group. Since 2002, he has been a professor of algorithms in bioinformatics at Tübingen University in Germany.

2 Motiviation and aims of the tutorial

The evolution of a species is usually represented by phylogenetic trees and many established methods exist for computing such trees from bio-molecular sequence data. The concept of a phylogenetic tree is clearly defined and familiar to most researchers in biology or bioinformatics.

However, when reticulate events such as hybridization, horizontal gene transfer or recombination play an important role in the evolution of some species, or when other events introduce incompatibilities, then phylogenetic trees may no longer suffice and phylogenetic networks may be required to provide an adequate representation of evolution.

Unfortunately, the concept of a phylogenetic network is clouded somewhat in confusion. There are three sources of confusion: Firstly, there are many different types of networks and methods for computing networks. Secondly, authors have repeatedly attempted to capture the term "phylogenetic network" by defining it to mean the special type of network that they are currently interested in. Thirdly, there are actually two different classes of phylogenetic networks, namely those that aim at simply providing a visualization of incompatible signals in a dataset, and those that provide an explicit scenario of reticulate evolution. We propose to name the former "implicit" and the latter "explicit" approaches.

In this tutorial we aim to give an overview over the different types of phylogenetic networks and methods and show how they are related to one another. We will look at the algorithms behind some of the methods. We will discuss which methods to apply to which type of data, and will discuss the strengths and weakness of different methods.

3 Targeted audience

This tutorial aims at researchers that have a reasonable understanding of phylogenetic analysis using trees and are interested in looking beyond tree-based methods at phylogenetic networks.

4 Tutorial outline

This is the structure of the tutorial:

- Introduction: a general overview of the different types of phylogenetic networks and there classification
- Phylogenetic trees: a quick review of important concepts
- Implicit networks: introduction to splits and split networks, implicit networks from sequences, distances and trees
- Explicit networks:
  - Hybridization and reticulate networks
  - Recombination networks
- Other types of networks

I have taught this tutorial twice before: at ISBM 2005 in Detroit, USA, and at GCB 2006 in Tübingen, Germany. I can provide a handout of approximately 45 pages, see `www-ab.informatik.uni-tuebingen.de` for further details.