

# Tutorial Proposal: Introduction to Phylogenetic Networks

Daniel H. Huson  
Center for Bioinformatics (ZBIT),  
Tübingen University,  
Sand 14, 72076 Tübingen, Germany,  
email: huson@informatik.uni-tuebingen.de

November 30, 2004

## 1 Brief description of instructor

I am Professor of Algorithms in Bioinformatics at Tübingen University. One of my active areas of research is the design and implementation of new algorithms for phylogenetic analysis, in particular for whole genome analysis and phylogenetic networks. I regularly teach graduate level courses on Algorithms in Bioinformatics, Sequence Analysis and Phylogeny. I have given approximately 80 conference presentations and invited talks, including at ISMB, RECOMB and WABI.

## 2 Title and expected goals, objectives and motivation

The tutorial is entitled *Introduction to phylogenetic networks*. The immense growth of bio-molecular data is providing new opportunities and new challenges for phylogenetic analysis. As more and more data becomes available, the complexity of evolution is becoming more and more apparent and phylogenetic trees do not always suffice, or are not always adequate, to represent the evolutionary relationships detected between different organisms.

Moreover, there is a lot of active research in bioinformatics aimed at providing non-tree-like models of evolution, and developing methods for reconstructing phylogenetic networks, rather than trees, from complex data sets. Such methods can be subdivided into two categories:

- methods that generate networks that represent some model of non-tree-like evolution, such as hybridization or recombination, or, generally speaking, reticulate evolution, and

- methods that generate networks that visualize incompatibility or conflicting signals in a data set, and are not directly interpretable as a model of evolution, for example, a “consensus network” that represents all splits that occur in a certain percentage of a collection of trees.

The main aim of this tutorial is to give an overview over the different types of phylogenetic networks that have been defined and are in use. These include ancestor recombination graphs, hybridization networks, galled trees, gt-networks, reticulate networks, splits graphs, consensus networks, neighbor-nets, bootstrap networks and median graphs. Rather than discuss each of these different concepts separately, we will show how they are related to each other and which ones are specializations of which.

### 3 Tutorial level

I would rate this as an *intermediate* tutorial.

### 4 Intended audience

The intended audience are people who have a reasonable understanding of phylogenetic analysis centered on trees and would like to become acquainted with phylogenetic networks. In particular, I want to target biologists who feel that their data sets require an analysis using networks rather than trees.

Participants of the tutorial should go home with:

- a good basic understanding of simple models of reticulate evolution,
- a overview over the many different variants of reticulate networks, and
- the ability to attack their own data set using different pieces of software that are available on the Web.

### 5 Detailed outline

The tutorial will have three parts. In the first part, I will discuss different models of evolution. First, I will remind the audience of how we usually model tree-like evolution. Then we will see how to modify this model to incorporate “reticulate” events such as hybridizations or recombinations.

Then second, with this model in mind, we will discuss different approaches that try to reconstruct the underlying “reticulate network” that generated a given set of data, such as haplotype data or gene sequences, etc. Although we will try to cover all conceptually important approaches, we will put special emphasize on methods for which reasonable implementations are available.

In this context, but somewhat separately, we will discuss some of the methods of the second type. Although these methods do not explicitly attempt to

reconstruct reticulate models of evolution, we will see that in practice some of these methods can give useful insights.

Thirdly, we will look at a number of biological data sets that clearly benefit from an analysis based on phylogenetic networks. For example, we may look at haplotype data, sequence data from plants in the presence of hybridization, and a whole genome data for a set of prokaryotes.

## Remarks

Please see [www-ab.informatik.uni-tuebingen.de](http://www-ab.informatik.uni-tuebingen.de) for details on my teaching, research and CV.