



Research Group

University of Tübingen  
Wilhelm Schickard Institute for Computer Science and  
ZBIT Center for Bioinformatics Tübingen

Algorithms

in

Bioinformatics

Prof. Dr. Daniel Huson

# jSplits - a Java Framework for Phylogenetic Trees and Networks

Tübingen, 4.7.2003

Copyright (c) 2008 Daniel Huson.

Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.2 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license can be found at <http://www.gnu.org/copyleft/fdl.html>

# Trees vs networks

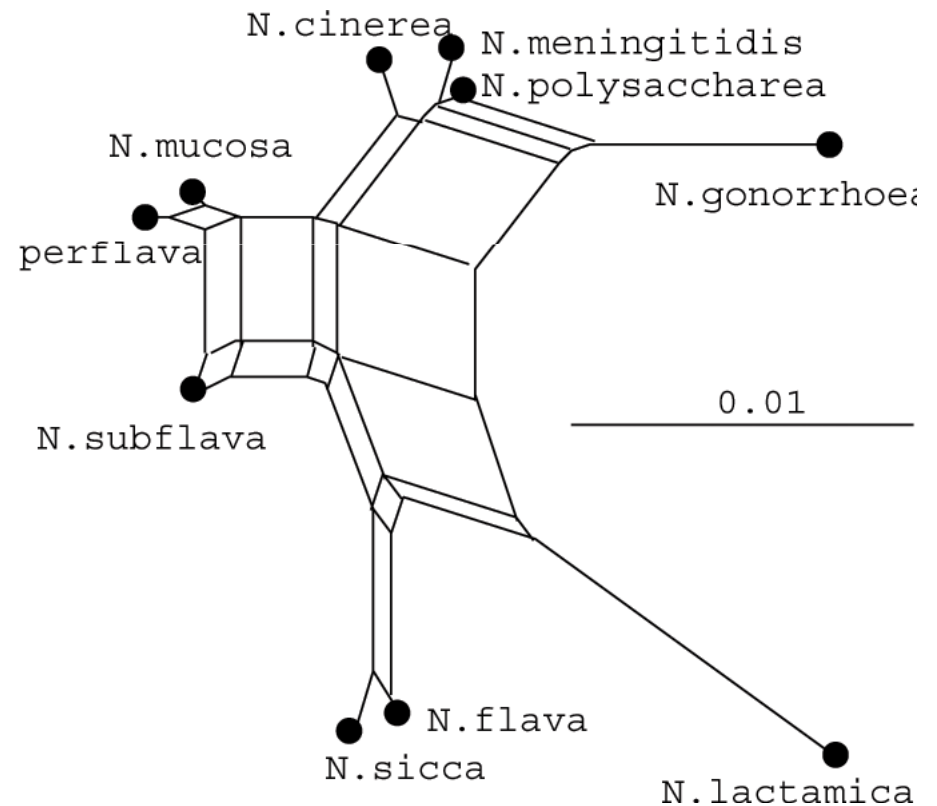
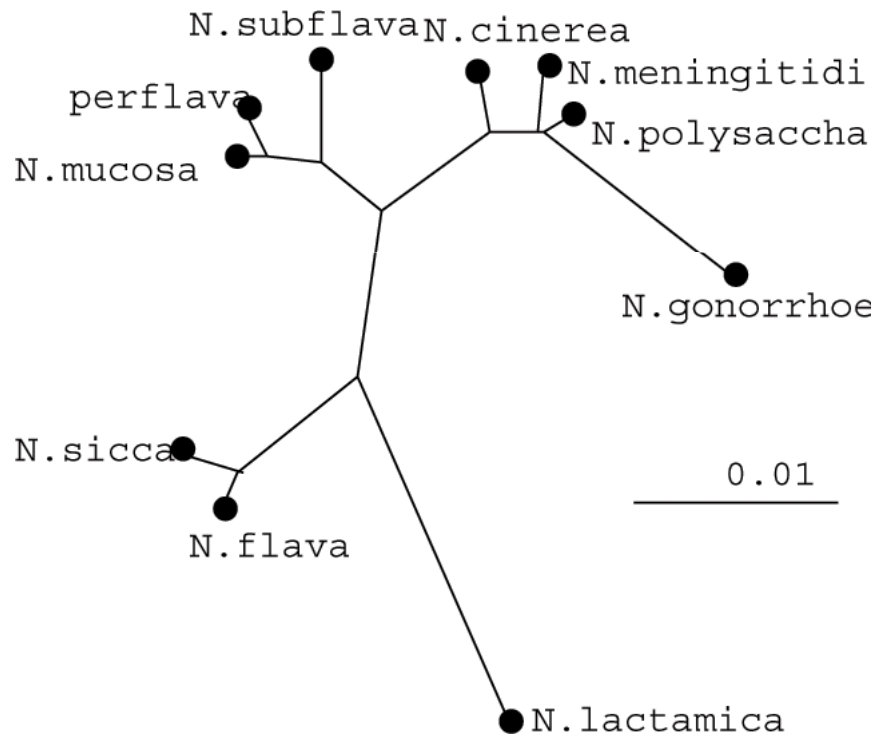
- Evolutionary relationships are usually represented by phylogenetic trees
- But, real data contain different and/or conflicting signals, and thus do not always clearly support a unique tree
- Hence, in some cases a phylogenetic network may be more appropriate...



# Trees vs networks

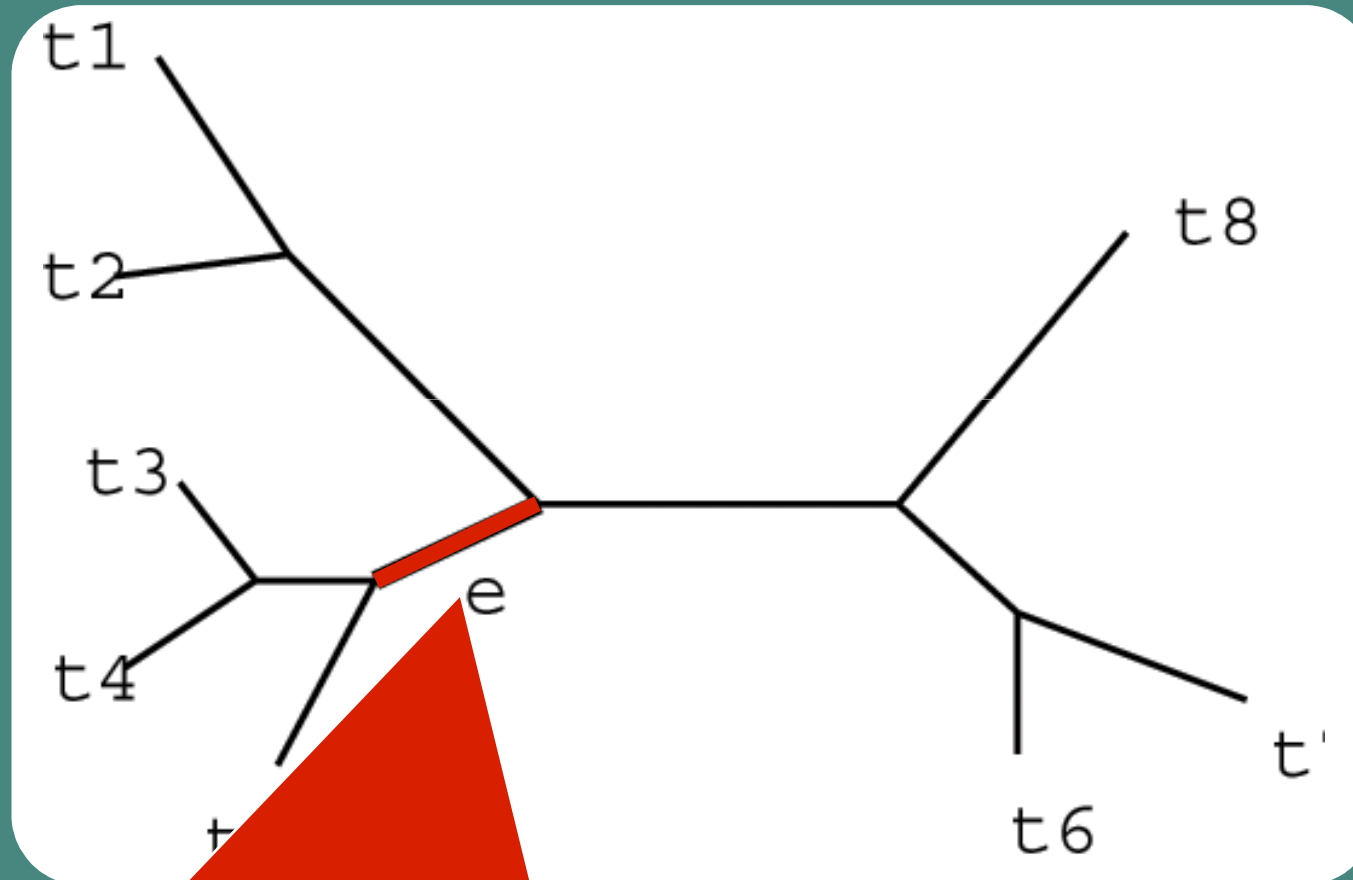
Computed using Neighbor-Joining

Computed using split decomposition



Neisseria phylogeny (Eddie Holmes, 1999)

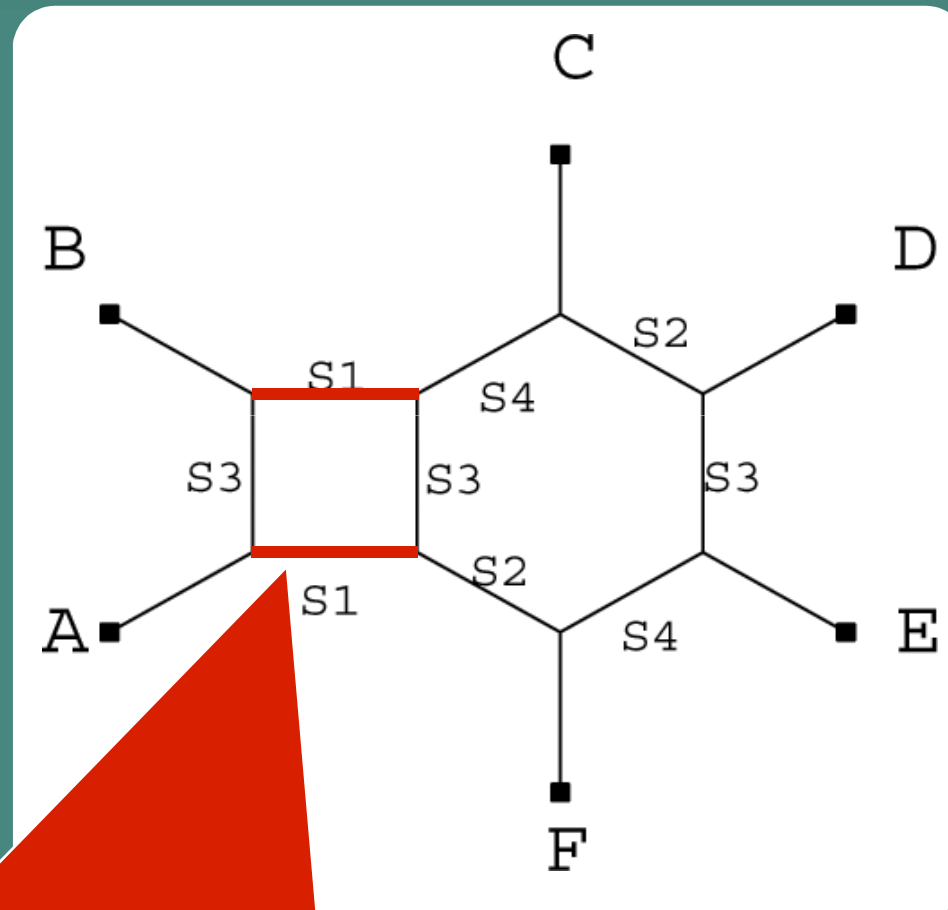
# Trees and splits



Edge  $e$  corresponds to split  $\{t_1, t_2, t_6, t_7, t_8\}$  vs  $\{t_3, t_4, t_5\}$

Daniel Huson, 2003

# Networks and splits



Cut-set of parallel edges defines split  $\{A, B\}$  vs rest

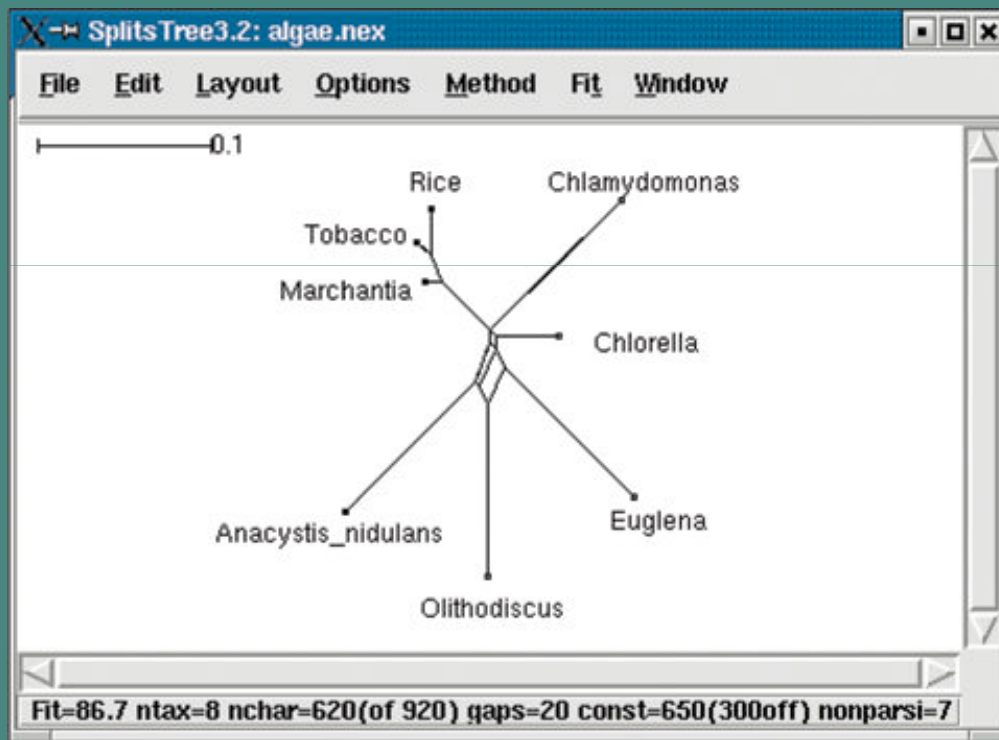
# Splits and graphs

Any given system  $\Sigma$  of splits can be represented by a *splits graph*  $G$ . Note that:

- $G$  is a tree iff  $\Sigma$  is compatible  
(e.g. Neighbor-Joining)
- $G$  is outer-planar iff  $\Sigma$  is circular  
(e.g. Neighbor-Net, Bryant & Moulton 2002)
- $G$  is usually planar or only mildly non-planar iff  $\Sigma$  is weakly compatible (e.g. Split Decomposition)
- $G$  is always subgraph of  $n$ -dim. hypercube  
(e.g. recoding of sequences, spectral analysis, median networks, consensus networks)

(Theory of splits worked out by Bandelt and Dress 1992)

# SplitsTree 3.2



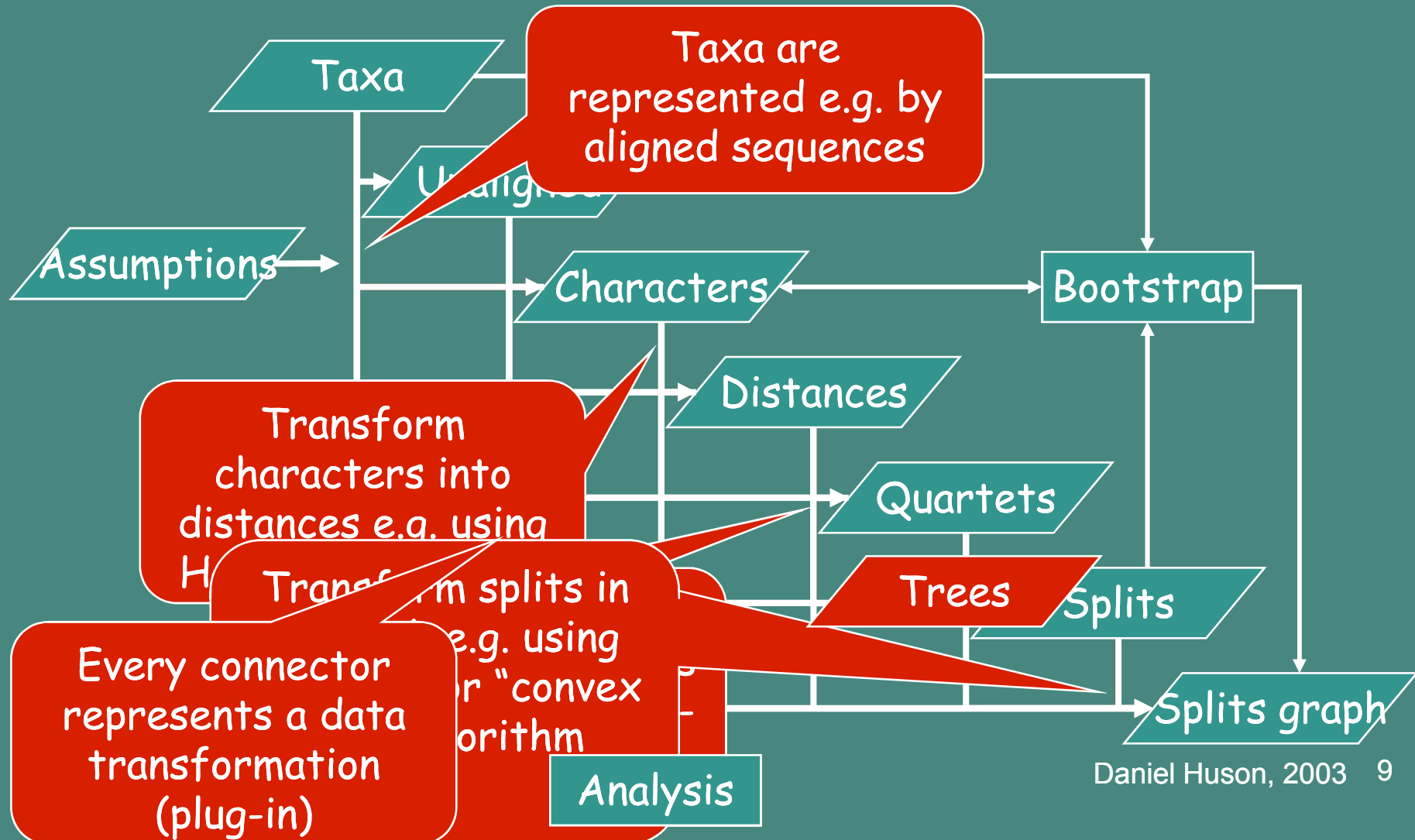
- Implements split decomposition and related methods
- First version developed with Rainer Wetzel in 1995
- Current version 3.2 in C++ using Tcl-Tk
- Runs under Linux, Unix, Windows and MacOS



# Design criteria for jSplits (SplitsTree4)

- Must run on any machine with minimal installation requirements
- GUI for interactive use, command-line for pipelines
- Open system, decentralized plug-in concept
- Based on splits, also including quartets etc
- Use Nexus file format
- Open source

# Data flow in jSplits

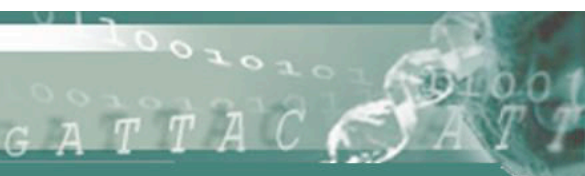


# Writing a new transformation

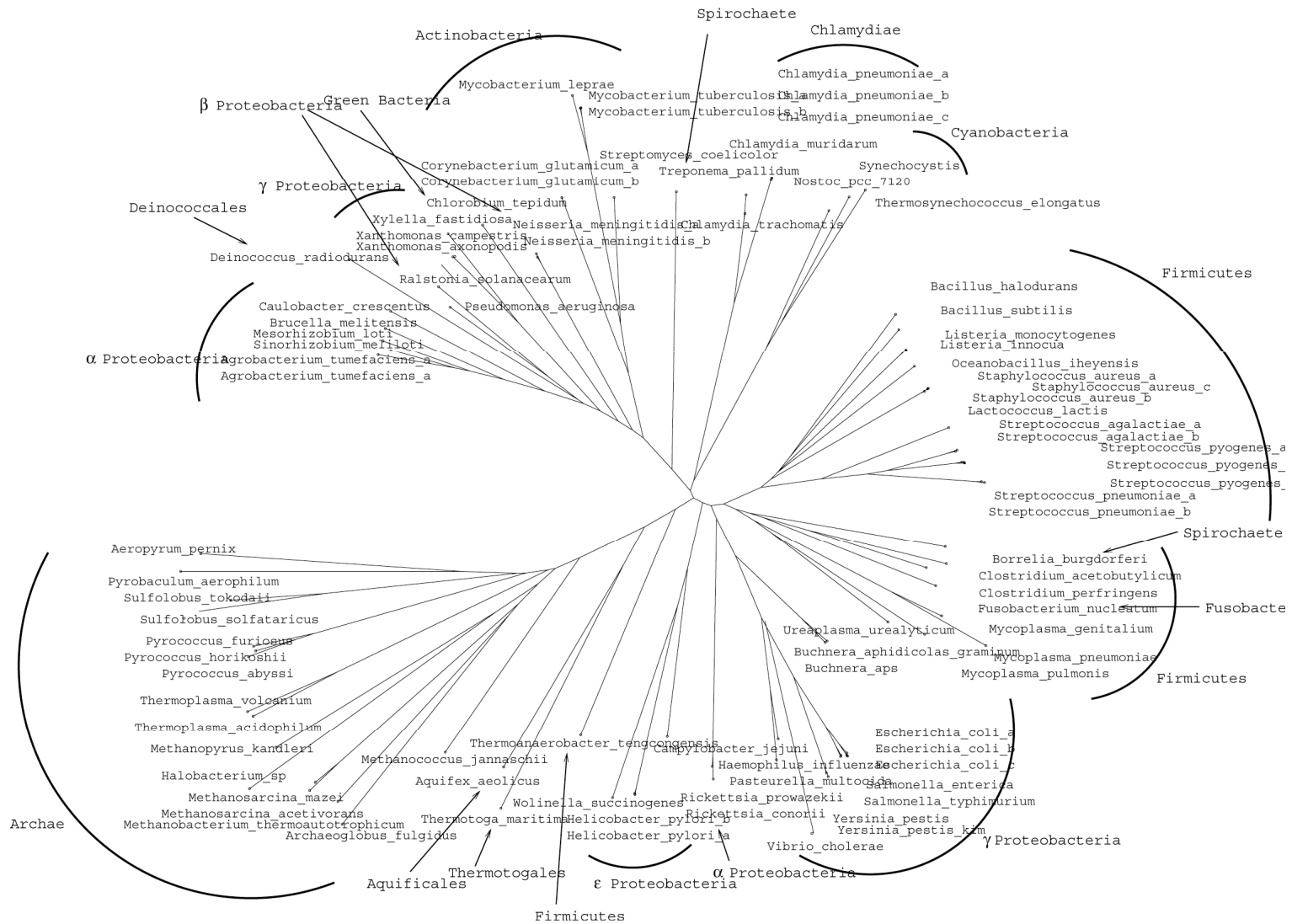
E.g., a new tree-building method "GreatTree" is provided to jSplits as follows:

```
public class GreatTree implements Distances2Tree
{
    // returns true, if GreatTree is applicable
    public boolean isApplicable (Taxa t, Distances d)
    {...}

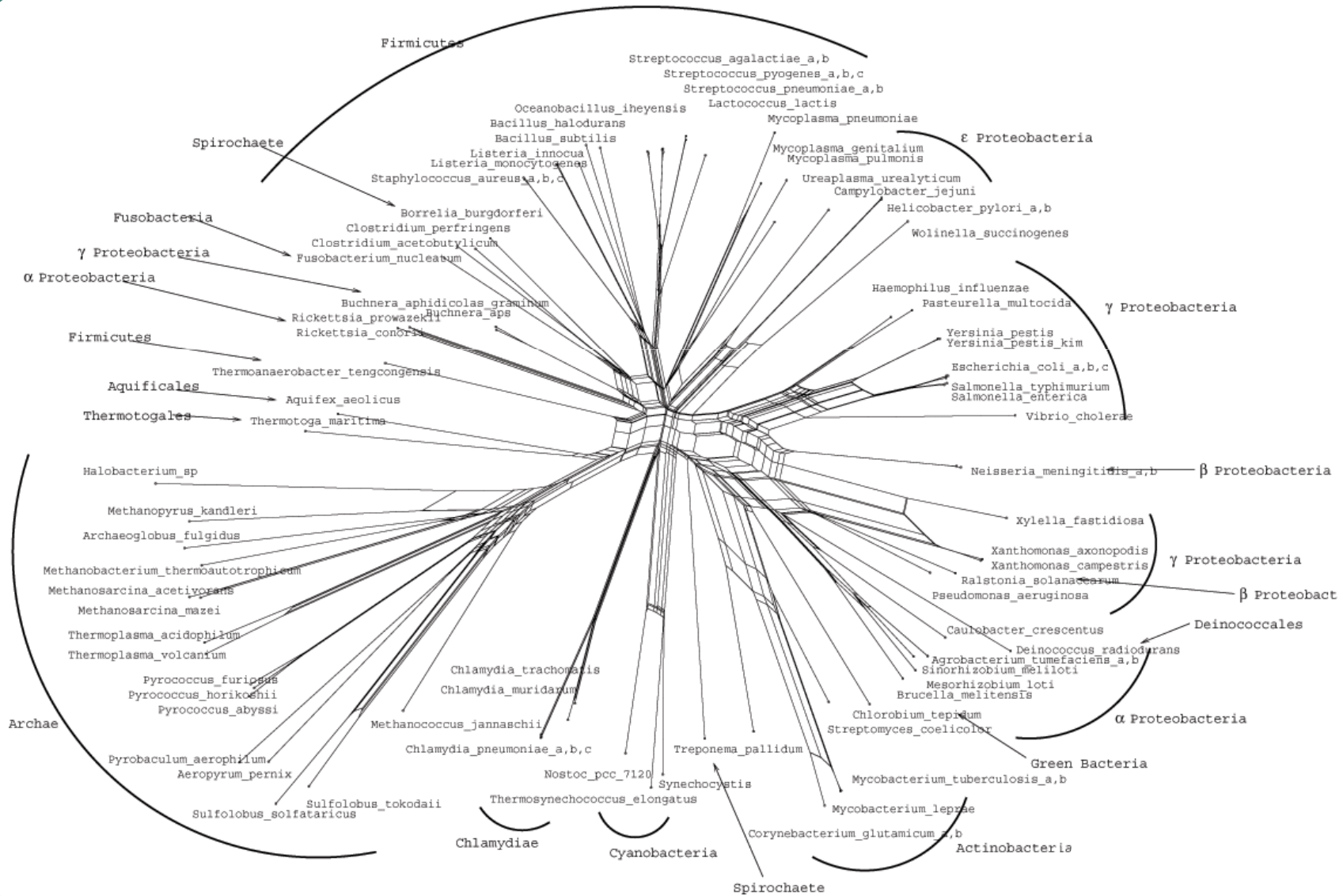
    // applies GreatTree and returns the tree
    public Tree apply (Taxa t, Distances d)
    {...}
}
```



# BioNJ on bacterial genomes



# N-Net on bacterial genomes



# Summary

- jSplits is an open frame-work for phylogenetic analysis
- Extensibility based on plugin design
- Built on splits, incorperates both tree and network methods
- Provides all popular distance-based tree building algorithms
- Provides network methods such as split decomposition, Neighbor-net and median networks

# Credits

- Authors: David Bryant and D.H.
- Additional programmers: Markus Franz and Michael Schröder
- Thanks to: Yuhui Wang