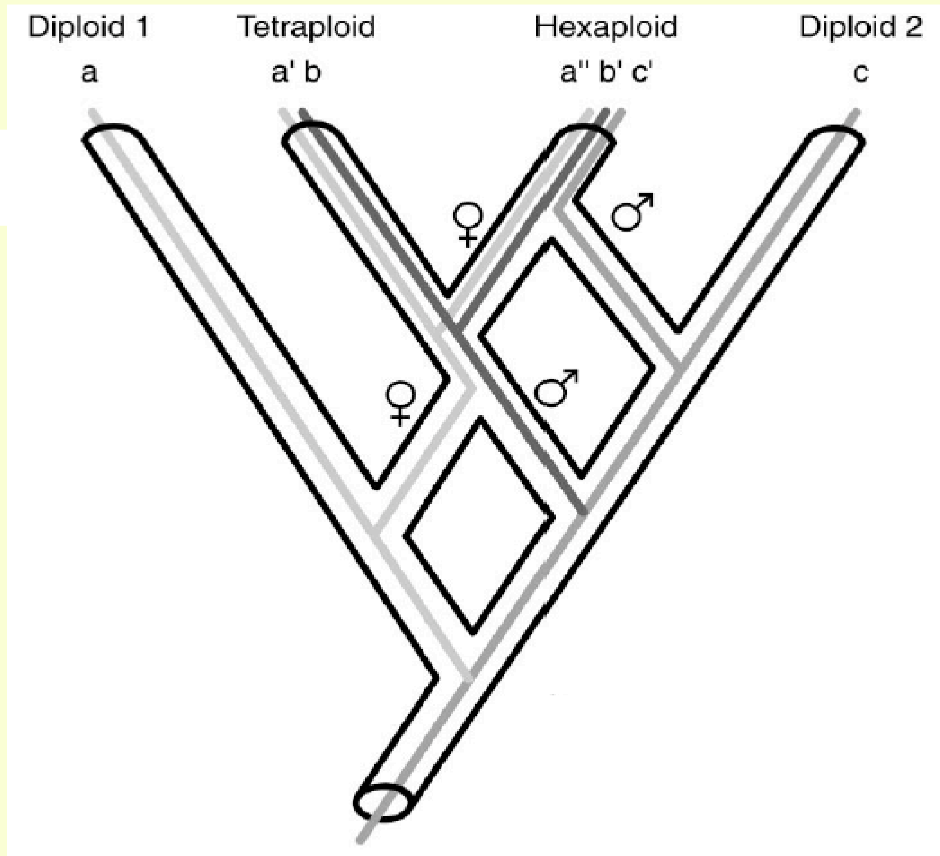
The complexity of deriving multi-trees from sets of bipartitions



Vincent Moulton, School of Computing Sciences

University of East Anglia

Joint work with

• **Dr. Katharina Huber, Martin Lott, Dr. Andreas Spillner**

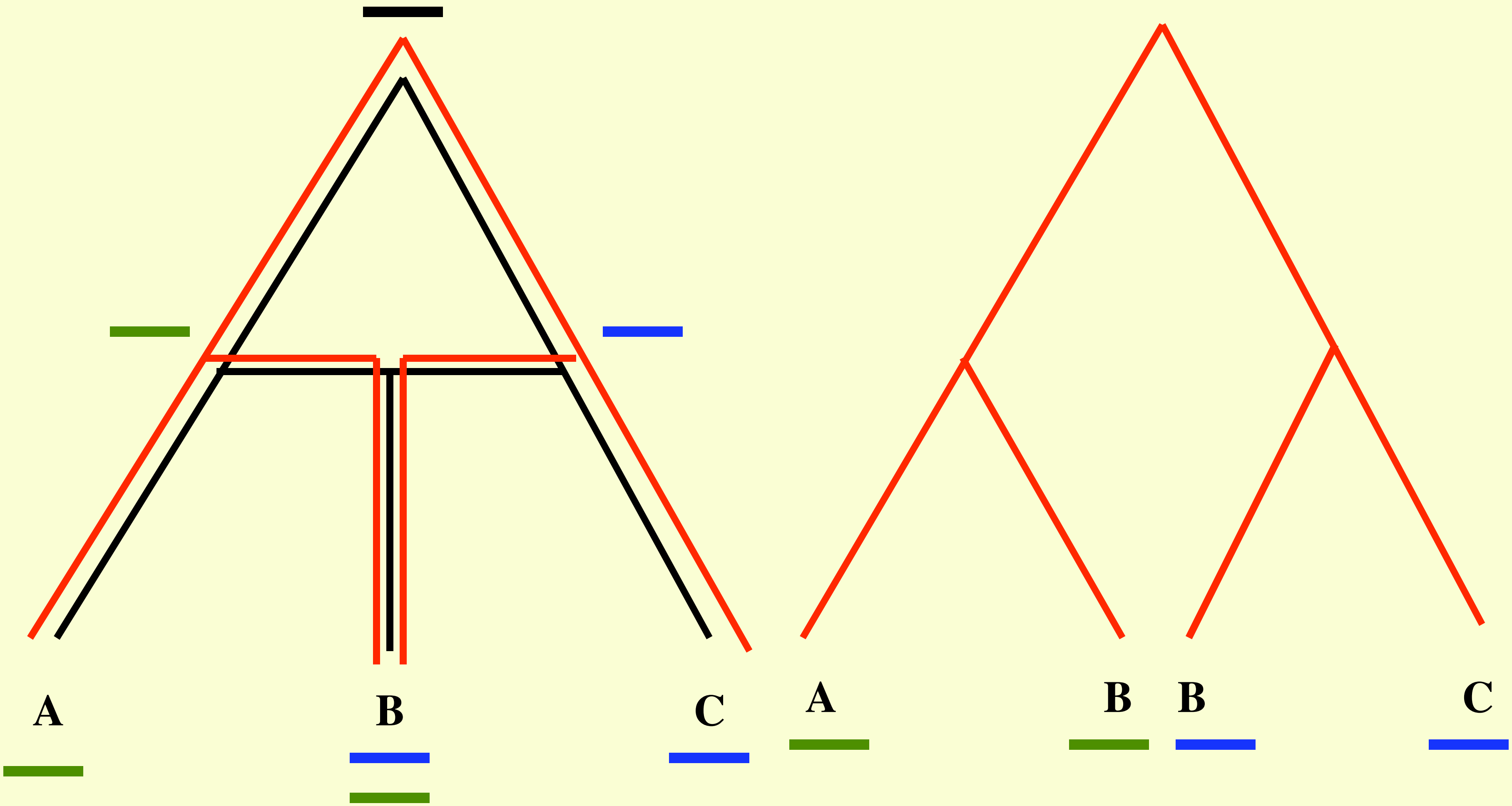
School of Computing Sciences,University of East Anglia

• **Prof. Bengt Oxelman, Anna Petri**

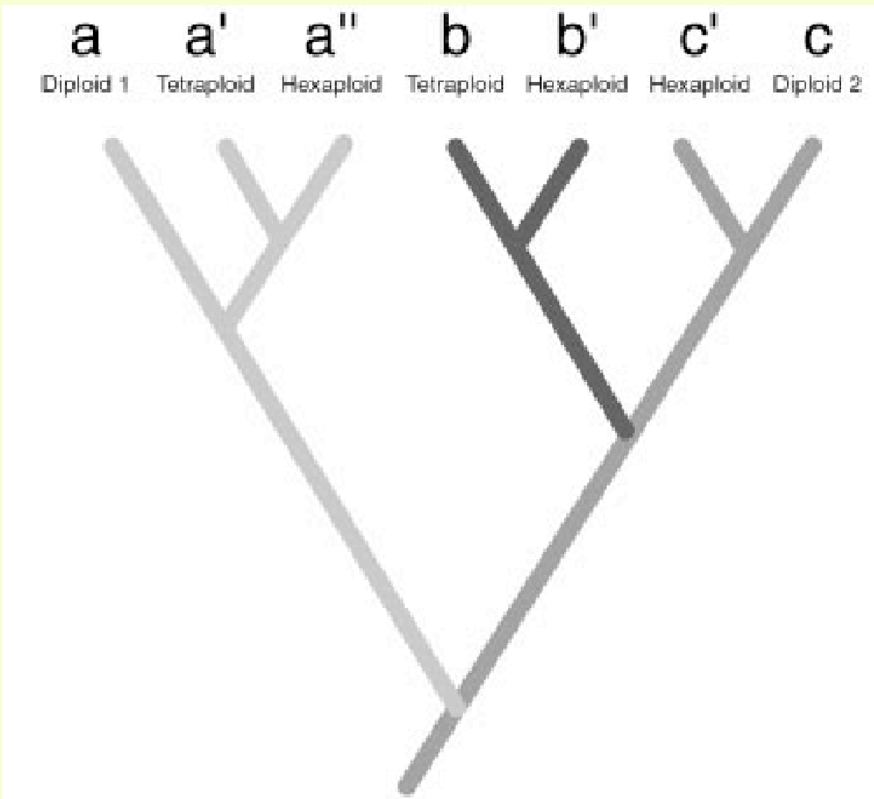
Department of Plant and Environmental Sciences, University of Gothenburg

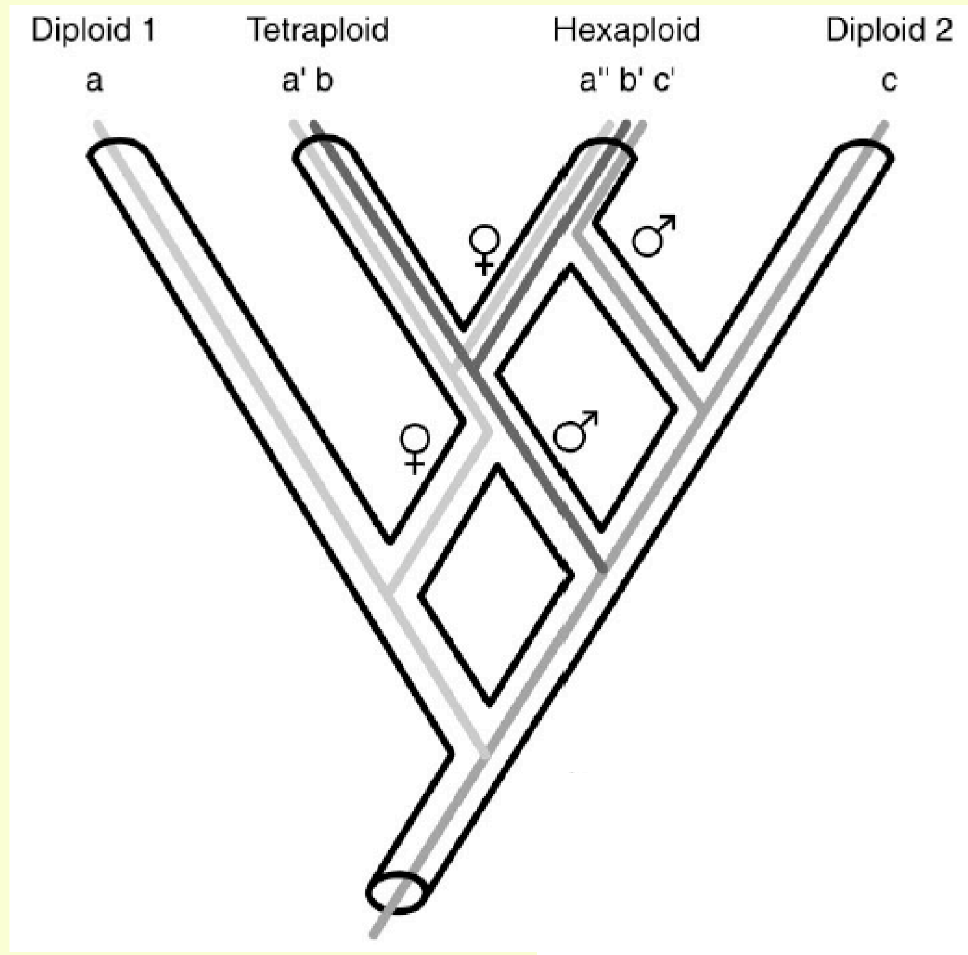
EPSRC

Modelling polyploidy
(multi-trees from networks)



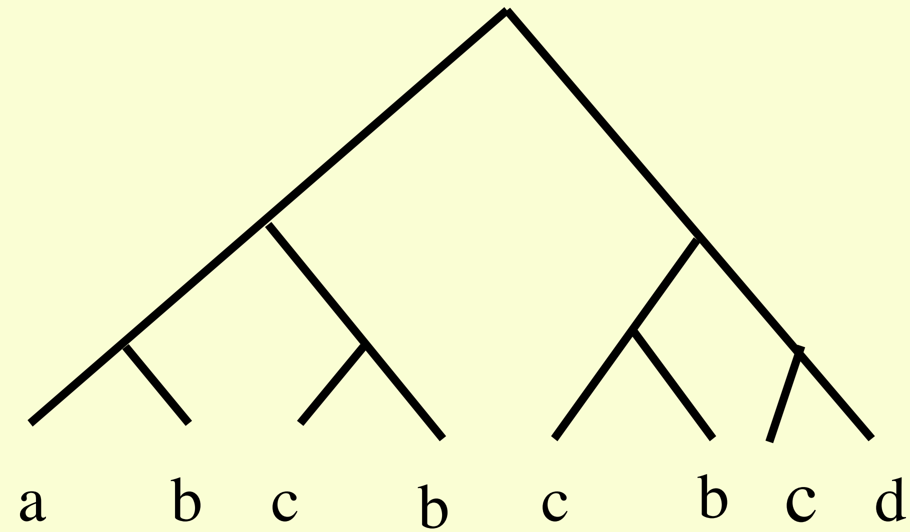
Networks from multi-trees?

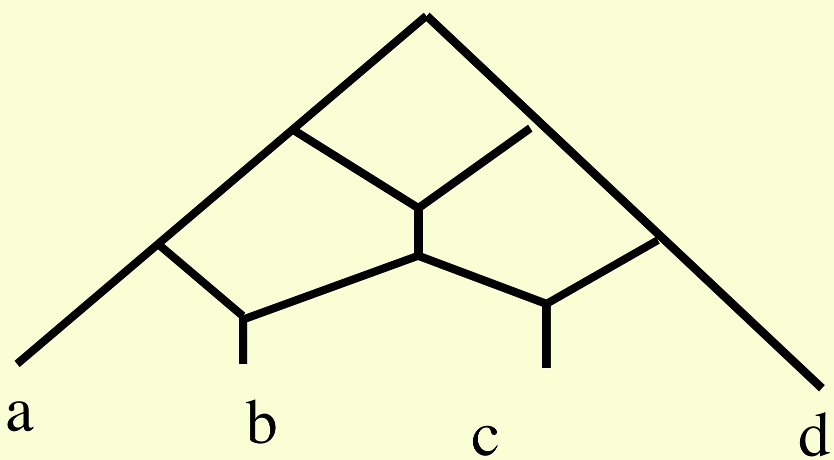


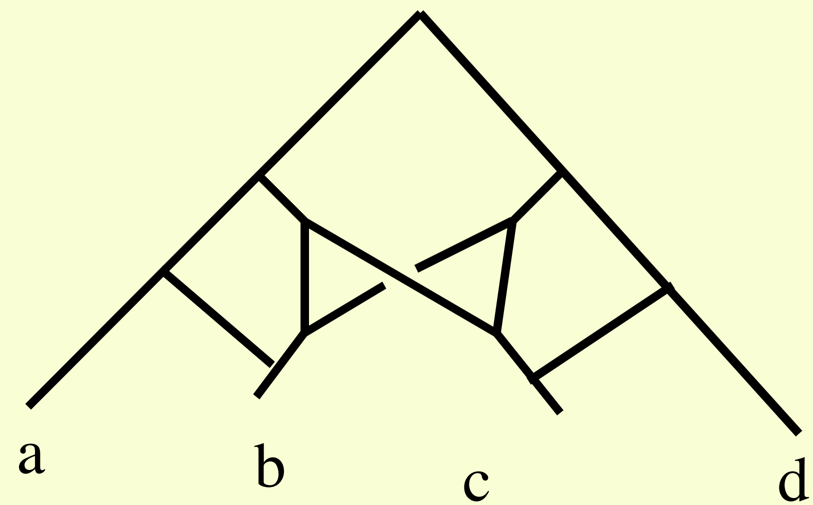


Smedmark et al., Systematic Biology, 2003

Multiple possibilties

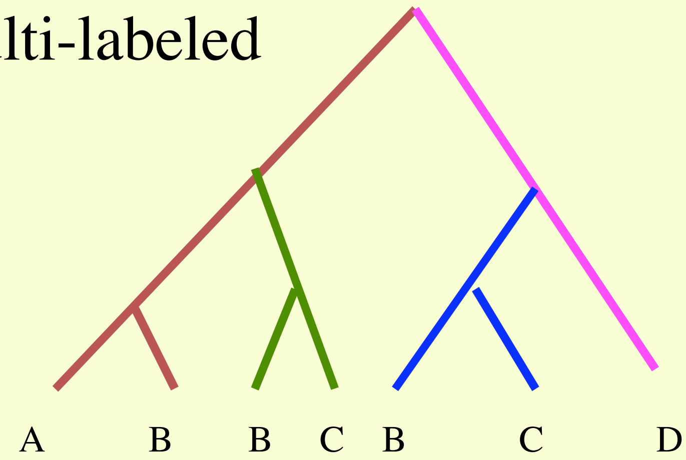


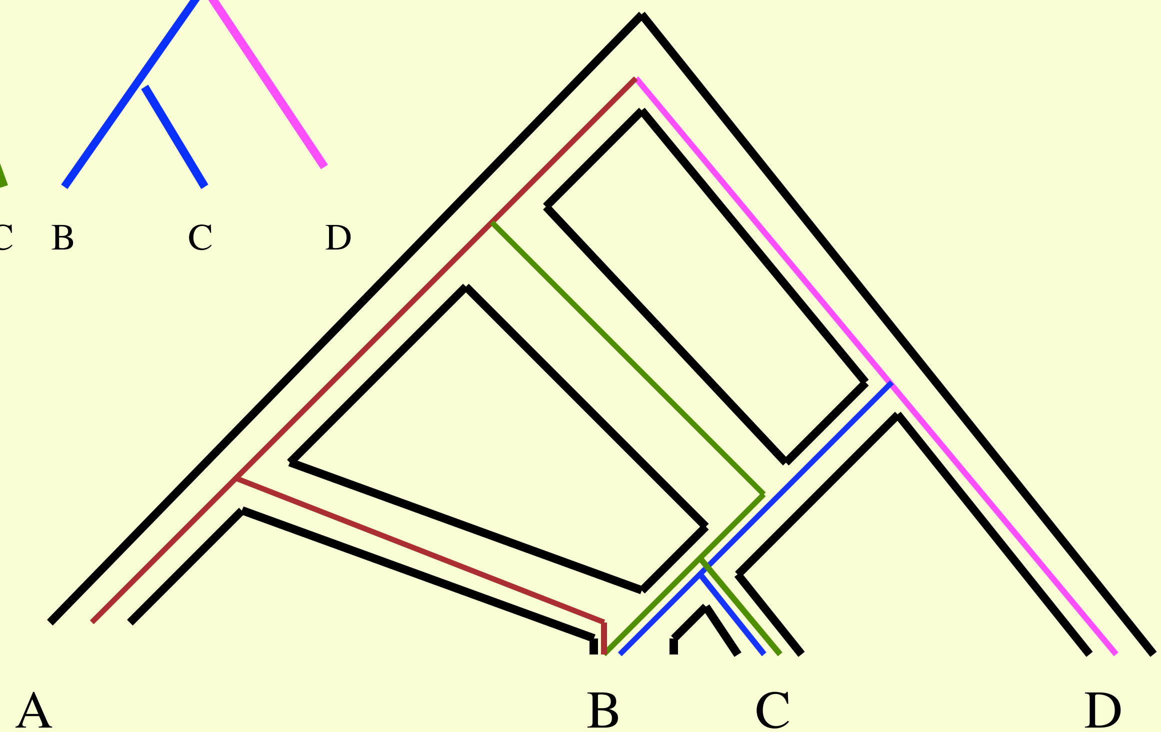




Aim

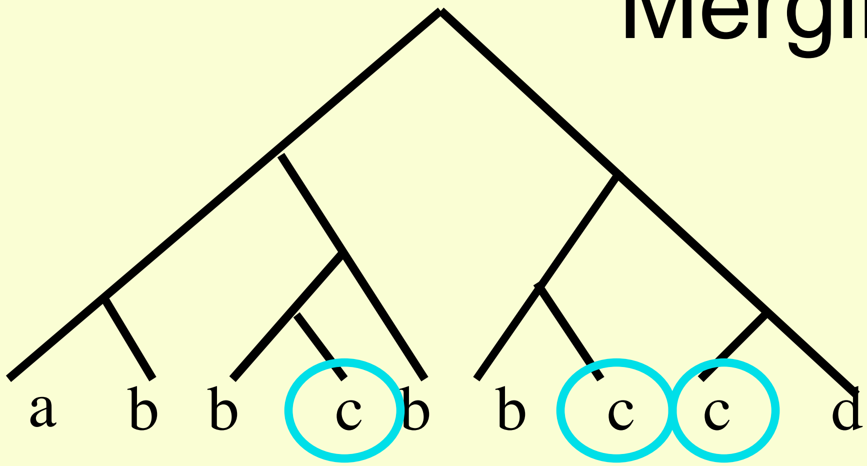
Given a multi-labeled tree *T* ….

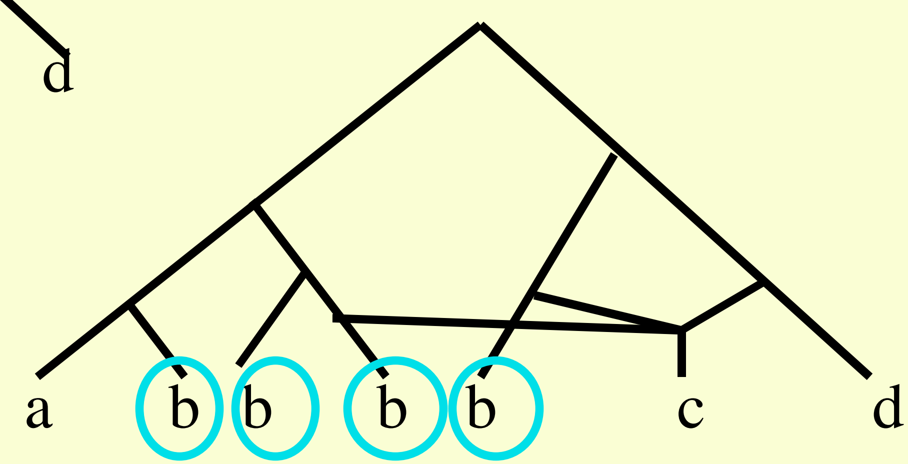


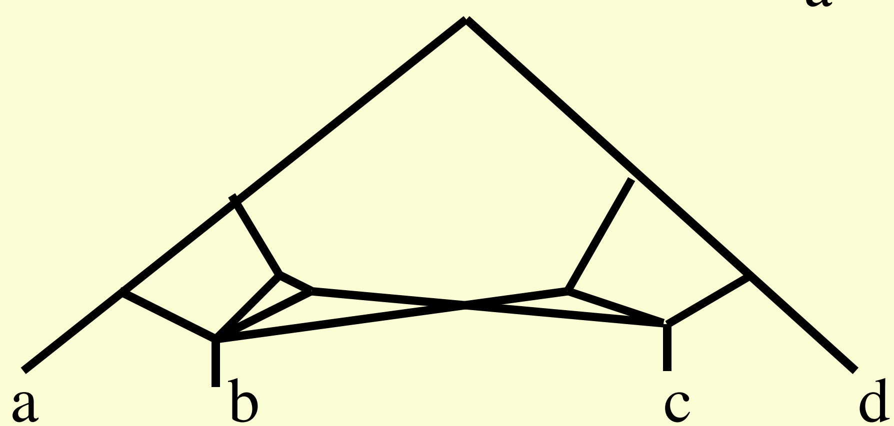


…. construct a “most parsimonious” reticulate network displaying *T*

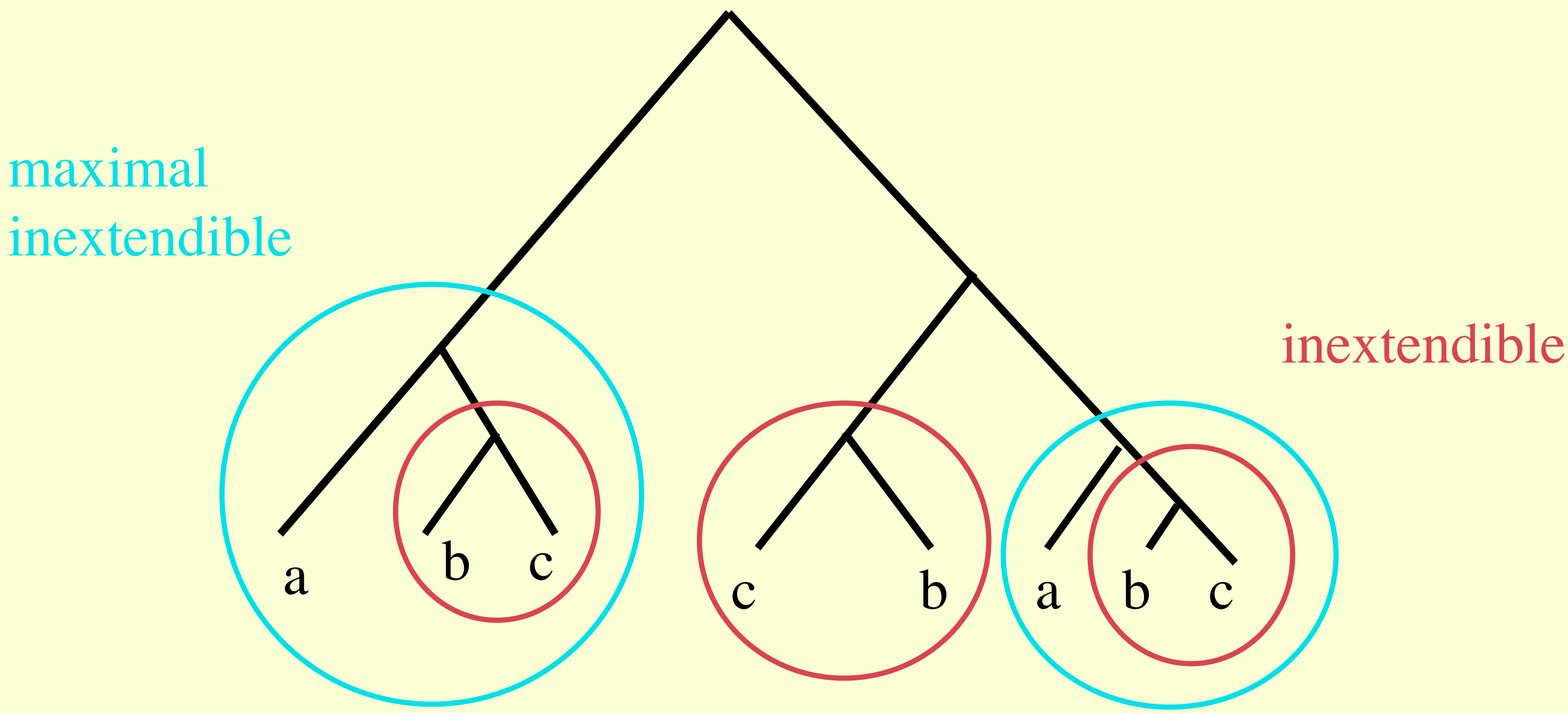
Merging leaves



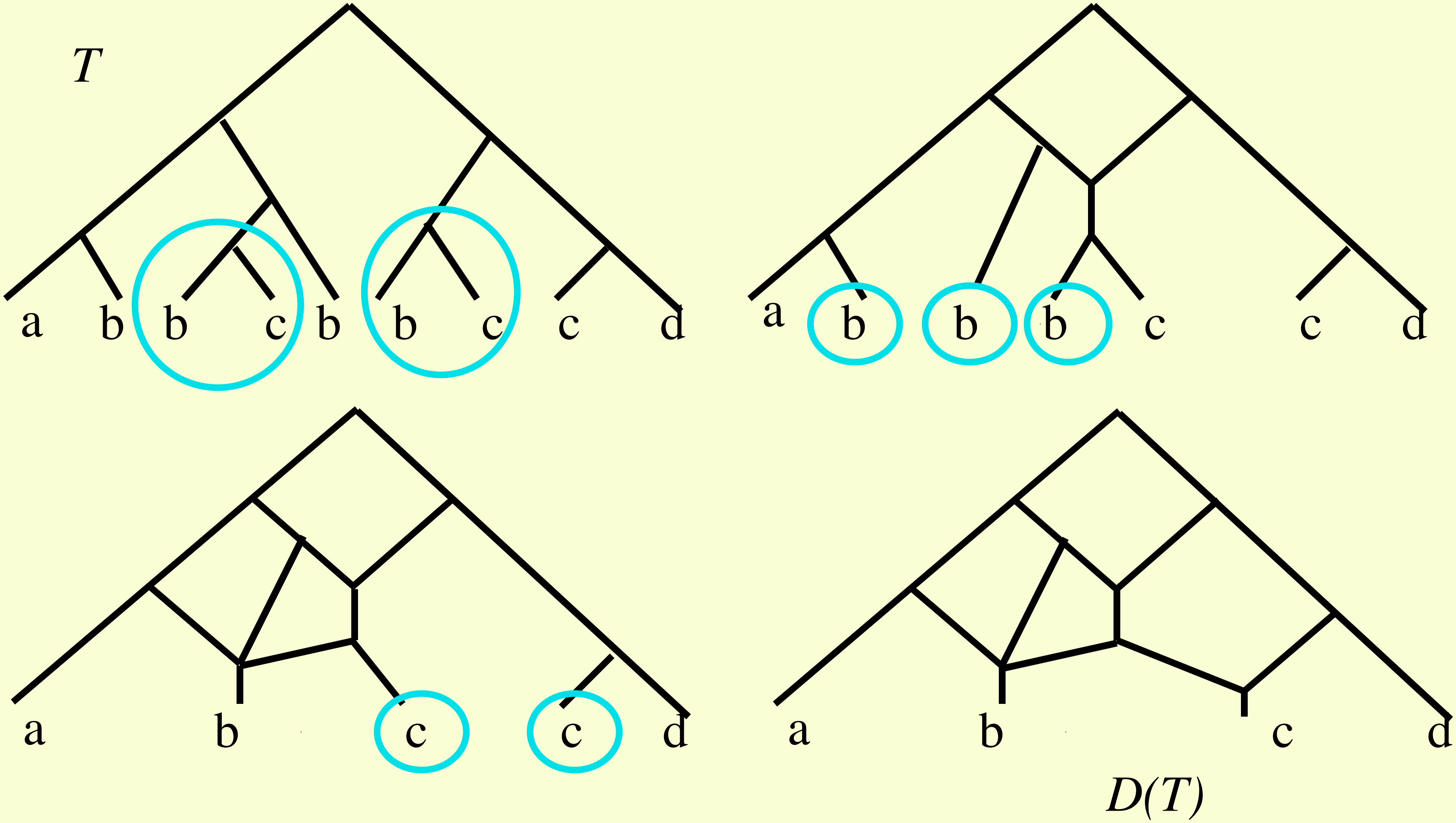




Inextendible subtrees

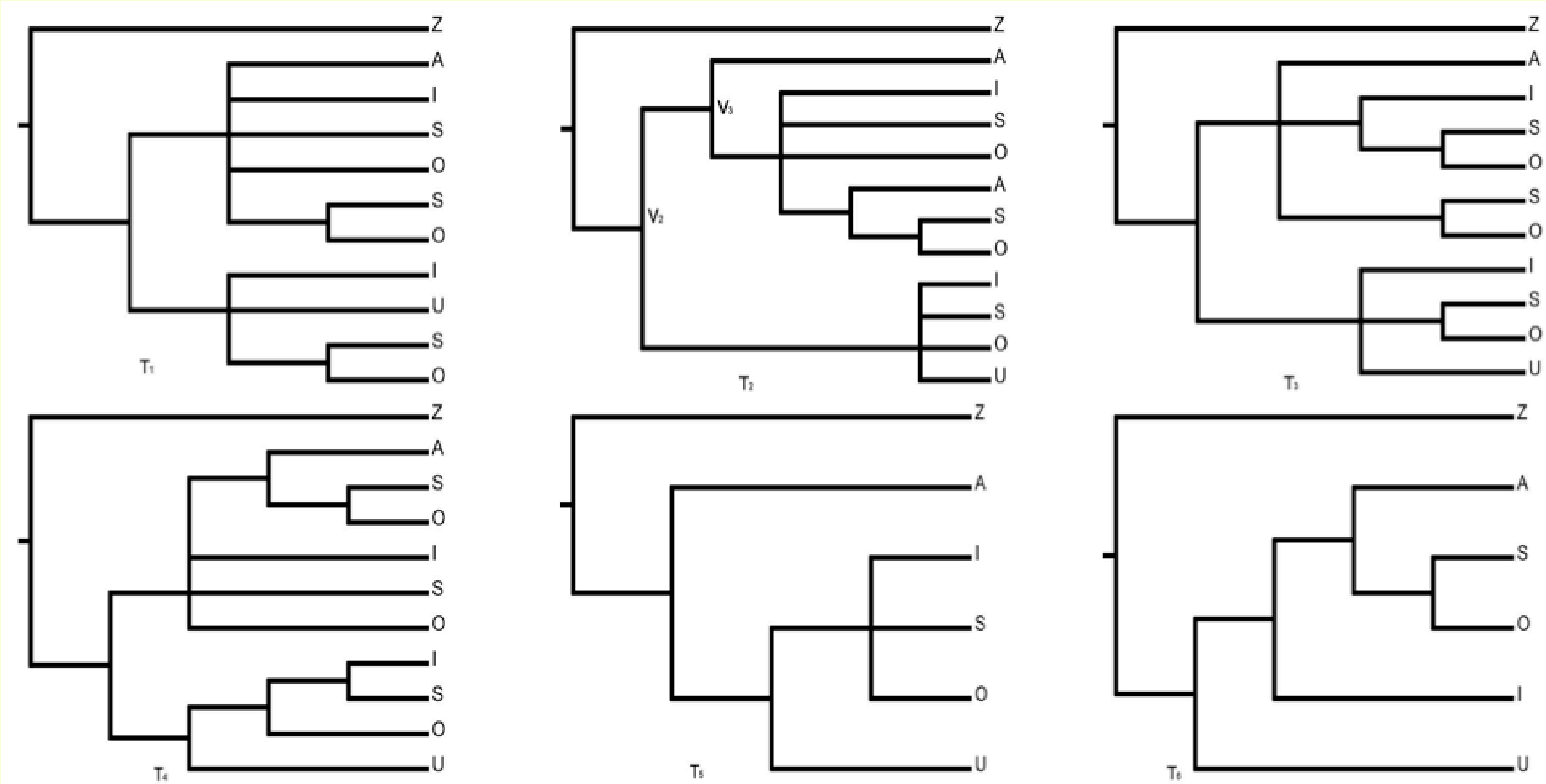


Construction

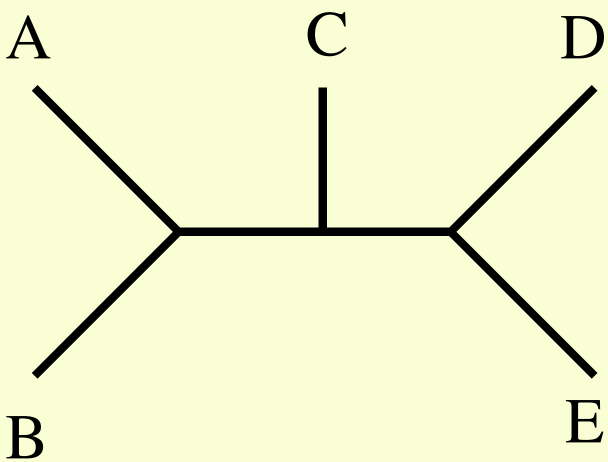


Theorem [Huber,Moulton, 2006] *D(T)* is "minimal" network displaying *T*.

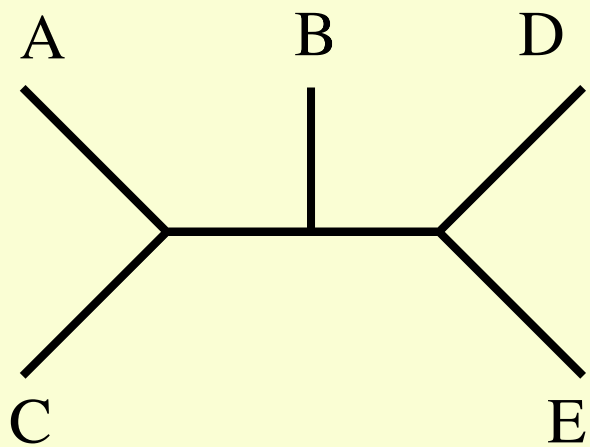
Question: How do we get the multi-tree?



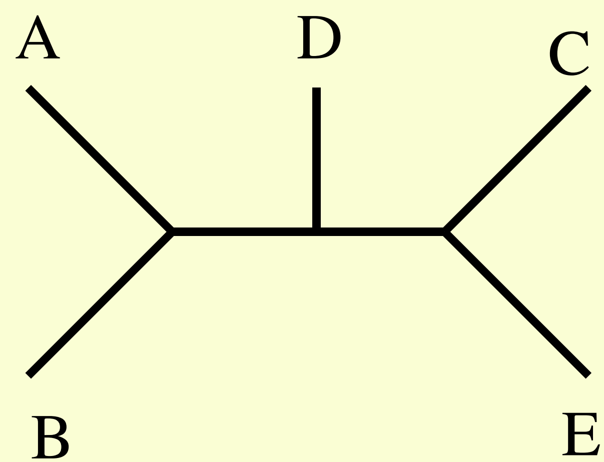
Consensus trees



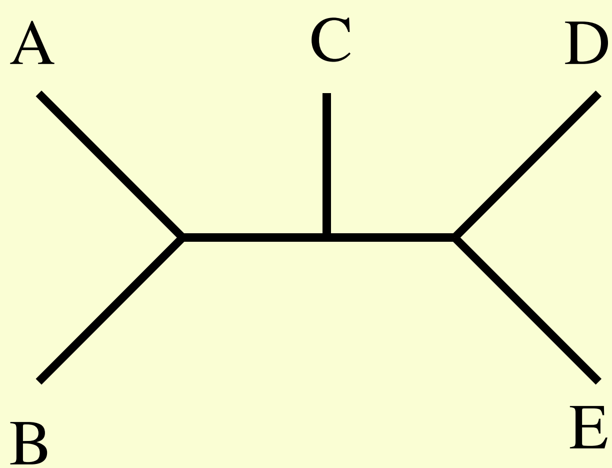
AB | CDE , ABC | DE



AC |BDE, ABC | DE



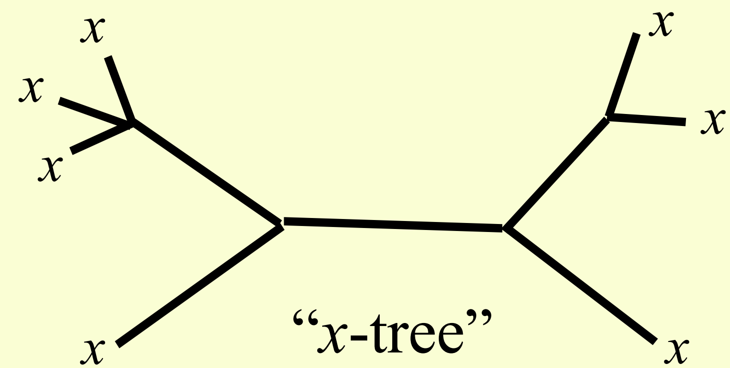
AB|DCE, ABD | CE



Problem!

Theorem

Given a set S of splits of a multi-set M, it is NP-hard to decide if S can be displayed by a multi-tree (even if the multiplicity of all elements in M is bounded by 3).



Idea for why this is the case:

$M = \{n\, x\}$

$S = \{ n_1\, x \mid (n-n_1)x\, , \ldots , n_m\, x \mid (n-n_m)\, x\}$

Deciding if we can display this set by a multi-tree is essentially equivalent to deciding if there is a subset of $\{n_1,.., n_m\}$ that adds up to $n/2$.

Useful result and conjecture

Given multiset $M = \{m(x)\ x\}_{x \text{ in } X}$, let

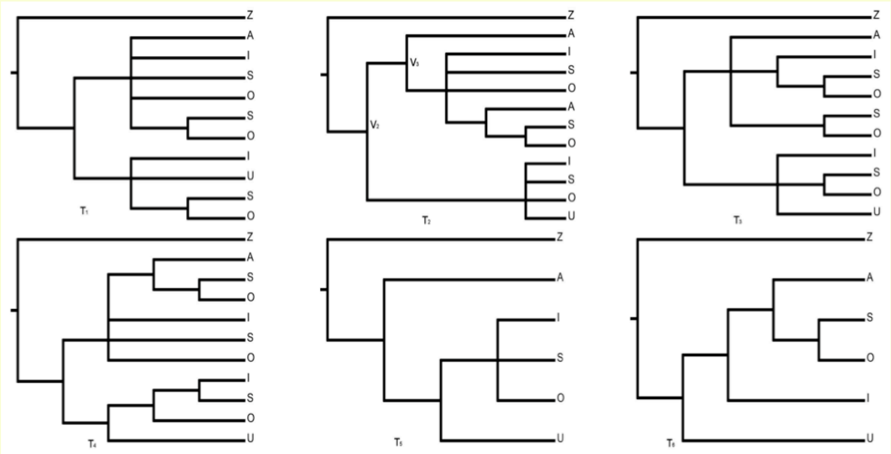
$$\Delta(M) = \Sigma_{\ x \text{ in } X}\,(m(x) - 1).$$

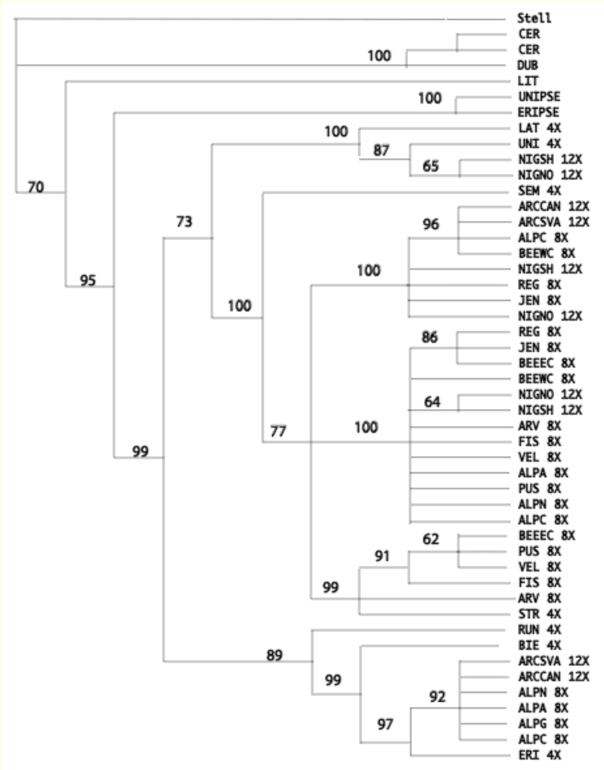
Theorem [Lott, Huber, Moulton, Spillner, in press]

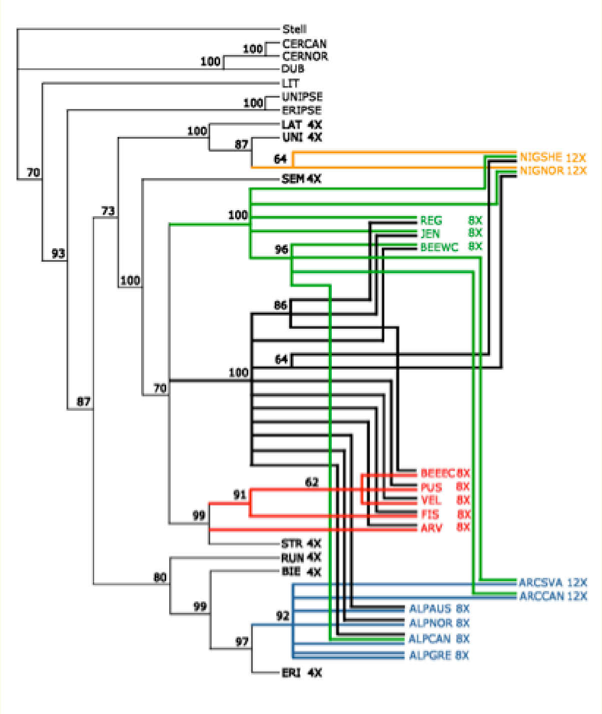
If every submultiset of size at most $m := max\{2\Delta, \Delta+2\}$ of a multiset of splits of M can be displayed by a multi-tree, then so can the whole collection.

Conjecture $m = \Delta+2$

Work in progress…







PADRE

Martin Lott

University of East Anglia
Computational Biology Laboratory

On this Site
- Welcome
- Members
- Research Topics
- Postgraduate Degrees
- Collaborations
- Downloads
- Seminars
- How to Find Us

Local Partners
- JIC Computational Biology Group

Related Websites
- UEA School of Computing Sciences
- John Innes Centre
- Institute of Food Research
- Norfolk & Norwich University Hospital
- Norwich Research Park

Phylogenetics Software - PADRE

PADRE: Package for Analyzing and Displaying Reticulate Evolution

Description | Availability | Manual | Test files | Support | References | Disclaimer

Description

PADRE is a Java software package allowing the computation and graphical representation of reticulate networks from Multi-Labelled (MuL) Trees. The network is constructed in stages which allows the user flexibility in choosing appropriate elements of the input tree for merging.

Input to the program takes the form of text files containing multi-labelled trees in Newick format as decribed by http://evolution.genetics.washington.edu/phylip/newicktree.html; see our test files for examples. The program is available as a complied JAR file which is platform independant and suitable for running on Windows, Linux and MacOS. Images may be exported to EPS, PNG and JPEG format suitable for inclusion in articles and scientific papers. All graphical representations displayed are initially scaled to fit the screen and nodes (including optionally the sub-trees for which they are the root) may be moved and resized in any direction.

If you use PADRE we would appreciate if you could cite the following paper:

- K. T. Huber, B. Oxelman, M. Lott and V. Moulton, *Reconstructing the evolutionary history of polyploids from multi-labelled trees*, Submitted

http://www.cmp.uea.ac.uk/~vlm/padre/