

# Assembling the network of life

Vincent Moulton

School of Computing Sciences  
University of East Anglia



Banff, January 2017

# Tree of life

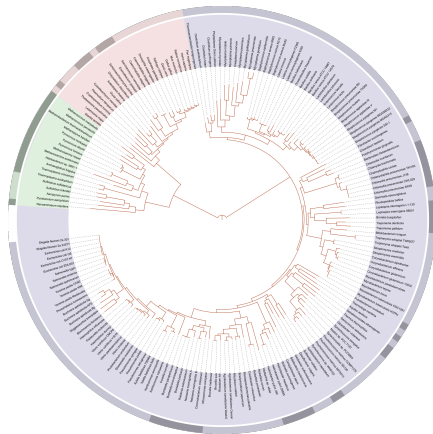
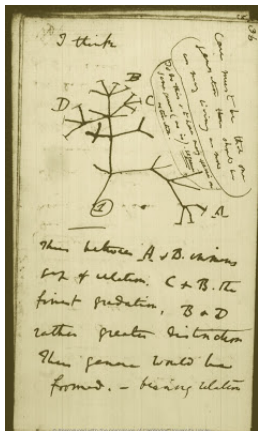


Figure: Darwin's tree and the Interactive Tree of Life (itol.embl.de)

# Network of Life?

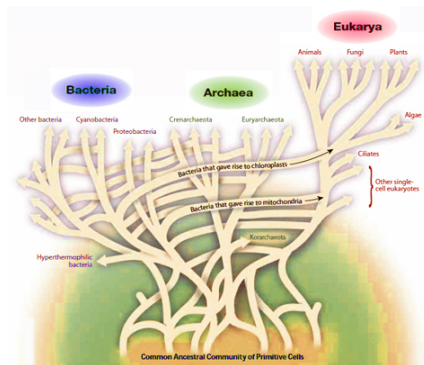
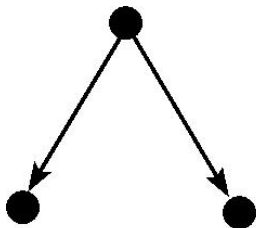


Figure: <http://www.nms.ac.jp/fesworld/top.html>

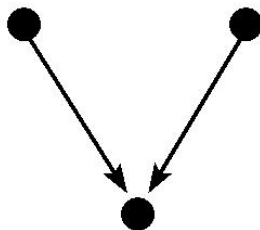
# Tree models can be inadequate

Phylogenetic trees are unable to represent the result of evolutionary processes such as

- ▶ recombination
- ▶ hybridization
- ▶ horizontal gene transfer

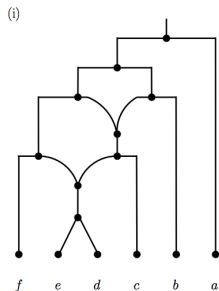


Speciation



Recombination

# Phylogenetic networks



A (rooted) phylogenetic network:

- ▶ a directed acyclic graph
- ▶ a unique root
- ▶ leaves are labelled by species
- ▶ no vertex with one parent and one child
- ▶ binary

**Central problem:** How to reconstruct phylogenetic networks from biological data?

# Example

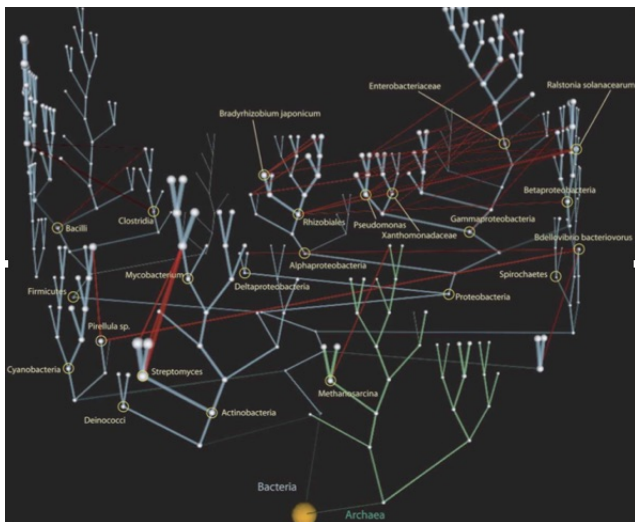
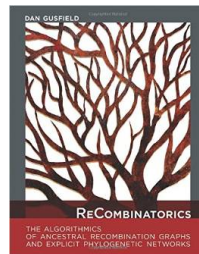
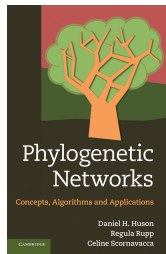
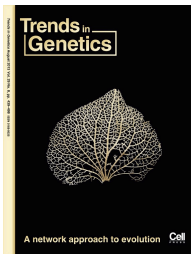
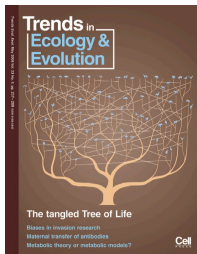
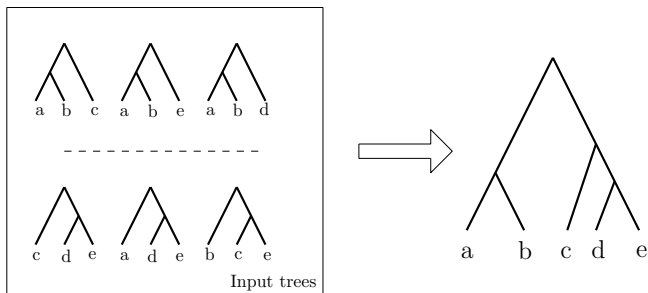


Figure: [Kunin, Goldovsky, Darzentas, Ouzounis, Genome Res. 2005]

# Active and growing area



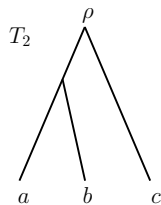
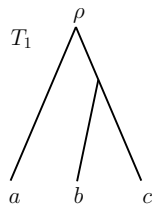
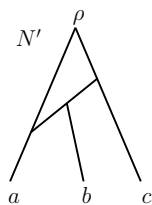
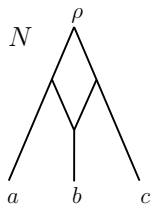
Idea: Assemble large trees from small trees



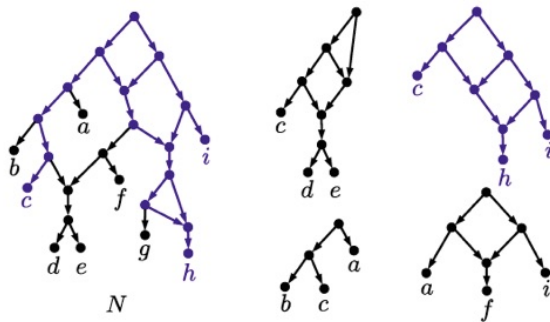
Could we use such an approach to build networks?



Networks are not defined by their subtrees



Could we build networks up from smaller networks?



# Trinets

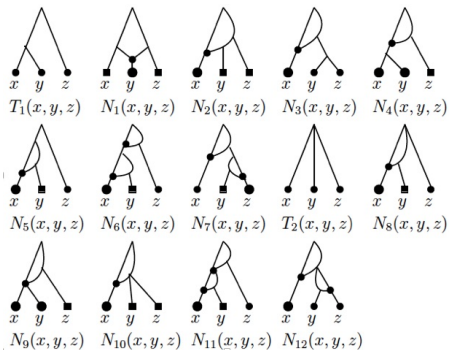
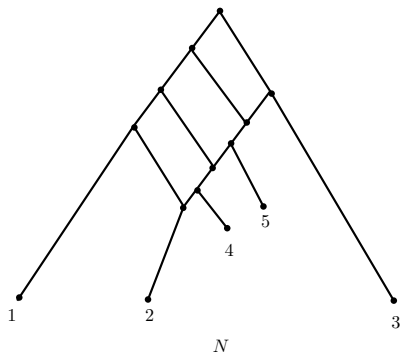


Figure: Trinets are 3-leaved networks.

# Trinets can define ...

Tree-child networks are determined by their trinets  
[van Iersel, Moulton 2014, J. Math. Biol.]



**Figure:** Tree-child: Each interior vertex has a child that has in-degree one (i.e., a tree vertex or leaf).

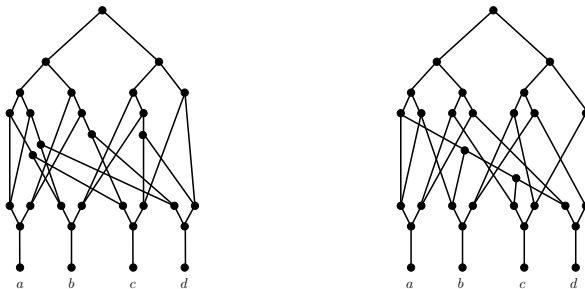
... but not always

Theorem (Huber-Iersel-Moulton-Wu, 2015, *Syst. Biol.*)

*For every  $n \geq 3$ , there exist two non-isomorphic phylogenetic networks  $N_1$  and  $N_2$  with  $n$  leaves such that*

- (i) they display the same set of subnetworks, and*
- (ii) they display the same set of trees.*

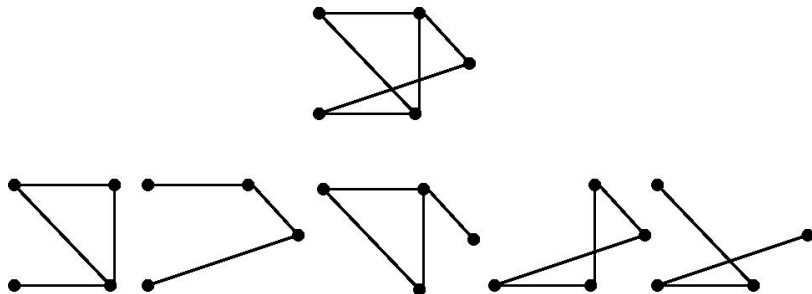
# First in an infinite family of counterexamples



**Figure:** Two different networks  $N_1$  and  $N_2$  that display the same set of subnetworks and trees.

# The reconstruction conjecture [Kelly, Ulam]

Open problem: Is a graph determined by its vertex-deleted subgraphs?



Note: False for *directed* graphs [Stockmeyer, J. Graph Theory 1977]

# Level-1 networks are tree-child

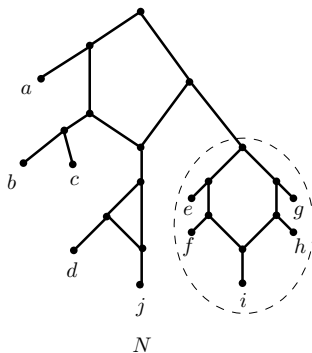


Figure: Level-1 network  $\equiv$  all undirected cycles are disjoint

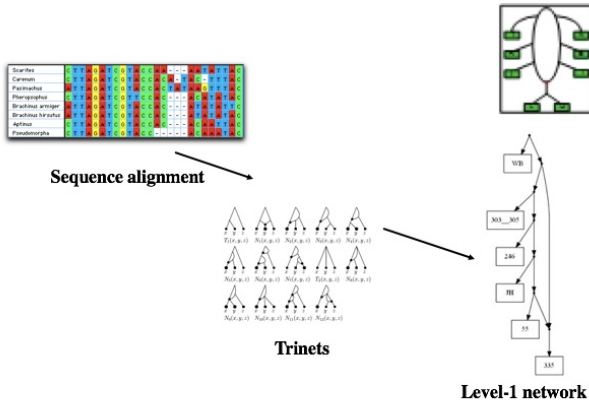


# Key difference between trees and level-1 networks

- ▶ Can decide in polynomial time if an arbitrary set of triplets is displayed by a tree (and, if so, can build it in polynomial time) [Aho 1984]
- ▶ NP-complete to decide if an arbitrary set of level-1 trinets is displayed by a level-1 network, but polynomial for “dense” sets [Huber, van Iersel, Moulton, Scornavacca, Wu, Algorithmica, 2016]

# TriLoNet (Trinet Level One Network)

Builds level-1 networks from sequence data by assembling trinetets



<https://www.uea.ac.uk/computing/trilonet#2> [Oldman et al., Mol Biol Evol. 2016]

# TriLoNet vs Lev1athon

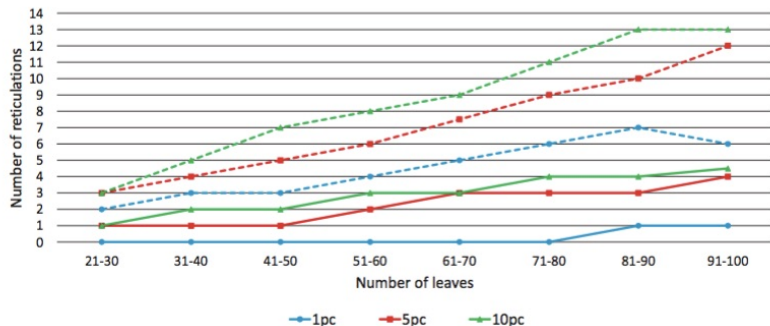


Figure: Simulated noise





Some open problems:

- ▶ For  $n = 4$ , our counter-example contains  $l = 4$  leaves but  $r = 12$  reticulations. Do there exist counterexamples with  $r < l$ ?
- ▶ A graph drawn uniformly at random is determined by its subgraphs with probability 1, as the size of the vertex set goes to infinity [Bollobás, 1990]. What about phylogenetic networks?
- ▶ Extensions to edge-weighted phylogenetic networks?
- ▶ Computing small likelihood-based networks from sequence data?

# Thanks

- ▶ Katharina Huber (UEA)
- ▶ Leo van Iersel (TU Delft, the Netherlands)
- ▶ James Oldman (UEA)
- ▶ Celine Scornavacca (Institut des Sciences de l'Evolution, Montpellier)
- ▶ Taoyang Wu (UEA)