

# Multirooted networks: What and why?

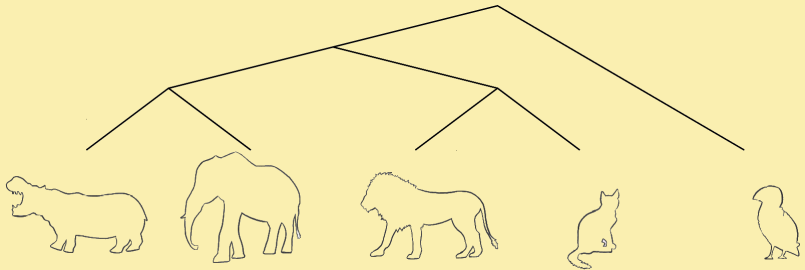
Guillaume Scholz



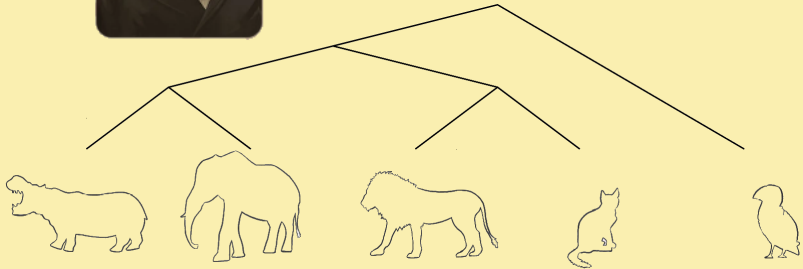
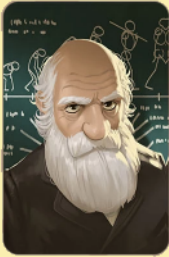
UNIVERSITÄT  
LEIPZIG

Bled 2024

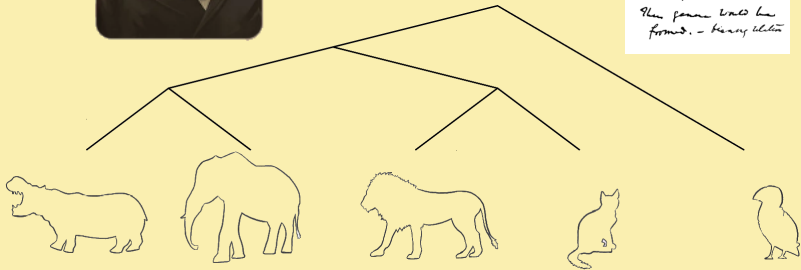
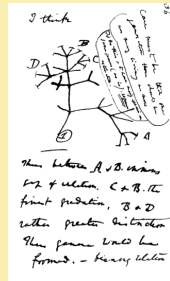
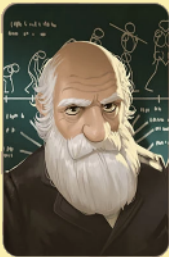
# Phylogenetic tree (~1850)



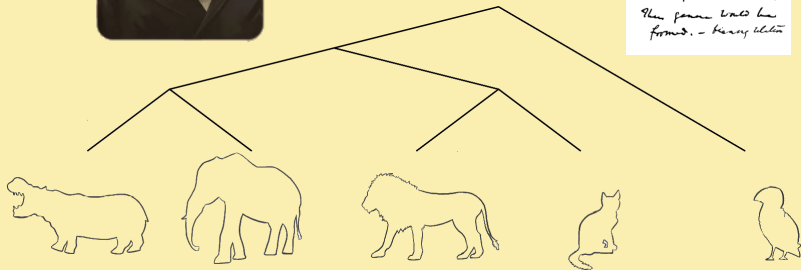
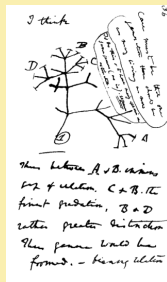
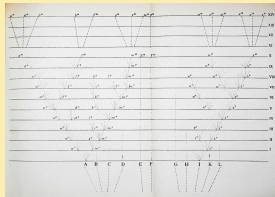
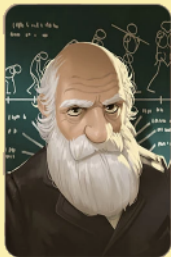
# Phylogenetic tree (~1850)



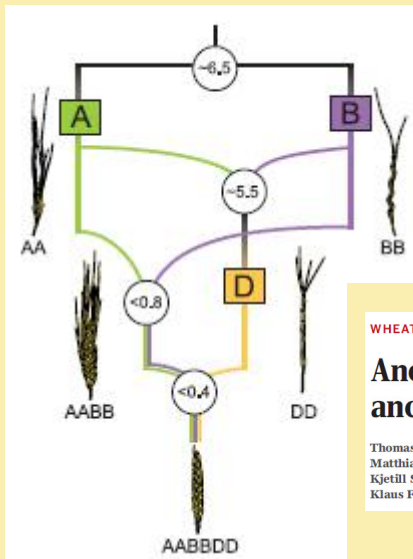
# Phylogenetic tree (~1850)



# Phylogenetic tree (~1850)



# Phylogenetic network (early 90s)



## WHEAT GENOME

### Ancient hybridizations among the ancestral genomes of bread wheat

Thomas Marcussen,<sup>1\*</sup> Simen R. Sandve,<sup>1\*†</sup> Lise Heier,<sup>2</sup> Manuel Spannagl,<sup>3</sup> Matthias Pfeifer,<sup>3</sup> The International Wheat Genome Sequencing Consortium,<sup>‡</sup> Kjetill S. Jakobsen,<sup>4</sup> Brande B. H. Wulff,<sup>5</sup> Burkhard Steuernagel,<sup>5</sup> Klaus F. X. Mayer,<sup>3</sup> Odd-Arne Olsen<sup>1</sup>

# Phylogenetic network (early 90s)

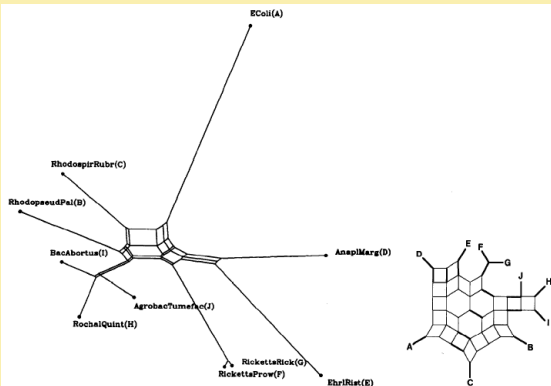


FIG. 4. A. A graphical representation of the split-decomposable part of the evolutionary distances between Rickettsiales and other eubacteria drawn to scale (data from Table 4 of Weisburg *et al.* (1991)). Taxon symbols are: (A) *Escherichia coli*; (B) *Rhodopseudomonas palustris*; (C) *Rhodospirillum rubrum*; (D) *Anaplasma marginale*; (E) *Ehrlichia risticii*; (F) *Rickettsia prowazekii*; (G) *Rickettsia rickettsii*; (H) *Rochalimaea quintana*; (I) *Bacillus abortus*; (J) *Agrobacterium tumefaciens*. B. A graph, isomorphic to the graph in A, but with all edges given the same length. Bold lines indicate links corresponding to the splits with isolation index larger than 10 (per 1000 sites). Taxon symbols are as in A.

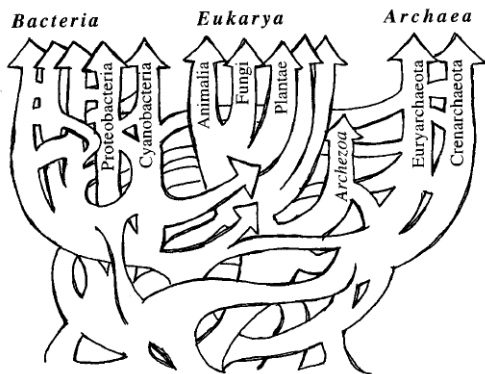
## Split Decomposition: A New and Useful Approach to Phylogenetic Analysis of Distance Data

HANS-JÜRGEN BANDEL<sup>\*</sup> AND ANDREAS W. M. DRES<sup>†</sup>

<sup>\*</sup>Mathematisches Seminar, Universität Hamburg, D-2000 Hamburg 13, Federal Republic of Germany; and <sup>†</sup>Fakultät für Mathematik, Universität Bielefeld, D-4800 Bielefeld 1, Federal Republic of Germany

# Phylogenetic network (early 90s)

Fig. 3. A reticulated tree, or net, which might more appropriately represent life's history. Martin (16), in a review covering many of the same topics as this one, has presented some striking colored representations of such patterns.



REVIEW

## Phylogenetic Classification and the Universal Tree

W. Ford Doolittle



## Networks: expanding evolutionary thinking

**Eric Bapteste<sup>1</sup>, Leo van Iersel<sup>2</sup>, Axel Janke<sup>3</sup>, Scot Kelchner<sup>4</sup>, Steven Kelk<sup>5</sup>, James O. McInerney<sup>6</sup>, David A. Morrison<sup>7</sup>, Luay Nakhleh<sup>8</sup>, Mike Steel<sup>9</sup>, Leen Stougie<sup>2,10</sup>, and James Whitfield<sup>11</sup>**

<sup>1</sup> Université Pierre et Marie Curie, Paris, France

<sup>2</sup> Centrum Wiskunde and Informatica, Amsterdam, The Netherlands

<sup>3</sup> Goethe University, Frankfurt am Main, Germany

<sup>4</sup> Idaho State University, Pocatello ID, USA

<sup>5</sup> Maastricht University, Maastricht, The Netherlands

<sup>6</sup> National University of Ireland, Maynooth, Ireland

<sup>7</sup> Sveriges Lantbruksuniversitet, Uppsala, Sweden

<sup>8</sup> Rice University, Houston TX, USA

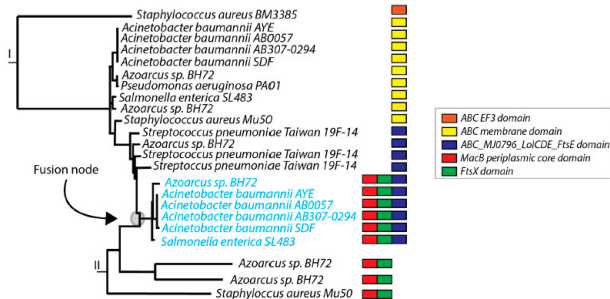
<sup>9</sup> University of Canterbury, Christchurch, New Zealand

<sup>10</sup> Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

<sup>11</sup> University of Illinois, Urbana IL, USA

**Networks allow the investigation of evolutionary relationships that do not fit a tree model. They are becoming a leading tool for describing the evolutionary relationships between organisms, given the comparative complexities among genomes.**

# Multi-rooted fusion graph (2015)



**Figure 5.** Two-rooted fusion graph. This two-rooted graph was constructed using the two phylogenetic trees from Figure 4. The trees were mid-point rooted and merged using Adobe Illustrator. The two roots are marked I and II. The grey dot, labelled "Fusion node" indicates the approximate location of the fusion event. The coloured squares display the domain architecture of the genes.

computation

ISSN 2079-3197

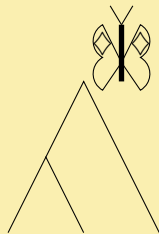
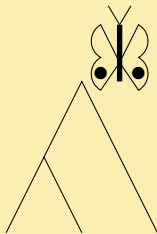
www.mdpi.com/journal/computation

Article

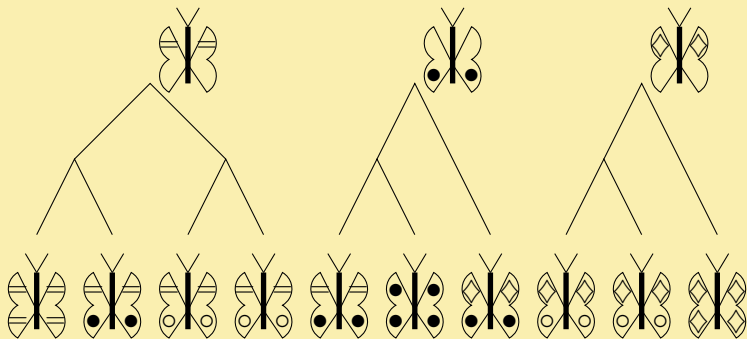
## Evolution by Pervasive Gene Fusion in Antibiotic Resistance and Antibiotic Synthesizing Genes

Orla Coleman <sup>†</sup>, Ruth Hogan <sup>†</sup>, Nicole McGoldrick <sup>†</sup>, Niamh Rudden <sup>†</sup> and James O. McInerney <sup>\*</sup>

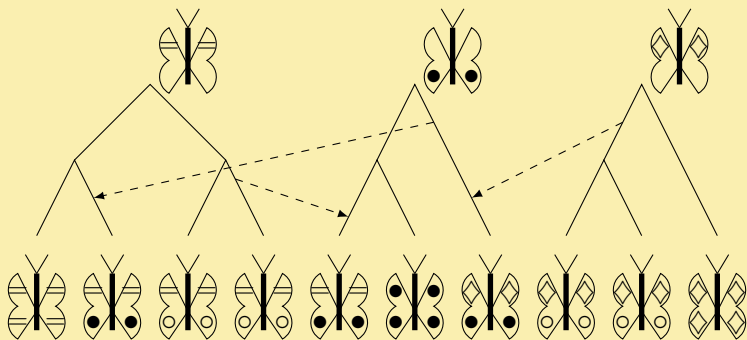
# Introgression



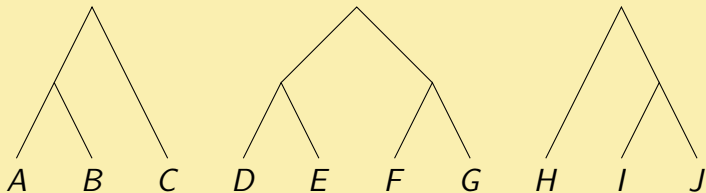
# Introgression



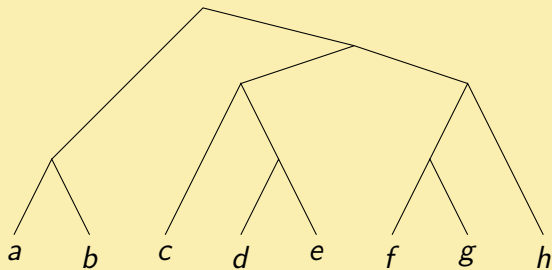
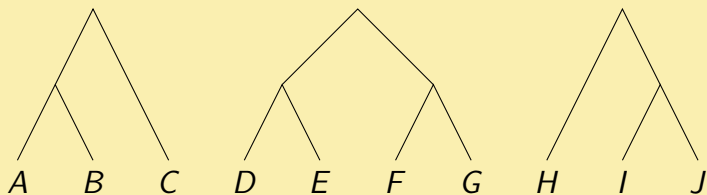
# Introgression



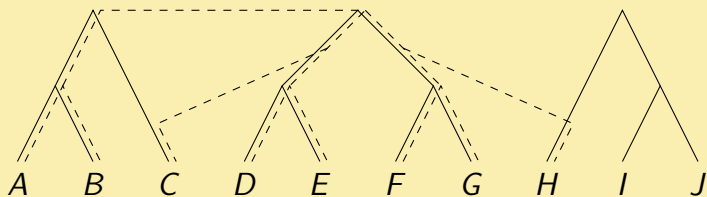
# Introgression detection algorithm



# Introgression detection algorithm

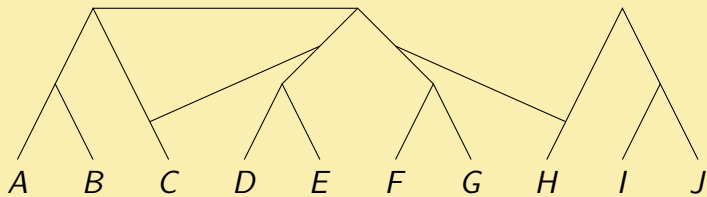


# Introgression detection algorithm





# Introgression detection algorithm



So...

Multiple-rooted networks can be used to represent:

So...

Multiple-rooted networks can be used to represent:

- Recombinations between members of different gene families

So...

Multiple-rooted networks can be used to represent:

- Recombinations between members of different gene families
- Introgression between members of distinct lineages

Multiple-rooted networks can be used to represent:

- Recombinations between members of different gene families
- Introgression between members of distinct lineages

They are useful when the common history of several families of taxa is unknown and/or irrelevant.

Multiple-rooted networks can be used to represent:

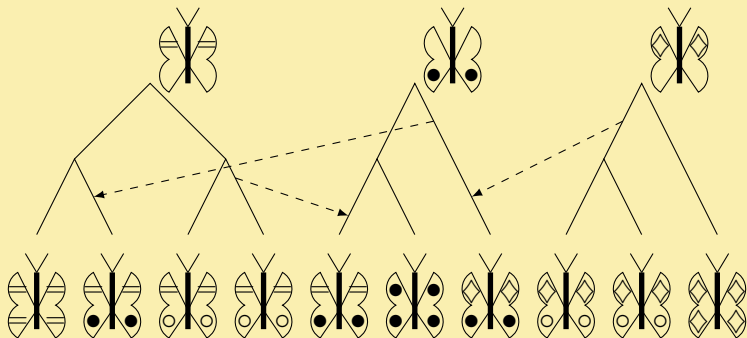
- Recombinations between members of different gene families
- Introgression between members of distinct lineages

They are useful when the common history of several families of taxa is unknown and/or irrelevant.

That's a good start!

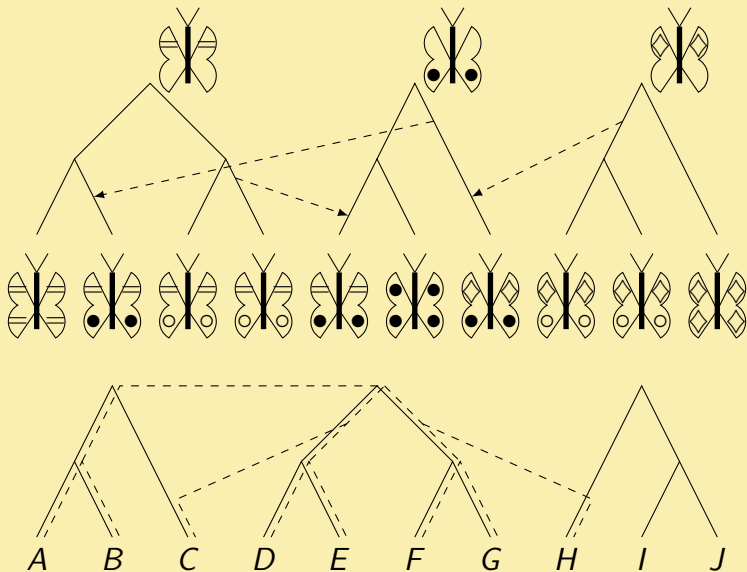
- GS, A.-A. Popescu, M. I. Taylor, V. Moulton and K. T. Huber. OSF-BUILDER: A new tool for reconstructing and representing phylogenetic histories involving introgression, *Systematic Biology* (2019) 68(5):717-729.
- K. T. Huber, V. Moulton and GS. Forest-based networks. *Bulletin of Mathematical Biology* (2022) 84: 119.
- K. T. Huber, V. Moulton and GS. Shared ancestry graphs and symbolic arboreal maps. (submitted to *SIAM Journal on Discrete Mathematics*).
- K. T. Huber, L. van Iersel, V. Moulton and GS. Is this network proper forest-based? (submitted to *Information Processing Letters*).
- GS. Representing distance-hereditary graphs with multi-rooted trees. (submitted to *Graphs and Combinatorics*).

# Forest-based networks

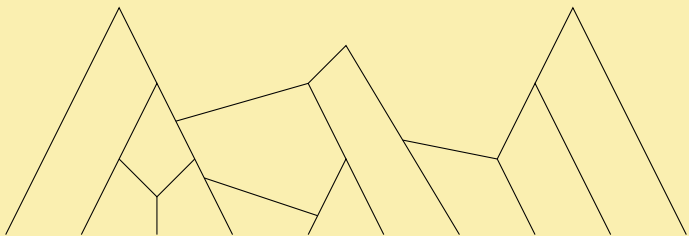




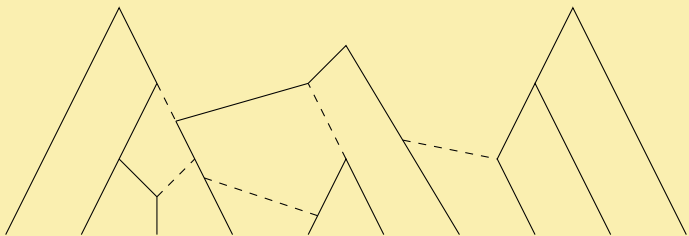
# Forest-based networks



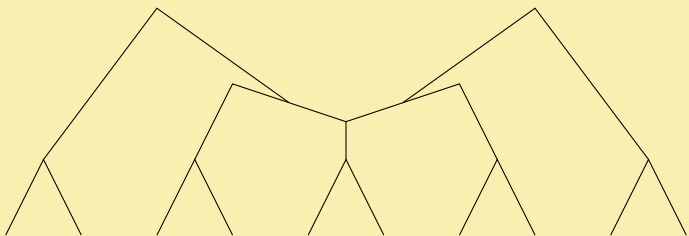
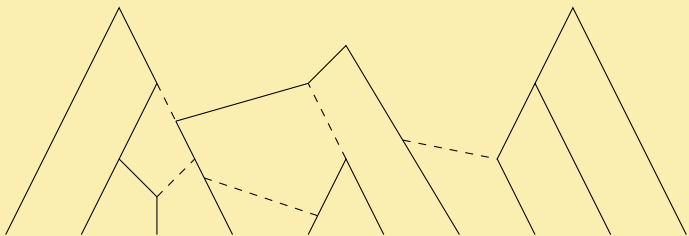
# Forest-based networks



# Forest-based networks



# Forest-based networks



# Forest-based networks

- Deciding whether a given multi-rooted network is forest-based is NP-hard.

# Forest-based networks

-Deciding whether a given multi-rooted network is forest-based is NP-hard.

even if the network has only two roots

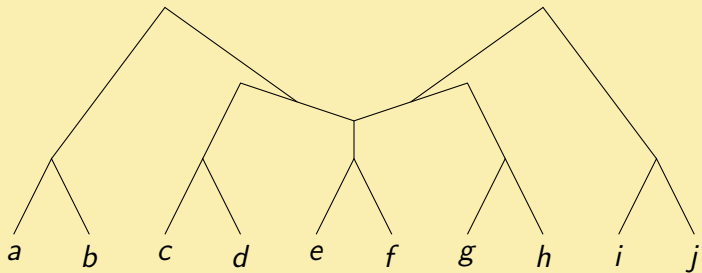
# Forest-based networks

-Deciding whether a given multi-rooted network is forest-based is NP-hard.

even if the network has only two roots

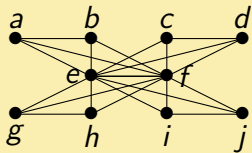
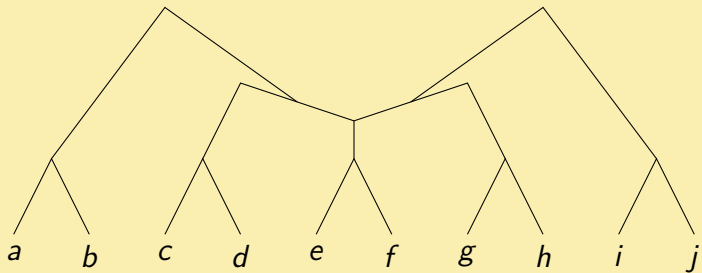
even when restricting to forests in which  
the number of trees coincides with  
the number of roots of the network

# Arboreal networks

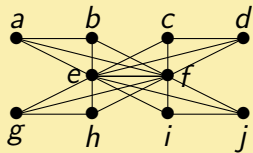
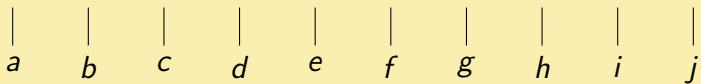




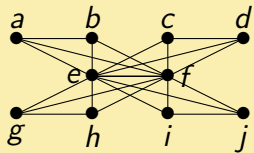
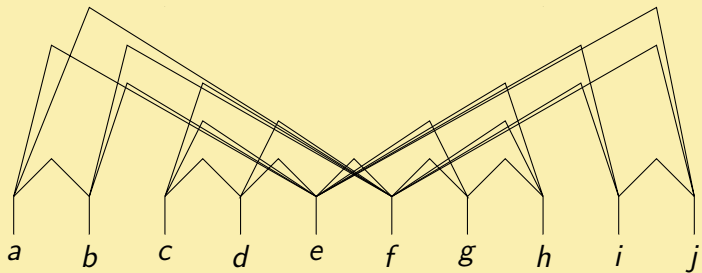
# Arboreal networks



# Arboreal networks



# Arboreal networks



# Arboreal networks

Any undirected graph  $G$  is the shared-ancestry graph of some mult rooted network.

# Arboreal networks

Any undirected graph  $G$  is the shared-ancestry graph of some mult rooted network.

If  $G$  is the shared-ancestry graph of some arboreal network  $N$ , then  $N$  has minimal number of roots and hybrid vertices.

# Arboreal networks

Any undirected graph  $G$  is the shared-ancestry graph of some mult rooted network.

If  $G$  is the shared-ancestry graph of some arboreal network  $N$ , then  $N$  has minimal number of roots and hybrid vertices.

$G$  is the shared-ancestry graph of some arboreal network  $N$  if and only if  $G$  is Ptolemaic.

# Arboreal networks

Any undirected graph  $G$  is the shared-ancestry graph of some mult rooted network.

If  $G$  is the shared-ancestry graph of some arboreal network  $N$ , then  $N$  has minimal number of roots and hybrid vertices.

$G$  is the shared-ancestry graph of some arboreal network  $N$  if and only if  $G$  is Ptolemaic.

In that case, such a network  $N$  can be built in polynomial time.

# Ptolemaic graphs

$G$  is Ptolemaic if the inequality:

$$d(u, v)d(x, y) + d(u, x)d(v, y) \geq d(u, y)d(v, x)$$

holds for all vertices  $x, y, u, v$ .



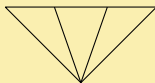
# Ptolemaic graphs

$G$  is Ptolemaic if the inequality:

$$d(u, v)d(x, y) + d(u, x)d(v, y) \geq d(u, y)d(v, x)$$

holds for all vertices  $x, y, u, v$ .

$G$  is Ptolemaic if it does not contain:



holes

gem

as induced subgraphs.

## **Applications**

Mult rooted networks provide an alternative to phylogenetic networks to represent complex evolutionary events (recombination, introgression, ...)

## **Applications**

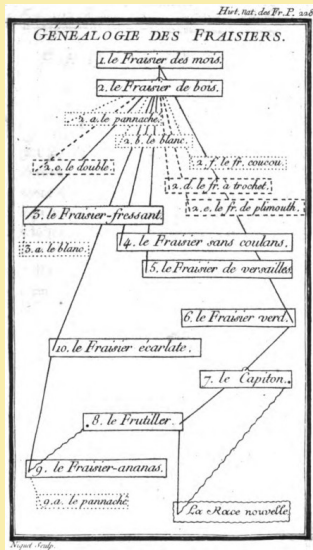
Mult rooted networks provide an alternative to phylogenetic networks to represent complex evolutionary events (recombination, introgression, ...)

## **Theory**

Mult rooted networks offer a new and exciting playground to mathematicians, with connections to graph theory, combinatorics, algorithmics, ...



# Actually...



Duchesne, 1766

# One last thing

We now have a **Graph Theory Seminar Series** in Leipzig!