Phylogenetic Networks

part of “Graphen und Netzwerke in der Biologie”

Sonja Prohaska

Computational EvoDevo
University Leipzig

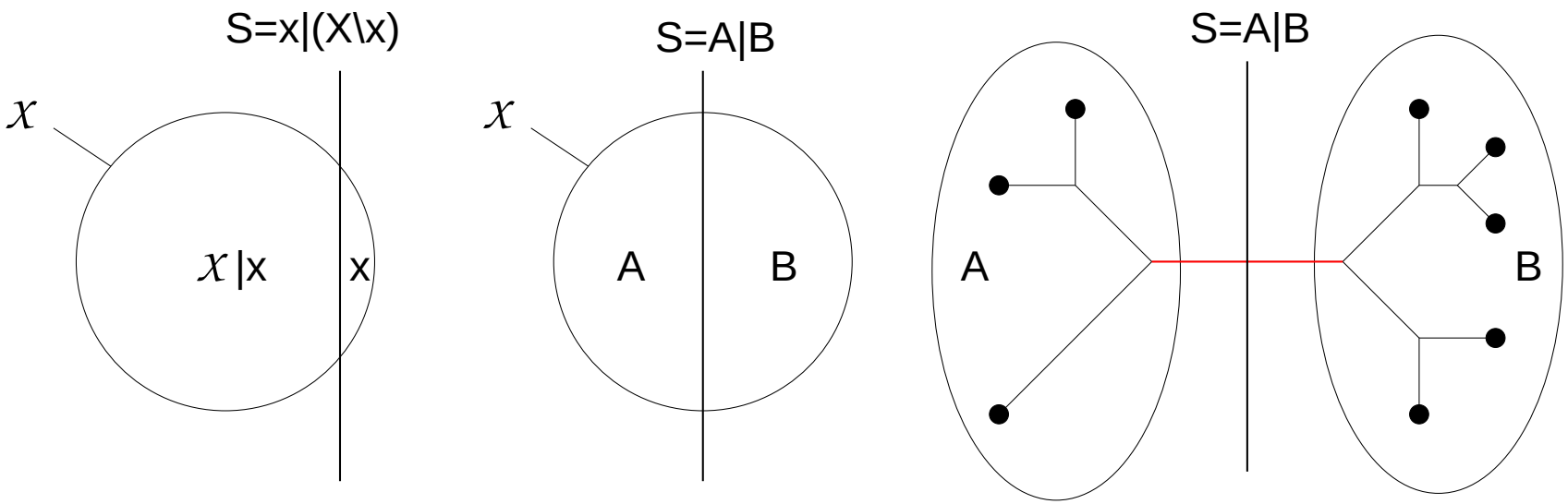
Leipzig, SS 2011

splits networks

- are unrooted phylogenetic networks
- contain as least as much information as a phylogenetic tree
- usually contain more information than a tree
- visualize information that is conflicting in tree representation
- monophyletic groups of sequences are more easily identified

What is a split?

A split $S = A|B$ is a bipartition of the sequences X into two non-empty, mutually exclusive subsets A and B. A and B are called *split parts*.



A *trivial split* $S = x|(X \setminus x)$ seperates one sequence x from the other sequences $X \setminus x$.

Relation Between Splits and Trees

Given a set of splits S on X, can it be represented by a tree?

Compartibility: two splits $S_1 = A_1|B_1$ and $S_2 = A_2|B_2$ are *compartible* if one of the four possible intersections of the split parts is empty (or splits don't "cross" each other).

general case

per def.

$A_1 \cap B_1 = \{\}$

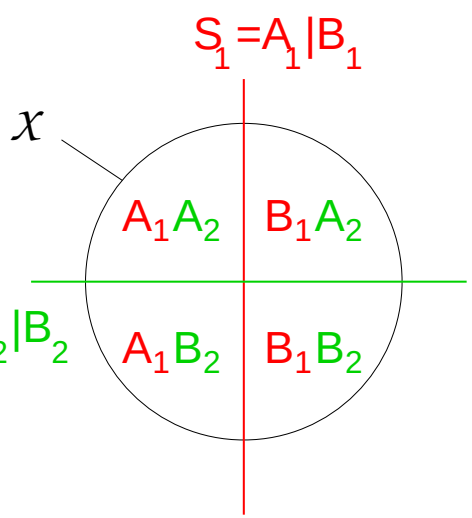
$A_2 \cap B_2 = \{\}$

$A_1 \cap A_2$

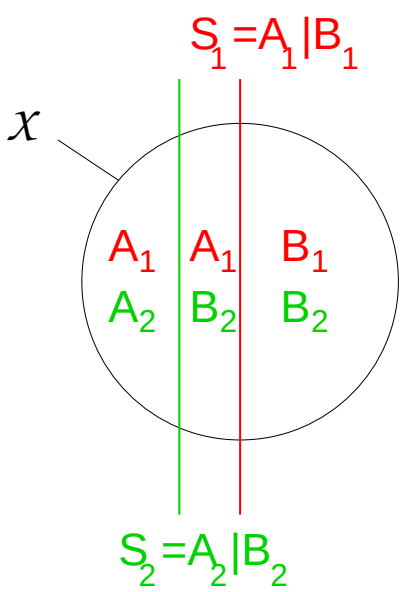
$B_1 \cap A_2$

$A_1 \cap B_2$

$B_1 \cap B_2$



compartible case



per def.

$A_1 \cap B_1 = \{\}$

$A_2 \cap B_2 = \{\}$

$A_1 \cap A_2$

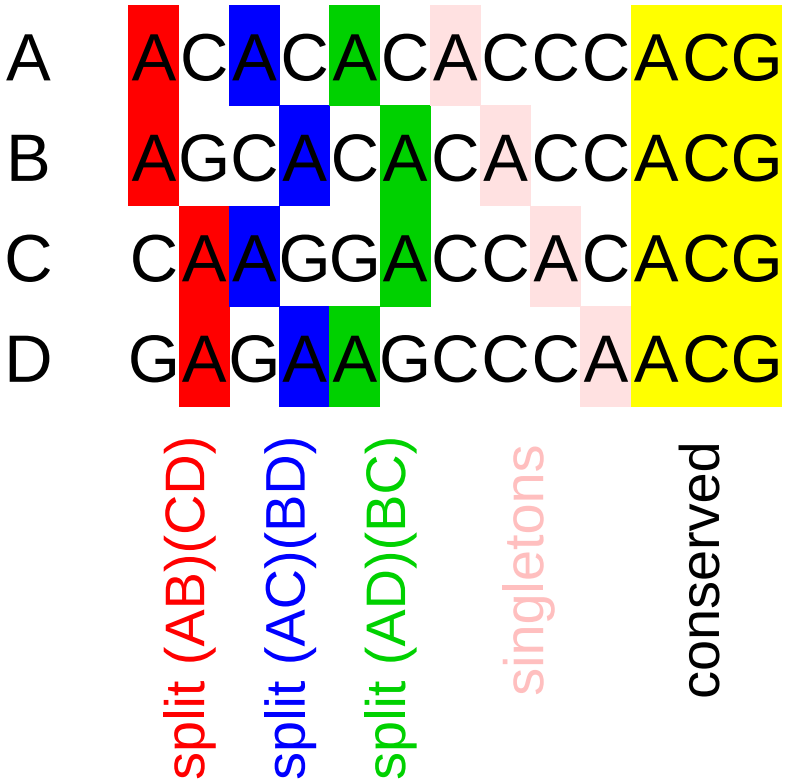
$B_1 \cap A_2 = \{\}$

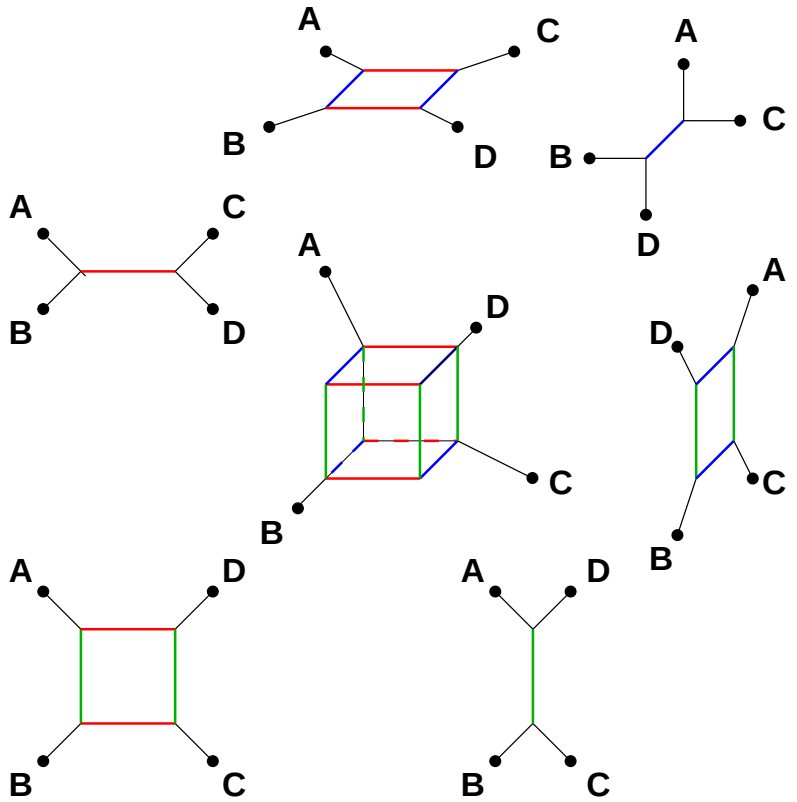
$A_1 \cap B_2$

$B_1 \cap B_2$

A set of splits is comartible if all pairs of splits are compartible.
A compartible set of splits can be represented as a tree.

The Information to be represented



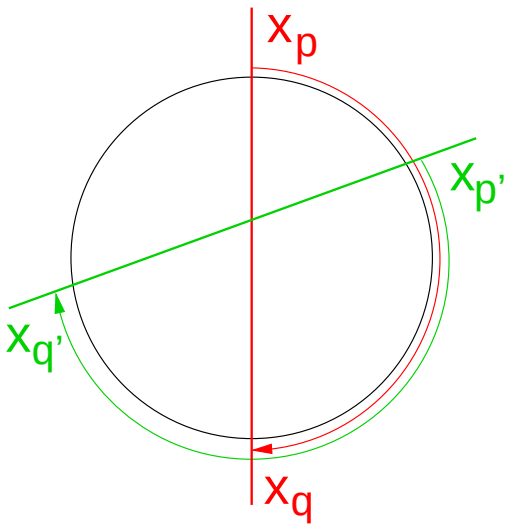


Circular splits

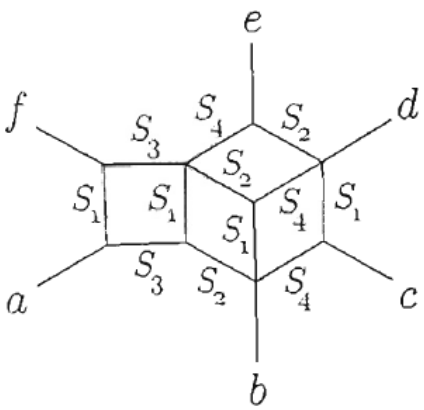
A set of splits S on X is called circular, if there exists a linear ordering $(x_1, x_2, ..., x_n)$ of the elements of X such that each split S_i has the form

$$S_i = \frac{x_p, x_{p+1}, ..., x_q}{X - x_p, x_{p+1}, ..., x_q} \quad (1)$$

for $1 < p <= q <= n$.



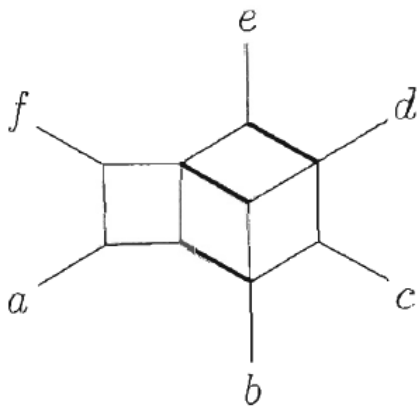
non-trivial splits



(a) Split network N

$S_1 = \{a, b, c\}|\{d, e, f\}$
$S_2 = \{a, e, f\}|\{b, c, d\}$
$S_3 = \{a, f\}|\{b, c, d, e\}$
$S_4 = \{a, b, f\}|\{c, d, e\}$

(b) Non-trivial splits



(c) Split $S_2 = \frac{\{a,e,f\}}{\{b,c,d\}}$

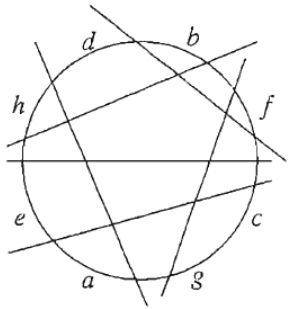
Figure 5.6 (a) A split network N representing all trivial splits on $\mathcal{X} = \{a, \ldots f\}$ and the four non-trivial splits listed in (b). All the edges representing a particular non-trivial split are labeled by that split. However, the labeling of edges by splits is usually omitted, as shown in (c). Edges representing the same split, such as the three edges shown in bold lines representing S_2, are drawn parallel and with the same length.

Is the splits system a circular splits system? Why?

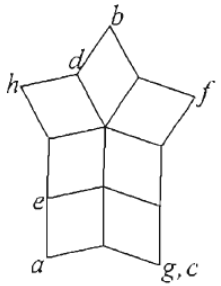
Circular Splits Systems Can be Representd by a Planar Network

$\{a,b,d,e,h\} \mid \{c,f,g\}$
$\{a,c,d,e,g,h\} \mid \{b,f\}$
$\{a,c,e,g\} \mid \{b,d,f,h\}$
$\{a,c,g\} \mid \{b,d,e,f,h\}$
$\{a,c,e,f,g\} \mid \{b,d,h\}$
$\{a,e,h\} \mid \{b,c,d,f,g\}$

(a) Circular splits



(b) Circular ordering

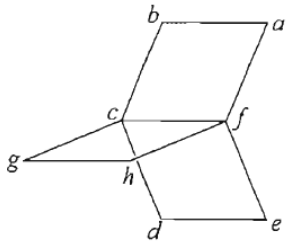


(c) Planar network

Figure 5.9 (a) A set of six circular splits $\mathcal{S}$ on $\mathcal{X} = \{a, b, \ldots, h\}$. (b) An arrangement of the taxa around a circle such that every split $S = A \mid B \in \mathcal{S}$ can be realized by a straight line through the circle that separates the two split parts A and B. A circular ordering is given by (a, g, c, f, b, d, h, e). (c) An outer-labeled planar split network representing $\mathcal{S}$.

$\{a,b\} \mid \{c,d,e,f,g,h\}$
$\{a,b,c,d,e,f\} \mid \{g,h\}$
$\{a,b,c,f,g,h\} \mid \{d,e\}$
$\{a,e,f,h\} \mid \{b,c,d,g\}$

(a) Non-circular splits



(b) Non-planar network

Figure 5.10 (a) A set of four non-circular splits $\mathcal{S}$ on $\mathcal{X} = \{a, b, \ldots, h\}$. (b) A non-planar split network representing $\mathcal{S}$.

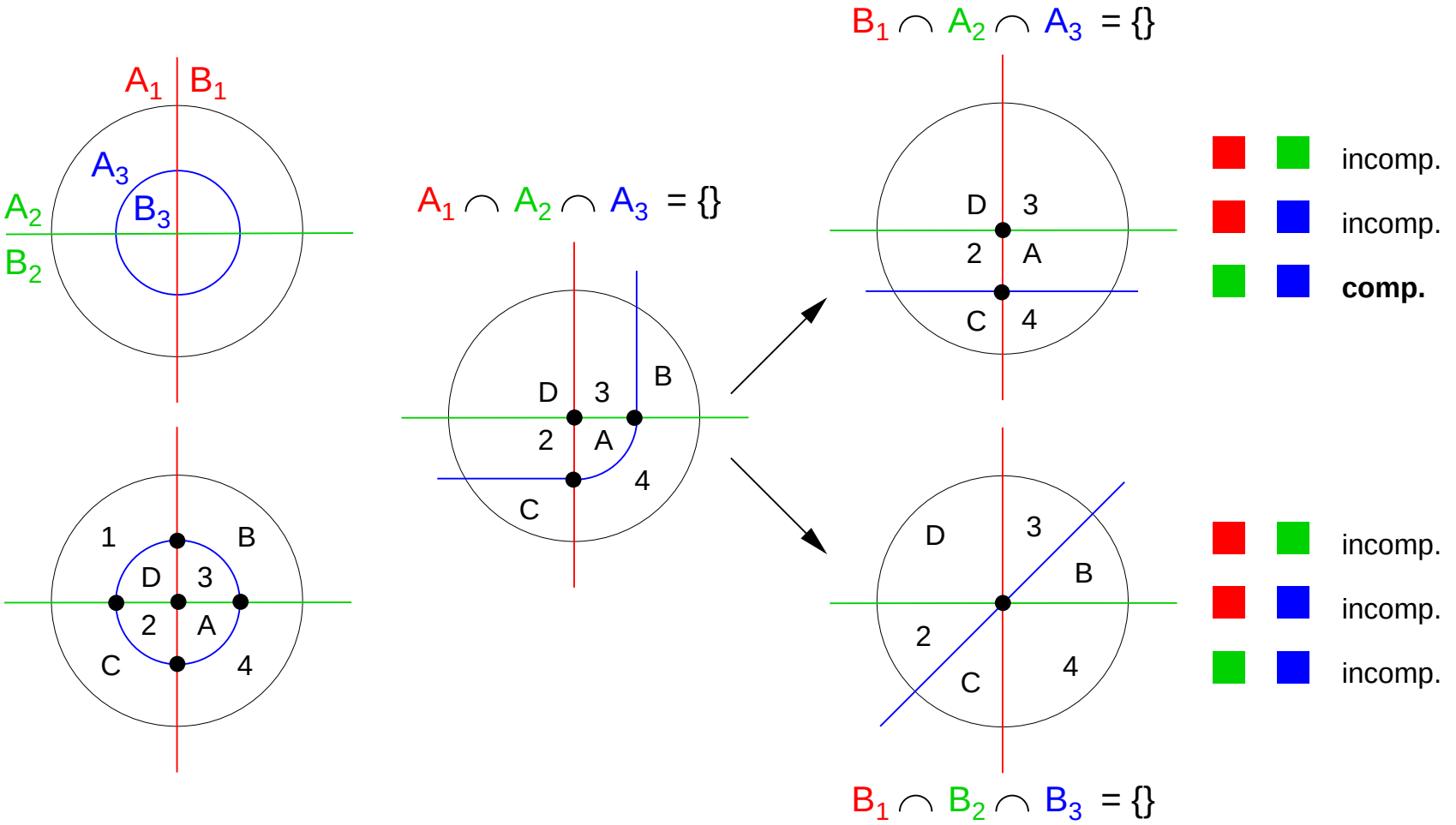
Weak compartibility

Three distinct splits are weakly compartible if

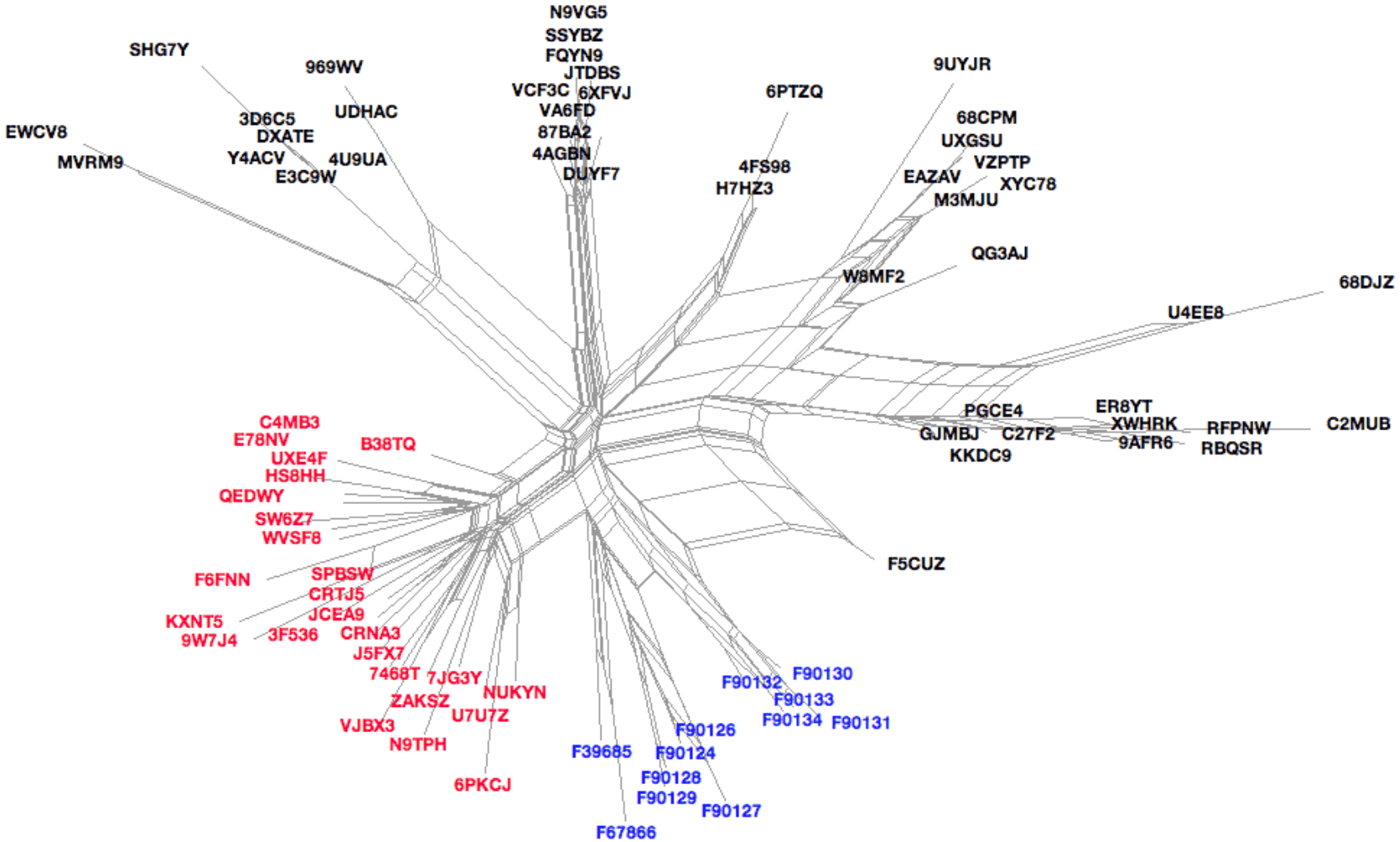
- at least one of the following intersections is empty:
  - $A_1 \cap A_2 \cap A_3$ (1)
  - $A_1 \cap B_2 \cap B_3$ (2)
  - $B_1 \cap A_2 \cap B_3$ (3)
  - $B_1 \cap B_2 \cap A_3$ (4)
- and symmetrically, at least one of the following intersections is empty:
  - $B_1 \cap B_2 \cap B_3$ (A)
  - $B_1 \cap A_2 \cap A_3$ (B)
  - $A_1 \cap B_2 \cap A_3$ (C)
  - $A_1 \cap A_2 \cap B_3$ (D)

A set of splits is weakly compartible if all triples are weakly compartible.

Weak compartibility



A Phylogenetic Network



Are the two group of sequences marked in red and blue supported by a split?

References

Phylogenetic Networks, Concepts, Algorithms and Applications by Daniel H Huson, Regula Rupp and Celine Scornavacca. Cambridge University Press, 2010. `http://www.cambridge.org/9780521755962`

Software **SplitsTree** `http://www.splitstree.org/`