

# DNA sequences II

Analyses of multiple sequence data datasets, incongruence tests, gene trees vs. species tree reconstruction, networks, detection of hybrid species

# DNA sequences II

- Test of congruence of multiple alignments [[ILD test](#), [Concaterpillar](#)]
- Visualization of gene tree incongruences using [NeighbourNet](#) [[Splitstree](#)]
- Reconstruction of [species tree](#) under multicoalescent model [[\\*BEAST](#)]
- Reconstruction of [species network](#) (ILS/hybridization) [[Phylonet](#)]

# Today's project

- 1) Evaluation of congruence of multiple sequence alignment
- 2) Identification of individuals that causing incongruence
- 3) Reconstruct of species tree
- 4) Reconstruction of species network

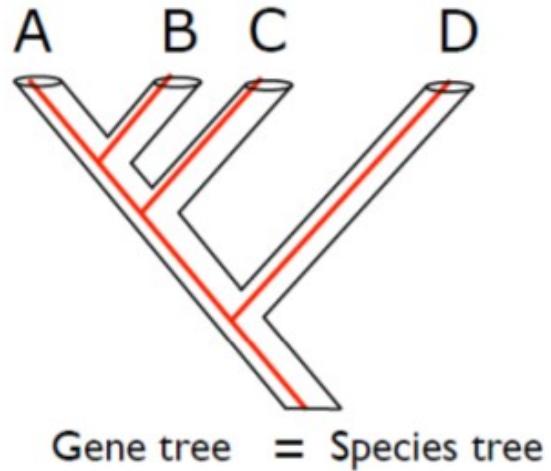
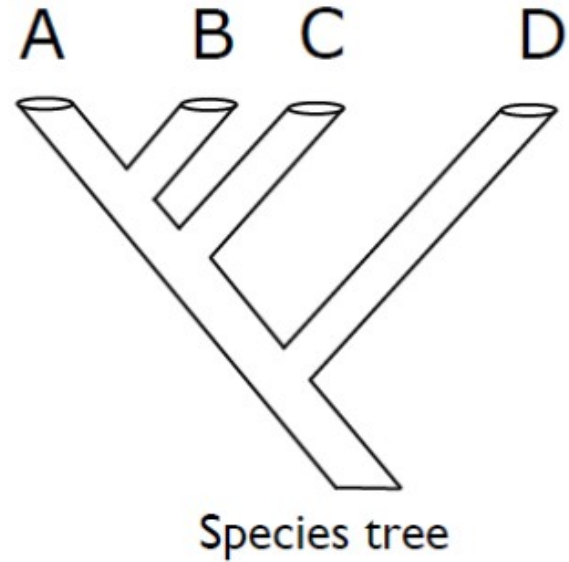
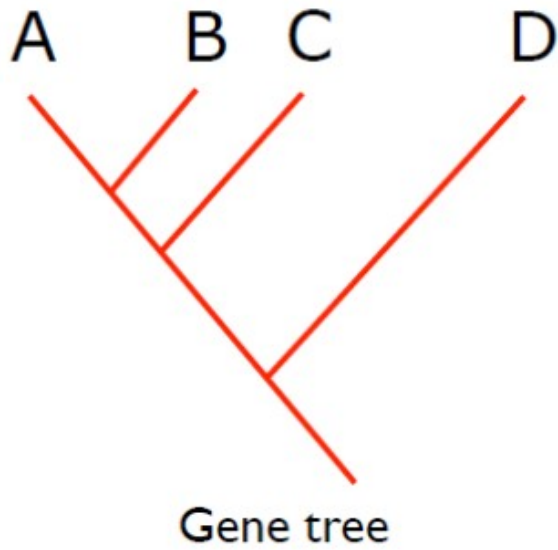
## AIMS:

- understand causes and consequences of differences between datasets
- choose the correct method for reconstruction of **species phylogeny**

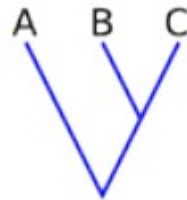
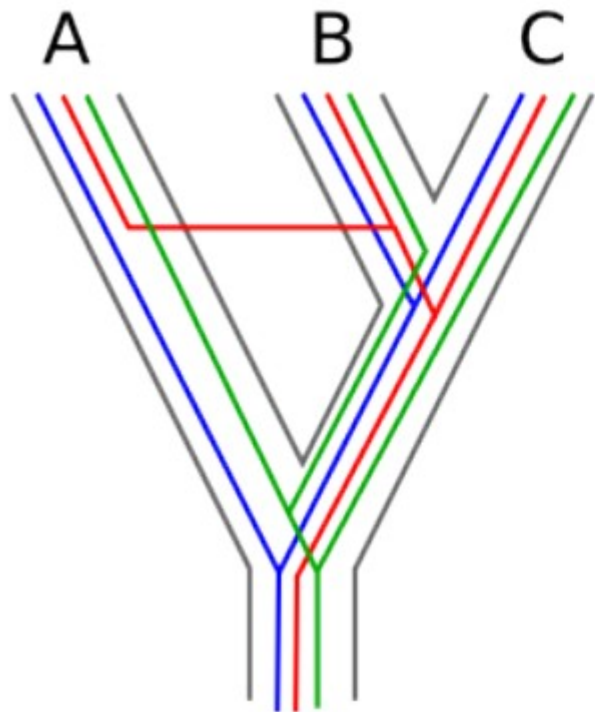
## Test input data

- Six alignments (from six independent markers) of *Curcuma* species

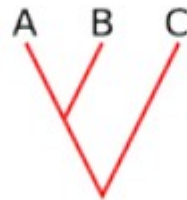
# Gene trees vs. Species tree



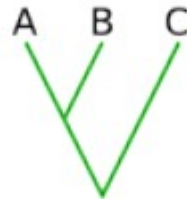
# (In)congruence of gene trees (alignments)



← Gene tree = species tree

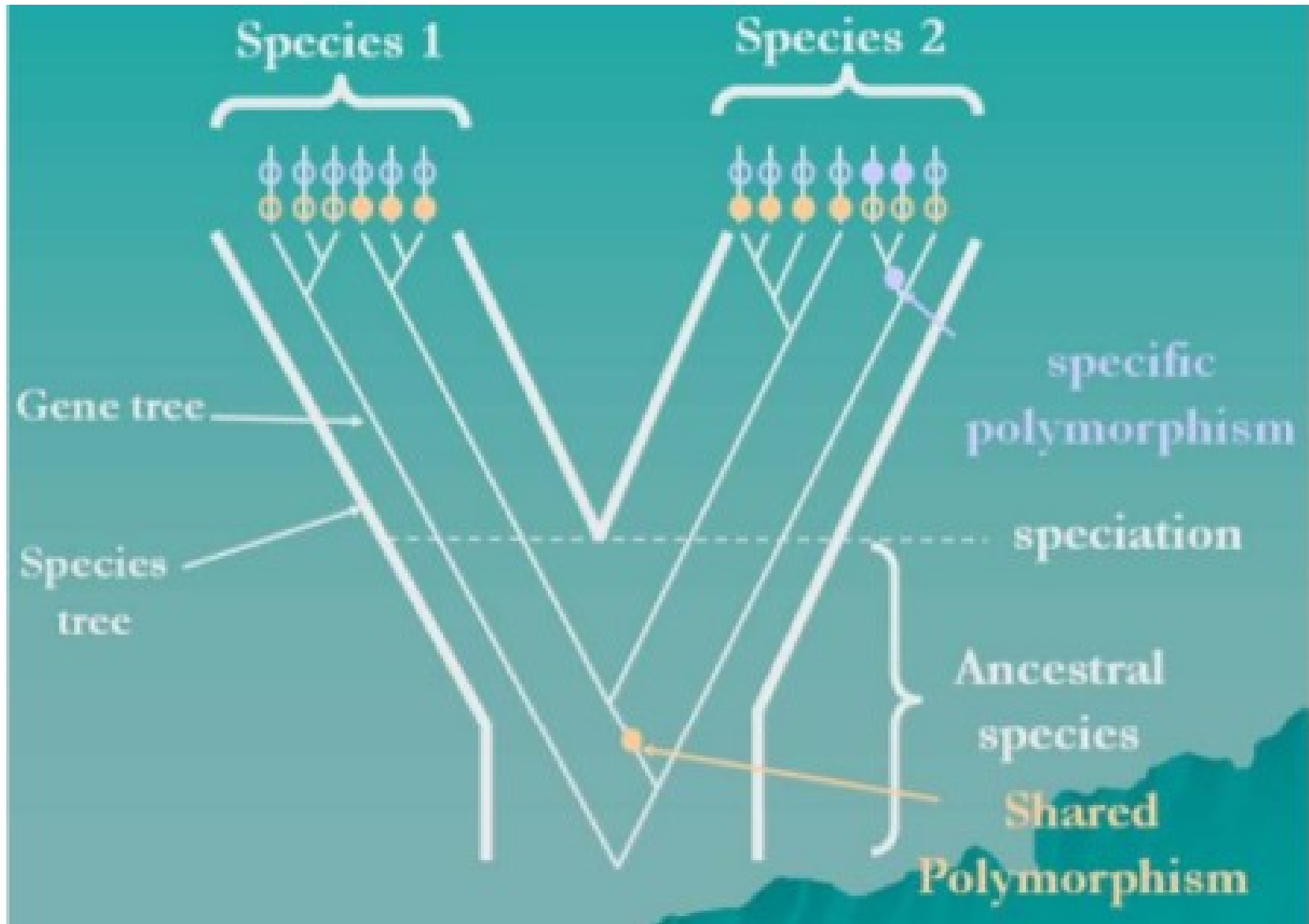


← Gene flow/hybridization

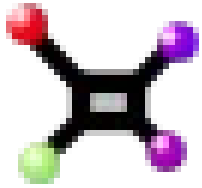


← Deep coalescence

Ancestral polymorphism ~ incomplete lineage sorting (ILS) ~  
deep coalescence



# Vizualization of conflicting datasets



NeighborNet- <http://www.splitstree.org/>

- distance based method
- powerful tool for conflict/hybrids detection

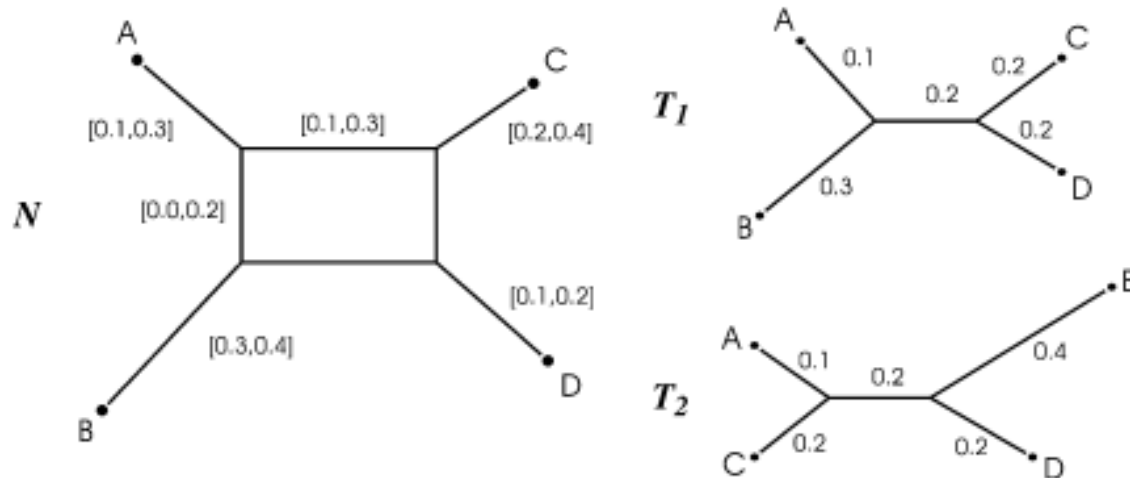


FIG. 6.—A simple confidence network  $N$  and two trees  $T_1, T_2$ . The network contains the tree  $T_1$  with the given branch lengths, but it does not contain tree  $T_2$  because the split  $AB|CD$  has weight 0 in tree  $T_2$  (i.e., it is not present), but the CI for the weight of  $AB|CD$  in the network does not include 0.

# Test of gene trees (dataset) incongruence

## Incongruence length difference (ILD) test (Farris et al., 1994)

- Test of homogeneity of datasets, H0: datasets are congruent
- Parsimony-based
- compares the length of the tree inferred from the combined data set to the combined length of the trees inferred for each locus in the data set
- rather outdated and questionable (e.g. Leigh, 2008)

## Shimodaira-Hasegawa (SH) test (Shimodaira, 2002)

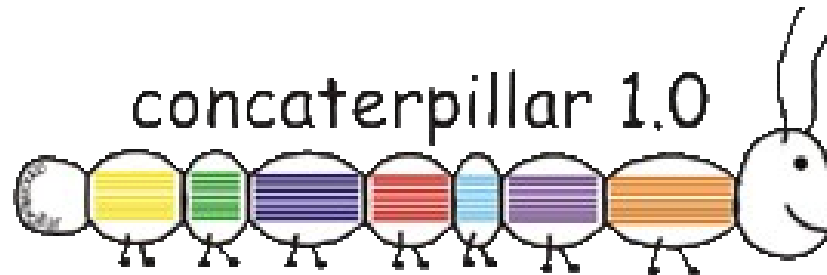
- Maximum-likelihood based
- determine whether individual markers reject the tree inferred from the concatenation of all markers
- also questionable (e.g. Leigh, 2008)



# Test of gene trees (dataset) incongruence

## Concaterpillar (Leigh, 2008)

- Test gene tree topology and branch length
- Maximum-likelihood based
- Can handle many markers in single analysis
- Running only on Linux/Mac



**CONGRUENCE TEST SELECTION** – might depends on the aims of the study, quality of data and amount of markers to be compared.

**READ** more about it, if you need it for your studies!

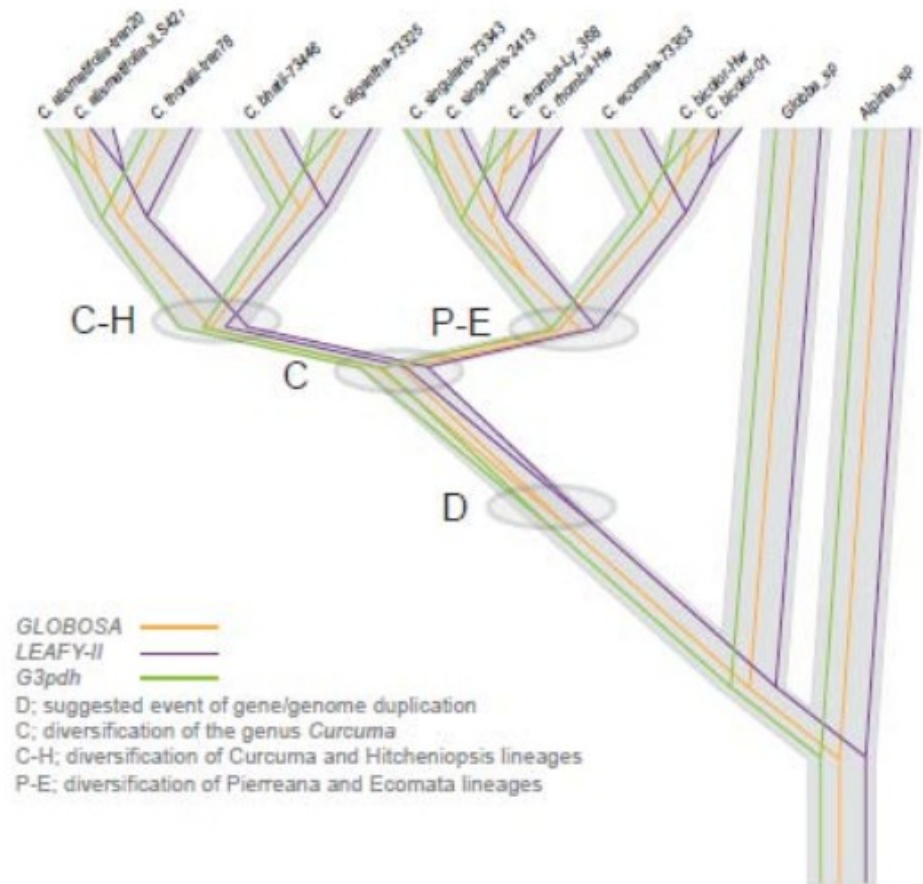
# Species tree reconstruction

- **X Concatenating** multiple alignments into single one --> analysing the data as a single locus?
  - Likely results in wrong topology (Kubatko & Degnan, 2007)
- MDC – **minimize deep coalescence**
  - Parsimony based methods
  - Do not take into account ILS
- **Multispecies coalescence models** (Heled and Drummond, 2010)
  - ML or Bayesian based searches
  - Accurate, powerful but not modeling hybridization
- just recently...
  - several models taking into account **ILS & hybridization**
  - ML or Bayesian based searches

# Species tree using multispecies coalescens

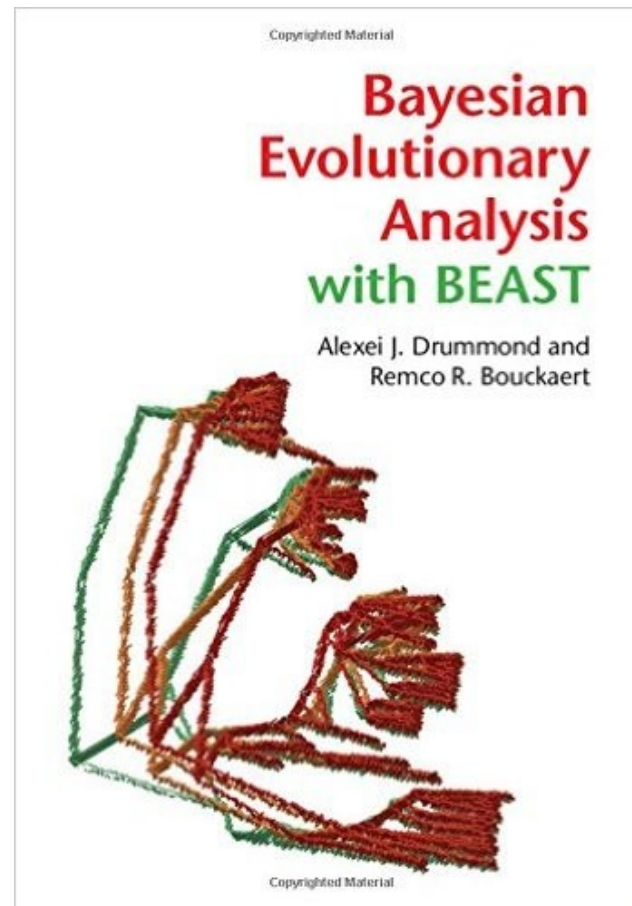
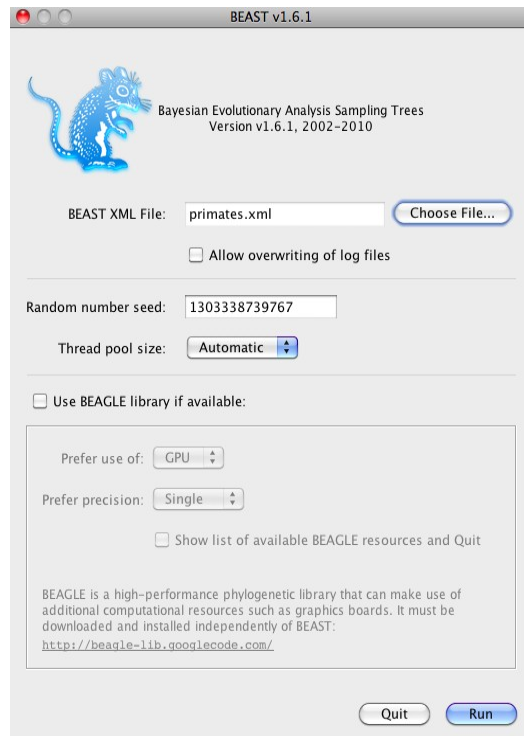
- \*BEAST (Heled and Drummond, 2010)

- Bayesian inference of the species tree model
- Do not allow for gene flow between studied species
- Based on alignments of multiple markers that can lead to different topology due to ILS



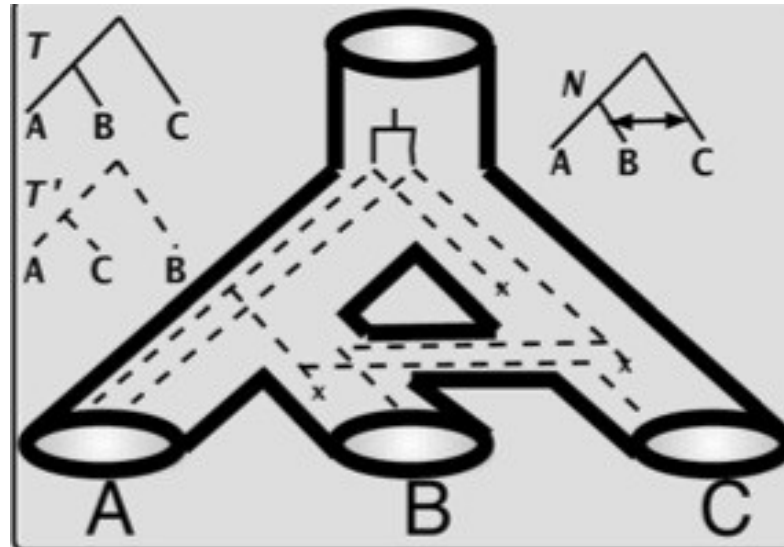
# BEAST software

- cross-platform program for Bayesian analysis of molecular sequences using MCMC
- \*BEAST is one of the modules
- valuable online support – google groups, tutorials  
<http://beast.bio.ed.ac.uk>
- Sister tool - BEAST2



# Species Networks

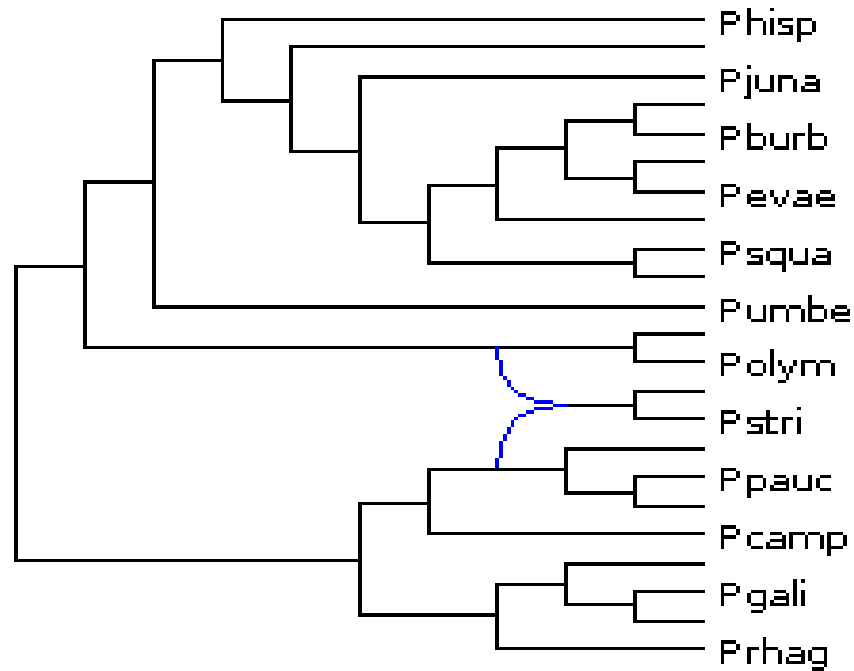
- Models allowing for **ILS and gene flow** – complex



- Only few models useful as an exploratory tool for multiple species
- PhyloNet software <https://bioinfo.cs.rice.edu/phyloNet>
  - Different tools to reconstruct networks
  - Exploratory species network reconstruction using Maximum pseudolikelihood approach
  - **Models user defined number of hybridization events**

# Species network visualization

- Dendroscope, <http://dendroscope.org/>



# Project

- 1) Evaluation of congruence of multiple sequence alignment
- 2) Identification of individuals that causing incongruence
- 3) Reconstruct of species tree
- 4) Reconstruction of species network

## AIMS:

- understand causes and consequences of differences between datasets
- choose the correct method for reconstruction of [species phylogeny](#)

....[FOLLOW THE MANUAL](#) and comes with question/comments to us!

**HAVE FUN!**