

ape & phangorn

Klaus Schliep

University of Massachusetts Boston

Evolution 2019, Providence, RI, June 21th 2016



A brief history:

- ▶ 1993: first release of R
- ▶ 1996: Ihaka & Gentleman, J. Comput. Graph. Statist., 5: 299–314
- ▶ 2000: R 1.0.0
- ▶ 2002: first release of ape by Emmanuel Paradis
- ▶ 2008: first release of phangorn
- ▶ June 2019: >200 packages depending on ape and 33 packages depend on phangorn on CRAN and hundreds more elsewhere: Bioconductor, GitHub, R-Forge, ...)

- ▶ I/O (Newick, NEXUS), manipulation of phylogenies
- ▶ Plot and annotate phylogenies
- ▶ Evolutionary distances from DNA sequences (Jukes–Cantor, Kimura, . . .)
- ▶ Distance-based phylogeny reconstruction (NJ, BIONJ, FastME, . . .)
- ▶ Generic bootstrap analyses (boot.phylo)
- ▶ A lot of tools for phylogenetic comparative methods (PCMs)
- ▶ Macroevolutionary analyses of phylogenies (γ -statistic, . . .)
- ▶ Sequence alignment with Clustal(- Ω), Muscle, T-Coffee

- ▶ Maximum likelihood phylogenetics with any kind of sequences (DNA, AA, codons, user-defined)
- ▶ Parsimony methods
- ▶ Evolutionary distances from AA sequences (WAG, Dayhoff, ...)
- ▶ Phylogenetic networks (Consensus networks, ...)
- ▶ Phylogenetic conflict/support (Lento plots)
- ▶ Distances between trees (Robinson-Foulds, SPR, Kuhner-Felsenstein, ...)

Acknowledgement

Thanks to Emmanuel Paradis for being such a great collaborator over the years!

Both packages have been possible without the contribution of a lot of people:

Simon Blomberg, Ben Bolker, Joseph Brown, Julien Claude, Hoa Sien Cuong, Richard Desper, Gilles Didier, Benoit Durand, Julien Dutheil, RJ Ewing, Olivier Gascuel, Thomas Guillerme, Christoph Heibl, Anthony Ives, Bradley Jones, Franz Krah, Daniel Lawson, Vincent Lefort, Pierre Legendre, Jim Lemon, Eric Marcon, Rosemary McCloskey, Johan Nylander, Rainer Opgen-Rhein, Andrei-Alin Popescu, Manuela Royer-Carenzi, Korbinian Strimmer, Damien de Vienne, Leonardo de Oliveira Martins, Alastair Potts, Tim W. White, Cyrill Stachniss, Michelle Kendall, Keren Halabi, Richel Bilderbeek, Liam Revell, Kristin Winchell & you?

Special thanks to all the volunteers at CRAN for their generous help and support.

And last but not least thanks to all the users who filed bug reports or even better bug fixes.



Splits, Trees & Networks

Motivation

- ▶ When analyzing phylogenetic data we usually expect the historical signal to match a tree.
- ▶ Often we have situations where we are confronted with many trees instead of a single tree.
- ▶ Trees from a bootstrap analysis or from MCMC run.
- ▶ There are many processes that can lead to conflicting signal:
 - ▶ trees with different historical signal, e.g. hybridization, recombination, incomplete lineage sorting, lateral gene transfer, etc.
 - ▶ trees with misleading signal (e.g. long branch attraction, compositional bias, changing patterns of variable sites).

Tools in *phangorn* and *ape*

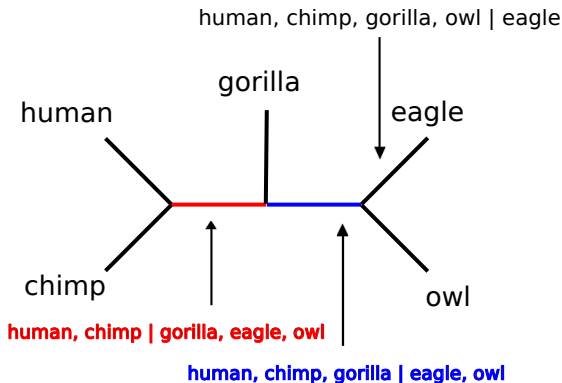
To display these trees or conflicting signal several tools are available:

- ▶ densiTree
- ▶ lento plots
- ▶ consensus trees and consensus networks
- ▶ neighborNet

The consensus tree method is implemented in *ape* all other functions in *phangorn*

Splits (Definition)

- ▶ A split is a bipartition of the taxa (labels, species) into two sets
- ▶ A bipartition of one taxa vs. the rest is known as a trivial split
- ▶ A split corresponds to an edge in a phylogenetic tree, removing the edge from a tree creates two sub-trees
- ▶ A tree only contains compatible splits

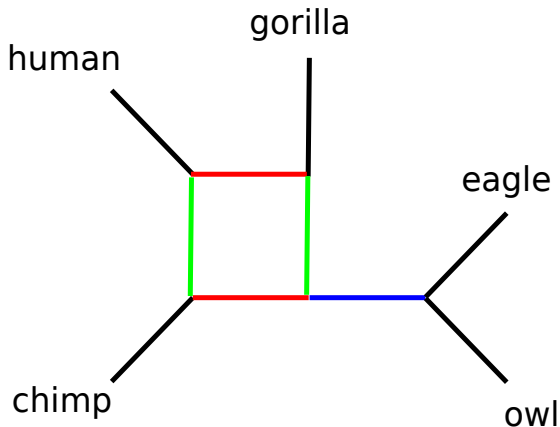


Incompatible Splits

Some collections of splits fit not on a tree but on a splits graph.

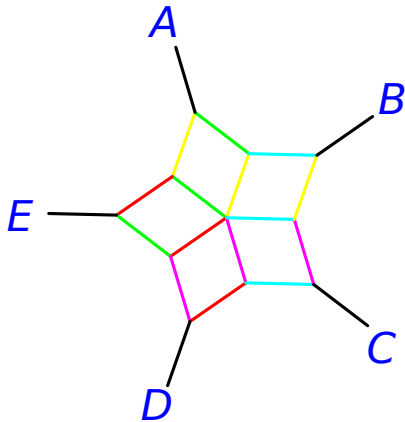
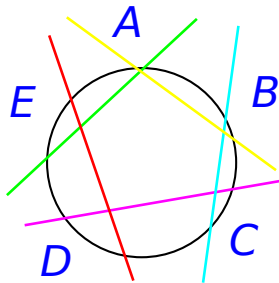
human, chimp | gorilla, eagle, owl

human, gorilla | chimp, eagle, owl



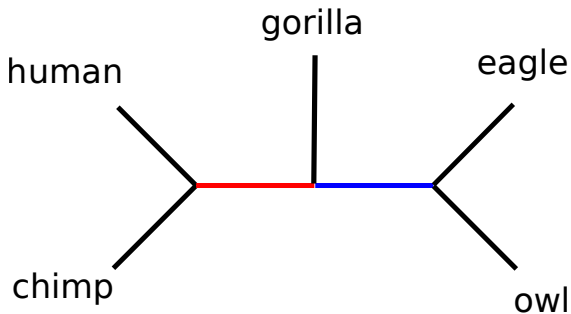
Circular Splits

Circular Splits can always displayed as a planar graph



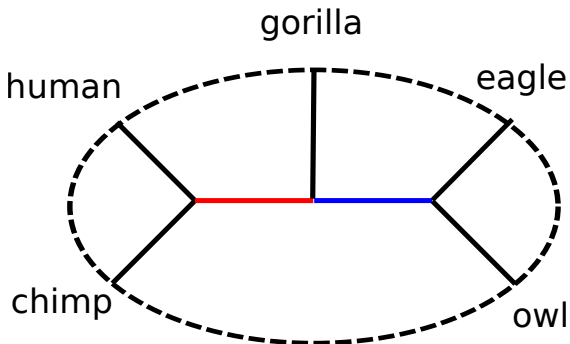
Trees are always circular

Trees only contain compatible splits and can always be displayed as a circular graph



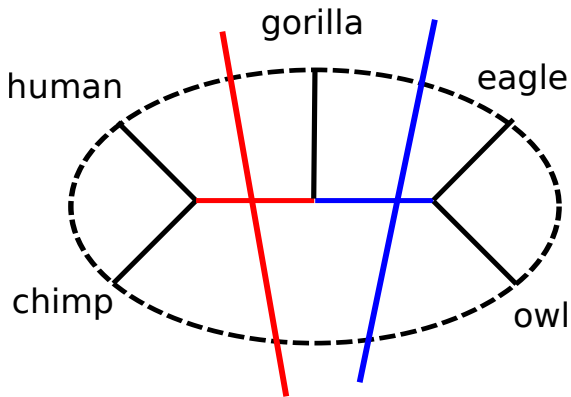
Trees are always circular

Trees only contain compatible splits and can always be displayed as a circular graph

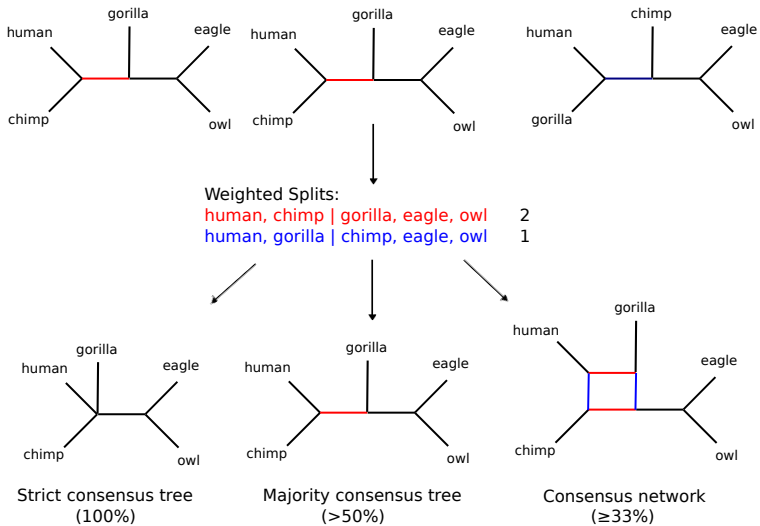


Trees are always circular

Trees only contain compatible splits and can always be displayed as a circular graph



Consensus Trees



Consensus Network

We can play with the threshold to adjust the number of edges we want to display

