

Homology and inference

Emily Jane McTavish

Life and Environmental Sciences
University of California, Merced

`ejmctavish@ucmerced.edu`, `twitter:snacktavish`

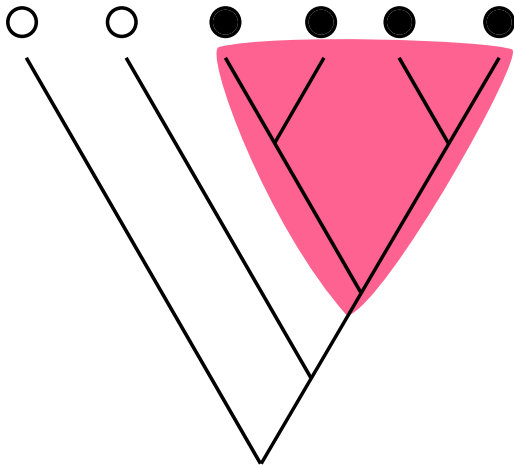
(With thanks to Mark Holder and Richard Edwards for slides)

The meaning of homology (**very roughly**):

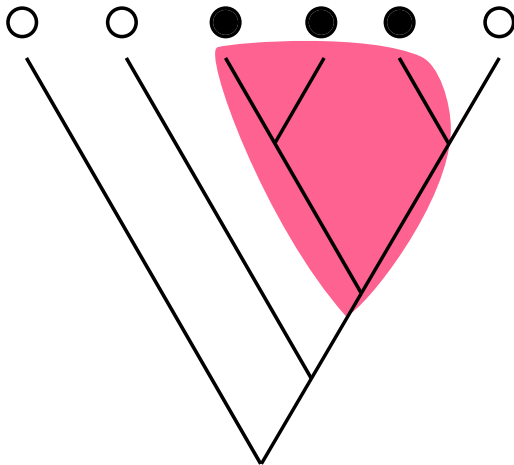
1. comparable (when applied to characters)
2. identical by descent (when applied to character states)

Ideally, each possible character state would arise once in the entire history of life on earth.

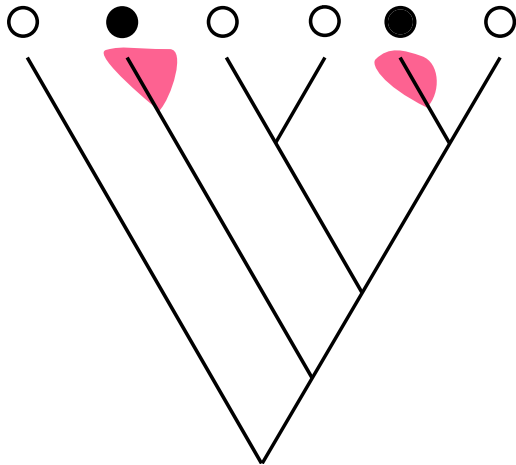
Instances of the filled character state are homologous
Instances of the hollow character state are homologous



Instances of the filled character state are homologous
Instances of the hollow character state are NOT homologous



Instances of the filled character state are NOT homologous
Instances of the hollow character state are homologous



Logical approach to phylogenetics

Premise: The following character matrix is correctly coded (character states are homologous in the strict sense):

	1
taxon A	Z
taxon B	Y
taxon C	Y

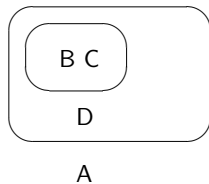
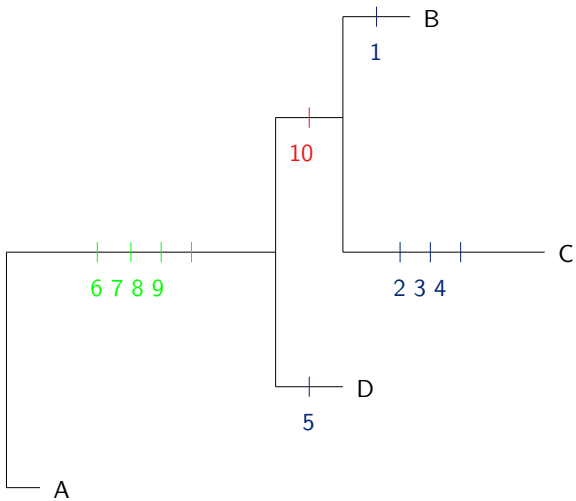
Is there a valid set of rules that will generate the tree as a conclusion?

Logical approach to phylogenetics (cont)

Rule: Two taxa that share a character state must be more closely related to each other than either is to a taxon that displays a different state.

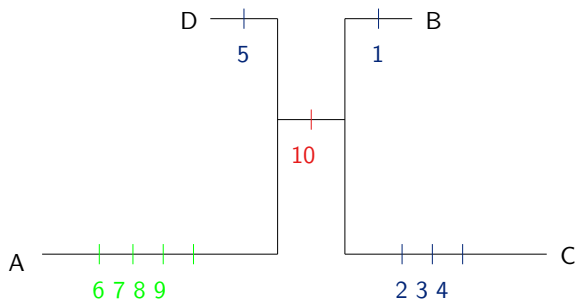
Is this a valid rule?

Taxon	Character #									
	1	2	3	4	5	6	7	8	9	10
A	0	0	0	0	0	0	0	0	0	0
B	1	0	0	0	0	1	1	1	1	1
C	0	1	1	1	0	1	1	1	1	1
D	0	0	0	0	1	1	1	1	1	0



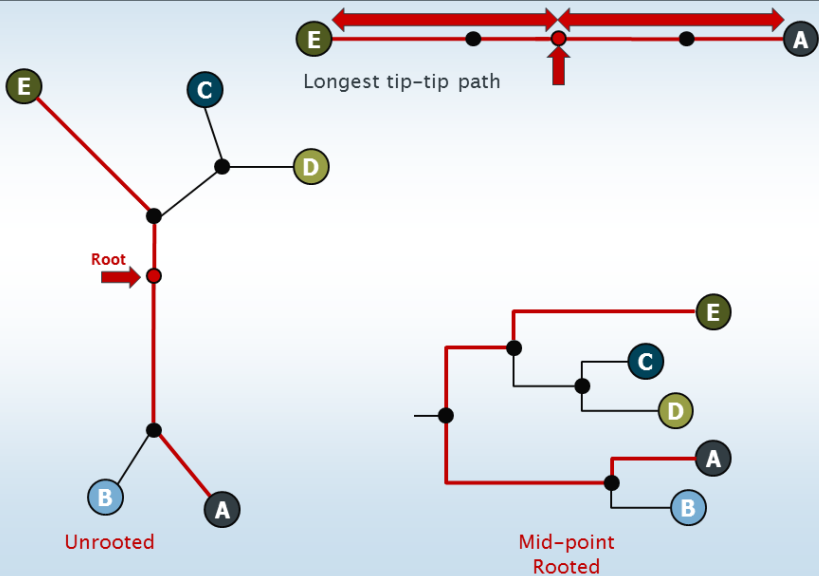
Interestingly, without polarization Hennig's method can infer unrooted trees. We can get the tree topology, but be unable to tell paraphyletic from monophyletic groups.

The outgroup method amounts to inferring an unrooted tree and then rooting the tree on the branch that leads to an outgroup.

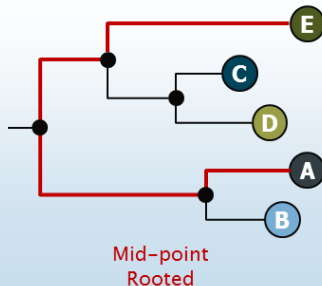
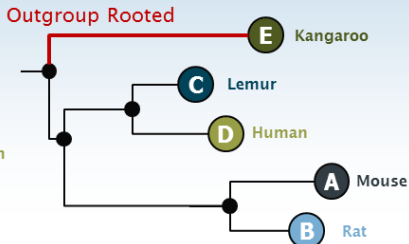
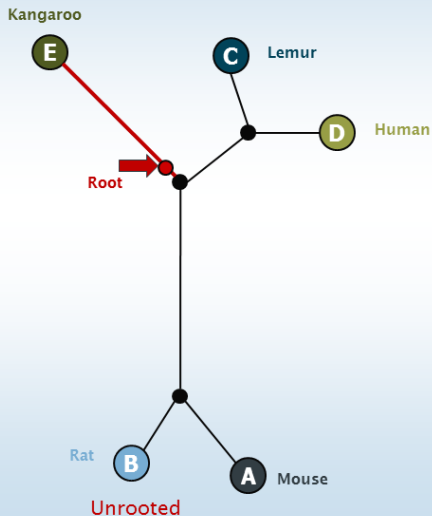


B	C
A	D

Mid-point Rooting



Outgroup Rooting



a brief digression into newick tree file format



Newick's Lobster House was the site of an historic 1986 meeting at which a standard was devised for storing descriptions of phylogenetic trees as strings. (Photo from Paul Lewis)

Note: $((1,2),3,4)$ is referred to as Newick or New Hampshire notation for the tree.

You can read it by following the rules:

- start at a node,
- if the next symbol is '(' then add a child to the current node and move to this child,
- if the next symbol is a label, then label the node that you are at,
- if the next symbol is a comma, then move back to the current node's parent and add another child,
- if the next symbol is a ')', then move back to the current node's parent.

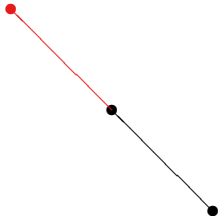
$((1,2),3,4)$



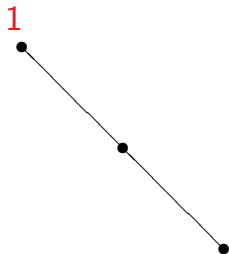
$((1,2),3,4)$



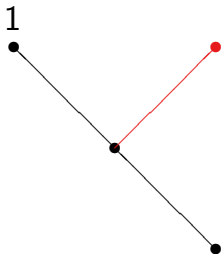
$((1,2),3,4)$



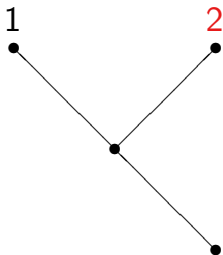
$((1,2),3,4)$



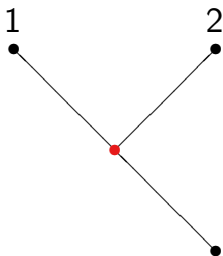
$((1,2),3,4)$



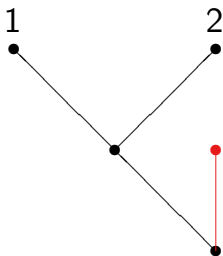
$((1,2),3,4)$



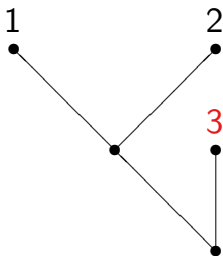
$((1,2),3,4)$



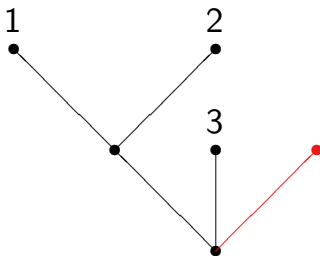
$((1,2),3,4)$



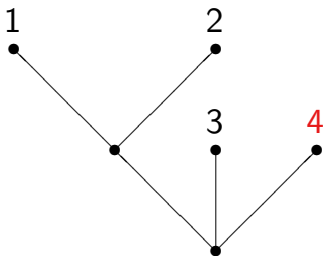
$((1,2),\textcolor{red}{3},4)$



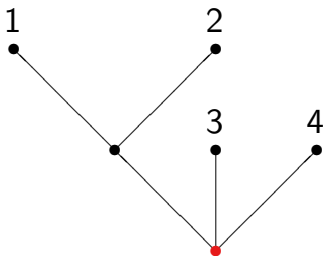
$((1,2),3,4)$



$((1,2),3,4)$



$((1,2),3,4)$



Newick

- ▶ Parenthetical tree format
- ▶ Rooted vs. unrooted trees are not differentiated
- ▶ Some programs interpret polytomy at root as 'unrooted'
- ▶ Branches and nodes not well differentiated
- ▶ A name can contain any characters except blanks, colons, semicolons, parentheses, and square brackets

Exercise

Create a newick tree file in your text editor with the content:

```
(( (C, (D,E)), (F,G), A), B);
```

Save it as 'example.tre'.

- ▶ Draw the tree by hand
- ▶ Write down all the splits in `..**` format.
- ▶ Re-root the tree. What rootings make the following true? Which cannot be true?
 - ▶ A is more closely related to G than it is to C
 - ▶ (C,D,E) is sister to (A,B,F,G)
 - ▶ (C,D) is sister to (A,B,E,F,G)
 - ▶ (C,D,E) is a paraphyletic group
 - ▶ (C,D,E) is a monophyletic group
 - ▶ (A,B,C) is a monophyletic group