

A quick review

Global alignment algorithm:

Needleman-Wunsch.

		G	A	A	T	C	
		0	-4	-8	-12	-16	-20
C		-4	-5	-9	-13	-12	-6
A		-8	-4	5	1	-3	-7
T		-12	-8	1	0	11	7
A		-16	-12	2	11	7	6
C		-20	-16	-2	7	11	17

Local alignment algorithm:

Smith-Waterman.

		A	A	G
		0	0	0
G		0	0	2
A		0	2	0
A		0	2	4
G		0	0	6
G		0	0	2
C		0	0	0

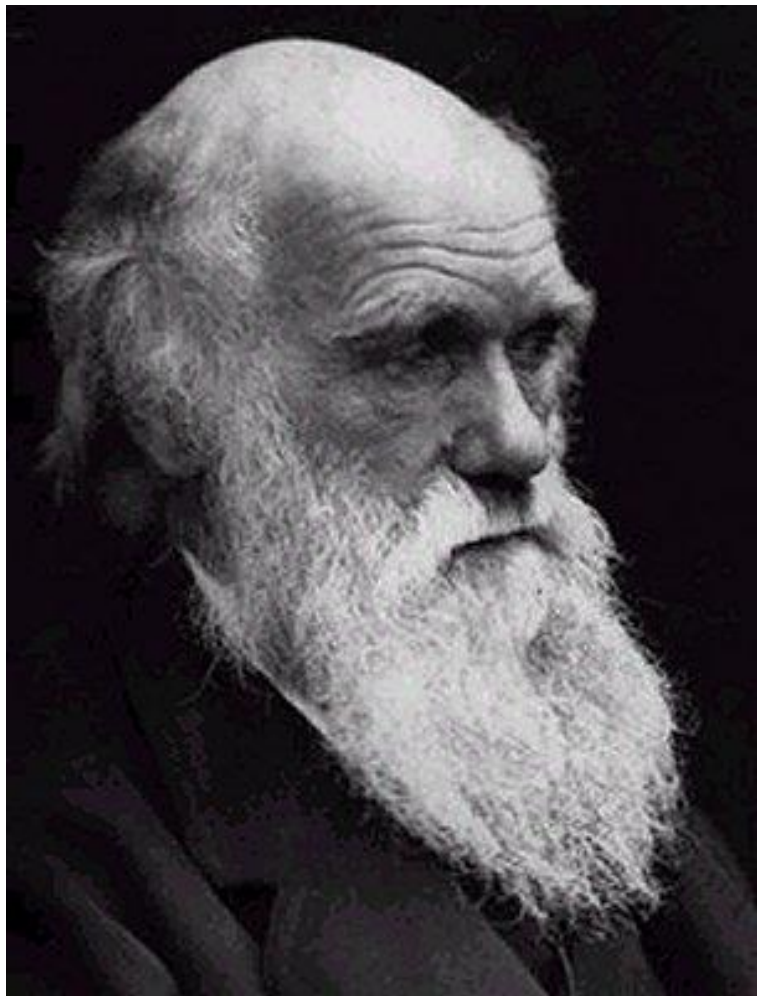
- Significance of similarity scores (P-values)
 - Empirical null score distribution
 - Extreme value distribution
- Multiple-testing correction (Bonferroni) and E-values

Phylogenetic Trees

Genome 373

Genomic Informatics

Elhanan Borenstein



36

I think

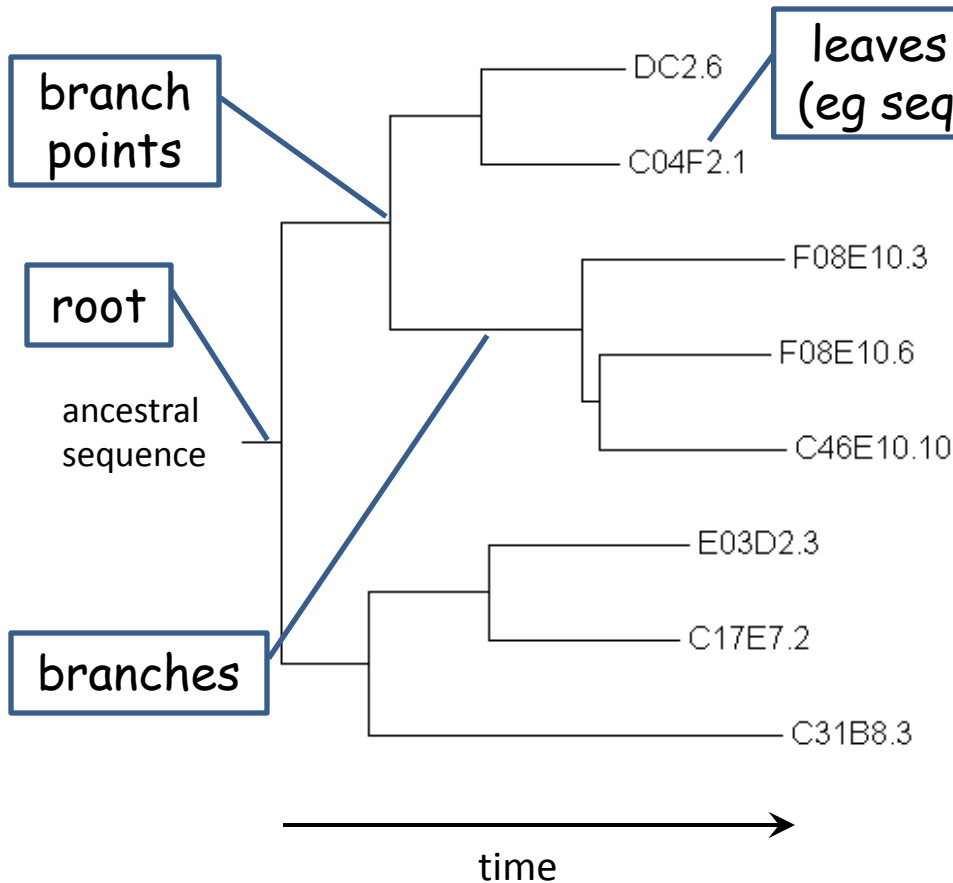
I do not think E. has any species in common with the rest.

Some may be shared in some cases than in others living in same place.

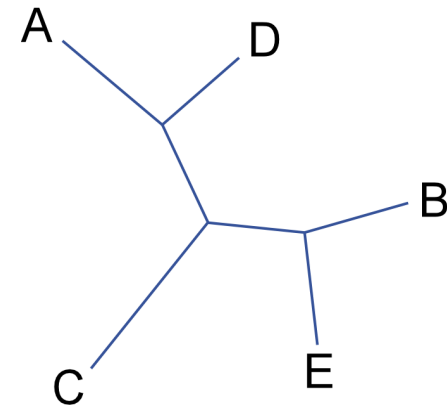
There between A & B. various sort of relation. C + B. The finest gradation, B & D rather greater distinction than genus would be formed. - bearing relation

Defining what a “tree” means

rooted tree (all real trees are rooted):



unrooted tree:
(used when the root isn't known):

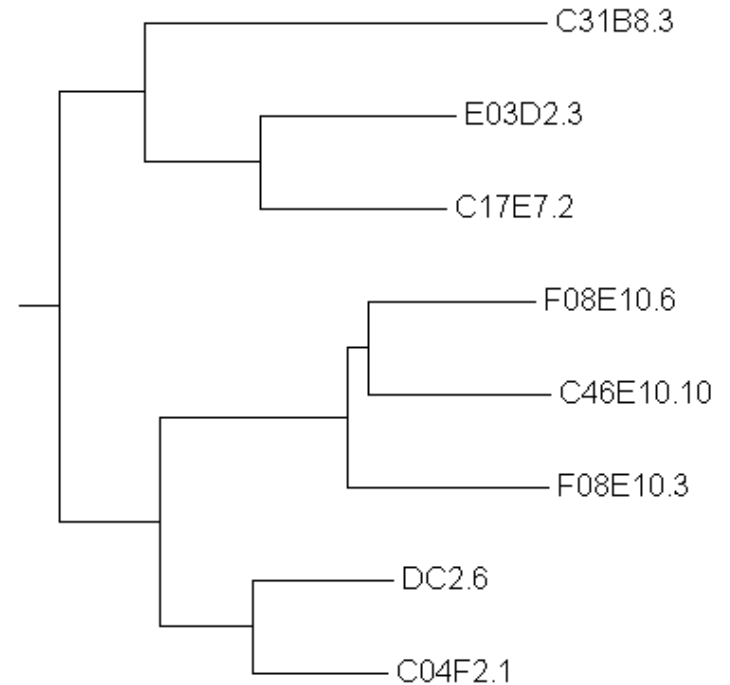
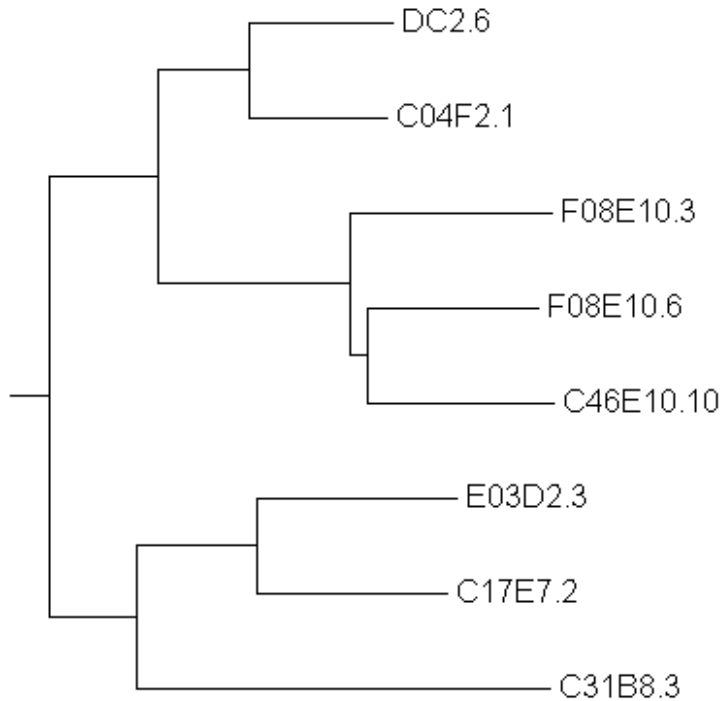


time radiates out from somewhere (probably near the center)

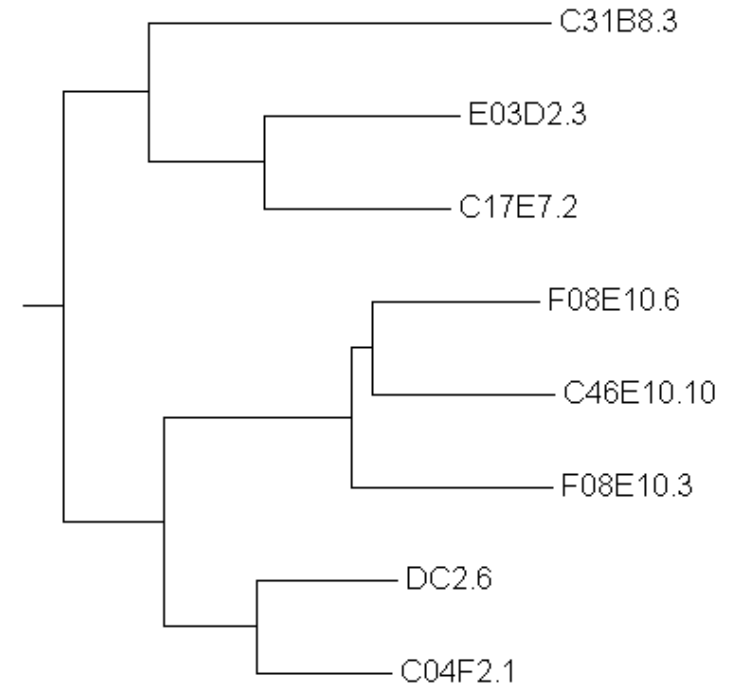
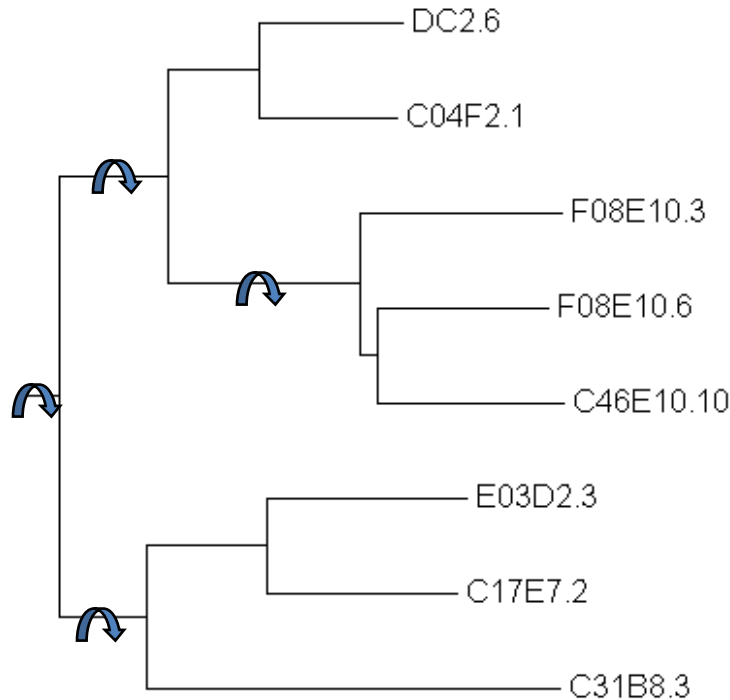
... sequence divergence is proportional to (horizontal) branch lengths

Note: A tree has topology and distances

Are these topologically different trees?



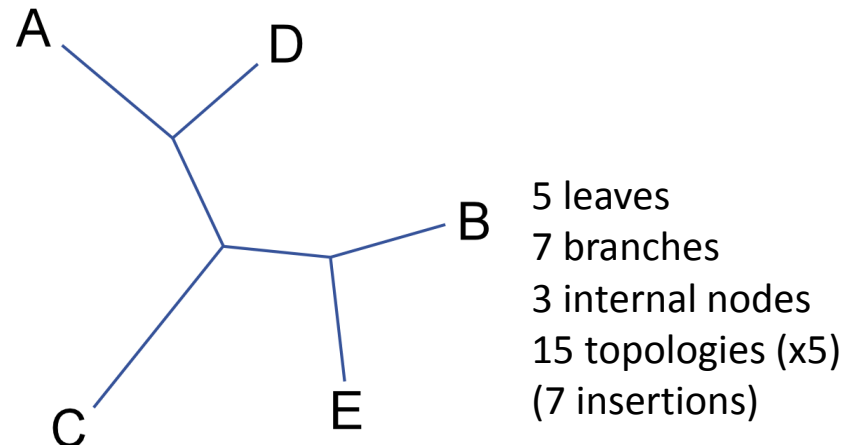
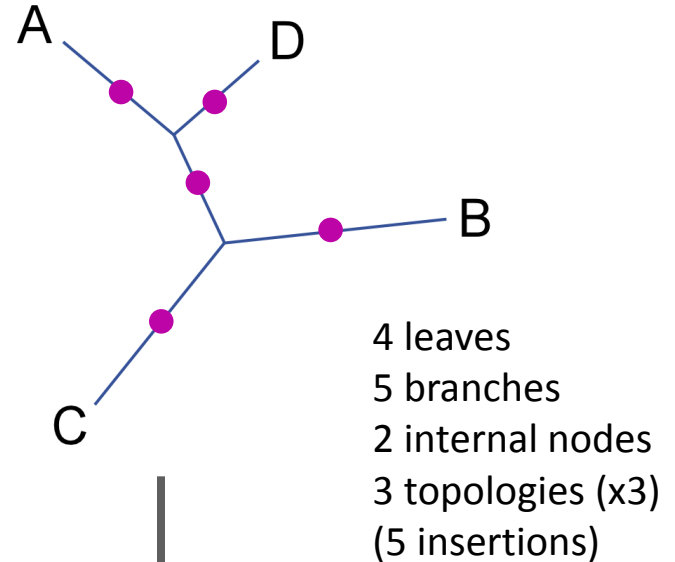
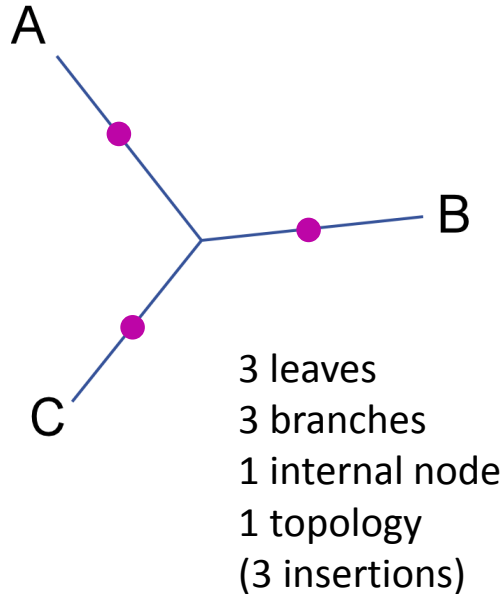
Are these topologically different trees?



Topologically, these are the *SAME* tree. In general, two trees are the same if they can be inter-converted by branch rotations.

Why is inferring phylogeny
a hard problem?

The number of tree topologies grows extremely fast



In general, an unrooted tree with **N** leaves has:

$2N - 3$ total branches

N leaf branches

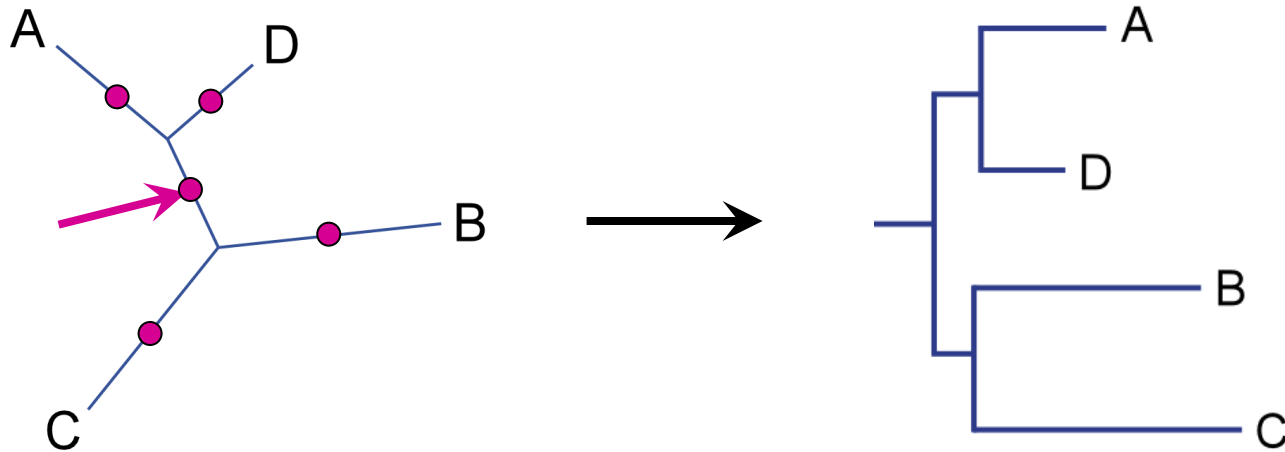
$N - 3$ internal branches

$N - 2$ internal nodes

$3 \cdot 5 \cdot 7 \cdot \dots \cdot (2N - 5) \sim O(N!)$ topologies

There are many rooted trees for each unrooted tree

For each unrooted tree, there are $2N - 3$ times as many rooted trees, where N is the number of leaves ($\#$ branches = $2N - 3$).



20 leaves - 564,480,989,588,730,591,336,960,000,000 topologies

How can you compute a tree?

- Many methods available, we will talk about:
 - Distance trees
 - Parsimony trees
- Others include:
 - Maximum-likelihood trees
 - Bayesian trees

