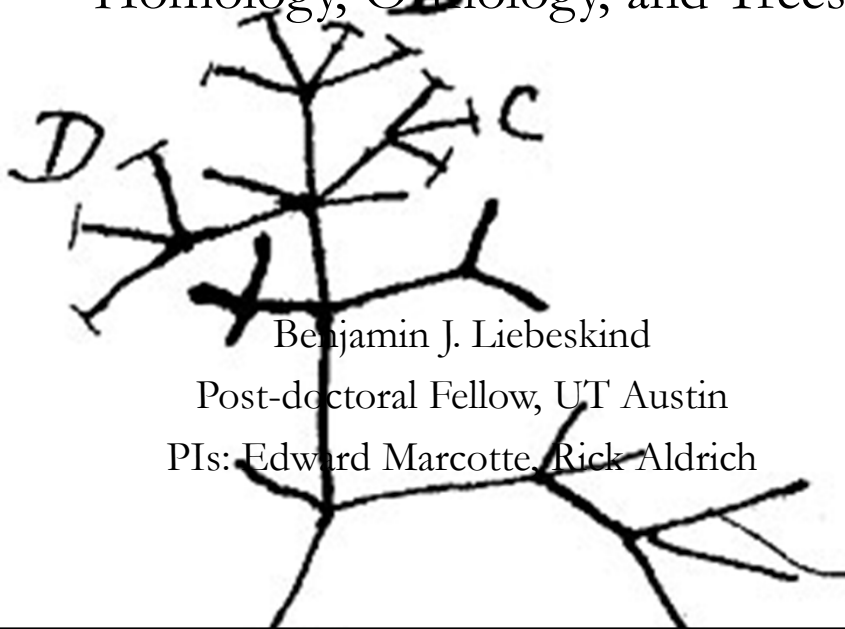


Homology, Orthology, and Trees

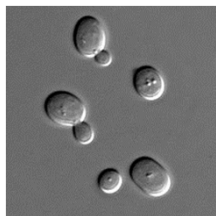


Benjamin J. Liebeskind

Post-doctoral Fellow, UT Austin

PIs: Edward Marcotte, Rick Aldrich

How do we compare proteins across species?



Gravitropism defects

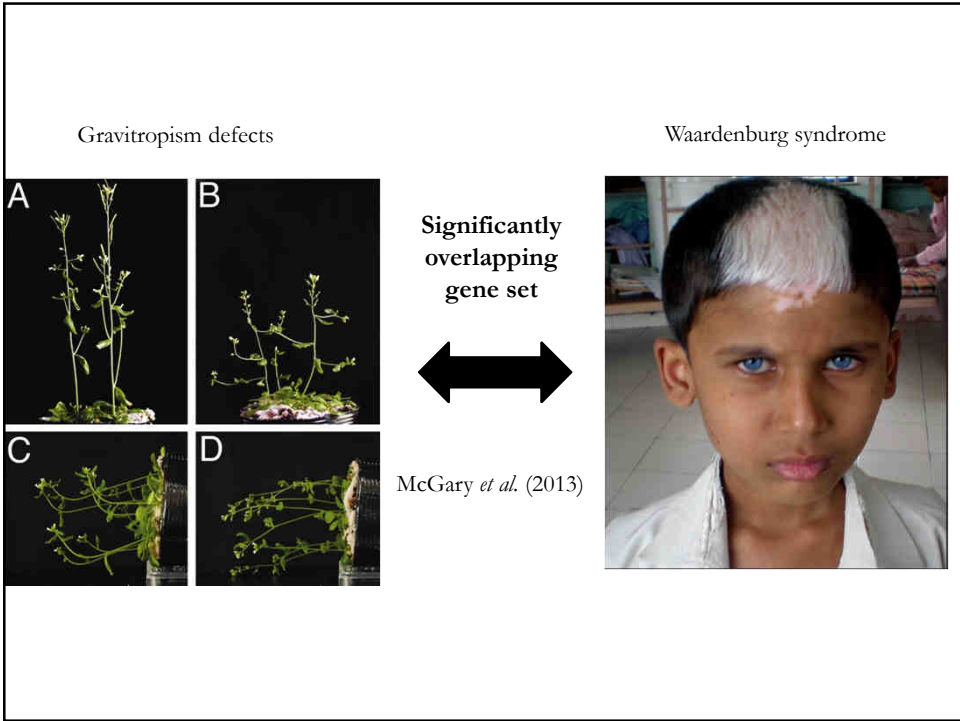
Waardenburg syndrome

A **B**

C **D**

Significantly overlapping gene set

McGary *et al.* (2013)



Homology



“Homologue...The same organ in different animals under every variety of form and function...Analogue...A part or organ in one animal which has the same function as another part or organ in a different animal”

- Richard Owen (1843)

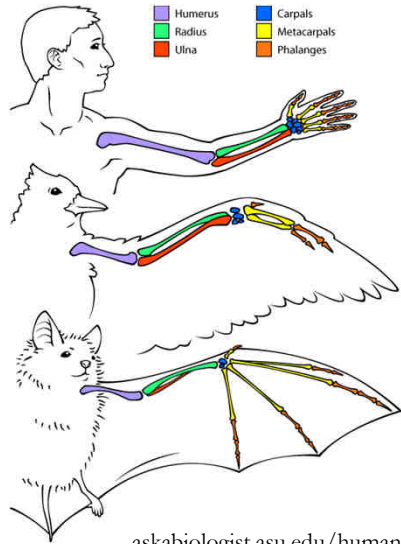
Homology

- Since Darwin:
 - Organs (or genes, or...) in two or more species that are similar due to common descent
 - I.e. they are descended from a similar organ in their most recent common ancestor

Homology

- Since Darwin:
 - Organs (or genes, or...) in two or more species that are similar due to common descent
 - I.e. they are descended from a similar organ in their most recent common ancestor
- **Note**
 - This means homology is binary
 - No such thing as % homology

Homology

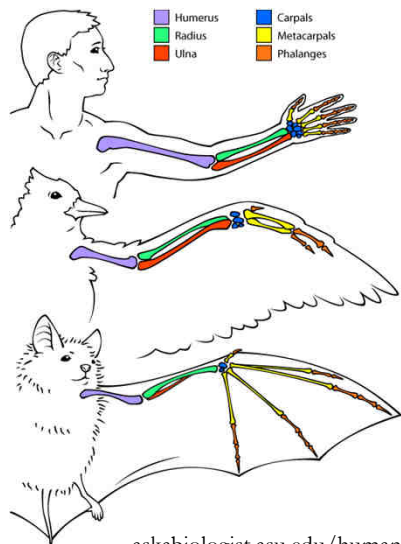


This is trickier than it sounds!!

Organs are not monolithic entities!!

askabiologist.asu.edu/human-bird-and-bat-bone-comparison

Homology



This is trickier than it sounds!!

Organs are not monolithic entities!!

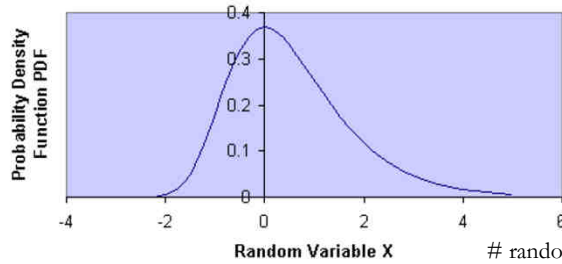
Bat wings and bird wings are homologous
as vertebrate forelimbs

But they are analogous as wings

askabiologist.asu.edu/human-bird-and-bat-bone-comparison

Homology of Genes

Alignment scores follow an *extreme value distribution*



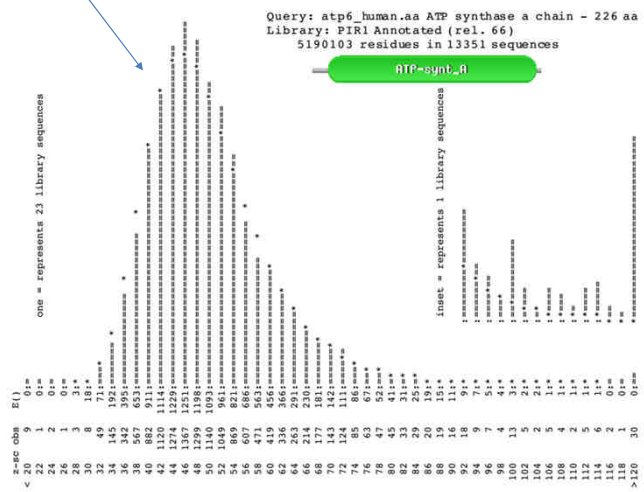
random trials & their average score

$$p(\text{max score} \leq X) \approx e^{-kNe^{\lambda(X-\mu)}}$$

Describe the shape & can be fit from a few trials

Alignment scores between *real* non-homologous sequences are indistinguishable from random

Real homologs have non-random scores



Slide courtesy of William Pearson

Homology

- Homologous gene families are very large
- How do we find one-to-one correspondence across species?

Human cycophilin family

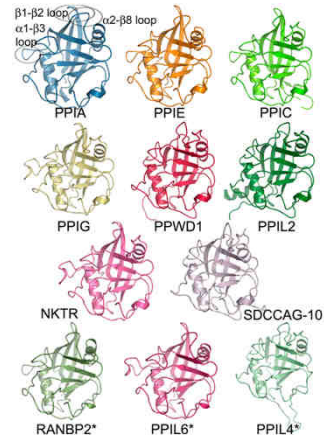


Fig. 2 in Davis et al., PLoS Biol 8(7): e1000439, 2010

Orthology

Walter Fitch



1929 - 2011

What is an Ortholog?

- Orthology is used nearly universally to compare *single* genes across species
- “Same gene in a different species”

What is an Ortholog?

- Orthology is used nearly universally to compare *single* genes across species
- “Same gene in a different species”

Terminology Barf

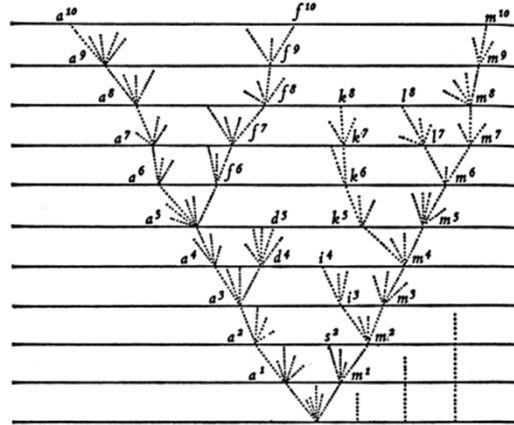
- Paralogs
- Inparalogs
- Outparalogs
- Xenolog
- Ohnologs
- Gametologs

What is an Ortholog?

- All these terms are meant to approximate the tree-like structure of gene evolution
- First an aside on trees...

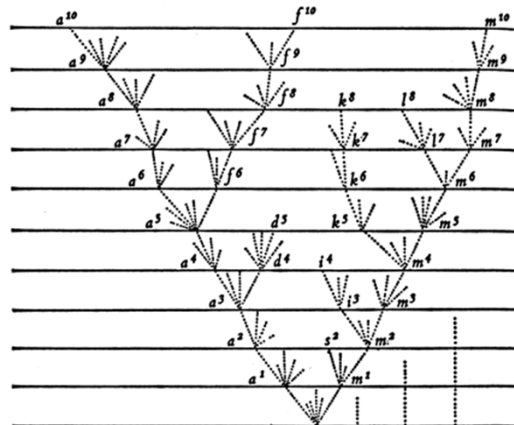
Basics of Phylogenetics

- Only figure in Darwin's "Origin of Species"

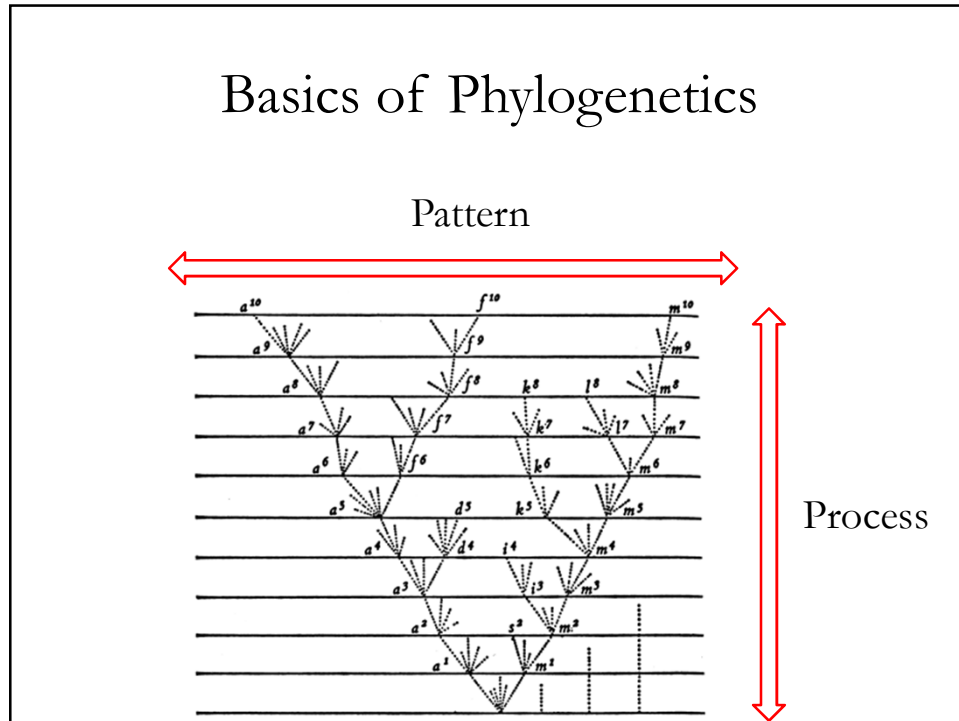


Basics of Phylogenetics

- Trees show the relationship between pattern and process



Basics of Phylogenetics



Basics of Phylogenetics

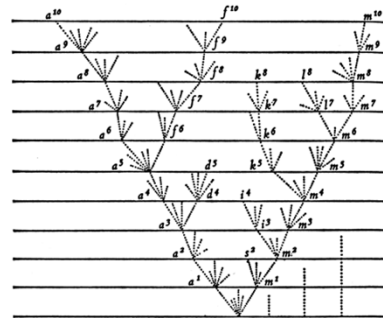
- Phylogenetic systematics (cladistics)
 - Organisms should be grouped by phylogenetic relationships
 - **Bifurcating trees**
- Key terms:
 - Clade
 - Monophyly
 - Paraphyly



Willi Hennig (1913 – 1976)

Basics of Phylogenetics

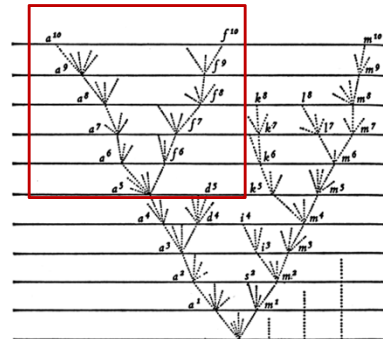
- Clade: an ancestor and all of its descendants



Basics of Phylogenetics

- Clade: an ancestor and all of its descendants

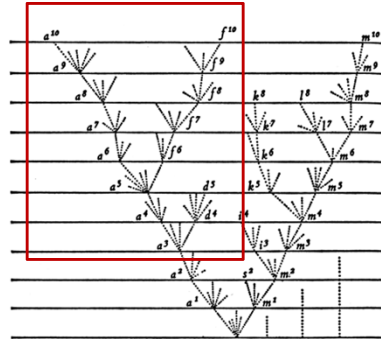
Clade!!



Basics of Phylogenetics

- Clade: an ancestor and all of its descendants

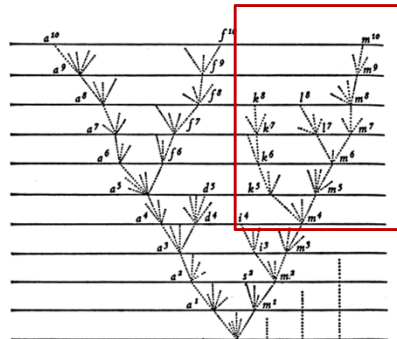
Clade!!



Basics of Phylogenetics

- Clade: an ancestor and all of its descendants

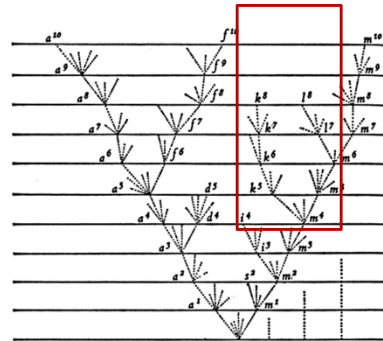
Clade!!



Basics of Phylogenetics

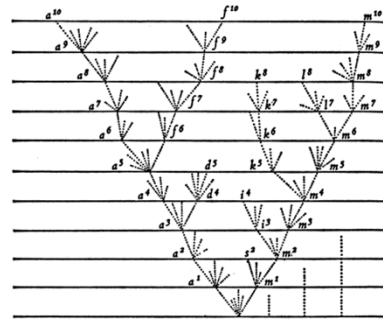
- Clade: an ancestor and all of its descendants

Not A Clade!!



Basics of Phylogenetics

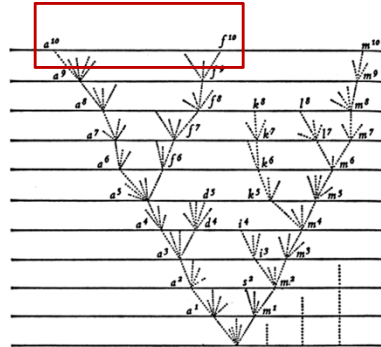
- Monophyletic group: organisms in a clade



Basics of Phylogenetics

- Monophyletic group: organisms in a clade

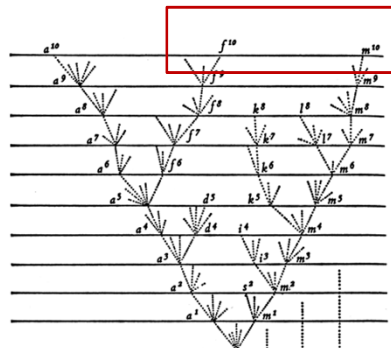
Monophyletic group!!



Basics of Phylogenetics

- Monophyletic group: organisms in a clade

Not a monophyletic group!!

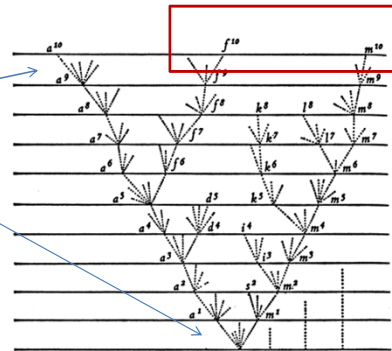


Basics of Phylogenetics

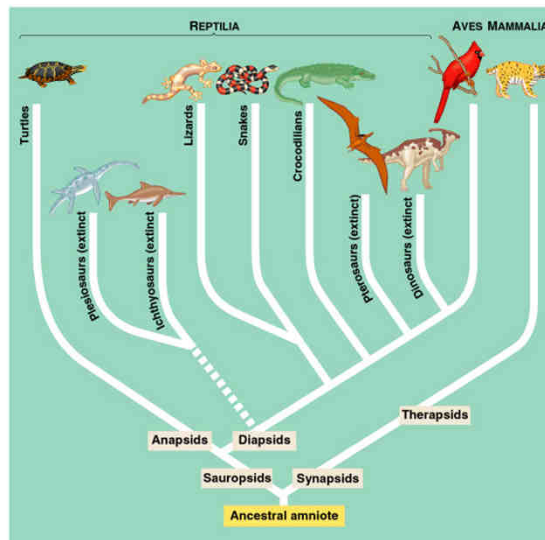
- Monophyletic group: organisms in a clade

- A group is *not* monophyletic if their most recent common ancestor has descendants that are not in the group

Not a monophyletic group!!



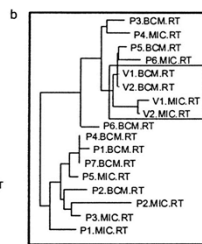
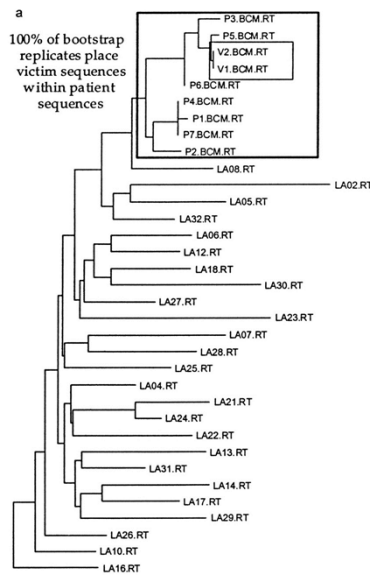
Basics of Phylogenetics



© 1999 Addison Wesley Longman, Inc.

Basics of Phylogenetics

- A doctor's girlfriend accuses him of injecting her with HIV. He said it was vitamin B12.
 - Who's right?
- Phylogenetics to the rescue!!



V: victim
 P: patient
 LA: Louisiana residents with HIV

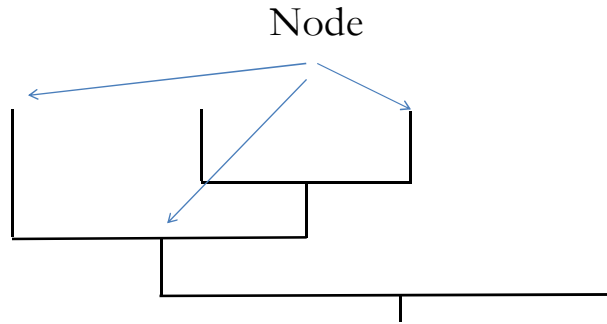
Michael L. Metzker et al. PNAS 2002;99:14292-14297



David Hillis

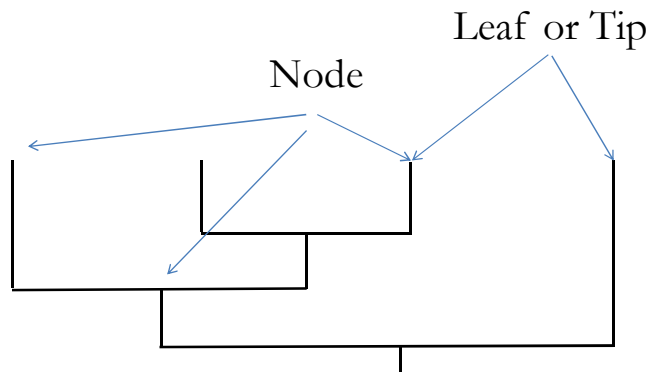
Basics of Phylogenetics

- Tree nomenclature: nodes



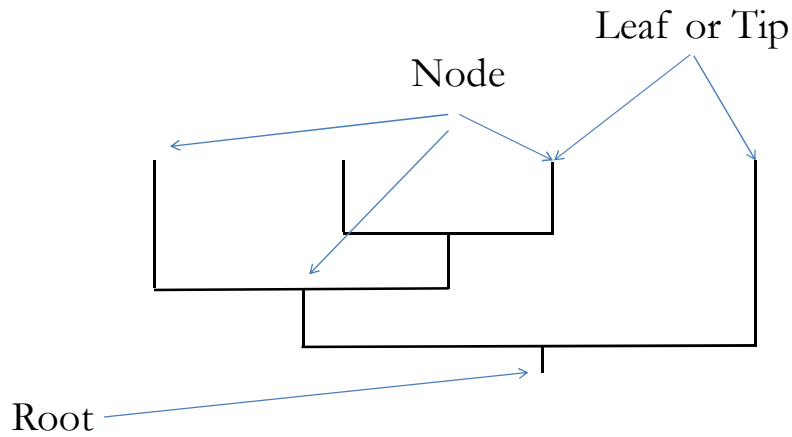
Basics of Phylogenetics

- Tree nomenclature: nodes



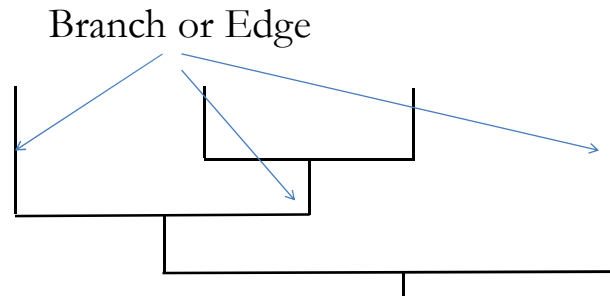
Basics of Phylogenetics

- Tree nomenclature: nodes



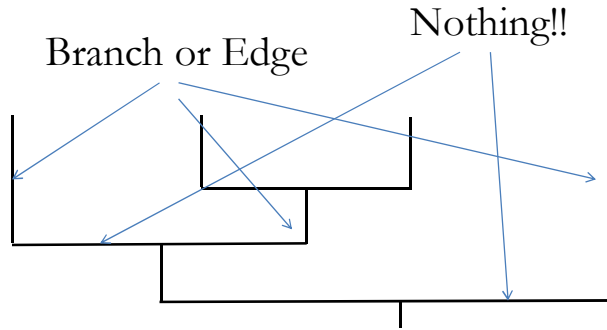
Basics of Phylogenetics

- Tree nomenclature: branches
 - Measures of evolutionary *rate*



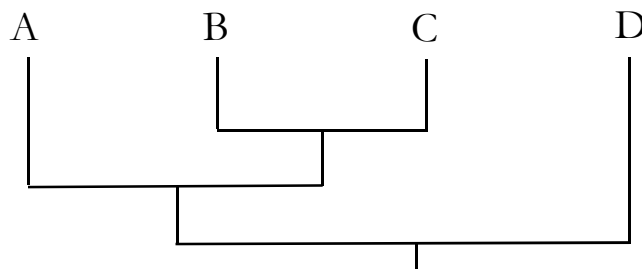
Basics of Phylogenetics

- Tree nomenclature: branches
 - Measures of evolutionary *rate*



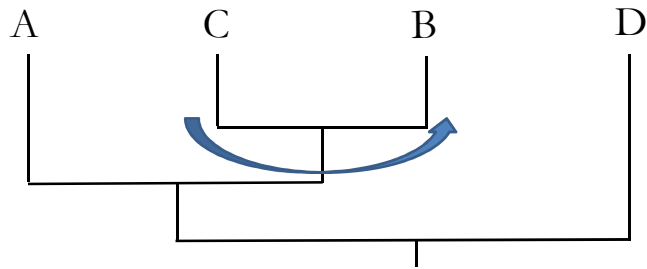
Basics of Phylogenetics

- Topology



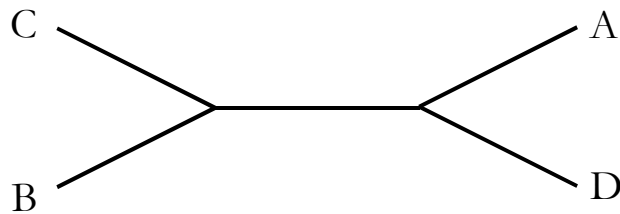
Basics of Phylogenetics

- Any node can be rotated without changing topology



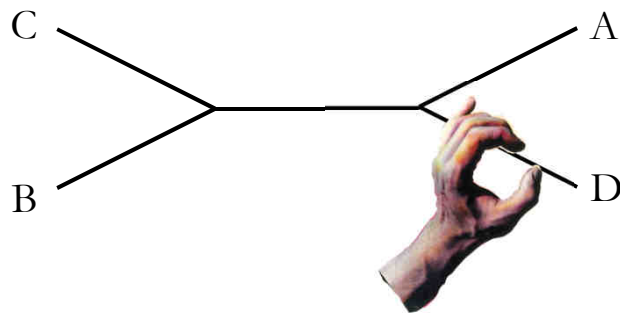
Basics of Phylogenetics

- The tree can be unrooted without changing topology



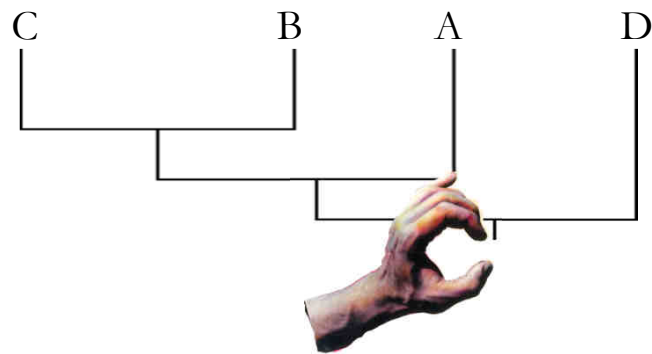
Basics of Phylogenetics

- The tree can be unrooted without changing topology



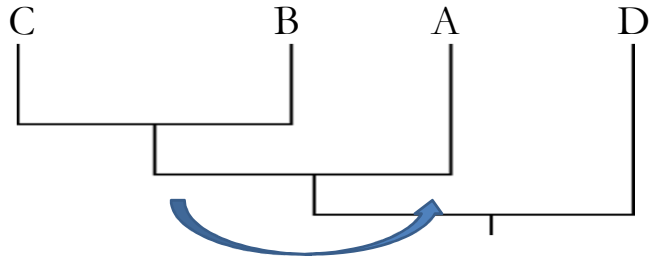
Basics of Phylogenetics

- The tree can be unrooted without changing topology



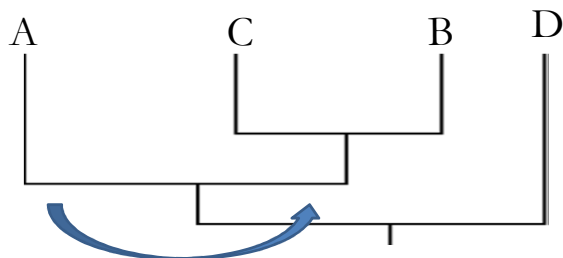
Basics of Phylogenetics

- The tree can be unrooted without changing topology



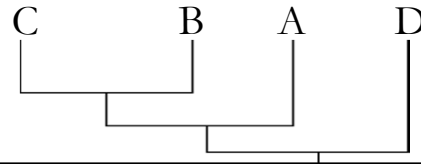
Basics of Phylogenetics

- The tree can be unrooted without changing topology



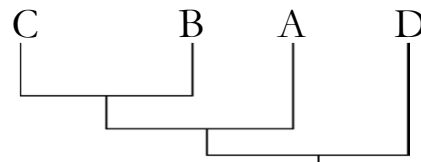
Basics of Phylogenetics

- Interpreting trees
 - Trees tell us the relative relatedness of leaf nodes



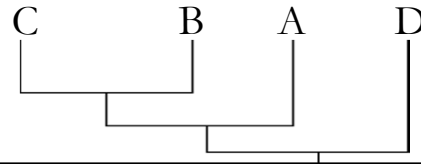
Basics of Phylogenetics

- Interpreting trees
 - Trees tell us the relative relatedness of leaf nodes
- Common misperceptions:
 - D is not the “ancestor” of any other leaf and is not necessarily an older lineage

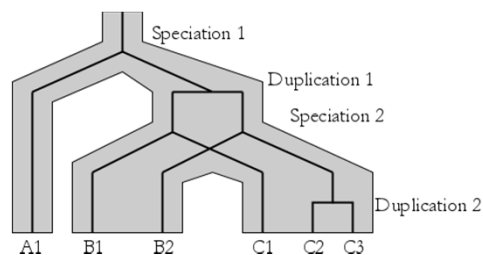


Basics of Phylogenetics

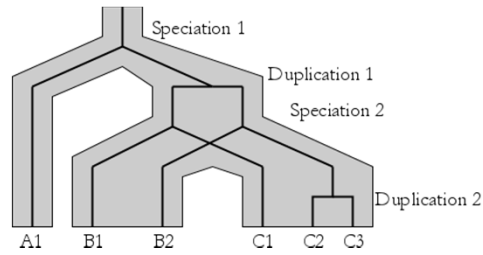
- Interpreting trees
 - Trees tell us the relative relatedness of leaf nodes
- Common misperceptions:
 - D is not the “ancestor” of any other leaf and is not necessarily an older lineage
 - The tree does not tell us that these tips are “related” (all organisms are related)



What is an Ortholog?

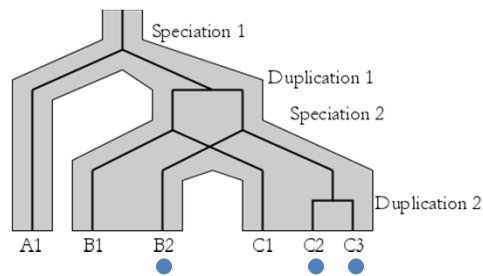


What is an Ortholog?



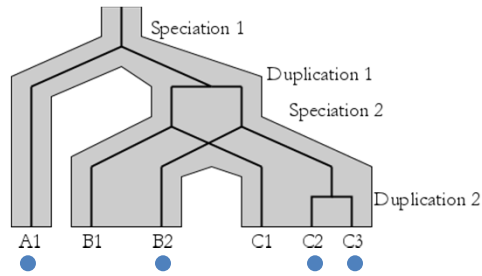
“Two genes whose common ancestor resides at a Y junction (speciation) are orthologous. Two genes whose common ancestor resides at a horizontal bar junction (gene duplications) are paralogous.” – Walter Fitch (2000) *Trends in Genetics*

What is an Ortholog?



“Two genes whose common ancestor resides at a Y junction (speciation) are orthologous. Two genes whose common ancestor resides at a horizontal bar junction (gene duplications) are paralogous.” – Walter Fitch (2000) *Trends in Genetics*

What is an Ortholog?



“Two genes whose common ancestor resides at a Y junction (speciation) are orthologous. Two genes whose common ancestor resides at a horizontal bar junction (gene duplications) are paralogous.” – Walter Fitch (2000) *Trends in Genetics*

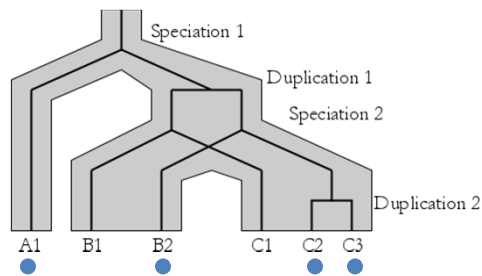
What is an Ortholog?

Deceptive simplicity:

Orthology is a pairwise relationship

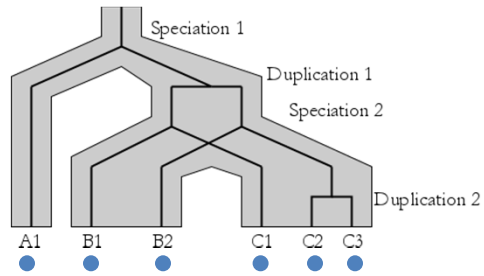
It is not transitive

Co-orthology is often abstracted from, with confusing results

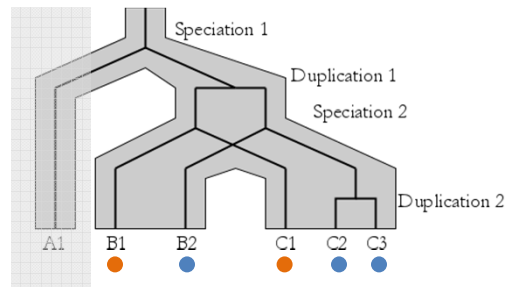


“Two genes whose common ancestor resides at a Y junction (speciation) are orthologous. Two genes whose common ancestor resides at a horizontal bar junction (gene duplications) are paralogous.” – Walter Fitch (2000) *Trends in Genetics*

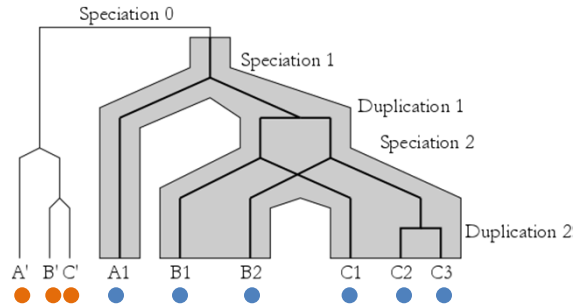
What is an Orthologous Group?



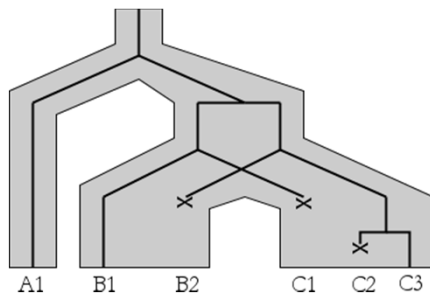
What is an Orthologous Group?



What is an Orthologous Group?



Uh Oh



-ologies

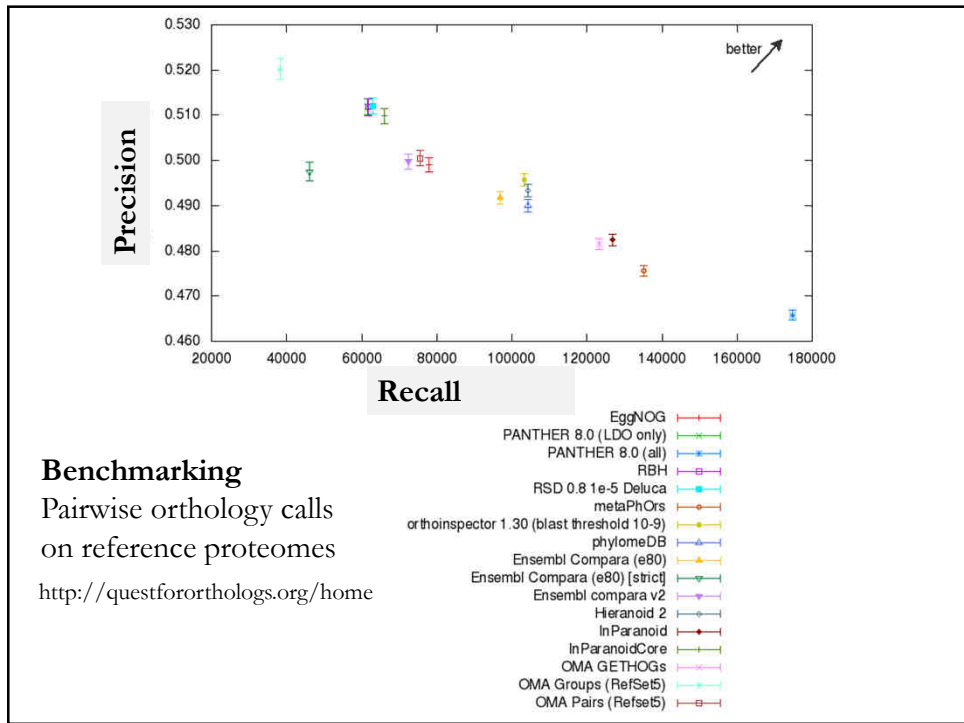
- Homology and Orthology give us a way to compare genes in different species
 - Homology contains whole families
 - Orthology is a direct comparison, usually denoting more functional similarity
- Both concepts are potentially slippery

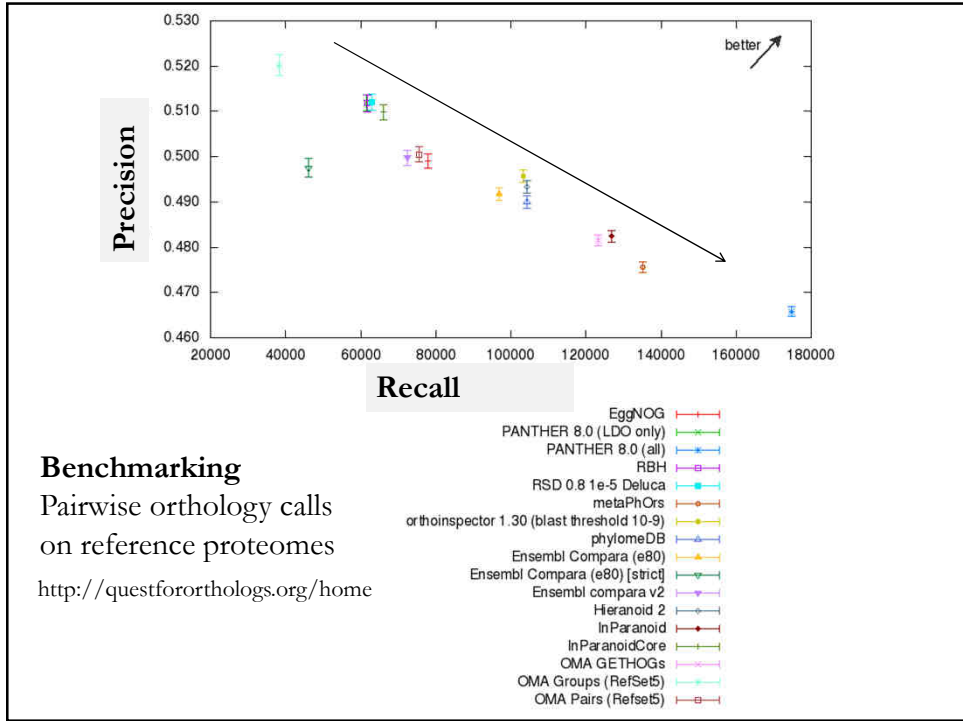
-ologies

- They are slippery because they try to “flatten” the true tree-like structure
- Orthology inference algorithms are fast
- Are they accurate?

-ologies

- They are slow to “flatten” the true tree
- Orthology inference is fast
- Are they accurate to infer true orthologs





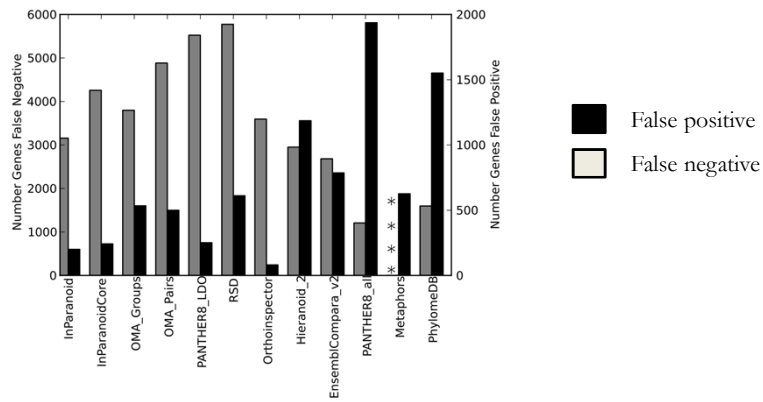
What's the way forward?

What's the way forward?

- Error propagation, handling uncertainty

What's the way forward?

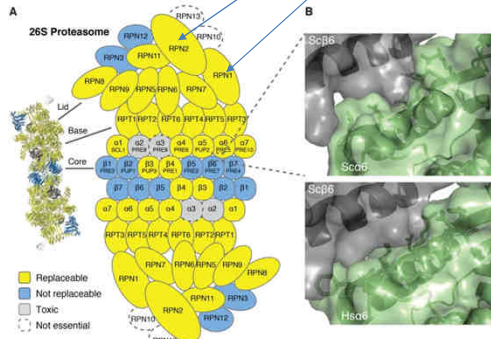
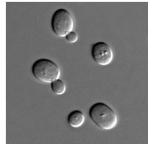
- Error propagation, handling uncertainty



Liebeskind, McWhite, Marcotte (2016)

And yet...

We can do amazing things with orthology, even with all the associated error



Kachroo, *et al.* (2015)

Thanks!