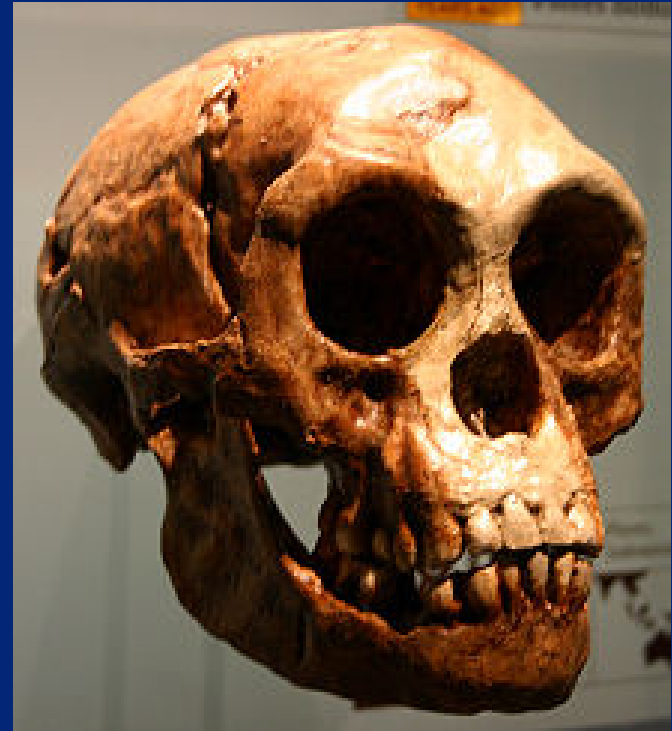


Outline

- Finishing up speciation
- Phylogenies—trees of species
 - What are phylogenies good for?
 - How to look at a phylogeny
 - Rooted versus unrooted trees
 - Clocklike versus non-clocklike trees
 - Appropriate data for phylogenies

Homo floresiensis

- One skull and 9 partial skeletons
- Around 3'6" tall and 55 lbs
- Brain of type specimen smaller than that of a chimpanzee
- Last dates initially thought to be 13,000 years ago, but now closer to 50,000



Pro species

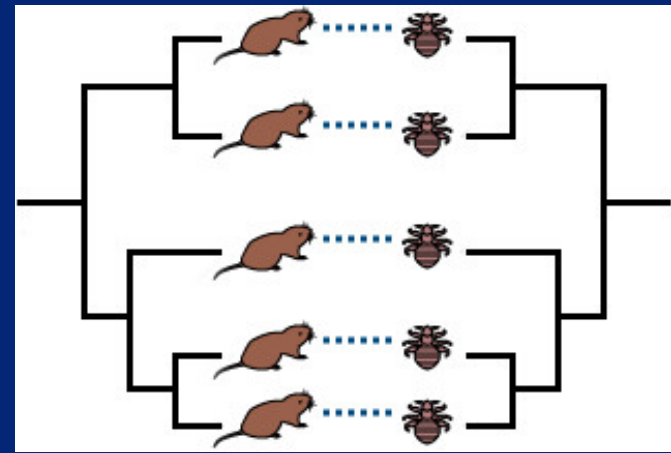
- Other Indonesian island taxa also smaller than average (mammoth, etc)
- *H. floresiensis* also has unusual limb and joint traits
- 700,000 y.o. fossil recently found could be an ancestor/intermediate with *H. erectus*; it's even smaller!

Anti species hypotheses

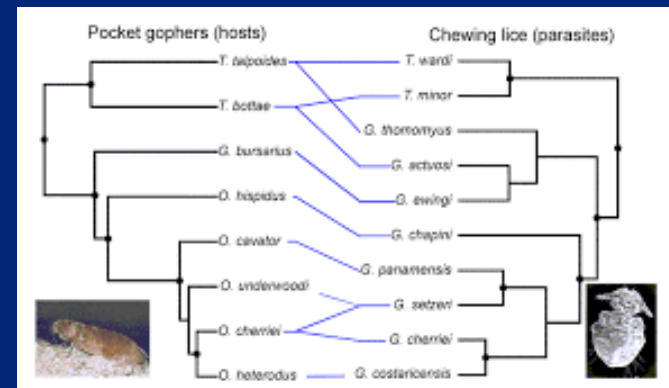
- Laron syndrome:
 - Autosomal recessive dwarfing syndrome
 - *H. floresiensis* smaller body and brain than typical for this, but genetic background could matter
 - (Side note: some modern humans with Laron syndrome have normal IQ)
- Cretinism (lack of functioning thyroid due to environmental factors)
- Question: are these hypotheses mutually exclusive with species status?

Cospeciation

- Host species and parasite species often speciate together
- Species trees of the two groups will look very similar
- Reproductive isolation of hosts may isolate parasites
- Adaptation of hosts may spur adaptation of parasites (or vice versa)



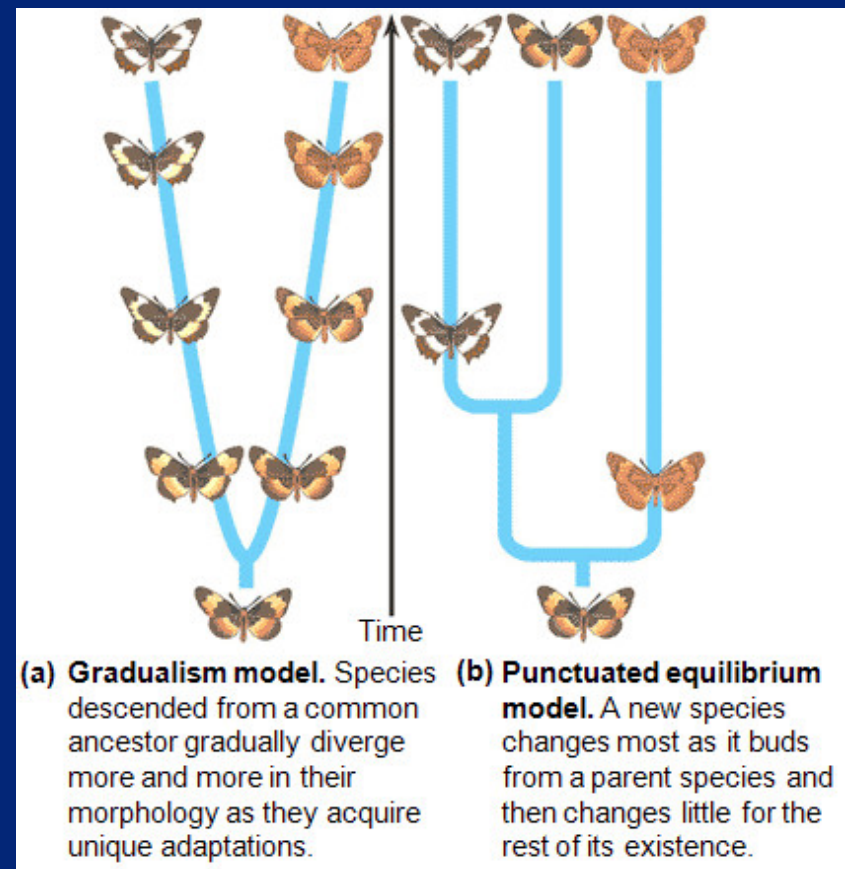
Idealized schematic



Real data

Gradual versus punctuated

- Classical model: species slowly accumulate differences
- Punctuated equilibrium model (Stephen Gould and colleagues)
 - Burst of change at speciation
 - Relative stasis elsewhere
- Favored by paleontologists, who find bursts of change in the fossil record



Some thoughts on punctuated evolution

- Paleontological species definition encourages this view because “gradual” species can’t be recognized as such
- Not all species evolve the same way
- Major changes in body or lifestyle probably require a burst of changes

Domesticated foxes

- Long-term breeding project started by Belyaev in Siberia, 1959
- Fox kits selected for sociability with humans
- Results not only sociable but oddly dog-like



Wild silver fox: image by Zefram



Russian domesticated foxes

Domesticated foxes

- Complex differences from wild foxes:
 - Very tame even when raised in cages
 - Domestic-like color patterns
 - Wagging tails, whining, barking
 - Can follow pointing finger or gaze
 - A few try to reproduce more than once/year
 - Differences in hormone levels, developmental timing
 - Differences in skull and body plan
- Major change in developmental pathway?

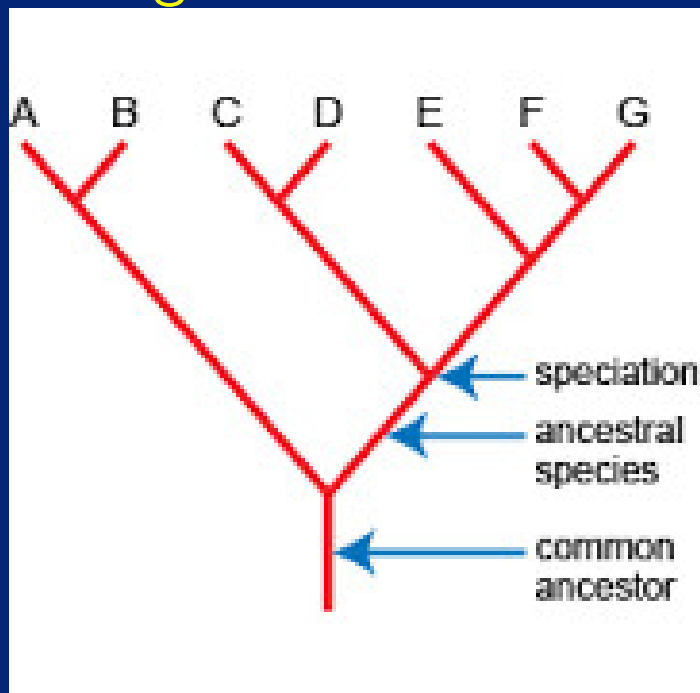
Stasis

- Some organisms appear to change very little over time:
 - ginkgo
 - coelacanth
 - horseshoe crab
- Other organisms change much more rapidly and diversify into multiple species
- Why? Ideas:
 - Change-resistant developmental “program”
 - Broad ecological niche
 - No improvements in easy reach
 - Cryptic species?

What is a phylogeny?

Phylogeny: a branching tree showing inferred relationships among species, populations, or individuals

Synonyms: evolutionary tree, cladogram



Other useful terms

- “Tree” – same as phylogeny
- Taxon, taxa – the units at the tips of the tree (species, populations, individuals)
- Clade – all taxa descending from a common ancestor
- Root – the common ancestor of the whole phylogeny

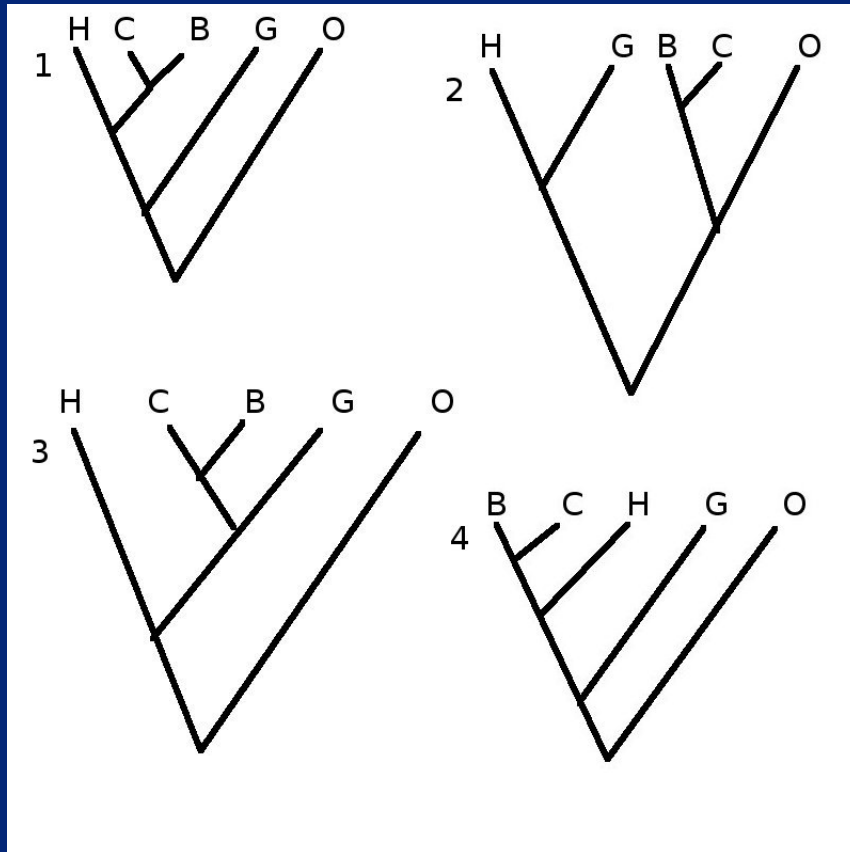
What are phylogenies good for?

- Relationships between organisms
- Dates of evolutionary events
- Evolutionary patterns—did some features evolve multiple times?
- Removing influence of phylogeny from ecological analyses (“comparative method”)
- Relationships among genes
- Patterns of speciation and diversification

How to look at a phylogeny

- Branching pattern shows pattern of relationships
- Right-left ordering is NOT significant; can be rearranged to emphasize or obscure points!
- Branch lengths may or may not be meaningful
- Biologists draw root at the bottom; math and CS types draw root at the top

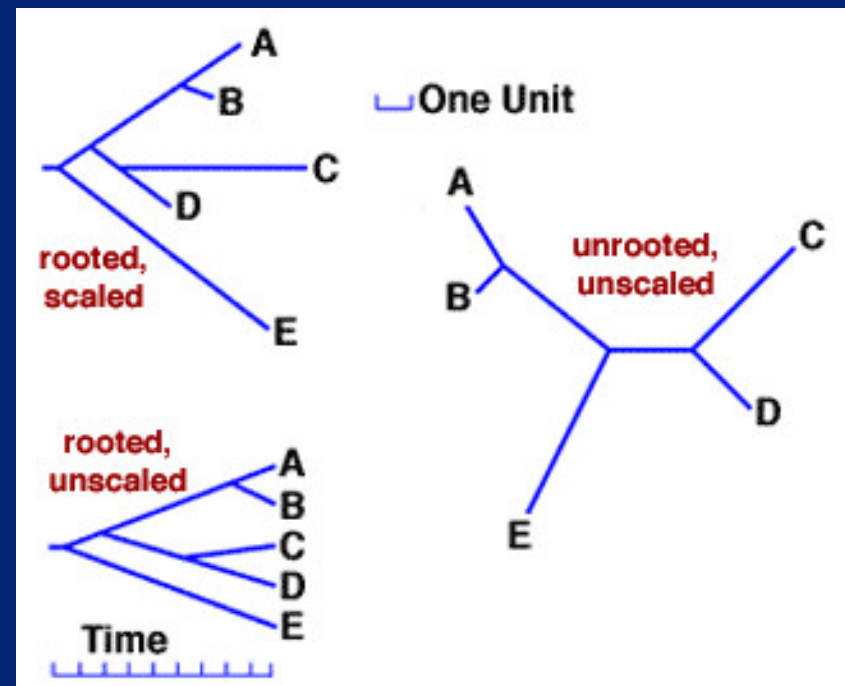
Practice problem



Two of these trees are the same (except for branch lengths).
Which two?

Rooted versus unrooted trees

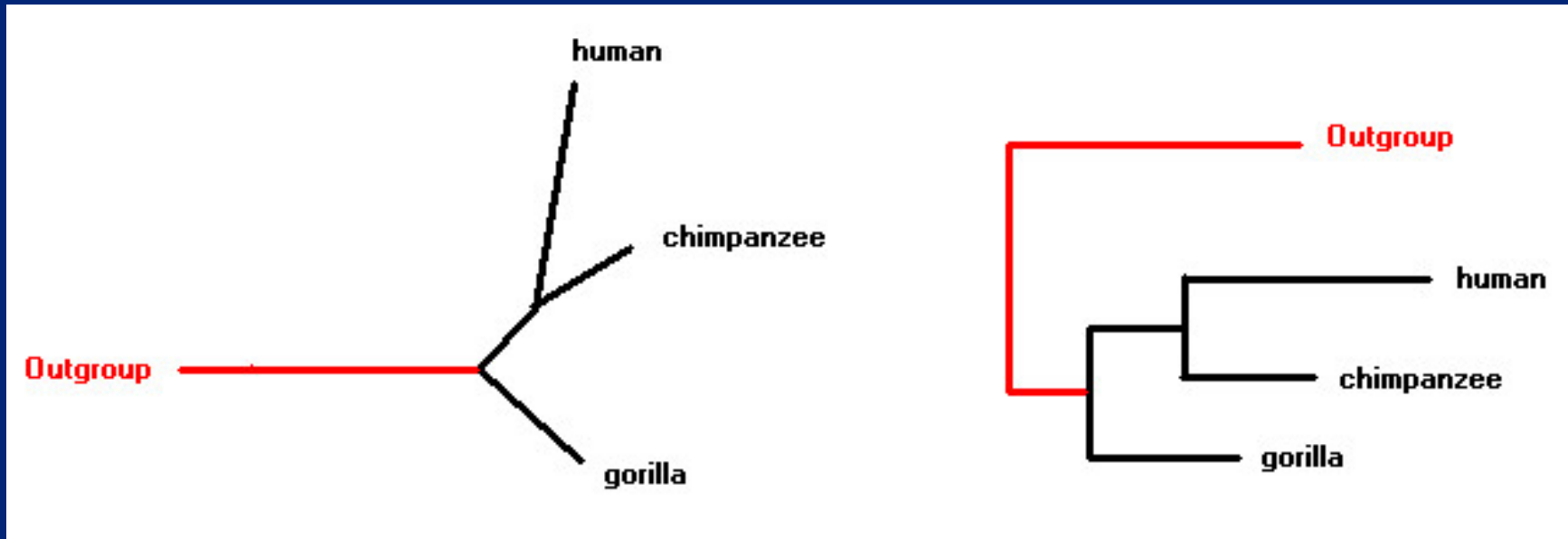
- A rooted tree (phylogeny) has a specific direction of evolution
- The root is the ancestral form from which the others evolved
- This is the most informative type of tree
- Unfortunately, most phylogeny inference methods produce unrooted trees



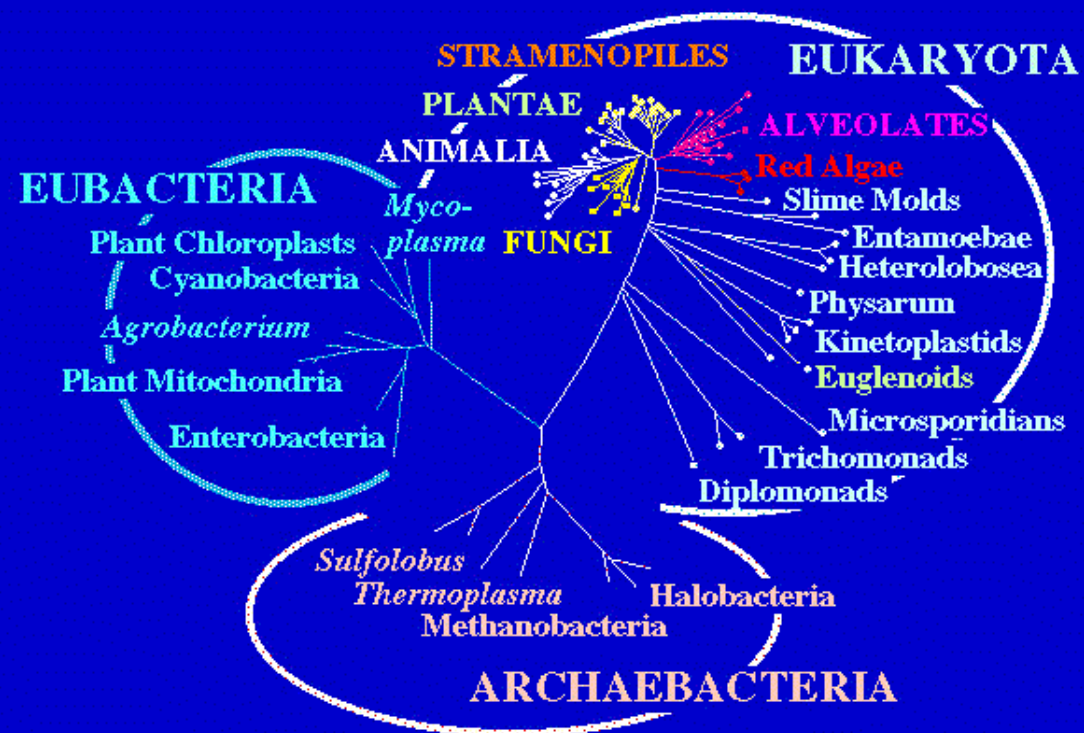
Rooted versus unrooted trees

- An unrooted tree corresponds to a collection of different rooted trees
- We don't know the direction of evolution
- Biological interpretation can be difficult without root
- Ways to root a tree:
 - Outgroup
 - Molecular clock

Outgroup rooting



- Outgroup – species known not to belong to clade
 - Wrong outgroup leads to wrong root
 - Too-distant outgroup leads to noise in data
- Some comparisons have no suitable outgroup



Molecular clock

- Can we assume rate of evolution the same on all branches?
- If so:
 - Root is point most distant from all tips
 - Branch length is proportional to time
 - If we can date a few points on tree, can date entire tree
- Clock may not hold:
 - Unequal generation time
 - Different selection constraints
 - Different mutation rates
- Clock assumption safest among closely related species

Appropriate data for phylogenies

- Good phylogenetic data has:
 - Enough variation to show relationships
 - Not so much variation that it randomizes signal
 - Ability to establish homology
 - *Relative freedom from convergent evolution*
 - Mode of evolution relatively well understood
 - If possible, a good clock
- No one type of data works for all problems

Some important dates in history

Origin of the universe	$-12^a \pm 2$
Formation of the solar system	-4.6 ± 0.4
First self-replicating system	-3.5 ± 0.5
Prokaryotic-eukaryotic divergence	-2.5 ± 0.3
Plant-animal divergence	-1.0
Invertebrate-vertebrate divergence	-0.5
Mammalian radiation beginning	-0.1

^aBillions of years ago

Protein family	PAMs ^a /100 res. /10 ⁸ years	Protein Lookback time ^b	
Pseudogenes	400	45 ^c	Primates, Rodents
Fibrinopeptides	90	200	Mammalian Radiation
Lactalbumins	27	670	Vertebrates
Ribonucleases	21	850	Animals
Hemoglobins	12	1.5 ^d	Plants/Animals
Acid Proteases	8	2.3	Prokaryotic/Eukaryotic
Triosephosphate isomerase	3	6	Archaea
Glutamate dehydrogenase	1	18	?

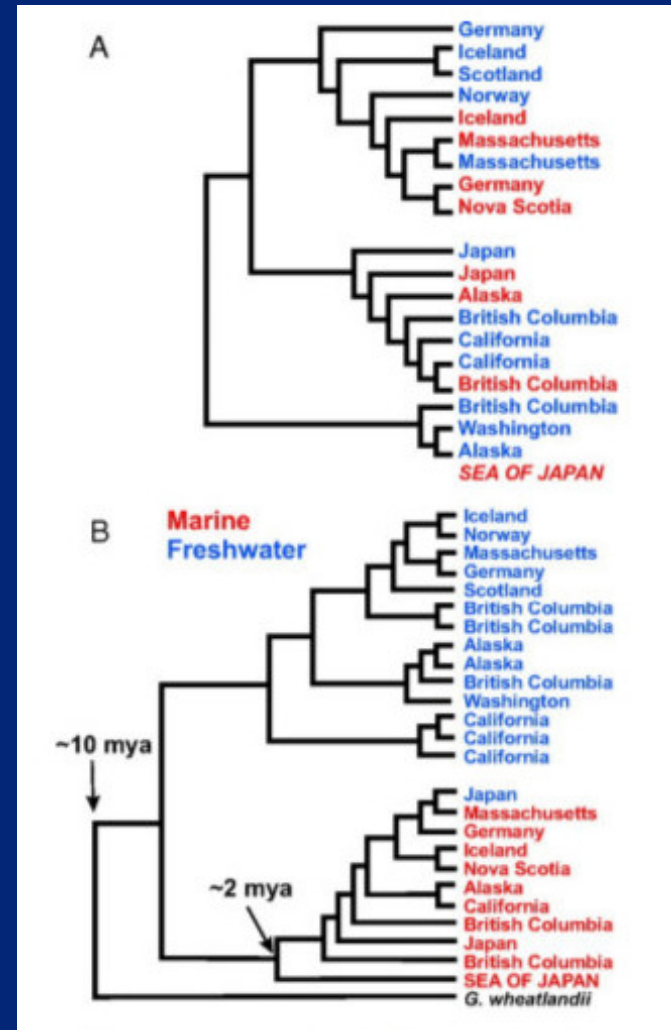
^aPAMs, point accepted mutations. ^bUseful lookback time, 360 PAMs, 15% identity. ^cMillions of years. ^dBillions of years.

Appropriate data for phylogenies

- Within species, between closely related species
 - Non-coding DNA and pseudogenes
 - Microsatellites
 - Very fast-evolving genes
 - mtDNA
- Between moderately similar species
 - Most protein-coding genes, especially housekeeping genes
- Between extremely dissimilar species
 - Ribosomal RNA
 - Very slow-evolving genes

Convergent evolution?

- Why not use loci involved in “exciting” traits of the species?
- Convergent evolution:
 - Two clades are under the same external pressure
 - They independently evolve the same response
 - Not a reliable indicator of relationships
- Upper figure is many random genes; lower is a gene involved in fresh/saltwater adaptation



Why phylogenies are hard

Tips	Topologies
3	3
4	18
5	180
6	2700
7	56700
8	1587600
9	57153600
10	2571912000
15	6958057668962400000
20	5644809895887305913369600000000
30	43684666131030695124646801986207638914406400000000000000
40	30273338299480073565463033645514572000429394320538625017078887219200000000000
50	3.28632×10^{112}
100	1.37416×10^{284}

Why phylogenies are hard

- In many cases tree search known to be “NP complete”
- No efficient algorithm is known—none may exist but this is unproven
- Solving any NP-complete problem solves ALL OF THEM
- Three consequences of such an algorithm
 - Reliably find the right phylogeny
 - Crack most/all current codes (business and military)
 - Difficult conversation with the NSA....
- Must use heuristic approximations which will sometimes fail (get the wrong tree)

Phylogeny methods

- Three major approaches to phylogeny inference
 - Parsimony (today)
 - Distance (Friday)
 - Maximum likelihood and Bayesian methods (Friday)

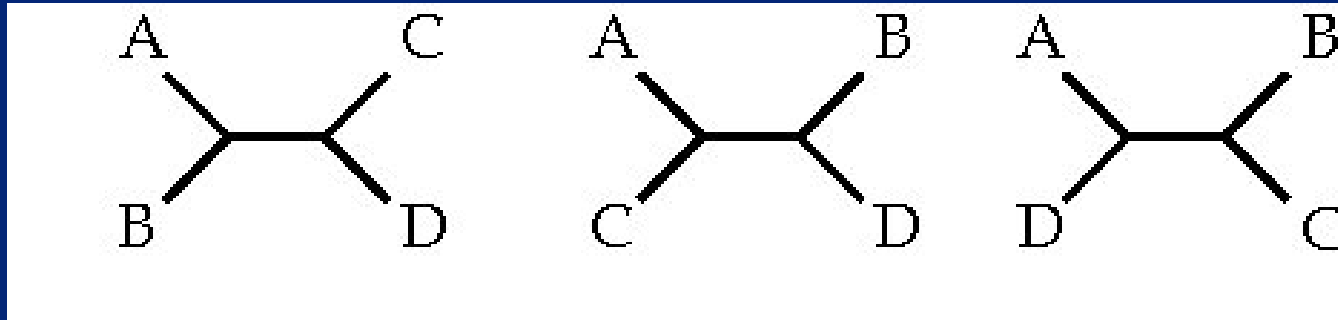
Parsimony methods

- (Philosophical) Principle of Parsimony: Make as few assumptions as possible
- (Phylogenetic) Principle of Parsimony: Prefer the tree that assumes the smallest number of evolutionary changes
- Assumes that changes are fairly rare and evenly distributed

Parsimony methods

- Advantages of parsimony:
 - No explicit mutational model required
 - Applicable to the widest variety of data—including morphological traits (all we have for fossils)
 - Moderately fast
- Disadvantages:
 - No explicit mutational model possible
 - Long branch attraction
 - Limited ability to put error bars on phylogeny estimate

Practice problem–parsimony

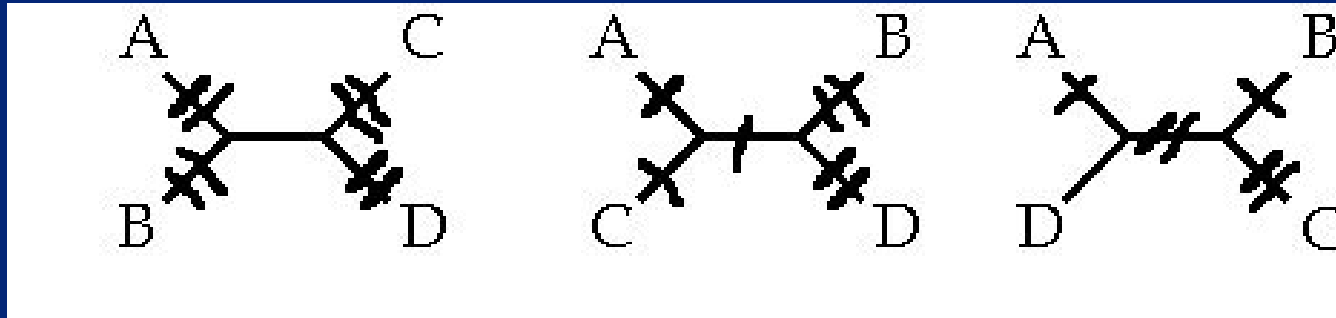


Taxon	1	2	3	4	5
A	A	A	C	G	A
B	T	A	A	T	T
C	T	A	A	G	A
D	A	C	C	G	T

How many changes are needed on each tree topology?

Which topology is preferred by parsimony?

Practice problem–parsimony



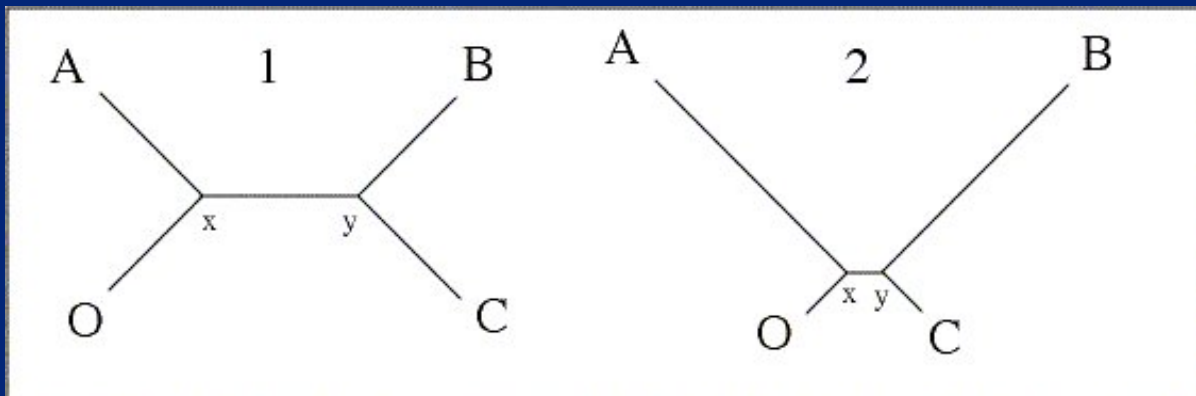
Taxon	1	2	3	4	5
A	A	A	C	G	A
B	T	A	A	T	T
C	T	A	A	G	A
D	A	C	C	G	T

How many changes are needed on each tree topology? 8, 7, 6

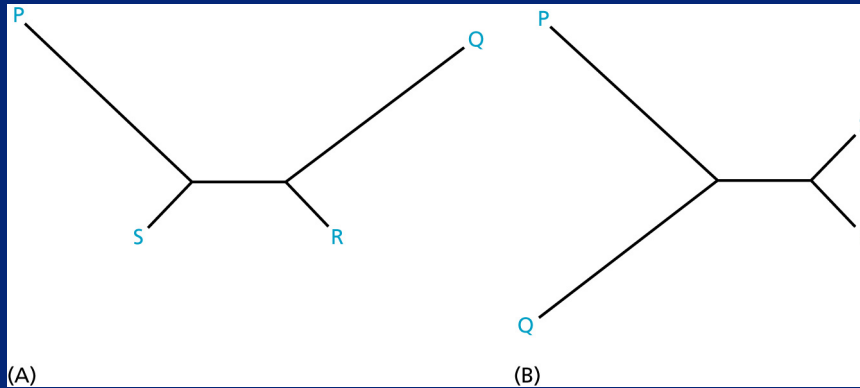
Which topology is preferred by parsimony? *Third topology*

Parsimony methods

- Felsenstein showed a four-tip tree which gives inconsistent results with parsimony
- “Inconsistent” – wrong answer becomes more probable as data increases
- With infinite data you would be 100% sure to get the wrong answer



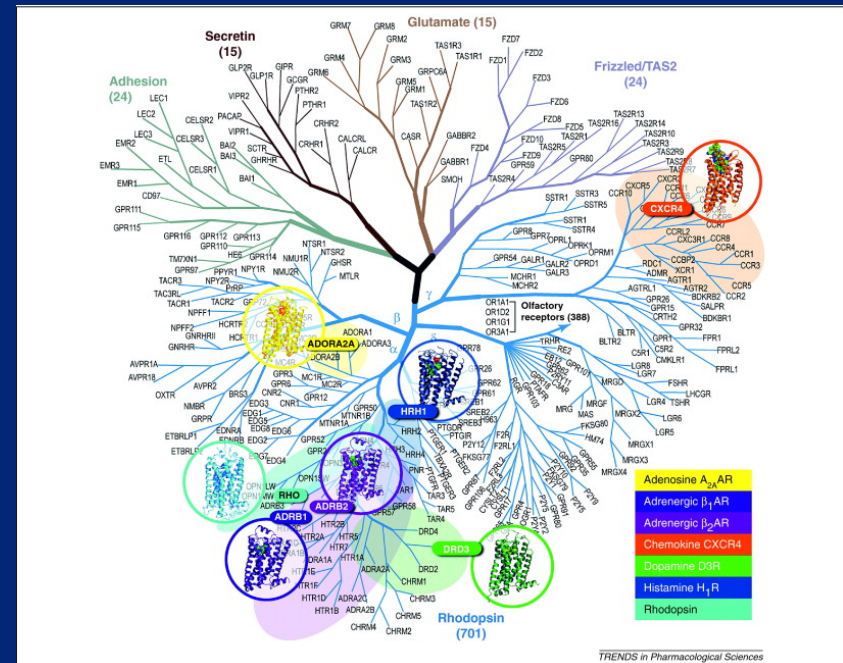
Long branch attraction



- Data from tree A leads to inference of (wrong) tree B
- Two convergent changes on the long branches are more likely than a single change on the short branches
- This violates the basic principle of parsimony
- Fast-evolving sites and data sets are particularly problematic

Betting on your trees

- Ken Rice makes parsimony trees of G-protein coupled receptors
 - Maximum likelihood too slow
 - Distance methods didn't perform well
- If new gene groups with:
 - Odor receptors – ignore
 - Neurotransmitter receptors – spend \$2K to validate



G-protein coupled receptor genes

One-minute responses

- Tear off a half-sheet of paper
- Write one line about the lecture:
 - Was anything unclear?
 - Did anything work particularly well?
 - What could be better?
- Leave at the back on your way out