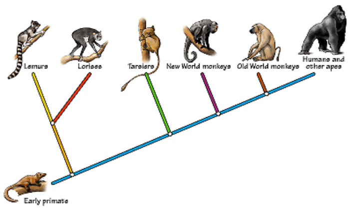
PHYLOGENETIC TREES

Terminology and
Maximum Parsimony

What is a phylogeny?

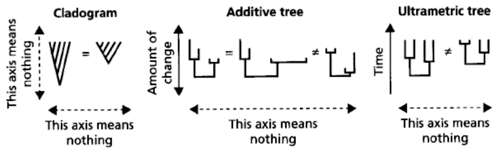
Phylogeny is the history of a group of taxa described as an evolutionary tree with a common ancestor at the base and descendent taxa as branch tips



descendent taxa

common ancestor

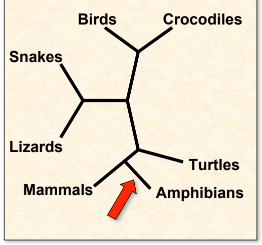
Different tree depictions contain different information



Terms to describe phylogenetic trees

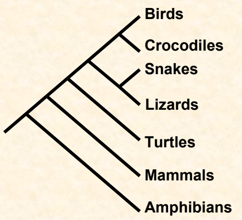
Remember ClustalW midpoint rooting.

Unrooted tree



Birds
Crocodiles
Snakes
Lizards
Turtles
Mammals
Amphibians

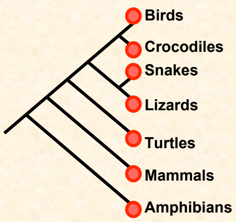
Rooted tree



Birds
Crocodiles
Snakes
Lizards
Turtles
Mammals
Amphibians

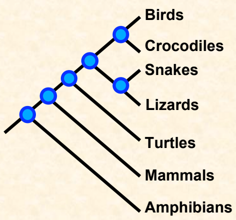
Terms to describe phylogenetic trees

Terminal nodes



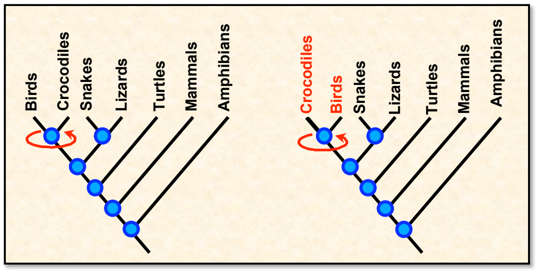
Birds
Crocodiles
Snakes
Lizards
Turtles
Mammals
Amphibians

Interior nodes



Birds
Crocodiles
Snakes
Lizards
Turtles
Mammals
Amphibians

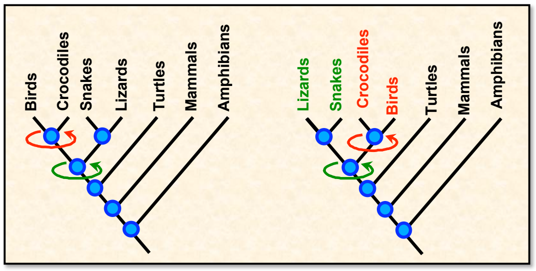
Taxa can rotate around interior nodes without changing relationships



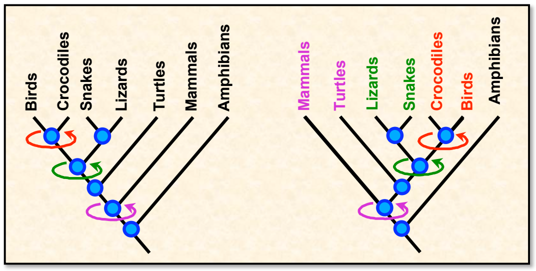
Birds
Crocodiles
Snakes
Lizards
Turtles
Mammals
Amphibians

Crocodiles
Birds
Snakes
Lizards
Turtles
Mammals
Amphibians

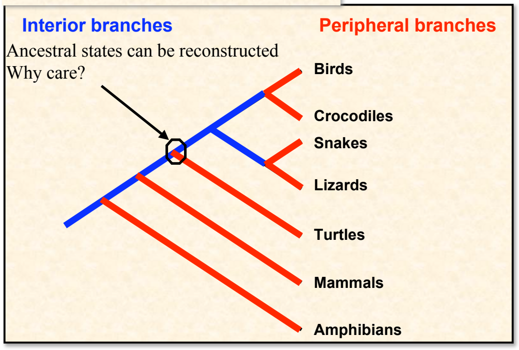
Taxa can rotate around interior nodes without changing relationships



Taxa can rotate around interior nodes without changing relationships



Terms to describe phylogenetic trees



Ancestral states can be reconstructed
Why care?

It is incorrect to say "high or low homology" or "percent homology."

Homology is a *hypothesis* that similarity of a trait in two or more species indicates descent from a common ancestor.

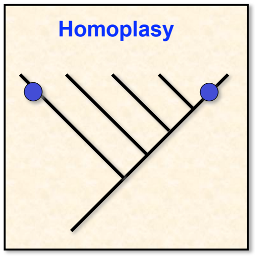
a trait could be a morphological, behavioral, or molecular character

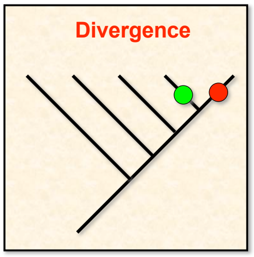
Homologous features are derived from common ancestors.

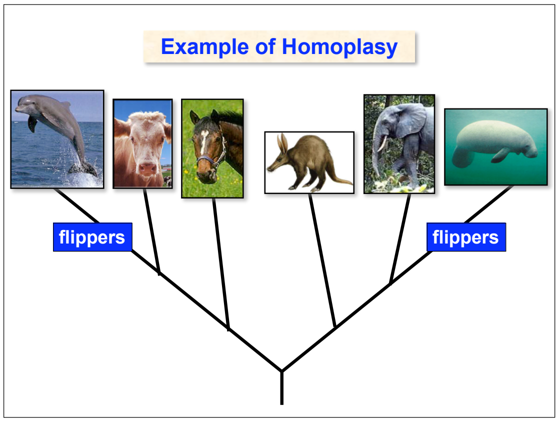
It seems reasonable to assume that things that look similar are related.

But, appearances can be misleading.

Two common reasons for deceptive appearances


Homoplasy

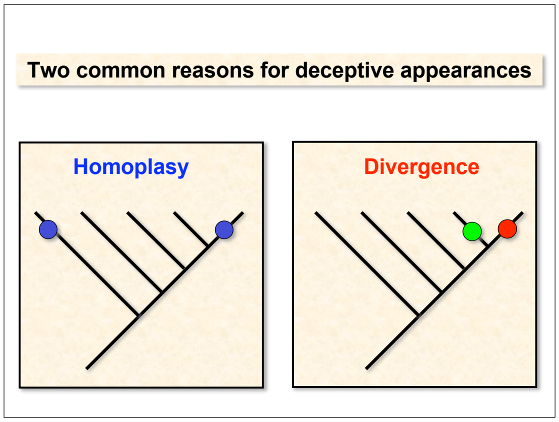

Divergence



Example of Homoplasy

flippers

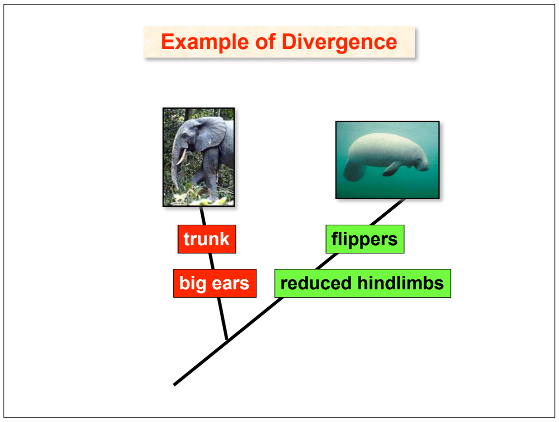
flippers



Two common reasons for deceptive appearances

Homoplasy

Divergence



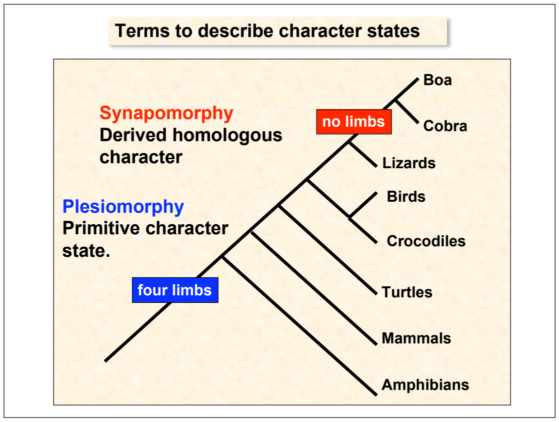
Example of Divergence

trunk

big ears

flippers

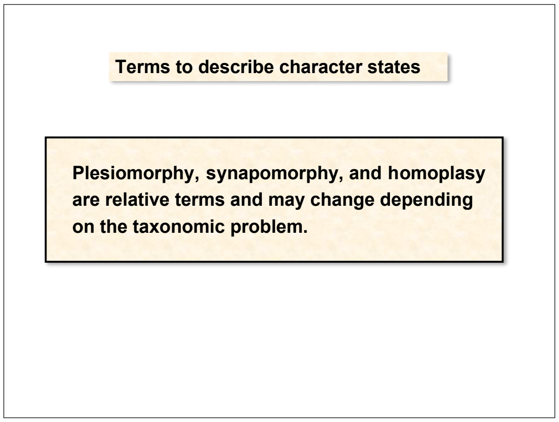
reduced hindlimbs



Terms to describe character states

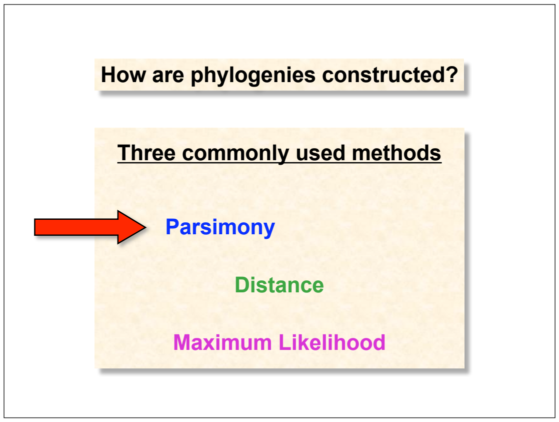
Synapomorphy
Derived homologous character

Plesiomorphy
Primitive character state.



Terms to describe character states

Plesiomorphy, synapomorphy, and homoplasy are relative terms and may change depending on the taxonomic problem.



How are phylogenies constructed?

Three commonly used methods

Parsimony

Distance

Maximum Likelihood

Parsimony Methods

Few assumptions:

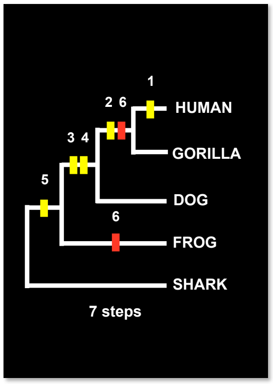
1. Simpler hypotheses are preferred to complicated ones.
2. Only synapomorphies are informative about relationships.

The basic idea of parsimony is to minimize the number of changes occurring on a phylogeny. Check all possible tree topologies and choose the shortest.

Morphology

	1	2	3	4	5	6
HUMAN	1	1	1	1	1	1
GORILLA	0	1	1	1	1	1
DOG	0	0	1	1	1	0
FROG	0	0	0	0	1	1
SHARK	0	0	0	0	0	0

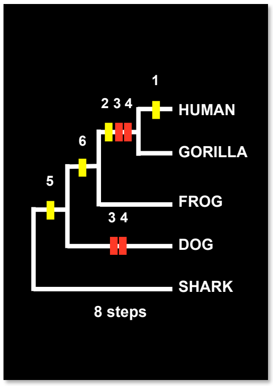
1. Bipedal
2. Position of orbits
3. Warm-blooded
4. Hair
5. Bony skeleton
6. Short tail



Morphology

	1	2	3	4	5	6
HUMAN	1	1	1	1	1	1
GORILLA	0	1	1	1	1	1
DOG	0	0	1	1	1	0
FROG	0	0	0	0	1	1
SHARK	0	0	0	0	0	0

1. Bipedal
2. Position of orbits
3. Warm-blooded
4. Hair
5. Bony skeleton
6. Short tail

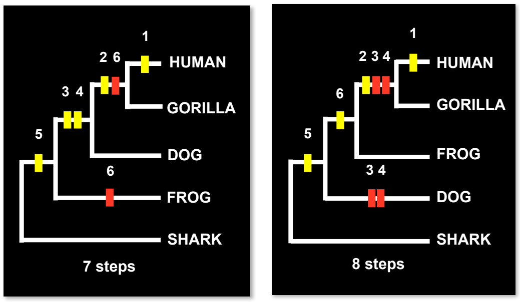


Morphology

	1	2	3	4	5	6
HUMAN	1	1	1	1	1	1
GORILLA	0	1	1	1	1	1
DOG	0	0	1	1	1	0
FROG	0	0	0	0	1	1
SHARK	0	0	0	0	0	0

1. Bipedal
2. Position of orbits
3. Warm-blooded
4. Hair
5. Bony skeleton
6. Short tail

Which tree is more parsimonious?

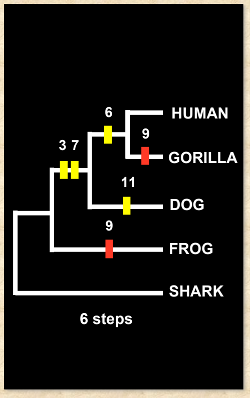


Molecular (DNA)

	1	2	3	4	5	6	7	8	9	10	11
HUMAN	C	A	C	T	A	C	T	C	A	T	T
GORILLA	C	A	C	T	A	C	T	C	G	T	T
DOG	C	A	C	T	A	T	T	C	A	T	C
FROG	C	A	T	T	A	T	A	C	G	T	T
SHARK	C	A	T	T	A	T	A	C	A	T	T

1. Aligned position 1
2. Aligned position 2
3. Aligned position 3
4. Aligned position 4
5. Aligned position 5

(and so on)



6 steps

Molecula (DN

	1	2	3	4	5	6	7	8	9	10	11
HUMAN	C	A	C	T	A	C	T	C	A	T	T
GORILLA	C	A	C	T	A	C	T	C	G	T	T
DOG	C	A	C	T	A	T	T	C	A	T	C
FROG	C	A	T	T	A	T	A	C	G	T	T
SHARK	C	A	T	T	A	T	A	C	A	T	T

1. Aligned position 1
2. Aligned position 2
3. Aligned position 3
4. Aligned position 4
5. Aligned position 5

(and so on)

How are phylogenies constructed?

Three commonly used methods

Parsimony

Distance

Maximum Likelihood

Distance Methods

Overall similarity indicates evolutionary relationships.

Taxa are grouped by the number of differences in character states.

No distinction is made between synapomorphies and plesiomorphies.

Distance methods start by calculating the distances between pairs of species.

	Morphology	Molecular (DNA)
HUMAN	1 1 1 1 1 1	C A C T A C T C A T T
GORILLA	0 1 1 1 1 1	C A C T A C T C G T T
DOG	0 0 1 1 1 0	C A C T A T T C A T C
FROG	0 0 0 0 1 1	C A T T A T A C G T T
SHARK	0 0 0 0 0 0	C A T T A T A C A T T

	Human	Gorilla	Dog	Frog
Human				
Gorilla	2			
Dog	5	5		
Frog	8	6	7	
Shark	9	9	6	3

How are phylogenies constructed?

Three commonly used methods

Parsimony

Distance

Maximum Likelihood

Maximum Likelihood Methods

ML methods find the tree most likely to produce the data observed given a specific model of how character states change.

ML methods are almost always used with only molecular data.

Which method should I use?

- Depends on the data being analyzed.
- Try all methods to determine if results are robust to model assumptions.
- My preference is for ML.

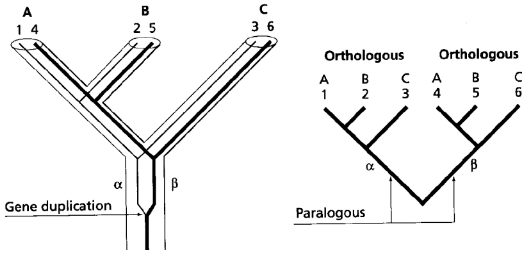
Some things we ignored today but have discussed

- Base composition
- Transition/transversion bias
- DNA substitution models
- Multiple substitutions
- Character weighting
- Choice of taxa

Does a tree generated from a gene sequence always give the species tree?

Gene tree versus species tree

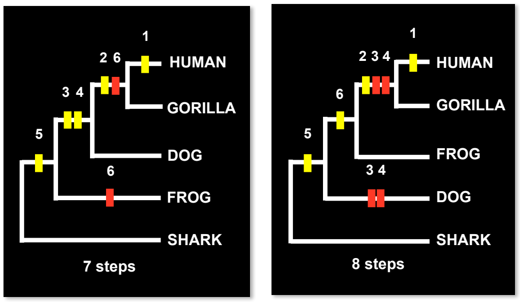
How many different genes should be used?



What gene(s) should we use?

- mtDNA
  - Bar code
- Nuclear genes
  - Introns
  - Exons
  - Non-coding
  - Neutral versus selected
- Chloroplast genes

Branching order, test for significance



Is the branching order significant?

- Bootstrap test for significance.
- Resample data with replacement
- Reconstruct phylogeneny
- Redo 100s of times
- Build consensus tree, bootstrap values are percent on times that branch was recovered.

ORIGINAL DATA

AAATCTGCTAGCATGCA
AAATCTGCTACCATGCA

BOOTSTRAPPED DATA SETS

AAAAACCCGGCTCAGCA
AAAAACCCGGCTCAGCA

AATTTTTCCCCATTGCC
AATTTTTCCCCATTGCC

AATTTCCAAACCCCTCA
AATTTCCAAACCCCTCA

AATTTGCCCCCCCCAAA
AATTTGCCCCCCCCAAA

AAATCCCGGTACTTGCA
AAATCCCGGTACTTGCA

AATTGGAAGCCAAGGCA
AATTGGAACCCAAGGCA

AAAATCCCCCATGGGAA
AAAATCCGCCATGGGAA

AATTGGCACCCCAAAAC
AATTGGCAGGCCAAAAC

AAATTCTTCCTGGCAAA
AAATTCTTCCTGGCAAA

AAAATTTCCACCATTCA
AAAATTTCCAGGATTCA

Questions

- Define maximum parsimony and maximum likelihood.
- Given two tree topologies, are they the same or different?
- Define homology.
- Given a data matrix and two tree topologies, calculate which is most parsimonious.
- Is a gene tree always the same as a species tree?
- What is one way to determine support for branching pattern?