

Phylogenetics

RL-V3 MPP: Course introduction

Rachel Warnock, Tim Brandler, Laura Mulvey
rachel.warnock@fau.de

Updated April 25, 2023

About this course

Thursdays 14:00–16:00 CET at Henke Str.

8 weeks of lectures + 6 weeks for project work

All lecture material available via the [Course website](#)

Course objectives

To learn the role of phylogenetics in palaeobiology

- Tree building
- Substitution models
- Dating trees
- Clock models
- Tree models
- Diversification rates
- Continuous trait evolution

We will apply these models in a Bayesian phylogenetic framework using the software [RevBayes](#)

Course evaluation

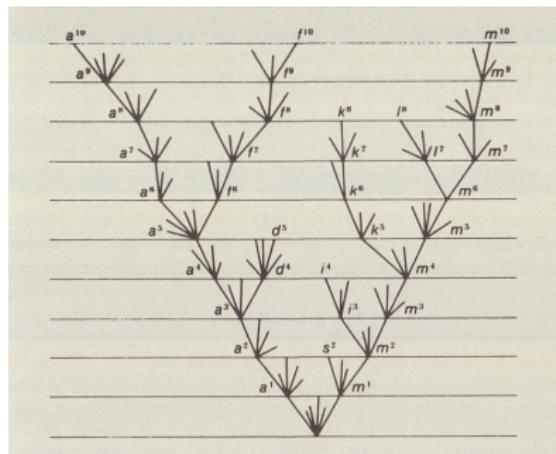
"Phylogenetics" is graded together with "Introduction to Statistical Modelling". Course code: RL-V3 MPP.

Evaluation consists of two components:

- Homework exercises that include reading and videos (6 in total)
- A written project report

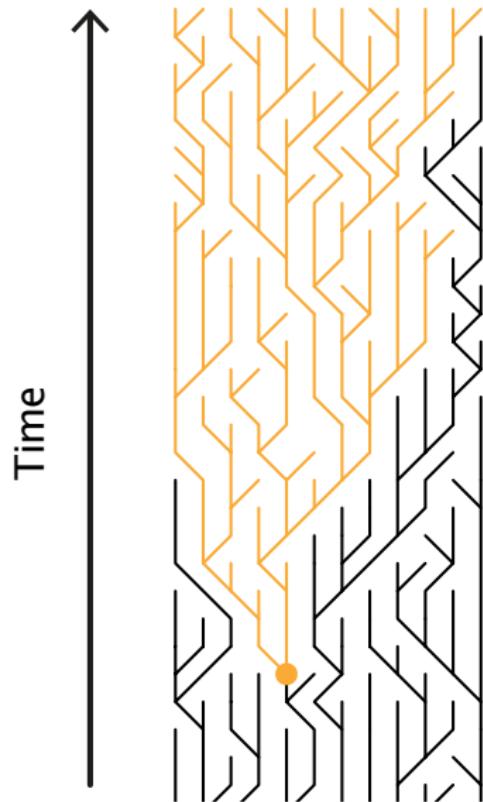
Today's objectives

- Recap “tree-thinking”
- Gain an understanding of the parsimony approach to tree-building and statistical inconsistency



From Darwin's *Origin of Species*

What is phylogenetics?



- populations
- species
- viruses
- cells
- languages

Data

- DNA
- morphology
- words
- cells

What is phylogenetics?

Phylogenetics allows us to study the relationships between entities that are related via an evolutionary process.

We can apply the same principles to any scenario where we have hierarchical (ancestor & descendant) relationships.

The data is anything that can tell us about the relationships between individuals.

What is phylogenetics?

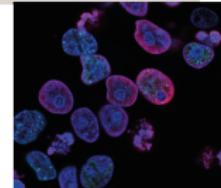
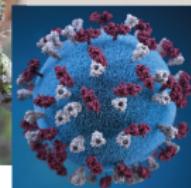
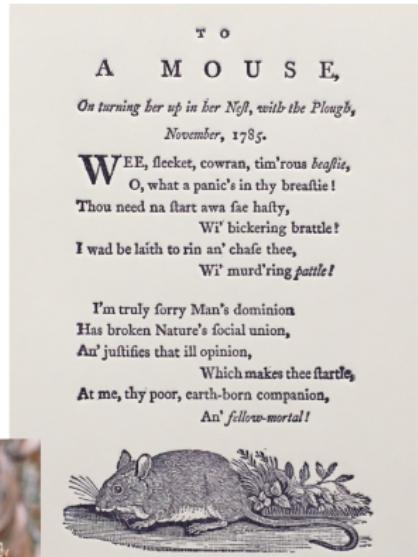


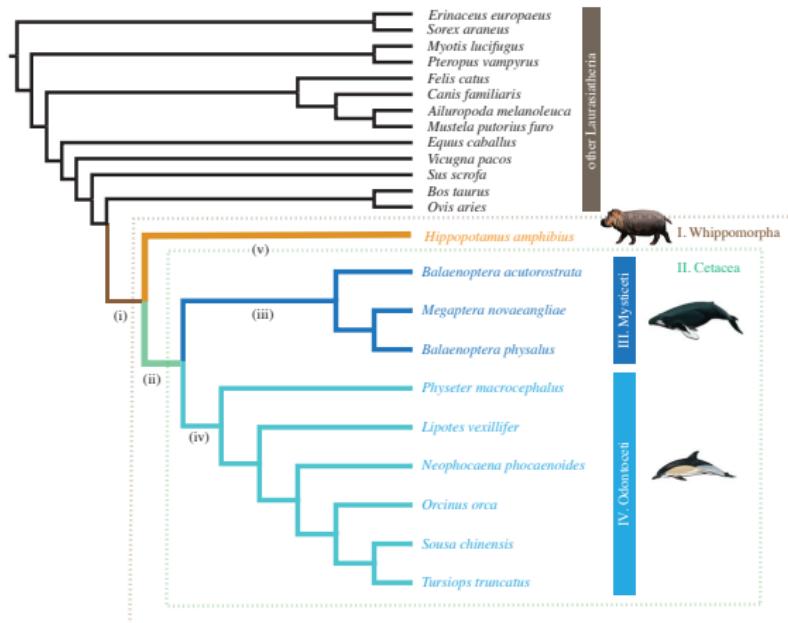
Image source upsplash.org

What are our data?

What can we learn from trees?

How are our favourite species related?

How does phylogeny reflect taxonomy?



What can we learn from trees?

Evolutionary relationships

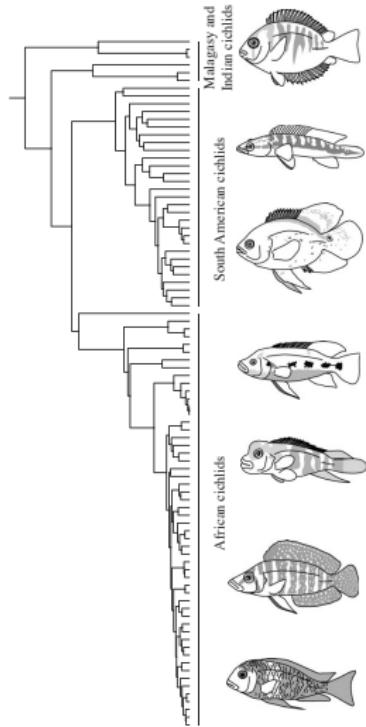


Image adapted from Friedmann et al. 2013. PRSB

What can we learn from trees?

Evolutionary relationships

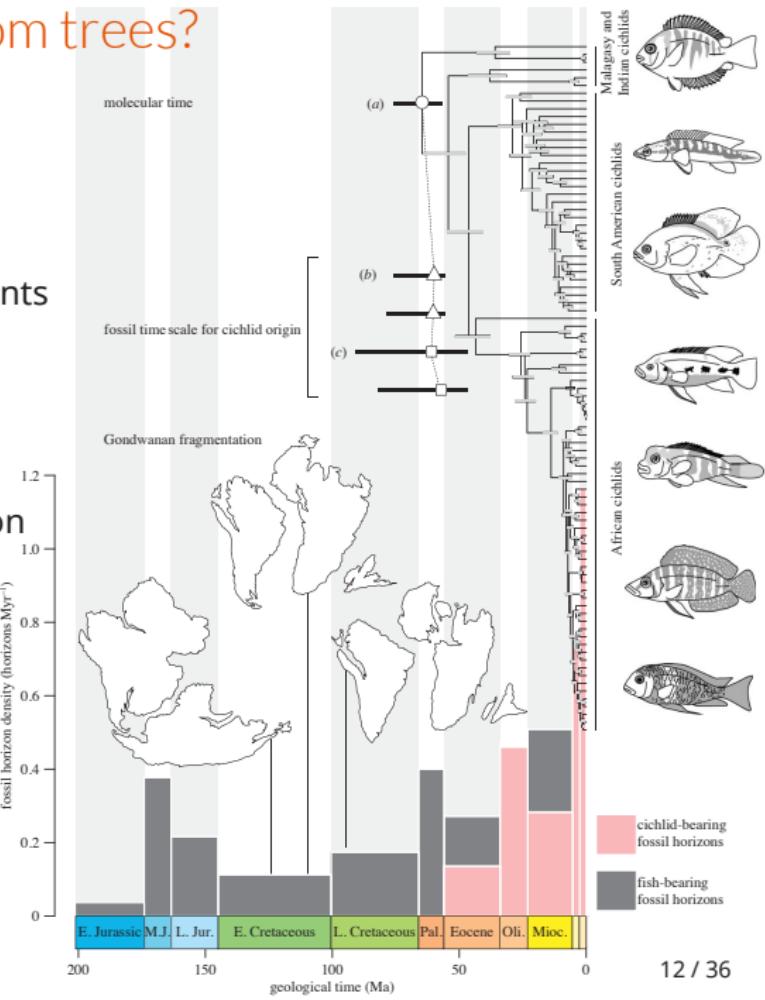
Timing of diversification events

Geological context

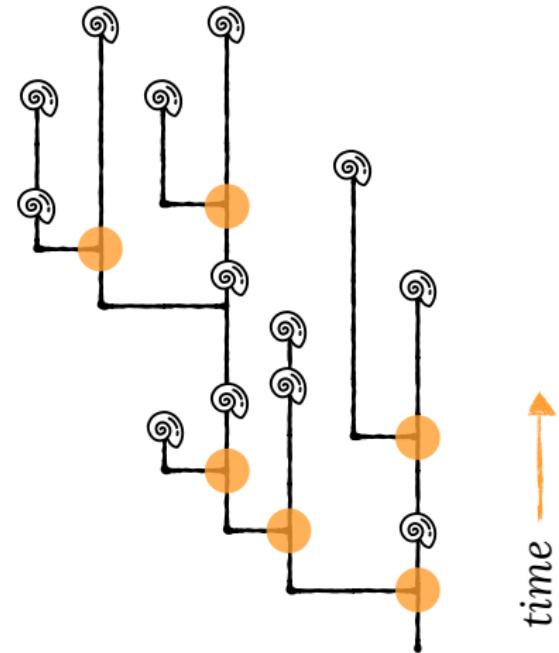
Rates of phenotypic evolution

Diversification rates

...



A phylogenetic tree captures part of evolutionary history that is otherwise not directly observable.

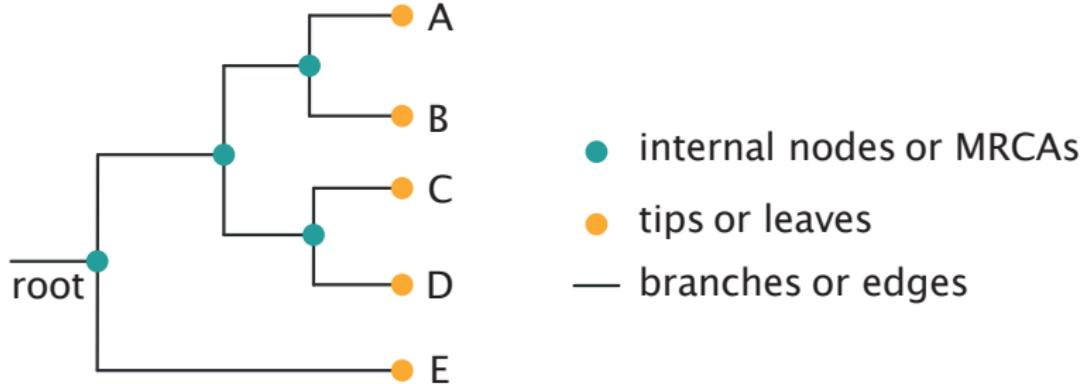


Nothing in biology makes sense except in the light of evolution
— Theodosius Dobzhansky (1973)

*Nothing in evolution makes sense except when seen in the
light of phylogeny* — Jay M. Savage (1997)

Where do we begin?

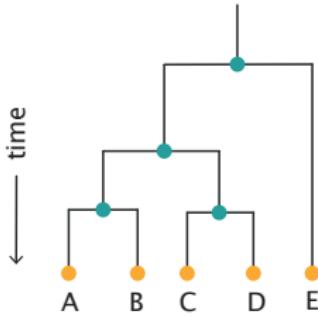
Some basic terms



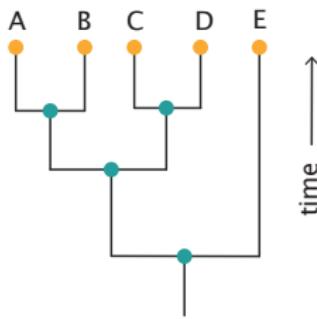
branch lengths = genetic distance OR time

Note: genetic distance = rate x time

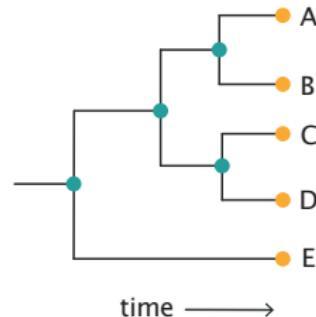
The direction of time



Computer science, maths



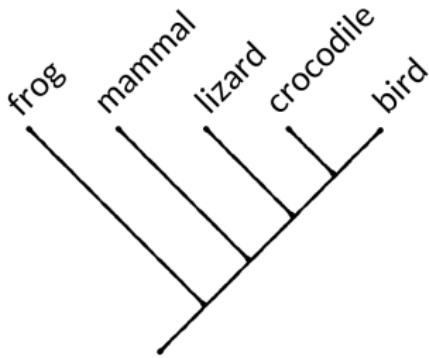
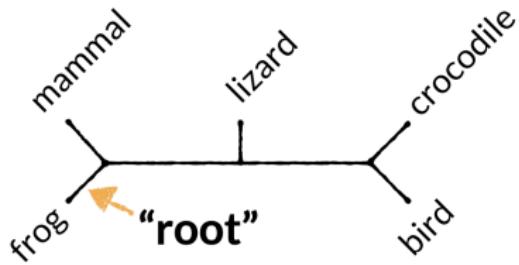
Geology



Evolutionary biology

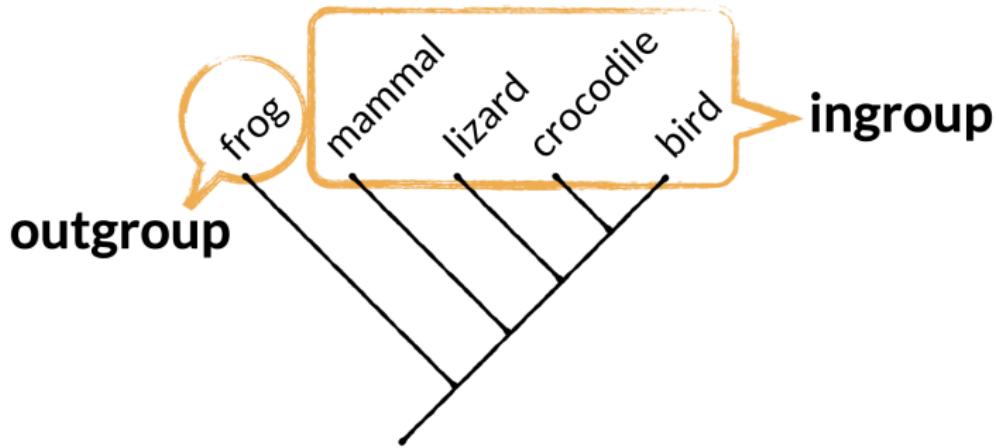
Tip: look for the root!

Rooted versus unrooted trees



Phylogenies are unrooted by default, because phylogenetic characters don't contain information about the direction of time.

Rooted versus unrooted trees

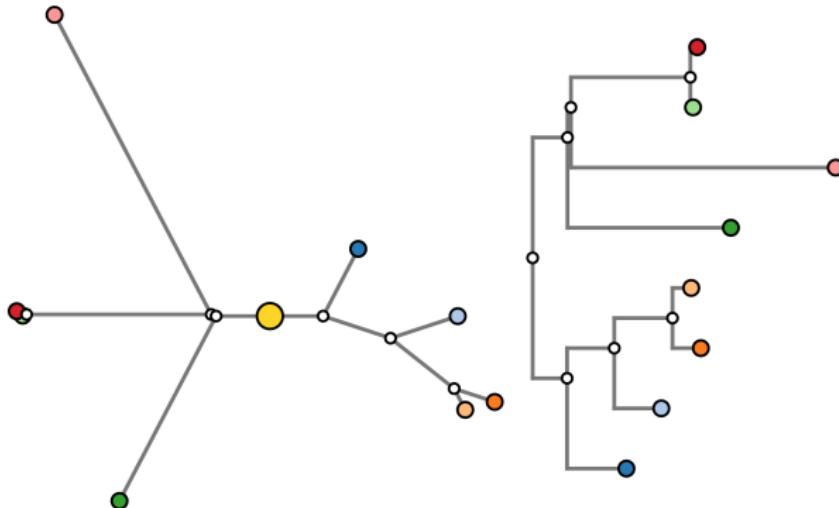


We have to find a way of breaking one of the branches in two, where the break represents the oldest point in our tree.

The most common approach is to use an outgroup – a taxon that we know is more distantly related than any of the taxon within the ingroup.

Rooted versus unrooted trees

By default phylogenies are not rooted.



We need an **outgroup** OR a model that includes **time**.

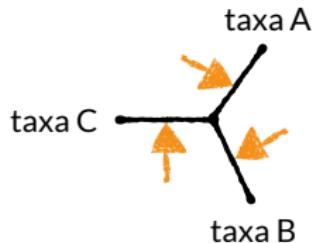
Use Art Poon's [online tool](#) to explore this further. Click [here](#) to learn more about reading trees.

Rooted versus unrooted trees

How many possible trees are there for 3 species?

Rooted versus unrooted trees

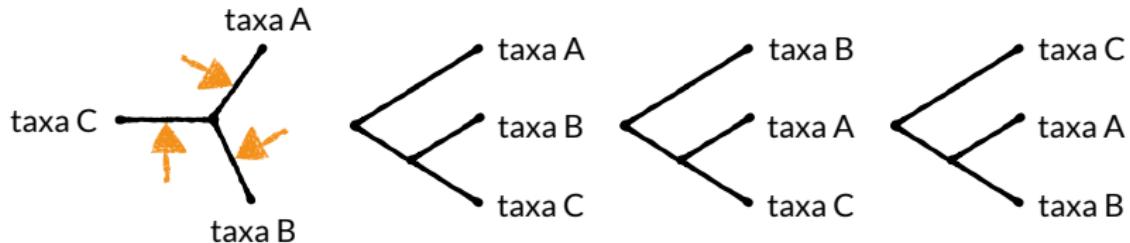
How many possible trees are there for 3 species?



unrooted = 1

Rooted versus unrooted trees

How many possible trees are there for 3 species?

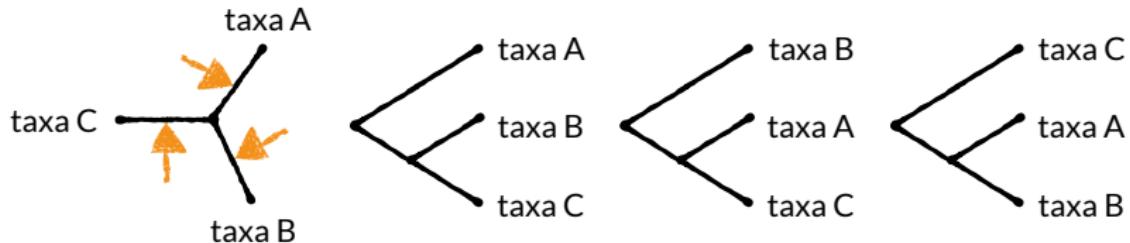


unrooted = 1

rooted = 3

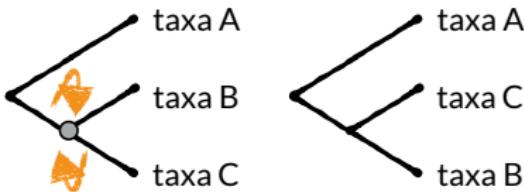
Rooted versus unrooted trees

How many possible trees are there for 3 species?



unrooted = 1

rooted = 3



Note these 2 trees are the same! B and C are more closely related.

What are our data?

Exercise

Character	taxa A	taxa B	taxa C	taxa D	taxa E
Lungs	0	1	1	1	0
Jaws	0	1	1	1	1
Feathers	0	0	1	0	0
Gizzard	0	0	1	1	0
Fur	0	1	0	0	0

- How many possible unrooted or rooted trees are there?
- What do you think the correct rooted tree should be?
- Write down your logic.

Exercise

Character	taxa A	taxa B	taxa C	taxa D	taxa E
Lungs	0	1	1	1	0
Jaws	0	1	1	1	1
Feathers	0	0	1	0	0
Gizzard	0	0	1	1	0
Fur	0	1	0	0	0

- How many possible trees are there?

There are a huge number of possible trees!

# species	# unrooted trees	# rooted trees
3	1	3
4	3	15
5	15	105
6	105	945
7	945	10395
8	10395	135135
9	135135	2027025
10	2027025	34459425

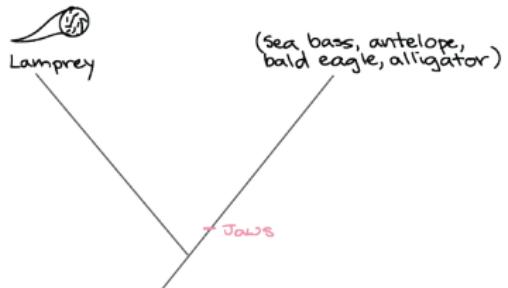
Exercise

Character	taxa A	taxa B	taxa C	taxa D	taxa E
Lungs	0	1	1	1	0
Jaws	0	1	1	1	1
Feathers	0	0	1	0	0
Gizzard	0	0	1	1	0
Fur	0	1	0	0	0

- What do you think the correct tree should be?

Exercise

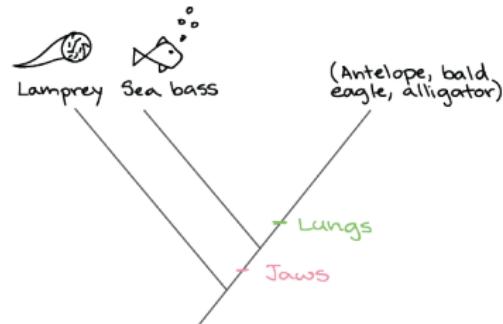
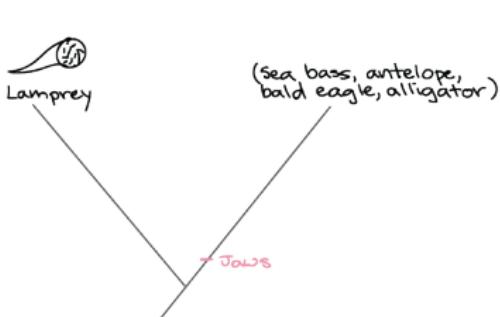
- What do you think the correct tree should be?



A = Lamprey, B = Antelope, C = Bald eagle, D = Alligator, E = Sea bass

Exercise

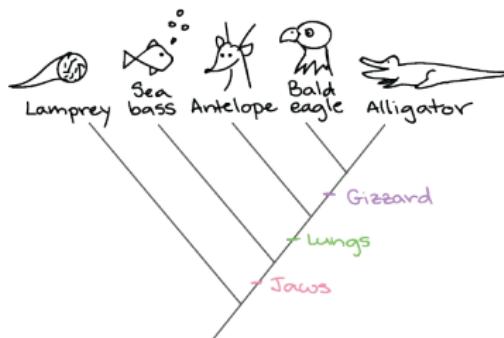
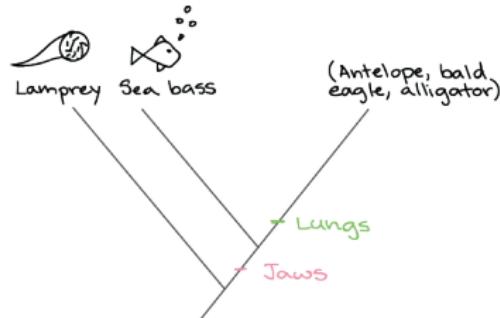
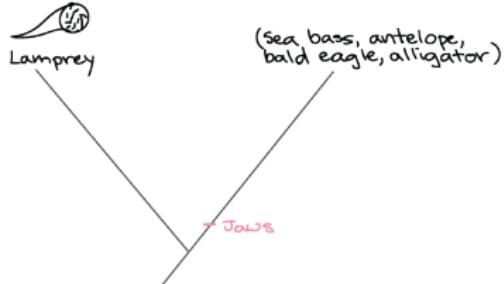
- What do you think the correct tree should be?



A = Lamprey, B = Antelope, C = Bald eagle, D = Alligator, E = Sea bass

Exercise

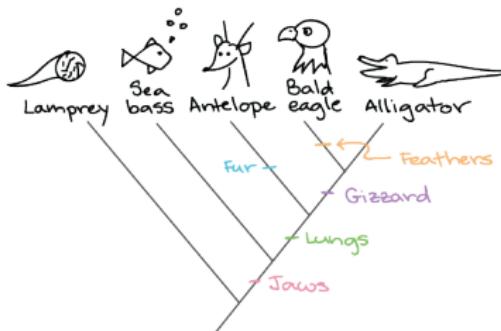
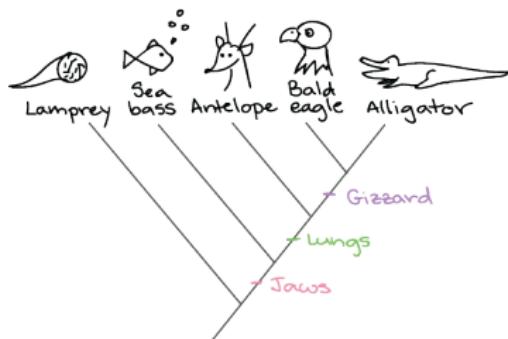
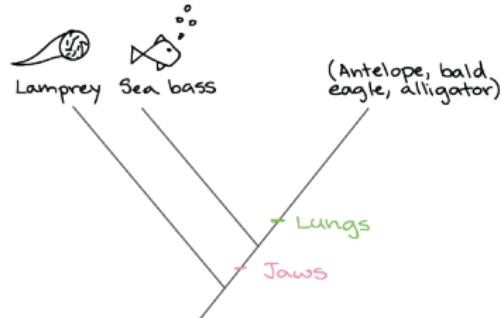
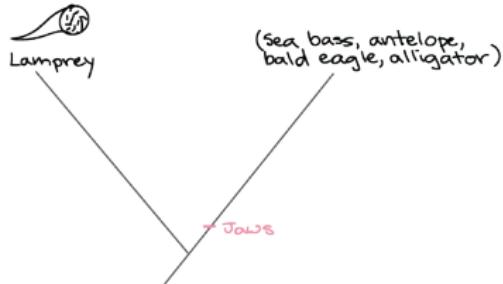
- What do you think the correct tree should be?



A = Lamprey, B = Antelope, C = Bald eagle, D = Alligator, E = Sea bass

Exercise

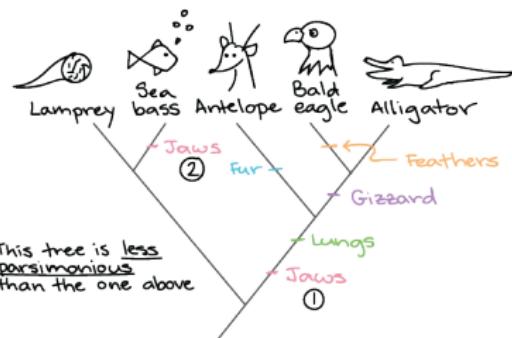
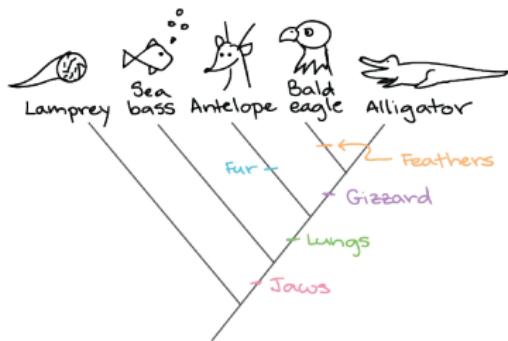
- What do you think the correct tree should be?



A = Lamprey, B = Antelope, C = Bald eagle, D = Alligator, E = Sea bass

Exercise

- What do you think the correct tree should be?



A = Lamprey, B = Antelope, C = Bald eagle, D = Alligator, E = Sea bass

Source [Khan Academy](#)

Exercise

- Write down your logic.

Exercise

- Write down your logic.
→ Most people intuitively assume the tree with the *fewest* changes is correct.

Exercise

- Write down your logic.
 - Most people intuitively assume the tree with the *fewest* changes is correct.
 - This approach to tree building is called **parsimony**.

How do we find the "best" tree?

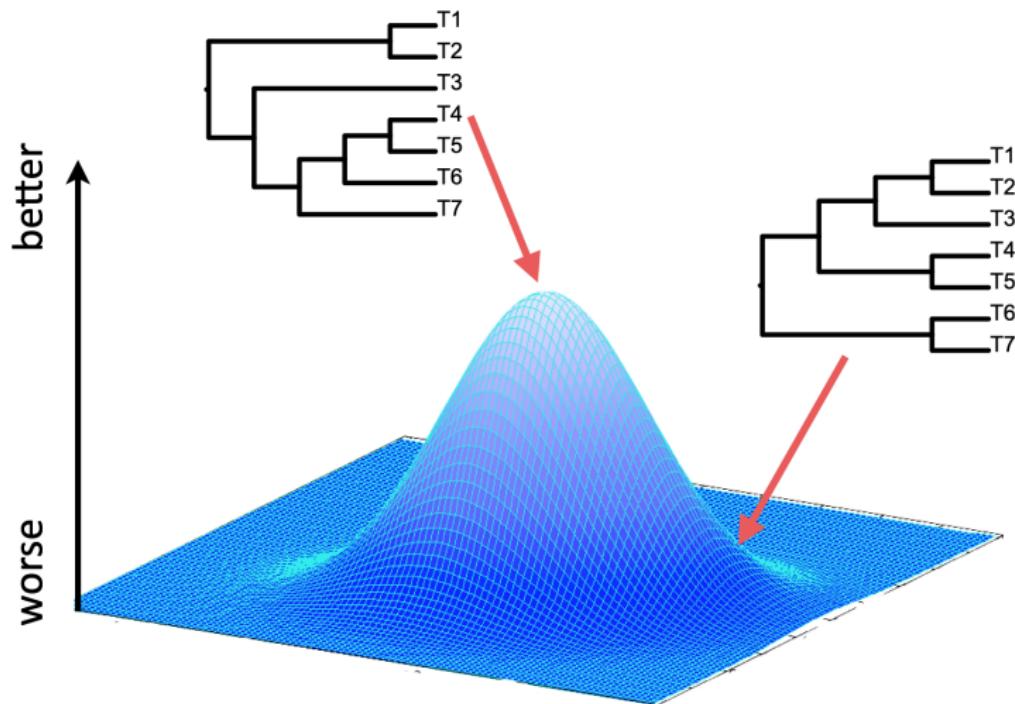


Image source: Tracy Heath

It depends how you measure "best"

Method	Criterion (tree score)
Maximum parsimony	Minimum number of changes
Maximum likelihood	Log likelihood score, optimised over branch lengths and model parameters
Bayesian	Posterior probability, integrating over branch lengths and model parameters

Both maximum likelihood and Bayesian inference are model-based approaches.

Parsimony

In reality, we don't know the true tree.

Maximum parsimony selects the tree (or trees) that require the fewest number of changes.

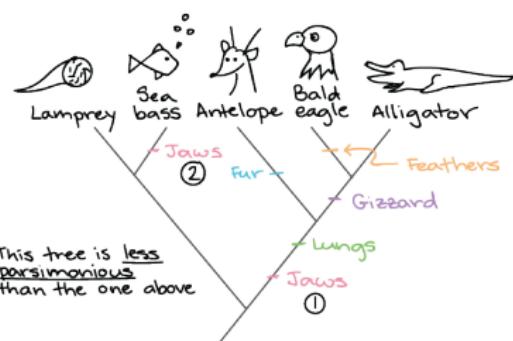
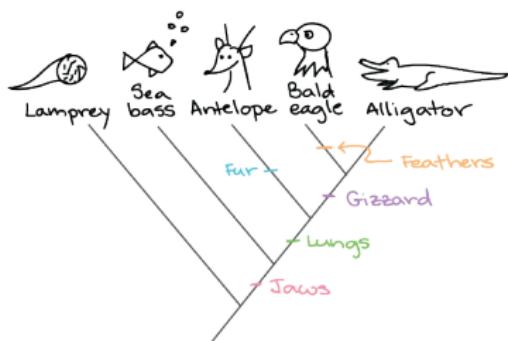
Given two trees, the one minimising the parsimony score (i.e., the minimum number of changes) is the better one.

Parsimony

In reality, we don't know the true tree.

Maximum parsimony selects the tree (or trees) that require the fewest number of changes.

Given two trees, the one minimising the parsimony score (i.e., the minimum number of changes) is the better one.

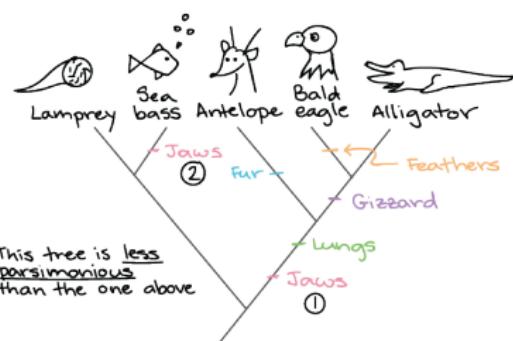
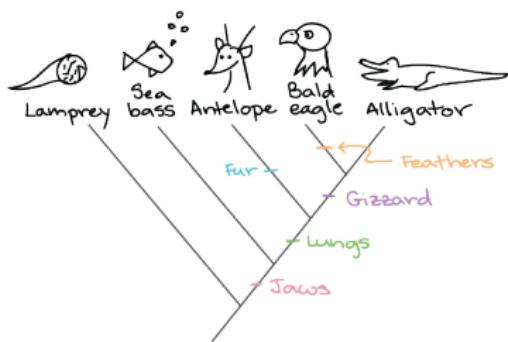


Parsimony

In reality, we don't know the true tree.

Maximum parsimony selects the tree (or trees) that require the fewest number of changes.

Given two trees, the one minimising the parsimony score (i.e., the minimum number of changes) is the better one.



Branch lengths = *number of observed changes or steps*

Parsimony

It is based on the **parsimony principle**: assume simpler explanations are better than complex ones.

Parsimony does not make **explicit** assumptions about the evolutionary process that generated the observed data.
→ However, the method makes **implicit** assumptions.

Exercise