

Phylogenetics

RL-V3 MPP: Course introduction

Rachel Warnock, Laura Mulvey

rachel.warnock@fau.de

May 3, 2022

About this course

Thursdays 16:00–18:00 CET at Henke Str. or via [Zoom](#)

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

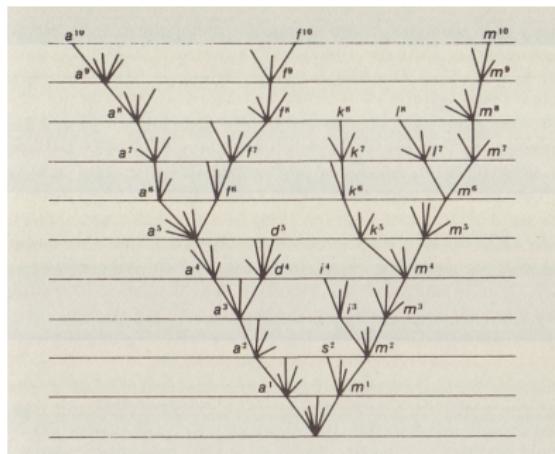
- Tree building
- Substitution models
- Dating trees
- Clock models
- Tree models
- Diversification rates
- Continuous trait evolution

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

Let's introduce ourselves!

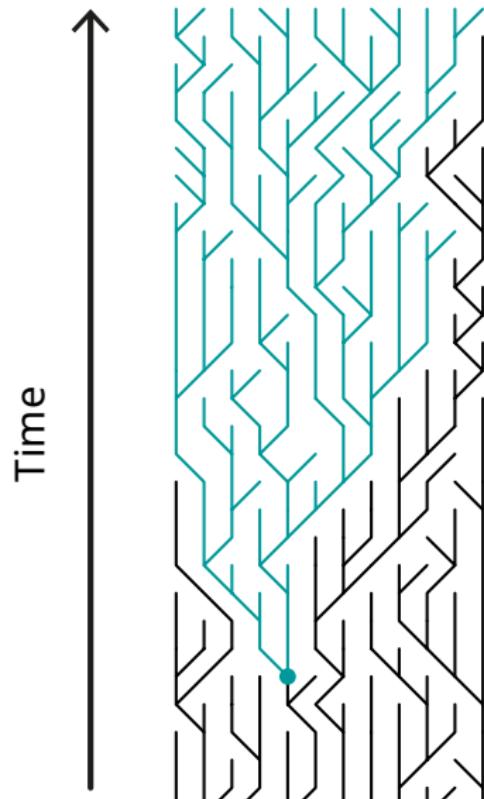
Today's objectives

- Begin “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

What is phylogenetics?

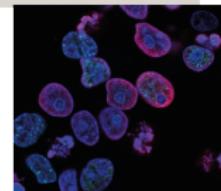
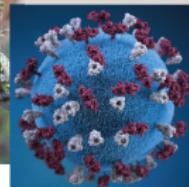
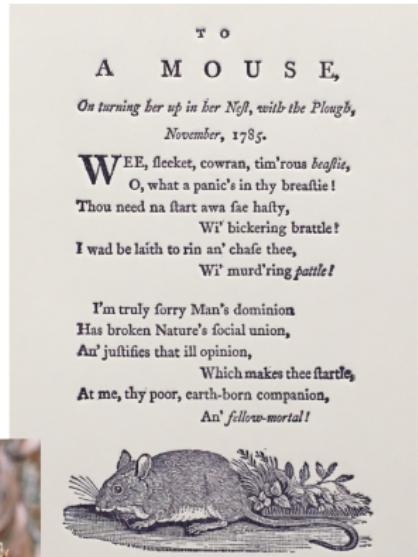
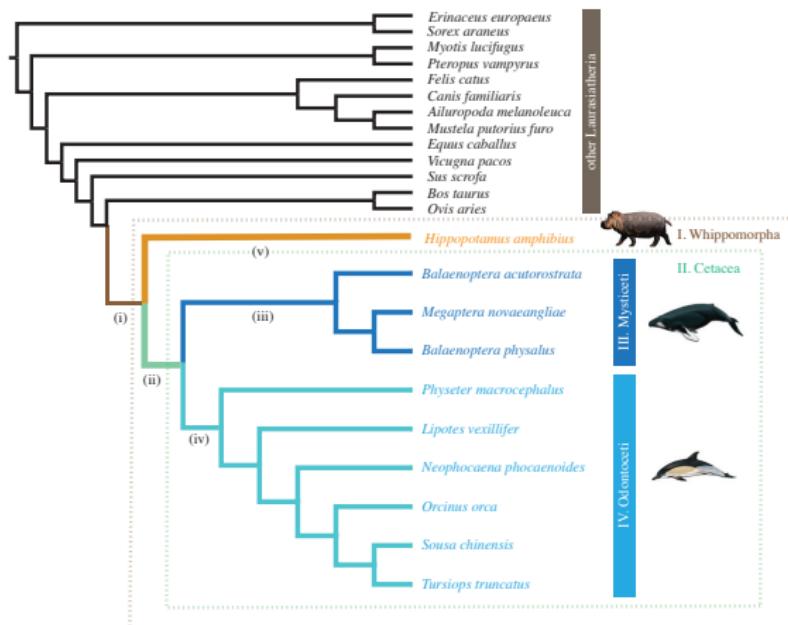


Image source upsplash.org

Why build trees?

How are our favourite species related?
How does phylogeny reflect taxonomy?



Tsagkogeorga et al. (2015) Royal Soc Open Science

Phylogenetic tools can be used to answer many questions

Reconstructing **evolutionary relationships** – what is the sequence of character change?

Estimating **evolutionary times** – how old are sponges? how was the earth's climate during the radiation of flowering plants?

Estimating **evolutionary rates** – have rates of character evolution changed over time?

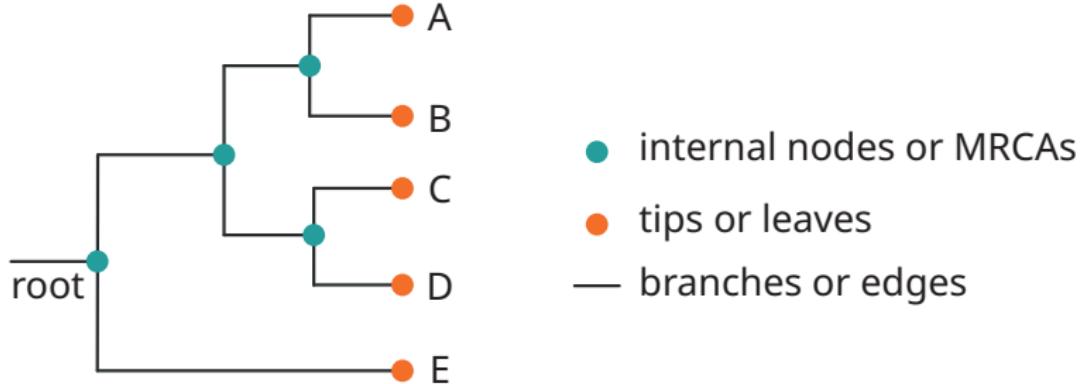
Inferring **evolutionary dynamics** – how do rates of virus evolution change between summer and winter? do smaller species have higher rates of speciation? is speciation rate higher in the tropics?

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?

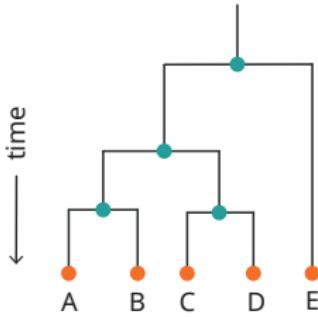
Some basic terms



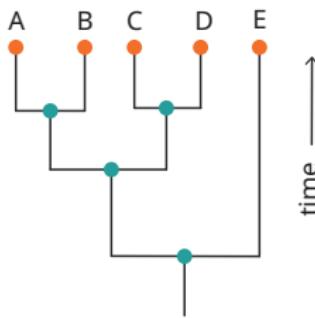
branch lengths = genetic distance OR time

Note: genetic distance = rate x time

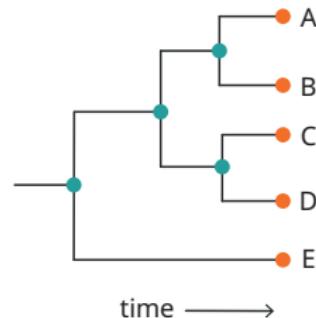
The direction of time



Computer science, maths



Geology



Evolutionary biology

Tip: look for the root!

Rooted versus unrooted trees

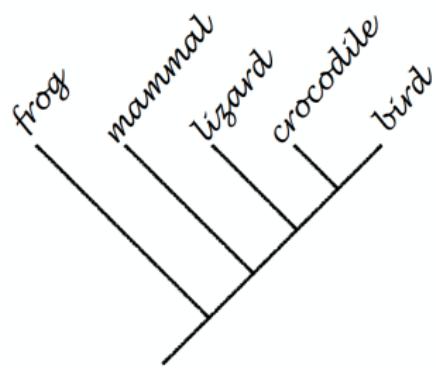
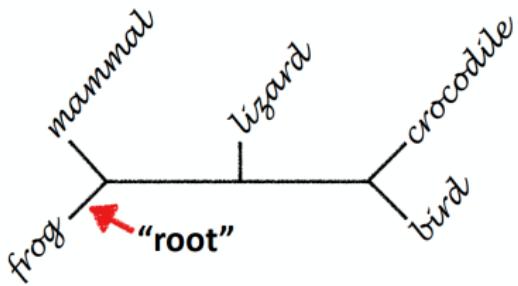


Image source: Phil Donoghue

Rooted versus unrooted trees

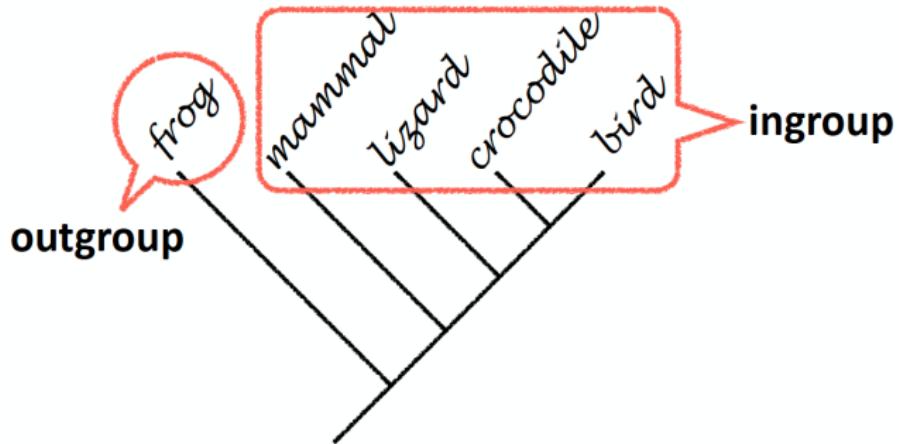
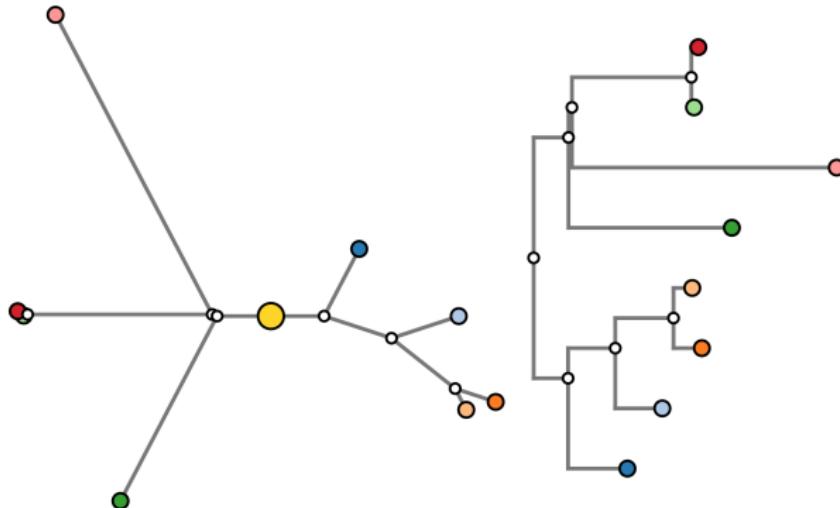


Image source: Phil Donoghue

Rooted versus unrooted trees

By default phylogenies are not rooted.



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

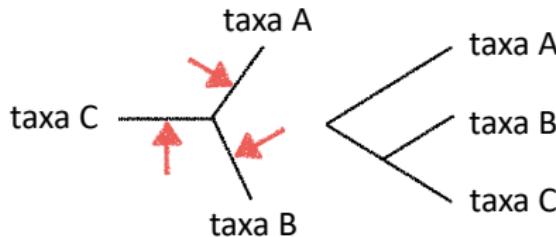
Use Art Poon's [online tool](#) to explore this further. Click [here](#) to learn more about reading trees.

Rooted versus unrooted trees

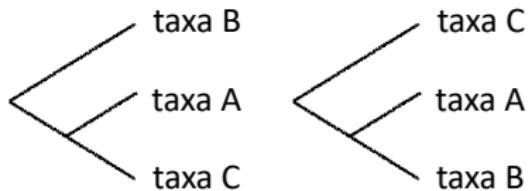
- How many possible trees are there for 3 species?

Rooted versus unrooted trees

- How many possible trees are there for 3 species?



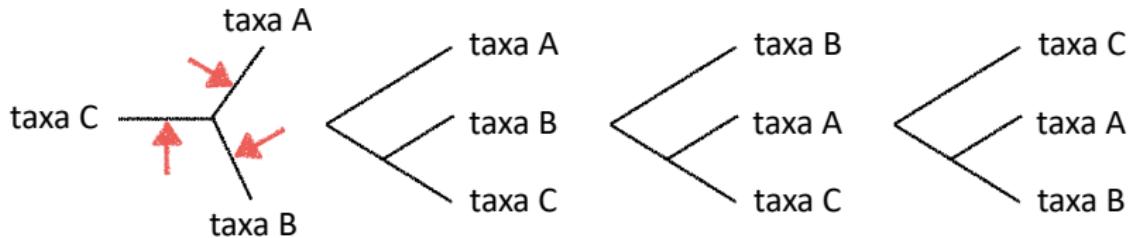
unrooted = 1



rooted = 3

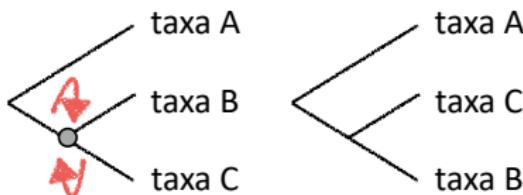
Rooted versus unrooted trees

How many possible trees are there for 3 species?



unrooted = 1

rooted = 3



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

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

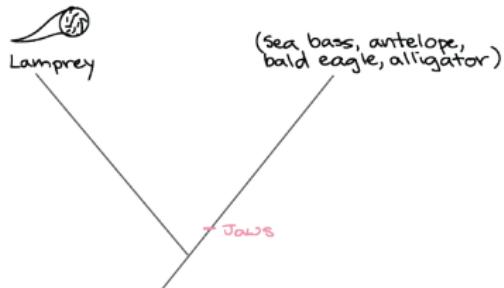
Exercise

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

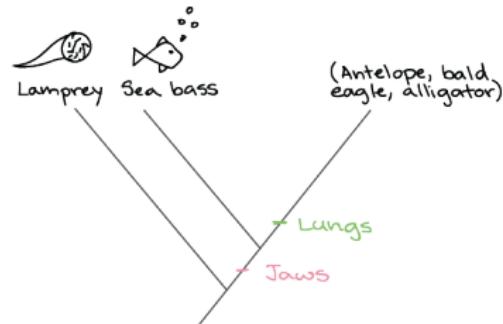
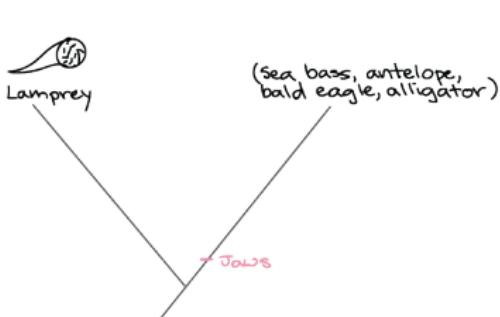
- What do you think the correct tree should be?



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

Exercise

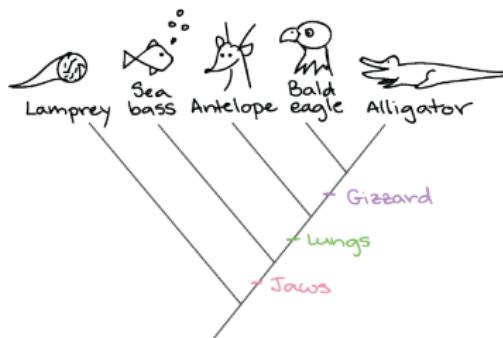
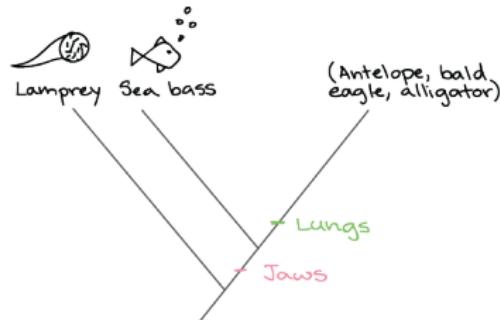
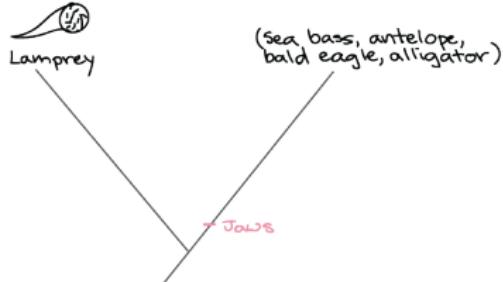
- What do you think the correct tree should be?



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

Exercise

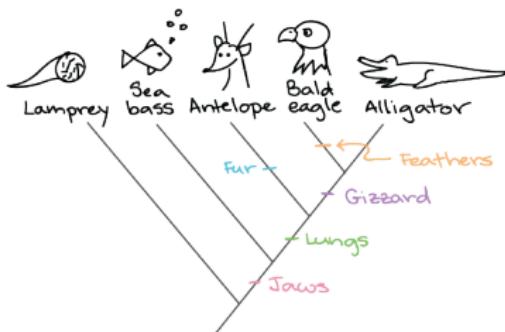
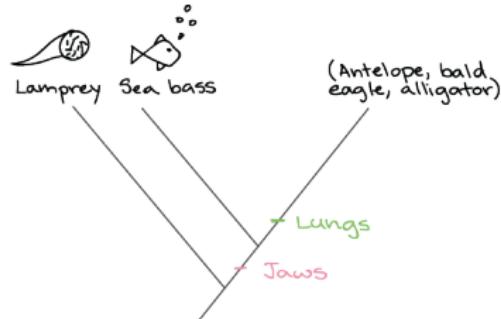
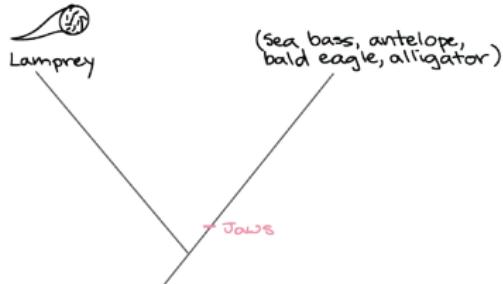
- What do you think the correct tree should be?



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

Exercise

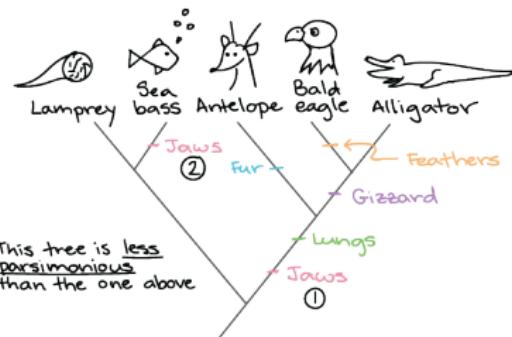
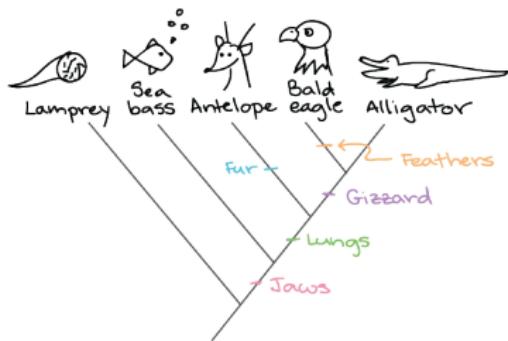
- What do you think the correct tree should be?



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?

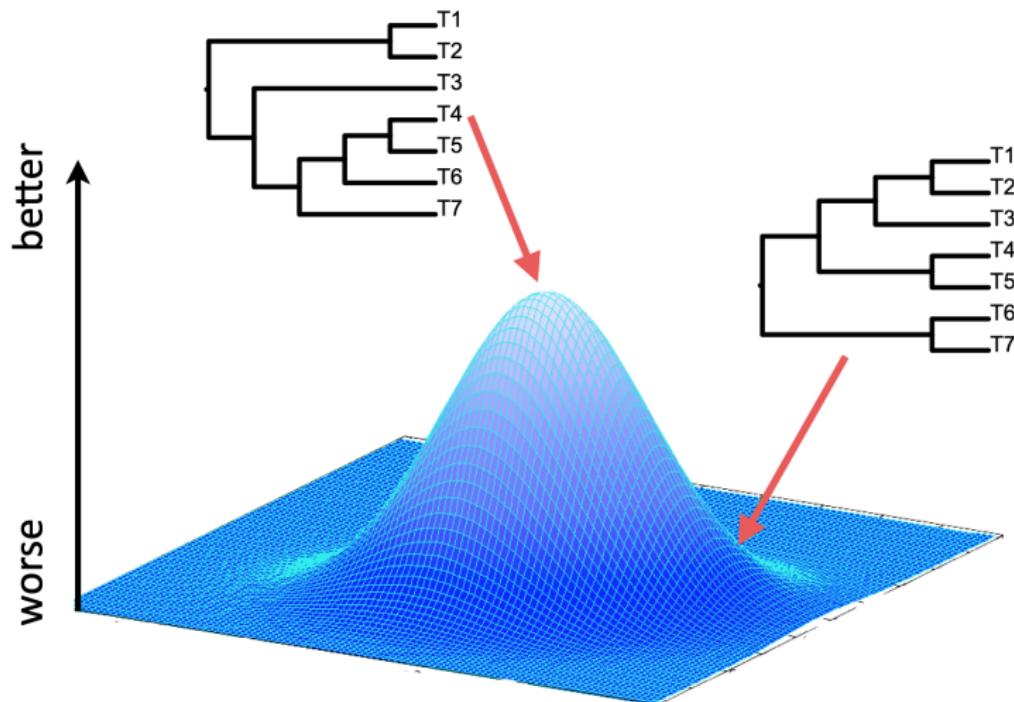


Image source: Tracy Heath

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

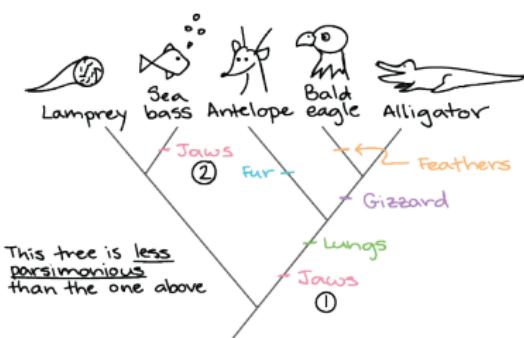
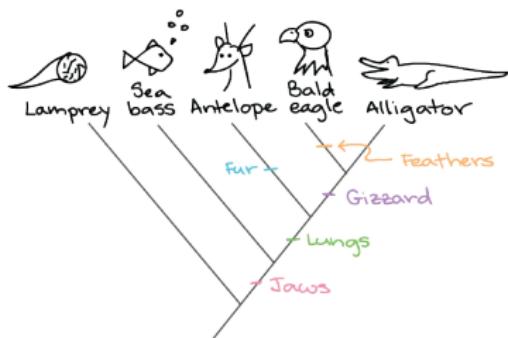
Maximum parsimony selects the tree (or trees) that require the fewest number of changes.

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

Parsimony

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

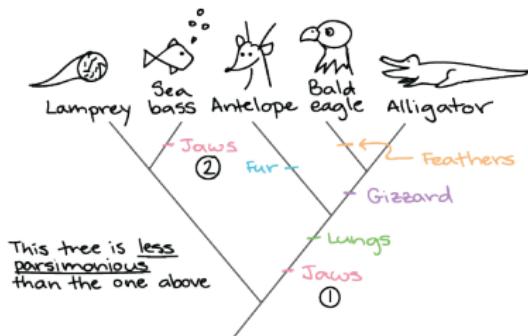
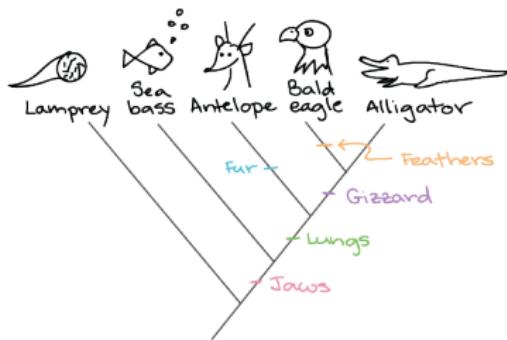
Given two trees, the one minimising the parsimony score (i.e. the minimum number of changes) is the better one.



Parsimony

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.

Convergence or homoplasy

Homoplasy: a trait that is found in two species, but not in their common ancestor.

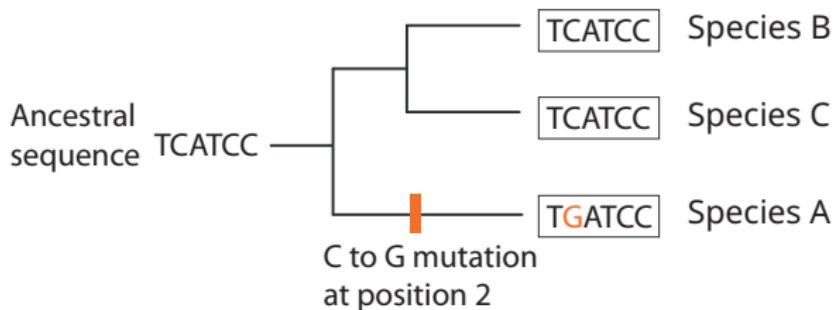


The bluebird, Pterosaur (extinct) and fruit bat: 3 different vertebrates independently lightened bones and transformed hands into wings.

Image source: [Convergent Evolution: an introduction](#)

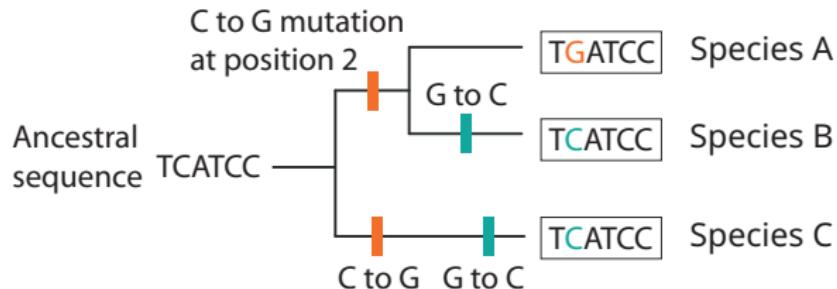
Molecular convergence

If we assume the simplest solution is correct, this could mislead our inference if the underlying process is more complex.



Molecular convergence

If we assume the simplest solution is correct, this could mislead our inference if the underlying process is more complex.



Parsimony

When we build a tree using parsimony and observe convergence, **ad hoc** explanations (e.g. convergence, reversals) are required to explain the patterns.

In the case of birds, pterosaurs and bats, we know based on other anatomical features that these taxa are distantly related but convergence can also interfere with our ability to recover the correct tree. In fact, this is very common.

Parsimony

Parsimony has been demonstrated to be statistically inconsistent.

An estimator is consistent if it is guaranteed to get the correct answer with an infinite amount of data.

Felsenstein (1978) demonstrated that in some situations, parsimony is inconsistent, i.e. it will recover the wrong tree, even with an infinite amount of data.

Long branch attraction

If you have long branches, or higher rates of evolution, the probability of a misleading parsimony due to convergence is much higher.

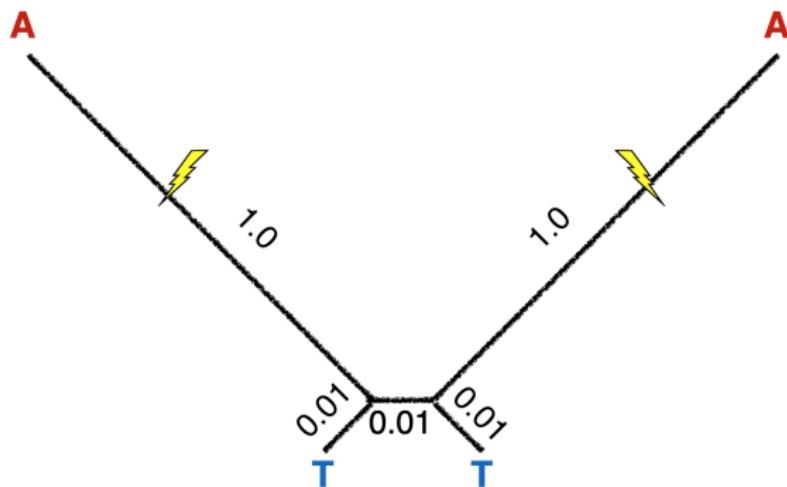


Image source: Tracy Heath

Long branch attraction

Parsimony is almost guaranteed to get the tree below wrong. It will incorrectly place two long branches together as sister lineages.

More data will make the problem worse, making this approach statistically inconsistent.

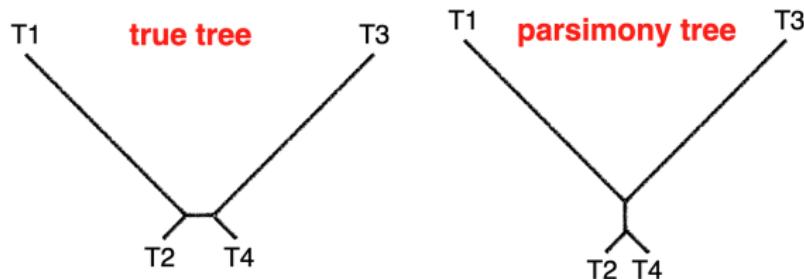
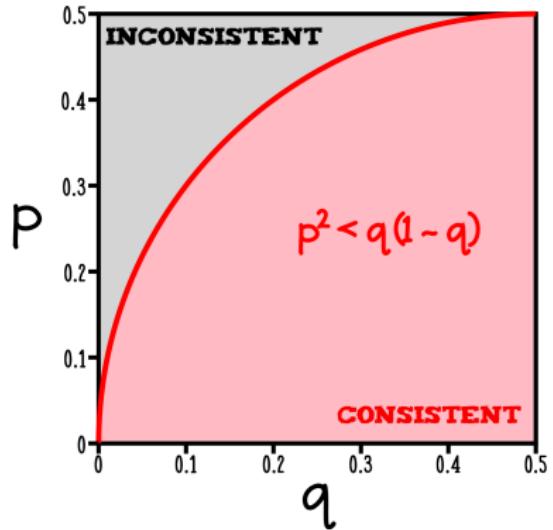
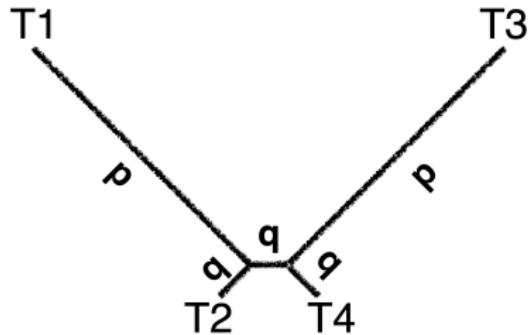


Image source: Tracy Heath

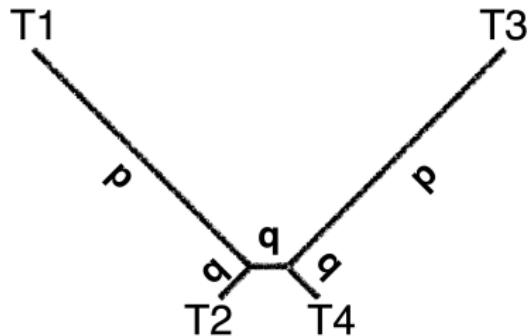
Long branch attraction



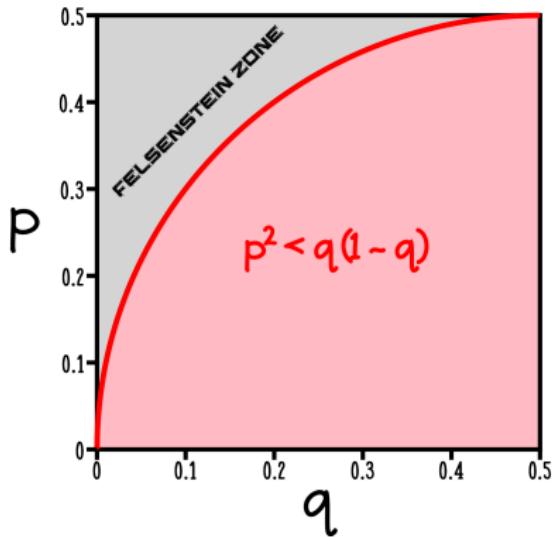
Here, the branch lengths represent probability (p, q) of change along that branch.

Felsenstein, *Inferring Phylogenies*, (2004), Image source: Tracy Heath

Long branch attraction



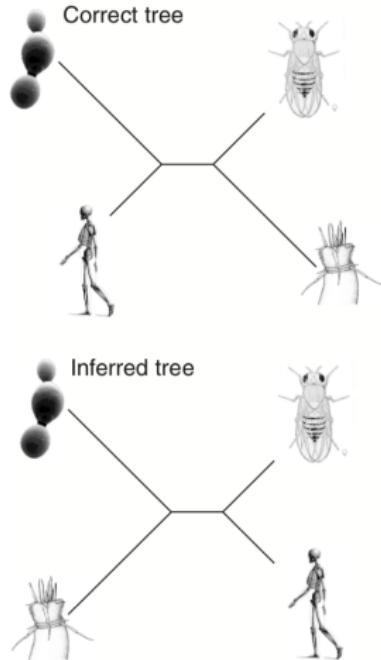
Here, the branch lengths represent probability (p, q) of change along that branch.



Felsenstein, *Inferring Phylogenies*, (2004), Image source: Tracy Heath

Long branch attraction

The relationship between nematodes, arthropods and chordates was misunderstood for a long time.

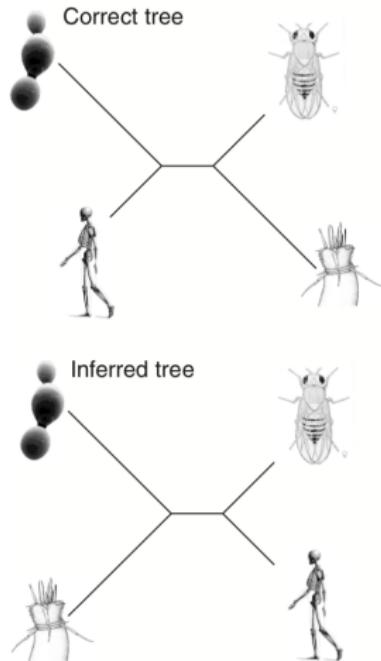


- outgroup = yeast
- Ecdysozoa
arthropods + nematodes,
ex. vertebrates
- Coelomata
arthropods + vertebrates,
ex. nematodes

Felsenstein (1978) *Systematic Zoology*, Image: Telford et al. (2005) *Current Biology*

Long branch attraction

The relationship between nematodes, arthropods and chordates was misunderstood for a long time.



- outgroup = yeast
- Ecdysozoa
arthropods + nematodes,
ex. vertebrates
- Coelomata
arthropods + vertebrates,
ex. nematodes

Note: this issue can affect all tree building methods!

Parsimony: advantages and disadvantages

The greatest advantage of parsimony is its beautiful simplicity

Computationally fast

Often produces sensible results

Parsimony: advantages and disadvantages

The greatest advantage of parsimony is its beautiful simplicity

Computationally fast

Often produces sensible results

Some argue that parsimony is assumption free

Others argue that parsimony does make assumptions, even if we don't know what they are

Yang (2014) Molecular Evolution: A Statistical Approach

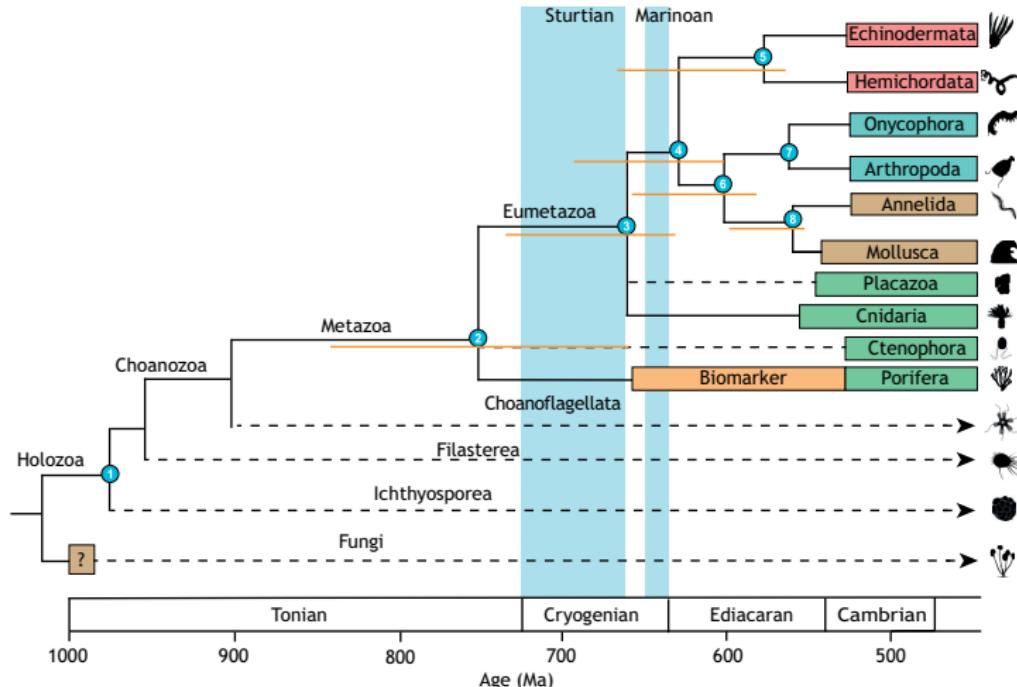
Parsimony vs. model-based approaches

Model-based approaches assume an explicit model of molecular or morphological evolution.

If evolutionary distance is small, model based approaches and parsimony typically recover the same tree.

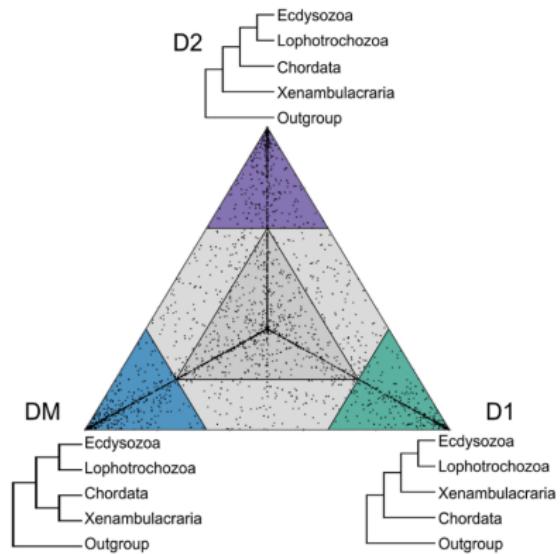
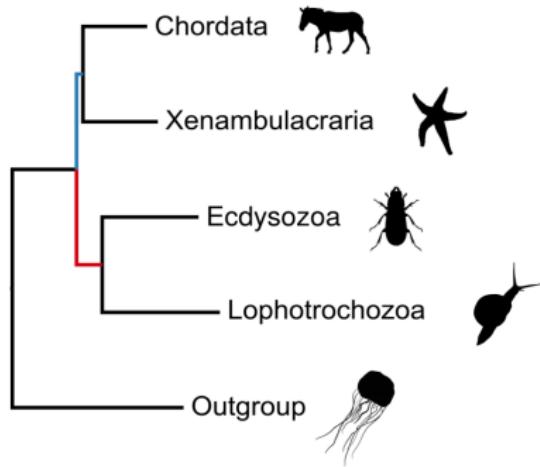
As distance increases, the amount of homoplasy (i.e. parallel or convergent changes) also increases, parsimony is more likely to recover the wrong tree.

Deuterostomes vs. protostomes - short branch attraction?



Erwin et al. (2020) Development

Deuterostomes vs. protostomes - short branch attraction?



Kapli et al. (2021) *Science Advances*

Take home

Parsimony is simple and intuitive but makes **implicit** assumptions about the evolutionary process.

In the following weeks we'll explore model-based approaches – these are more flexible and make **explicit** assumptions → we must take care to understand these!

Suggested listening / reading

Check out this fascinating [interview](#) with Joseph Felsenstein by Mary Kuhner! Joe played an important role in establishing the field of statistical phylogenetics.

Kapli et al. (2021) Lack of support for Deuterostomia prompts reinterpretation of the first Bilateria. *Science Advances*.

Homework

Try installing the software [RevBayes](#)