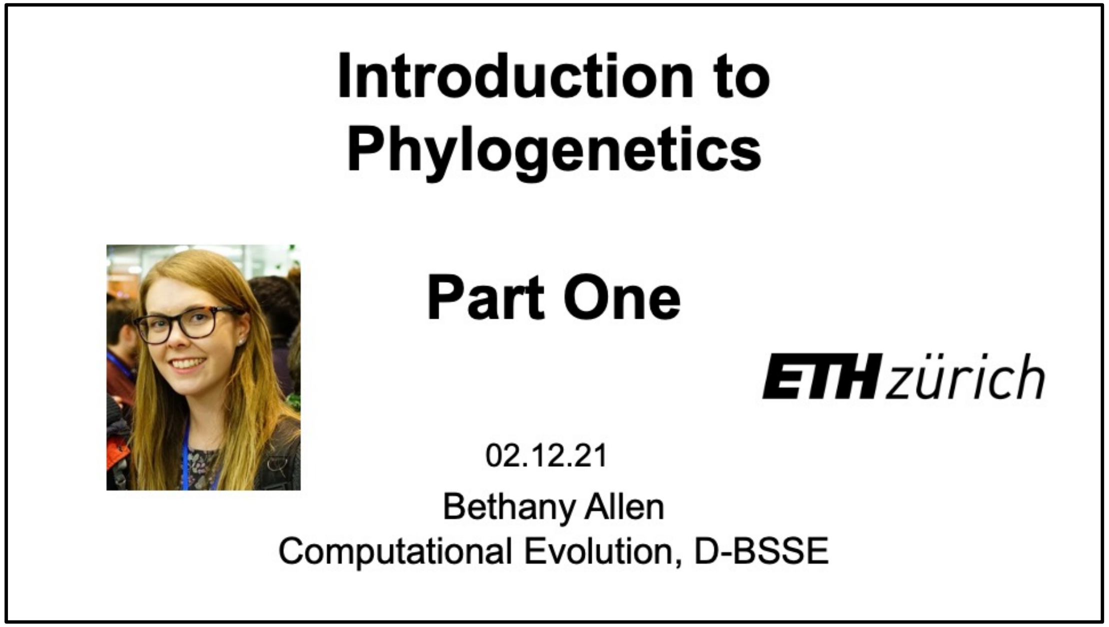


Introduction to Phylogenetics

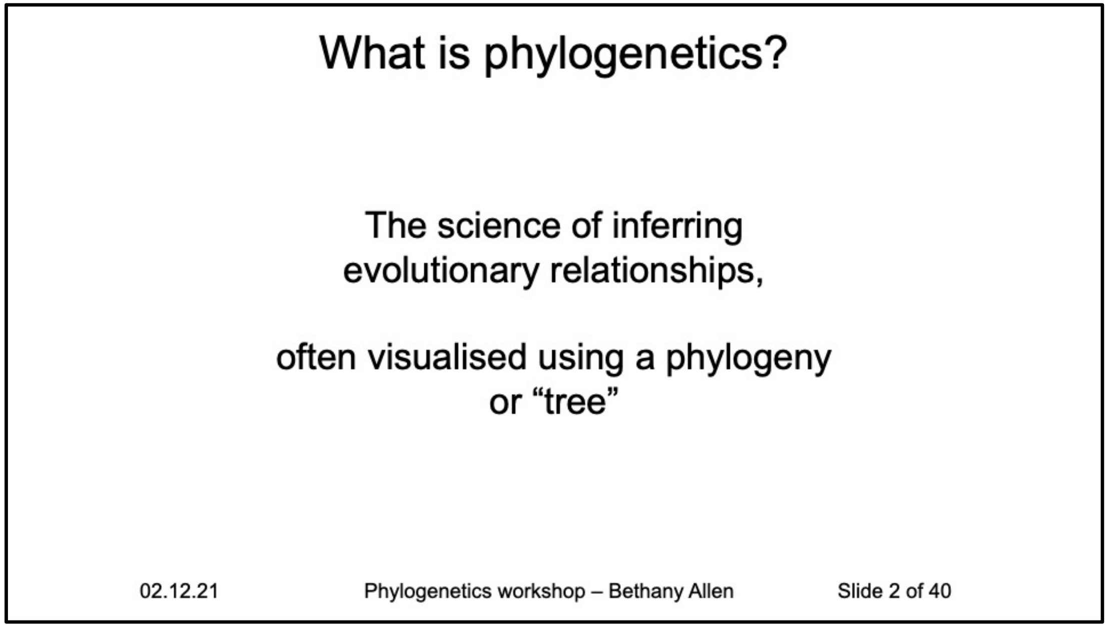
Part One

ETH zürich

02.12.21

Bethany Allen
Computational Evolution, D-BSSE

My background is in palaeontology, so my interests lie in understanding evolutionary processes in deep time – this is the perspective in which this presentation is given.

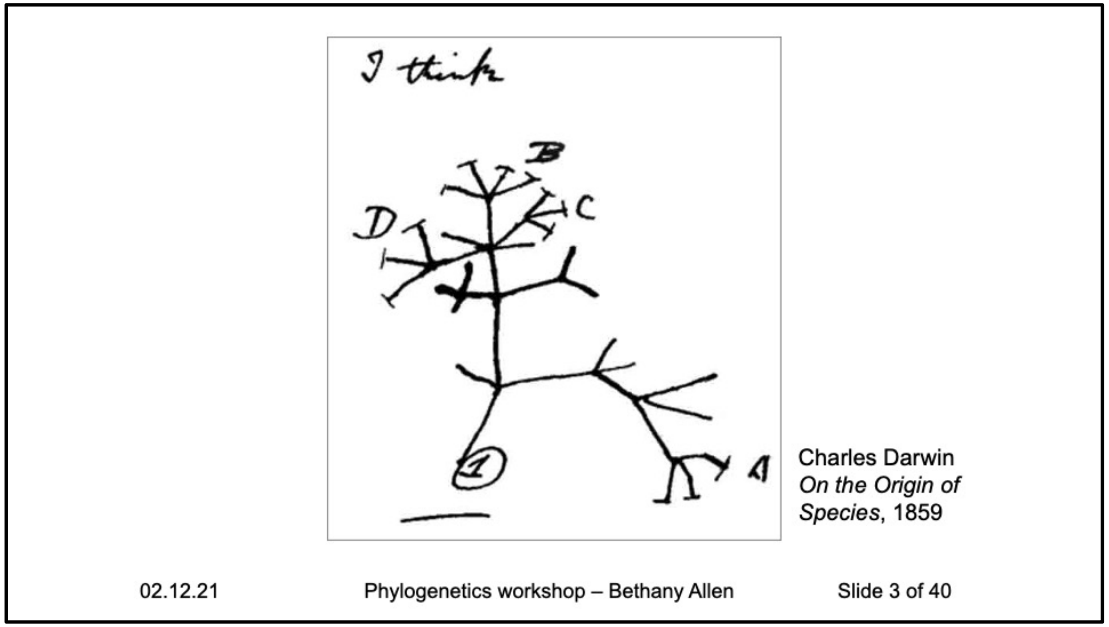


What is phylogenetics?

The science of inferring
evolutionary relationships,

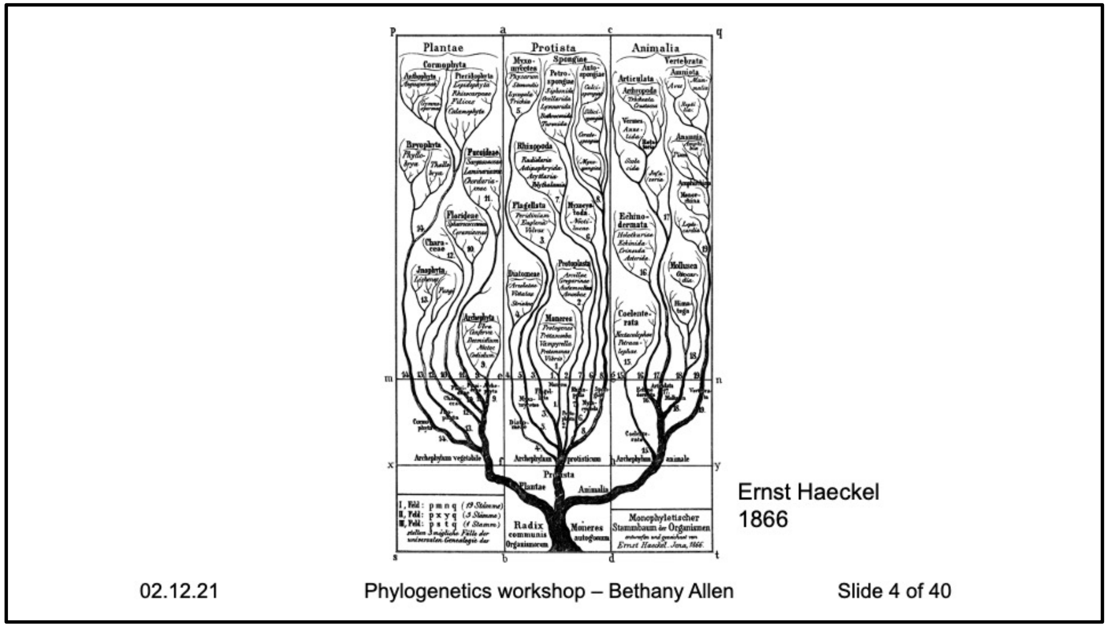
often visualised using a phylogeny
or “tree”

Tree-based terminology is pervasive in phylogenetics, so if in doubt, think about what the term might mean in gardening!



Charles Darwin
On the Origin of Species, 1859

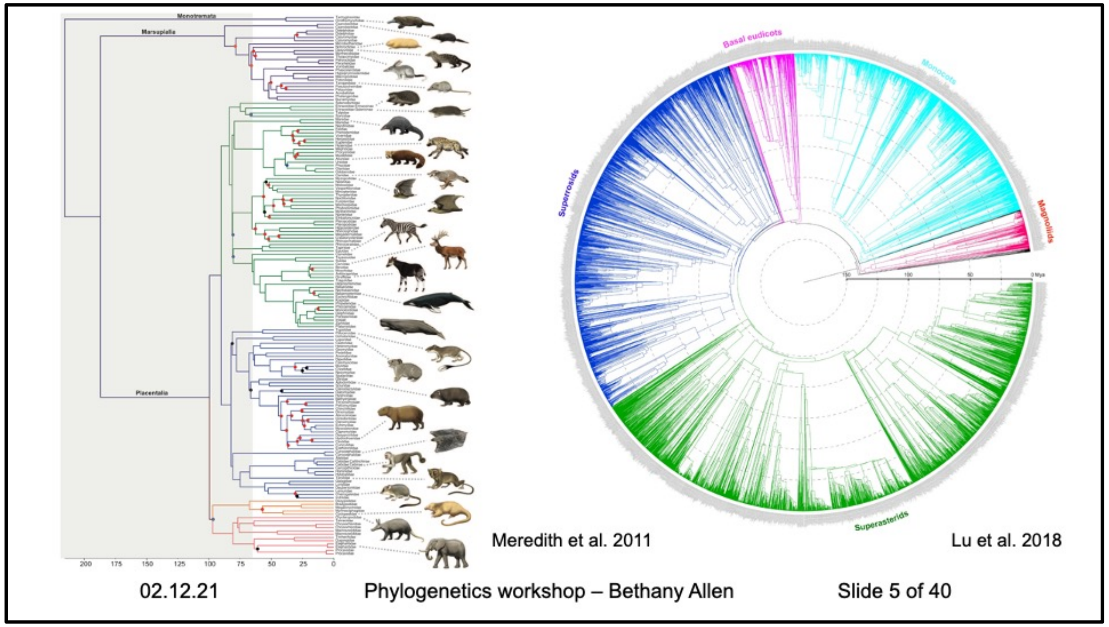
Darwin described natural selection, and formed the basis of evolutionary theory, in this famous book. It includes this figure, captioned “I think”, and shows evolution as a branching process linking together different species. Under the assumption that all life has a universal common ancestor, it follows that all living things can be described in this way, with some having diverged more recently, and others diverging further in the past.



Ernst Haeckel
1866

02.12.21 Phylogenetics workshop – Bethany Allen Slide 4 of 40

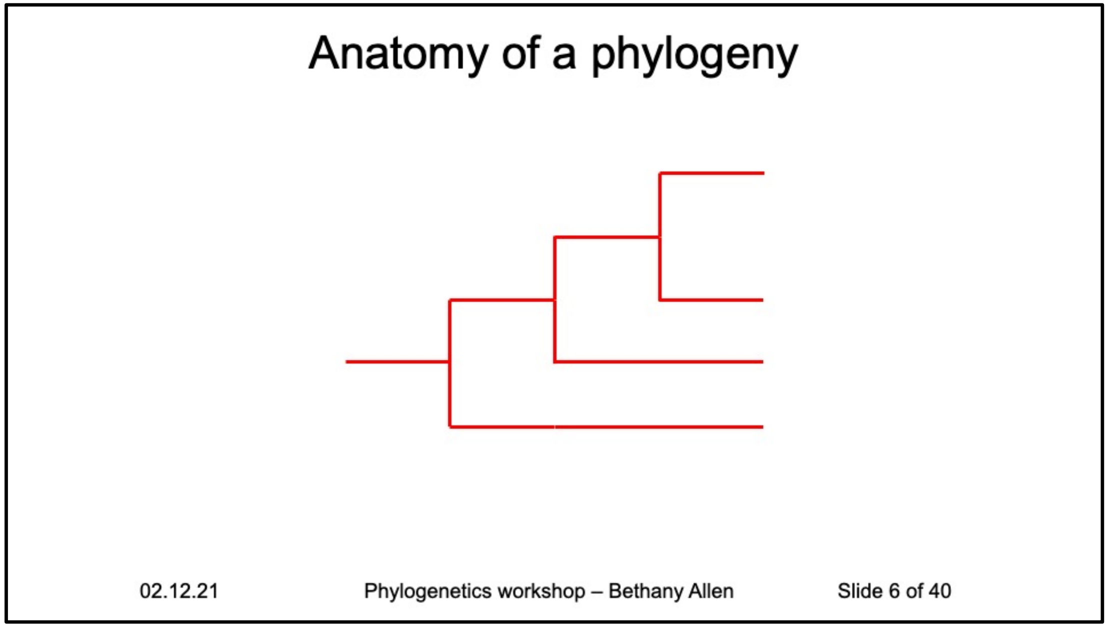
Haeckel was the first person to call such figures ”phylogenies”. Famous for his artistry as much as his science, Haeckel further developed the idea of making these diagrams more formal and scientific.



Meredith et al. 2011

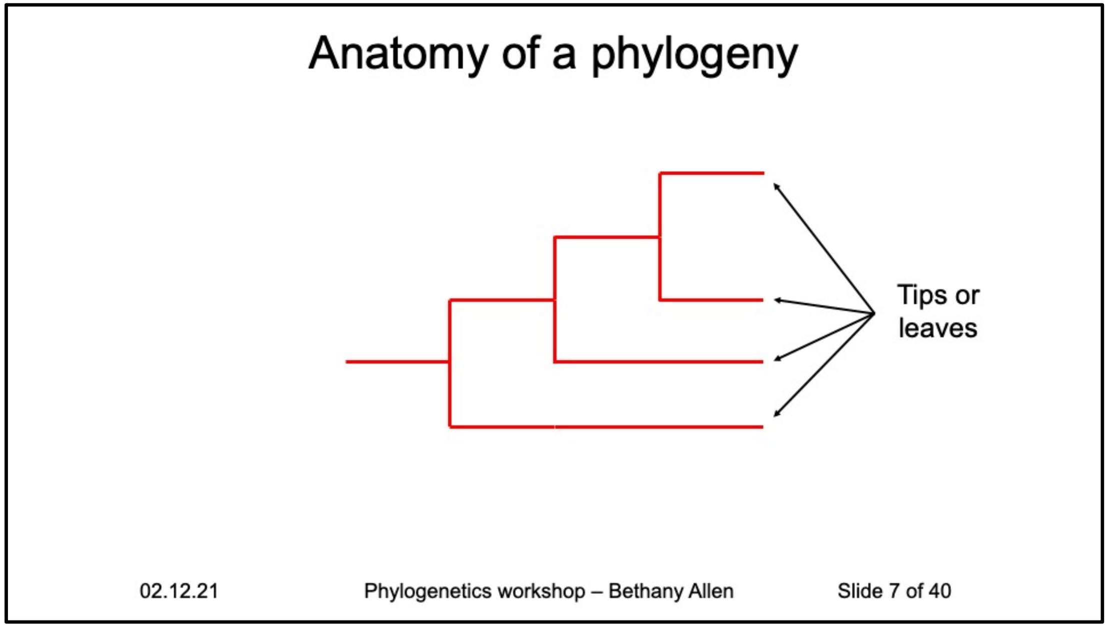
Lu et al. 2018

Since then, phylogenetics has seen considerable development as a scientific field. Here are two examples of recently published phylogenies, on the left for all living mammals, and on the right for all angiosperms in China. These show the two most popular forms for figures, the classic phylogram and the (more aesthetically pleasing?) circular/fanned phylogeny.



Anatomy of a phylogeny

This is a simple phylogeny which I will use to define key features and discuss some important points.

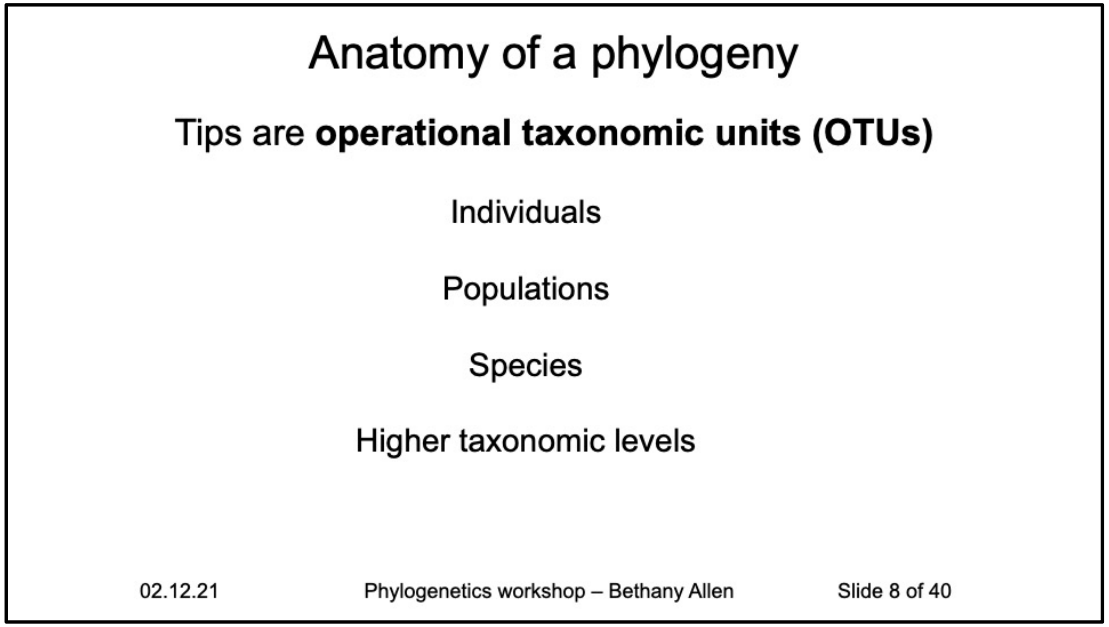


Anatomy of a phylogeny

Tips or leaves

02.12.21 Phylogenetics workshop – Bethany Allen Slide 7 of 40

Tips correspond to each of the entities being compared within the phylogeny.



Anatomy of a phylogeny

Tips are **operational taxonomic units (OTUs)**

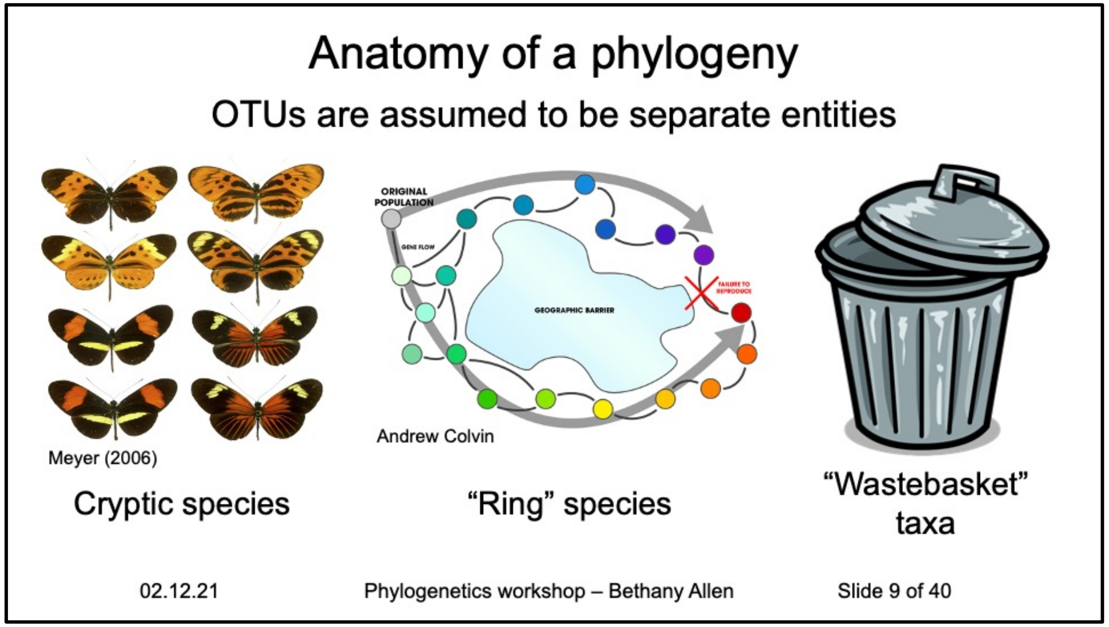
Individuals

Populations

Species

Higher taxonomic levels

Depending on the aim of the research, tips can correspond to many different things. For example, many people in my research group are interested in viral genome evolution, and use viral sequences taken from individual patients to track disease spread over time. Samples can be drawn from different populations to look at genetic differences within a species across space. Species are the typical OTUs we think of in phylogenies, but it’s worth remembering that they are often represented in the phylogeny by molecular data collected from a single individual of that species; morphological data is more likely to be averaged across, or supplemented using, multiple specimens. This is also true of higher taxa, which are often represented by a single species, and therefore potentially by a single individual.



Anatomy of a phylogeny

OTUs are assumed to be separate entities

ORIGINAL POPULATION
GENE FLOW
GEOGRAPHIC BARRIER
FAILURE TO REPRODUCE

Meyer (2006)

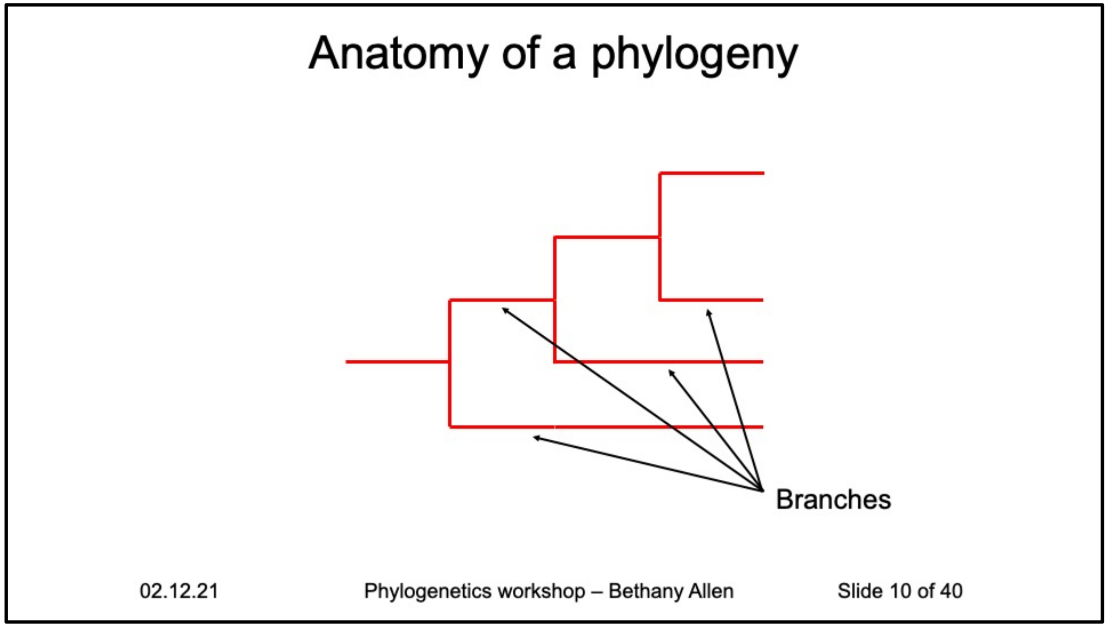
Andrew Colvin

Cryptic species

“Ring” species

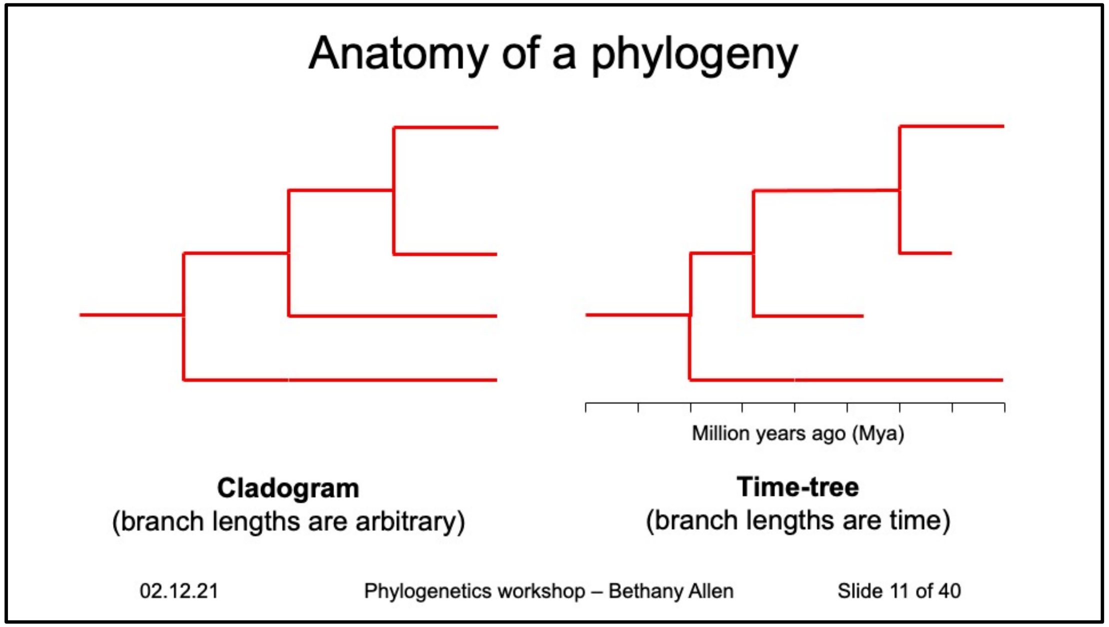
“Wastebasket” taxa

We consider our tips to be internally cohesive as a biological entity but separated from other such entities. Is this a realistic assumption? Biology is messy and not easy to summarise in this way. Some examples of violations include cryptic species, which are reproductively separate but look very similar to each other. These are impossible to identify or do anything about in the fossil record. There are also ring species: these exist as a series of local populations spread across a large area, with adjacent populations able to interbreed successfully, while geographically distant populations cannot. “Wastebasket” taxa are common in the fossil record, and consist of species which are difficult to categorise correctly, often because they are very similar morphologically, and therefore represent a placeholder for uncertainty rather than a cohesive, evolutionary entity.



Anatomy of a phylogeny

Branches are the horizontal parts of the tree.

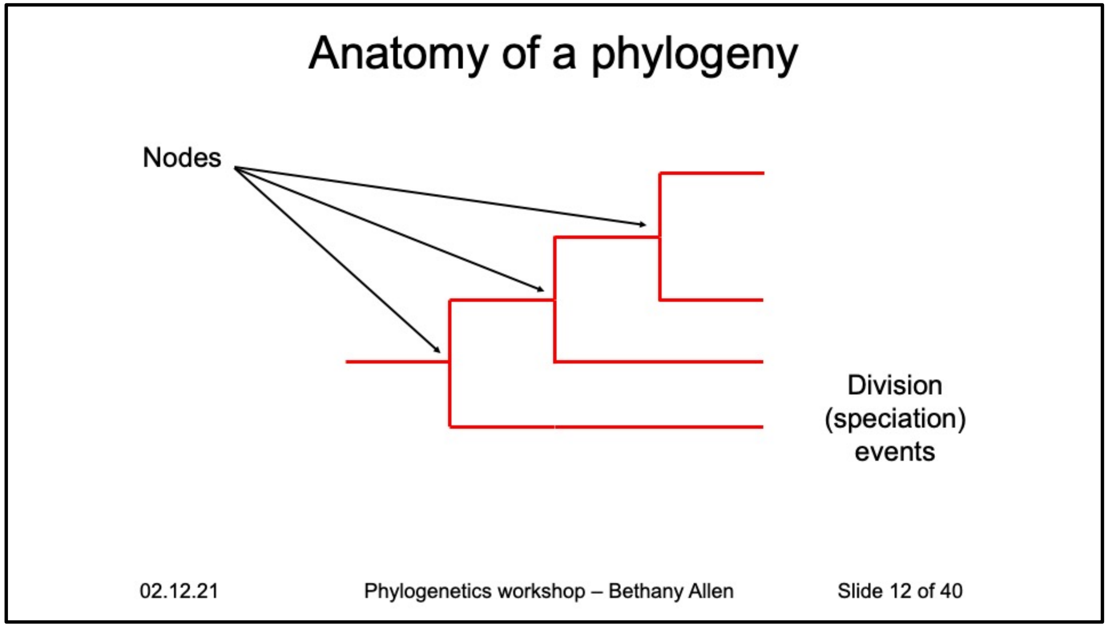


Anatomy of a phylogeny

Cladogram
(branch lengths are arbitrary)

Time-tree
(branch lengths are time)

Cladograms only indicate the relationships between tips, so branch lengths are arbitrary. However, branches can also be scaled to evolutionary change, and often then used to reflect absolute time. More on this later.

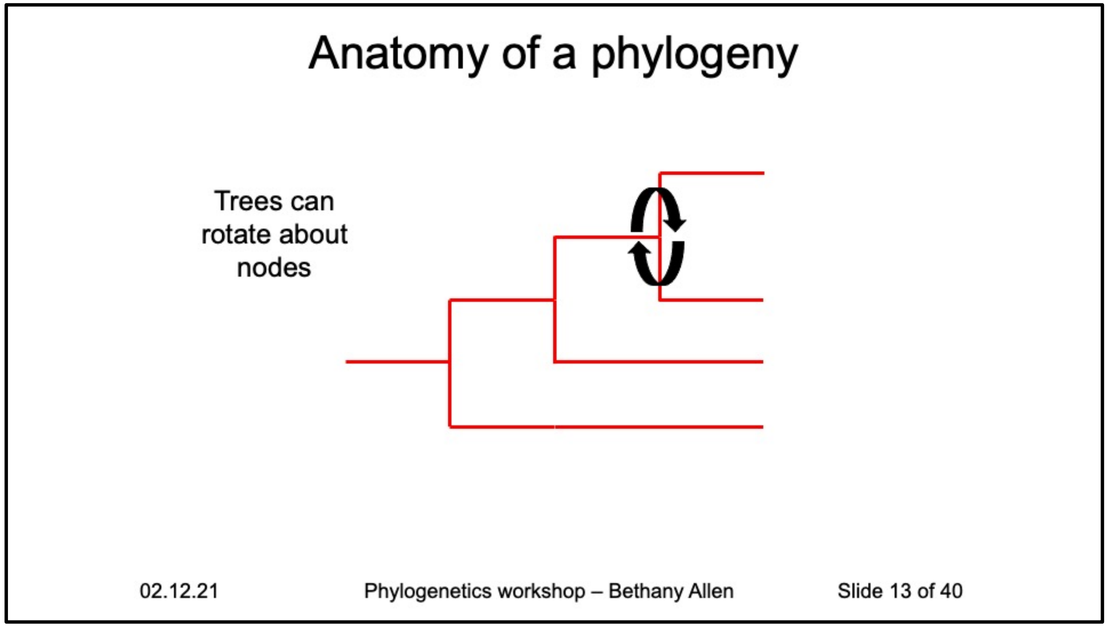


Anatomy of a phylogeny

Nodes

Division (speciation) events

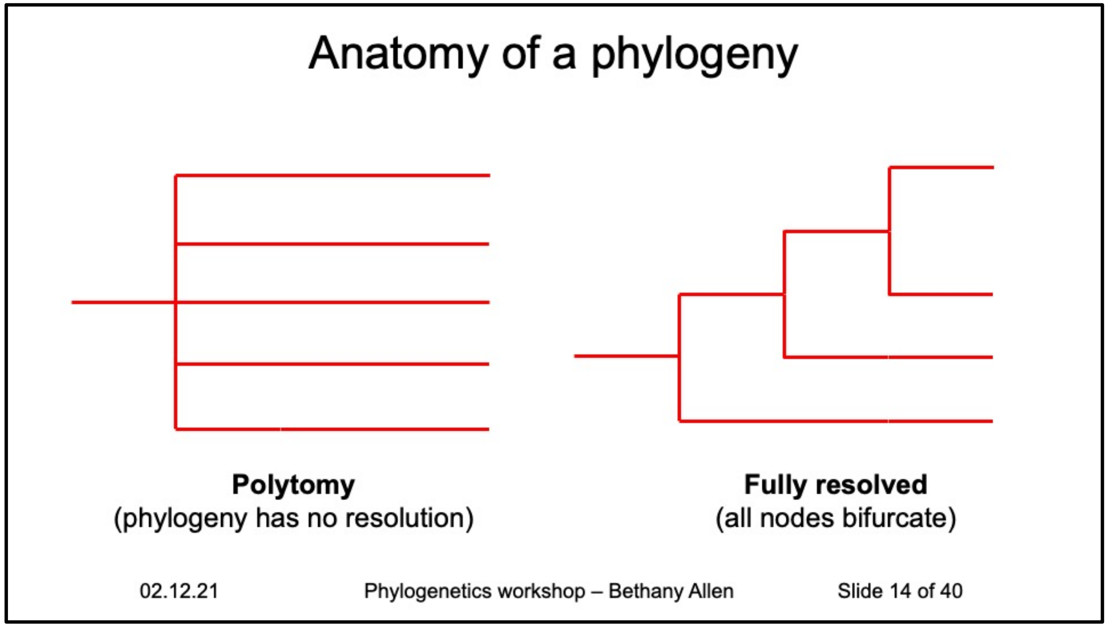
Evolutionary trajectories join (or split, depending on your direction of interest) at nodes.



Anatomy of a phylogeny

Trees can rotate about nodes

A brief note on how phylogenies are presented – all trees can be rotated about any node, so the order in which tips are presented in a tree is usually chosen for ease of labelling or aesthetic purposes.

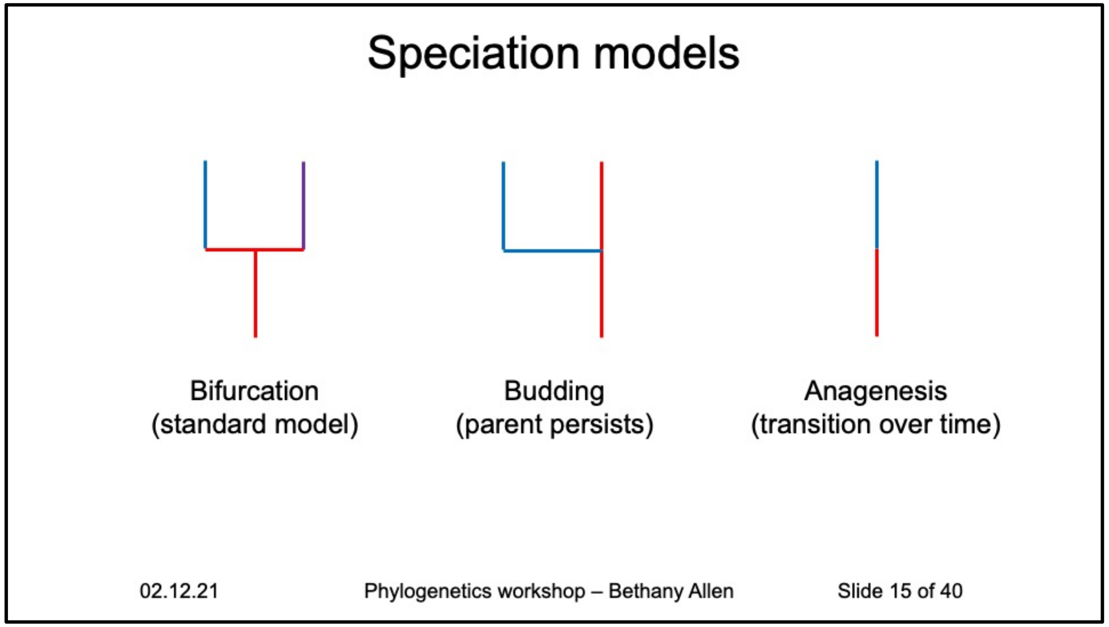


Anatomy of a phylogeny

Polytomy
(phylogeny has no resolution)

Fully resolved
(all nodes bifurcate)

Most phylogenetic analyses aim to present a fully bifurcating tree, where only two lineages split at any single point. However, poor signal in the data can sometimes prevent this level of resolution. Uncertainty in the order of bifurcations can be expressed by collapsing these events into a polytomy, informally known as a “rake”.



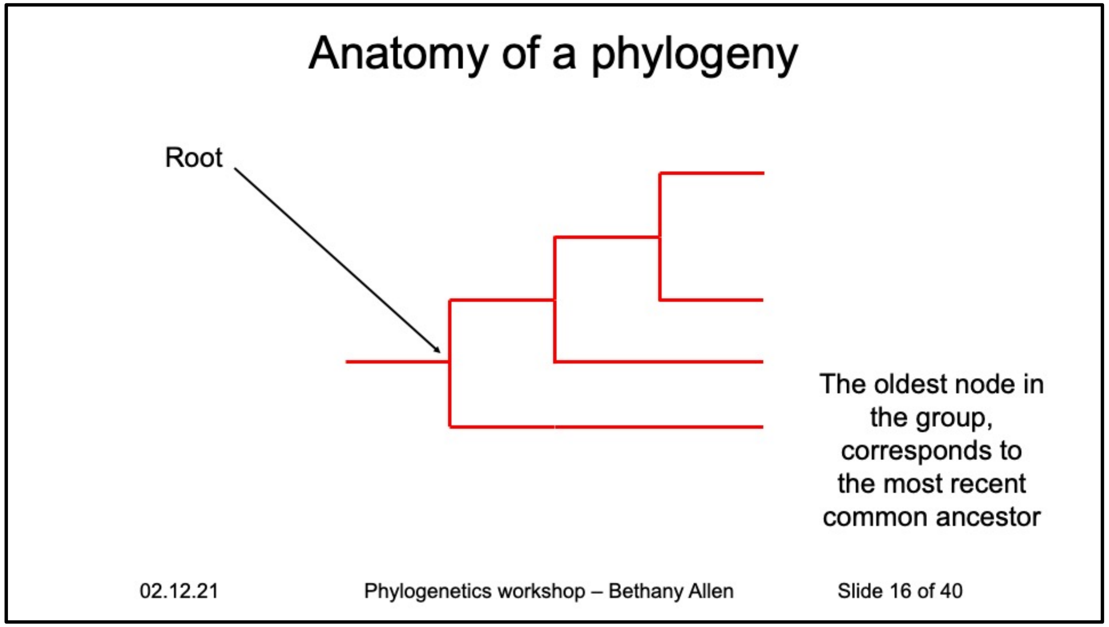
Speciation models

Bifurcation
(standard model)

Budding
(parent persists)

Anagenesis
(transition over time)

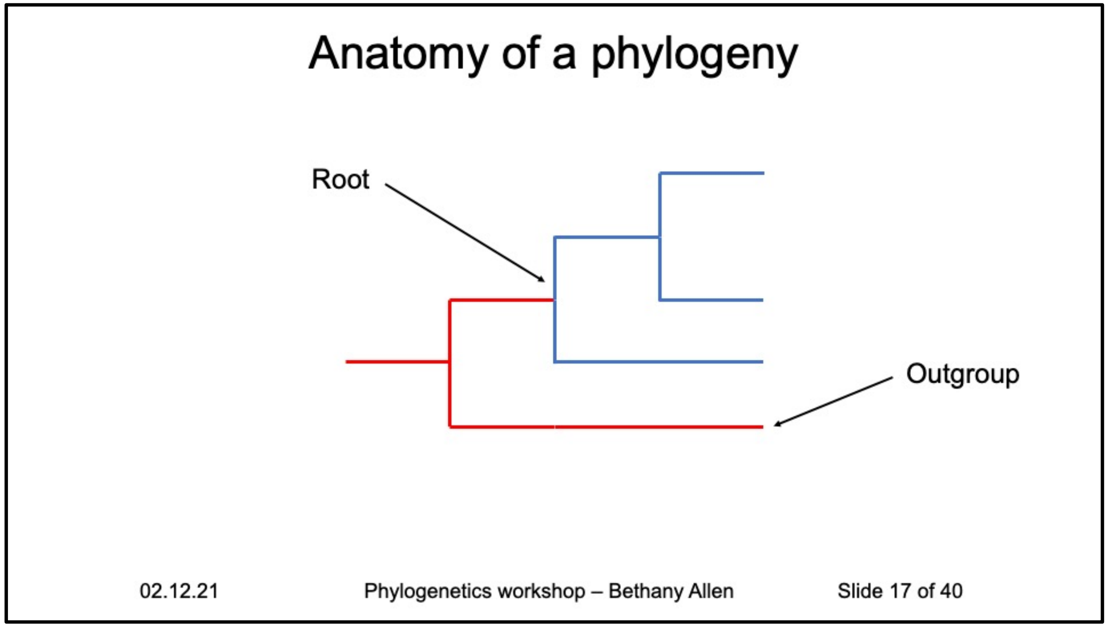
As a quick aside, phylogenies are generally built to assume a bifurcating model of evolution, where one species splits into two different species. However, other models exist, which are arguably closer to the truth. These include budding, where the parent species remains alongside a newly evolved child, and anagenesis, where one species transitions into another along the same lineage, forming two different “chronospecies”. Applying these modelswhen building phylogenies requires different assumptions and implementations, but is possible.



Anatomy of a phylogeny

The oldest node in the group, corresponds to the most recent common ancestor

The root age is often a major characteristic of interest in time-scaled analyses.

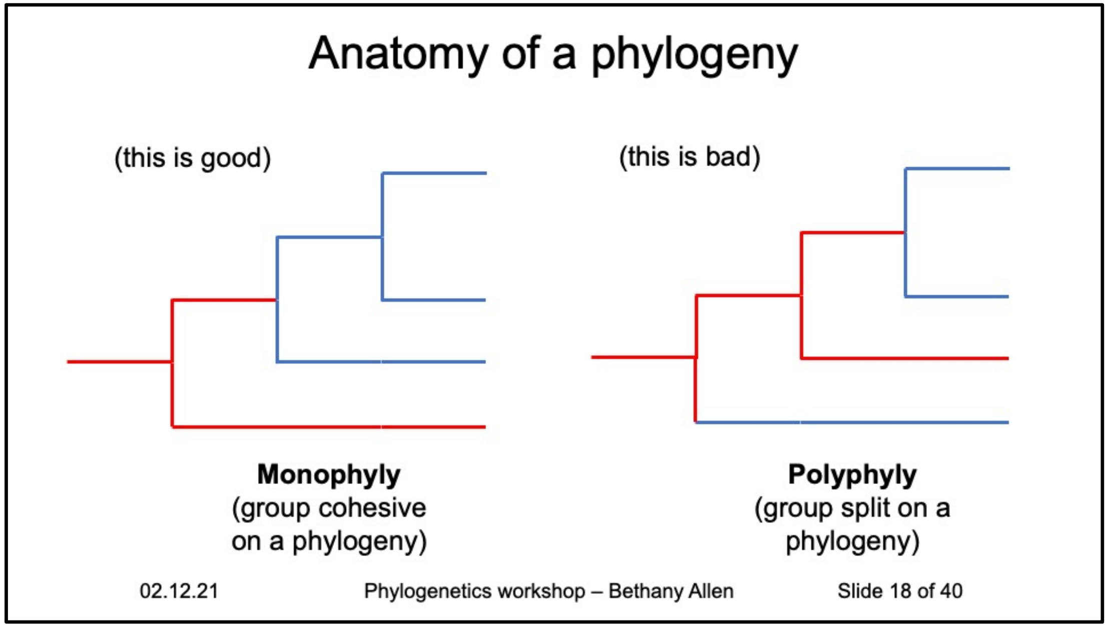


Anatomy of a phylogeny

Root

Outgroup

A phylogeny draws relationships between its tips, but especially in a cladogram, information around how the phylogeny is oriented is not inherently included. In most phylogenetic studies, organisms from outside the group of interest are included as an "outgroup", which serves to anchor the tree by indicating where the root is.



Anatomy of a phylogeny

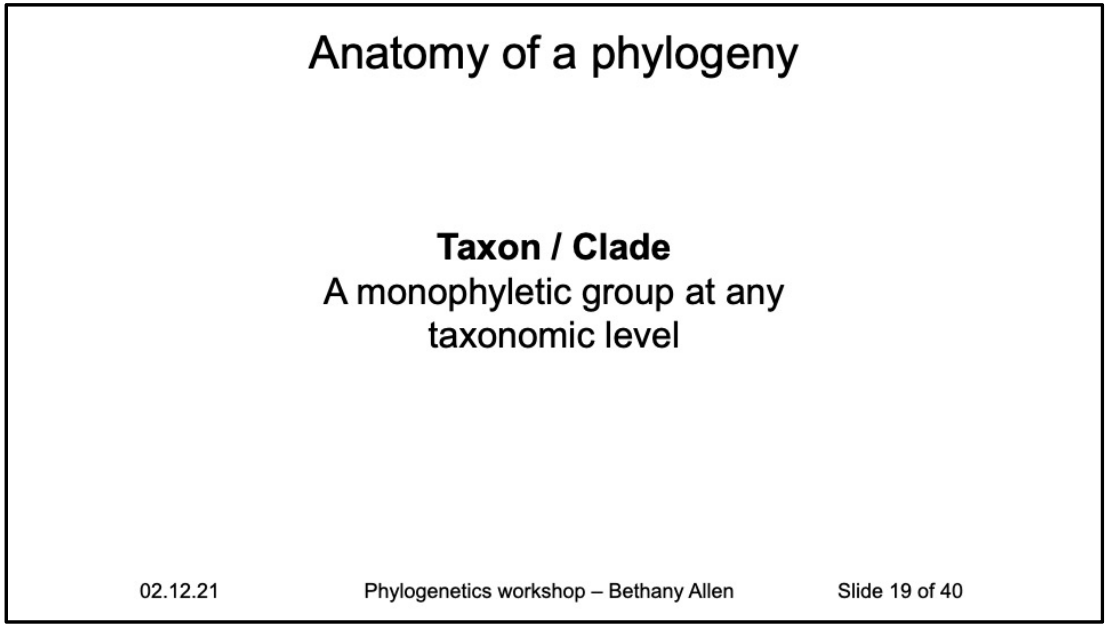
(this is good)

(this is bad)

Monophyly
(group cohesive
on a phylogeny)

Polyphyly
(group split on a
phylogeny)

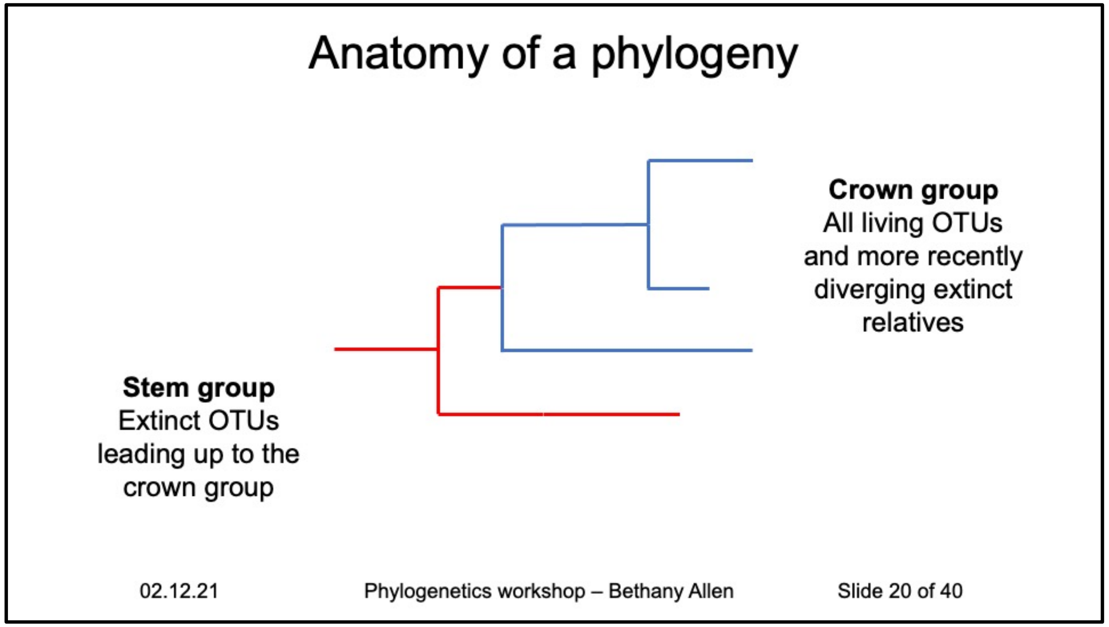
When designating groups of higher taxa, the aim should be to achieve mohophyly, where all tips included in the group fall together on the phylogeny and no tips from within the cluster are excluded. This seems simple, but is not the case in many colloquial groups such as fishes, worms, marine reptiles, and whales. It’s also the reason why we describe dinosaurs as “non-avian dinosaurs”, because birds are nested within this group; “non-avian dinosaurs” are extinct but dinosaurs are not, because birds are still around today.



Anatomy of a phylogeny

Taxon / Clade
A monophyletic group at any taxonomic level

Groups of organisms are often described as taxa or clades; these terms refer to a grouping at any taxonomic level, but should be monophyletic in their designation.

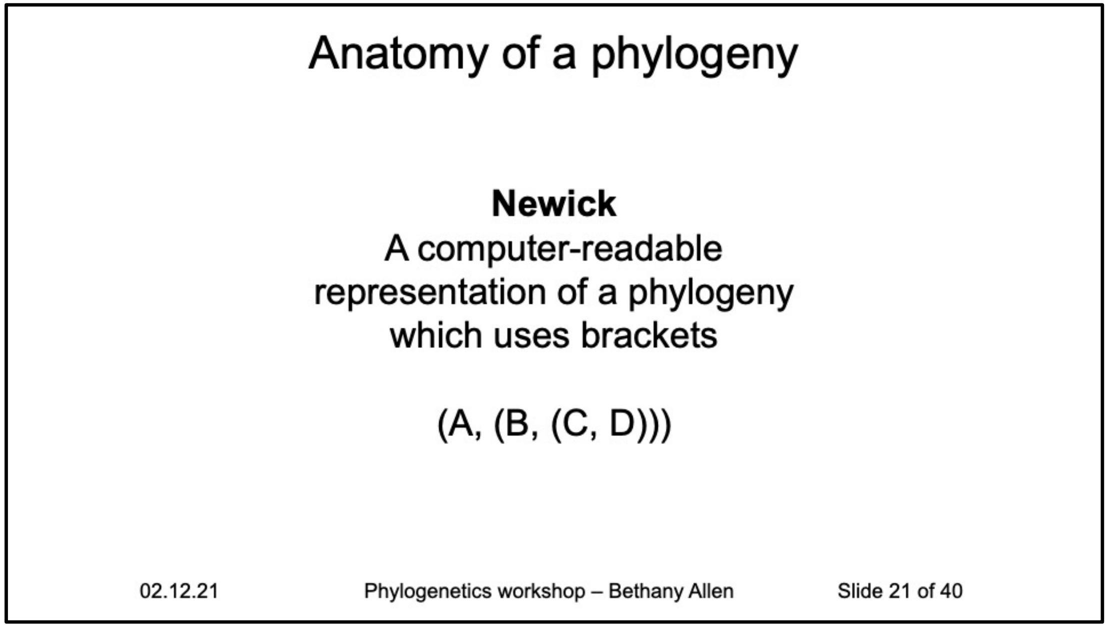


Anatomy of a phylogeny

Crown group
All living OTUs
and more recently
diverging extinct
relatives

Stem group
Extinct OTUs
leading up to the
crown group

”Stem group” and “crown group” are terms you might see in papers about clade evolution; they are not always consistently used but are broadly considered as defined here.



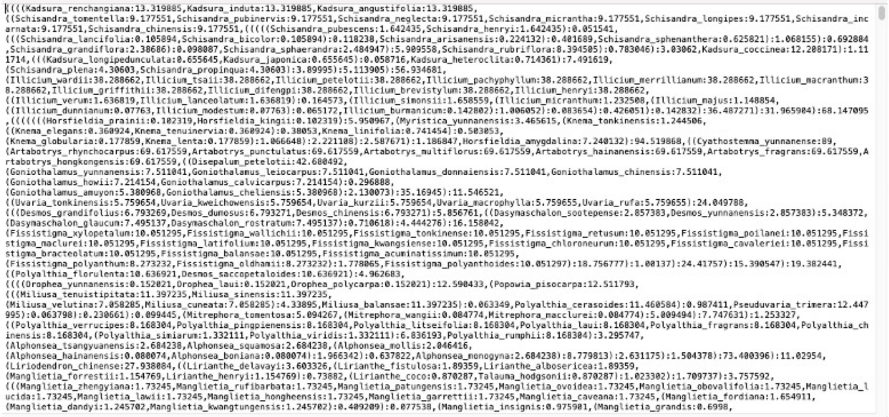
Anatomy of a phylogeny

Newick
A computer-readable
representation of a phylogeny
which uses brackets

(A, (B, (C, D)))

Newick format is used to store phylogenies in a computer-readable form and transfer them between computer programs.

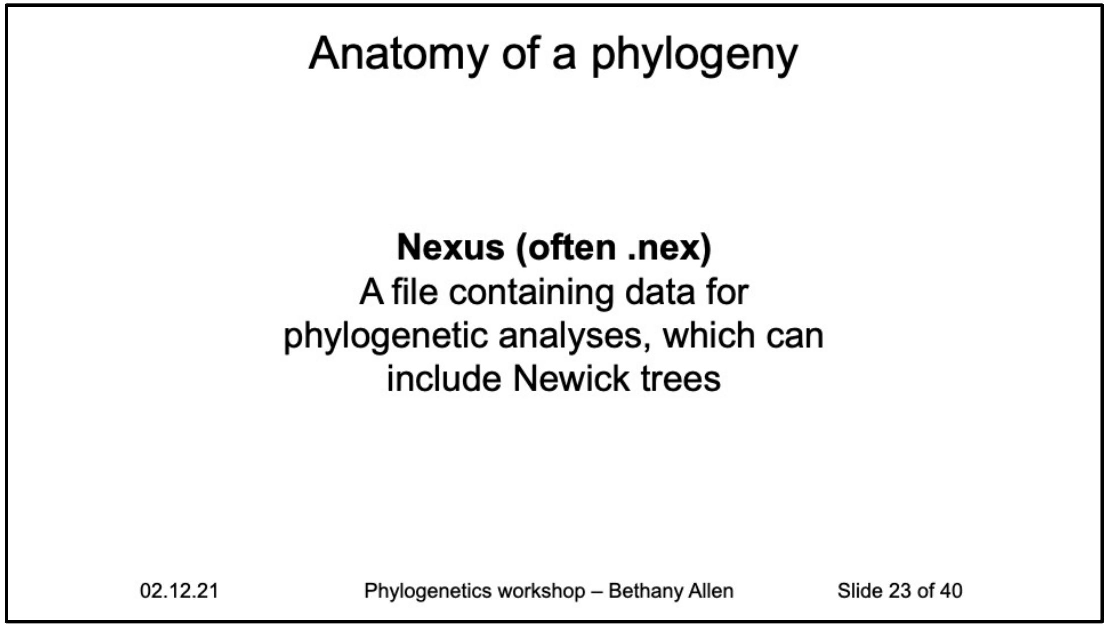
Anatomy of a phylogeny



```
((((Kadsura_renchangiana:13.319885,Kadsura_induta:13.319885,Kadsura_angustifolia:13.319885,
((Schisandra_tomentella:9.177551,Schisandra_pubinervis:9.177551,Schisandra_neglecta:9.177551,Schisandra_micrantha:9.177551,Schisandra_longipes:9.177551,Schisandra_inc
arnata:9.177551,Schisandra_chinensis:9.177551,(((((Schisandra_pubescens:1.642435,Schisandra_henryi:1.642435):0.051541,
(((Schisandra_lancifolia:0.105894,Schisandra_bicolor:0.105894):0.118238,Schisandra_arisanensis:0.224132):0.401689,Schisandra_sphenanthera:0.625821):1.068155):0.692884
,Schisandra_grandiflora:2.38686):0.098087,Schisandra_sphaerandra:2.484947):5.909558,Schisandra_rubriflora:8.394505):0.783046):3.03062,Kadsura_coccinea:12.208171):1.11
1714,(((Kadsura_longipedunculata:0.655645,Kadsura_japonica:0.655645):0.058716,Kadsura_heteroclita:0.714361):7.491619,
(Schisandra_plena:4.30603,Schisandra_propinqua:4.30603):3.89995):5.113905):56.934681,
(Illicium_wardii:38.288662,Illicium_tsaii:38.288662,Illicium_petelotii:38.288662,Illicium_pachyphyllum:38.288662,Illicium_merrillianum:38.288662,Illicium_macranthum:3
8.288662,Illicium_griffithii:38.288662,Illicium_difengpi:38.288662,Illicium_brevistylum:38.288662,Illicium_henryi:38.288662,
((Illicium_verum:1.636819,Illicium_lanceolatum:1.636819):0.164573,(Illicium_simonsii:1.658559,(Illicium_micranthum:1.232508,(Illicium_majus:1.148854,
((Illicium_dunnianum:0.07763,Illicium_modestum:0.07763):0.065172,Illicium_burmanicum:0.142802):1.006052):0.083654):0.426051):0.142832):36.487271):31.965904):68.147095
,((((((Horsfieldia_prainii:0.102319,Horsfieldia_kingii:0.102319):5.950967,(Myristica_yunnanensis:3.465615,(Knema_tonkinensis:1.244506,
((Knema_elegans:0.360924,Knema_tenuinervia:0.360924):0.38053,Knema_linifolia:0.741454):0.503053,
(Knema_globularia:0.177859,Knema_lenta:0.177859):1.066648):2.221108):2.587671):1.186847,Horsfieldia_amygdalina:7.240132):94.519868,((Cyathostemma_yunnanense:89,
(Artabotrys_rhynchocarpus:69.617559,Artabotrys_punctulatus:69.617559,Artabotrys_multiflorus:69.617559,Artabotrys_hainanensis:69.617559,Artabotrys_fragrans:69.617559,A
rtabotrys_hongkongensis:69.617559,((Disepalum_petelotii:42.680492,
(Goniothalamus_yunnanensis:7.511041,Goniothalamus_leiocarpus:7.511041,Goniothalamus_donnaiensis:7.511041,Goniothalamus_chinensis:7.511041,
(Goniothalamus_howii:7.214154,Goniothalamus_calvicarpus:7.214154):0.296888,
(Goniothalamus_amuyon:5.380968,Goniothalamus_cheliensis:5.380968):2.130073):35.16945):11.546521,
((Uvaria_tonkinensis:5.759654,Uvaria_kweichowensis:5.759654,Uvaria_kurzii:5.759654,Uvaria_macrophylla:5.759655,Uvaria_rufa:5.759655):24.049788,
(((Desmos_grandifolius:6.793269,Desmos_dumosus:6.793271,Desmos_chinensis:6.793271):5.856761,((Dasymaschalon_sootepense:2.857383,Desmos_yunnanensis:2.857383):5.348372,
(Dasymaschalon_glaucum:7.495137,Dasymaschalon_rostratum:7.495137):0.710618):4.444276):16.158042,
(Fissistigma_xylopetalum:10.051295,Fissistigma_wallichii:10.051295,Fissistigma_tonkinense:10.051295,Fissistigma_retusum:10.051295,Fissistigma_poilanei:10.051295,Fissi
stigma_maclurei:10.051295,Fissistigma_latifolium:10.051295,Fissistigma_kwangsiense:10.051295,Fissistigma_chloroneurum:10.051295,Fissistigma_cavaleriei:10.051295,Fissi
stigma_bracteolatum:10.051295,Fissistigma_balansae:10.051295,Fissistigma_acuminatissimum:10.051295,
(Fissistigma_polyanthum:8.273232,Fissistigma_oldhamii:8.273232):1.778065,Fissistigma_polyanthoides:10.051297):18.756777):1.00137):24.41757):15.390547):19.382441,
((Polyalthia_florulenta:10.636921,Desmos_saccopetaloides:10.636921):4.962683,
((((Orophea_yunnanensis:0.152021,Orophea_laui:0.152021,Orophea_polycarpa:0.152021):12.590433,(Popowia_pisocarpa:12.511793,
(((Miliusa_tenuistipitata:11.397235,Miliusa_sinensis:11.397235,
(Miliusa_velutina:7.058285,Miliusa_cuneata:7.058285):4.33895,Miliusa_balansae:11.397235):0.063349,Polyalthia_cerasoides:11.460584):0.987411,Pseuduvaria_trimera:12.447
995):0.063798):0.230661):0.099445,(Mitrephora_tomentosa:5.094267,(Mitrephora_wangii:0.084774,Mitrephora_macclurei:0.084774):5.009494):7.747631):1.253327,
((Polyalthia_verrucipes:8.168304,Polyalthia_pingpienensis:8.168304,Polyalthia_litseifolia:8.168304,Polyalthia_laui:8.168304,Polyalthia_fragrans:8.168304,Polyalthia_ch
inensis:8.168304,(Polyalthia_simiarum:1.332111,Polyalthia_viridis:1.332111):6.836193,Polyalthia_rumphii:8.168304):3.295747,
(Alphonsea_tsangyuanensis:2.684238,Alphonsea_squamosa:2.684238,(Alphonsea_mollis:2.046416,
(Alphonsea_hainanensis:0.080074,Alphonsea_boniana:0.080074):1.966342):0.637822,Alphonsea_monogyna:2.684238):8.779813):2.631175):1.504378):73.400396):11.02954,
(Liriodendron_chinense:27.938084,((Lirianthe_delavayi:3.603326,(Lirianthe_fistulosa:1.89359,Lirianthe_albosericea:1.89359,
(Manglietia_forrestii:1.154769,Lirianthe_henryi:1.154769):0.73882,(Lirianthe_coco:0.870287,Talauma_hodgsonii:0.870287):1.023302):1.709737):3.757592,
(((Manglietia_zhengyiana:1.73245,Manglietia_rufibarbata:1.73245,Manglietia_patungensis:1.73245,Manglietia_ovoidea:1.73245,Manglietia_obovalifolia:1.73245,Manglietia_l
ucida:1.73245,Manglietia_lawii:1.73245,Manglietia_hongheensis:1.73245,Manglietia_garrettii:1.73245,Manglietia_caveana:1.73245,(Manglietia_fordiana:1.654911,
(Manglietia_dandyi:1.245702,Manglietia_kwangtungensis:1.245702):0.409209):0.077538,(Manglietia_insignis:0.975981,(Manglietia_grandis:0.6998,
```

Liu et al. 2021

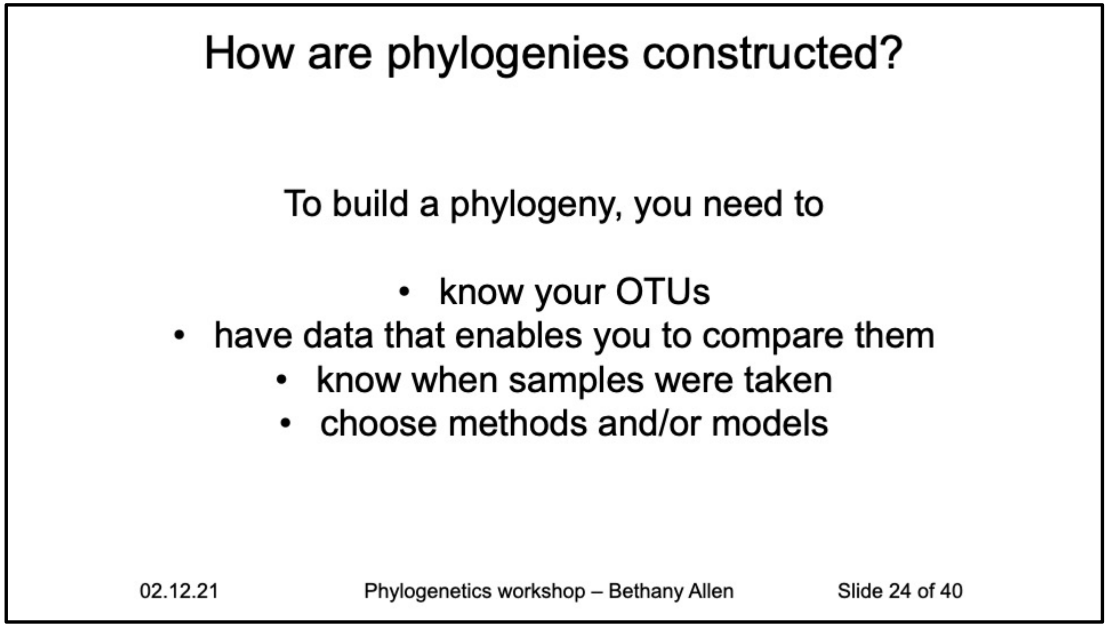
This is an example of a Newick text file, containing a phylogeny of Chinese angiosperms. You can see numbers here which indicate the branch lengths.



Anatomy of a phylogeny

Nexus (often .nex)
A file containing data for phylogenetic analyses, which can include Newick trees

Nexus files can contain lots of different types of data relevant to phylogenetic analyses, each of which have a standard format and are clearly marked in the file, to be read by computer programs.



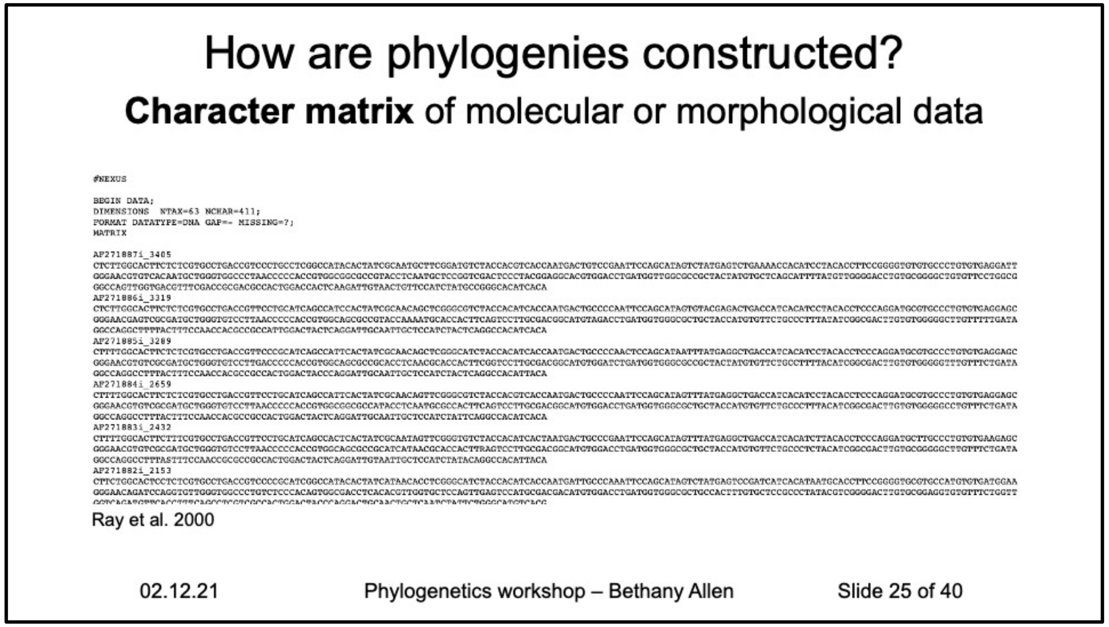
How are phylogenies constructed?

To build a phylogeny, you need to

- know your OTUs
- have data that enables you to compare them
- know when samples were taken
- choose methods and/or models

02.12.21 Phylogenetics workshop – Bethany Allen Slide 24 of 40

It might sound simple but often it’s not!



How are phylogenies constructed?

Character matrix of molecular or morphological data

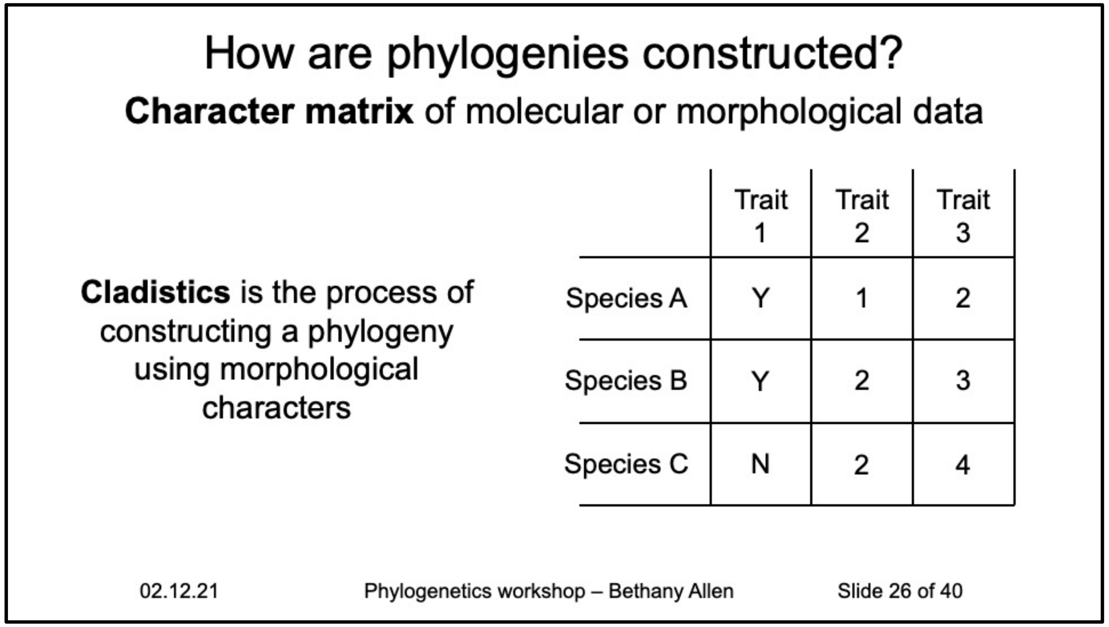
```
#NEXUS

BEGIN DATA;
DIMENSIONS  NTAX=63 NCHAR=411;
FORMAT DATATYPE=DNA GAP=- MISSING=?;
MATRIX

AF271887i_3405
CTCTTGGCACTTCTCTCGTGCCTGACCGTCCCTGCCTCGGCCATACACTATCGCAATGCTTCGGATGTCTACCACGTCACCAATGACTGTCCGAATTCCAGCATAGTCTATGAGTCTGAAAACCACATCCTACACCTTCCGGGGTGTGTGCCCTGTGTGAGGATT
GGGAACGTGTCACAATGCTGGGTGGCCCTAACCCCCACCGTGGCGGCGCCGTACCTCAATGCTCCGGTCGACTCCCTACGGAGGCACGTGGACCTGATGGTTGGCGCCGCTACTATGTGCTCAGCATTTTATGTTGGGGACCTGTGCGGGGCTGTGTTCCTGGCG
GGCCAGTTGGTGACGTTTCGACCGCGACGCCACTGGACCACTCAAGATTGTAACTGTTCCATCTATGCCGGGCACATCACA
AF271886i_3319
CTCTTGGCACTTCTCTCGTGCCTGACCGTTCCTGCATCAGCCATCCACTATCGCAACAGCTCGGGCGTCTACCACATCACCAATGACTGCCCCAATTCCAGCATAGTGTACGAGACTGACCATCACATCCTACACCTCCCAGGATGCGTGCCCTGTGTGAGGAGC
GGGAACGAGTCGCGATGCTGGGTGTCCTTAACCCCCACCGTGGCAGCGCCGTACCAAAATGCACCACTTCAGTCCTTGCGACGGCATGTAGACCTGATGGTGGGCGCTGCTACCATGTGTTCTGCCCTTTATATCGGCGACTTGTGTGGGGGCTTGTTTTTGATA
GGCCAGGCTTTTACTTTCCAACCACGCCGCCATTGGACTACTCAGGATTGCAATTGCTCCATCTACTCAGGCCACATCACA
AF271885i_3289
CTTTTGGCACTTCTCTCGTGCCTGACCGTTCCCGCATCAGCCATTCACTATCGCAACAGCTCGGGCATCTACCACATCACCAATGACTGCCCCAACTCCAGCATAATTTATGAGGCTGACCATCACATCCTACACCTCCCAGGATGCGTGCCCTGTGTGAGGAGC
GGGAACGTGTCGCGATGCTGGGTGTCCTTGACCCCCACCGTGGCAGCGCCGCACCTCAACGCACCACTTCGGTCCTTGCGACGGCATGTGGATCTGATGGTGGGCGCCGCTACTATGTGTTCTGCCTTTTACATCGGCGACTTGTGTGGGGGTTTGTTTCTGATA
GGCCAGGCCTTTACTTTCCAACCACGCCGCCACTGGACTACCCAGGATTGCAATTGCTCCATCTACTCAGGCCACATTACA
AF271884i_2659
CTTTTGGCACTTCTCTCGTGCCTGACCGTTCCTGCATCAGCCATTCACTATCGCAACAGTTCGGGCGTCTACCACGTCACCAATGACTGCCCCAATTCCAGCATAGTTTATGAGGCTGACCATCACATCCTACACCTCCCAGGATGCGTGCCCTGTGTGAGGAGC
GGGAACGTGTCGCGATGCTGGGTGTCCTTAACCCCCACCGTGGCGGCGCCATACCTCAATGCGCCACTTCAGTCCTTGCGACGGCATGTGGACCTGATGGTGGGCGCTGCTACCATGTGTTCTGCCCTTTACATCGGCGACTTGTGTGGGGGCCTGTTTCTGATA
GGCCAGGCCTTTACTTTCCAACCACGCCGCCACTGGACTACTCAGGATTGCAATTGCTCCATCTATTCAGGCCACATCACA
AF271883i_2432
CTTTTGGCACTTCTTTCGTGCCTGACCGTTCCTGCATCAGCCACTCACTATCGCAATAGTTCGGGTGTCTACCACATCACTAATGACTGCCCGAATTCCAGCATAGTTTATGAGGCTGACCATCACATCTTACACCTCCCAGGATGCTTGCCCTGTGTGAAGAGC
GGGAACGTGTCGCGATGCTGGGTGTCCTTAACCCCCACCGTGGCAGCGCCGCATCATAACGCACCACTTRAGTCCTTGCGACGGCATGTGGACCTGATGGTGGGCGCTGCTACCATGTGTTCTGCCCTCTACATCGGCGACTTGTGCGGGGGCTTGTTTCTGATA
GGCCAGGCCTTTASTTTCCAACCGCGCCGCCACTGGACTACTCAGGATTGTAATTGCTCCATCTATACAGGCCACATTACA
AF271882i_2153
CTTCTGGCACTCCTCTCGTGCCTGACCGTCCCCGCATCGGCCATACACTATCATAACACCTCGGGCATCTACCACATCACCAATGATTGCCCAAATTCCAGCATAGTCTATGAGTCCGATCATCACATAATGCACCTTCCGGGGTGCGTGCCATGTGTGATGGAA
GGGAACAGATCCAGGTGTTGGGTGGCCCTGTCTCCCACAGTGGCGACCTCACACGTTGGTGCTCCAGTTGAGTCCATGCGACGACATGTGGACCTGATGGTGGGCGCTGCCACTTTGTGCTCCGCCCTATACGTCGGGGACTTGTGCGGAGGTGTGTTTCTGGTT
GGTCAGATGTTCACCTTTCAGCCTCGTCGCCACTGGACTACCCAGGACTGCAACTGCTCAATCTATTCTGGGCATGTCACC
```

Ray et al. 2000

Data for comparing OTUs is typically a character matrix. This example is Hepatitis C sequences from patients in Egypt. The best molecular data to use is usually one or more slow-evolving genes (strings of G, C, A, and T). Molecular data can also come from protein structure, where the code indicates amino acids. The idea is that differences accrue over time and can therefore be used to indicate the relative evolutionary distances between sequences. Large changes such as insertions and deletions can drastically alter genomes, so usually some form of alignment processing has to take place before the phylogenetic analysis to identify where these might have occurred.



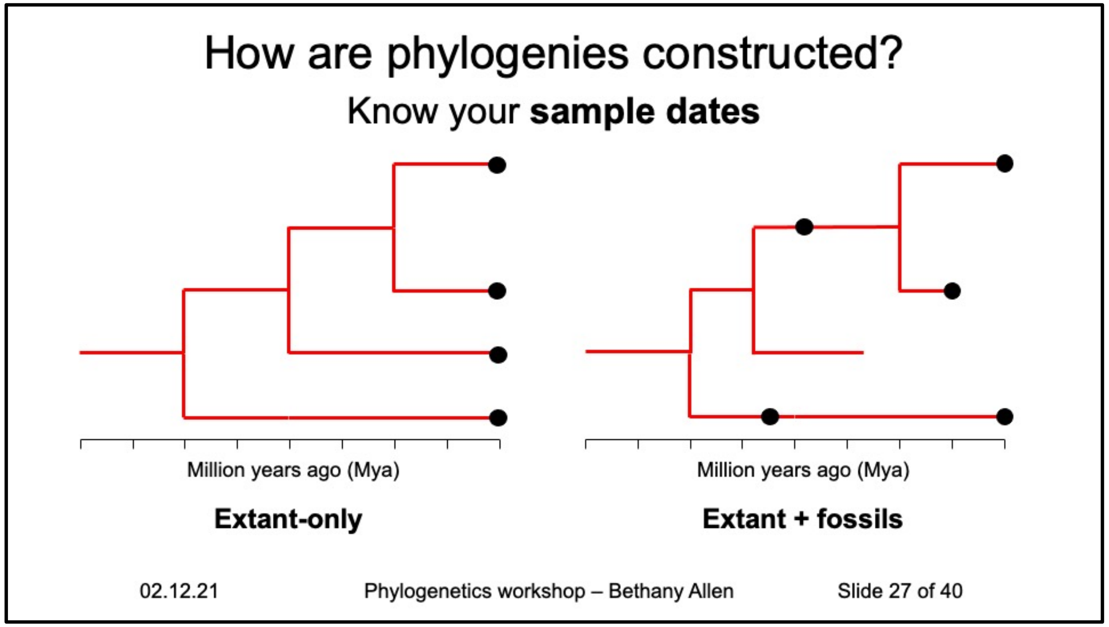
How are phylogenies constructed?

Character matrix of molecular or morphological data

Cladistics is the process of constructing a phylogeny using morphological characters

	Trait 1	Trait 2	Trait 3
Species A	Y	1	2
Species B	Y	2	3
Species C	N	2	4

Collecting morphological data is difficult and highly subjective. Various guidelines exist to outline good practice. Traits can be binary or multi-state, but generally a single scheme should be adhered to throughout a character matrix. Continuous traits are rarely used for phylogenetic analyses, but are sometimes categorised into states. Ideally characters should be chosen that are not typical of convergent evolution (where organisms evolve similar body forms across different clades because they are highly functional). Especially with fossils, it is common to have large amounts of missing data, and this can impact on the resolution of the phylogeny.



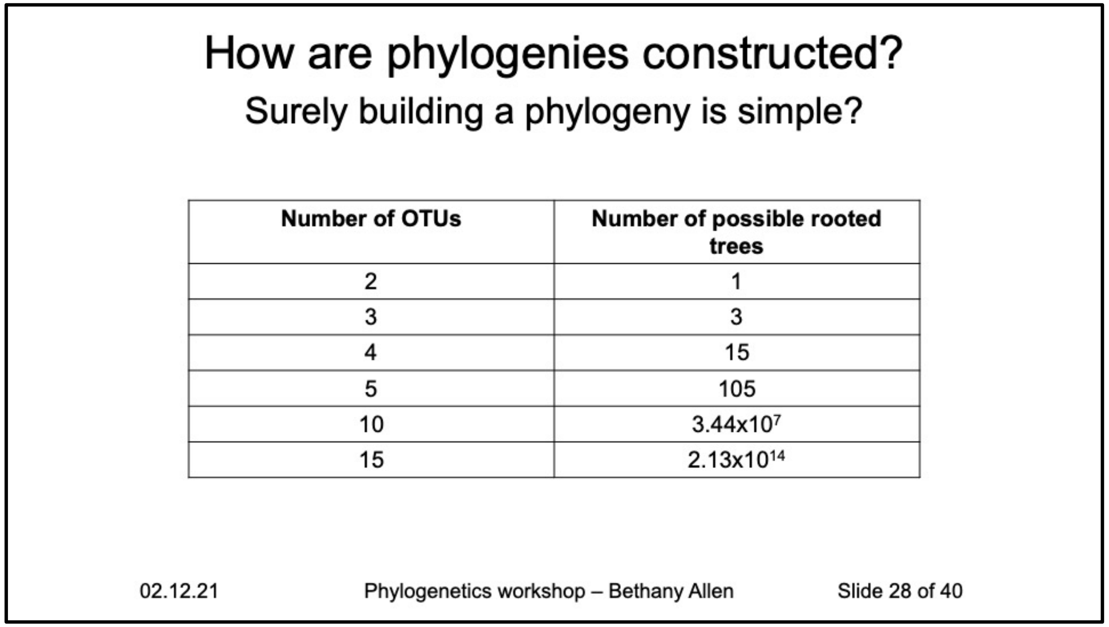
How are phylogenies constructed?

Know your **sample dates**

Extant-only

Extant + fossils

This is straightforward if you are only using extant taxa in your analysis, but if fossil data is being included, it is important to have a good understanding of data sources and uncertainty when considering age.

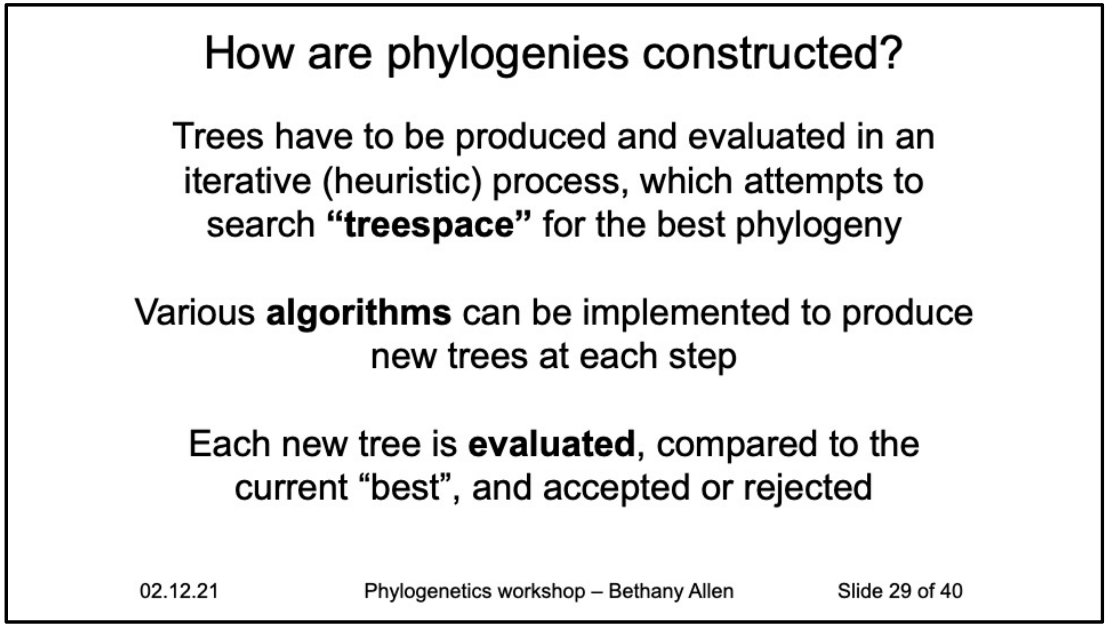


How are phylogenies constructed?

Surely building a phylogeny is simple?

Number of OTUs	Number of possible rooted trees
2	1
3	3
4	15
5	105
10	3.44×10^{7}
15	2.13×10^{14}

Building a phylogeny intuitively might seem simple, but unfortunately it isn't. To be certain of having identified the most realistic phylogeny based on the data, you would want to compare every possible permutation of divergences. However, this number is incredibly large, even for what we might consider to be a fairly small phylogeny. As a result, we cannot hope to compare every option and must take a different approach to try and identify our "best" tree.



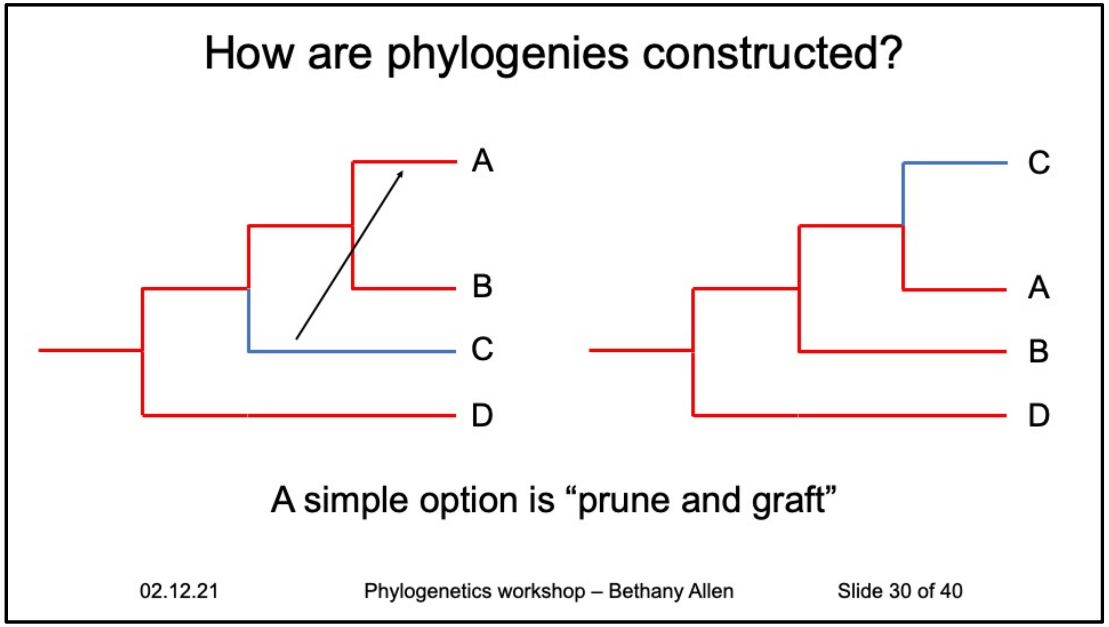
How are phylogenies constructed?

Trees have to be produced and evaluated in an iterative (heuristic) process, which attempts to search **“treespace”** for the best phylogeny

Various **algorithms** can be implemented to produce new trees at each step

Each new tree is **evaluated**, compared to the current “best”, and accepted or rejected

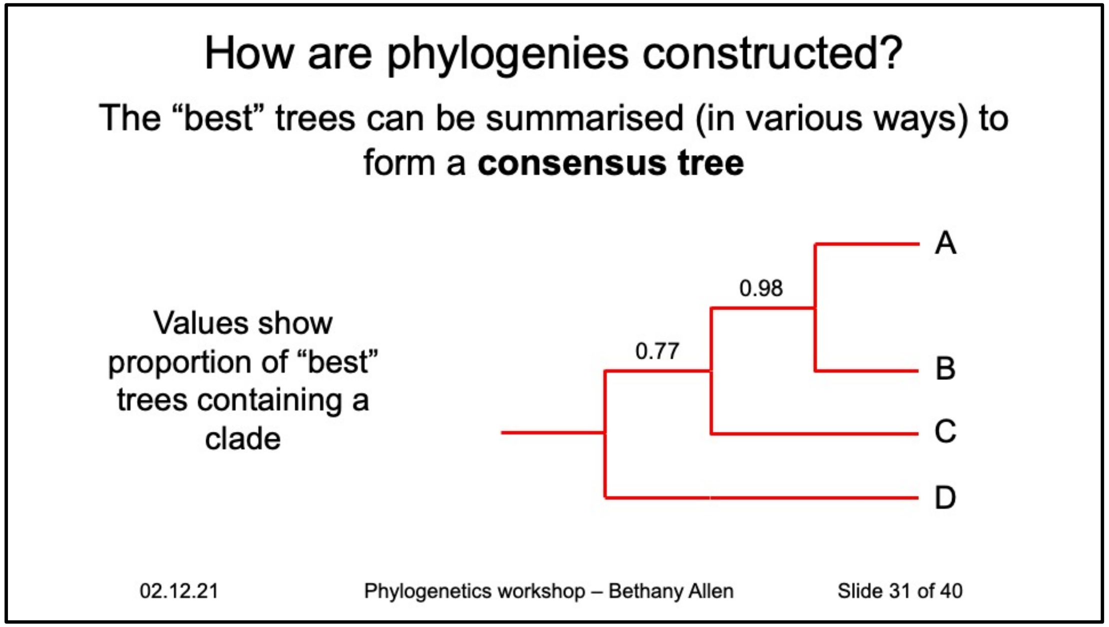
“Treespace” is a somewhat theoretical entity which describes possible phylogenies and their relationships to one another. I have put “best” in quote marks, which is something I will talk about in much more detail shortly.



How are phylogenies constructed?

A simple option is “prune and graft”

“Prune and graft” involves selecting a branch, or cluster of branches, and transplanting it across the tree. This means that the next tree examined is generally similar (close in treespace) to the last.

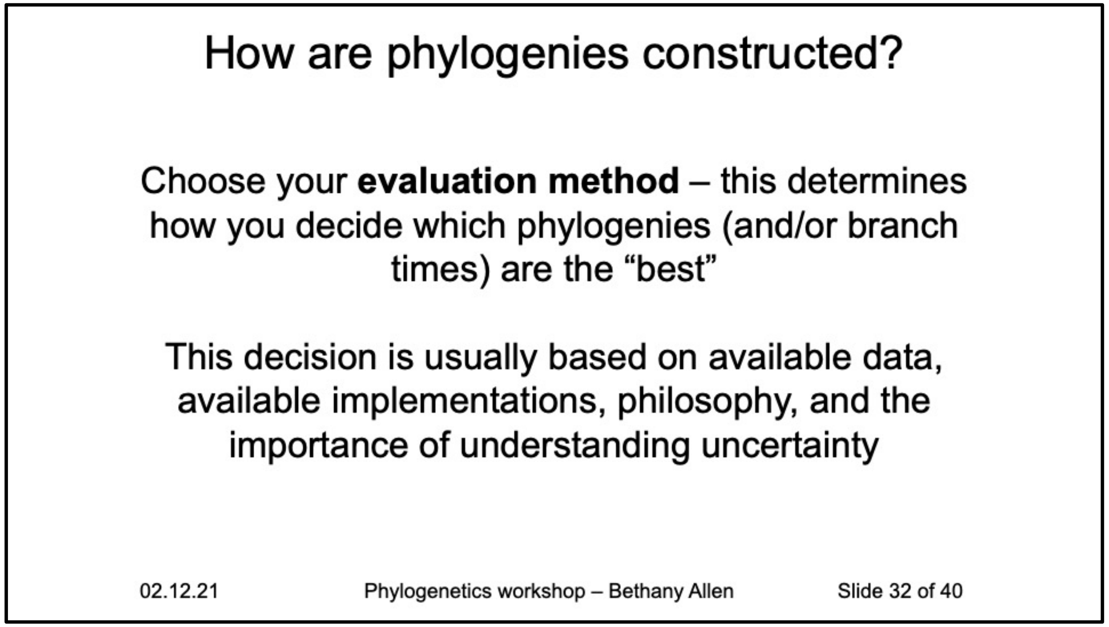


How are phylogenies constructed?

The “best” trees can be summarised (in various ways) to form a **consensus tree**

Values show proportion of “best” trees containing a clade

The consensus tree is typically what you will see in the publication, it summarises the relationships discovered in the analyses and puts an estimate of uncertainty on the nodes presented.

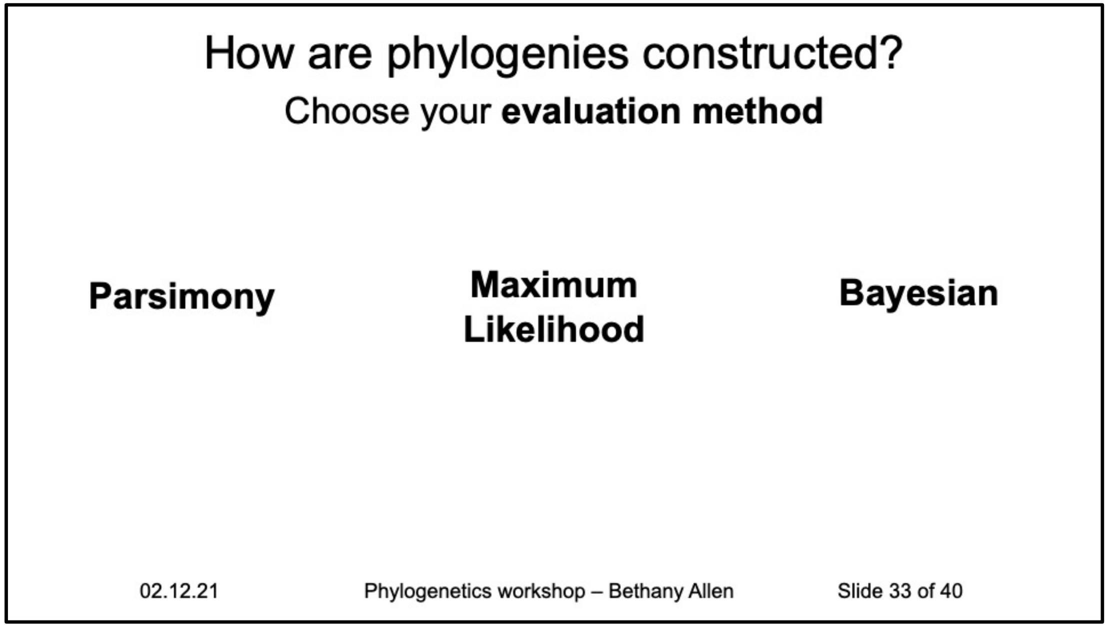


How are phylogenies constructed?

Choose your **evaluation method** – this determines how you decide which phylogenies (and/or branch times) are the “best”

This decision is usually based on available data, available implementations, philosophy, and the importance of understanding uncertainty

I will return to this discussion in more detail in Part Two.



How are phylogenies constructed?

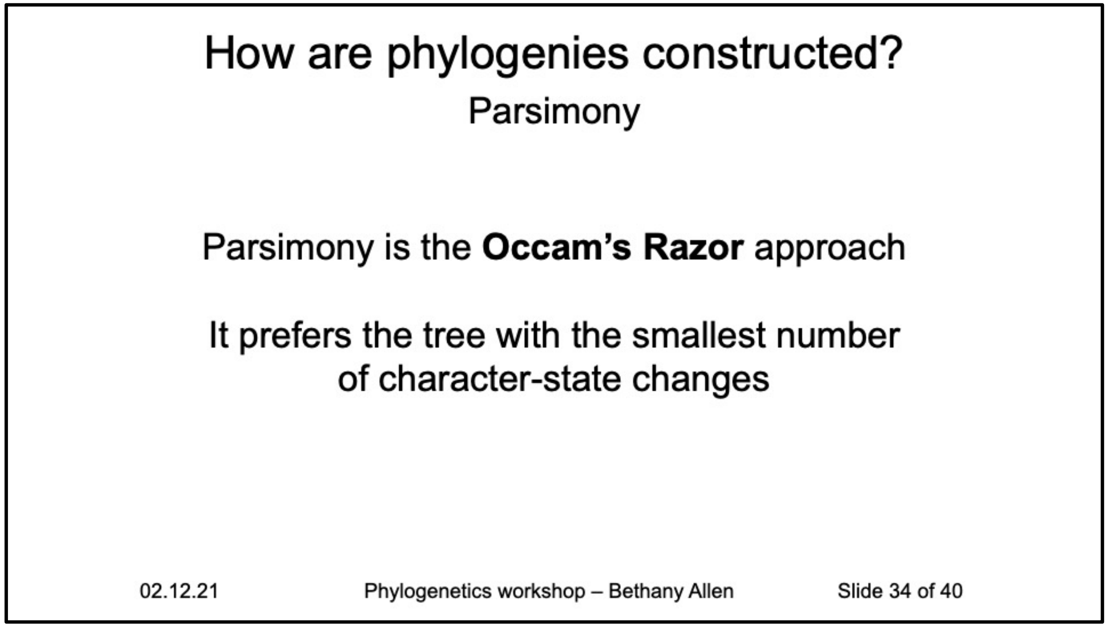
Choose your **evaluation method**

Parsimony

Maximum Likelihood

Bayesian

Other methods exist but these are the three main (popular) options, each of which I will describe.



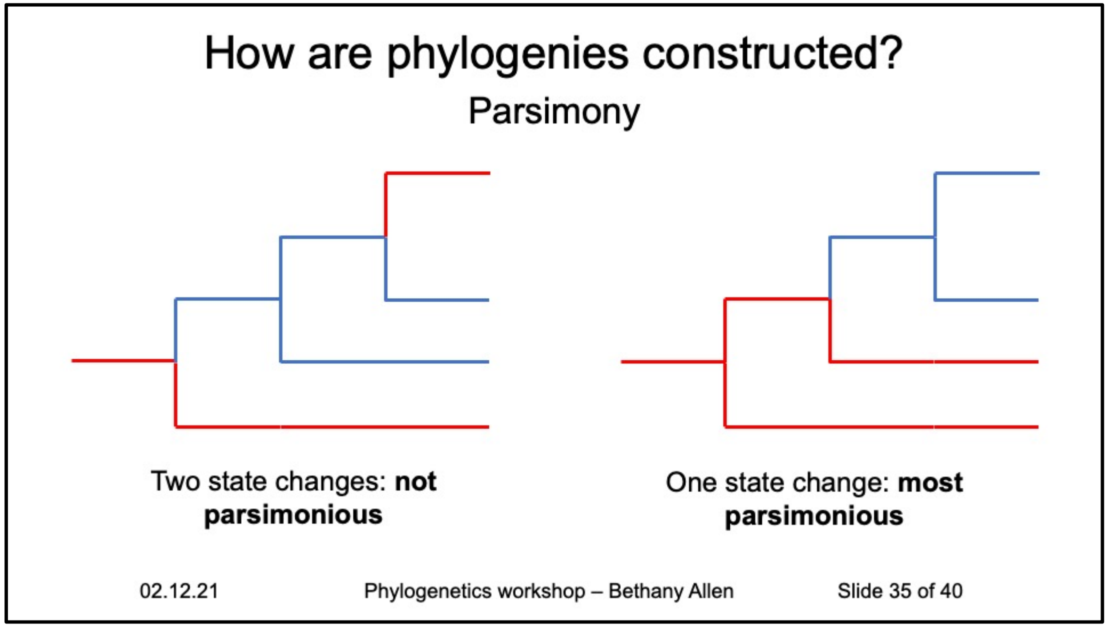
How are phylogenies constructed?

Parsimony

Parsimony is the **Occam's Razor** approach

It prefers the tree with the smallest number of character-state changes

Occam's Razor is a philosphical approach which suggests that the simplest explanation is the best one. In this application, ”simplest” describes the smallest number of state changes. This is based on the assumption that state changes are generally one-directional, which is a very strong and not always sensible assumption. While it might be argued that this is often the case with morphological datasets (traits that are gained don't tend to be lost again), this is much more controversial for molecular datasets.



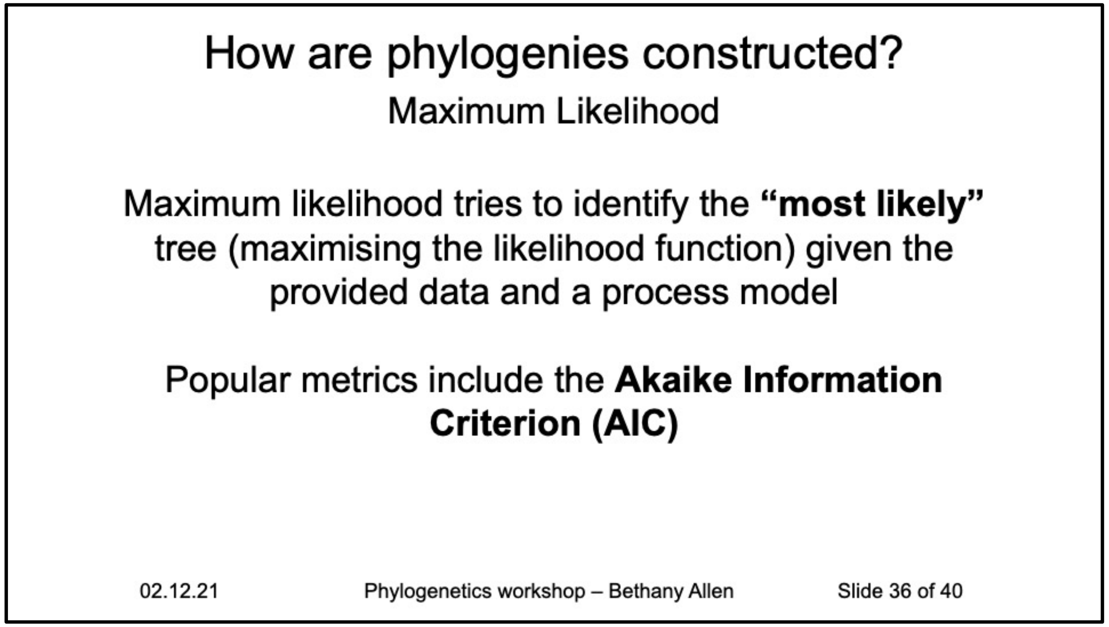
How are phylogenies constructed?

Parsimony

Two state changes: **not parsimonious**

One state change: **most parsimonious**

The left tree shows a situation where the phylogeny suggests that a trait is evolved and then lost again; this is two state changes, and so this phylogeny would not be chosen as the “best” using a maximum parsimony approach. The tree that is preferred is the one in which the smallest number of state changes, summed across all traits, takes place across the phylogeny.



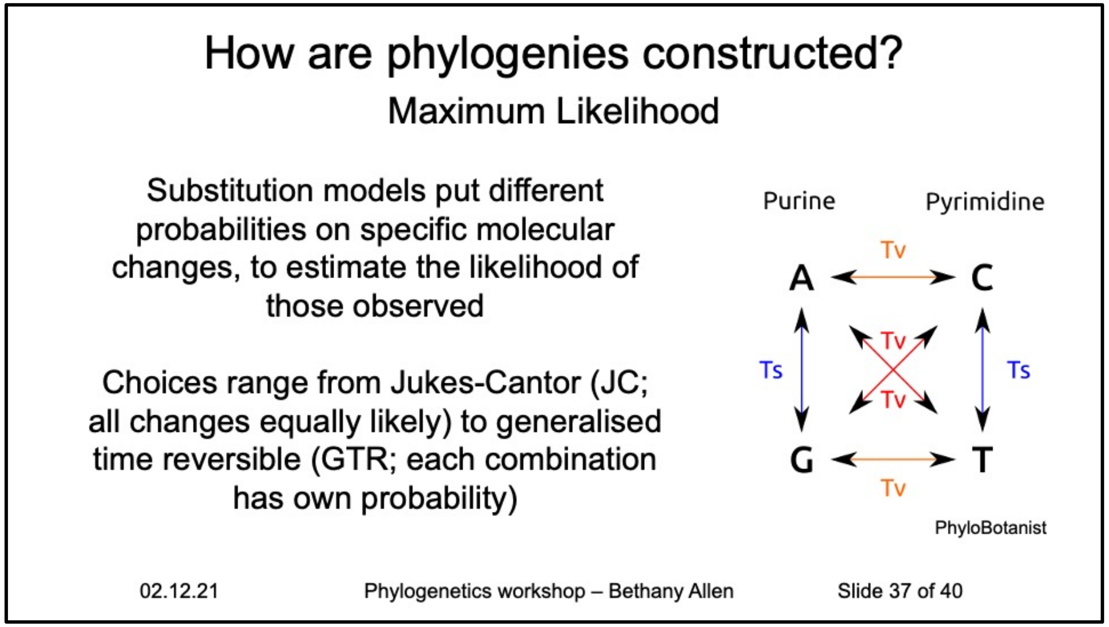
How are phylogenies constructed?

Maximum Likelihood

Maximum likelihood tries to identify the **“most likely”** tree (maximising the likelihood function) given the provided data and a process model

Popular metrics include the **Akaike Information Criterion (AIC)**

Maximum likelihood hinges on the assumption that the chosen model of evolution is correct, as the assessment of likelihood is based on that model. Choice of assessment metric can also be important, as these vary in how they respond to different sample sizes, how they penalise complexity, etc.



How are phylogenies constructed?

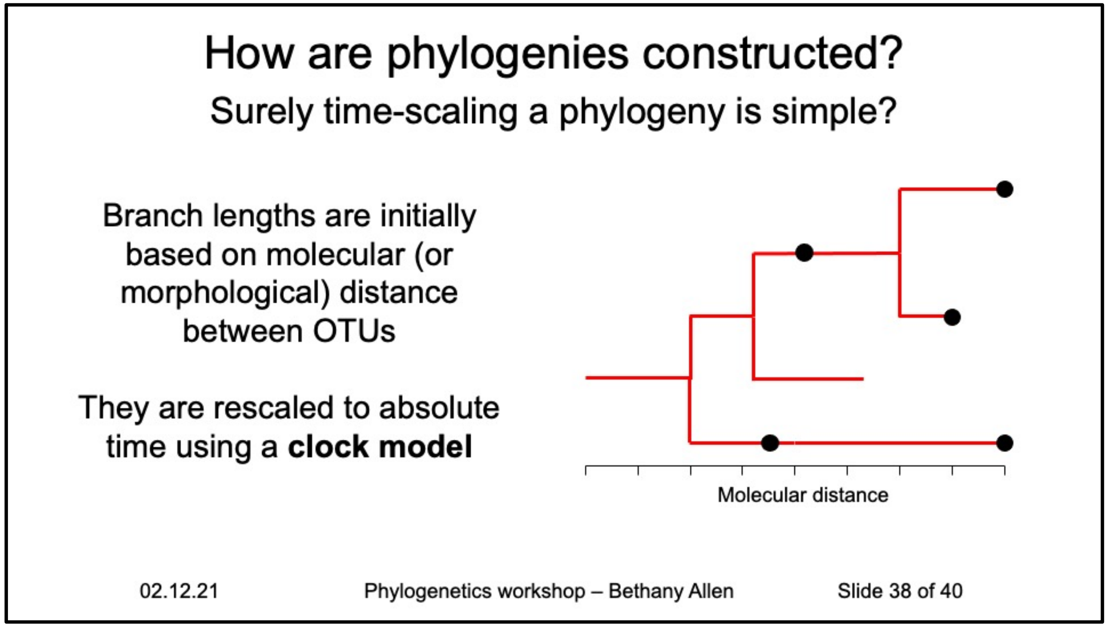
Maximum Likelihood

Substitution models put different probabilities on specific molecular changes, to estimate the likelihood of those observed

Choices range from Jukes-Cantor (JC; all changes equally likely) to generalised time reversible (GTR; each combination has own probability)

PhyloBotanist

The evolutionary model is typically a substitution model for different molecular transitions. Models for transitions in morphological evolution are controversial, but are often simple as a result.



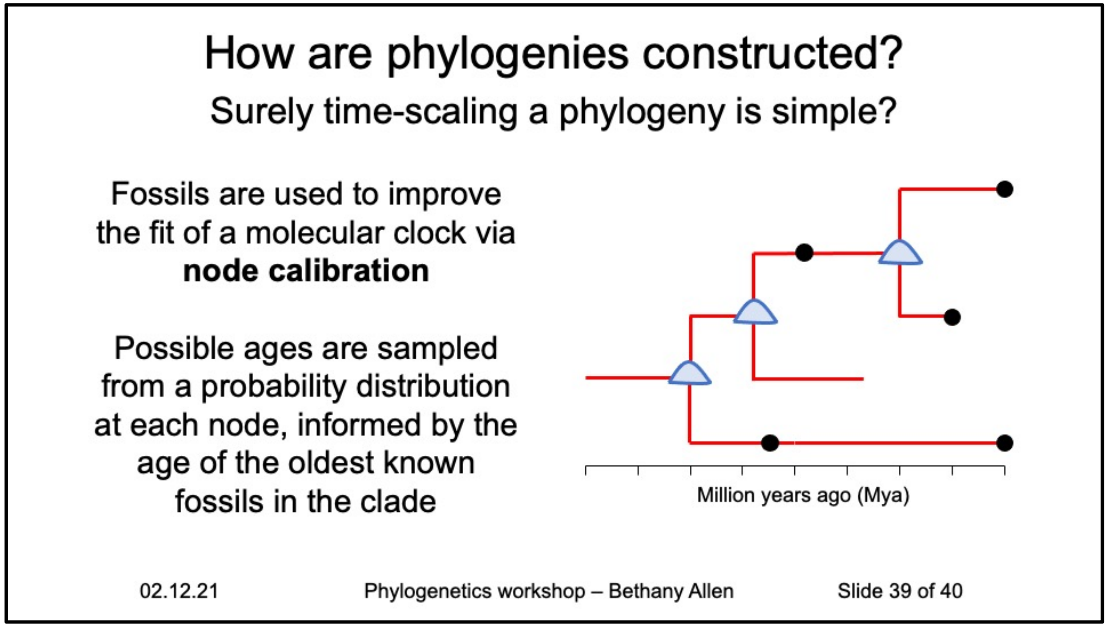
How are phylogenies constructed?

Surely time-scaling a phylogeny is simple?

Branch lengths are initially based on molecular (or morphological) distance between OTUs

They are rescaled to absolute time using a **clock model**

These methods create a cladogram which can then be scaled if required. The first clock models were fixed, however is has become clear that rates of molecular evolution are far from consistent, and so now clocks which are more flexible tend to be preferred, usually called “relaxed” clocks. Such clocks can be applied to allow different rates of molecular evolution on each branch of the tree.



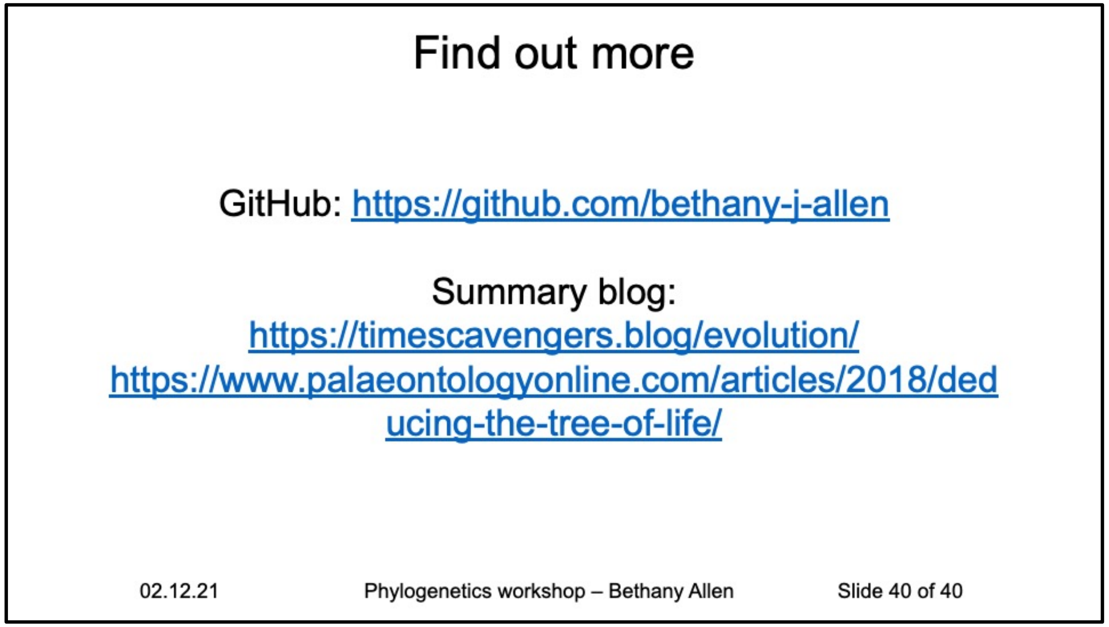
How are phylogenies constructed?

Surely time-scaling a phylogeny is simple?

Fossils are used to improve the fit of a molecular clock via **node calibration**

Possible ages are sampled from a probability distribution at each node, informed by the age of the oldest known fossils in the clade

Node calibration can be really important when using a relaxed clock model, to provide a constraint on how much movement is available in branch lengths. However, node calibration is very difficult for many reasons, and is often considered somewhat of a dark art. Extensive justifications of their temporal positioning and distribution shape should be provided when they are used.



Find out more

GitHub: https://github.com/bethany-j-allen

Summary blog:
https://timescavengers.blog/evolution/
https://www.palaeontologyonline.com/articles/2018/deducing-the-tree-of-life/