Tree-thinking and introduction to phylogenetics

2025 Workshop on Molecular Evolution – Tracy Heath

Nothing in biology makes sense except in the light of evolution.

— Theodosius Dobzhansky, 1973

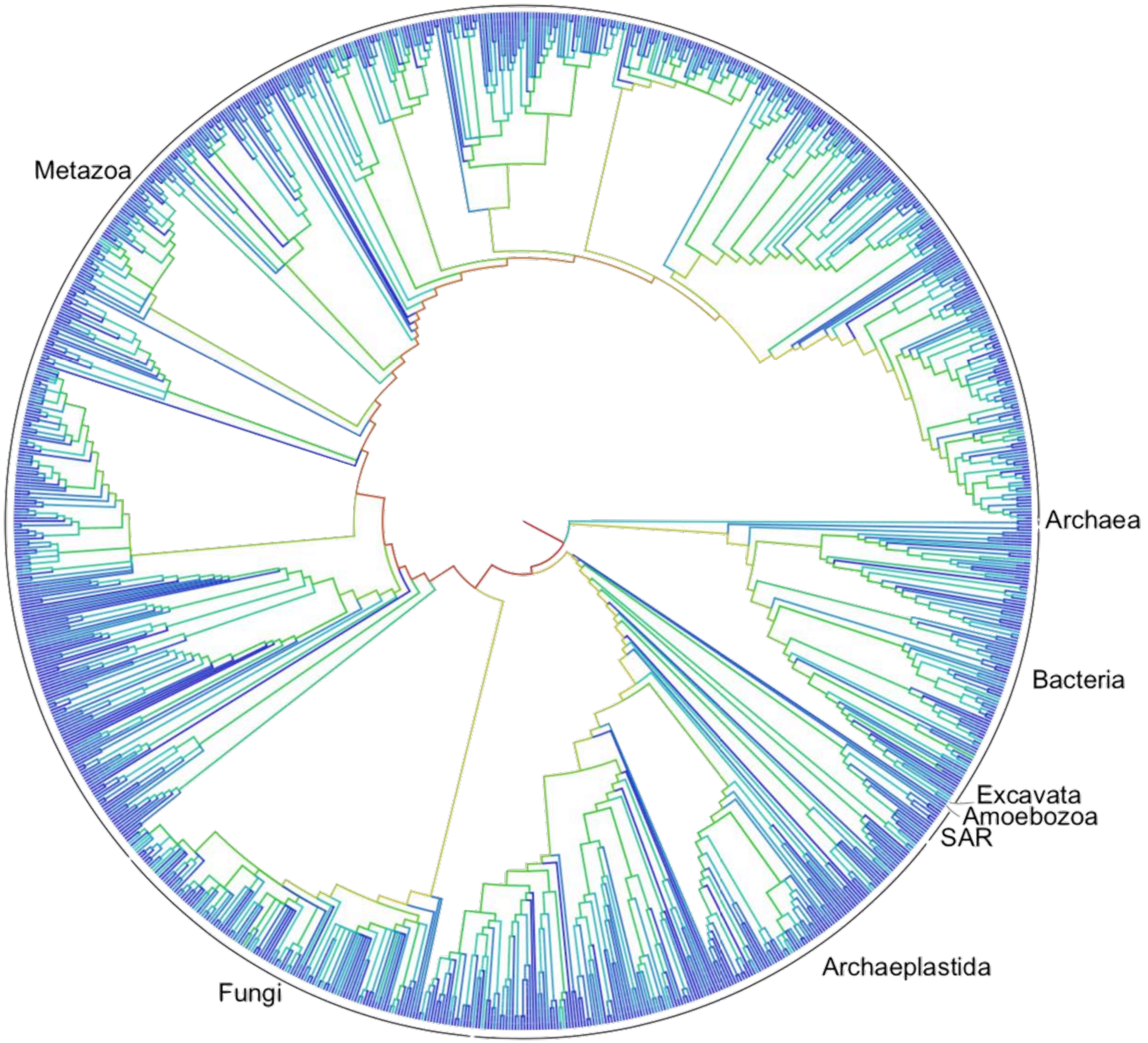
Nothing in evolution makes sense except when seen in the light of phylogeny.

— Jay Savage, 1997

The Tree of Life

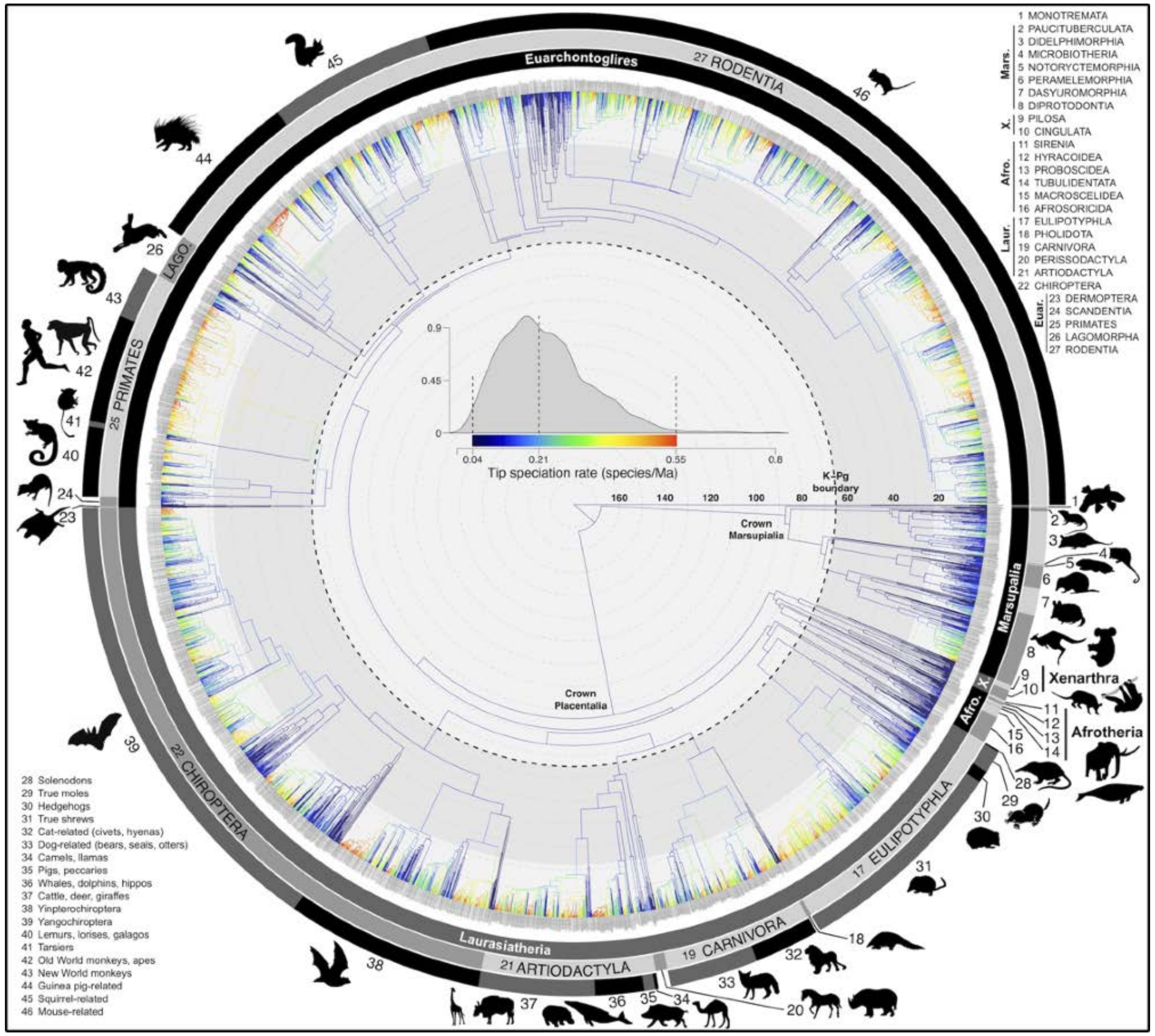
Three billion years the Tree has grown
From replicators' first seed sown
To branches rich with progeny:
The wonder of phylogeny.

excerpt from the poem "The Tree of Life"
by David Maddison



Open Tree of Life Synthetic Tree from Hinchliff et al. (PNAS 2015)

Systematics and Phylogenies



The study of biodiversity

How are species delimited?

How are species organized and distributed?

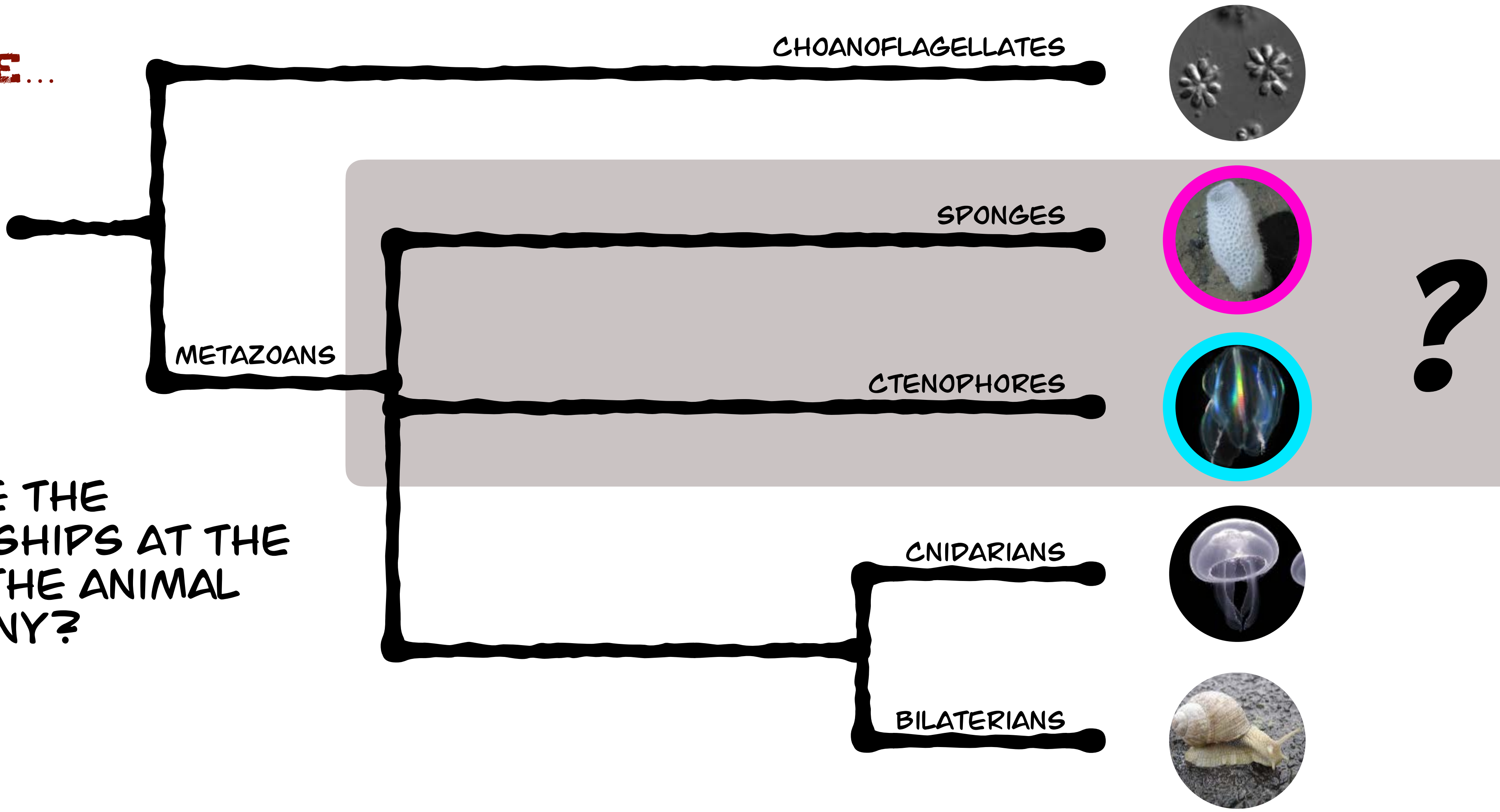
What evolutionary and ecological processes produced the patterns of biodiversity observed on Earth?

Phylogeny of mammals from Upham et al. (PLoS Biology 2019)

“Resolving” the Tree of Life

We can never know the *true* phylogeny of life on Earth…

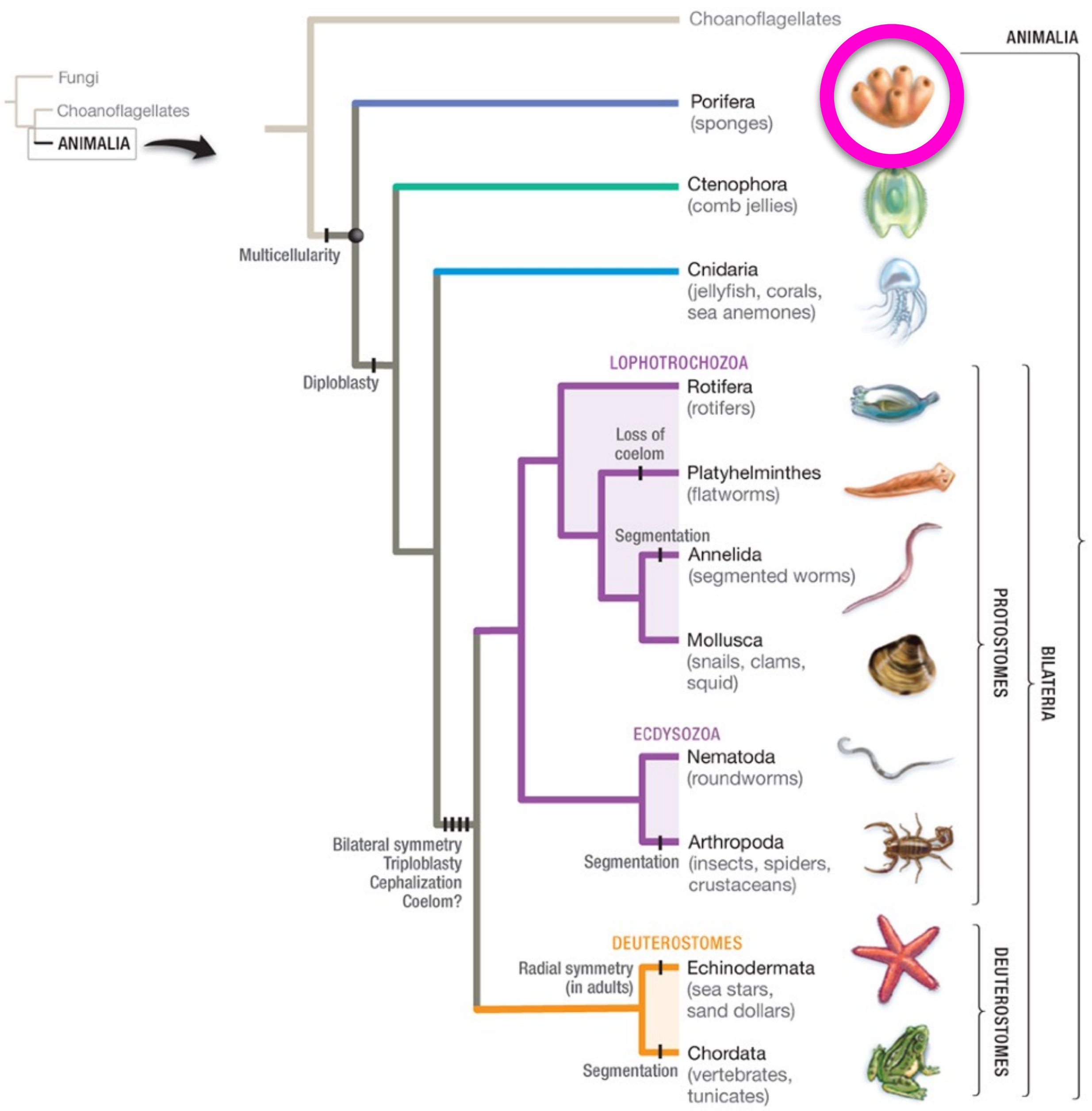
FOR EXAMPLE…



WHAT ARE THE RELATIONSHIPS AT THE ROOT OF THE ANIMAL PHYLOGENY?

…but we can work to consider all of the data and use rigorous inference methods to produce reliable estimates.

The Animal Phylogeny

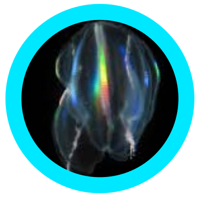


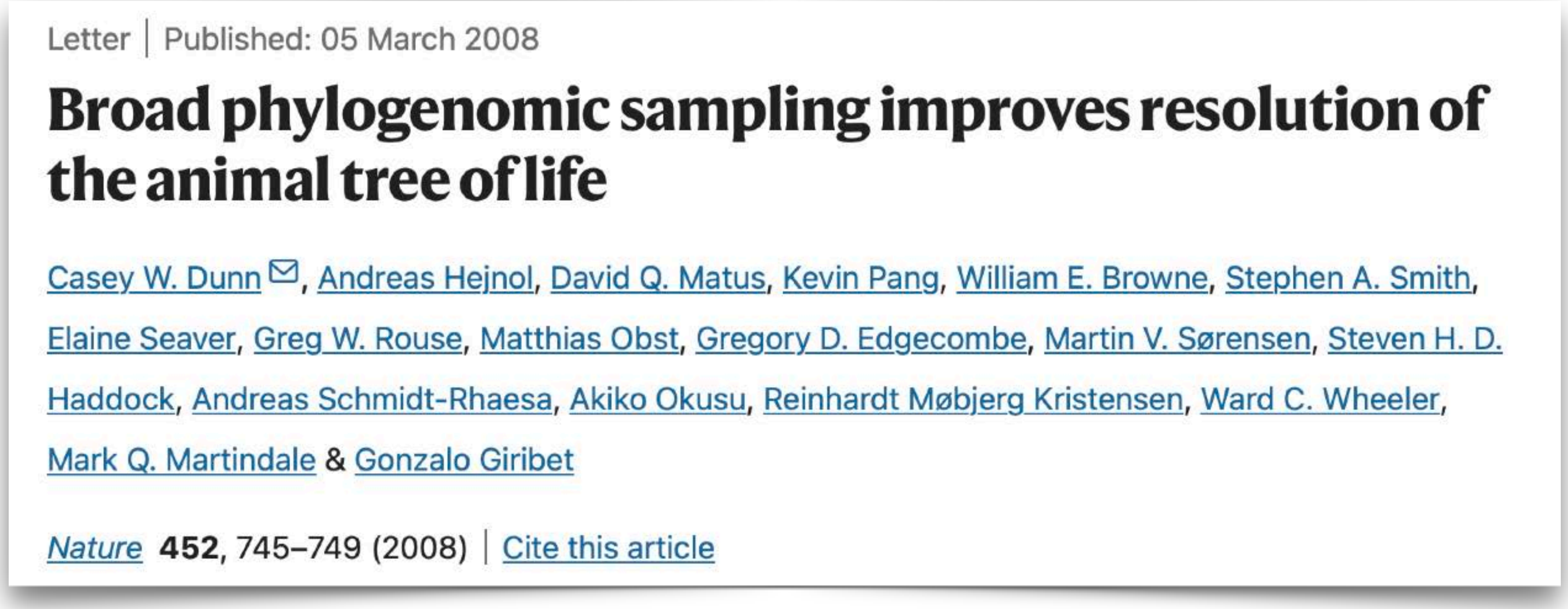
Until recently, there was little debate about the relationships at the root of the animal tree

Morphological evidence indicated that sponges were sister to all other animal lineages

Animal phylogeny from *Biological Science 8th Edition* by Freeman et al. (2024)

Ctenophores Spark Debate



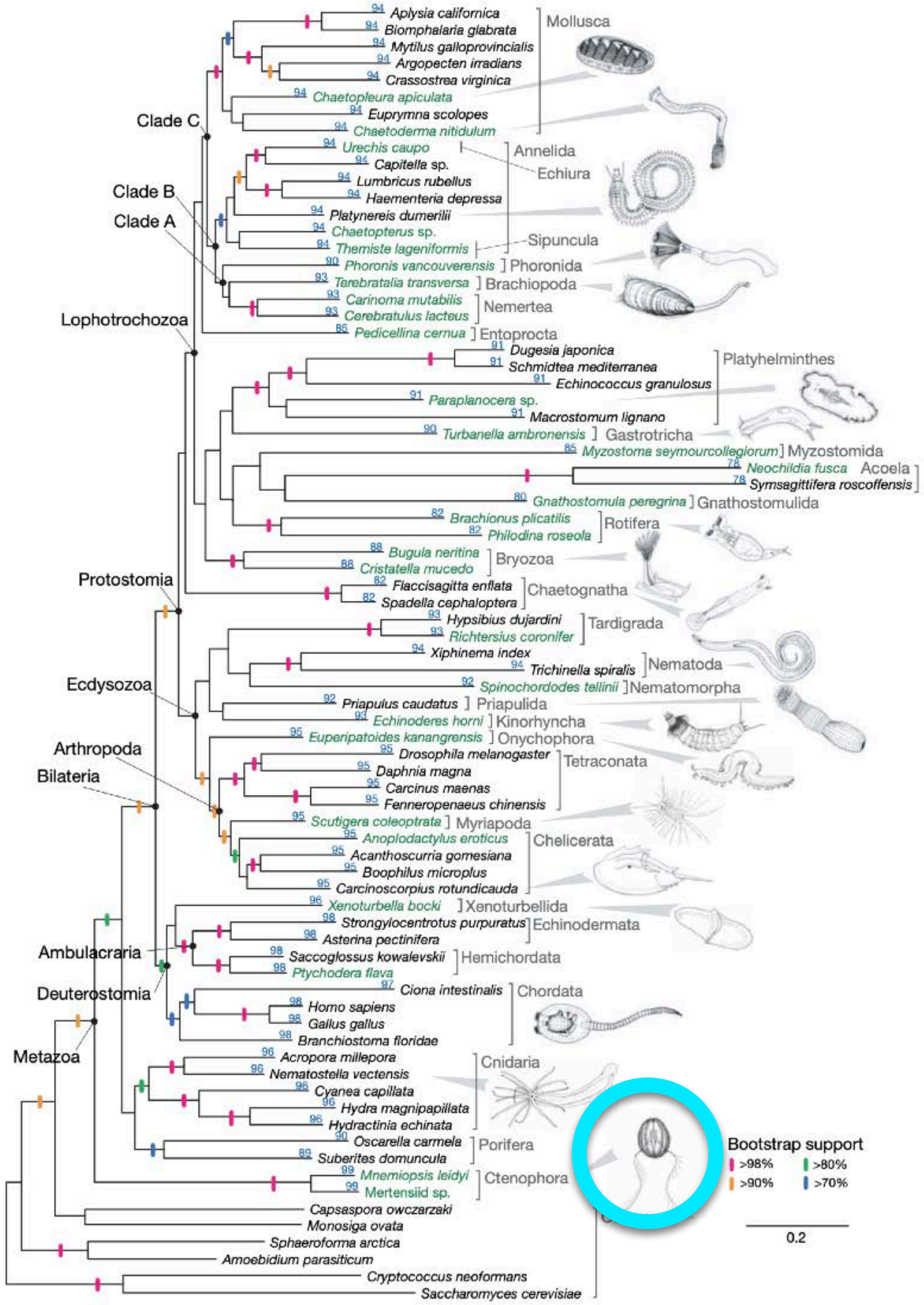


Letter | Published: 05 March 2008

Broad phylogenomic sampling improves resolution of the animal tree of life

Casey W. Dunn, Andreas Hejnol, David Q. Matus, Kevin Pang, William E. Browne, Stephen A. Smith, Elaine Seaver, Greg W. Rouse, Matthias Obst, Gregory D. Edgecombe, Martin V. Sørensen, Steven H. D. Haddock, Andreas Schmidt-Rhaesa, Akiko Okusu, Reinhardt Møbjerg Kristensen, Ward C. Wheeler, Mark Q. Martindale & Gonzalo Giribet

Nature **452**, 745–749 (2008) | Cite this article



CreatureCast video: http://www.nytimes.com/2014/01/23/science/earth/moving-without-feet-to-do-so.html

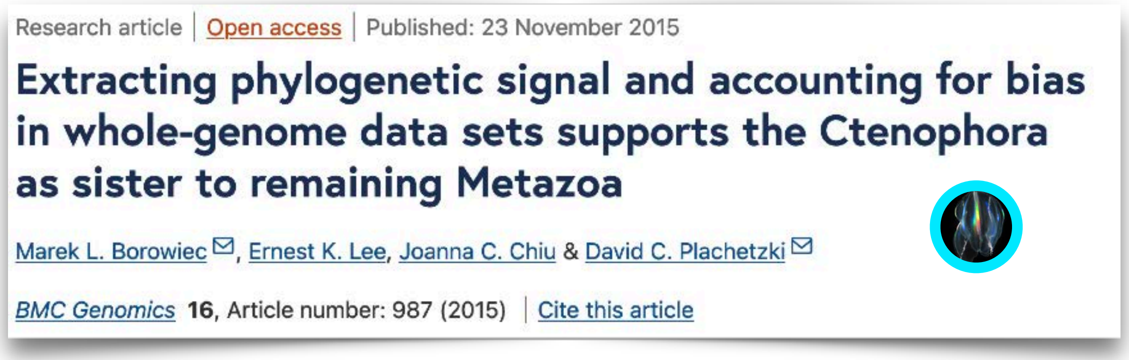
Who’s my sister?

PNAS

Error, signal, and the placement of Ctenophora sister to all other animals

Nathan V. Whelan[a,1], Kevin M. Kocot[b], Leonid L. Moroz[c], and Kenneth M. Halanych[a]

April 20, 2015 | 112 (18) 5773-5778 | https://doi.org/10.1073/pnas.1503453112



Research article | Open access | Published: 23 November 2015

Extracting phylogenetic signal and accounting for bias in whole-genome data sets supports the Ctenophora as sister to remaining Metazoa

Marek L. Borowiec ✉, Ernest K. Lee, Joanna C. Chiu & David C. Plachetzki ✉

BMC Genomics **16**, Article number: 987 (2015) | Cite this article

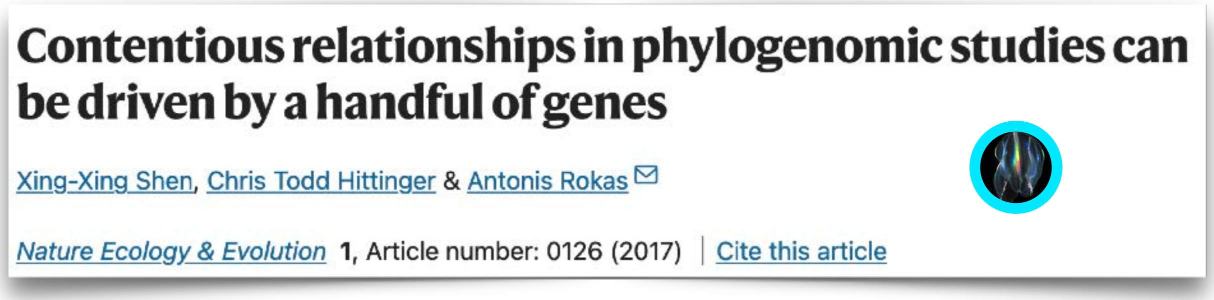
---

Current Biology
Article

ARTICLE · Volume 27, Issue 7, P958-967, April 03, 2017 ·

A Large and Consistent Phylogenomic Dataset Supports Sponges as the Sister Group to All Other Animals

Paul Simion[1,17,18] · Hervé Philippe[2,3,17] ✉ · Denis Baurain[4] · Muriel Jager[1] · Daniel J. Richter[5,6,7] · Arnaud Di Franco[2] · Béatrice Roure[2,3] · Nori Satoh[8] · Éric Quéinnec[1] · Alexander Ereskovsky[9,10] · Pascal Lapébie[11] · Erwan Corre[12,13] · Frédéric Delsuc[14] · Nicole King[5] · Gert Wörheide[15,16] · Michaël Manuel[1,19]

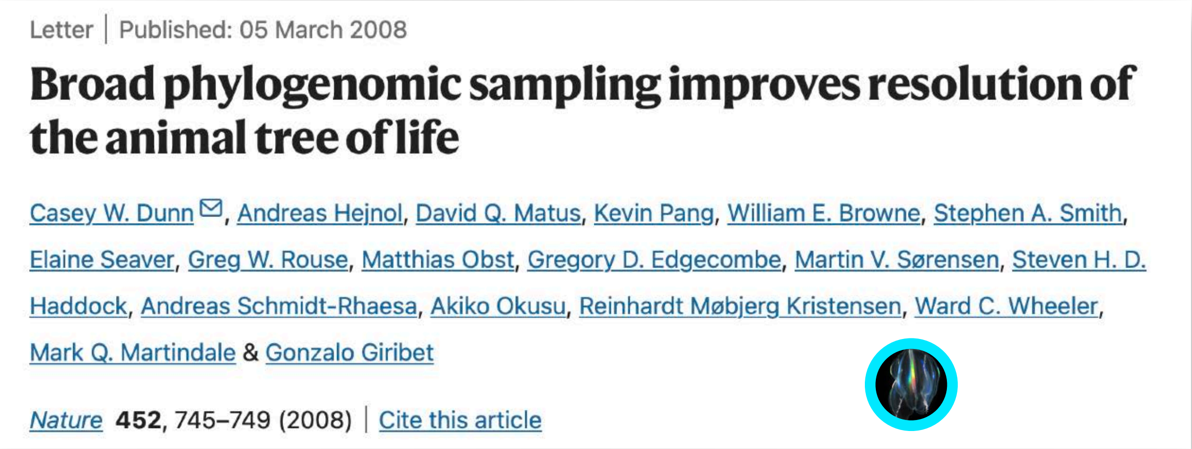


Contentious relationships in phylogenomic studies can be driven by a handful of genes

Xing-Xing Shen, Chris Todd Hittinger & Antonis Rokas ✉

Nature Ecology & Evolution **1**, Article number: 0126 (2017) | Cite this article

---

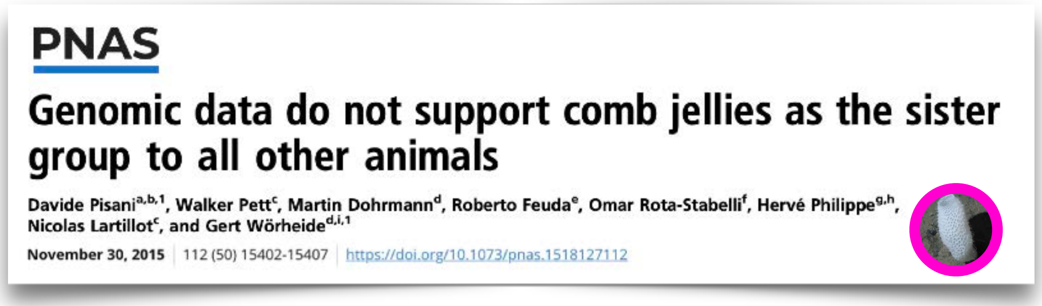


Letter | Published: 05 March 2008

Broad phylogenomic sampling improves resolution of the animal tree of life

Casey W. Dunn ✉, Andreas Hejnol, David Q. Matus, Kevin Pang, William E. Browne, Stephen A. Smith, Elaine Seaver, Greg W. Rouse, Matthias Obst, Gregory D. Edgecombe, Martin V. Sørensen, Steven H. D. Haddock, Andreas Schmidt-Rhaesa, Akiko Okusu, Reinhardt Møbjerg Kristensen, Ward C. Wheeler, Mark Q. Martindale & Gonzalo Giribet

Nature **452**, 745–749 (2008) | Cite this article



PNAS

Genomic data do not support comb jellies as the sister group to all other animals

Davide Pisani[a,b,1], Walker Pett[c], Martin Dohrmann[d], Roberto Feuda[e], Omar Rota-Stabelli[f], Hervé Philippe[g,h], Nicolas Lartillot[c], and Gert Wörheide[d,i,1]

November 30, 2015 | 112 (50) 15402-15407 | https://doi.org/10.1073/pnas.1518127112

The Guardian

Evolution row ends as scientists declare sponges to be sister of all other animals

Nicola Davis

Thu 30 Nov 2017 12.00 EST

Current Biology
Report

REPORT · Volume 27, Issue 24, P3864-3870.E4, December 18, 2017 ·

Improved Modeling of Compositional Heterogeneity Supports Sponges as Sister to All Other Animals

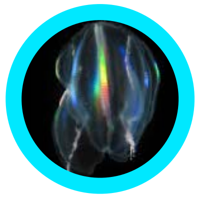
Roberto Feuda[1,11] · Martin Dohrmann[2,11] · Walker Pett[3] · Hervé Philippe[4,5] · Omar Rota-Stabelli[6] · Nicolas Lartillot[7] · Gert Wörheide[2,8,9] ✉ · Davide Pisani[10,12]

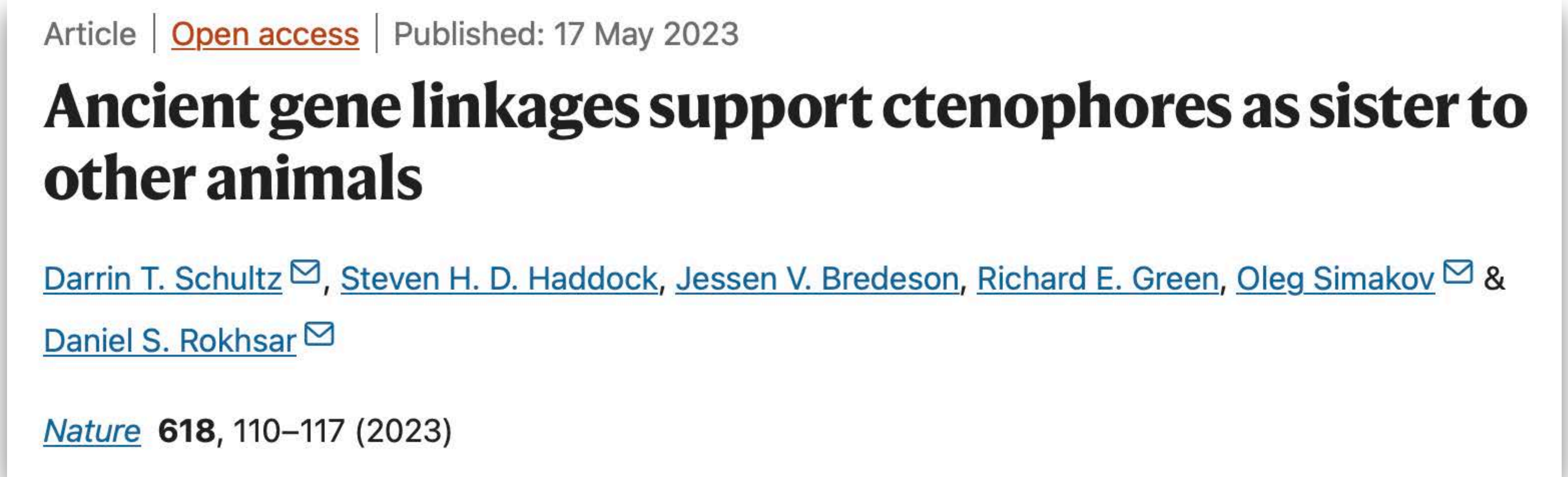
Evidence for sponges as sister to all other animals from partitioned phylogenomics with mixture models and recoding

Anthony K. Redmond & Aoife McLysaght ✉

Nature Communications **12**, Article number: 1783 (2021) | Cite this article

New Evidence



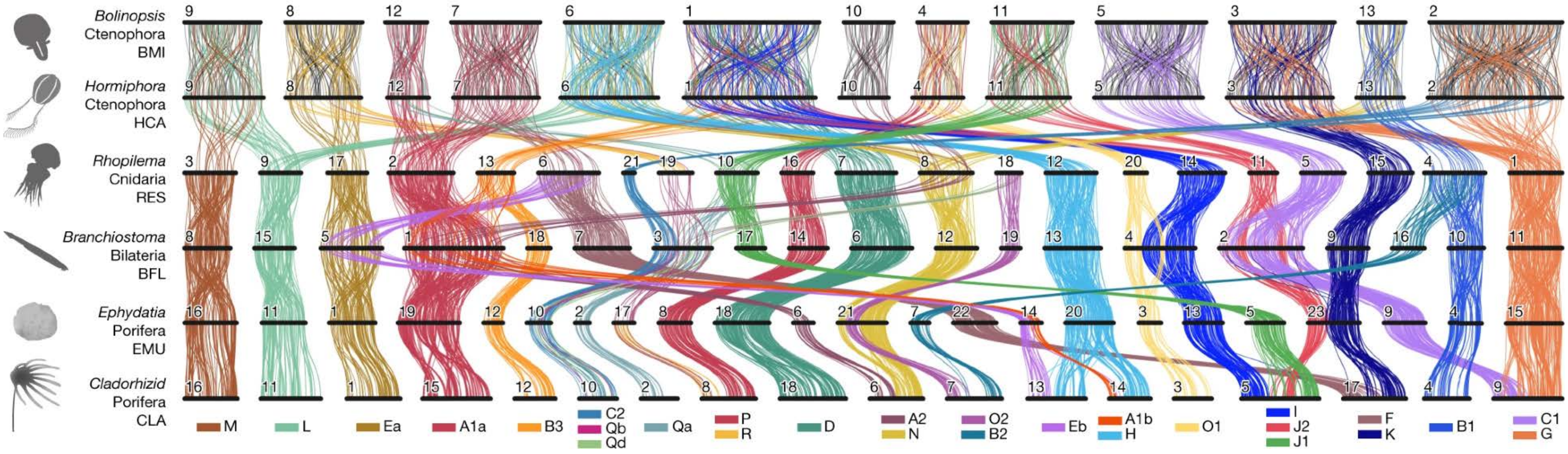


Article | Open access | Published: 17 May 2023

Ancient gene linkages support ctenophores as sister to other animals

Darrin T. Schultz, Steven H. D. Haddock, Jessen V. Bredeson, Richard E. Green, Oleg Simakov & Daniel S. Rokhsar

Nature **618**, 110–117 (2023)



Large-Scale Gene Order Changes Are Rare

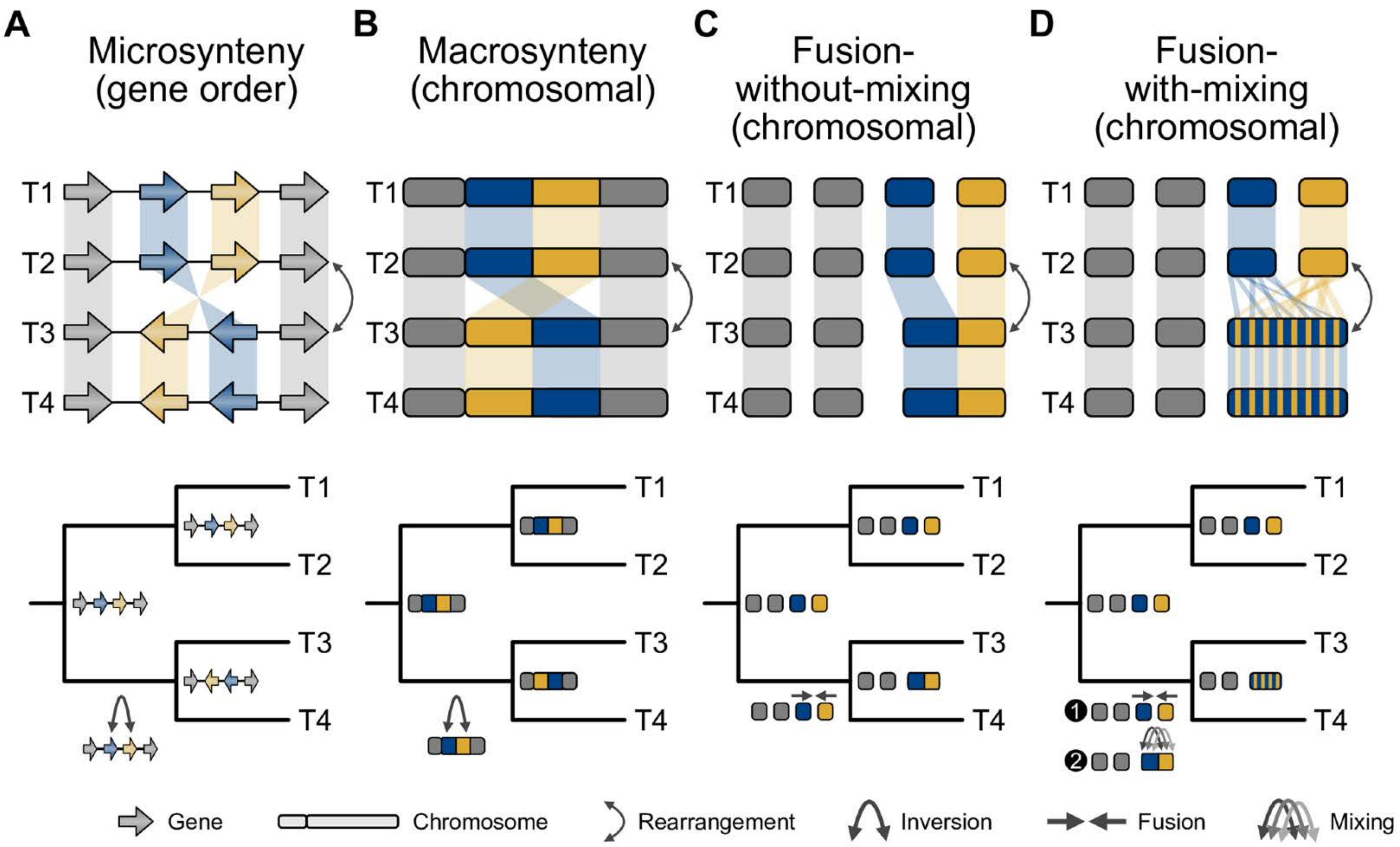
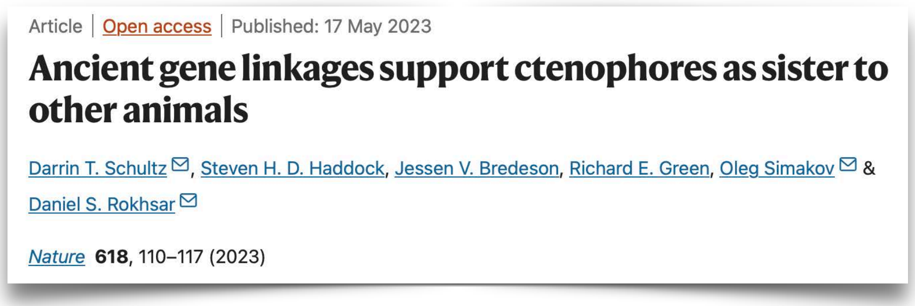


Figure from Steenwyk & King ([PLoS Biology, 2024]())

Conserved Synteny

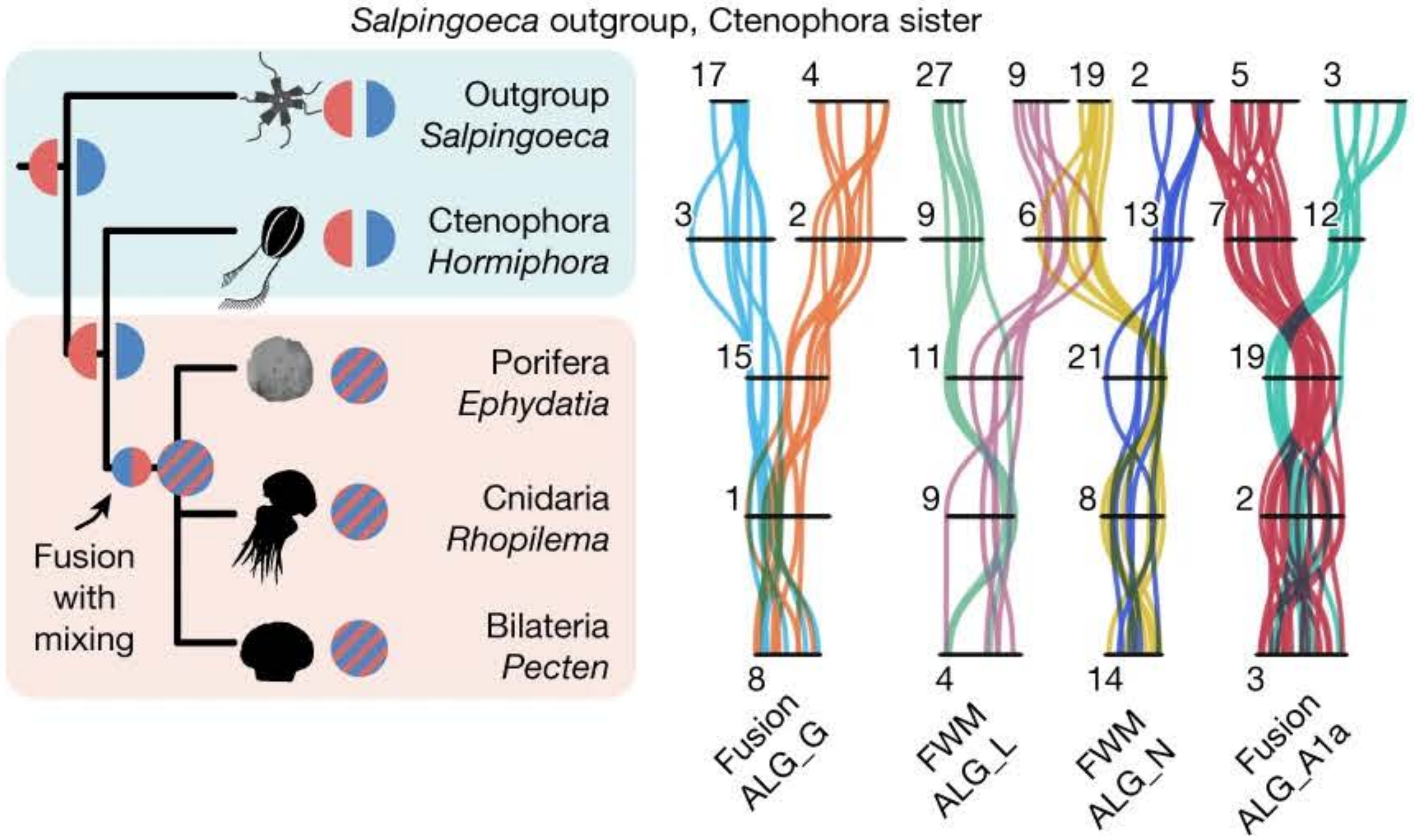


Article | Open access | Published: 17 May 2023

Ancient gene linkages support ctenophores as sister to other animals

Darrin T. Schultz, Steven H. D. Haddock, Jessen V. Bredeson, Richard E. Green, Oleg Simakov & Daniel S. Rokhsar

Nature **618**, 110–117 (2023)

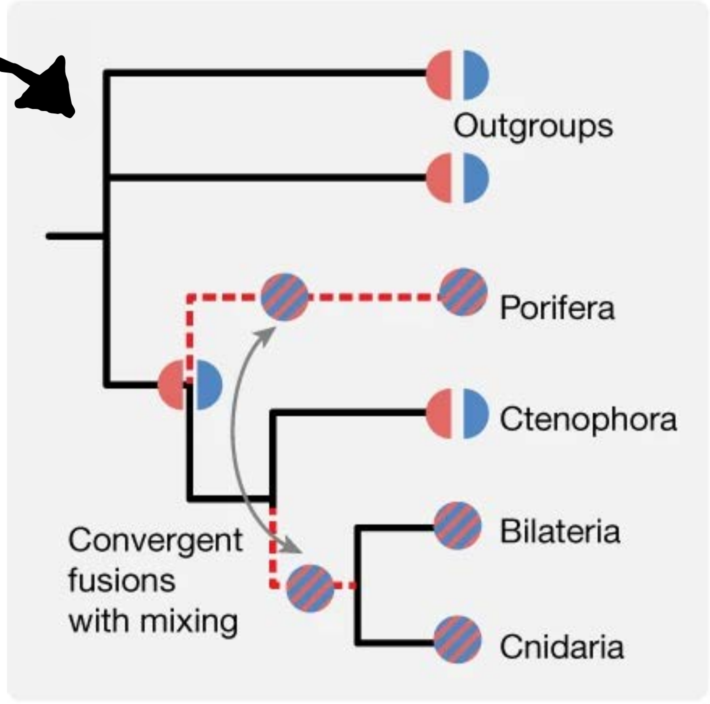

Salpingoeca outgroup, Ctenophora sister

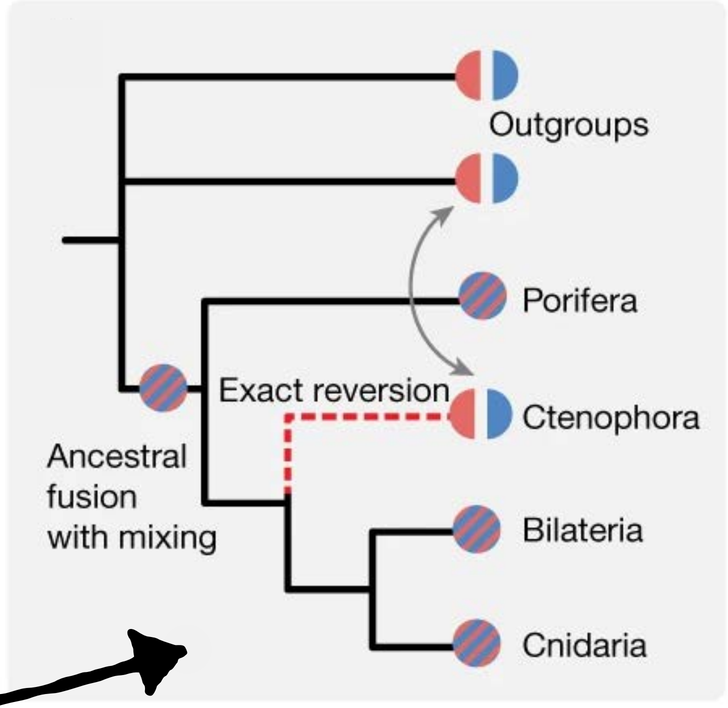
Unlikely changes are needed for these patterns to exist if sponges are sister to all other animals

Convergent fusions followed by mixing

or

Exact unmixing followed by fissions

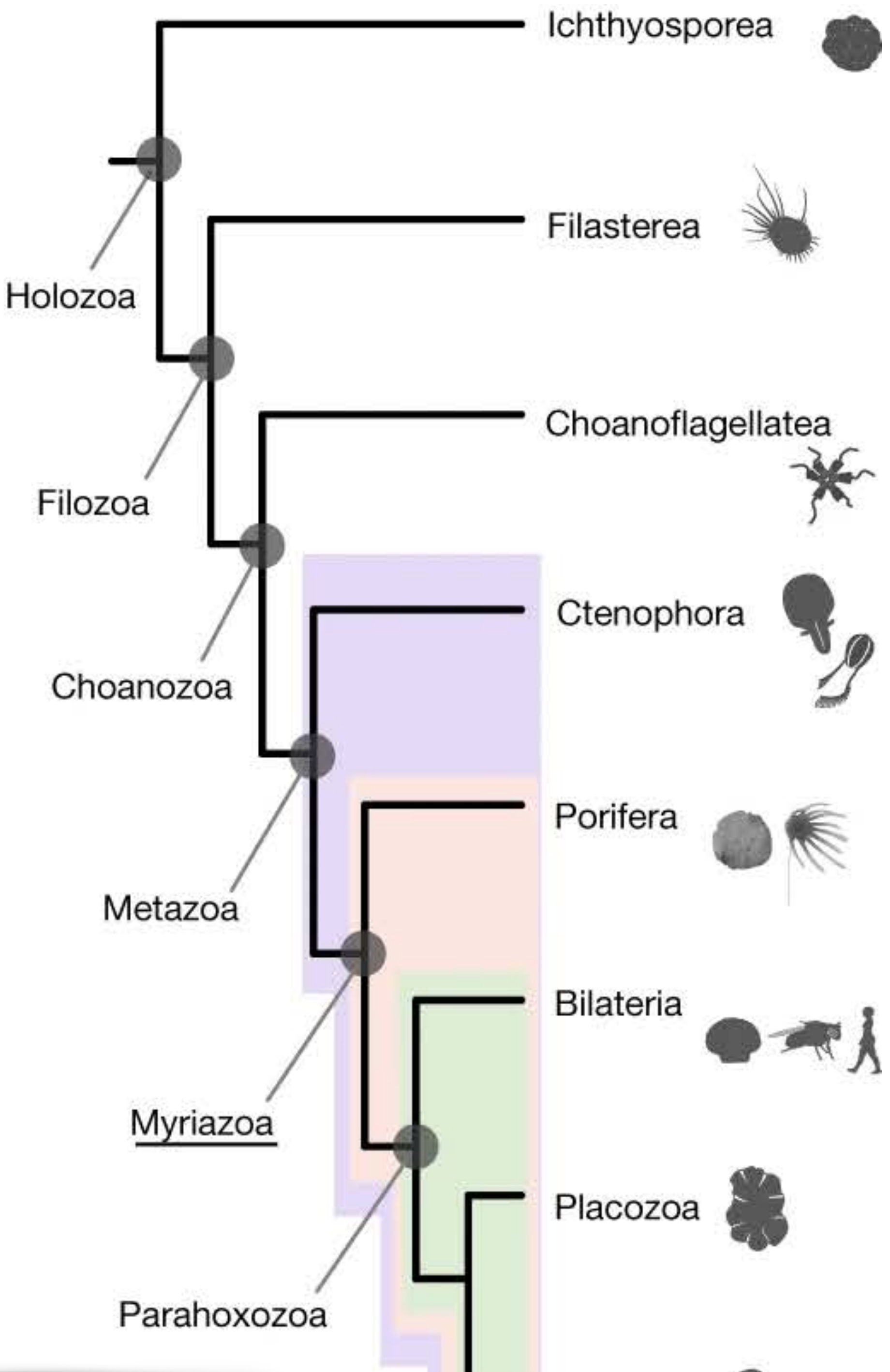




Phylogenies Give Evolutionary Context

Understanding the complex history of animals requires more research on ctenophore & sponge biology since these groups are't well studied & often misrepresented

The evolution of the neuromuscular system is more complicated than we originally assumed



Syncytial nerve net in a ctenophore adds insights on the evolution of nervous systems

PAWEL BURKHARDT, JEFFREY COLGREN, ASTRID MEDHUS, LEONID DIGEL, BENJAMIN NAUMANN, JOAN J. SOTO-ANGEL, EVA-LENA NORDMANN, MARIA Y. SACHKOVA, AND MAIKE KITTELMANN Authors Info & Affiliations

SCIENCE • 20 Apr 2023 • Vol 380, Issue 6642 • pp. 293-297 • DOI: 10.1126/science.ade5645

And new data and/or methods may lead to new insights…

Sponge v comb jellies: which was evolution's first trailblazer?

Michael Marshall

Sat 9 Sep 2023 08.00 EDT

Figure from Schultz et al. (Nature, 2023)

Misconceptions

A jelly is the ultimate ancestor of all animal life, finds study

For nearly 100 years, scientists have thought a sponge was the first type of animal on the planet, but new genetic technique points to a creature like today's comb jellies

Ian Johnston Science Correspondent • Monday 10 April 2017 16:02 BST •

INDEPENDENT

600-800 million years of evolution gave us ctenophores, sponges, jellyfish, and humans… so it is not correct to say “our ancestors were comb jellies”

SMART NEWS

Comb Jellies May Be the World’s Oldest Animal Group

A new study suggests that ancestors of comb jellies, not sponges, were the first to break off from the common ancestor of all animals

Margaret Osborne
Daily Correspondent
May 18, 2023

Smithsonian MAGAZINE

The most-recent common ancestor of all known ctenophores is younger than the ancestor of sponges, bilaterians, & cnidarians. A lot has happened during the history of animals that we don’t know…

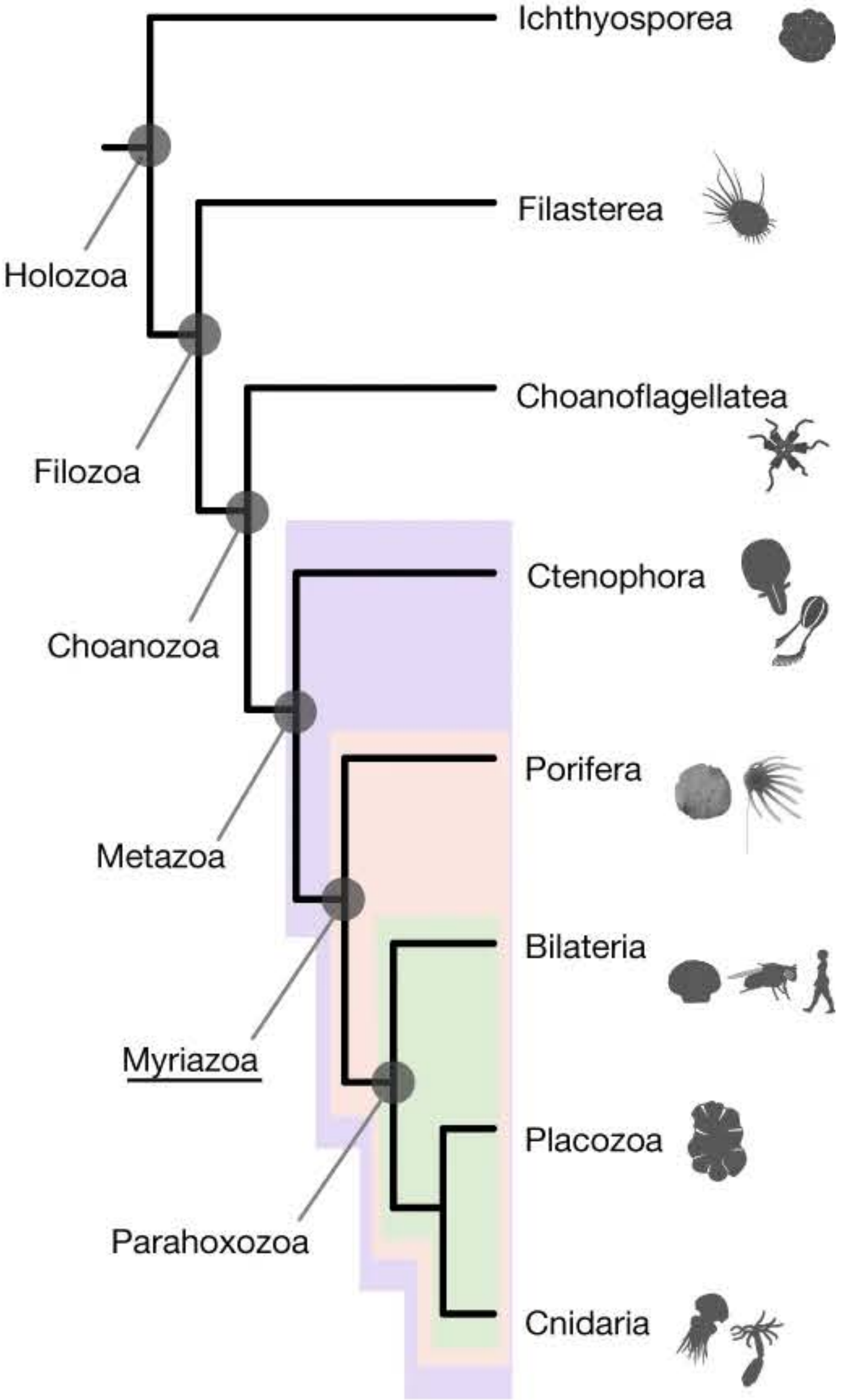
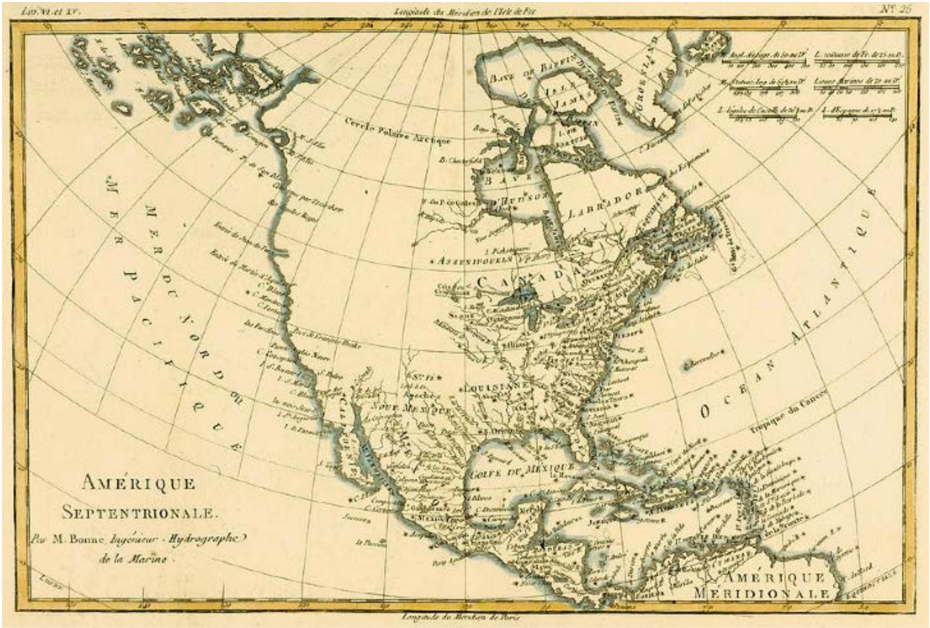


Figure from Schultz et al. (Nature, 2023)

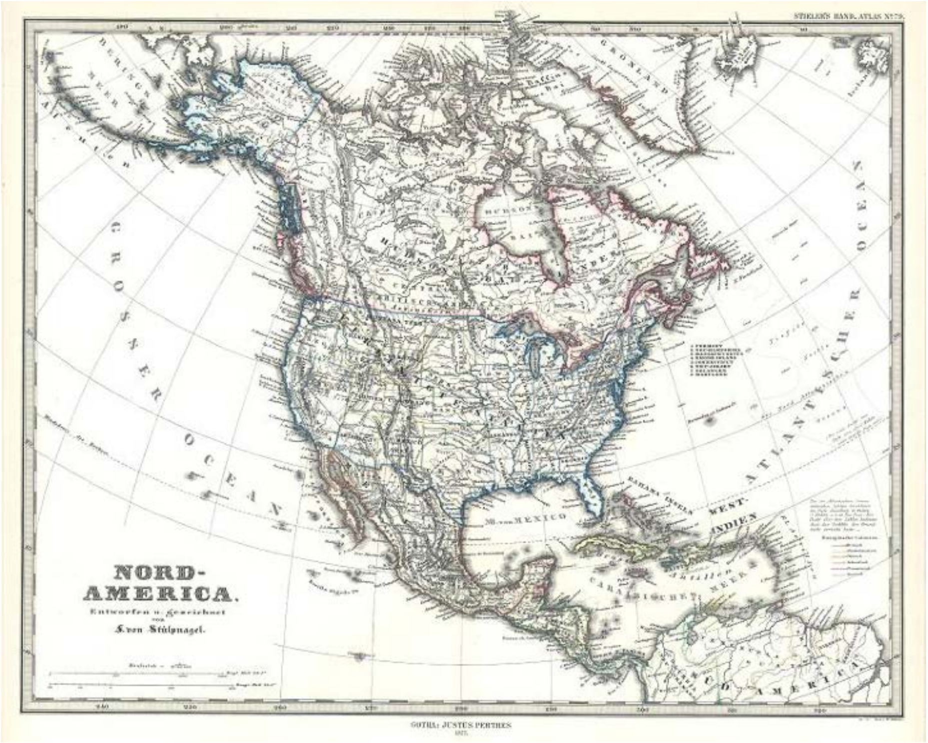
Estimates Change with More Data

1780



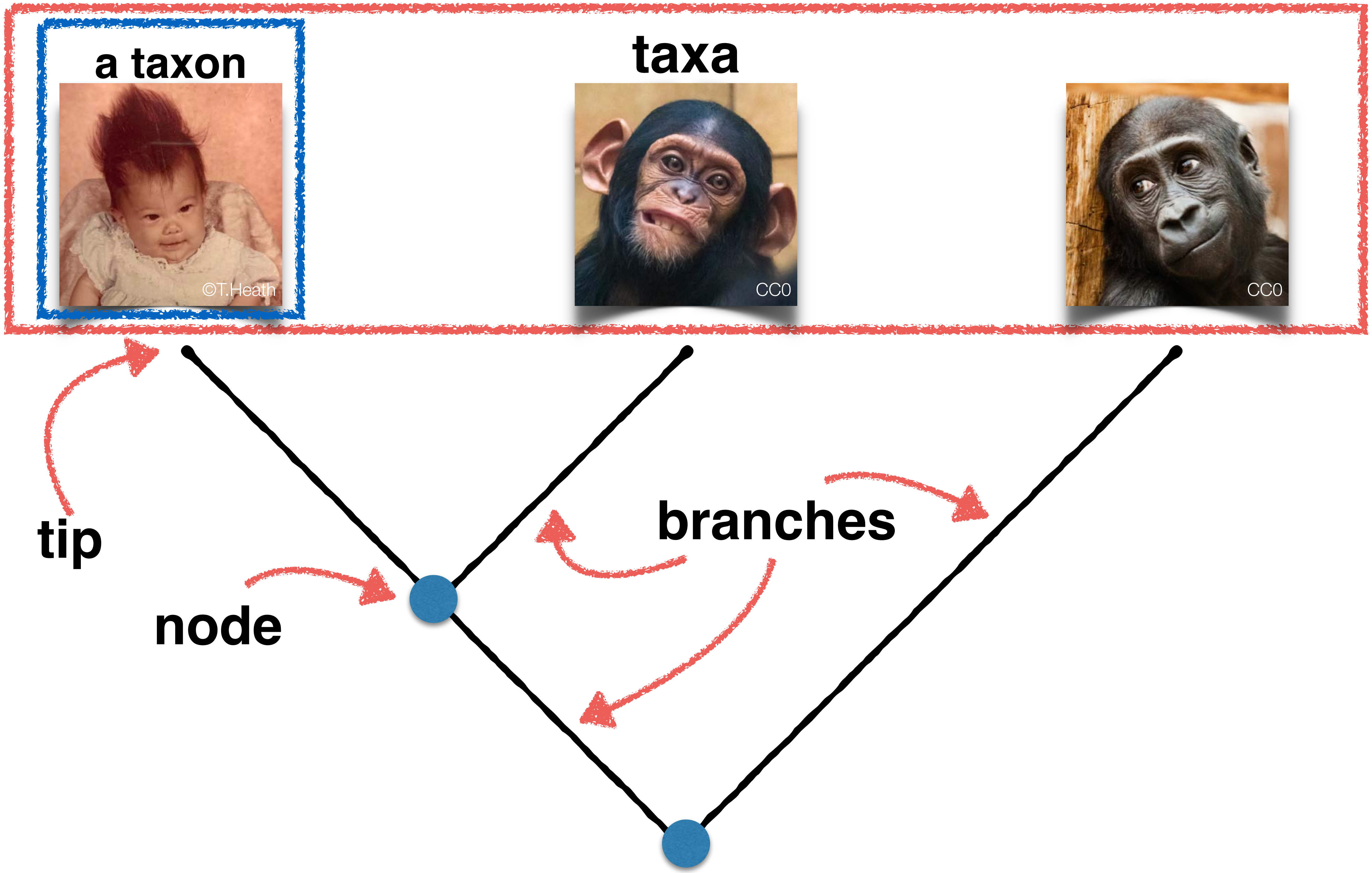
Like maps, phylogenetic trees are **estimates** that can change (and improve) with **more data and better methods**

1877



The Metazoa phylogeny may change again…

Phylogeny Terminology & Concepts



a taxon

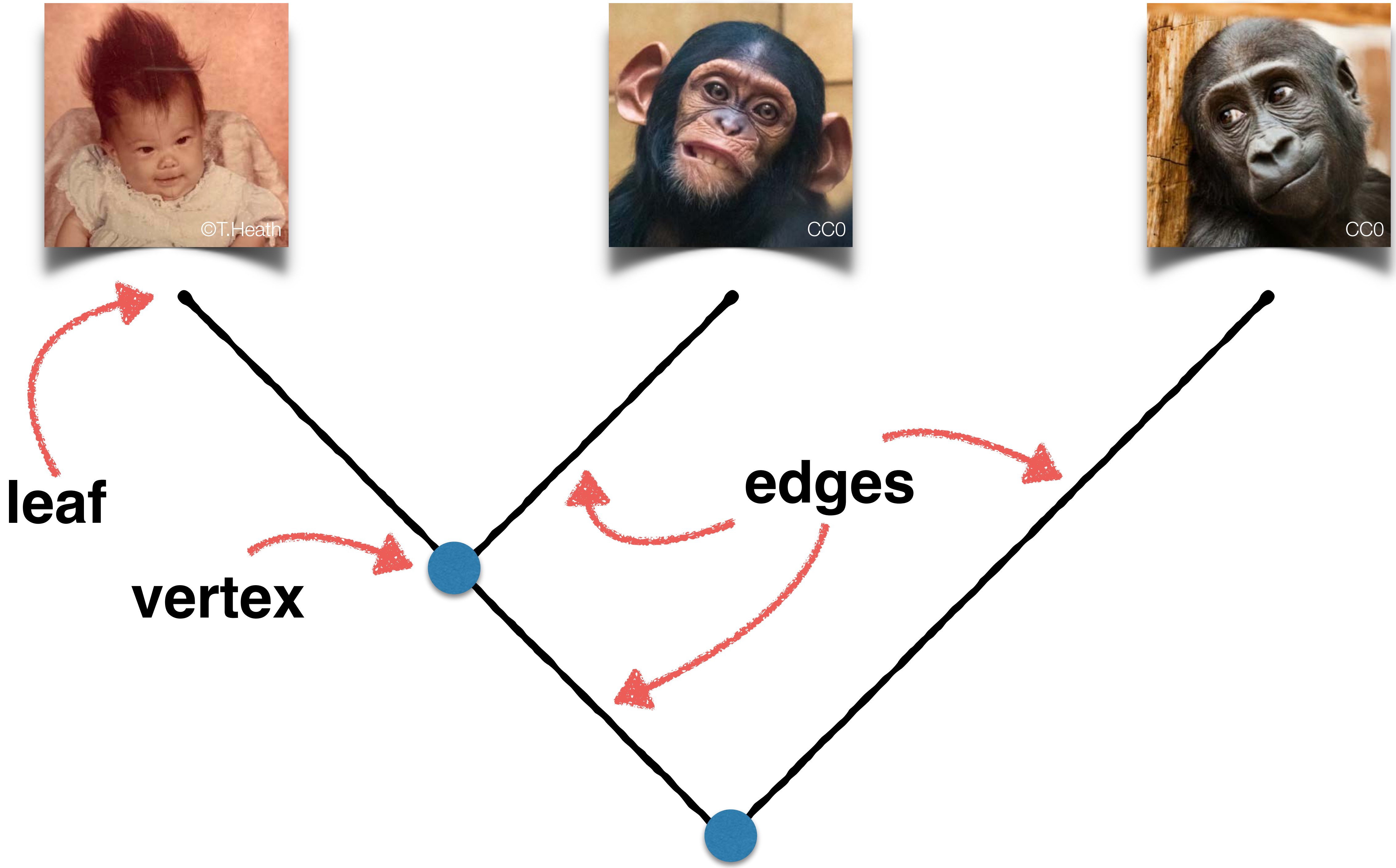
taxa

tip

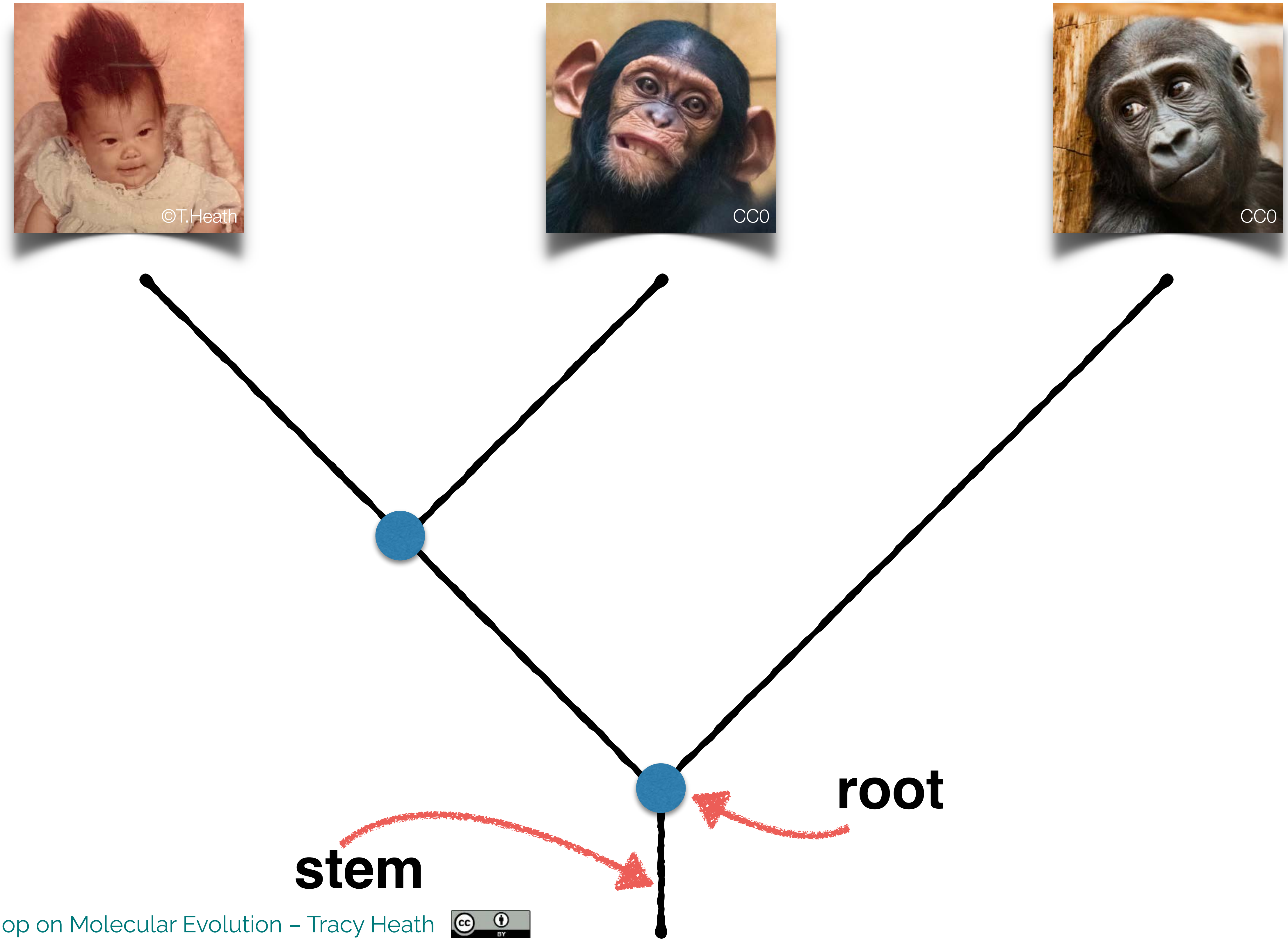
node

branches

Phylogeny Terminology & Concepts



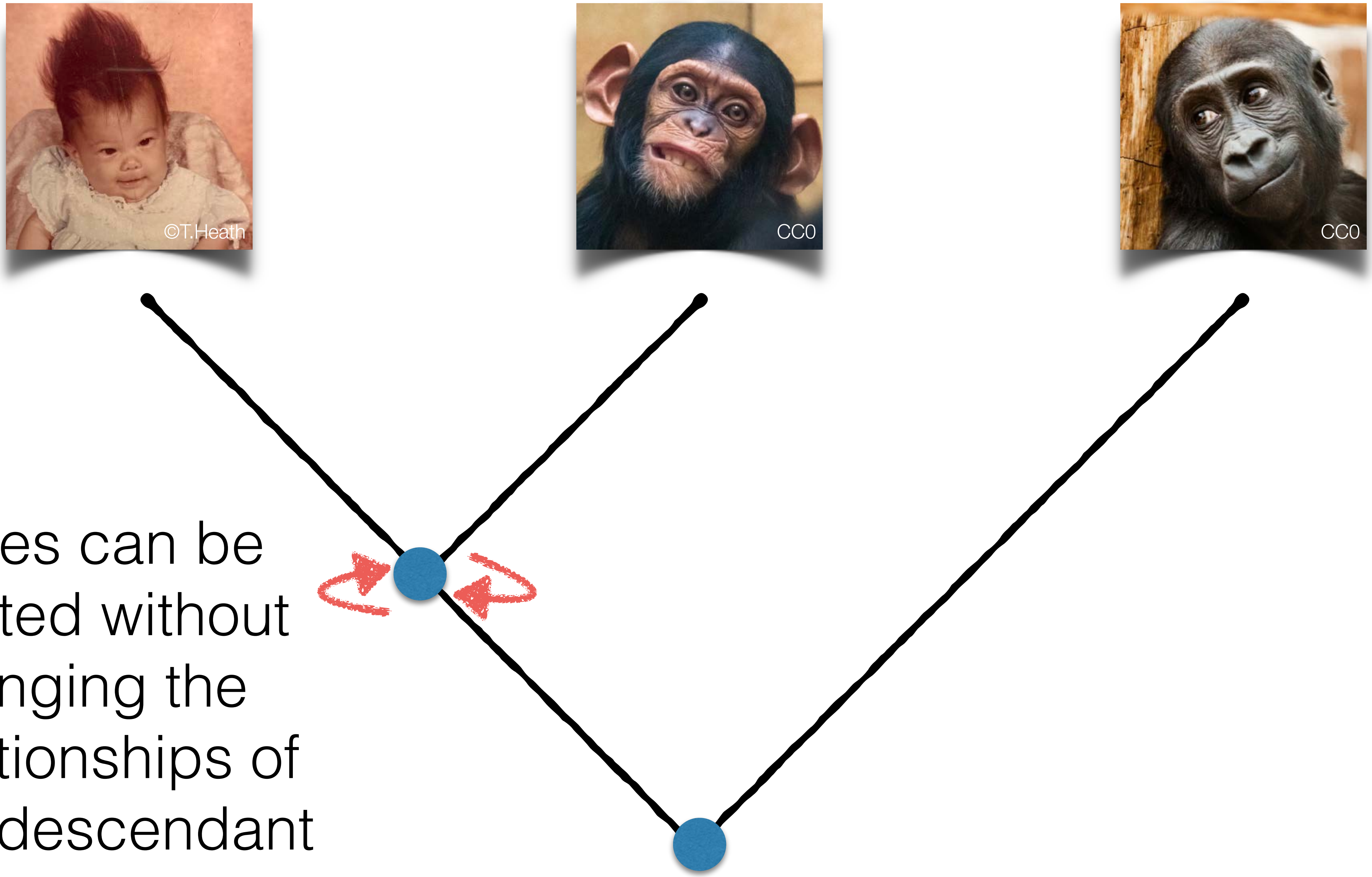
Phylogeny Terminology & Concepts



root

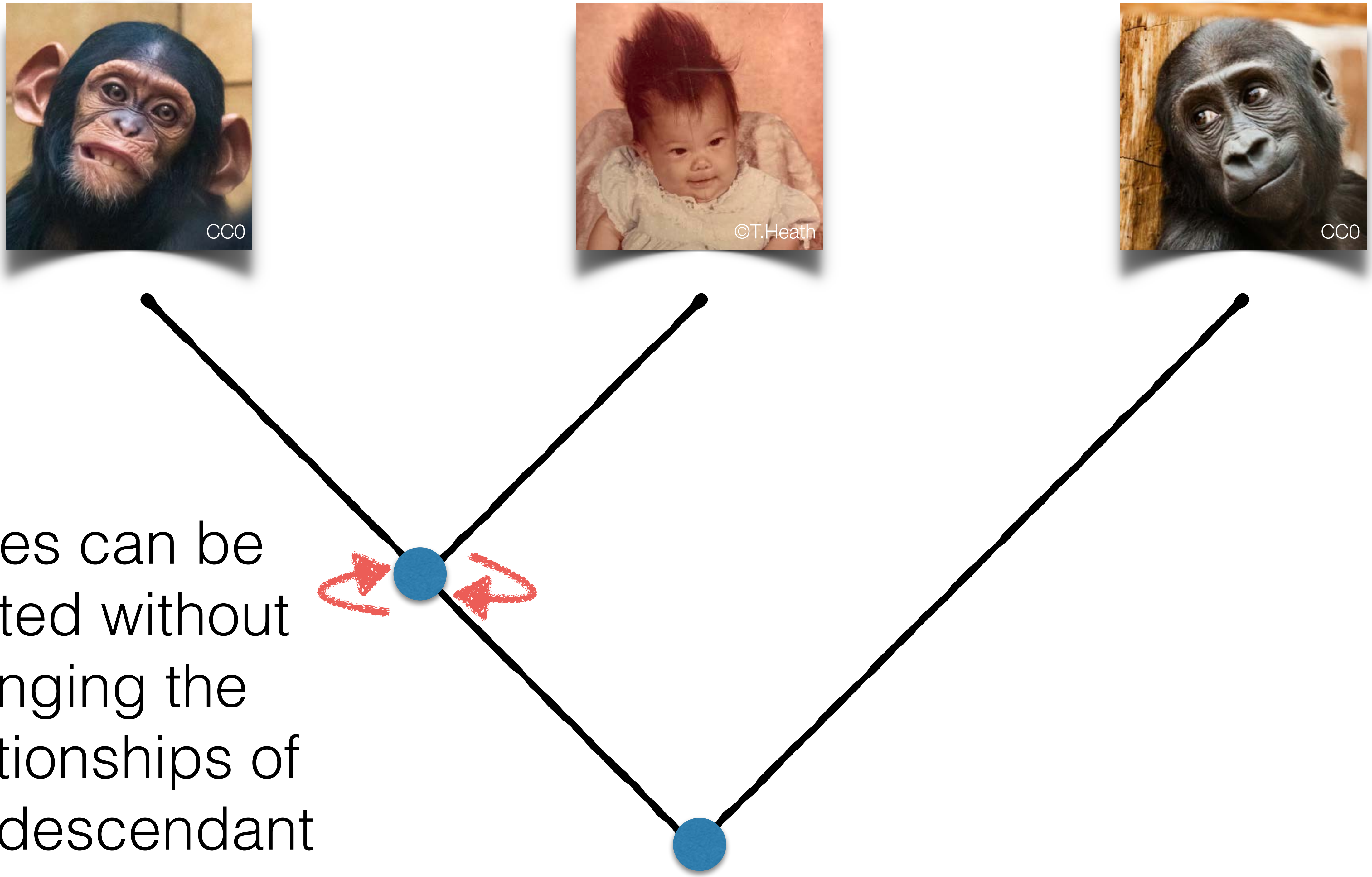
stem

Phylogeny Terminology & Concepts



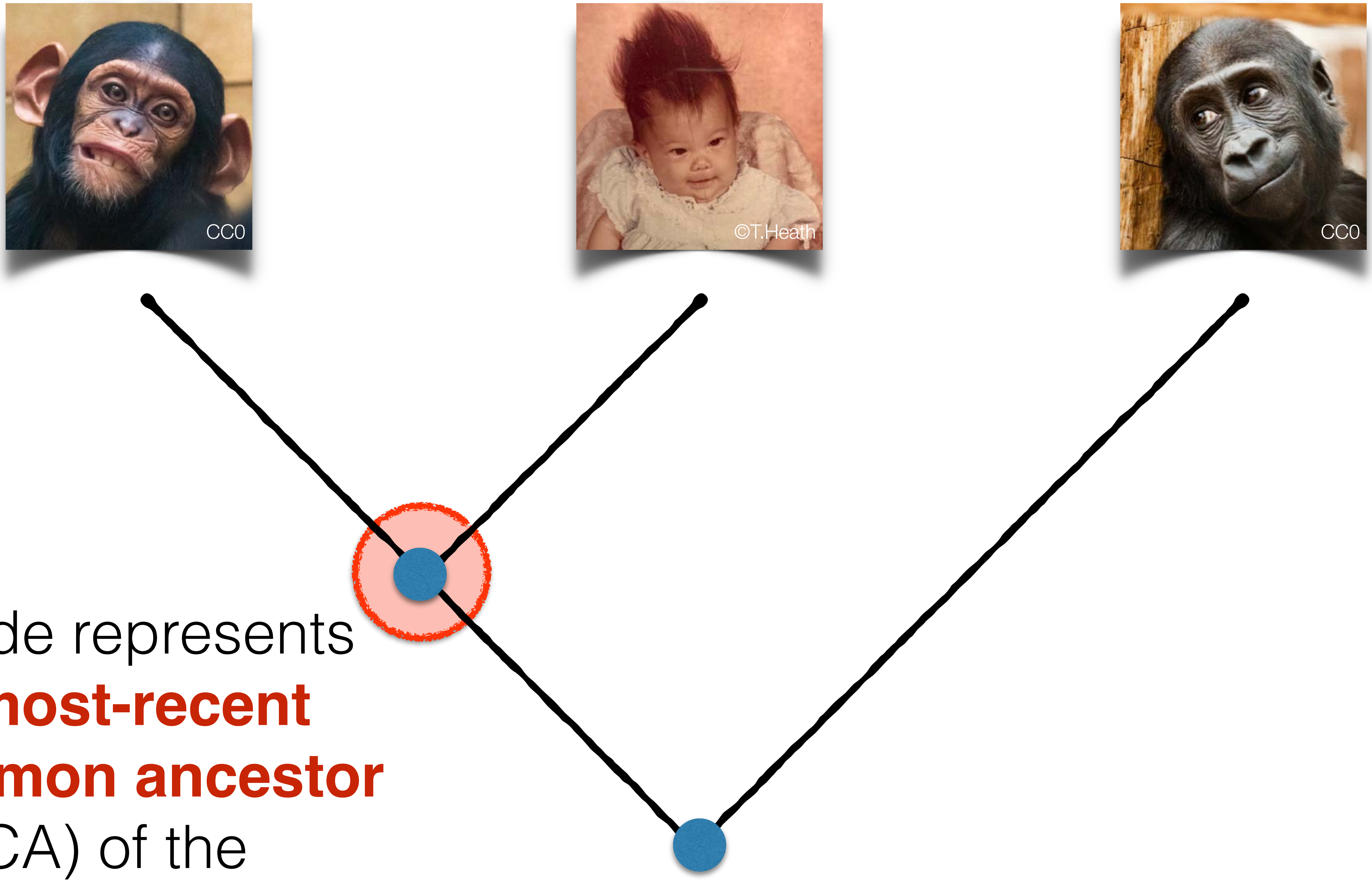
nodes can be rotated without changing the relationships of the descendant branches

Phylogeny Terminology & Concepts



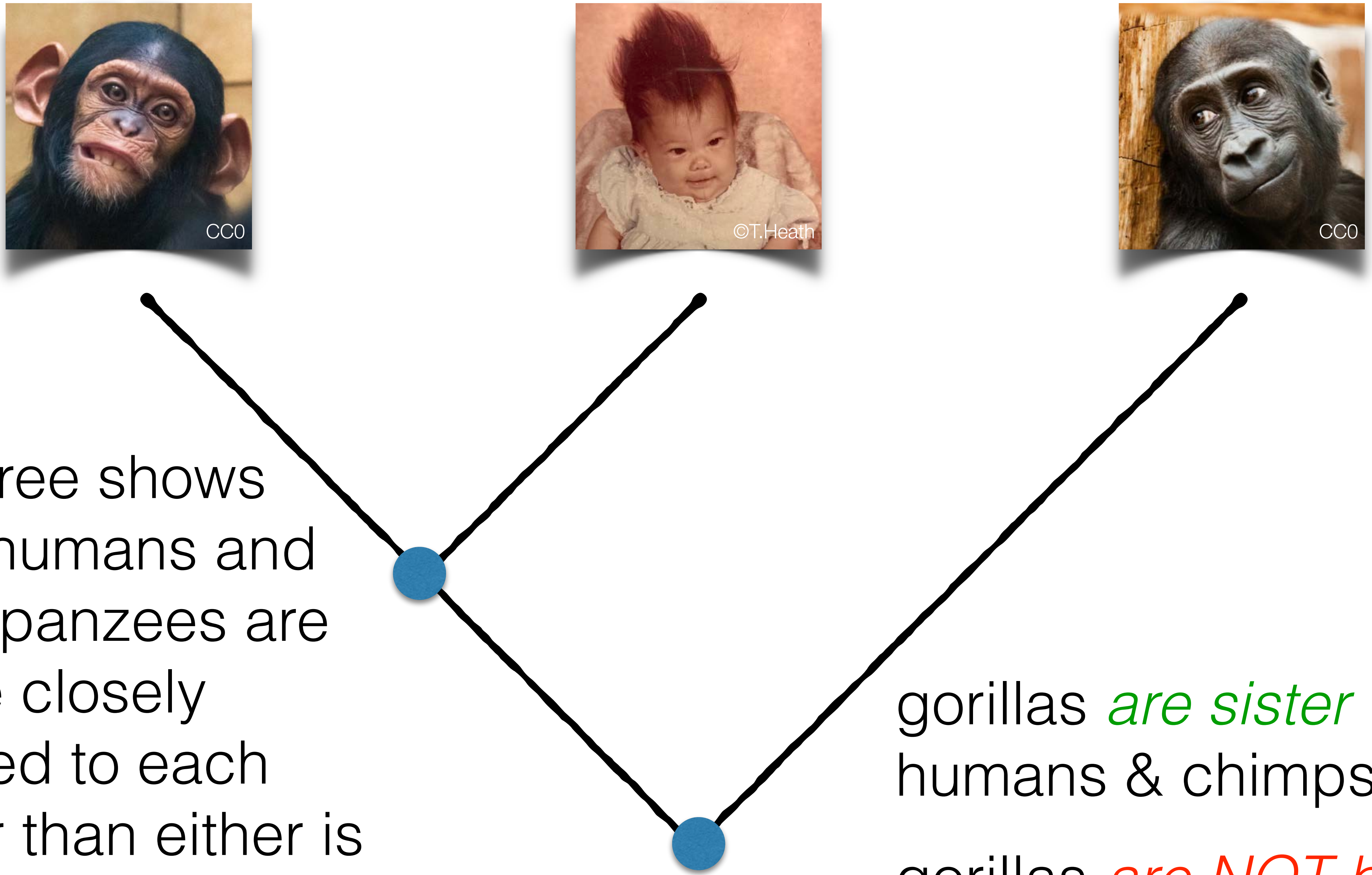
nodes can be rotated without changing the relationships of the descendant branches

Phylogeny Terminology & Concepts



a node represents the **most-recent common ancestor** (MRCA) of the descendant taxa

Phylogeny Terminology & Concepts

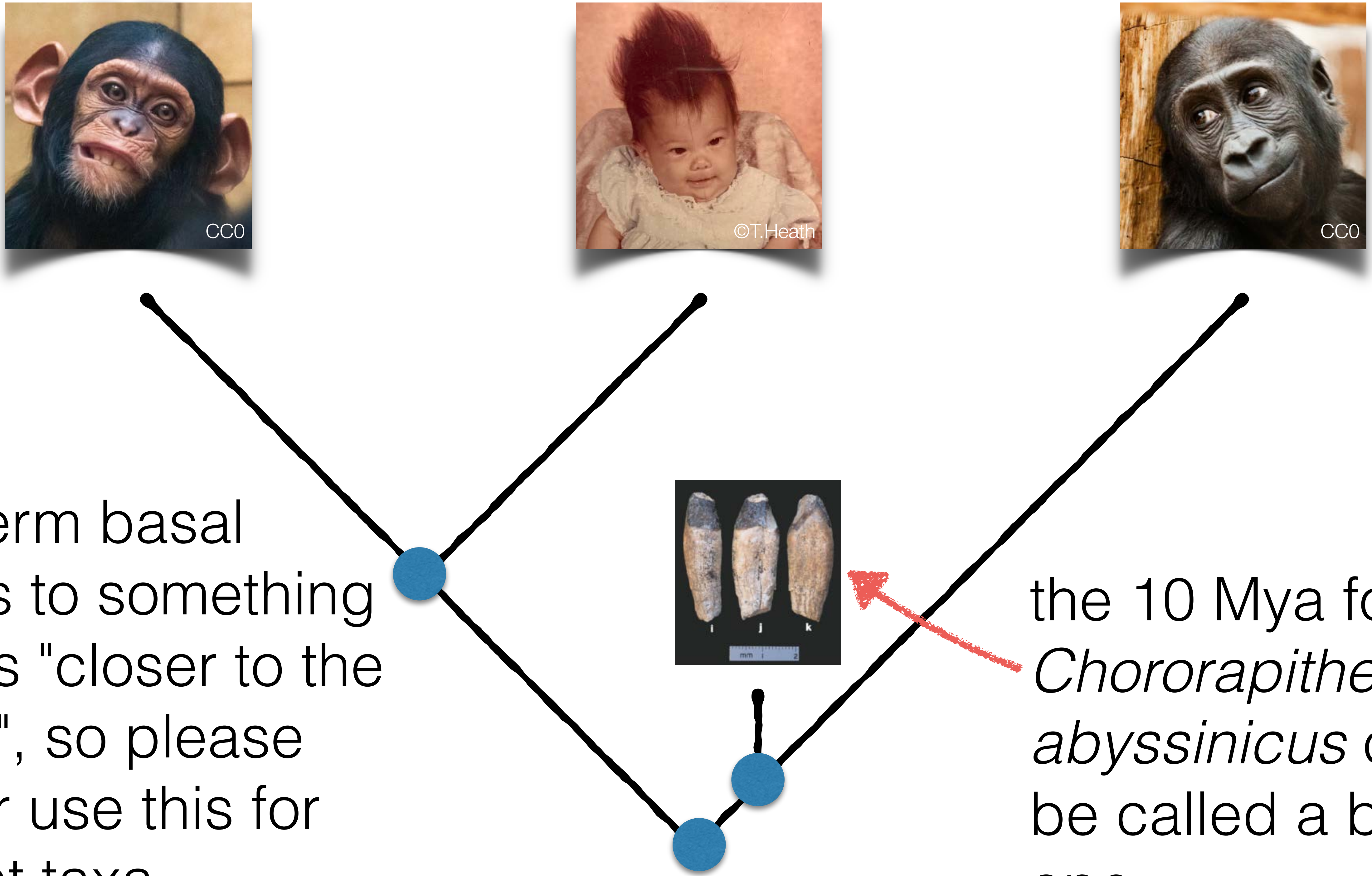


this tree shows that humans and chimpanzees are more closely related to each other than either is to gorillas

gorillas *are sister to* humans & chimps

gorillas *are NOT basal* to chimps & humans

Phylogeny Terminology & Concepts

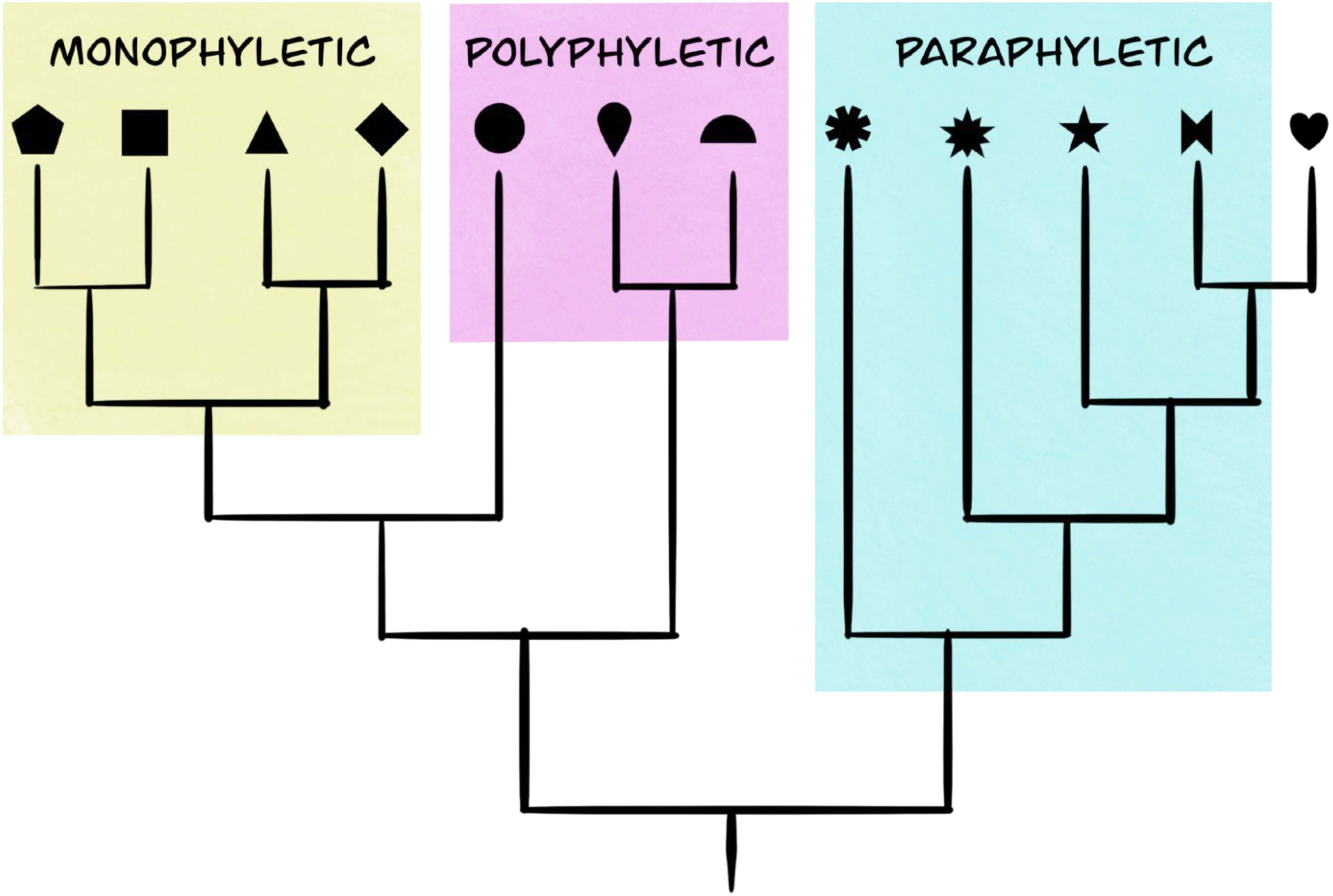


the term basal refers to something that is "closer to the base", so please never use this for extant taxa

the 10 Mya fossil *Chororapithecus abyssinicus* can be called a basal ape (Sewa et al 2007)

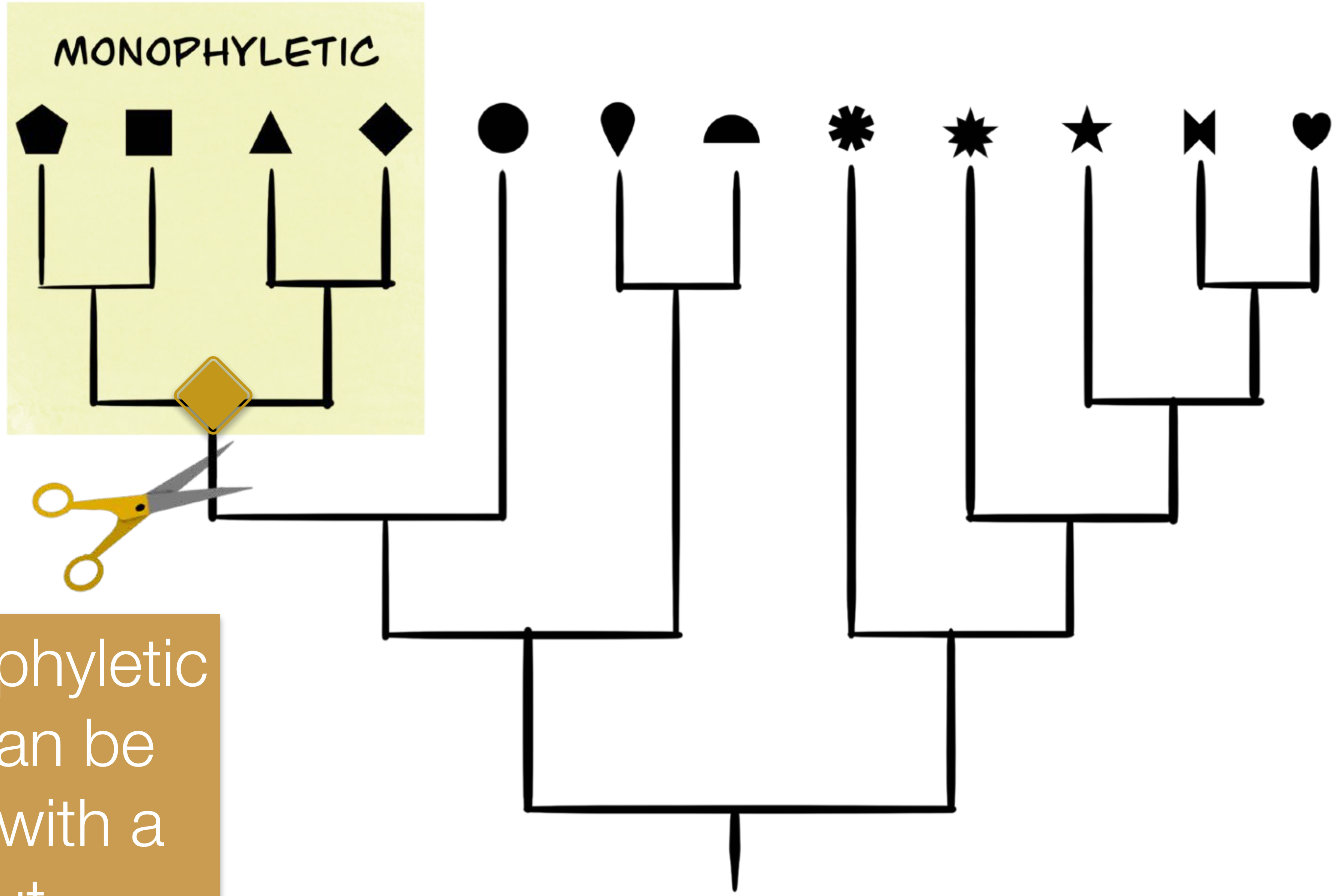
see: http://for-the-love-of-trees.blogspot.com/2016/09/the-ancestors-are-not-among-us.html

Groups in a Phylogeny



Monophyly

a monophyletic group includes an ancestor and all of its descendants

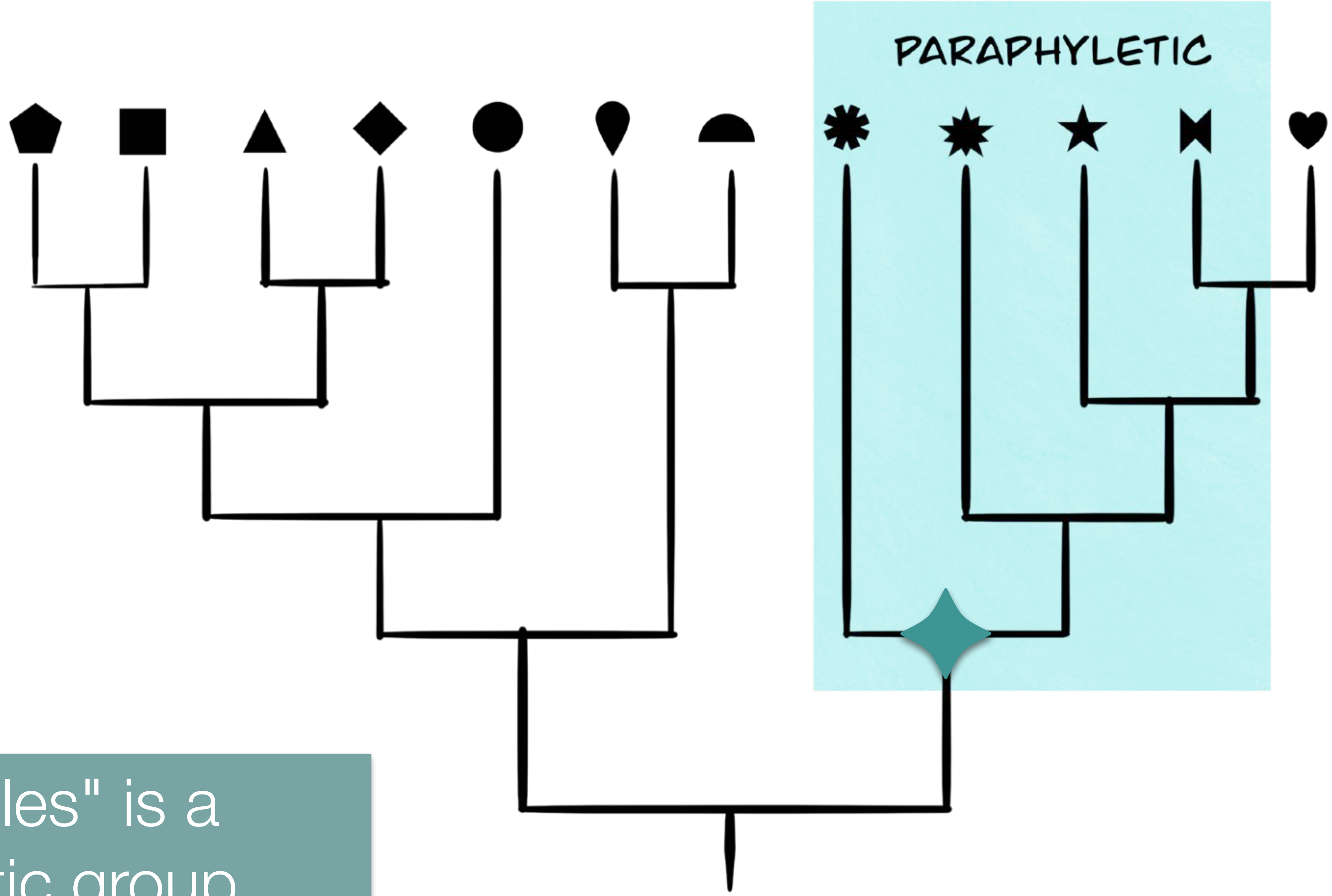


a monophyletic group can be pruned with a single cut

monophyletic group = clade

Paraphyly

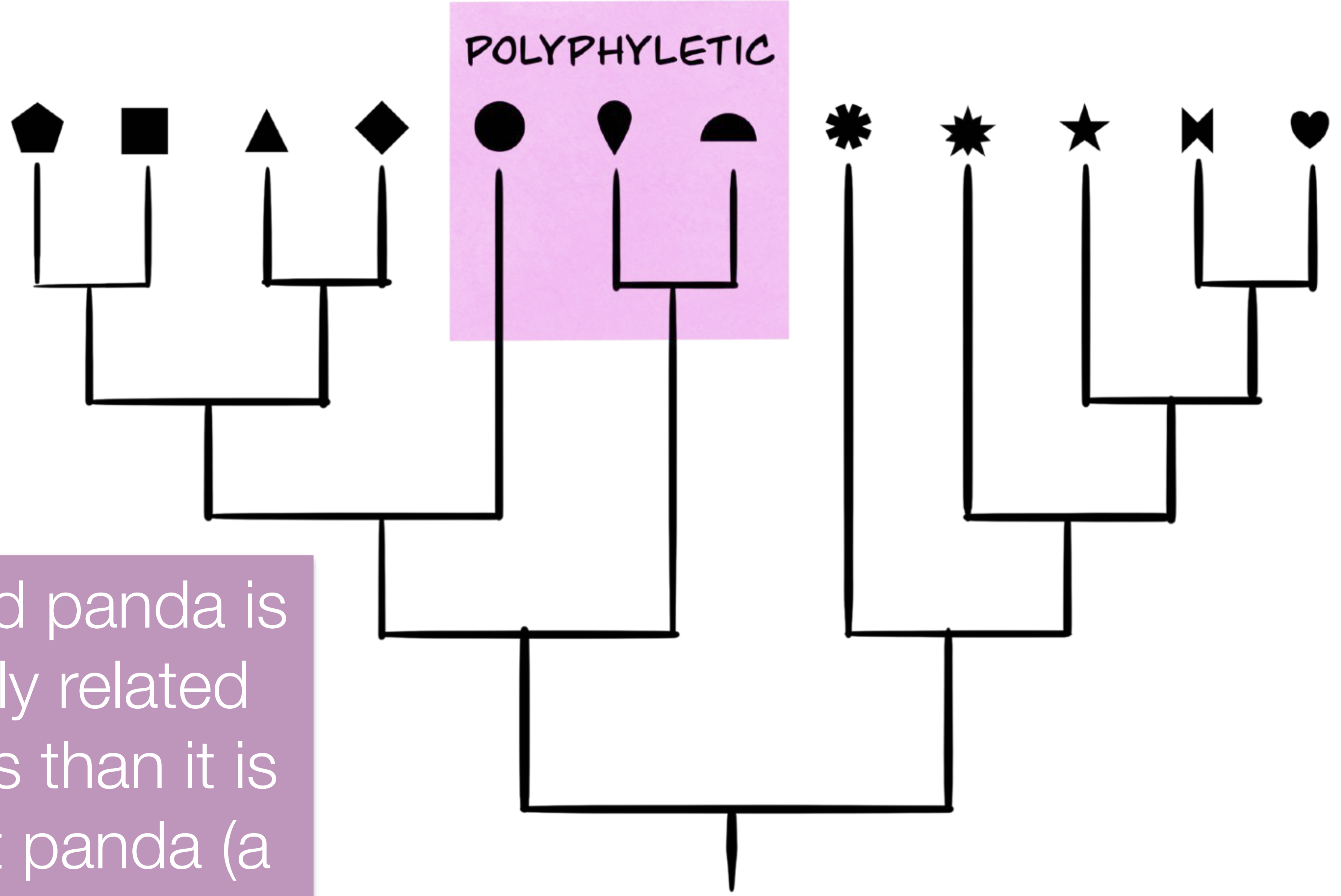
a paraphyletic group includes an ancestor and a subset of its descendants



e.g., "reptiles" is a paraphyletic group unless it includes birds

Polyphyly

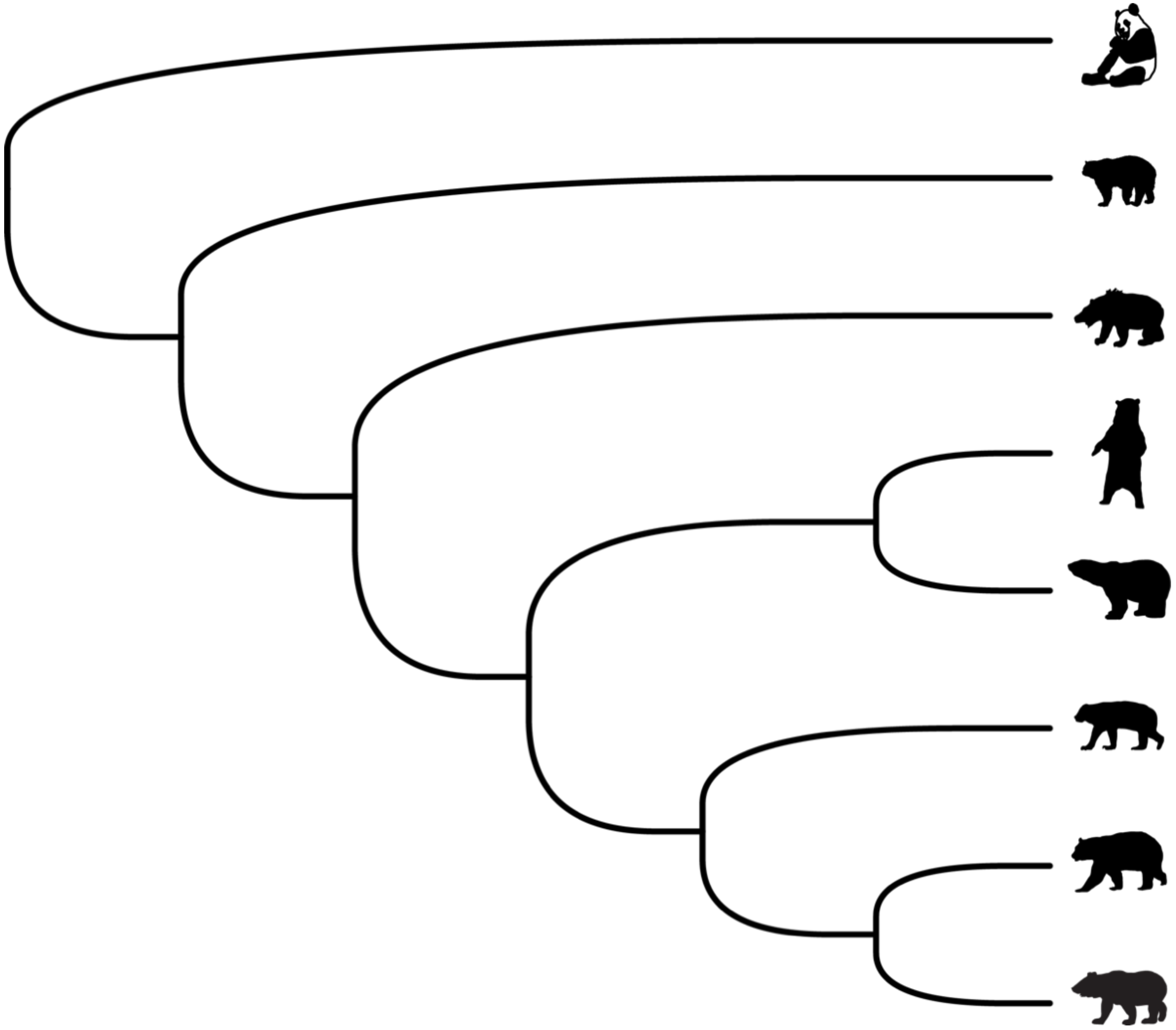
a polyphyletic group includes a set of taxa, but not their common ancestor



e.g., the red panda is more closely related to raccoons than it is to the giant panda (a bear), so "pandas" is a polyphyletic group

Phylogeny: Branch Lengths

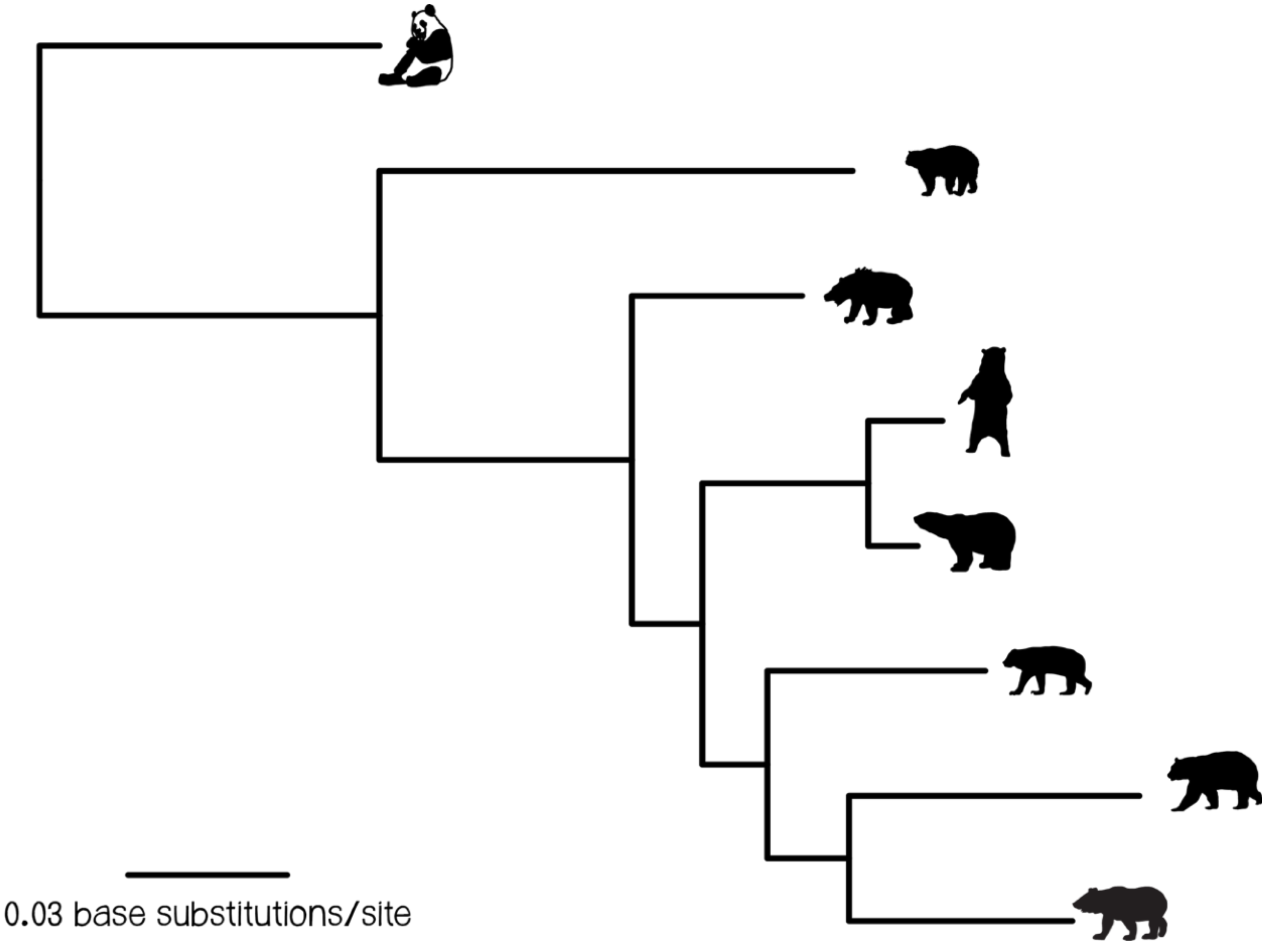
can have no meaning & just show the pattern of relationships



(figure adapted from Heath et al. 2014)

Phylogeny: Branch Lengths

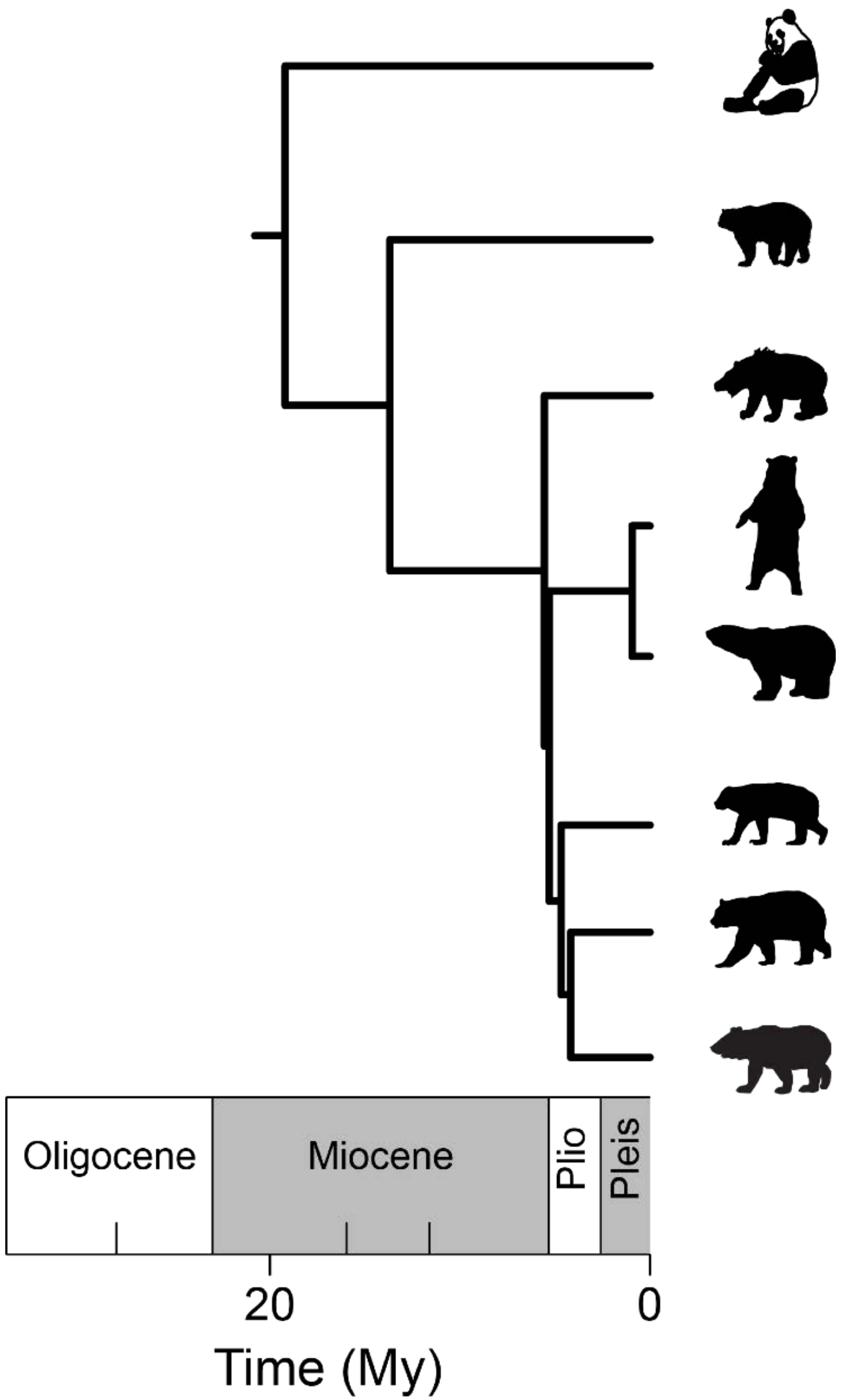
can represent the amount of **genetic difference**



(figure adapted from Heath et al. 2014)

Phylogeny: Branch Lengths

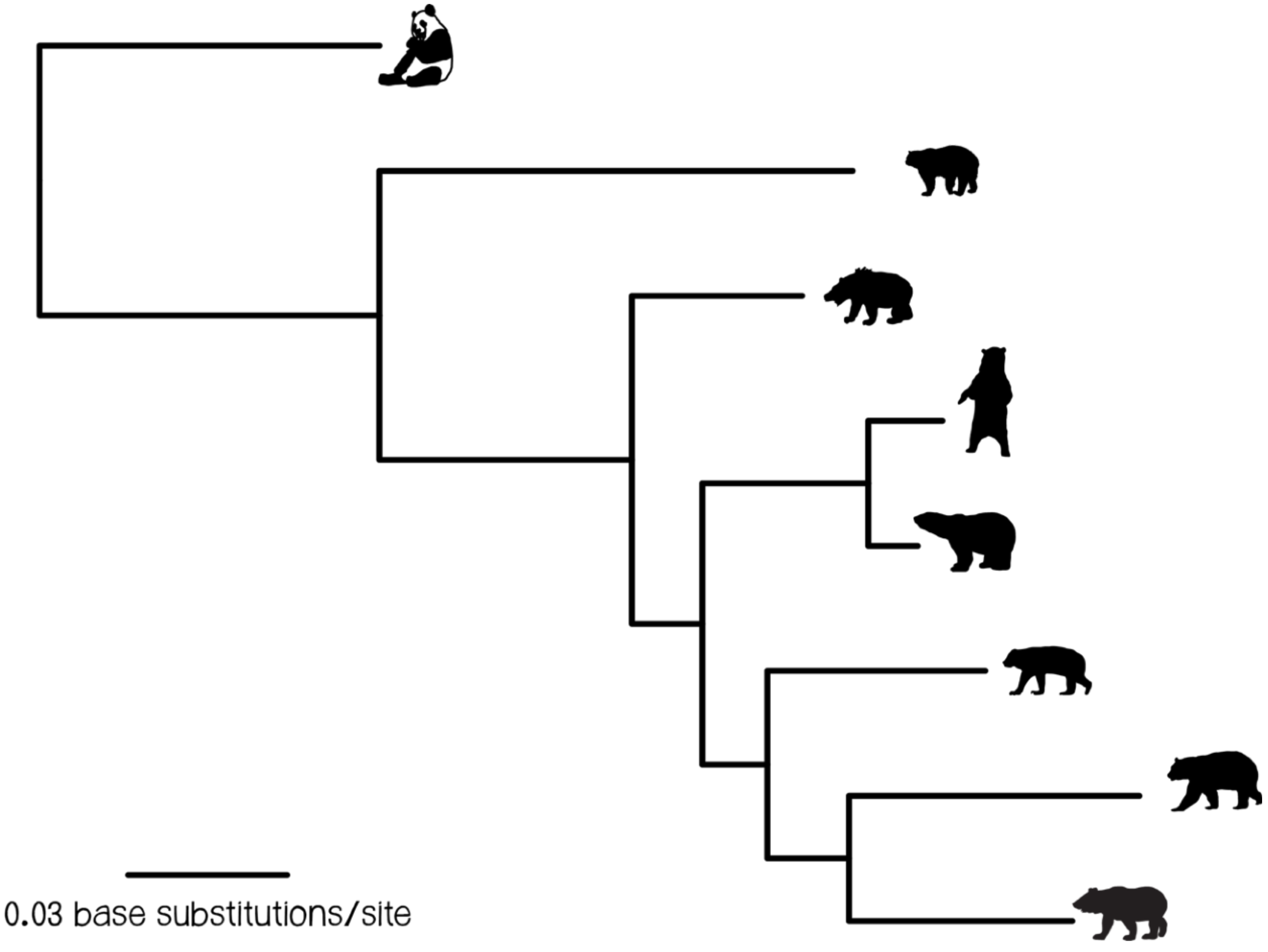
can represent the
duration of time
between nodes



(figure adapted from Heath et al. 2014)

Phylogeny: Rooting

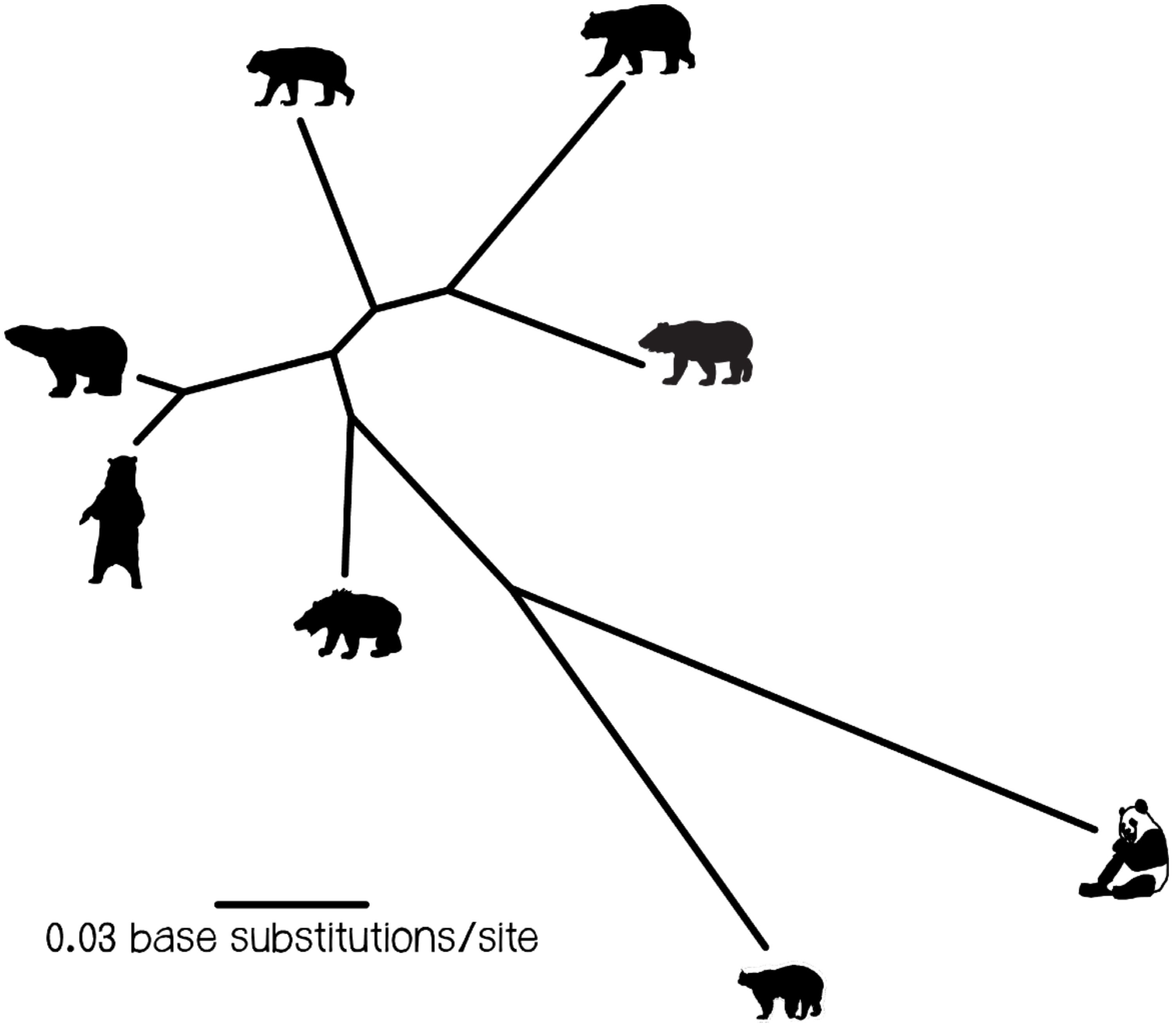
a tree can be **rooted** to show the direction or relative timing of divergence



(figure adapted from Heath et al. 2014)

Phylogeny: Rooting

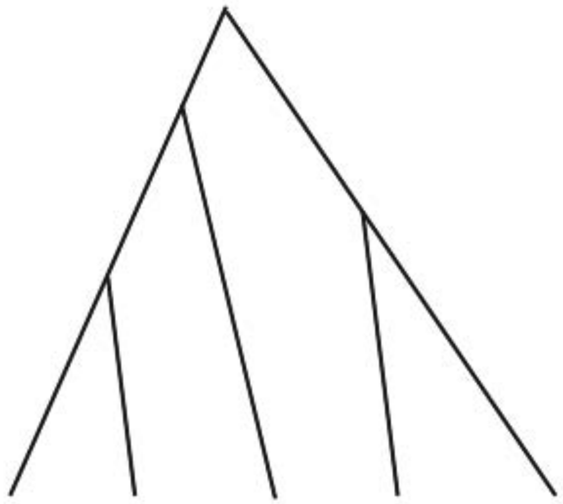
a tree can be **unrooted** showing only relationships among lineages



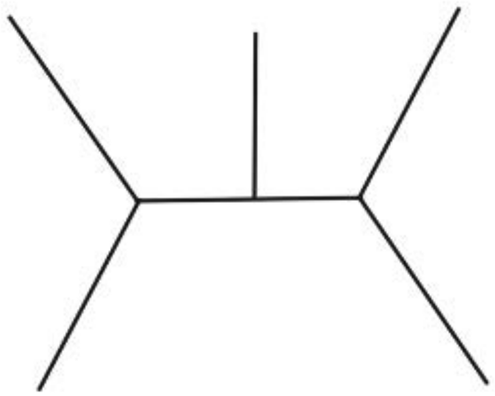
(figure adapted from Heath et al. 2014)

Resolved Phylogeny

a tree can be **bifurcating** or **binary** when all nodes split into only 2 descendants (such a tree is also called "resolved")



fully-resolved tree

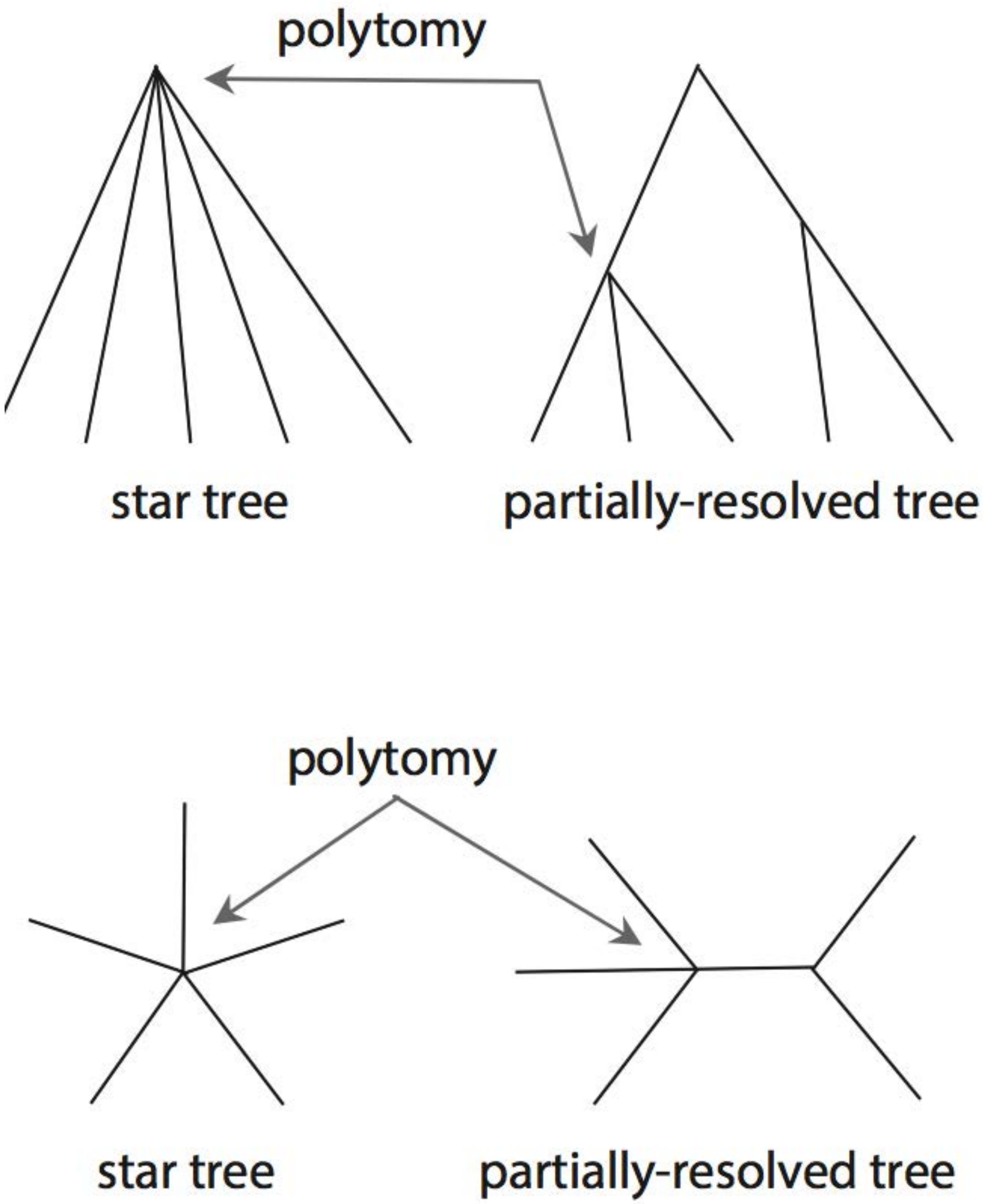


fully-resolved tree

(Yang 2014)

Unresolved Phylogeny

polytomies allow us to represent unresolved nodes

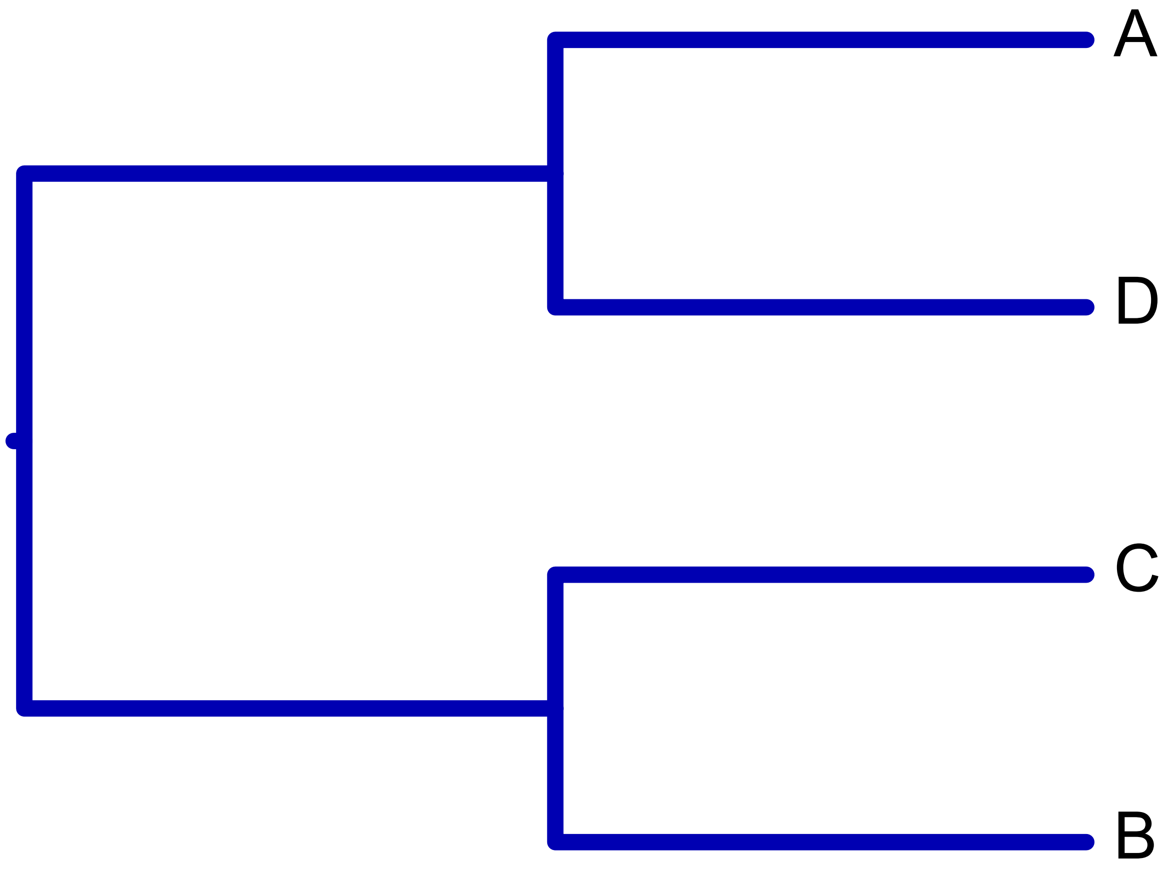


(Yang 2014)

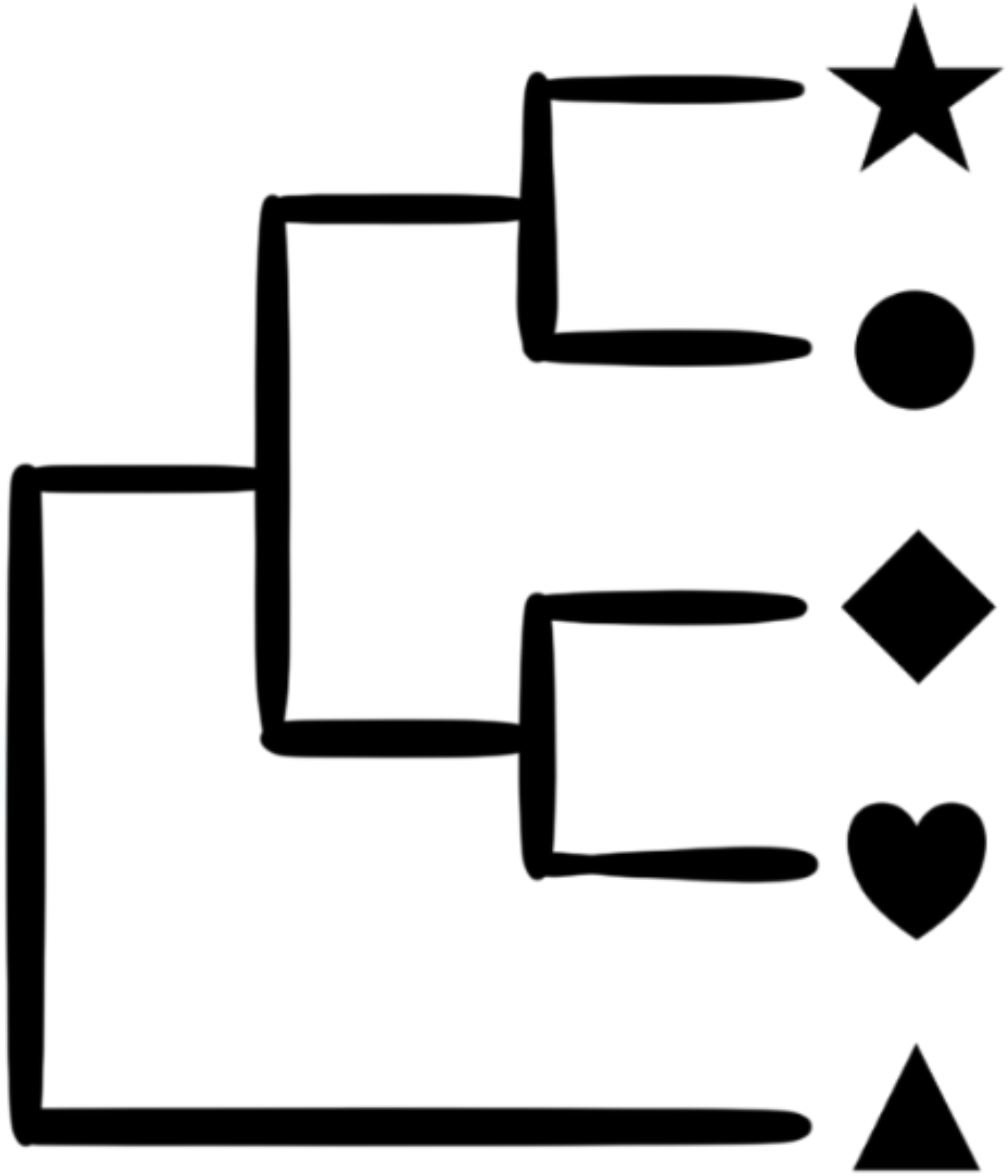
Representing Trees

`((A,D),(C,B));`

we can represent trees using [Newick format](), which uses sets of nested parentheses



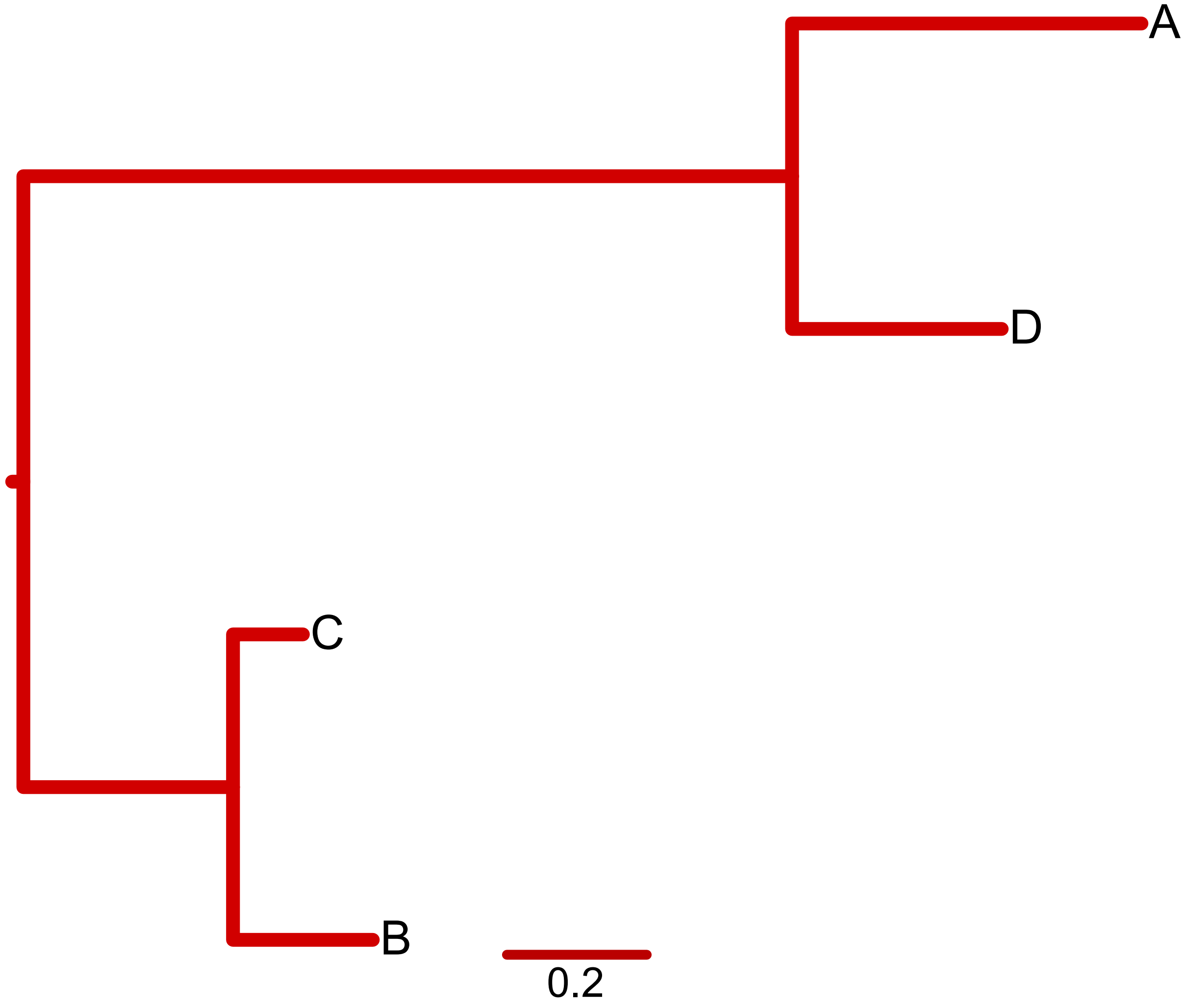
What is the Newick representation for this tree?



Representing Trees

`((A:0.5,D:0.3):1.1,(C:0.1,B:0.2):0.3);`

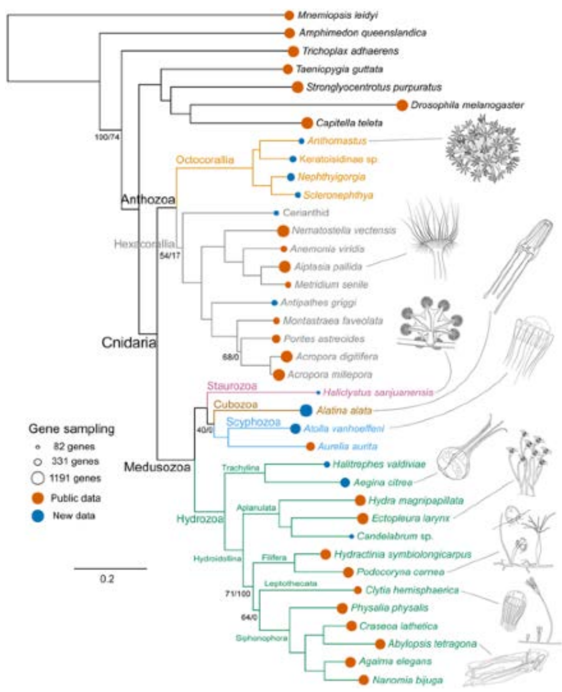
this format also accommodates branch information



Draw the tree

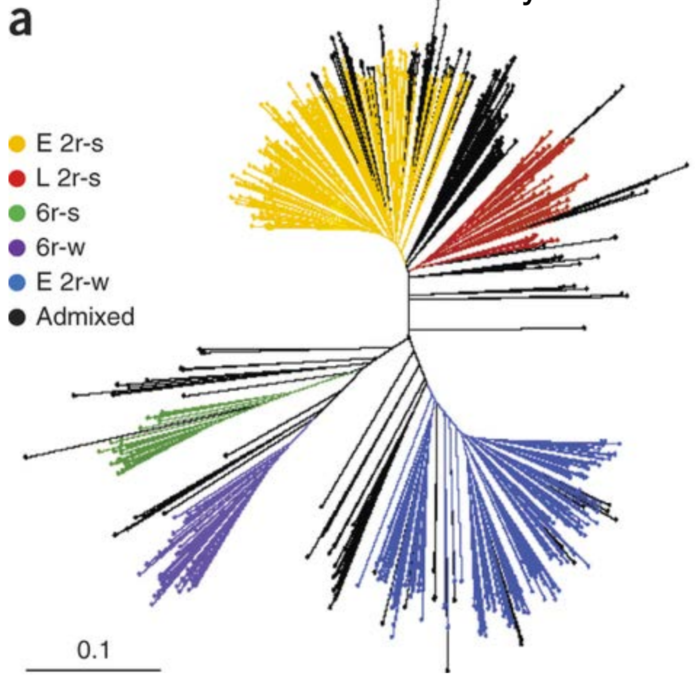
```
((((A,C),D),(B,E)),F);
```

Phylogenies

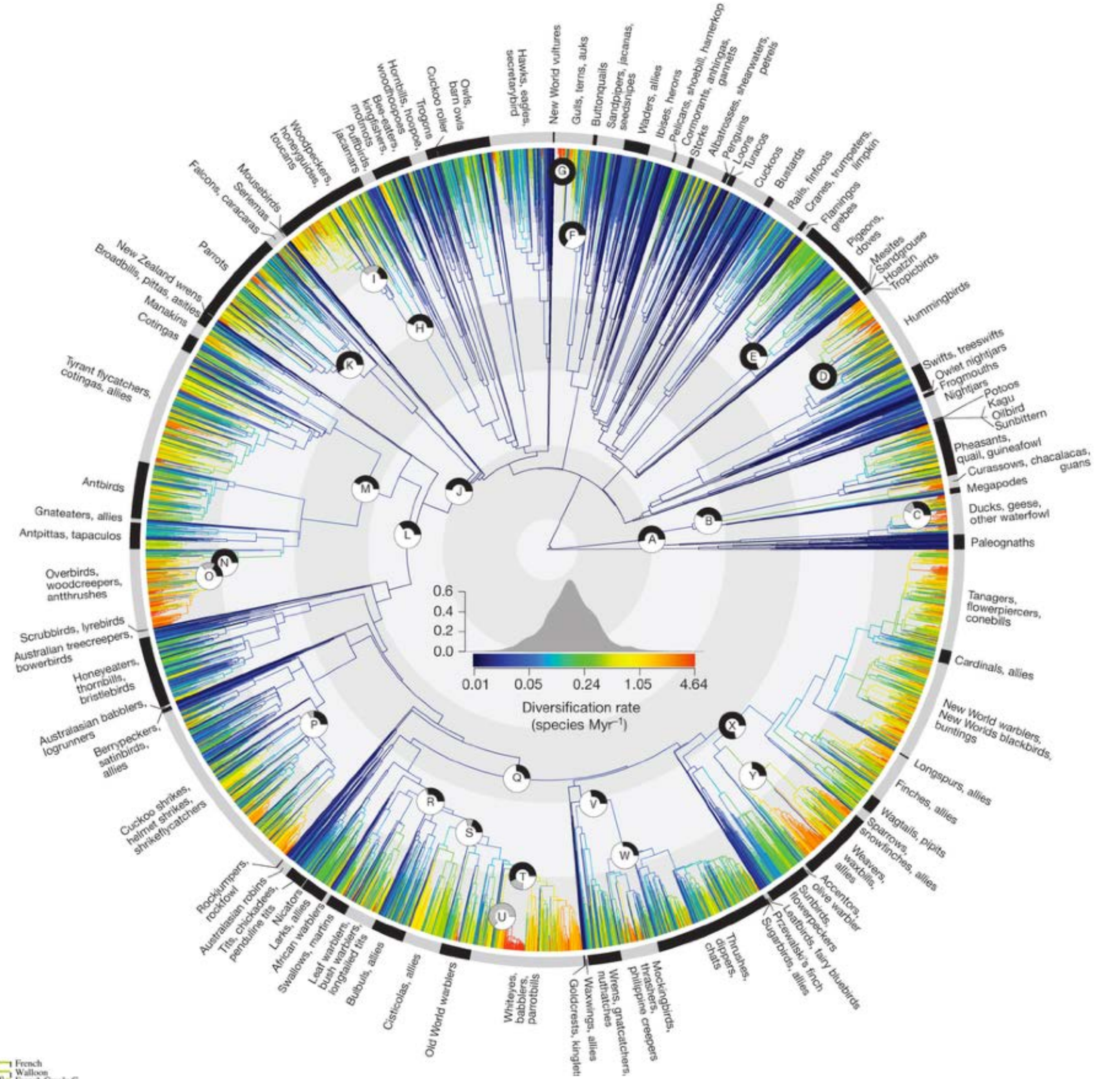


doi.org/10.1371/journal.pone.0139068

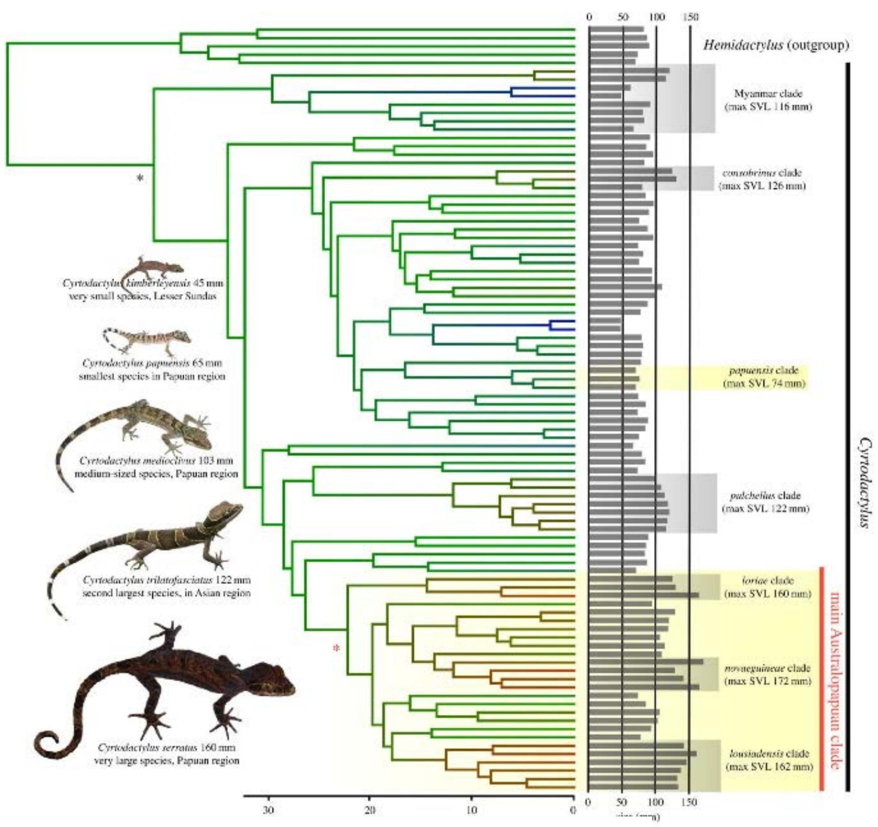
Genetic structure in barley



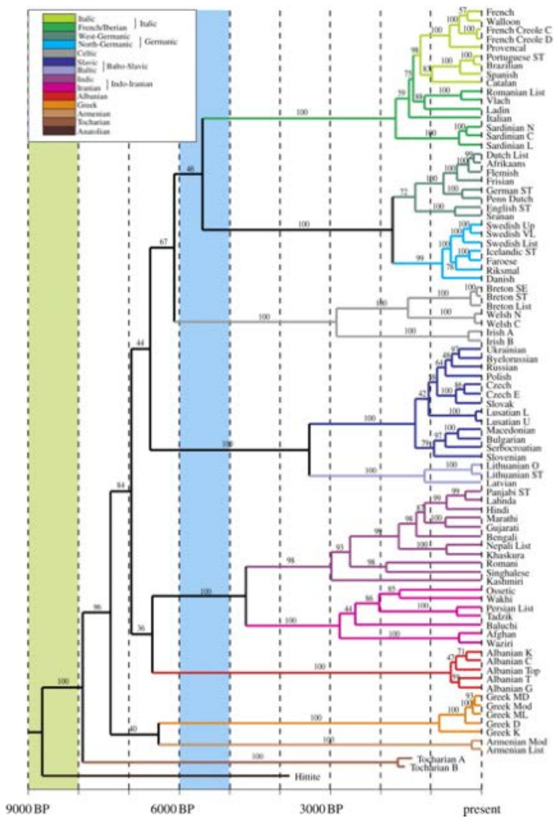
doi.org/10.1038/ng.2447



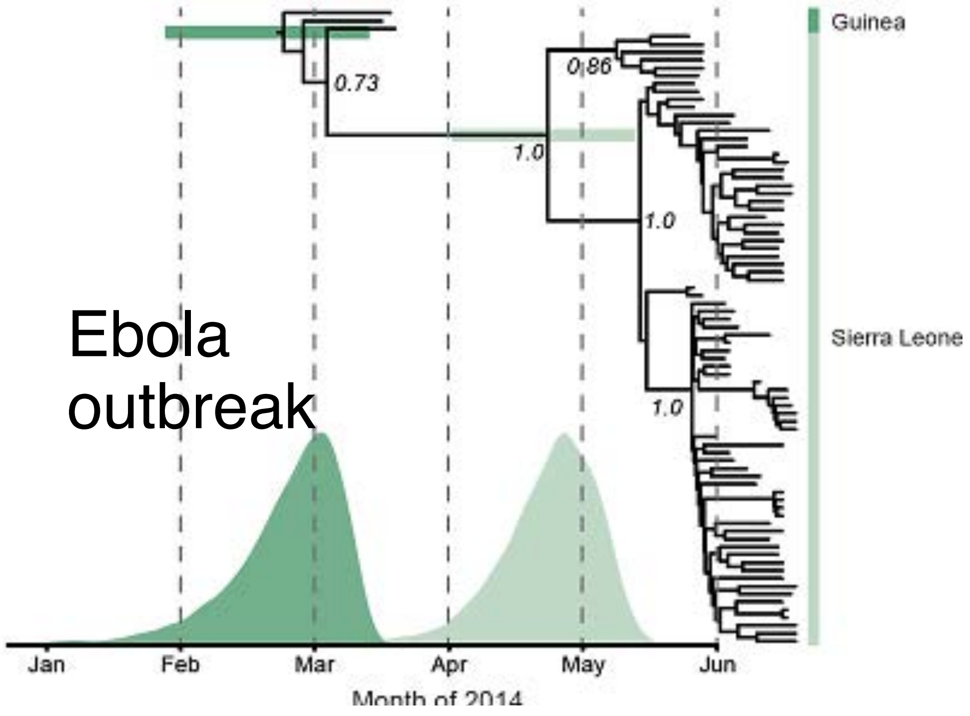
doi.org/10.1038/nature11631



doi.org/10.1098/rsbl.2014.0479

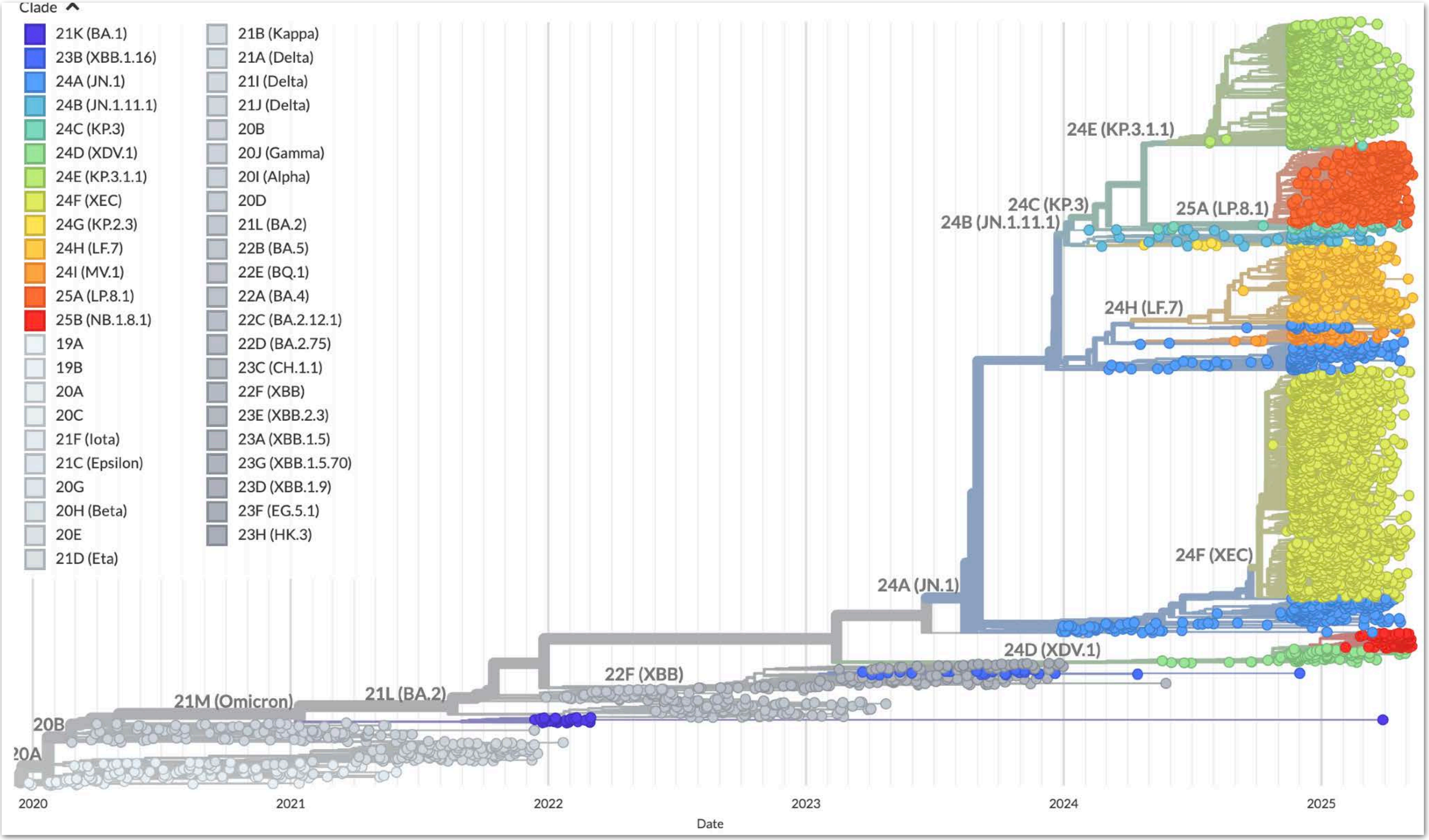


doi.org/10.1098/rstb.2010.0378


Ebola outbreak

doi.org/10.1126/science.1259657

Phylogenetics & Epidemiology



https://nextstrain.org/ncov/gisaid/global

Inferring Phylogenies

to estimate a phylogeny, we need to start with data

the data are observations of character states for a set of taxa

taxa	character 1	character 2	character 3
T1	**pointed**	**blue**	**present**
T2	**pointed**	**blue**	**present**
T3	**round**	**blue**	**absent**
T4	**round**	**black**	**absent**

a column in the matrix is a **character**

the form that character takes is its **state**

Inferring Phylogenies

to estimate a phylogeny, we need to start with data

taxa	character 1	character 2	character 3
T1	A	T	C
T2	A	T	T
T3	G	T	G
T4	G	T	G

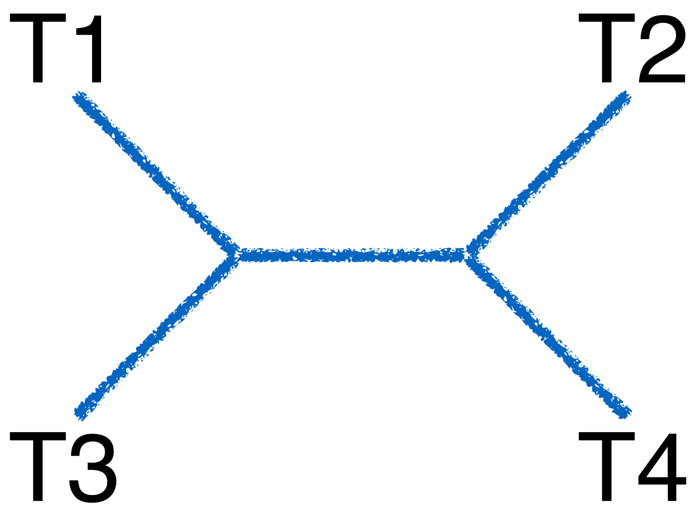
the data are observations of character states for a set of taxa

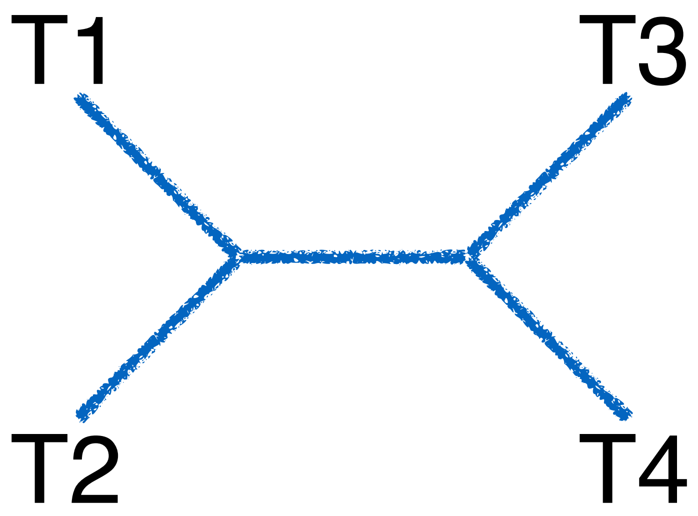
discrete characters can be molecular or morphological

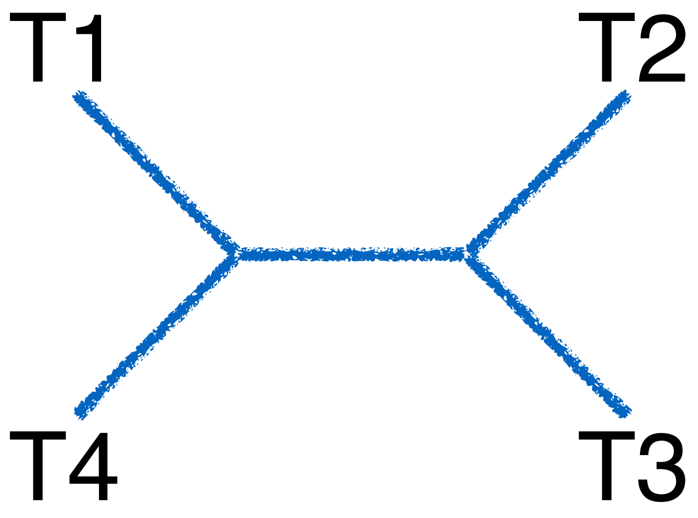
Inferring Phylogenies

we infer phylogenies by evaluating tree topologies

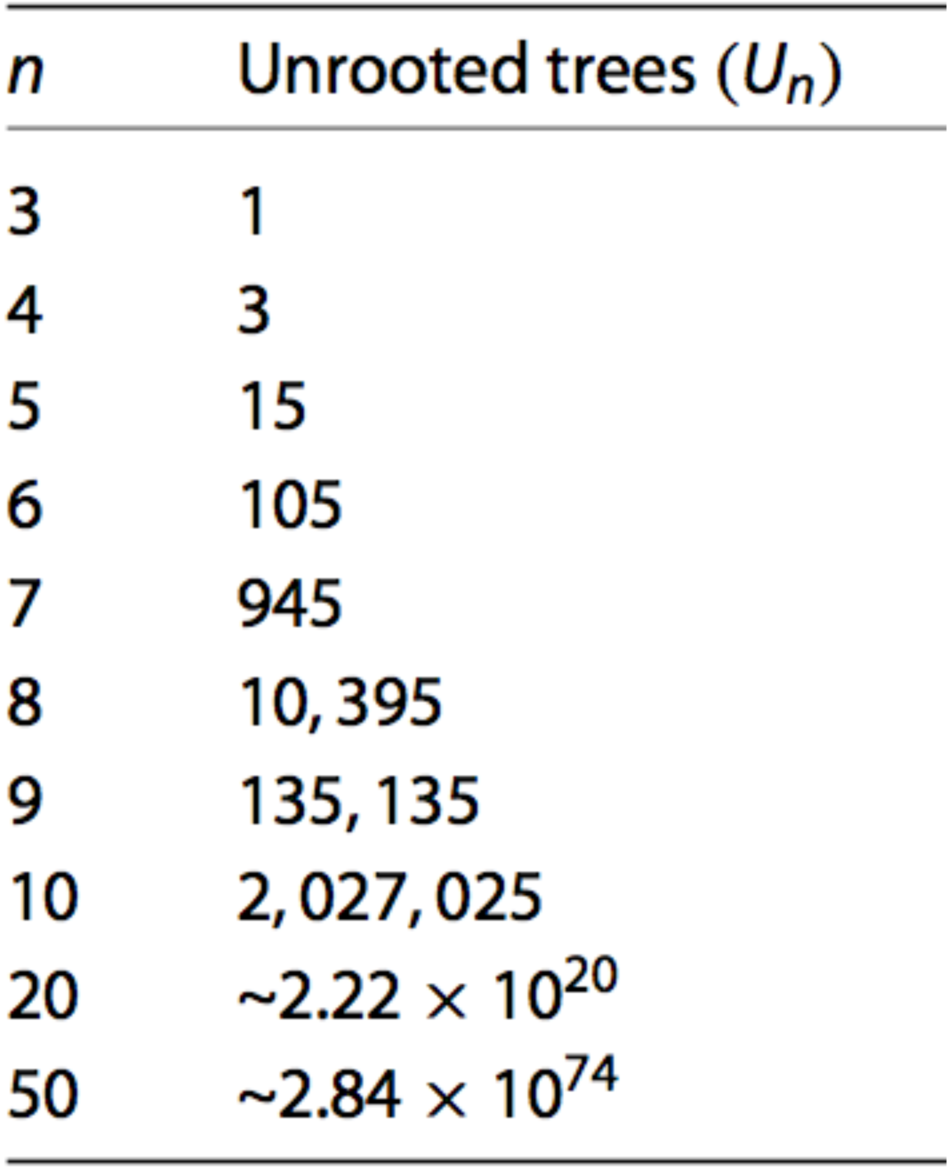
for 4 taxa we can evaluate all possible unrooted topologies (there are only 3)







How many trees?

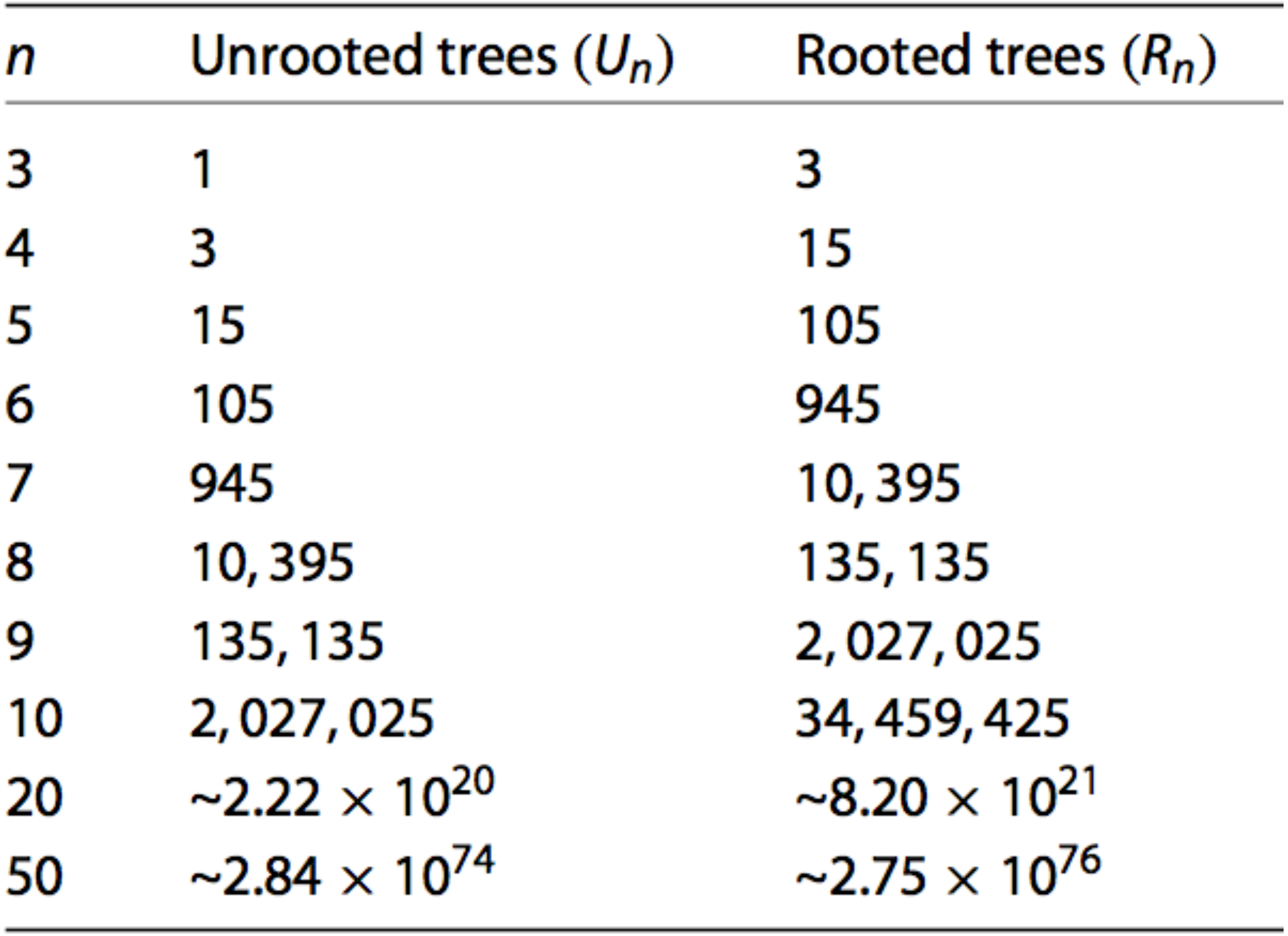


n	Unrooted trees (U_n)
3	1
4	3
5	15
6	105
7	945
8	10,395
9	135,135
10	2,027,025
20	~2.22×10^{20}
50	~2.84×10^{74}

$$N = \frac{(2t-5)!}{2^{t-3}(t-3)!}$$

(Yang 2014)

How many trees?



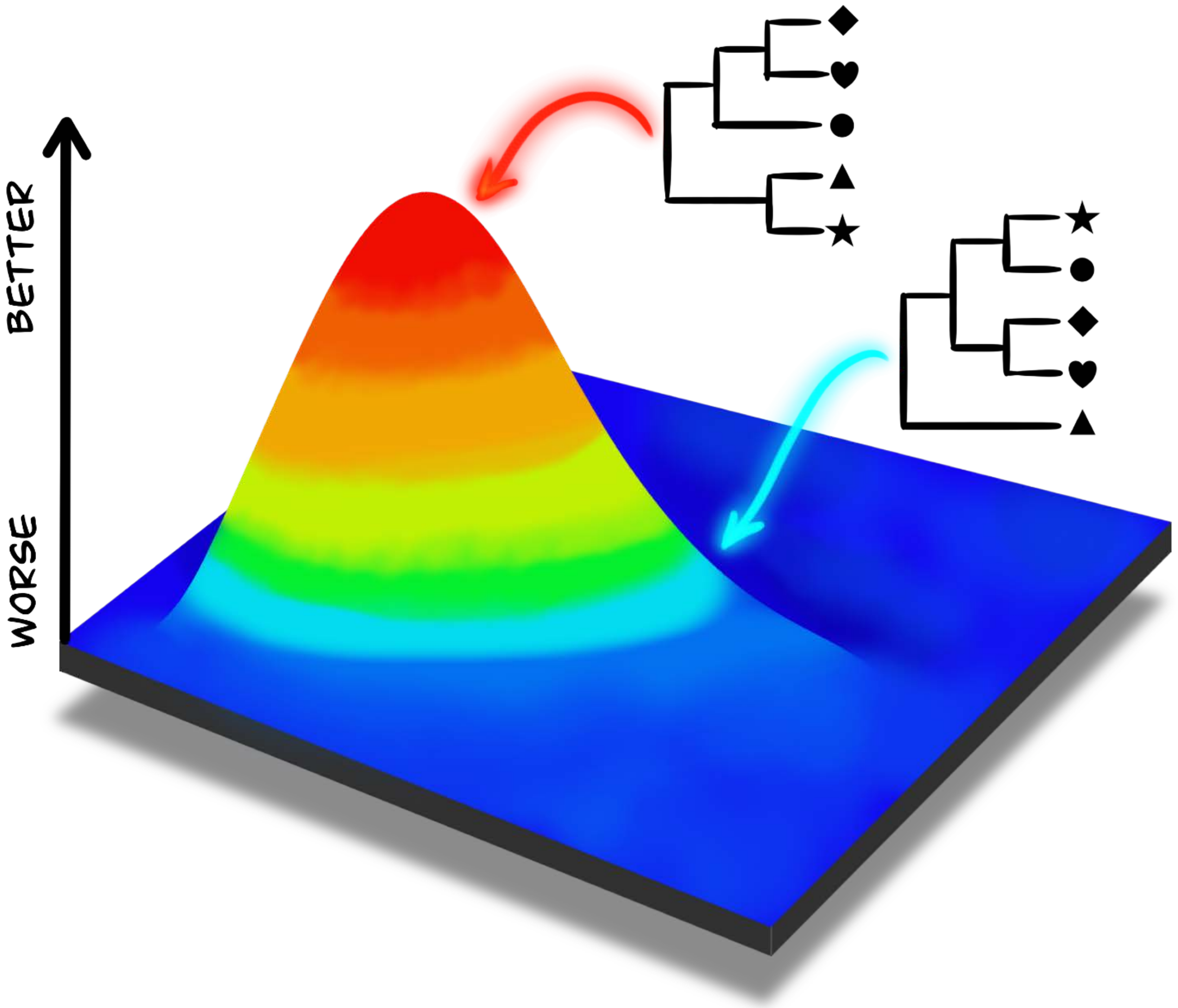
n	Unrooted trees (U_n)	Rooted trees (R_n)
3	1	3
4	3	15
5	15	105
6	105	945
7	945	10,395
8	10,395	135,135
9	135,135	2,027,025
10	2,027,025	34,459,425
20	$\sim 2.22 \times 10^{20}$	$\sim 8.20 \times 10^{21}$
50	$\sim 2.84 \times 10^{74}$	$\sim 2.75 \times 10^{76}$

$$N = \frac{(2t-3)!}{2^{t-2}(t-2)!}$$

at 51 taxa, the number of trees exceeds the number of particles in the observable universe

(Yang 2014)

How to find the "best" tree?



It depends on how you measure "best"

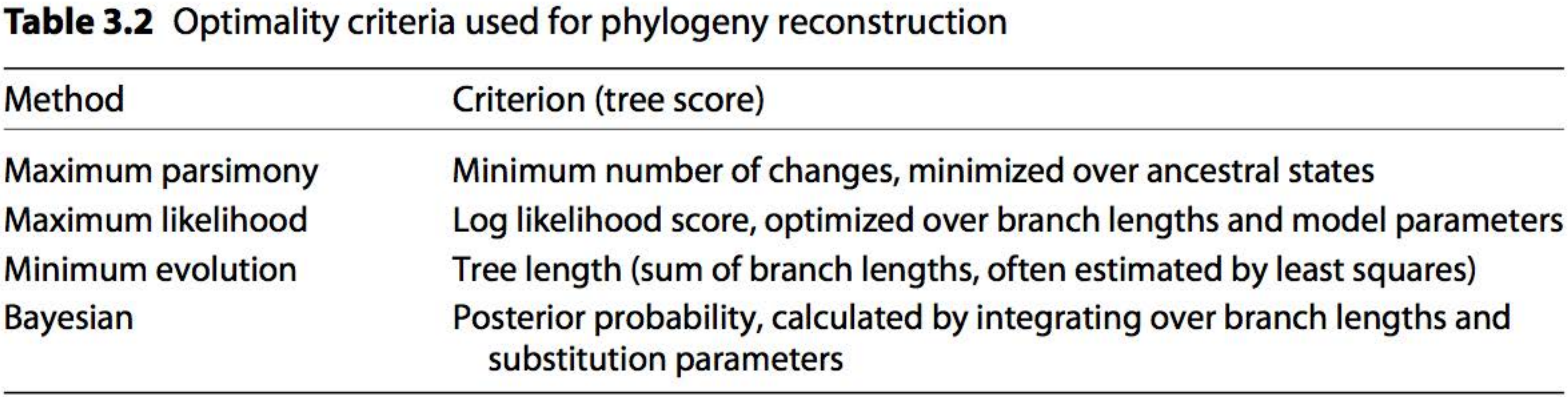
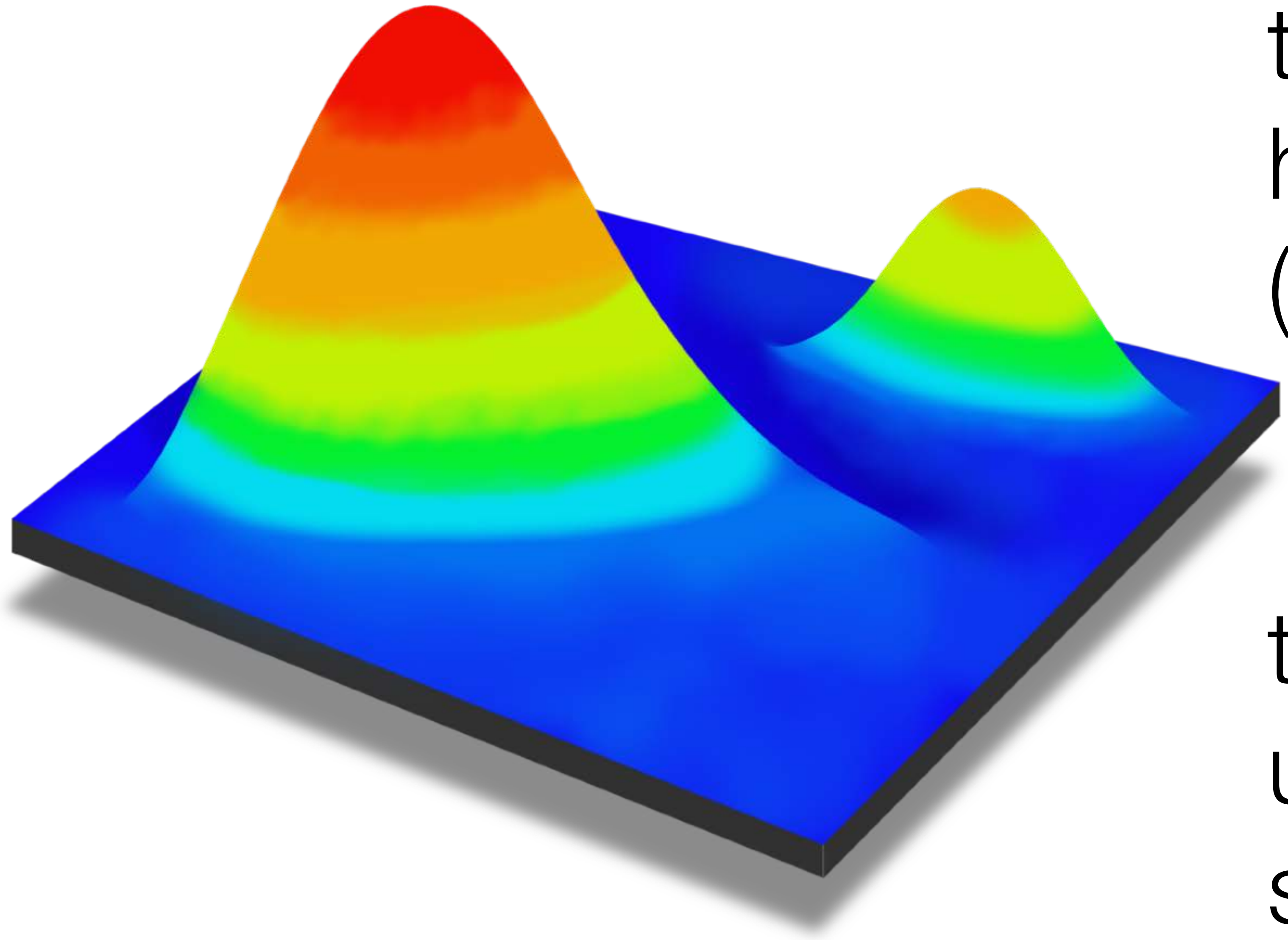


Table 3.2 Optimality criteria used for phylogeny reconstruction

Method	Criterion (tree score)
Maximum parsimony	Minimum number of changes, minimized over ancestral states
Maximum likelihood	Log likelihood score, optimized over branch lengths and model parameters
Minimum evolution	Tree length (sum of branch lengths, often estimated by least squares)
Bayesian	Posterior probability, calculated by integrating over branch lengths and substitution parameters

(Yang 2014)

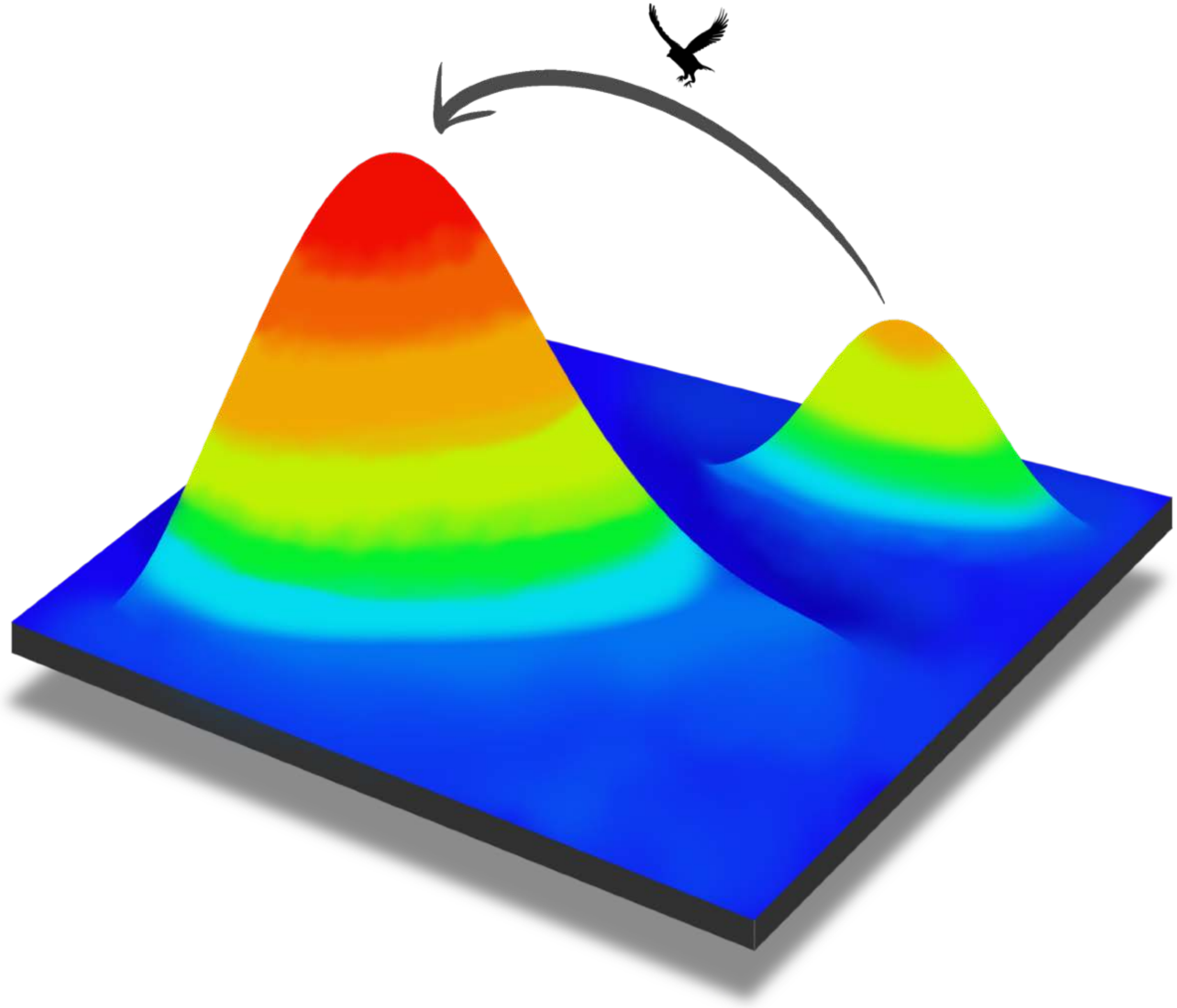
Searching Tree Topologies



a number of heuristic tree-search algorithms have been developed (Yang 2014, Ch. 3)

these methods allow us to evaluate just a subset of the possible trees

Searching Tree Topologies



importantly, we need tree-search methods that can find the global optimum

these approaches are useful for maximum parsimony, maximum likelihood, and Bayesian methods