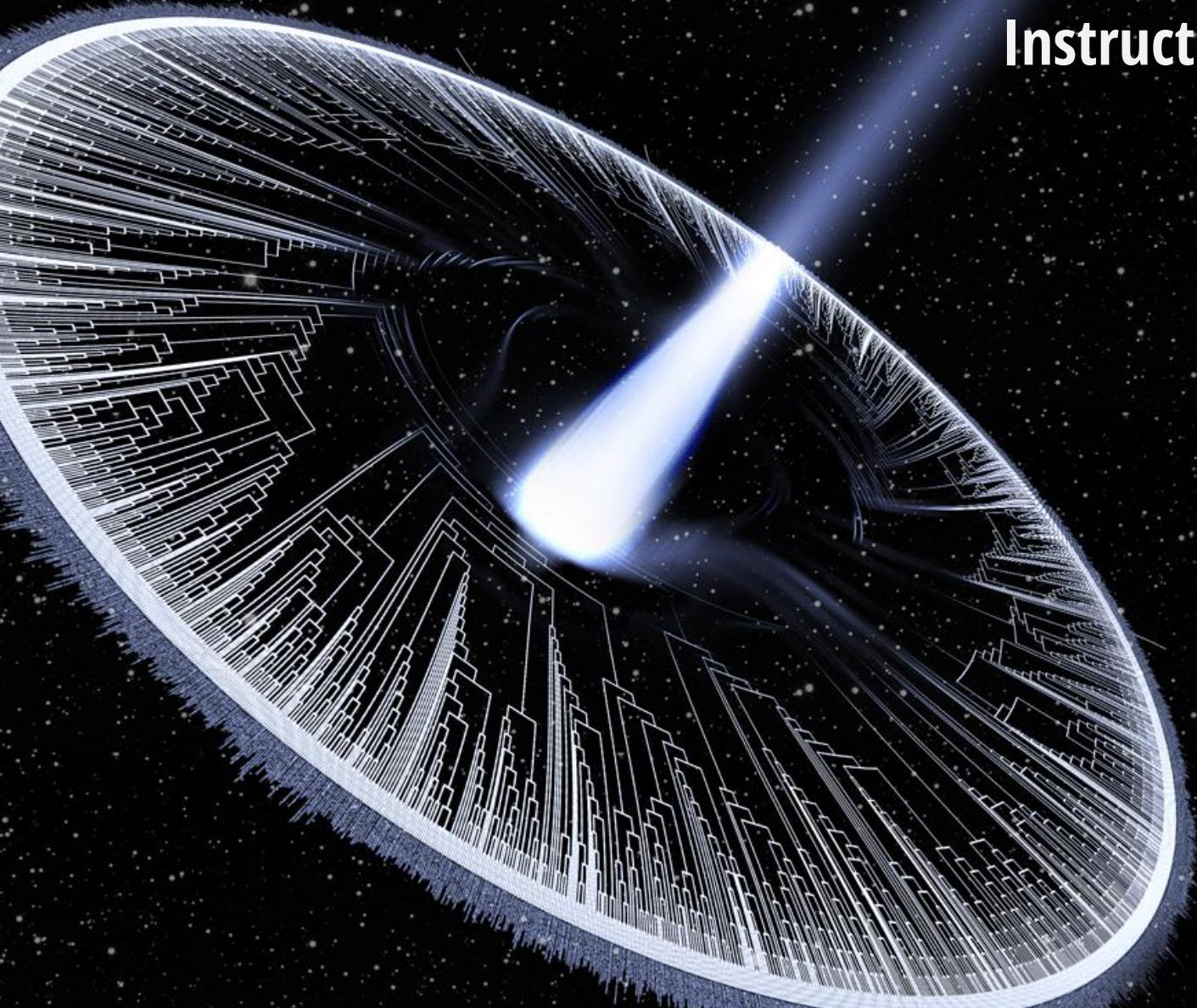


Macroevolution

Instructor: Josef Uyeda
juyeda@vt.edu
uyedalab.com



Goals of the course:

- 1. Learn the core of phylogenetic models**
- 2. Learn how to use trees**
- 3. Learn about the big questions in macroevolution**
- 4. What are your goals?**

Syllabus - see Canvas site

***Subject to change**

Introductions

- 1. Your name, department, year, program**
- 2. A hobby or interest outside your job**
- 3. What topic you most want to learn about in this class.**

Why phylogeny?

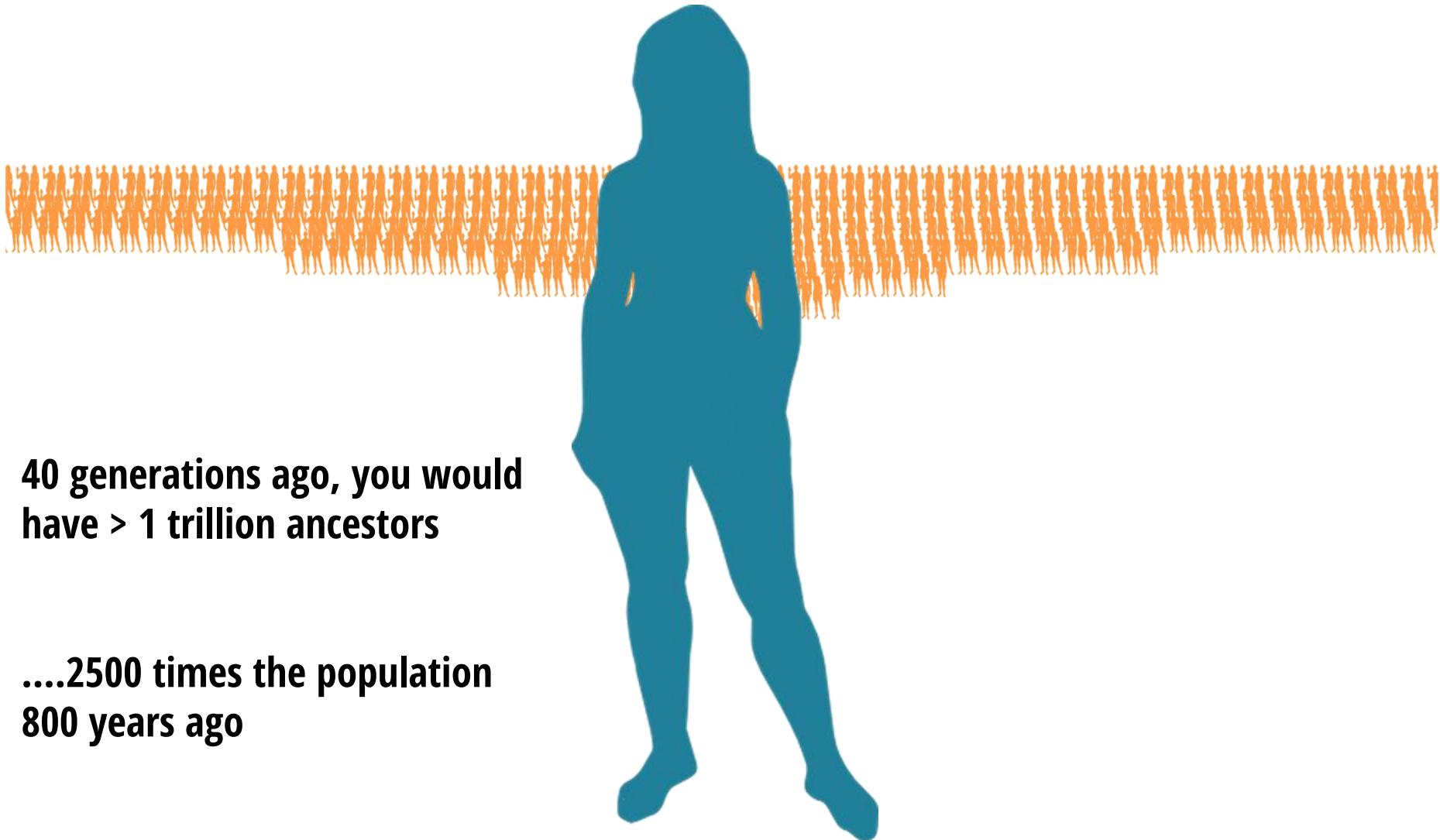
What does a phylogeny represent?

www.OneZoom.org/life/

www.pollev.com/josefuyeda941

How many ancestors do you have?

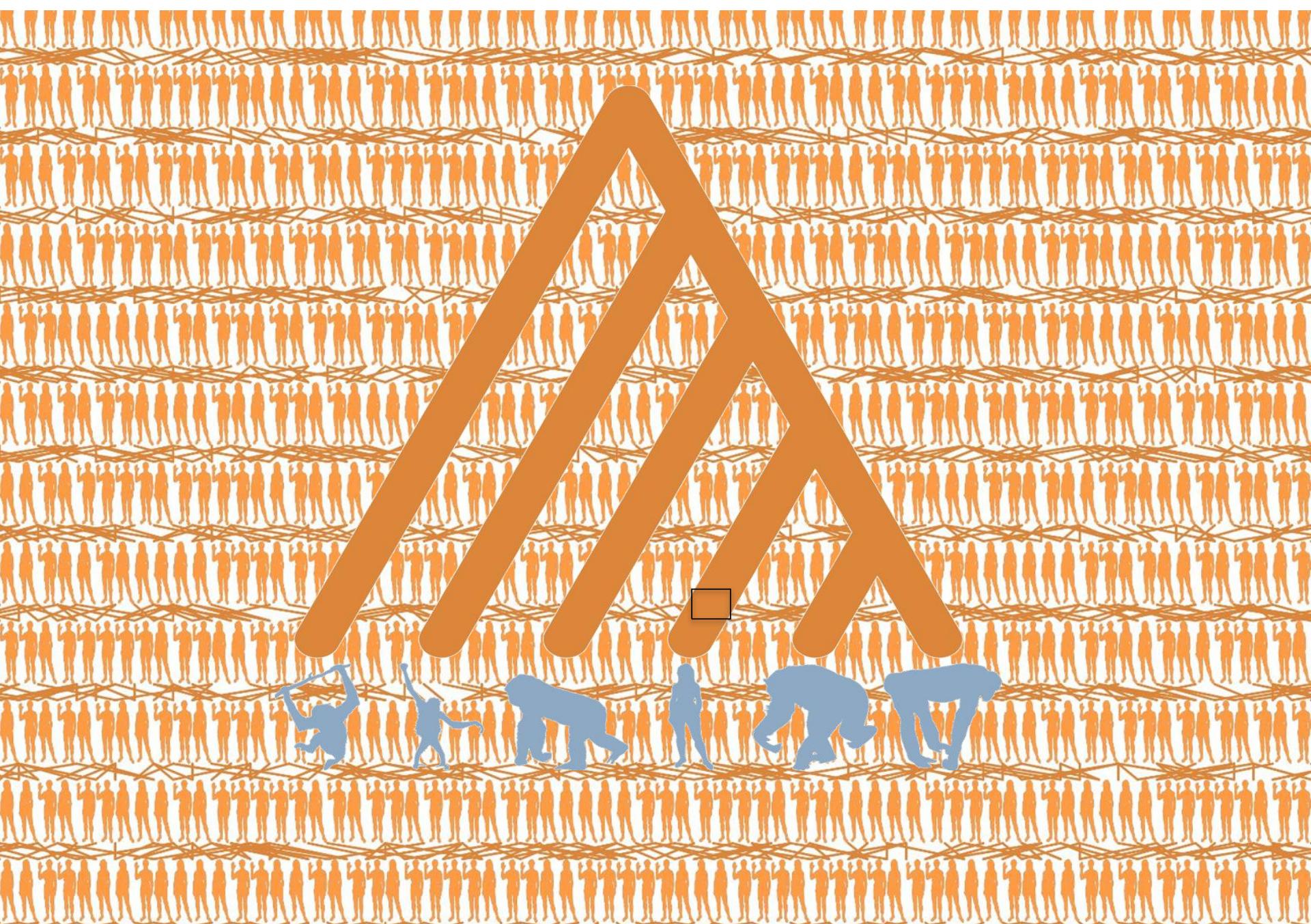
- A) Consider your personal family tree (i.e. a pedigree). How many parents do you have? How many grandparents? Great grandparents? What's the formula?
- B) Use the formula in the question A above we can get our number of ancestors after say, 25 generations (25 generations for humans is ~ 400-900 years).
- C) What are phylogenies? Are they a) hypotheses b) real facts that exist in nature or c) a metaphor?

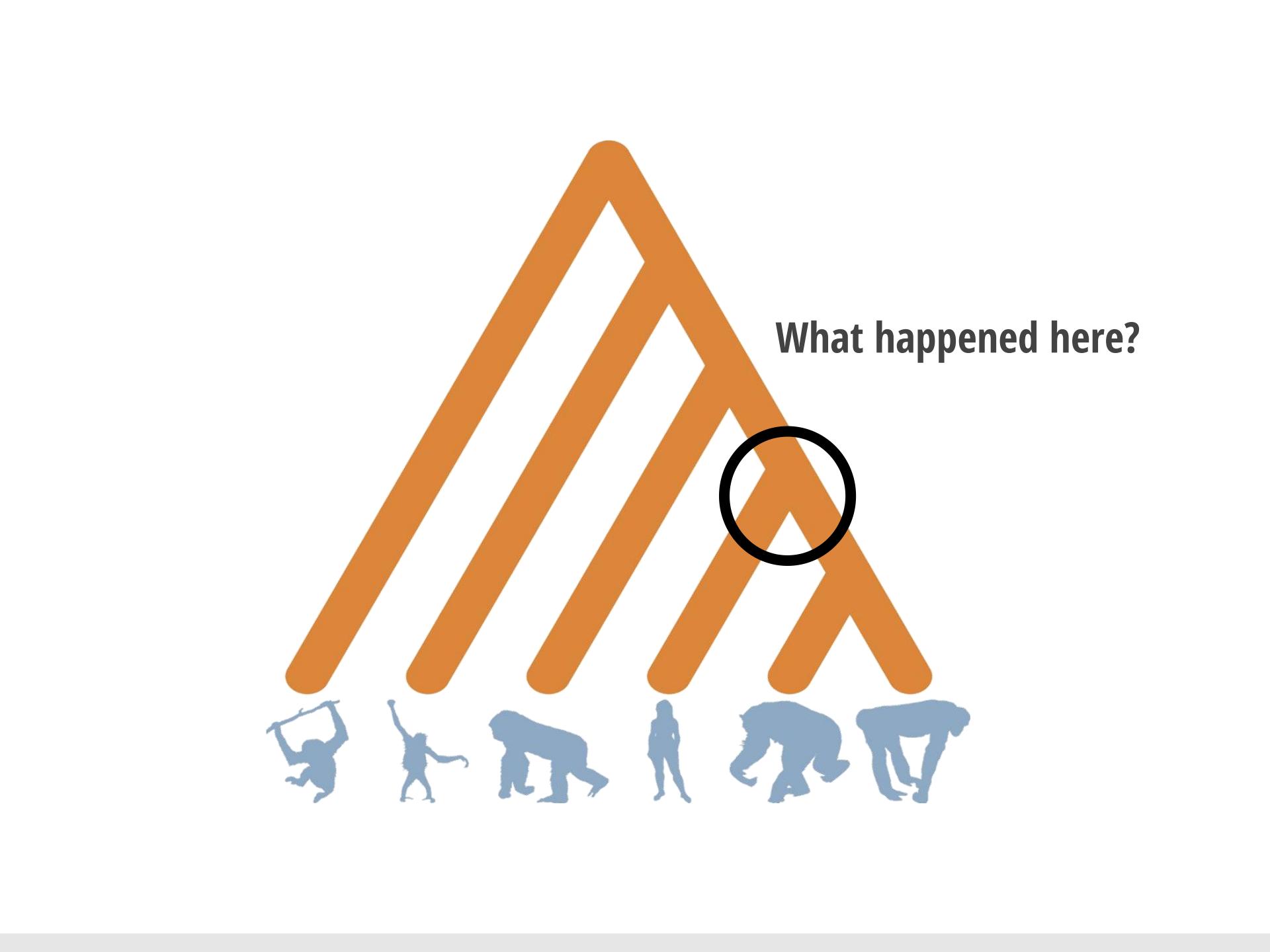


**40 generations ago, you would
have > 1 trillion ancestors**

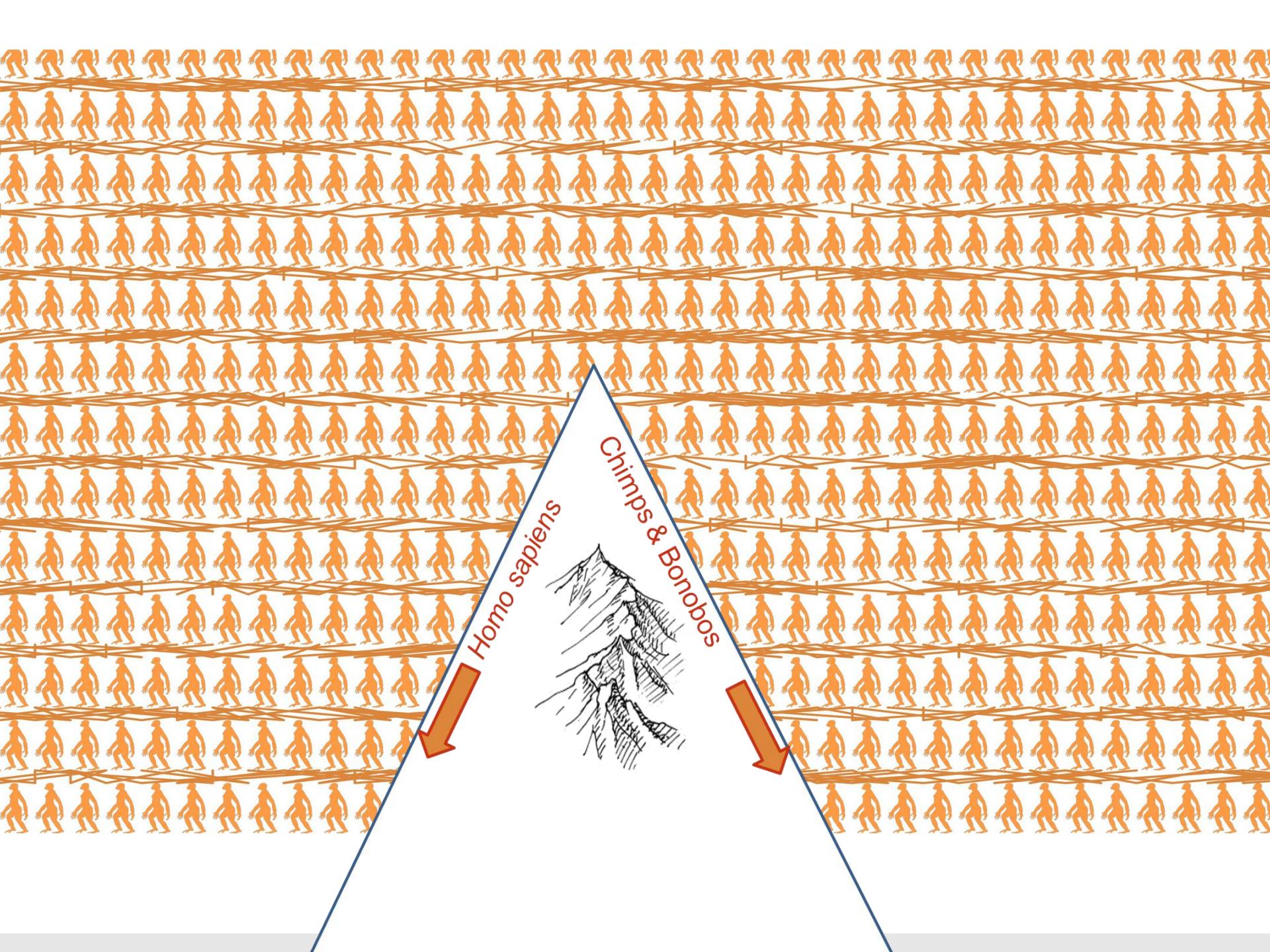
**....2500 times the population
800 years ago**





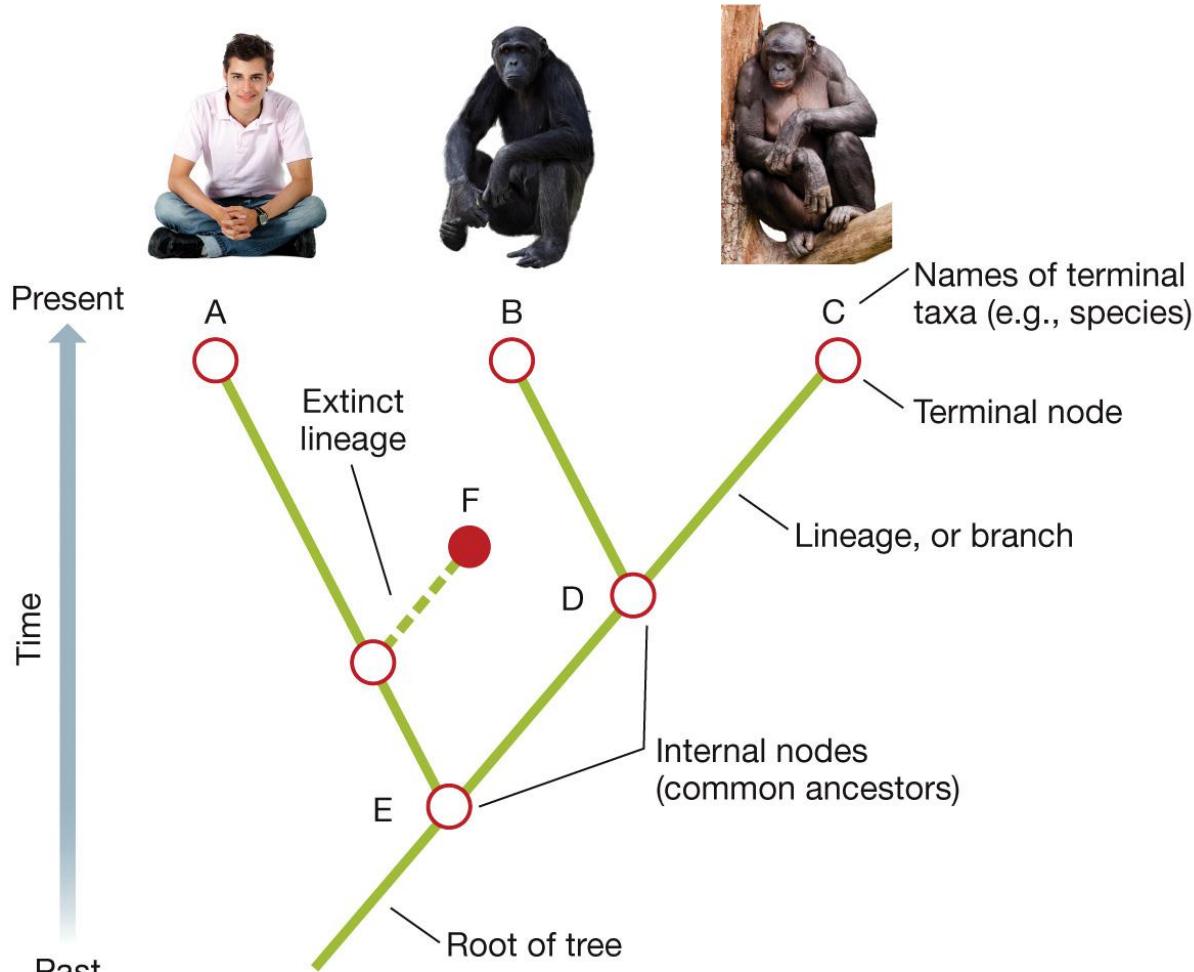


What happened here?



Terminology

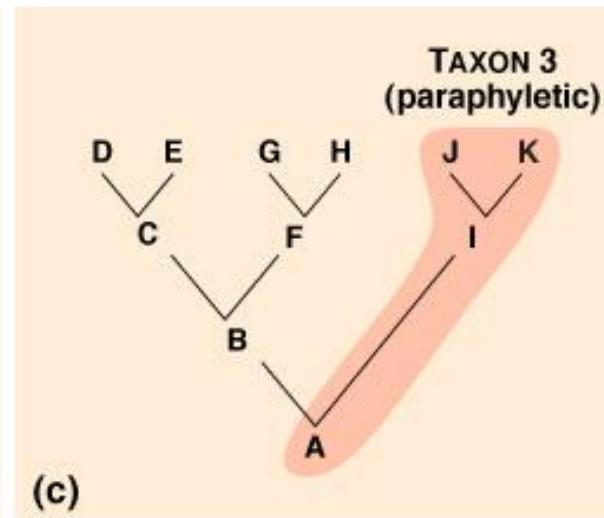
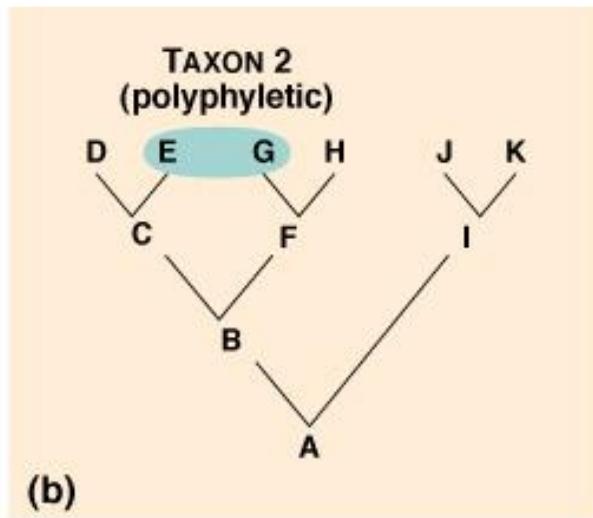
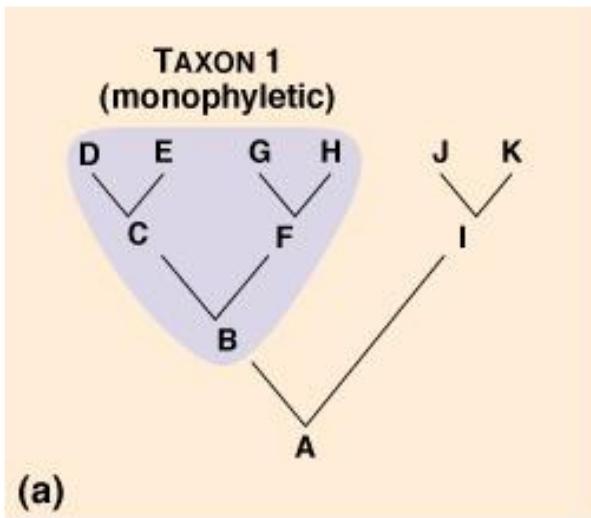
Root
Branch
Node
Tip
Outgroup
Clade
Monophyletic
Paraphyletic
Polyphyletic



EVOLUTION 4e, Figure 2.6
© 2017 Sinauer Associates, Inc.

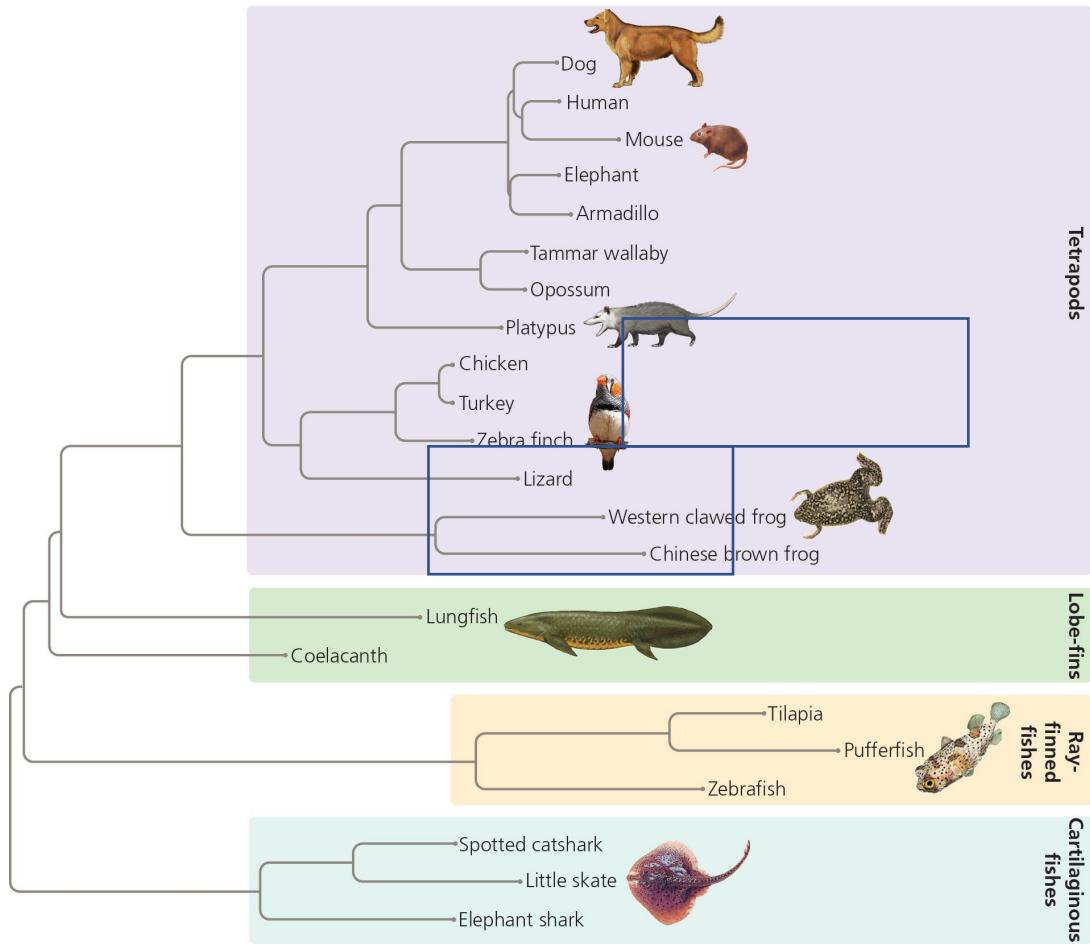
Taxonomy & Systematics

Goal: Monophyly



The phylogeny depicted unambiguously demonstrates that Lungfish are basal to /more primitive than frogs

- A. True
B. False



Assignment I: Canvas self-assessment

Read Baum & Smith, and complete the questions in Assignment I. If you are registered in the class, you can take a canvas quiz (retake up to 10 times until you get 100%) to assess your understanding.

I'm going to assume you understand basic terminology and can read & interpret trees going forward.

Macroevolution, systematics, phylogenetics, taxonomy, biodiversity science, oh my

What is macroevolution?

**We seek to understand the processes that explain
biodiversity**

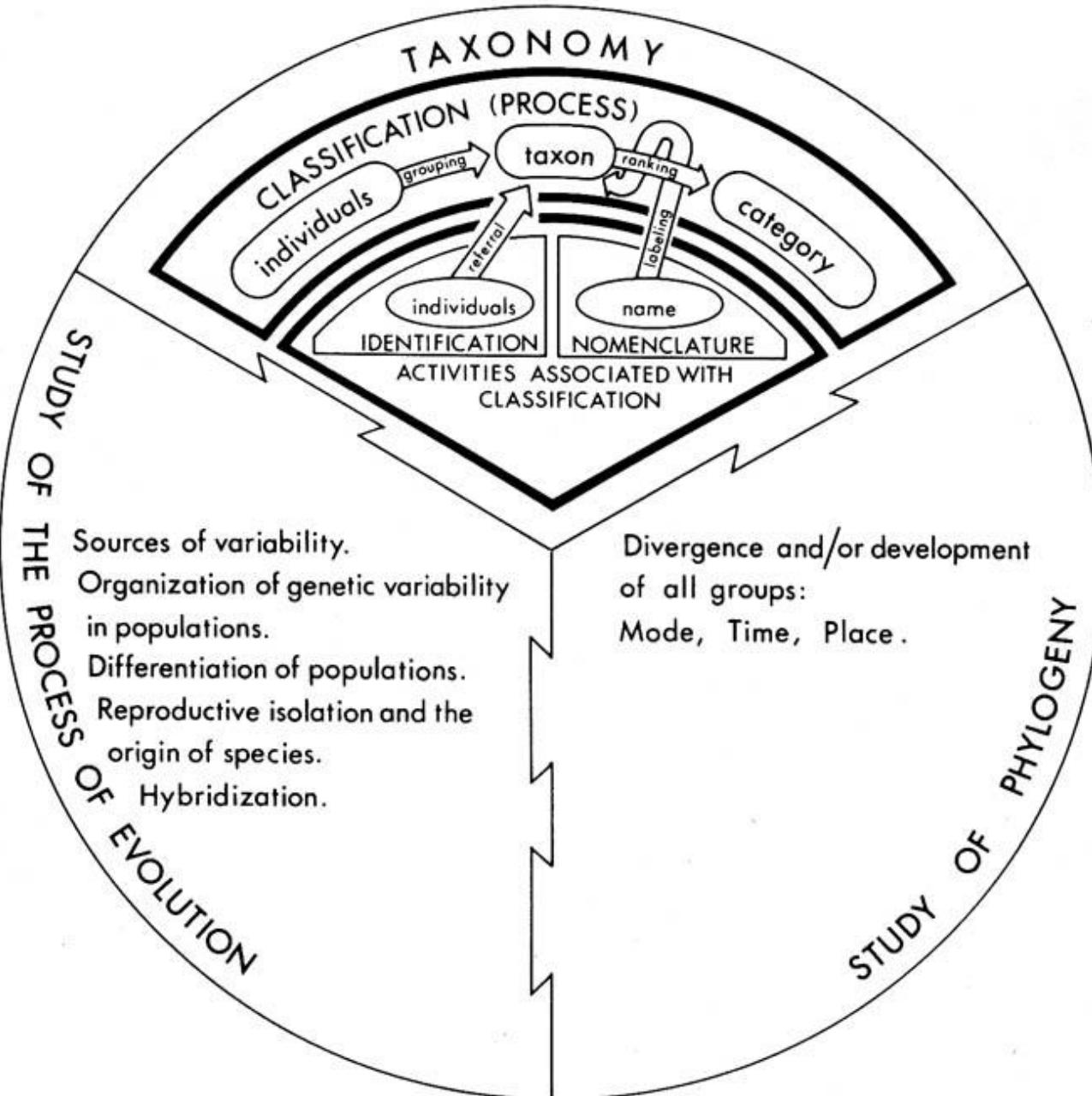
My terminology

Inferring trees from data- Phylogenetics

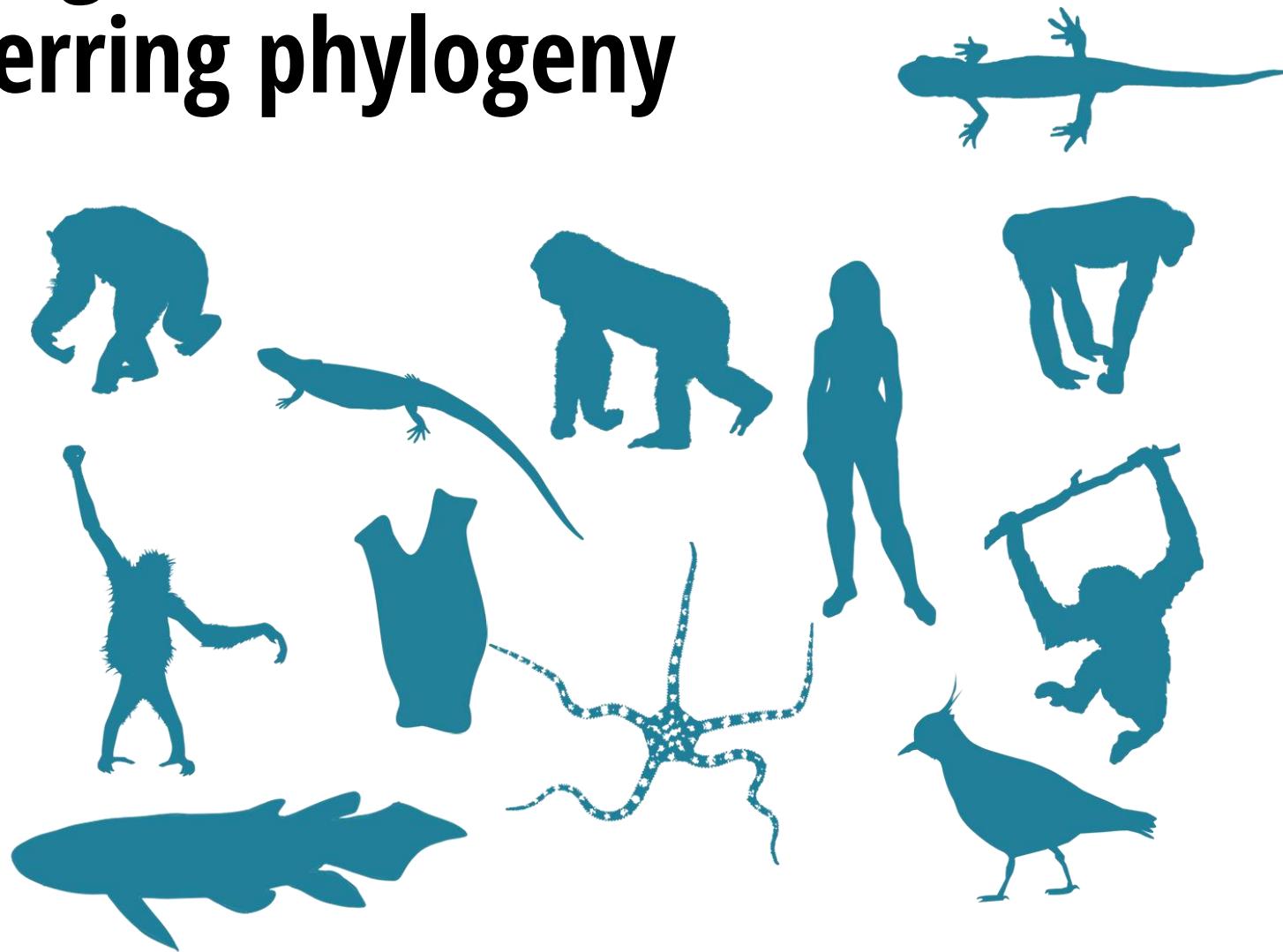
Using trees to study evolution- Phylogenetic comparative methods

Asking questions over deep time- Macroevolution

SYSTEMATICS



Going backwards- Inferring phylogeny



Tree space is huge!!

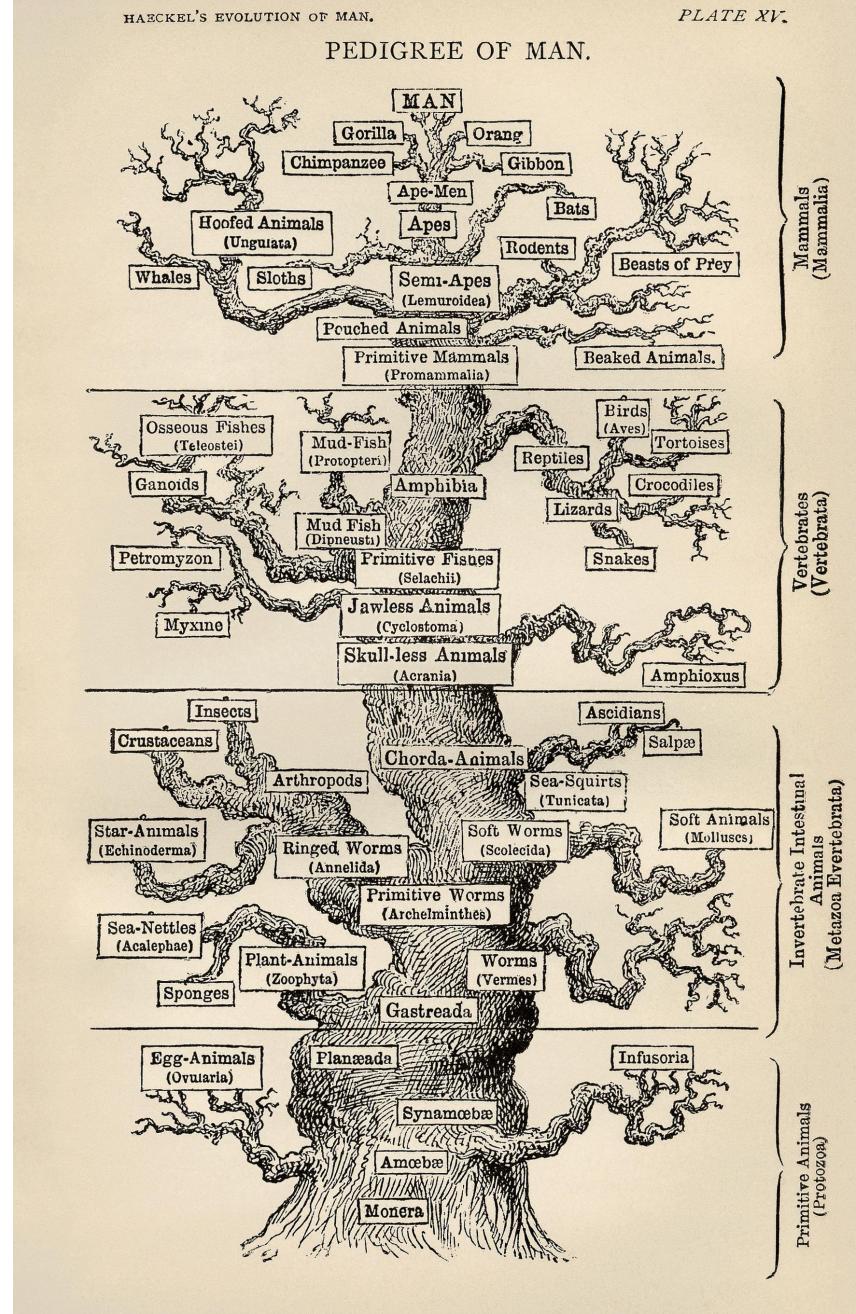
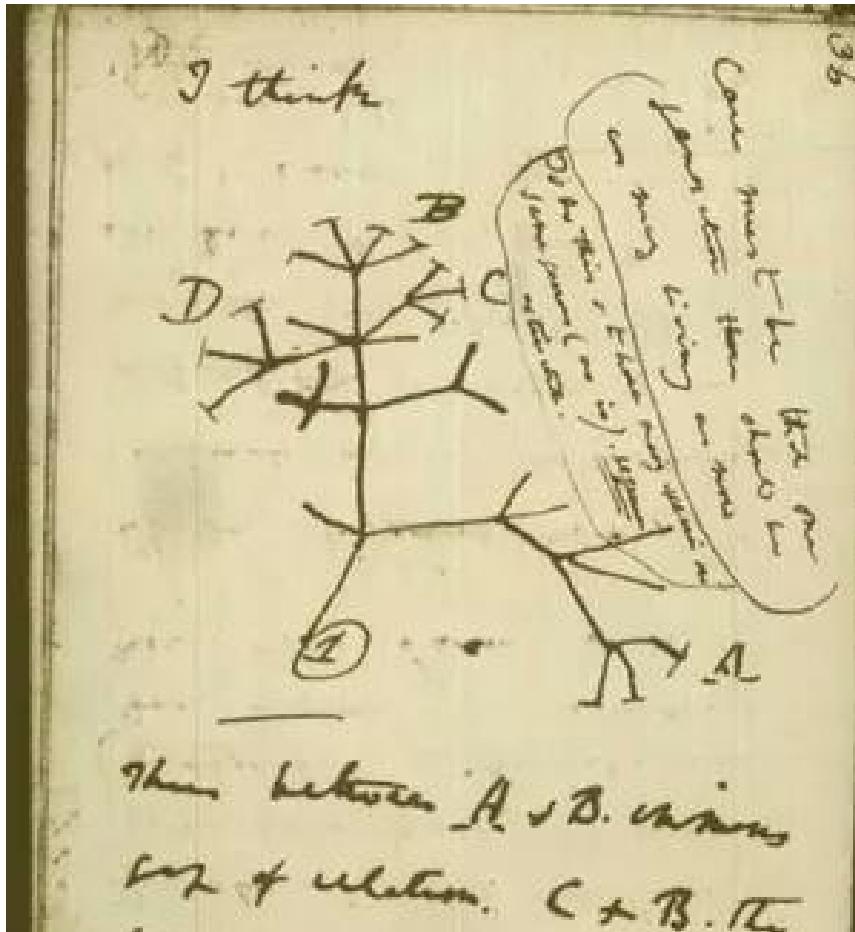
Inferring phylogenies is incredibly challenging statistical and computational task that has put phylogenetics at the forefront of many advances in statistics & computer science. This course will cover the story of phylogenetics and the basic models, as they are shared with phylogenetic comparative methods. However, for in depth “how-to” on building trees from molecular data, take Anton Suvorov’s course next semester!

Table 3.1: The number of rooted, bifurcating, labeled trees for n species, for various values of n . The numbers for more than 20 species are approximate.

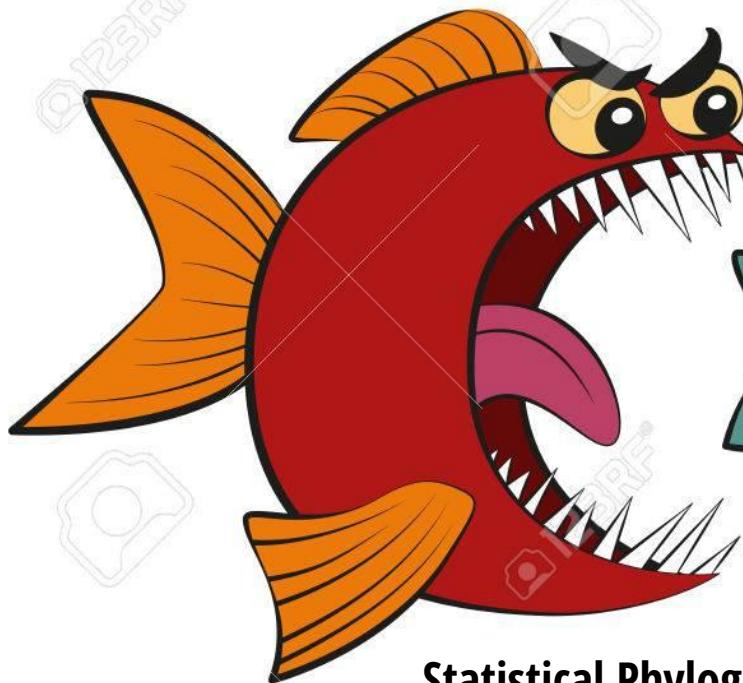
Species	Number of trees
1	1
2	1
3	3
4	15
5	105
6	945
7	10,395
8	135,135
9	2,027,025
10	34,459,425
11	654,729,075
12	13,749,310,575
13	316,234,143,225
14	7,905,853,580,625
15	213,458,046,676,875
16	6,190,283,353,629,375
17	191,898,783,962,510,625
18	6,332,659,870,762,850,625
19	221,643,095,476,699,771,875
20	8,200,794,532,637,891,559,375
30	4.9518×10^{38}
40	1.00985×10^{57}
50	2.75292×10^{76}

**How can we find the “true” tree?
(Is there a true tree?!?)**

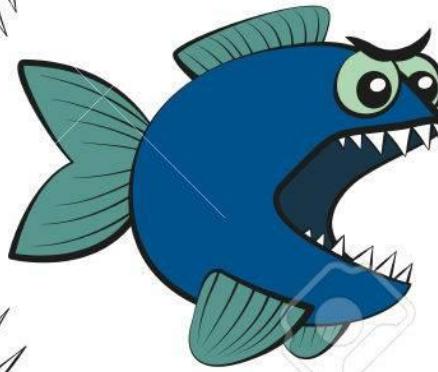
History of phylogenetics



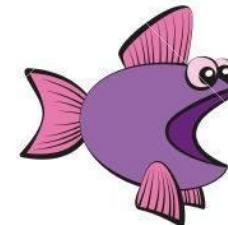
History of phylogenetic inference



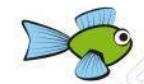
Statistical Phylogenetics
(1970s- present)



Cladistics
(1960s-90s)



Phenetics
(1950s-70s)

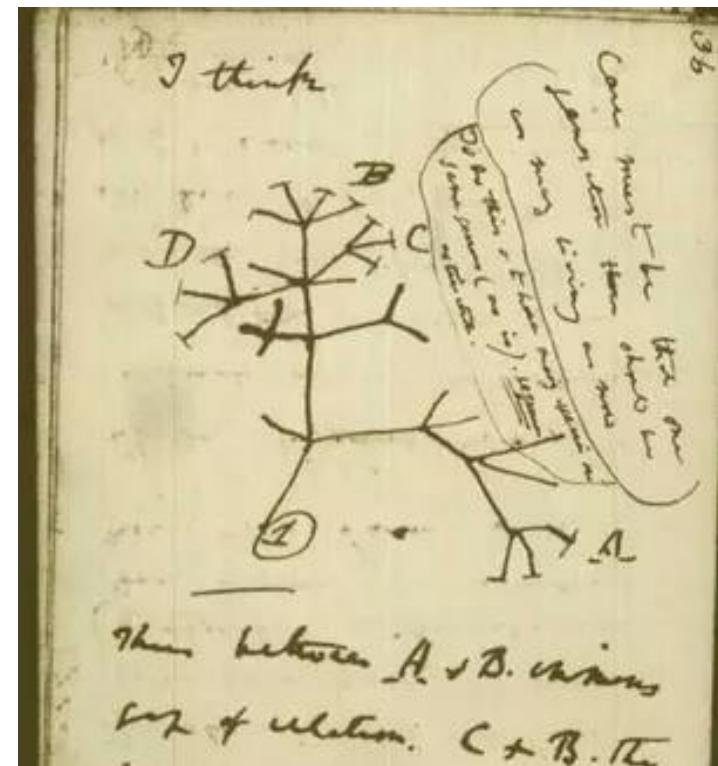


Evolutionary
Taxonomy
(Pre-1960s)

History of phylogenetics: Evolutionary Taxonomy

No objective methodology

Relies on expert opinion & a lifetime of study



History of phylogenetics: Phenetics

Numerical Taxonomy & distance methods - measure overall similarity

How do you measure similarity?

What are the drawbacks of this approach?

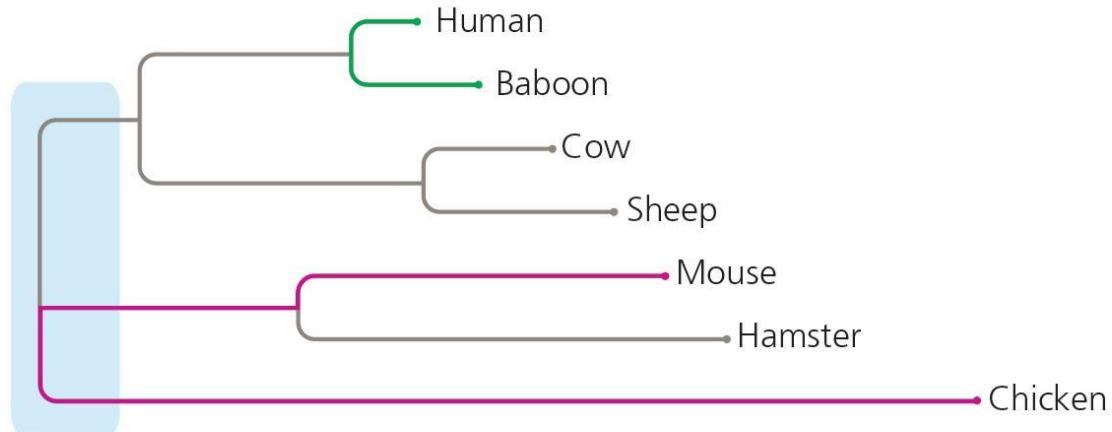
Sokal, Sneath & Rohlf (1960's) argued phenetics provided much needed reproducibility

A

Human	MV H L T P E E K S A V T A L W G K V N V D E V G G E A L G R L L V V Y P W T Q R F F E S F G D L S
Baboon	• • • • • • • • • • • • • N • D • • • • •
Cow	• - - • • A • • A • • • F • • • K •
Sheep	• - - • • A • • A • • • G F • • • K • • • • A • H • • •
Mouse	• • • • • D A • • A • • S G • • • • A • Y D • • •
Hamster	• • • • • D A • • A L • • G • • • • A • A • • A • • • • • • • • • • • • • • • • • • • H • • •
Chicken	• • • W • A • • Q L I • G • • • • A • C • A • • A • I • • • • A • • • • A • • • N • •

B

	Baboon	Cow	Sheep	Mouse	Hamster	Chicken
Human	2	6	9	8	9	13
Baboon	7	10	7	10	13	
Cow		3	11	12	16	
Sheep			12	9	15	
Mouse				7	16	
Hamster					14	

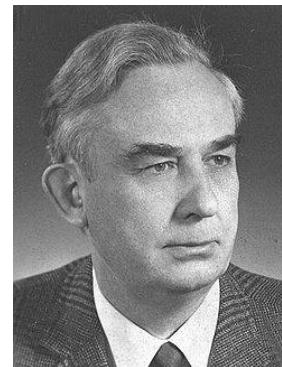
C

History of phylogenetics: Cladistics

"Phylogenetic systematics"

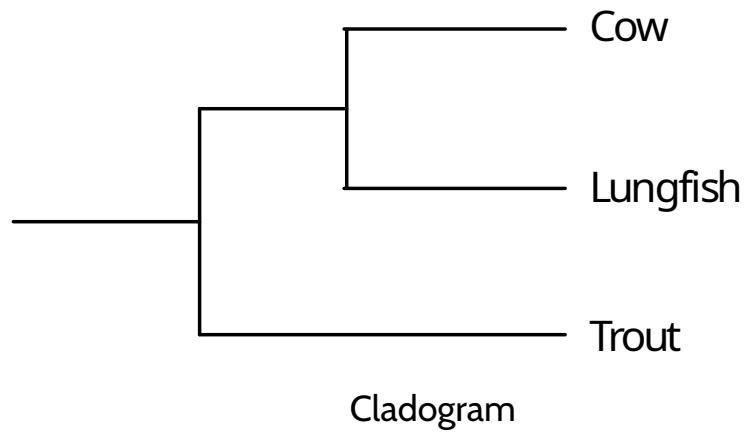
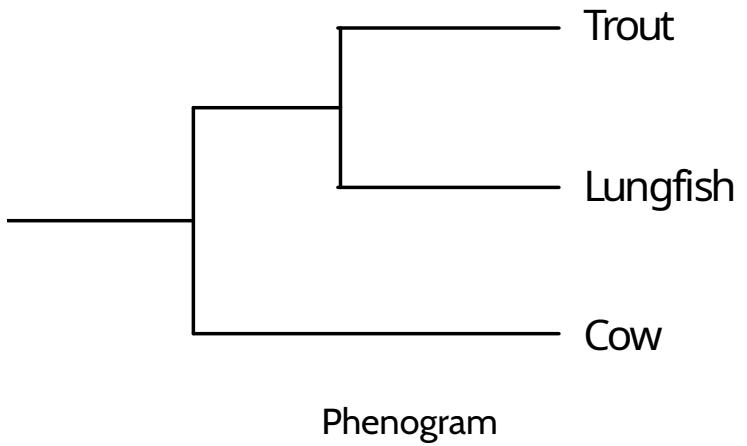
Argued that grouping should not be based on *overall trait similarity*, but only on *shared derived traits (synapomorphies)*

Taxonomy based on *monophyly*



Willi Hennig
(1966)

Parsimony & Occam's Razor- The simplest explanation is probably the correct one



What's the phylogenetic relationship?





Bones

Bones

Parsimony



Bones



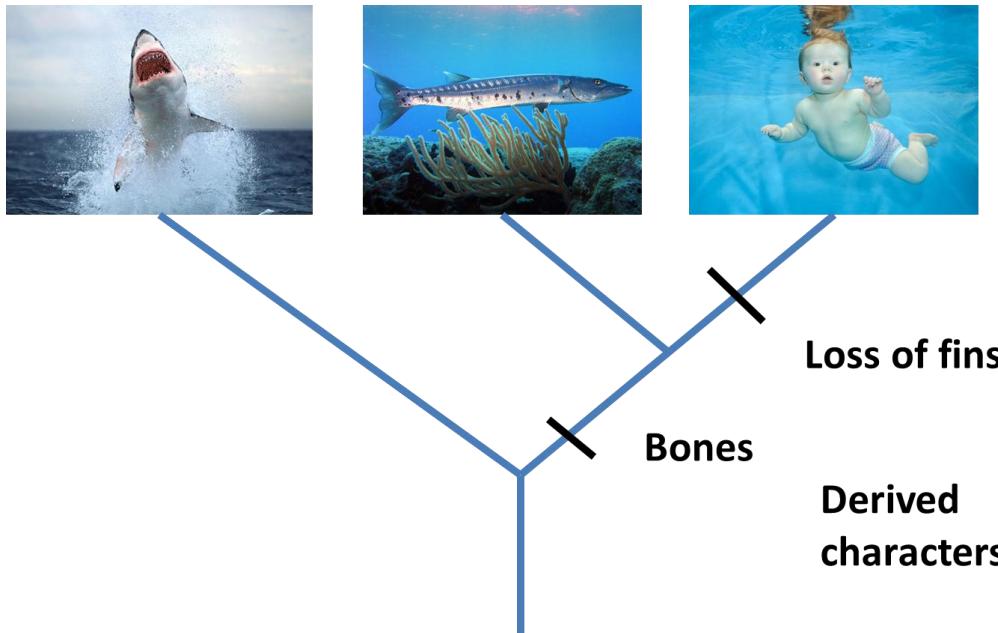
Bones

Fins

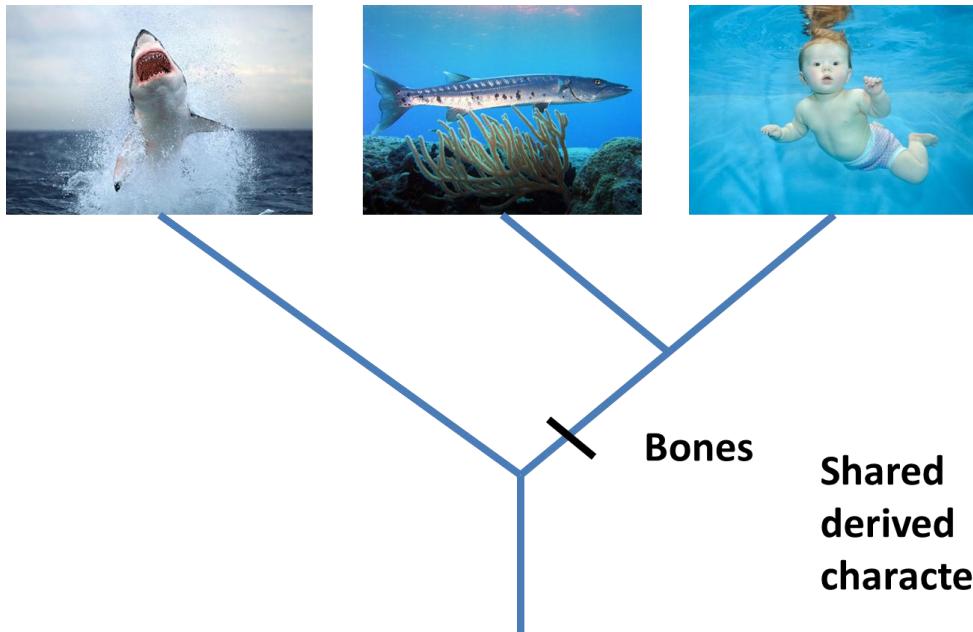
Loss of fins

A phylogenetic tree diagram showing the relationships between three organisms. The tree has a central node from which three branches extend to the right. The top branch leads to a shark image, the middle branch leads to a barracuda image, and the bottom branch leads to a baby swimming image. Each branch is labeled with a trait: 'Bones' is labeled on the middle branch, 'Fins' is labeled on the bottom branch, and 'Loss of fins' is labeled on the top branch. A thick black horizontal bar is positioned below the labels 'Bones' and 'Fins'.

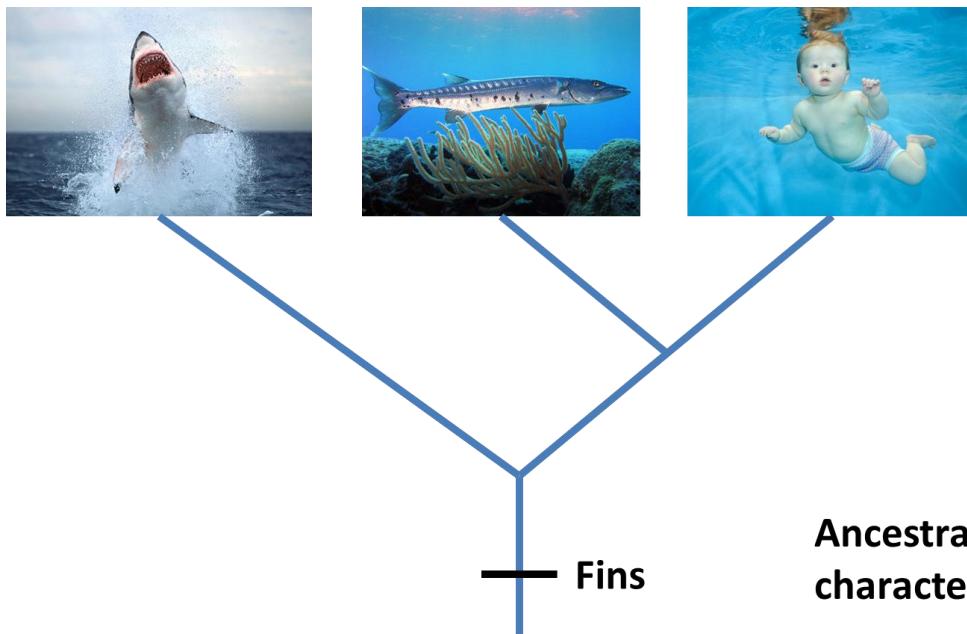
Derived Characters = Apomorphies



Shared derived characters = Synapomorphies



Shared ancestral characters = Symplesiomorphies



Overall similarity takes all characters equally

Cladistics only considers synapomorphies, and excludes autapomorphies & symplesiomorphies

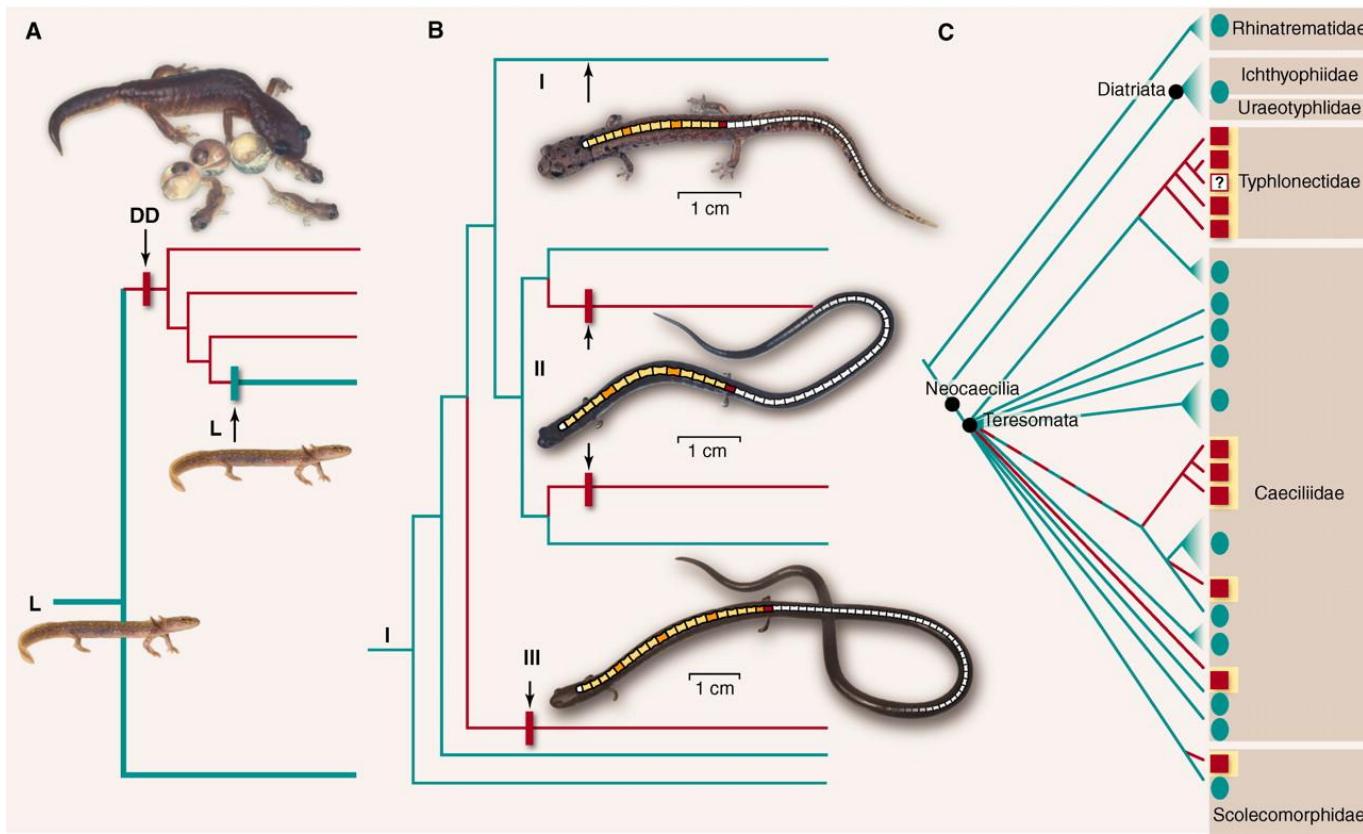
Best supported hypothesis: Minimizing # of steps

Critical role of outgroup to distinguish derived vs. ancestral characters

Also note that characters are defined as derived/ancestral relative to the taxa being considered (the ingroup)

Conflict = Homoplasy

Evolutionary reversals and/or convergence



Parsimony informative vs. uninformative

Must have at least 2 taxa with shared traits in 2 different states.

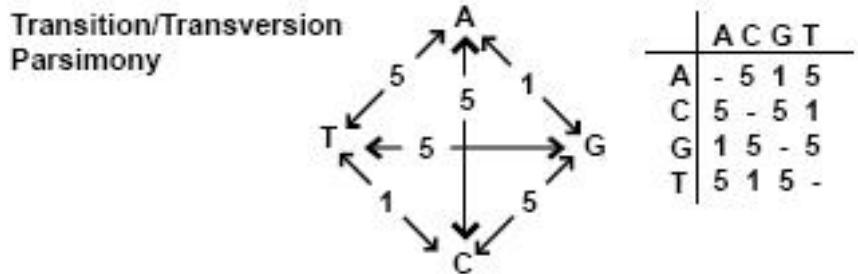
TABLE 7.7 Examples of informative and uninformative characters, with ? used to represent uncertain or missing character states.

Taxon	Characters														
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	1	0	1	0	?	1	0	1	0	0	0	0	1	0
3	0	1	1	2	1	0	2	0	1	1	0	0	2	0	
4	1	1	1	1	?	1	2	0	1	0	1	1	2	1	
5	1	1	2	3	1	1	0	0	1	2	?	2	2	2	
6	1	0	2	2	1	?	1	0	1	0	?	3	2	?	

Generalized Parsimony

Step matrices

To				
	Ordered	Unordered	Dollo	Camin-Sokal
	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3
F 0	- 1 2 3	- 1 1 1	- M 2M 3M	- 1 2 3
r 1	1 - 1 2	1 - 1 1	1 - M 2M	i - 1 2
o 2	2 1 - 1	1 1 - 1	2 1 - M	i i - 1
m 3	3 2 1 -	1 1 1 -	3 2 1 -	i i i -



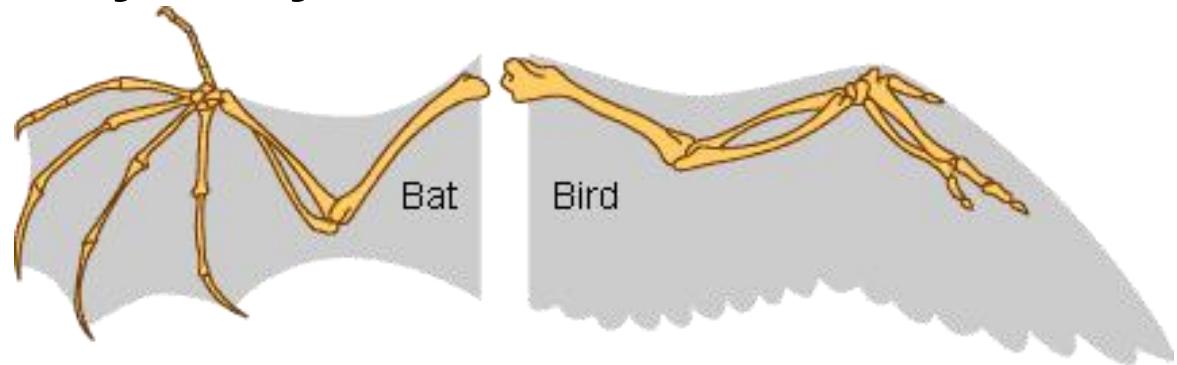
By Peter Forey

<https://www.palass.org/publications/newsletter/cladistics-palaeontologists/cladistics-palaeontologists-part-4-optimisation>

Homology

Primary homology - Initial hypothesis of homology due to similarity

Secondary homology - Implied homology single origin on phylogeny by parsimony analysis



*Underlines importance of character construction - often surprisingly subjective
"Reciprocal illumination" Hennig 1966*

History of phylogenetic inference: Statistical phylogenetics

Luca Cavalli-Sforza, Anthony Edwards and later, Joseph Felsenstein

Probabilistic models of trait evolution

**Showed statistical issues with parsimony,
e.g. "the Felsenstein zone"**

**-> Highlighted the importance of
simulation!**

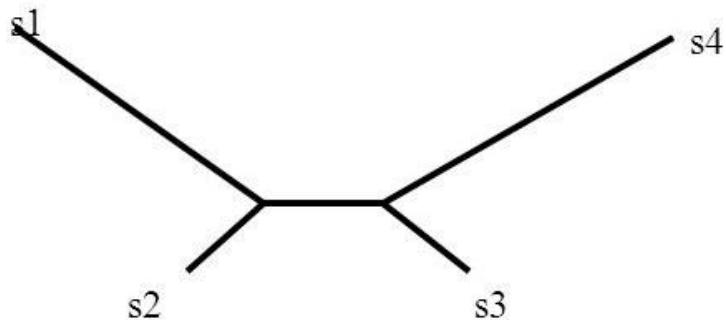
Maximum Likelihood & Bayesian inference



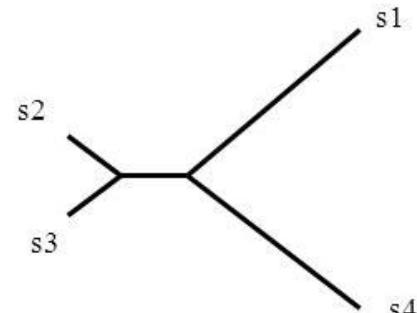
Joe Felsenstein

The Felsenstein Zone & "Long-branch attraction"

True Tree



Reconstructed Tree



More generally, parsimony fell out of favor for molecular data due to more flexibility in incorporating substitution models in statistical phylogenetics

My oversimplification of the problem...

Under the cladistics philosophy (model), homoplasy is a “mistake” in character construction

Coding independently evolved traits the same, look closer and you'll see they're different!! (e.g. wings of birds & wings of bats)

But for DNA, complete convergence is real (a “T” is a “T” is a “T”, no matter how it became that)

As similarity by convergence becomes more common than similarity by shared ancestry, parsimony fails, and no “looking closer” at the character will save it.

Parsimony main assumption "Occam's Razor" has unclear implications for what this means for the assumed evolutionary process

Ultimately -> rare changes or "no common mechanism"
(Tuffley and Steele 1997)

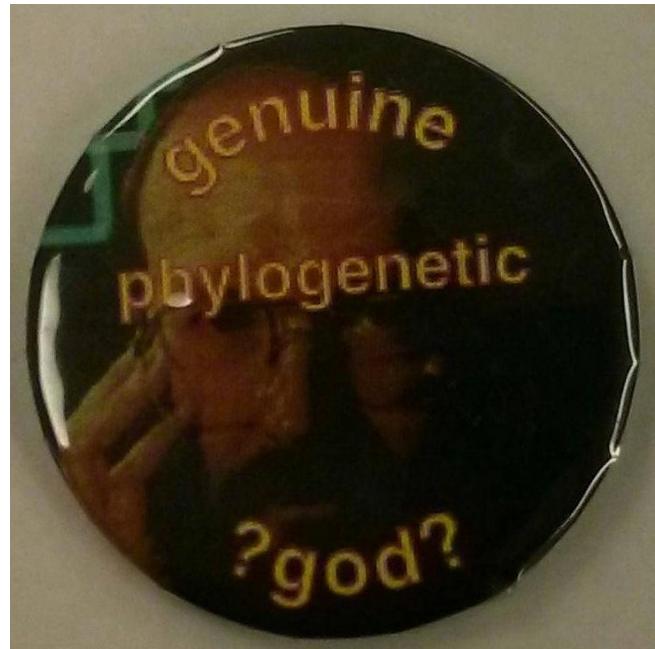
Inconsistent estimator - An estimator that gets more and more certain of the wrong answer as more data is added

Conflict in science

MATT SIMON SCIENCE 02.03.16 07:00 AM

TWITTER NERD-FIGHT REVEALS A LONG, BIZARRE SCIENTIFIC FEUD

#ParsimonyGate



The epistemological paradigm of this journal is parsimony. There are strong philosophical arguments in support of parsimony versus other methods of phylogenetic inference (e.g. Farris, [1983](#)).

The high citation index of *Cladistics* shows that the journal is publishing some of the most ground-breaking empirical and theoretical research on the history of life, and we remain committed to the publication of outstanding systematics research. As a community of scientists, the Willi Hennig Society is always open to new methods and ideas, and to well-reasoned criticisms of old ones. However, we do not hold in special esteem any method solely because it is novel or purportedly sophisticated.

Phylogenetic data sets submitted to this journal should be analysed using parsimony. If alternative methods are also used and there is no difference among the results, the author should defer to the principles of the Society and present the tree obtained by parsimony. Unless there is a pertinent reason to include multiple trees from alternative methods, a tree based on parsimony is sufficient as an intelligible, informative and repeatable hypothesis of relationships, and articles should not be cluttered with multiple, often redundant, trees produced from other methods. If alternative methods give different results and the author prefers an unparsimonious topology, he or she is welcome to present that result, but should be prepared to defend it on philosophical grounds.

In keeping with numerous theoretical and empirical discussions of methodology published in this journal, we do not consider the hypothetical problem of statistical inconsistency to constitute a philosophical argument for the rejection of parsimony. All phylogenetic methods, including parsimony, may produce inconsistent or otherwise inaccurate results for a given data set. The absence of certain truth represents a philosophical limit of empirical science.

Cladistics will publish research based on methods that are repeatable, clearly articulated and philosophically sound. We believe these guidelines implement the vision of Willi Hennig ([1965](#), p. 97), who said, "(i)nvestigation of the phylogenetic relationship between all existing species and the expression of the results of this research in a form which cannot be misunderstood, is the task of phylogenetic systematics."

Parsimony still commonly used for:

Morphological analyses (though probabilistic models exist here too)

Starting trees

Understanding history of trait change (ancestral state reconstructions)