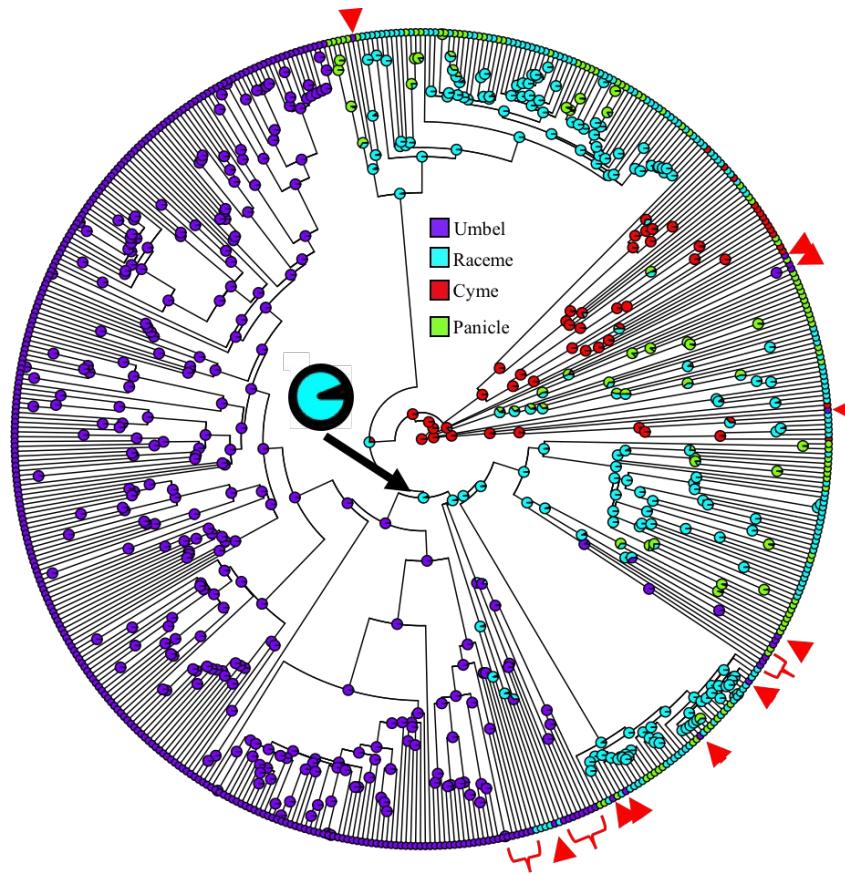


# Evolutionary and Ecology



# Module Outline:

- Day 1: Phylogenetics Intro, Alignment & Parsimony
- Day 2: Maximum Likelihood/ Bayesian Methods
- Day 3: Introduction to Ecology + MaxENT

# Instructors



Jesus Martinez-Gomez  
Ph.D Candidate



Eugenio Valderrama  
Post Doc

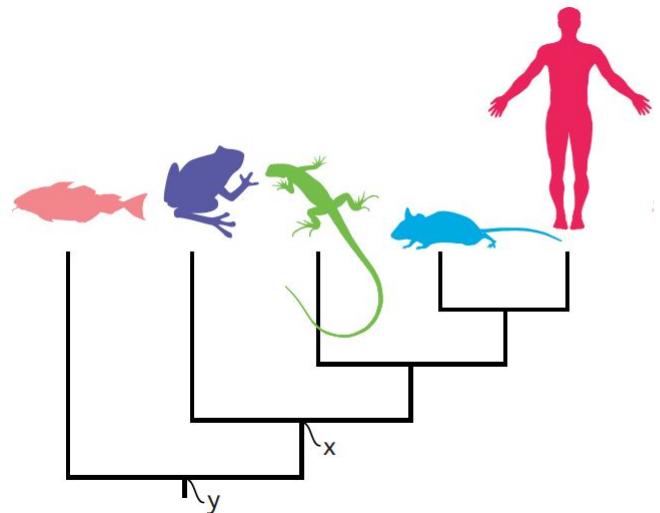
# Day 1: Outline

- General Concepts
  - What is phylogenetics
  - Evolutionary Theory
  - Homology
- Tree Thinking:
  - Reading Trees
- How to Infer a Phylogeny
  - Data
  - DNA alignment
  - Inference Methods
  - Parsimony
- Summary
  - Why should you care
  - Lab Introduction
  - Resources
  - Glossary of Terms

# Introduction: What is Phylogenetics

**Phylogenetics:** is the study of the evolutionary history and relationships among individuals or groups of organisms (e.g. species, or populations)

- Wikipedia

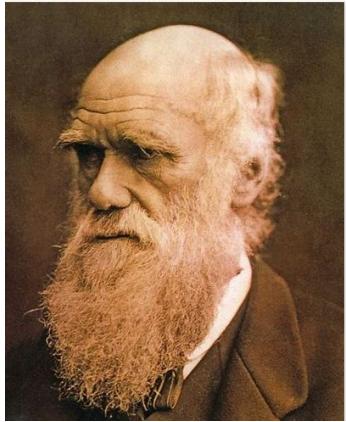


# **Introduction: What is Phylogenetics**

## **Types of questions you can answer**

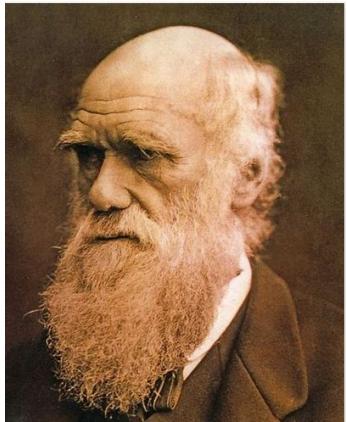
- Phylogenetics:
  - What is the evolutionary relationships of these organisms?
  - What is the evolutionary relationship of these gene (e.g., orthologs, paralogs) ?
- Phylogenetic Comparative Methods:
  - What is the ancestral condition of a trait or gene function?
  - How fast to certain functions evolve?

# Introduction: Evolutionary Theory

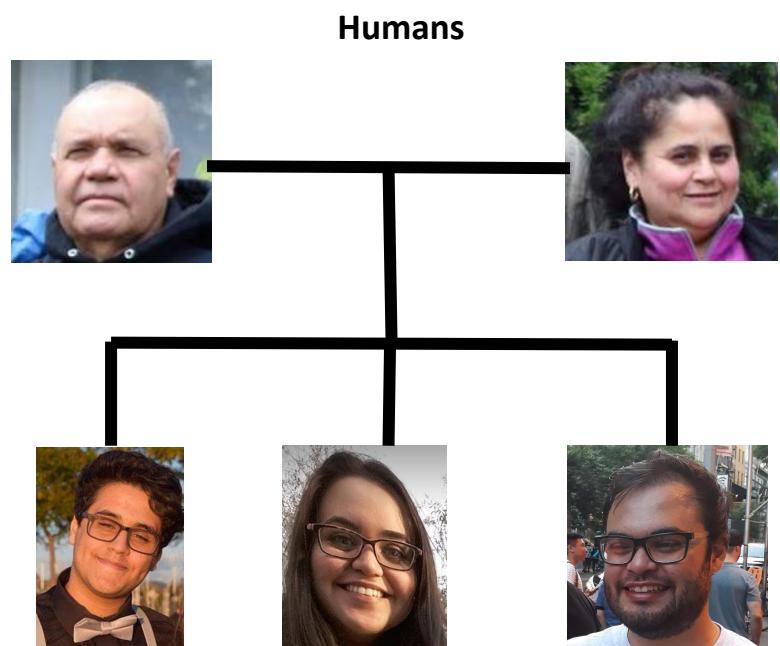


**Charles Darwin**  
**Evolution** - Descent with  
Modification

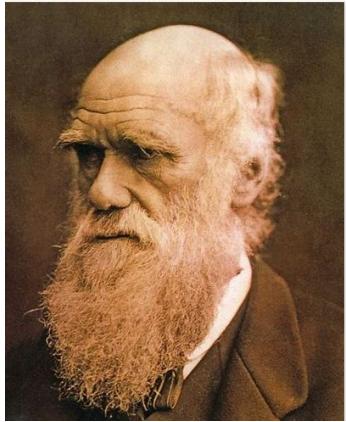
# Introduction: Evolutionary Theory



**Charles Darwin**  
**Evolution** - Descent with  
Modification

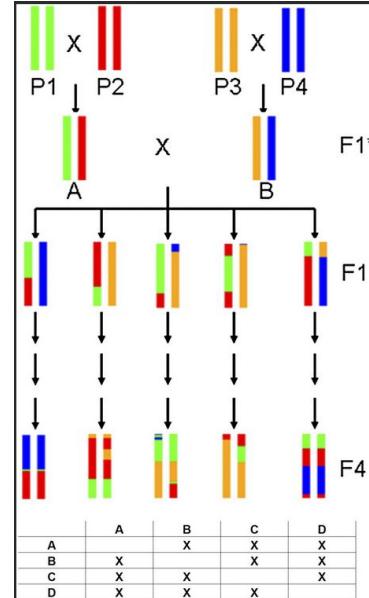


# Introduction: Evolutionary Theory



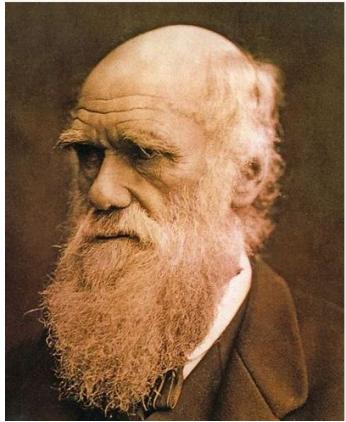
**Charles Darwin**  
**Evolution** - Descent with  
Modification

**Arabidopsis Recombinant Inbreed Line (RIL)**

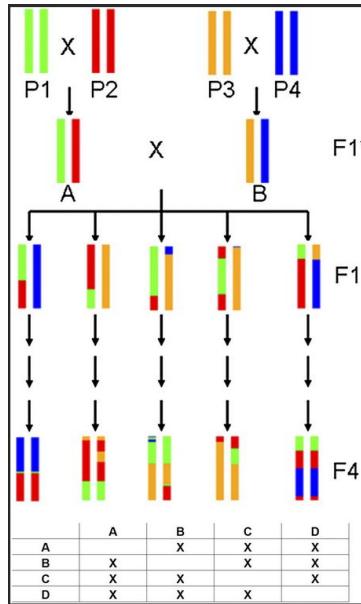


# Introduction: Evolutionary Theory

On a population level we observe that individuals are related by ancestry. We can extrapolate out to all living life to state that all biological life is related.



**Charles Darwin**  
**Evolution - Descent with Modification**

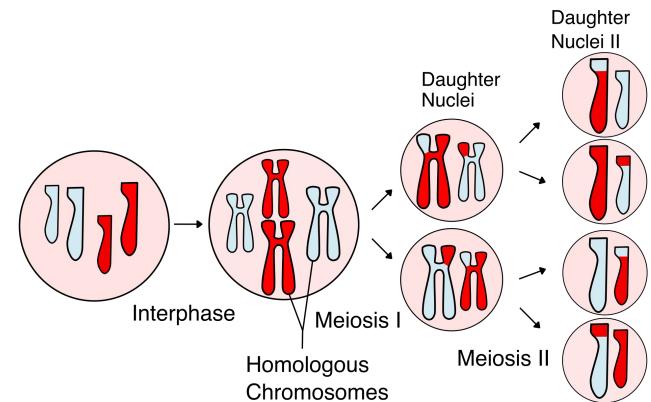
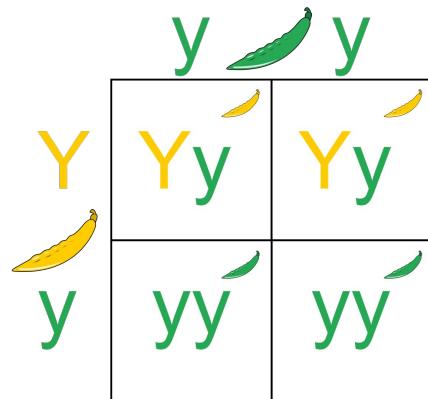


# Introduction: Evolutionary Theory

Patterns of genetic inheritance are predictable and molecular biology provides the mechanisms, DNA replication and meiosis.



**Gregor Mendel**  
Gene - Unit of Inheritance



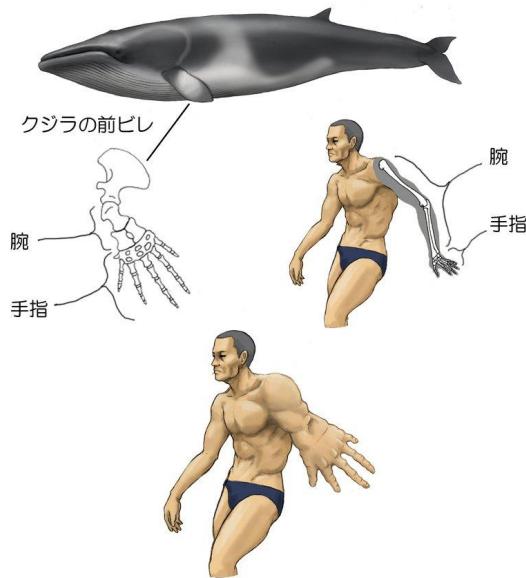
# Introduction: Homology

If we accept Darwinian evolution, we can make one very important argument, a homolog statement.

- **Homology:** Similarities between individuals due to evolutionary history
- **A Homology statement:** An hypothesis that two or more structures are similar because of descent with modification.

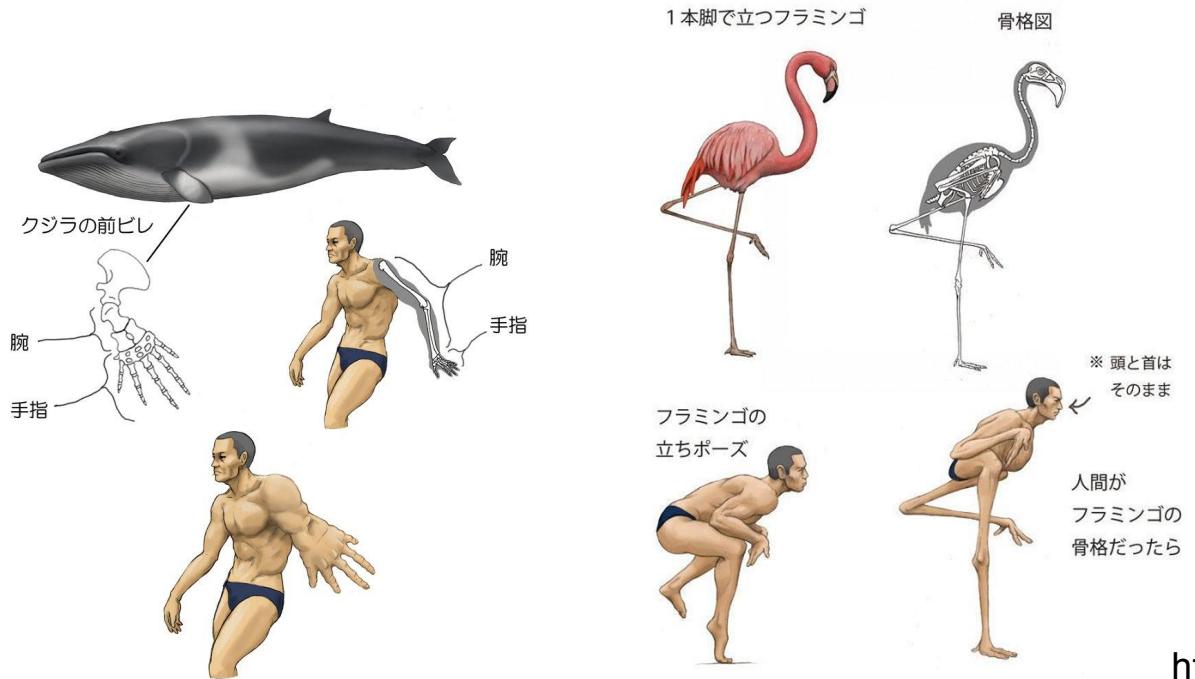
# Introduction: Homology

If we accept Darwinian evolution, we can make one very important argument, a homolog statement.



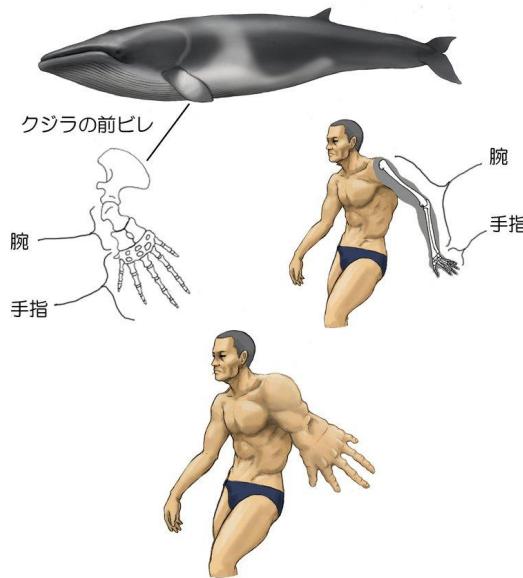
# Introduction: Homology

If we accept Darwinian evolution, we can make one very important argument, a homolog statement.



# Introduction: Homology

Homology statements are **hypothesis**. Unlike population level, we were not there to observe evolution because it occurred of millions of years therefore we need to build evidence.



**Homology Statement:** The human forearm and whale flipper are homologous. My evidence for that is:

- The number and types of bones are similar
- Bone arrangement is similar
- Similar developmental pattern

# Introduction: Homology

Homology statements are **hypothesis**. Unlike population level, we were not there to observe evolution because it occurred millions of years therefore we need to build evidence.

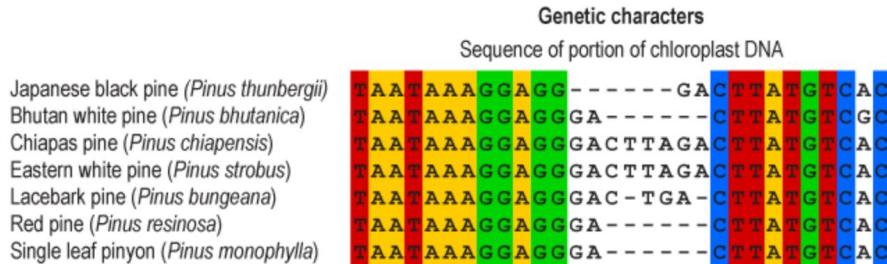
**Question:** This is an alignment of a chloroplast gene. What is some evidence that they are homologous?

Genetic characters	
Sequence of portion of chloroplast DNA	
Japanese black pine ( <i>Pinus thunbergii</i> )	T A A T A A A G G A G G ----- G A C T T A T G T C A C
Bhutan white pine ( <i>Pinus bhutanica</i> )	T A A T A A A G G A G G G A ----- C T T A T G T C G C
Chiapas pine ( <i>Pinus chiapensis</i> )	T A A T A A A G G A G G G A C T T A G A C T T A T G T C A C
Eastern white pine ( <i>Pinus strobus</i> )	T A A T A A A G G A G G G A C T T A G A C T T A T G T C A C
Lacebark pine ( <i>Pinus bungeana</i> )	T A A T A A A G G A G G G A C - T G A - C T T A T G T C A C
Red pine ( <i>Pinus resinosa</i> )	T A A T A A A G G A G G G A ----- C T T A T G T C A C
Single leaf pinyon ( <i>Pinus monophylla</i> )	T A A T A A A G G A G G G A ----- C T T A T G T C A C

# Introduction: Homology

Homology statements are **hypothesis**. Unlike population level, we were not there to observe evolution because it occurred millions of years therefore we need to build evidence.

**Question:** This is an alignment of a chloroplast gene. What is some evidence that they are homologous?



**Homology Statement:** This region of chloroplast DNA is homologous for all these pine trees. My evidence for that is:

- The genes have similar sequence
- The protein they encoded have similar function
- The genes are all found in the chloroplast genome,

# Introduction: Summary

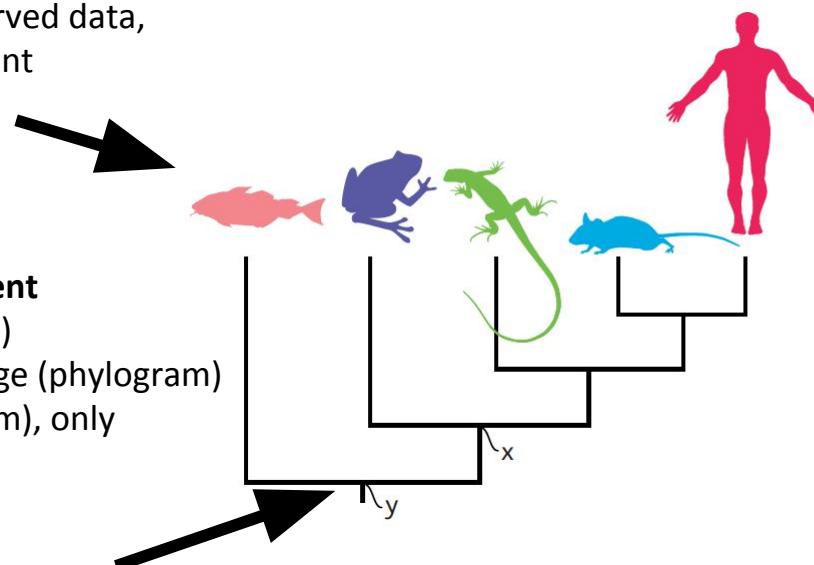
- All biological life is related.
- Because all biological life is related we can make homology statements
- Phylogenetics is ultimately a test of homology

# Day 1: Outline

- General Concepts
  - What is phylogenetics
  - Evolutionary Theory
  - Homology
- Tree Thinking:
  - Reading Trees
- How to Infer a Phylogeny
  - Data
  - DNA alignment
  - Inference Methods
  - Parsimony
- Summary
  - Why should you care
  - Lab Introduction
  - Resources
  - Glossary of Terms

# Tree Thinking: Reading Trees

**Tips:** Represent observed data,  
can be extinct or extant



**Branches can represent**

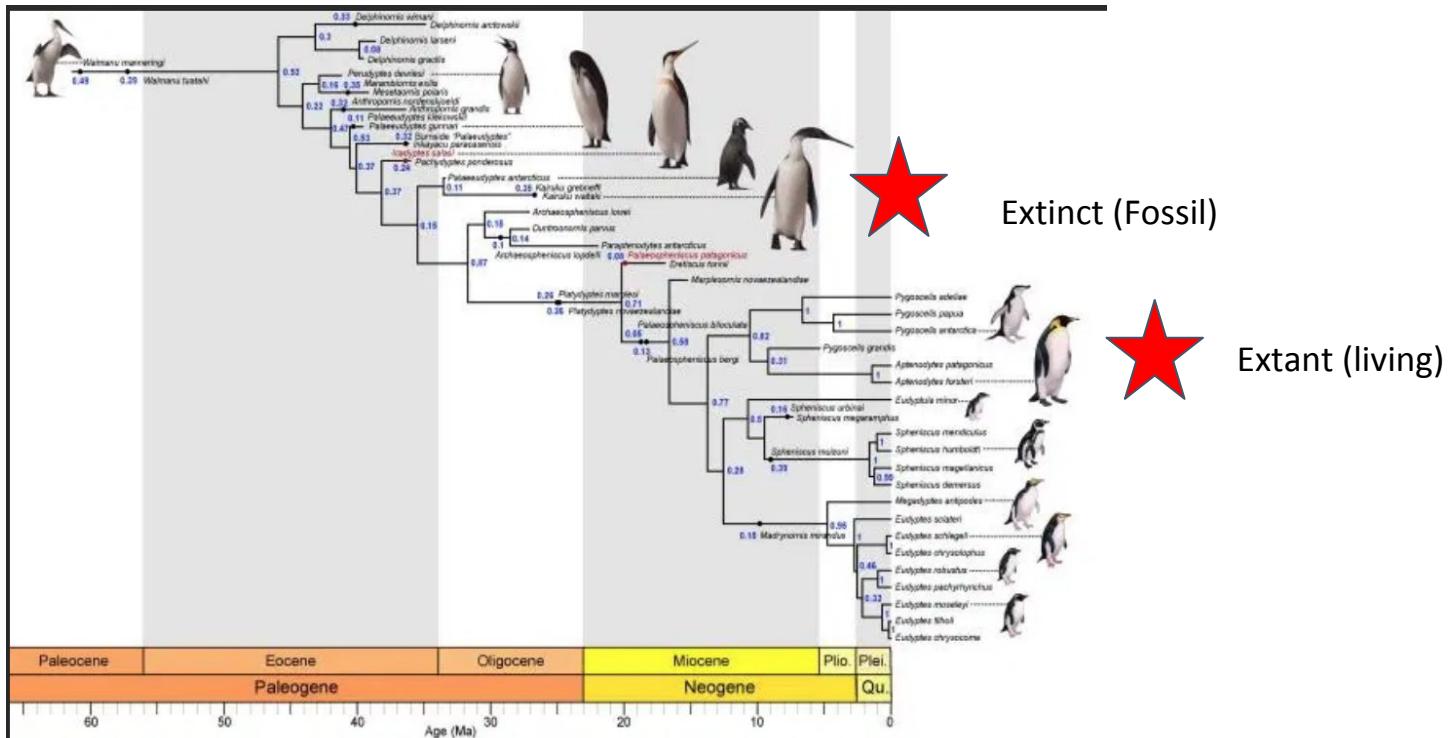
- Time (chronogram)
- Evolutionary change (phylogram)
- Nothing (cladogram), only relationships

**Internal Nodes:**

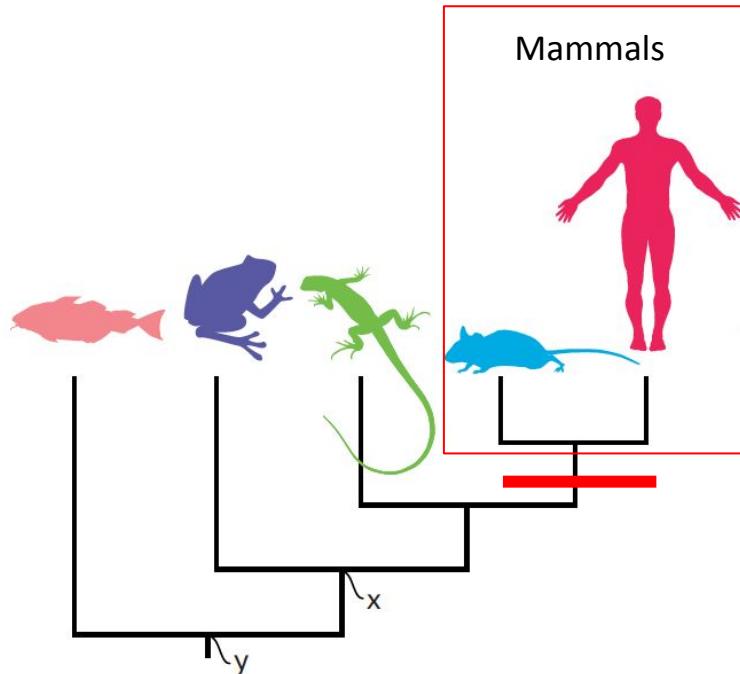
(Represent hypothetical ancestors)

# Tree Thinking: Reading Trees

An example of chronogram



# Tree Thinking: Reading Trees



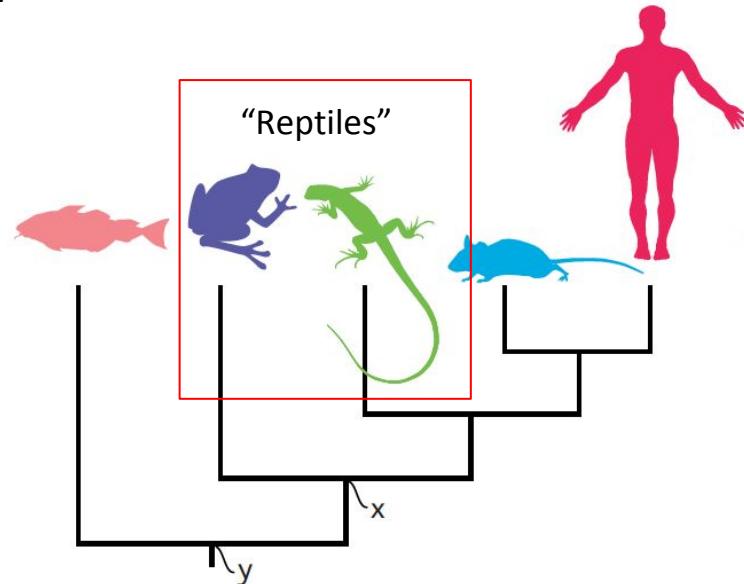
**Monophyletic (Clade):**  
An ancestor (node) and **all** its descendants.

**Synapomorphy:** A shared derived trait.

- Fur
- Mammary glands

# Tree Thinking: Reading Trees

**Question:** Is this group a clade?

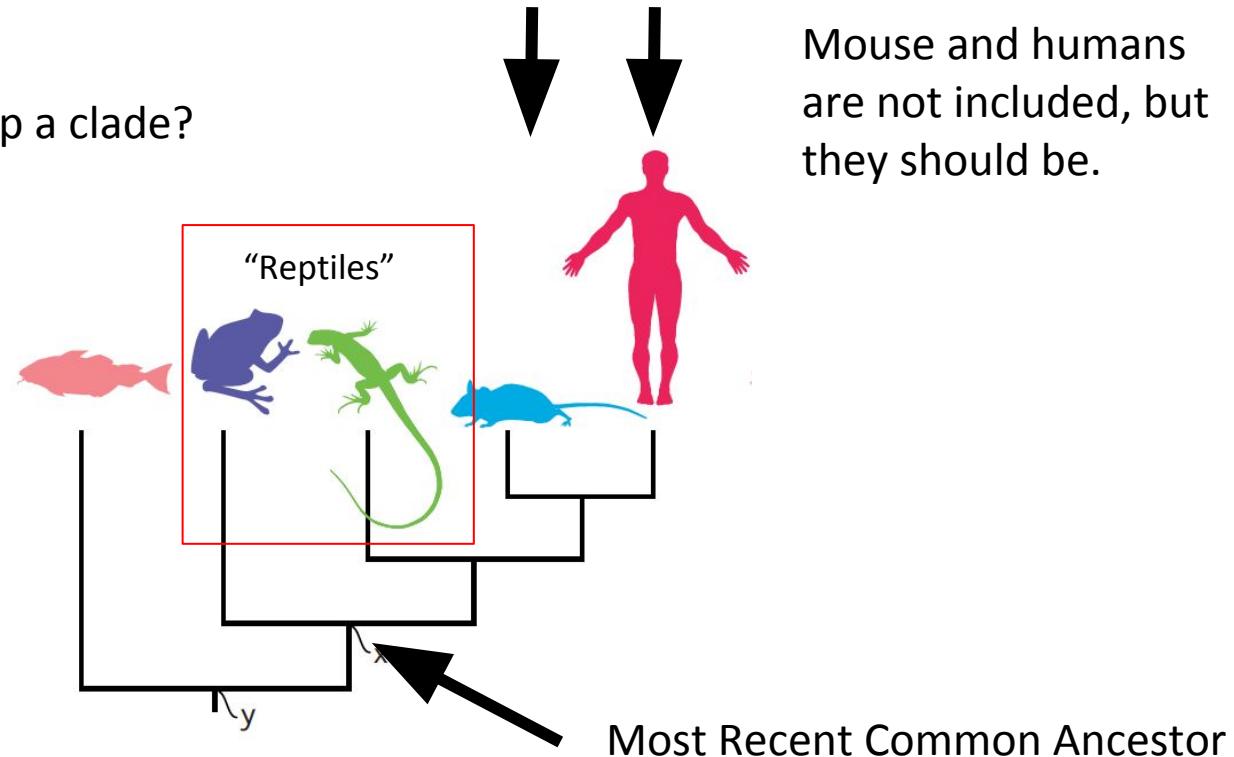


# Tree Thinking: Reading Trees

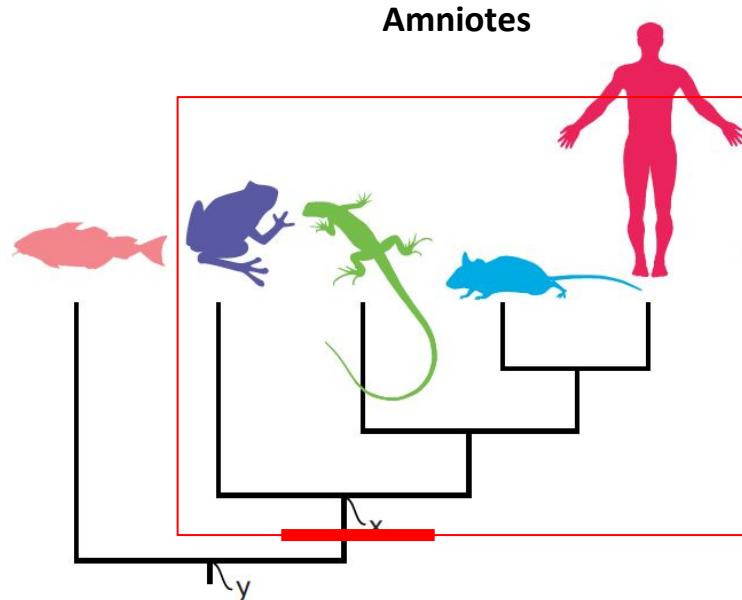
**Question:** Is this group a clade?

No! Does not not include all descended from the most recent common ancestor

Mouse and humans are not included, but they should be.



# Tree Thinking: Reading Trees

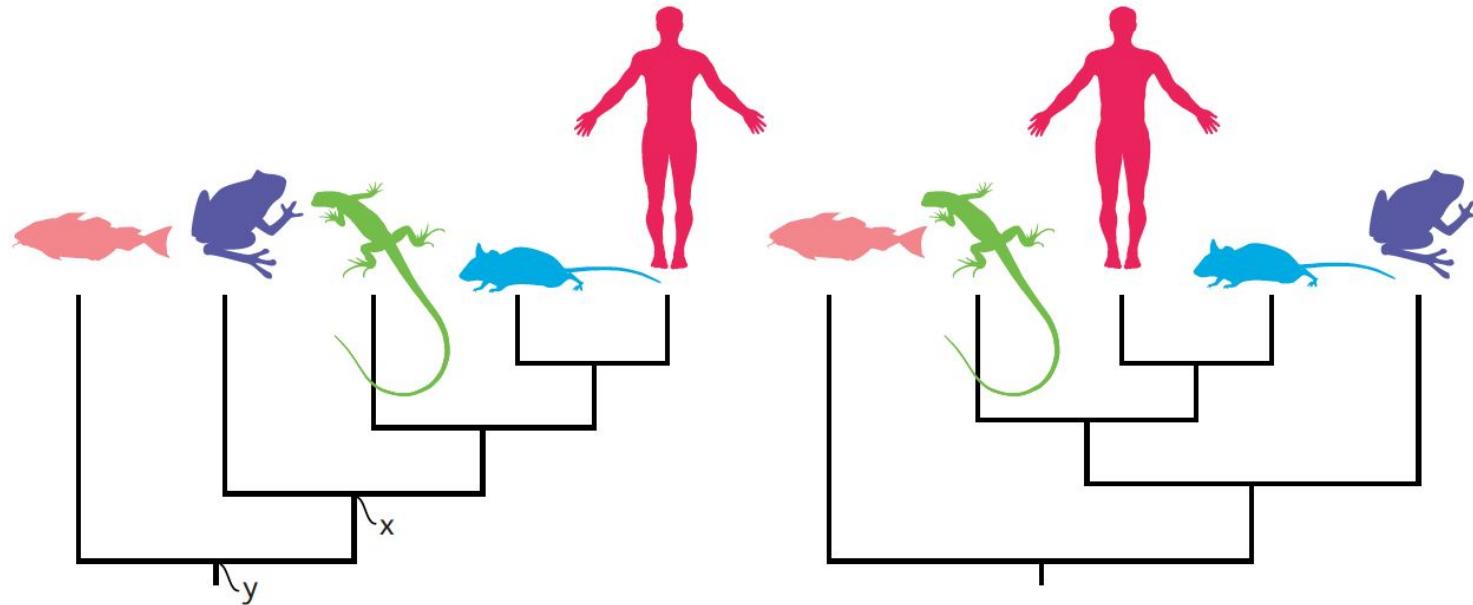


**Synapomorphy:** A shared derived trait.

- Lay eggs on land or retain egg in mother

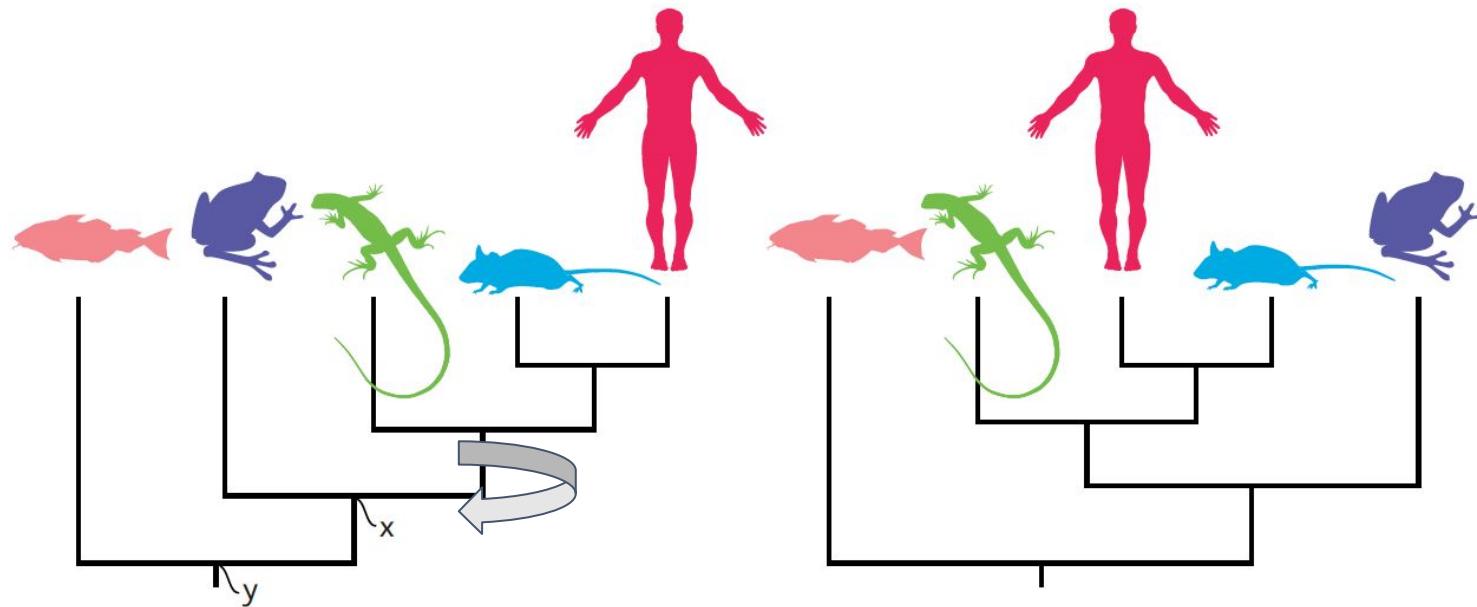
# Tree Thinking: Reading Trees

**Question:** Do these two tree show the same or different evolutionary history?



# Tree Thinking: Reading Trees

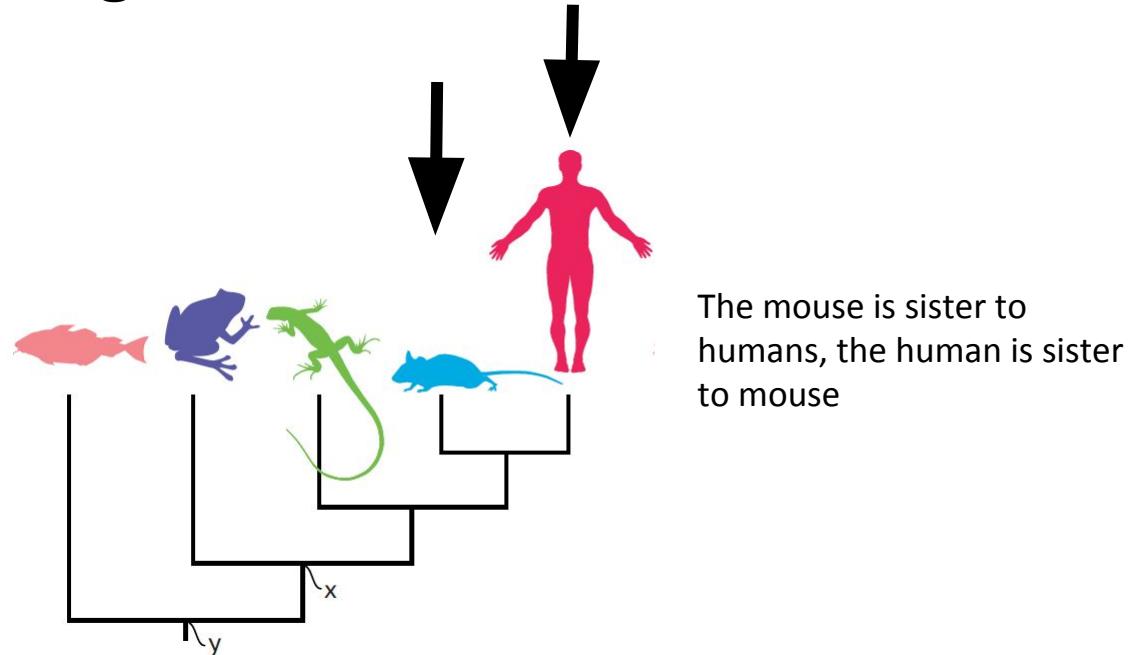
They show the **same** evolutionary history. The nodes are just rotated



# Tree Thinking: Reading Trees

## Sister Taxa:

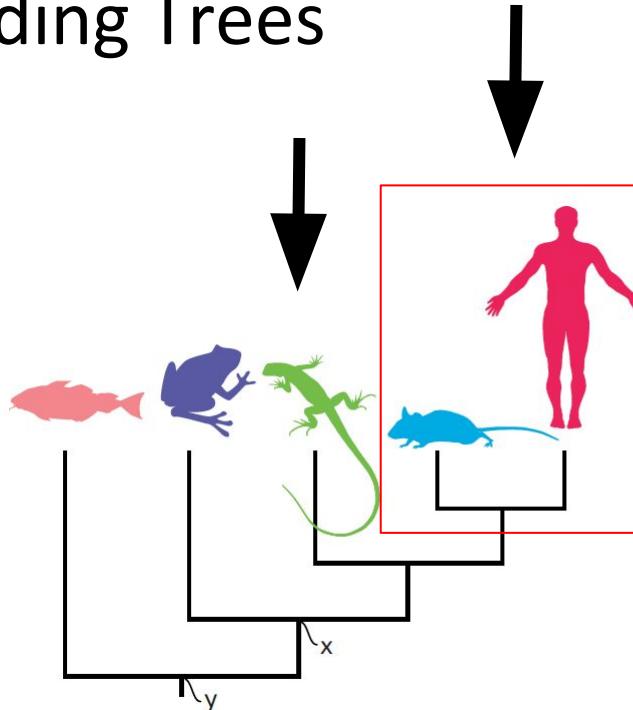
Two groups that share a common ancestor. They can be individuals or clades



# Tree Thinking: Reading Trees

## Sister Taxa:

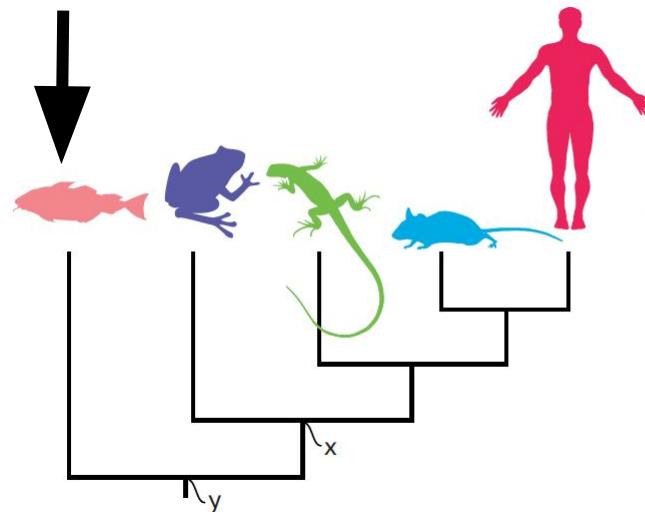
Two groups that share a common ancestor. They can be individuals or clades



The lizard is sister to the clade containing mouse+humans. The clade containing mouse+humans is sister to lizard

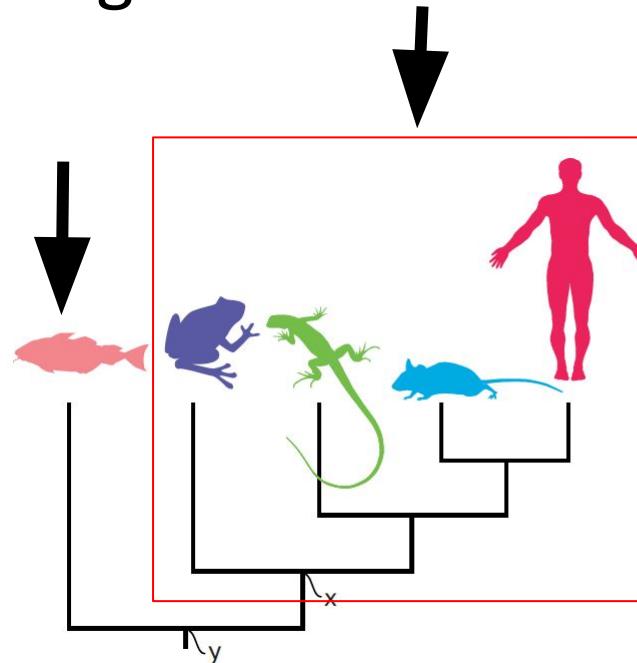
# Tree Thinking: Reading Trees

**Question:** What is the sister to fish?



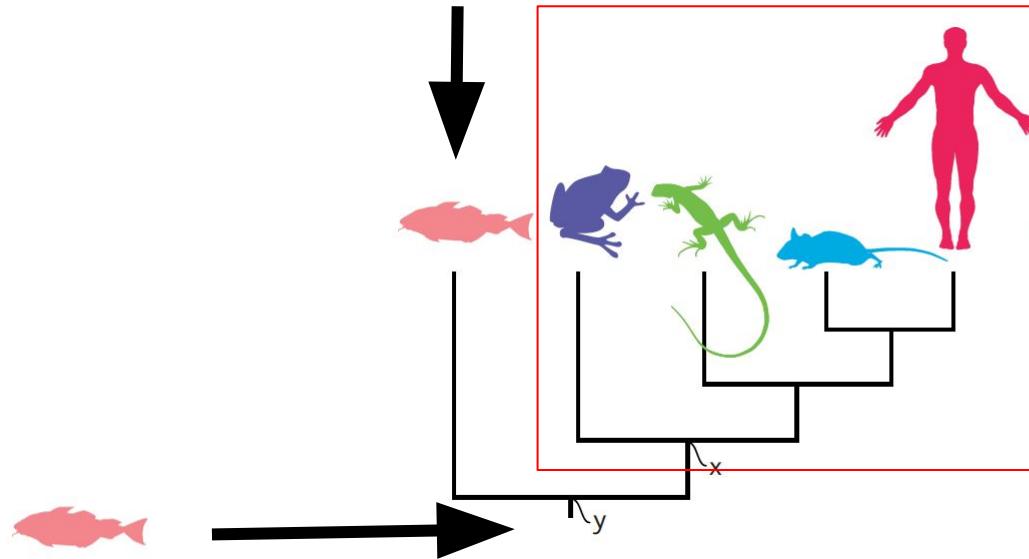
# Tree Thinking: Reading Trees

**Question:** What is the sister to fish?  
The clade containing everyone else!



# Tree Thinking: Reading Trees

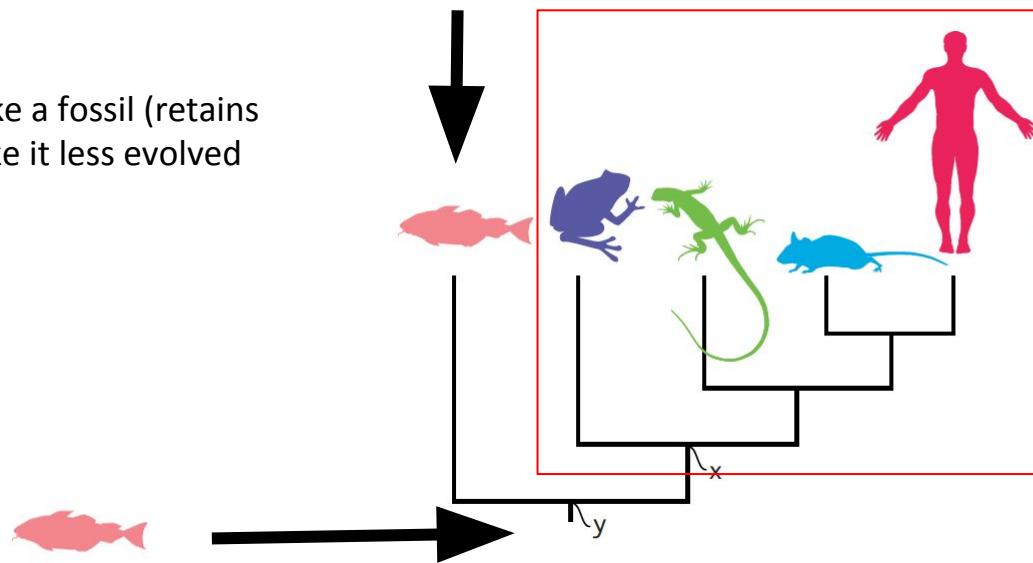
In the tree of life you often see a similar pattern depicted here, where you have a lineage with low number of species, sister to a large clade with many species. Often times the word, **basal** or **early diverging** will be used to describe this low - number of species lineage. This is incorrect!



# Tree Thinking: Reading Trees

In the tree of life you often see a similar pattern depicted here, where you have a lineage with low number of species, sister to a large clade with many species. Often times the word, **basal** or **early diverging** will be used to describe this low - number of species lineage. This is incorrect!

Just because an extant lineage looks like a fossil (retains plesiomorphic characters) doesn't make it less evolved than other extant lineages.



# Tree Thinking: Reading Trees

Ginkgos, the beautiful gymnosperm whose leaves are turning a magnificent gold color right now, are sometimes referred to a “living fossils.” This is because a fossil of a “Gingko”-like fossil was discovered before the living plant was. Why is it, technically incorrect to call a living Ginkgo a “living fossil”?



# Tree Thinking: Reading Trees

Phylogenies can be visualized in many different ways, but they all have nodes, tips and branches



# Day 1: Outline

- General Concepts
  - What is phylogenetics
  - Evolutionary Theory
  - Homology
- Tree Thinking:
  - Reading Trees
- How to Infer a Phylogeny
  - Data
  - DNA alignment
  - Inference Methods
  - Parsimony
- Summary
  - Why should you care
  - Lab Introduction
  - Resources
  - Glossary of Terms

# How to infer a phylogeny: Data

Multiple Homology statements can be used to infer ancestry hypothesis.

**Every single column in a matrix (an alignment is a matrix), is a homology statement**

Molecular Data:  
(DNA, RNA, Amino Acid)

Japanese black pine (*Pinus thunbergii*)  
Bhutan white pine (*Pinus bhutanica*)  
Chiapas pine (*Pinus chiapensis*)  
Eastern white pine (*Pinus strobus*)  
Lacebark pine (*Pinus bungeana*)  
Red pine (*Pinus resinosa*)  
Single leaf pinyon (*Pinus monophylla*)

Genetic characters						
Sequence of portion of chloroplast DNA						
T	A	A	A	G	G	G
T	A	A	A	G	G	G
T	A	A	A	G	G	G
T	A	A	A	G	G	G
T	A	A	A	G	G	G
T	A	A	A	G	G	G
T	A	A	A	G	G	G

Morphological Data:  
(Morphological features)

	number of vascular bundles per needle	sheath around needle bundle (1=straight, 2=curling back)	number of needles per bundle	seed wing (0=absent, 1=detachable, 2=permanent)
Japanese black pine ( <i>Pinus thunbergii</i> )	2	1	2	2
Bhutan white pine ( <i>Pinus bhutanica</i> )	1	2	5	1
Chiapas pine ( <i>Pinus chiapensis</i> )	1	2	5	1
Eastern white pine ( <i>Pinus strobus</i> )	1	2	5	1
Lacebark pine ( <i>Pinus bungeana</i> )	1	2	3	2
Red pine ( <i>Pinus resinosa</i> )	2	1	2	2
Single leaf pinyon ( <i>Pinus monophylla</i> )	1	2	1	0

# How to infer a phylogeny: DNA alignment

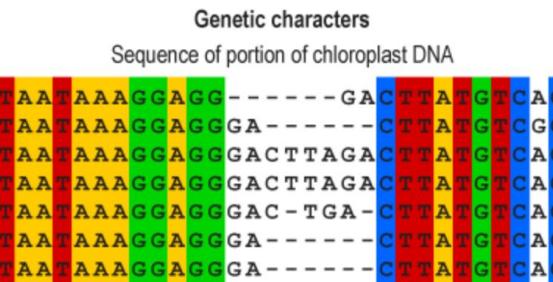
Many algorithms have been developed to align DNA

- Local Alignment - Attempt to align parts of sequence
- Global Alignment - Attempt to align sequence along its entire length

Algorithms use a set of rules to find the ‘best’ alignment. Importantly the best alignment does not entail the correct homology.

Molecular Data:  
(DNA, RNA, Amino Acid)

Japanese black pine (*Pinus thunbergii*)  
Bhutan white pine (*Pinus bhutanica*)  
Chiapas pine (*Pinus chiapensis*)  
Eastern white pine (*Pinus strobus*)  
Lacebark pine (*Pinus bungeana*)  
Red pine (*Pinus resinosa*)  
Single leaf pinyon (*Pinus monophylla*)



# How to infer a phylogeny: DNA alignment

Algorithms use a set of rules to find the ‘best’ alignment. Importantly the best alignment does not entail the correct homology.

**Question:** Which alignment is correct?

two substitutions  
one indel

**Sequence 1:** ATGCGTCGTT

||•|| |||•

**Sequence 2:** ATCCG-CGTC

no substitutions  
five indels

**Sequence 1:** AT--GCGTCGTT

|| ||||||

**Sequence 2:** ATCCGCGTC---

# How to infer a phylogeny: DNA alignment

Algorithms use a set of rules to find the ‘best’ alignment. Importantly the best alignment does not entail the correct homology.

**Question:** Which alignment is correct?

two substitutions  
one indel

**Sequence 1:** ATGCGTCGTT  
                  ||•||  |||•  
**Sequence 2:** ATCCG-CGTC

no substitutions  
five indels

**Sequence 1:** AT--GCGTCGTT  
                  ||  |||||  
**Sequence 2:** ATCCGCGTC---

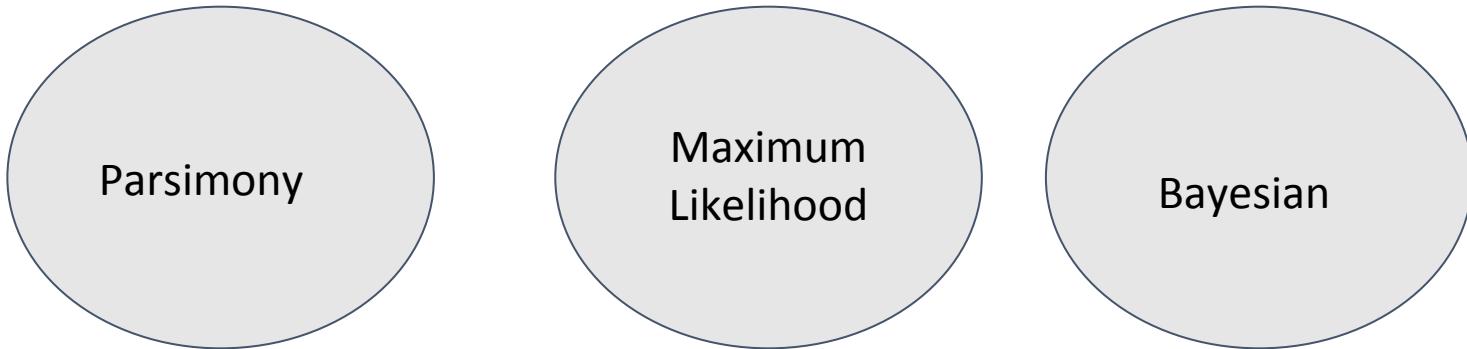
**Trick question (sorry):**

- 1) Both could be correct depending on algorithm.
- 2) Neither could be correct specifying the ‘true’ evolutionary relationships.

# How to infer a phylogeny: Inference Methods

- Phylogenetics aims to test **hypothesis of homologies** (matrix or alignment) by inferring a phylogeny.
- Phylogenetics use a **criterion** to determine which phylogeny fits the data best.

# How to infer a Phylogeny: Inference Methods



Parsimony

Maximum  
Likelihood

Bayesian

# How to infer a Phylogeny: Inference Methods

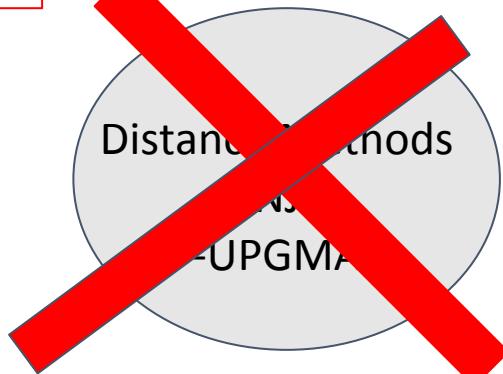
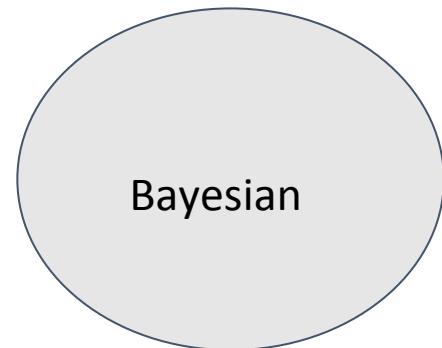
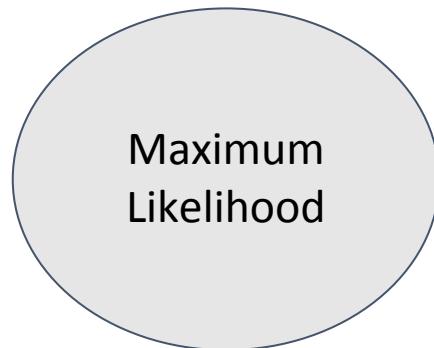
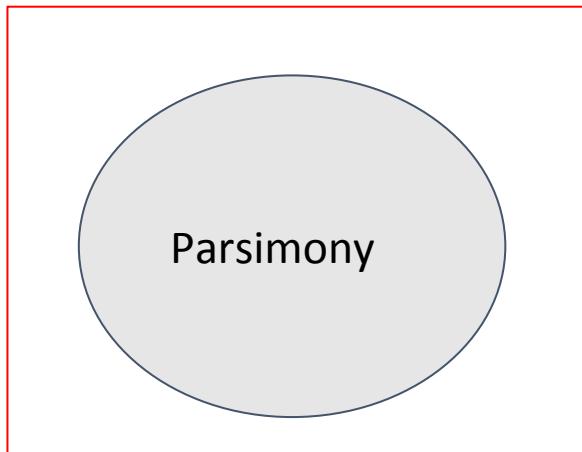
Parsimony

Maximum  
Likelihood

Bayesian

Distance Methods  
NJ  
UPGMA

# How to infer a Phylogeny: Inference Methods



# How to infer a Phylogeny: Parsimony

- Sometimes called **Occam's razor**
  - Loose Translation: “The simpler the better”
- Parsimony is used as a **criterion** to identify the best phylogeny
- Parsimony Programs attempt to find the tree(s) which require(s) the smallest number of character state changes to explain the observed data.

# How to infer Phylogenies: Parsimony

Searching through all possible tree space is a difficult problem as the number of possible trees grows with the number of tips.

4	15	enumerable by hand
5	105	enumerable by hand on a rainy day
6	945	enumerable by computer
7	10395	still searchable very quickly on computer
8	135135	a bit more than the number of hairs on your head
9	2027025	population of Houston, Texas
10	34459425	more than the number of combinations in the National Lottery
12	$13.7 \times 10^9$	probably the limit for searching by computers
20	$8.20 \times 10^{21}$	$\approx$ the number of stars in the observable universe
48	$3.21 \times 10^{70}$	$\approx$ the number of particles in the universe
698	$10^{5242}$	the number of influenza genomes I recently analysed

# How to infer a phylogeny: Summary

- Different types of data used to infer phylogenies
- How alignment algorithms make homology statements of DNA
- What Parsimony is and how it works.

# Day 1: Outline

- General Concepts
  - What is phylogenetics
  - Evolutionary Theory
  - Homology
- Tree Thinking:
  - Reading Trees
- How to Infer a Phylogeny
  - Data
  - DNA alignment
  - Inference Methods
  - Parsimony
- Summary
  - Why should you care
  - Lab Introduction
  - Resources
  - Glossary of Terms

# Summary: Why should you care?

- You can harness the power of evolution in your research. Tree thinking and evolutionary thinking provides a framework to make predictions about the natural world
- If your research focus on genomics you will inevitably make an alignment. Be cautious of what assumptions the algorithm is making because reviewers will be!
- If you are studying a particular gene, inferring a gene tree is a relatively simple way of to add a new dimension to your research (higher impact factor).
- So you don't look like a fool when reading paper.

# **Summary: Resource - Phylogenetics/ Computational Courses at Cornell**



Phylogeny of Green Plants  
Kevin Nixon

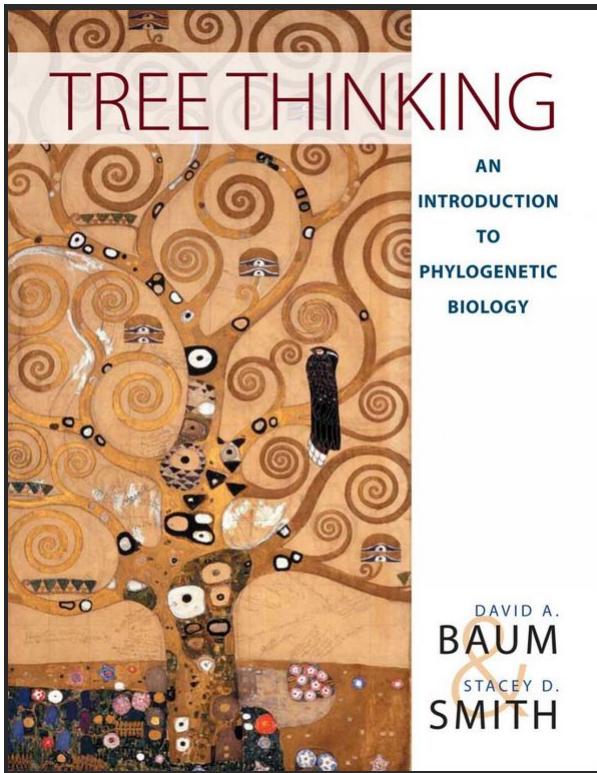


Model Based Phylogenetics  
Corrie Moreau



Computing for Plant Biologist  
Gaurav Moghe

# Summary: Resources - Tree Thinking



Baum, D.A. and Smith, S.D., 2013. *Tree thinking: an introduction to phylogenetic biology*. Greenwood Village (CO): Roberts.

# Summary: Resources - The Tree Room

[https://evolution.berkeley.edu/evolibrary/article/0\\_0\\_0/evotrees\\_intro](https://evolution.berkeley.edu/evolibrary/article/0_0_0/evotrees_intro)

 **Understanding Evolution**  
your one-stop source for information on evolution

EVOLUTION 101 TEACHING MATERIALS RESOURCE LIBRARY SUPPORT THIS PROJECT

glossary | home | Google Custom Search



**The Tree Room**

[THE TREE ROOM](#)  
[Primer on trees](#)  
[Tree misinterpretations](#)  
[Field guide to evolutionary trees](#)  
[How to build a tree](#)  
[Trees matter](#)  
[For teachers](#)  
[For museums and zoos](#)  
[Credits](#)

[print](#)

## The Tree Room

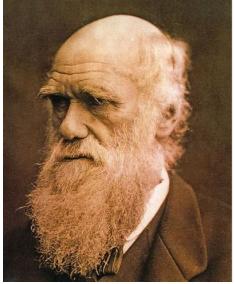
Over 150 years ago, Darwin chose the tree of life as a metaphor for one of the most powerful ideas in biology: the relatedness of all living things. The power of that idea can be seen today in the ubiquity of evolutionary trees (also called [phylogenetic trees](#)) in all biological disciplines, from studies of newly discovered [species](#) to cutting-edge cancer research. Correspondingly, evolutionary trees have made their way into textbooks, museum exhibits, and the media. To understand modern biology, we all need to understand how evolutionary trees can be read and used.

The Tree Room provides a wide variety of tools for teaching and learning about evolutionary trees in both classrooms and informal science education settings. You can explore the following sections:

- [Primer](#) — Learn the basics about understanding evolutionary trees. If you are not very familiar with trees, you'll want to start here.
- [Misinterpretations and intuitive ideas about evolutionary trees](#) — Find out how evolutionary trees are commonly misinterpreted.
- [Field guide to trees](#) — Confused by a complicated tree from a textbook or the Web? Use our interactive guide to help unpack it.
- [How to build a tree](#) — Learn about the methods that scientists use to reconstruct evolutionary history.
- [Trees matter](#) — Discover the many practical applications of evolutionary trees.
- [For teachers](#) — Get quick tips for incorporating trees into your classroom and browse a special selection of lessons for teaching about trees.
- [For museums and zoos](#) — Find design tips and other tools to help you effectively incorporate trees into your exhibit.

**The Tree Room**  
page 1 of 1  
[next >](#)

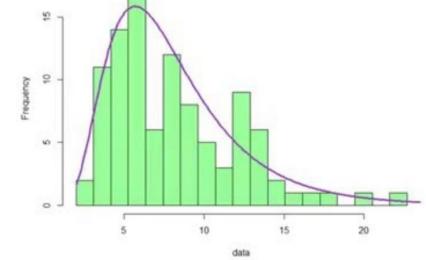
# Lab Introduction



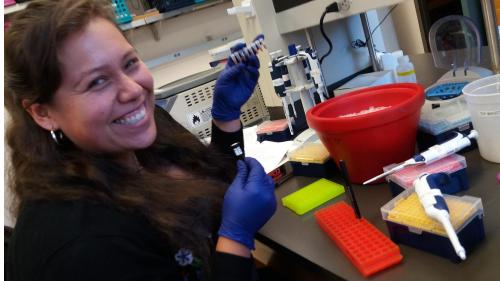
# Evolutionary Theory



# Computational Science



## Statistics



# Molecular Biology



Natural History

# Lab Introduction: File Format

- Fasta File (.fasta/.fas)
  - For sequence aligned or unaligned
- Newick File (.tree/.tre/.nwk)
  - For any phylogeny
- TNT formate (.tnt)
  - This is specific to the program TNT which we will use on day 1.
- Nexus File (.nexus/.nex)
  - For either sequences, phylogeny, other character matrix or all of the above

# **Important note on homework:**

- All homework due November 3rd.
- Please please please email Eugenio and I if you have questions, well set up a time to zoom. The sooner the better!
-

# Summary: Glossary

**Phylogenetics:** is the study of the evolutionary history and relationships among individuals or groups of organisms (e.g. species, or populations)

**Homology:** Similarity due to common ancestry

**Phylogeny:** A representation of the evolutionary relationships between groups.

**Monophyletic (clade):** An ancestor (node) and all its descendants.

**Synapomorphies:** A shared derived trait. They define clades.

**Sister Taxa:** Two groups that share a common ancestor. They can be individuals or clades

# Summary: Glossary

**Extant:** Living Species

**Extinct:** Dead Species

**Parsimony:** In evolutionary biology, a method to infer phylogenies by minimizing the number of changes