We are now recruiting Learning Assistants for Spring 2020!

Biology 212-Principles of Biology II

Section 1—MWF 3:10-4:00
Dr. Howell and Dr. Srivastava

Section 2,3 —TR 12:40-2:00
Dr. Manz and Dr. Sakaguchi

Section 4 – TR 2:10-3:30
Dr. Kukday

- Do you enjoy helping others learn?
- Are you good at explaining new ideas to others?
- Do you need to review fundamental concepts in biology before taking entrance exams?

LAs receive 2 credits of BIOL 491 or GEN 492.

These can count as 400-level courses towards your degree.

To express interest or ask questions, please fill out this form by Nov 24:
https://forms.gle/CkKgmAva2EsKtfjw8

COMPUTATIONAL ACTIVITY 6

November 21, 2019

Building Phylogenies: Exercise

Review what you remember about phylogenies to answer the worksheet questions and build trees with:

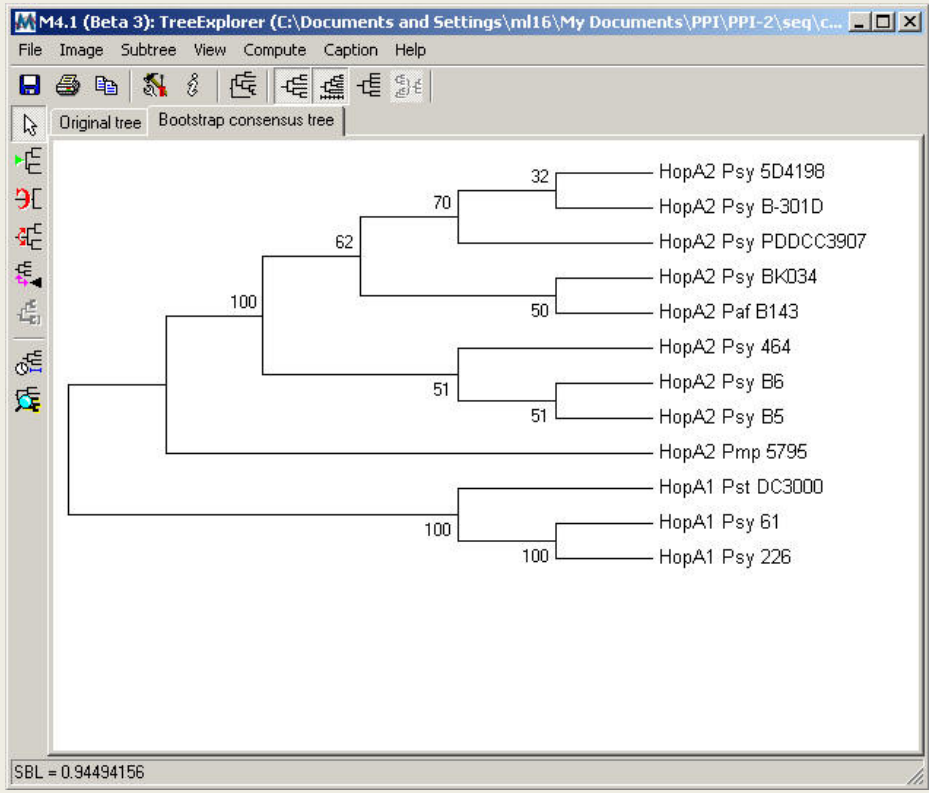
- Character/trait data
- Sequence data

WORKING WITH MEGA

MEGA 7

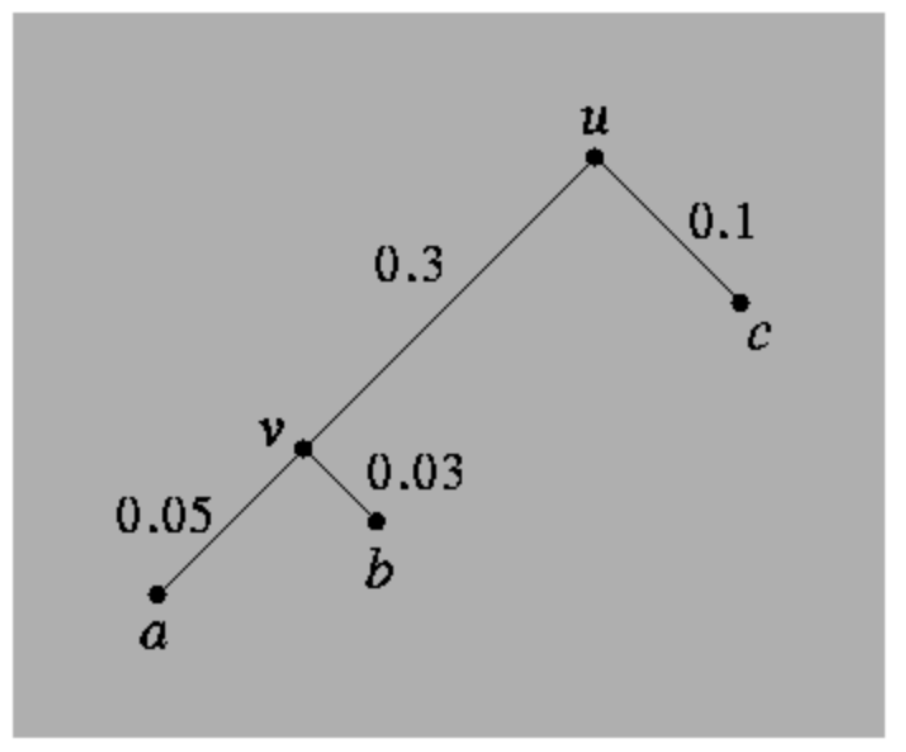
Molecular Evolutionary Genetics Analysis

- Phylogenetic software
  - *Distance and character based methods*
  - *No Bayesian methods*
- Data on GitHub
- Take a look at the sequences



- What organism are you working with?
- What does “Conserved Sites” mean?
- What does “Variable Sites” mean?
- What about “Parsim-Informative”?
- What sequence information will be helpful for building a tree?
- The statistics tab can tell you about the sequences. Look at the Nucleotide Composition. What is this telling you?

Pairwise Distances



- What sequences appear to be closer?
- Which appear to be different?

Building a Tree in MEGA

- Run a Maximum Likelihood Tree in the Phylogeny tab of MEGA.
- Which species are being called sister species?
- Now let’s run a neighbor-joining tree. What is different between the two trees?
- Try out the other methods available on MEGA. What can you conclude?

METHODS OF BUILDING PHYLOGENIES

Classified into two categories:

Distance methods: use some measure of evolutionary distance between pairs of OTUs and construct a tree based on those distances (do not use sequence data directly)

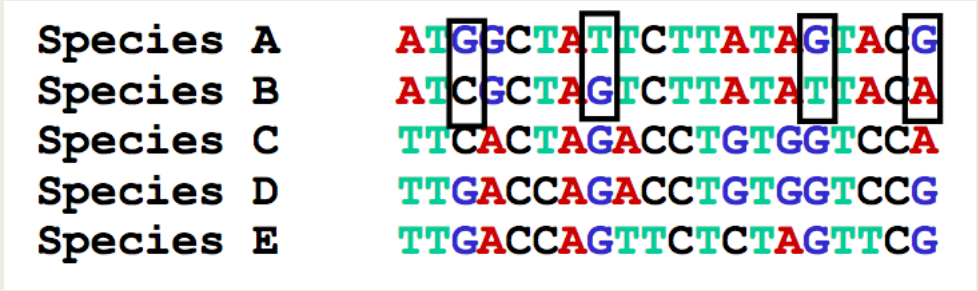
- UPGMA
- Neighbor-Joining

Character state methods: use the actual sequence data, where each site is considered a character with one of 4 possible states for DNA (A, C, G, T)

- Parsimony
- Maximum likelihood
- Bayesian

Distance-Based: Calculating Distances

- **Uncorrected Distance:** Count the changes between sequences and divide by the sequence length. Ex: Dist(A,B) = ?



  - *Tends to underestimate genetic distance*

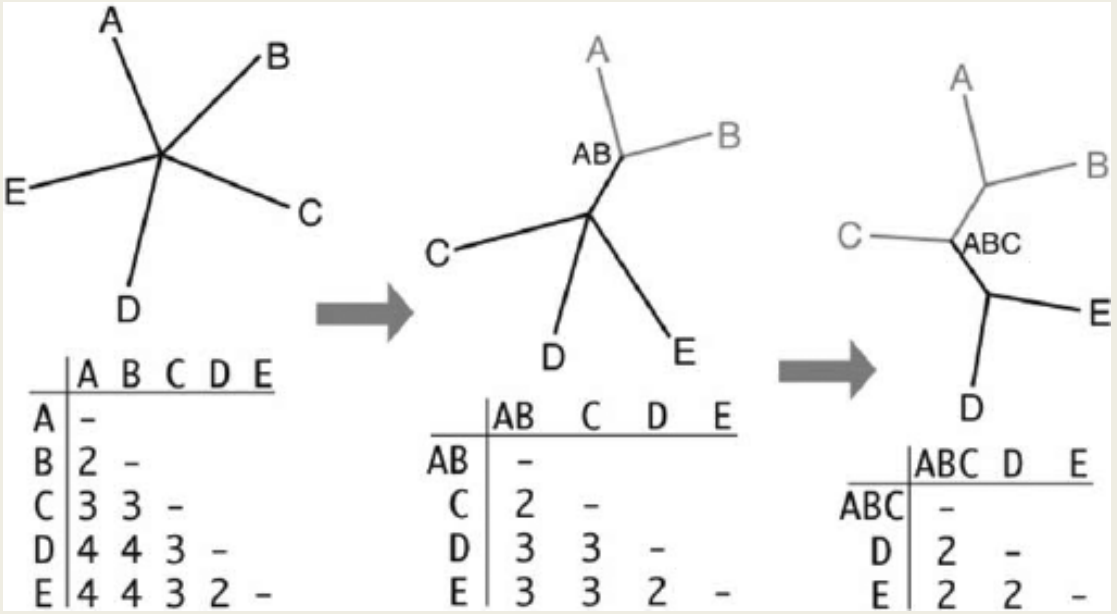
- **Jukes-Cantor Distance:** Allows the setting of the substitution rate. Assumes that any nucleotide change is equally likely and that all nucleotides occur at equal frequency.

$$K(A,B) = -3/4 \ln [1 - 4/3 D(A,B)]$$

Neighbor-Joining Method

1. Find the smallest distance using your chosen distance estimator.
2. Those two species will be pulled away from the star, as they are considered closely related.
3. Recalculate distances to the new group AB
4. Find the smallest distance again.
5. Pull out the species that now have the lowest distance.
6. Repeat

- Does not assume molecular clock

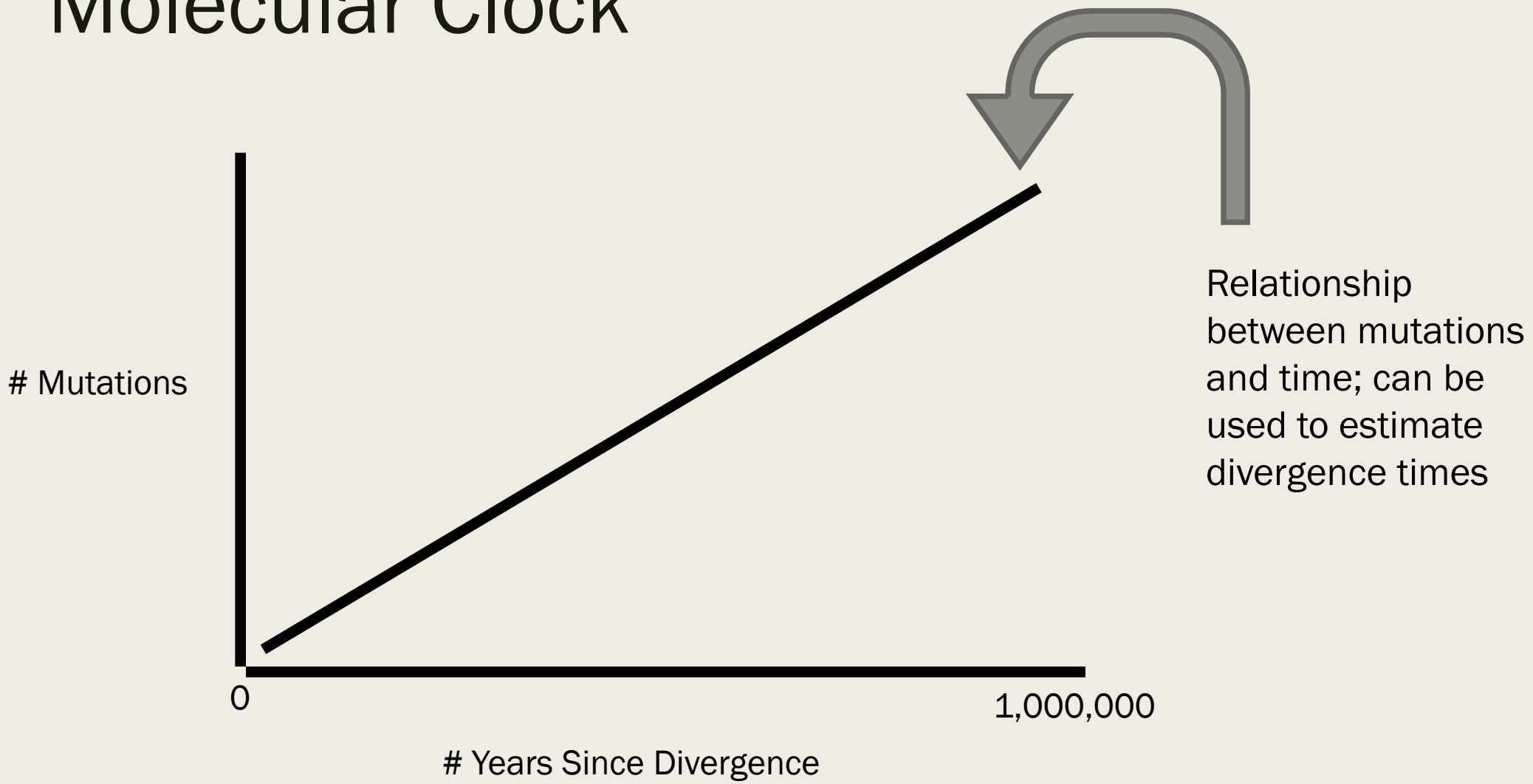


	A	B	C	D	E
A	-				
B	2	-			
C	3	3	-		
D	4	4	3	-	
E	4	4	3	2	-

	AB	C	D	E
AB	-			
C	2	-		
D	3	3	-	
E	3	3	2	-

	ABC	D	E
ABC	-		
D	2	-	
E	2	2	-

Molecular Clock



Relationship between mutations and time; can be used to estimate divergence times

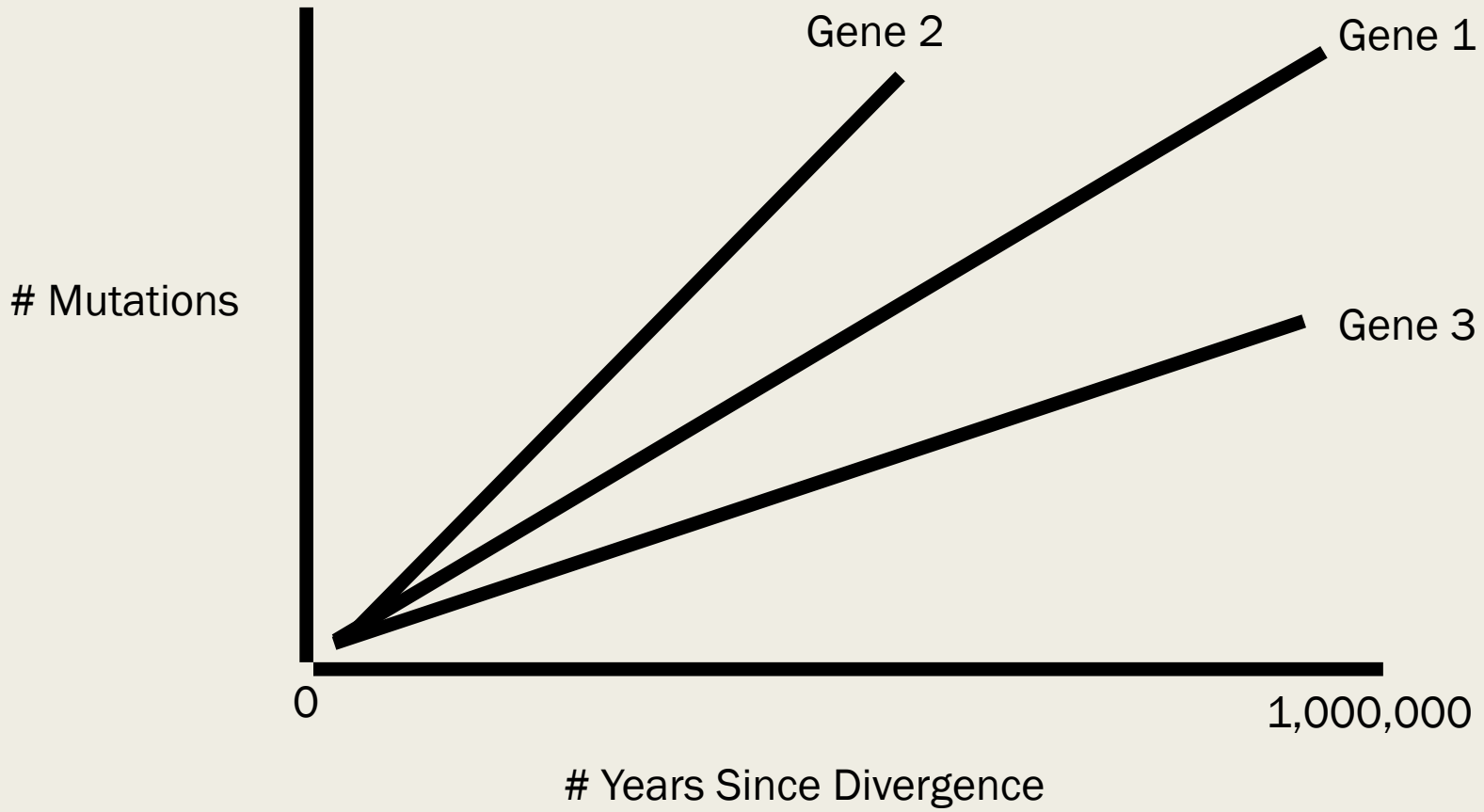
UPGMA: Unweighted Pair Group Method with Arithmetic mean

- Assumptions
  - *Constant substitution rate*
    - Over time
    - Over lineages

- Similar to Neighbor-Joining, but always produces a rooted tree

- Not a robust method, why?

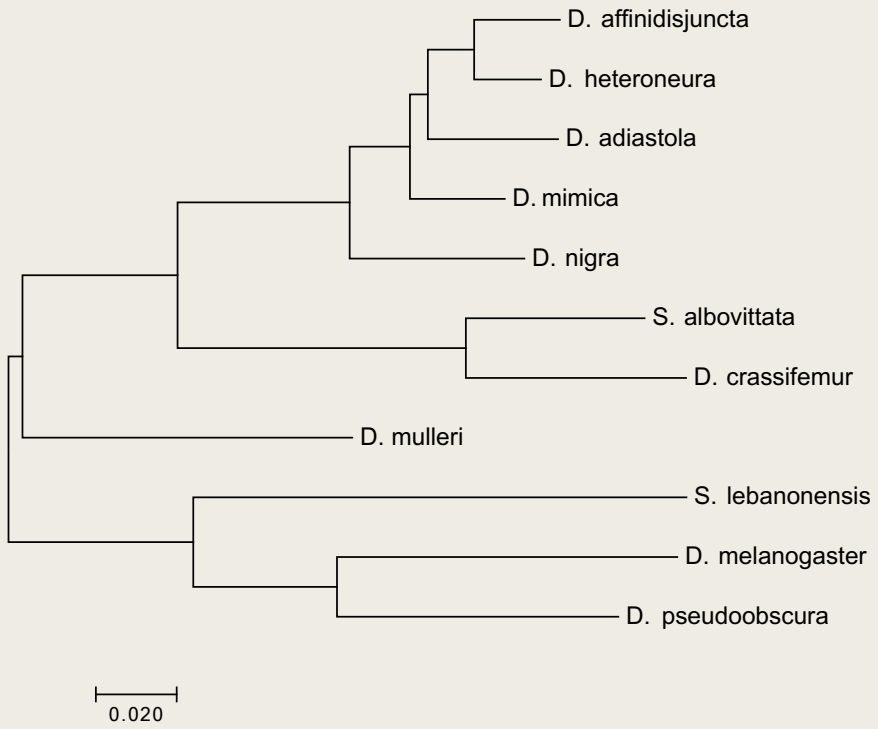
Molecular Clock: Problematic?



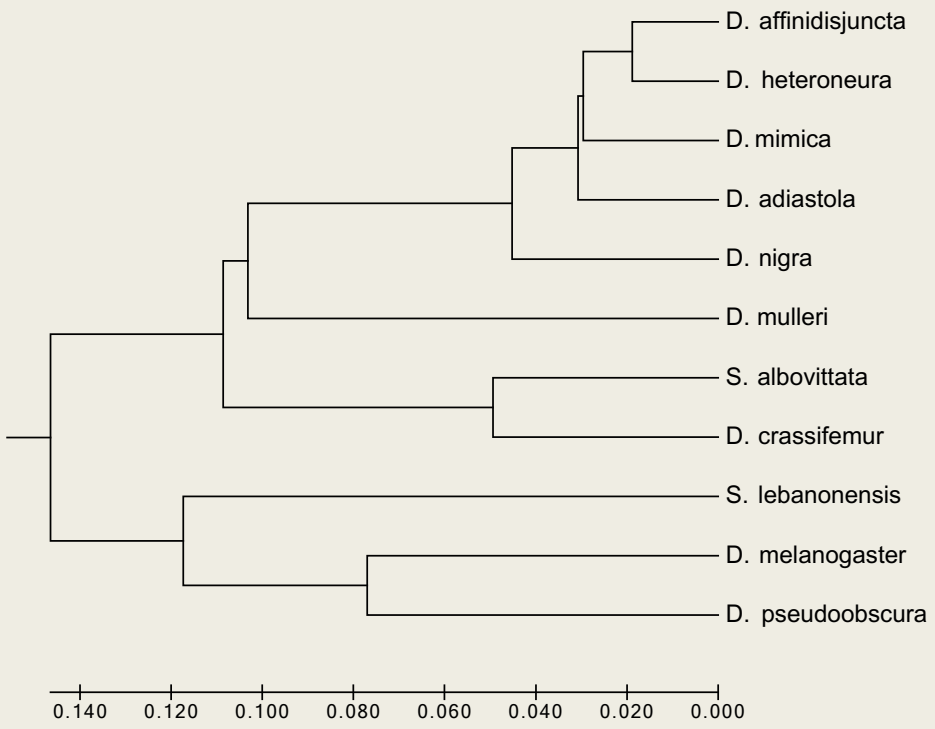
Relationship between mutations and time can be dependent on the genes used and the species included!

Which should we prefer?

NJ



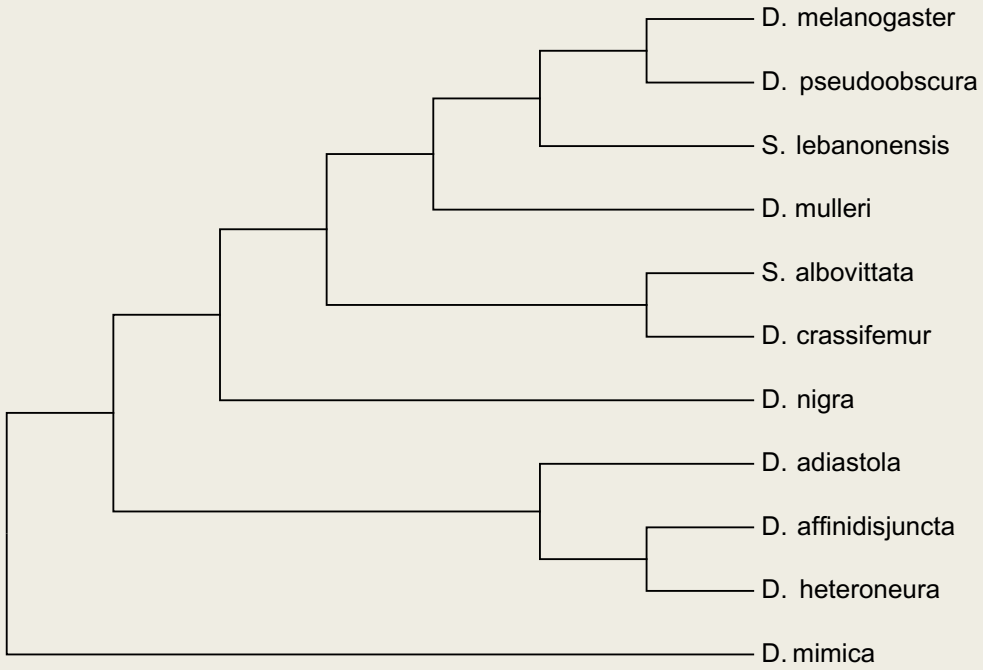
UPGMA



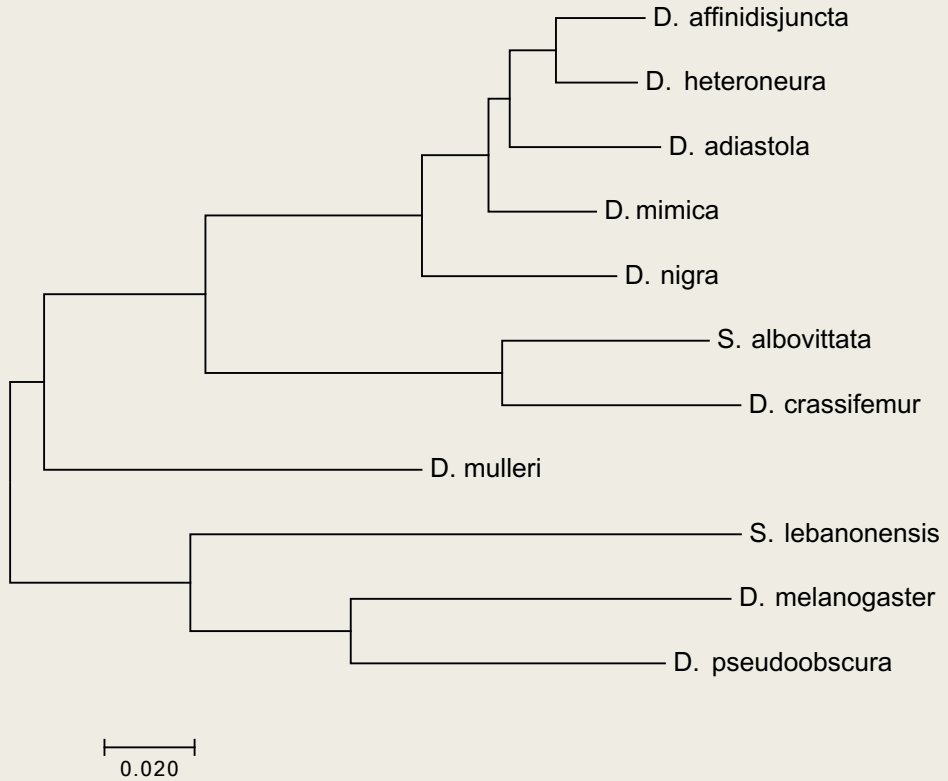
Character-Based

- **Parsimony Methods**
  - ***Maximum Parsimony:*** *(reviewed in class) Try to minimize branch lengths by minimizing the number of mutations between closely related sequences.*
  - ***Minimum Evolution:*** *Tries to fit the assumption that the tree with the smallest sum of branch length estimates is the true tree.*
- **Assumptions**
  - *No reversals*
  - *No convergence/parallel evolution*
  - *No homoplasy*

Max Parsimony

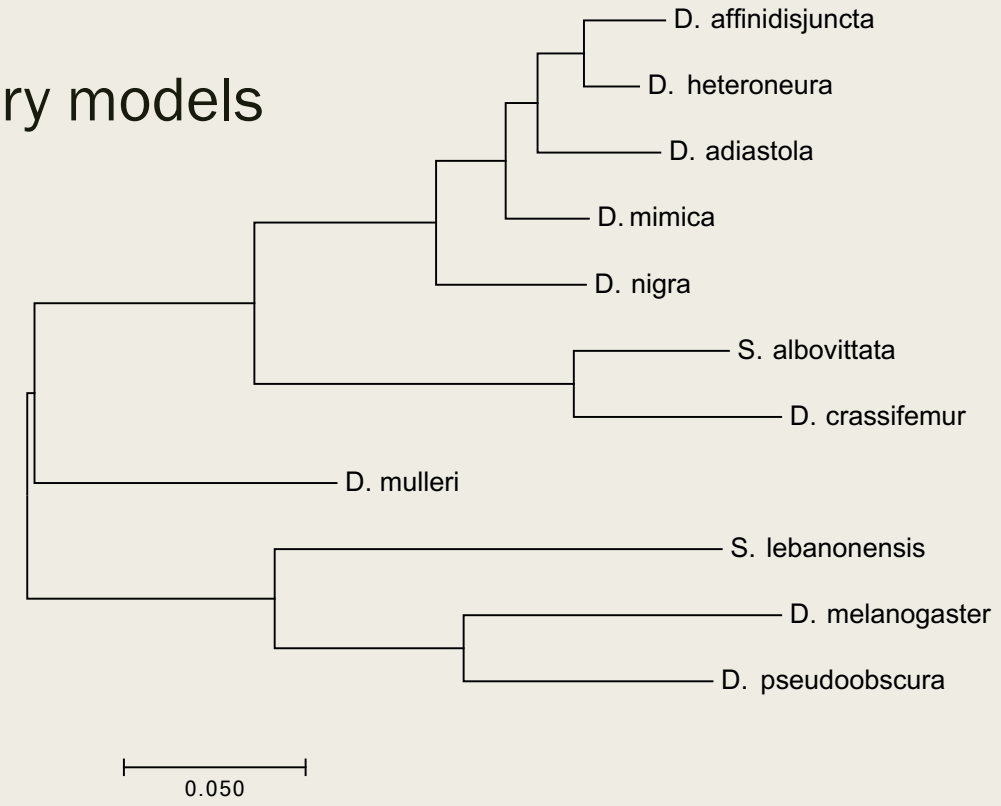


Min Evolution



Maximum Likelihood

- Choose the tree which makes the data the most “probable”; in other words, choose the tree which is best able to predict the data that you have
- Can use different evolutionary models
- Robust
- Computationally intensive

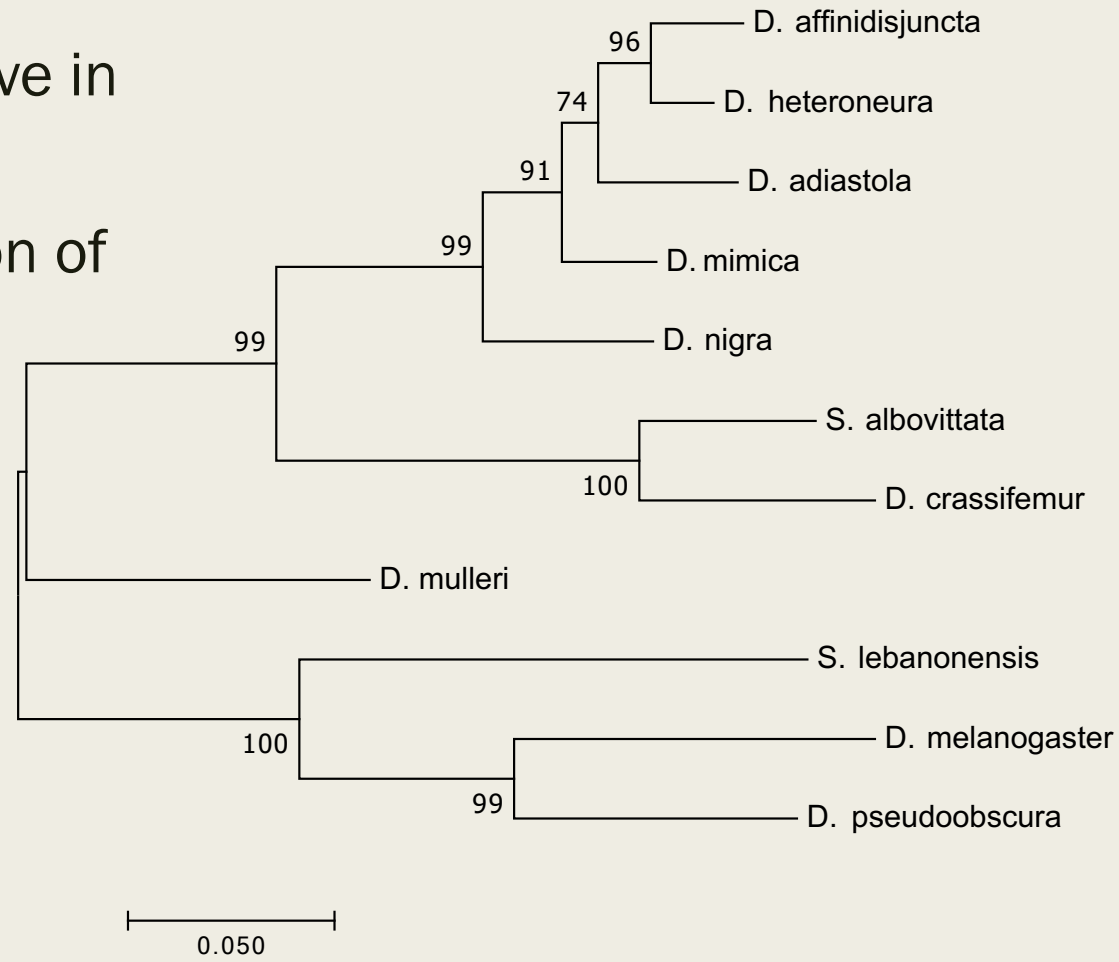


Hypothesis Testing Using Trees

- Form hypothesis of evolutionary relationships between species
  - *Trait data*
  - *Fossils*
  - *Biogeography*
- Collect genetic or trait data
- Choose an appropriate tree-building method
- Bootstrapping

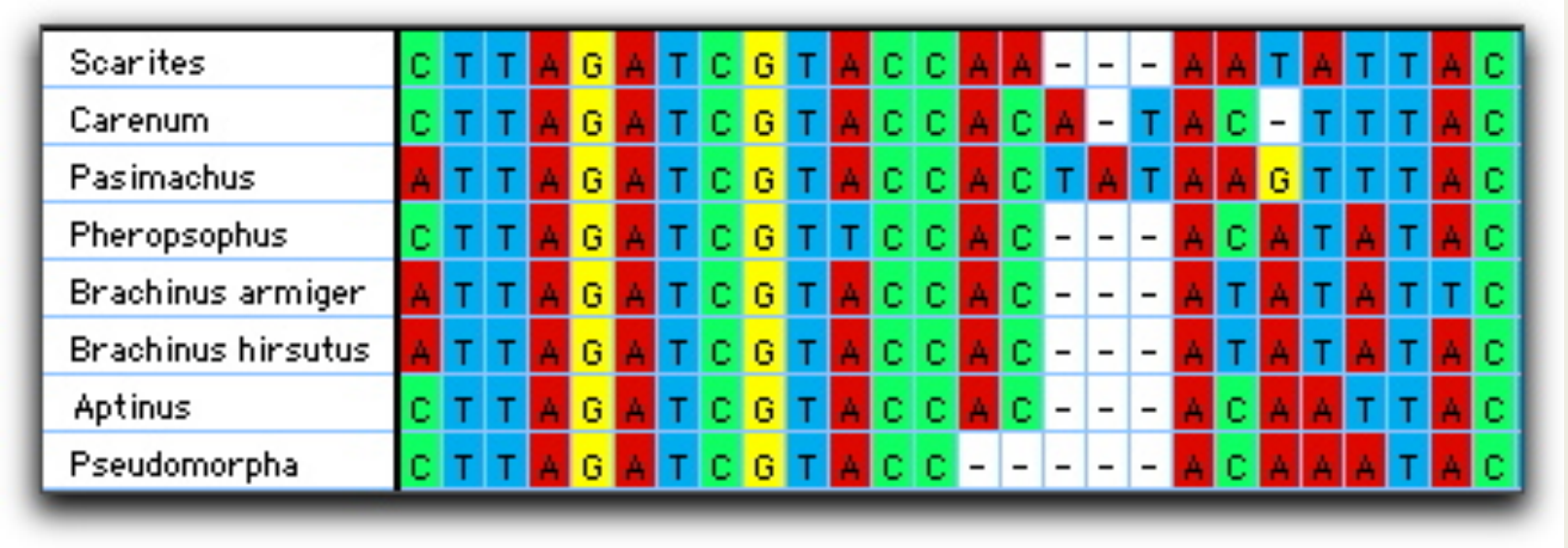
Bootstrapping

- How much confidence should we have in a certain tree?
- 100% ➔ high support for that portion of the tree
- Which trees are similar? Which are different?



Using Mega with GenBank: Alignment

- Alignment



Steps

1. Create new DNA alignment
2. Use the Web: GenBank option
3. Do a nucleotide search for “sloth cytochrome b”.
4. Click an entry and select Add to Alignment in the top right.
5. Add names and click OK.
6. Go back and now click the entry for Bradypus tridactylus
7. Select sequences and align
8. Export alignment in MEGA format