Laboratory 1

Create a Phylogeny

In lectures you heard a number of times about the vast diversity of life that has evolved over the past few billions years. We discussed life in its major domains and spent a some time on the major eukaryotic groups—plants and animals. We explained the idea of natural selection as the major force creating this diversity and mentioned Darwin and Wallace as the major contributors to this scientific hypothesis. We looked at the "family tree" of birds and hypothesized why, despite being overall similar, they have notably different beaks! They descended from a common ancestor, and show selection for beaks that helped them eat different prey, reproduce, and maintain this helpful variation. This is the explanation for evolution and for extinction.

Something to ponder: while the diversity of animals seems nearly endless, today the Earth is experiencing the greatest mass extinction in its long history. During this lab period, **20 species will probably go extinct**, mostly due to habitat loss and degradation of the environment. Most of these are considered small and unimportant by many people. However, I do not know what is "unimportant" and why our species is more important then others...

1.1 Background

Phylogeny is the genealogy (i.e., "family tree") of organism. In other words, the phylogeny represents the ancestor-descendant relationships. The inference of phylogeny is one of the foci of evolutionary biologists. It is also one of the most difficult tasks that these scientists undertake. Because one can never replay the "tape of history", one can never "know" the true phylogeny. Species are subject to extinction and parallel evolution, and these phenomena obscure phylogeny. At best, a "phylogeny" is really a "phylogenetic hypothesis." In no other field is appreciation of the strengths and weaknesses of the scientific method more appropriate.

The basis for inferring phylogeny is the **synapomorphy** = *a shared derived characteristic.* For example, the presence of a cranium (= skull) is a synapomorphy for the *Vertebrata.* No other animals have a skull. Thus, this synapomorphy supports a phylogenetic hypothesis that states that all vertebrates (fishes, mammals, amphibians, etc.) are more closely related to each other than to any other groups of animals. For example, fish and frogs are more closely related to each other than they are to crayfish.

To infer a synapomorphy, one must have some idea what the ancestral (**plesiomorphic**) state of that character is. For example, the ancestral amphibian had lungs. Because there is a large group of salamanders that do not have lungs, we call "lunglessness" a synapomorphy that unites all lungless salamanders into a single family. Lunglessness is a *derived condition* that all these species of salamanders share.

Using synapomorphies, one can place species on a "family tree" or phylogeny. Because traits are subject to natural selection, it is uncommon to find concrete synapomorphies that define 100% of the group.

Terrestrial vertebrates are called *Tetrapoda* in reference to a significant synapomorphy: presence of four limbs. However, snakes are a member of the *Tetrapoda* even though they lack external evidence of legs. Yet there are numerous other characteristics that clearly make snakes reptiles, thus members of the *Tetrapoda.* Use of only one characteristic (presence or absence of four limbs) would lead one to an incorrect phylogeny that did not include the snakes with the reptiles. Use of as many characters as possible is critical to developing a rigorous phylogenetic hypothesis.

Loss of legs (synapomorphy of tetrapods) in the snakes is an example of what is referred to as *evolutionary reversal.* The best phylogenetic hypothesis is one in which the number of evolutionary events (achievements, reversals *etc.*) is minimal. This is known as the **principle of parsimony**. There is a consensus among evolutionary biologists that the most parsimonious phylogenetic hypothesis is the one that is most likely to represent the true genealogy of a group of organisms.

* * *

Most organisms have common names, such as the "red maple" or the "red-winged blackbird". However, these common names are often misleading. Many different species are called the same thing in different parts of world, and many identical species are called different names. Formal Latin names are used by scientists to establish a unique name for each species on the Earth: "red maple" is *Acer rubrum* whereas "red-winged blackbird" is *Agelaius phoeniceus.* Each Latin name is made, approved and used by scientists worldwide.

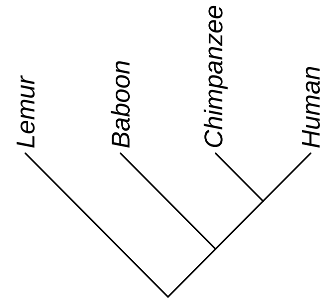
Every species name consists of two parts: the first part is the **generic name** (or genus, e.g., *Homo*); the second part is the **specific epithet** (e.g., *sapiens.*) This *Linnaean binomial system of nomenclature* was introduced by Carolus Linnaeus in the XVIII century and has been in use ever since.

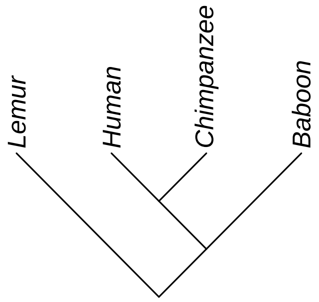
The study and practice of biological classification is known as **taxonomy**. Species and genera are **taxonomic groups** (taxa). Taxonomic groups have *ranks*: they may be species, genera, families, orders, classes, phyla, kingdoms or stay between them like subclass or superfamily.

Biologists would like taxonomy to reflect phylogeny. For example, all the frogs in *Ranidae* are hypothesized to be more closely related to one another than to any frogs from other groups. However, because phylogeny is difficult to infer, taxonomy is always changing. As scientists' opinions of phylogeny change, so does taxonomy. As more information is gathered, phylogenetic hypotheses may change; this often results in a change in taxonomy. This is the reason that textbooks often present different taxonomies. This is also the reason that the taxonomy presented in any textbook will not be the same one that is found in textbooks in 20 years.

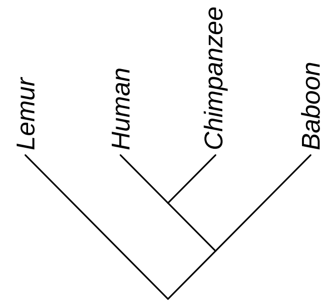
1.2 Understanding phylogenetic trees

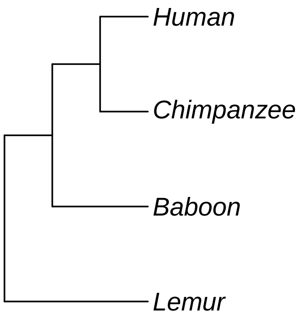
1. Phylogenetic tree is a *dichotomously branching diagram* where all terminals (tips representing objects) are connected. Phylogenetic tree is *related with time*; start of the diagram is always older then end. All terminals must be labeled. Nodes which have more than two branches are typically not allowed.
2. Tree edges may be freely rotated in any direction. For example, these trees are same:



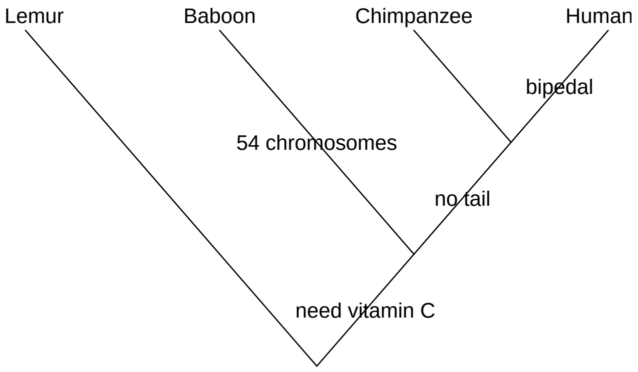


3. Direction of branches also does not matter. These trees are same:





4. Apomorphies (and reversals) could be shown as edge labels:



5. Phylogenetic tree is directly related with taxonomy (classification):
   - Suborder Strepsirrhini: lemur
   - Suborder Haplorrhini:
     - Superfamily Cercopithecoidea: baboon
     - Superfamily Hominoidea:
       - Genus *Pan*: chimpanzee
       - Genus *Homo*: human

1.3 Lets Practice! How to make a phylogeny tree

The instructor will explain phylogeny method on any appropriate example (Sesame Street creatures, random stuff from a classroom) but the following instruction will explain it on the example of imaginary planet:

Planet Aqua is entirely covered with shallow water. The ocean is inhabited with various flat organisms (see the figure). These creatures (let us call them "kubricks") can photosynthesize and/or eat other organisms or their parts (which match with their mouths), and move (only if they have no stalks). We need to make a phylogeny of kubrick species A–G because kubrick H is an **outgroup** and by definifion, has only plesiomorphic (ancestral) characters.

1. Start with the group you need to classify, assign the name for every object (member of the group).

   (This is already done, see above.)

2. Find as many characters as possible. Working number of characters is $N \geq n+1$ where n is the number of objects. Therefore, in this example we need $N \geq 8$. Review all characters and make the *character list*.

   Note that plesiomorphies and *autapomorphies* (unique derived characters) do not help to make a tree so **concentrate on finding synapomorphies**, derived characters which unite two or more (but not all) objects.

   For every character, guess its plesiomorphic (ancestral) and apomorphic (derived) states, either (1) with the help of outgroup, or (2) using the other methods like comparative anatomy, paleontology or embryology. Encode plesiomorphic character state as "0", and apomorphic as "1":

   Character list:

   1. Mouth(s) presents 1, no mouth 0

   2. One mouth 1, not 0

   3. Triangle mouth 1, not 0

   4. Square mouth 1, not 0

   5. Stalk 1, no stalk 0

   6. No photosynthesis 1, photosynthesis 0

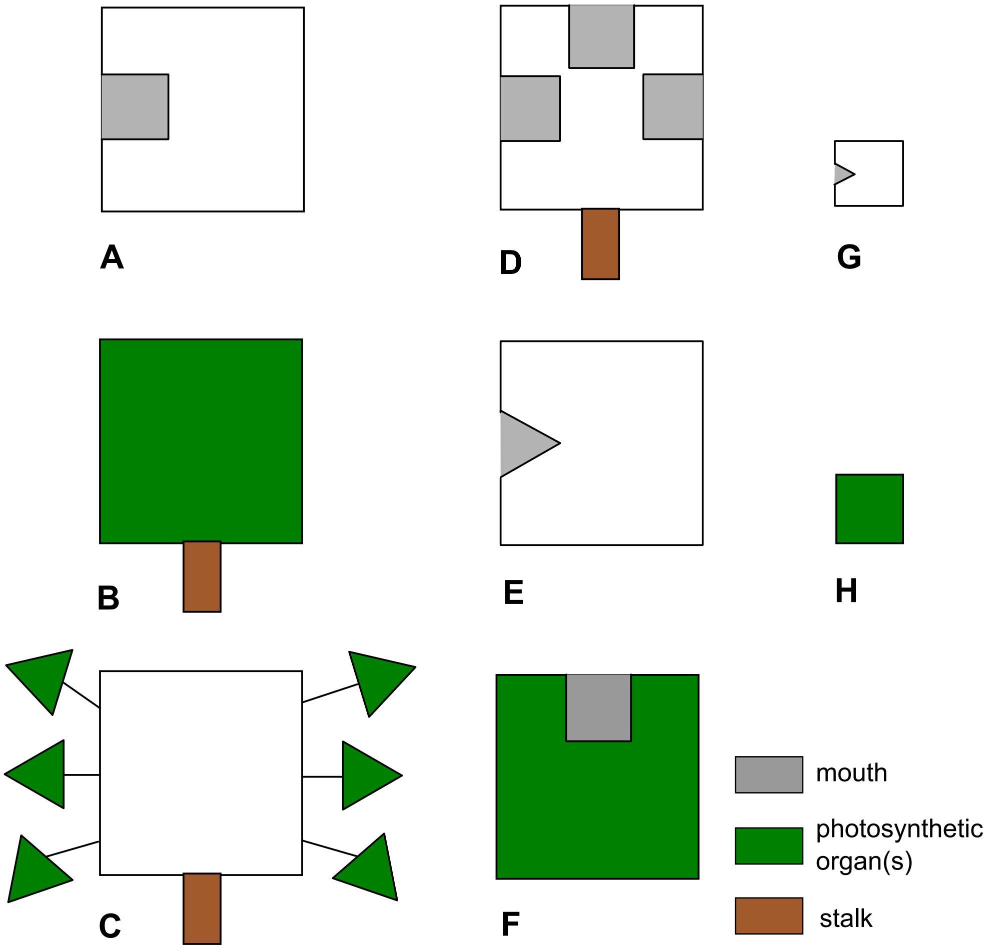


Figure 1.1: Kubricks.

7. Body big 1, body small 0

8. Mouth on top 1, no mouth on top 0

Note character #6: since outgroup *is* photosynthetic, we must make the absence of photosynthesis apomorphic state.

Make the **Character Table** where rows are names of objects, columns are characters, and each cell contains 0 or 1:

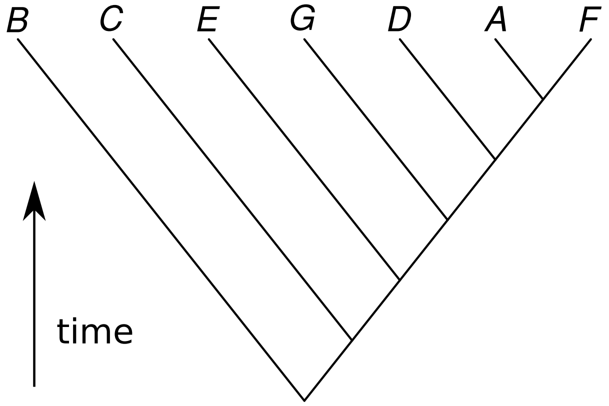
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Kubrick A	1	1	0	1	0	1	1	0
Kubrick B	0	0	0	0	1	0	1	0
Kubrick C	0	0	0	0	1	0	1	0
Kubrick D	1	0	0	1	1	1	1	1
Kubrick E	1	1	1	0	0	1	1	0
Kubrick F	1	1	0	1	0	0	1	1
Kubrick G	1	1	1	0	0	1	0	0

Without such table, we cannot calculate the length of our future tree (see below).

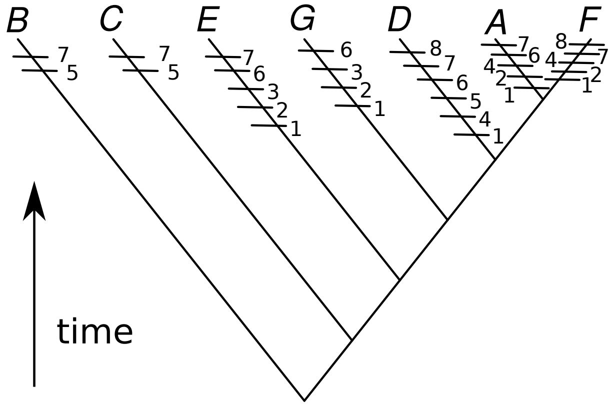
Note that *outgroup is not here*; this is because we already polarized our characters and therefore do not need outgroup anymore.

3. Start to draw the tree, preferably from the most ancestral ("primitive") member (which has most zeroes), this allows to pinpoint the **root** of our tree.

   Most ancestral member is the first branch. Continue to grow the tree attaching branches, preferably (but not necessarily) placing most similar objects closer. Actually, you may even attach members at random because this tree will be optimized anyway. All terminal branches should end with objects.

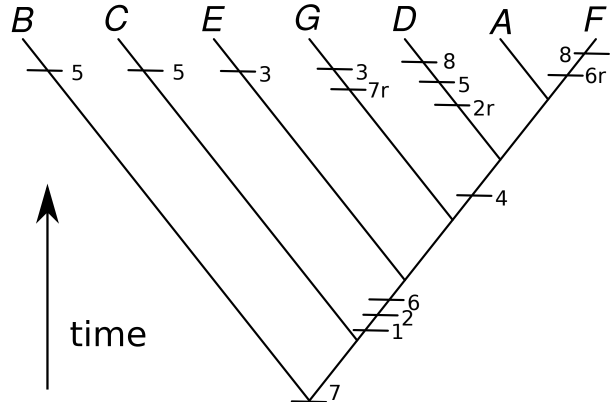


4. Look at the complete tree and label all apomorphies from the character table as tickmarks with labels. Now calculate the length of tree which is a **number of evolutionary events**, or simply number of tickmarks.



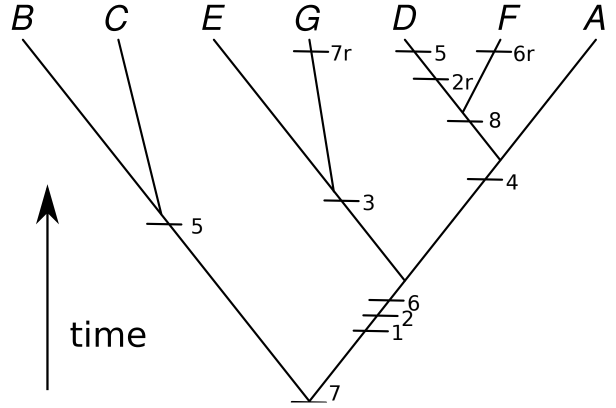
   Length of tree is simply the number of tickmarks, so $L = 29$ (this is exactly how many "ones" are in the table). This is a bit too much since good tree should have length $L < 2N$ where N is the number of characters. In our case, $L < 16$ is strongly recommended.

5. You will now likely see ways to improve your tree! Start to think how to make the tree shorter. First, decrease length without altering tree. You can shift characters down, deeper in time. Then all descendants from this point should have this character. If some do not have it, you may introduce *reversal*, the loss of character. Label reversals with "r" letter, like "2r":



Length of this tree is the number of all tickmarks so $L = 15$ now.

6. Then start to alter the actual tree. The basic method is to prune (take off) any branch (except root) and attach it to another place, re-label tree and calculate the length again. If the length is shorter, this is *more parsimonious* tree. Try to find the *most parsimonious* (shortest) tree, i.e. the tree with as low number of tickmarks as possible. Do not forget that tree branches may rotate freely.



Now $L = 12$! This is likely the *most parsimonious tree.* If you find the shorter one, you will receive extra points!

1.4 Procedure—More practice!

1. We will work in small teams today.
2. Every team should (1) choose 7 *different* plastic dinosaur plastic toys, (2) assign a name for each, (3) make character table, (4) make initial tree, (5) label apomorphies and reversals (if any); (6) calculate its length, and (7) reorganize the tree to make your phylogeny **as parsimonious as possible**.
   (a) To name dinosaurs, use shortcut of their scientific name (e.g., "Drom" for Dromeosauridae). Check the image below. We have:

Pteranodontidae (dinosaur "bats")

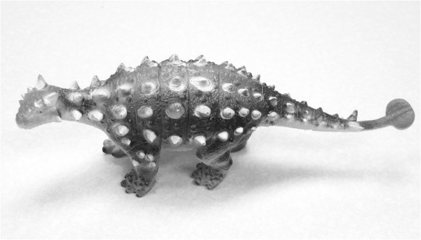
Dromaeosauridae ("raptors")

Tyrannosauridae and relatives

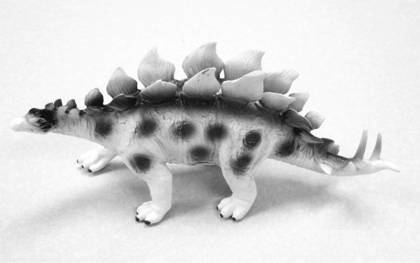
Diplodocidae ("titans")

Hadrosauridae ("duck-billed")

Pachycephalosauria ("skull-domed")



Ankylosauria ("club-tailed")



Stegosauria (with osteoderms)

Ceratopsidae ("frilled")

Disclaimer. Not all toys are imaged here. These toys are not supposed to serve as proper reconstructions of dinosaur groups mentioned above.

   (b) To polarize character states (i.e. decide which of them are plesiomorphic and which are apomorphic), you may use either the outgroup like imaginary "turtle" or "salamander", they both should work for our dinosaurs today. Alternatively, you can employ comparative and paleontology arguments (developmental arguments are not applicable here.)

   You must state clearly which approach you took and why you call some characters apomorphic (derived, "1") and some plesiomorphic (ancestral, "0"). Avoid terms "advanced" and "primitive", phylogeny does not work with these!

Lab 1 report

Your name ____

1. Write your dinosaur *character list* and draw the *character table* below. How did you determine which character state is plesiomorphic and which is apomorphic? Explain. (5 pts).

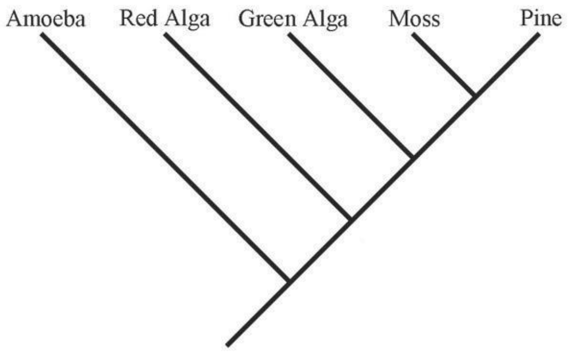
   Then, draw **two** of your dinosaur phylogeny trees (if there is no space, use reverse side or add one page). Do not forget to label all apomorphies and reversals (if any). Note the length of trees. Circle the **most parsimonious** tree. Why is this tree most parsimonious? Explain. (5 pts).

2. On next pages, **answer basic three thinking assignment** (5 pts, ½ point per answer).

<u>Basic Tree Thinking Assessment</u>

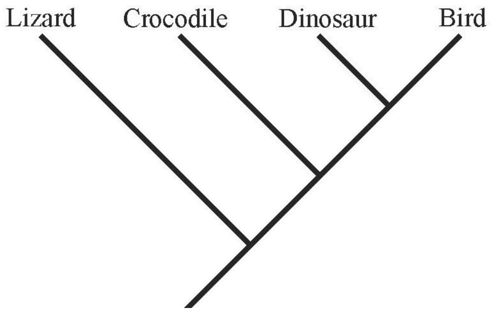
David A. Baum, Stacey DeWitt Smith, Samuel S. Donovan

This quiz includes a number of multiple-choice questions you can use to test yourself on your ability to accurately interpret evolutionary trees. Insofar as real biological examples have been used they are accurate based on current knowledge.



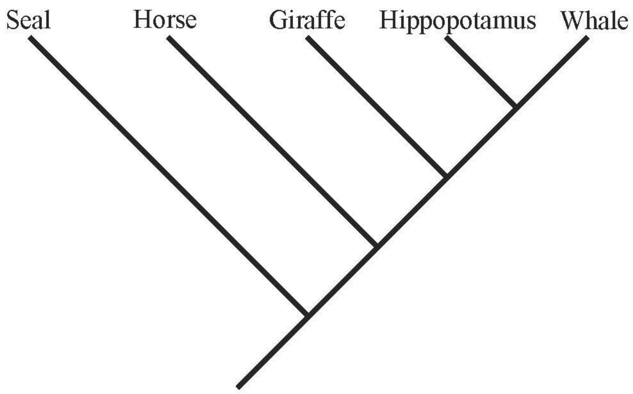
1) By reference to the tree above, which of the following is an accurate statement of relationships?

a) A green alga is more closely related to a red alga than to a moss
b) A green alga is more closely related to a moss than to a red alga
c) A green alga is equally related to a red alga and a moss
d) A green alga is related to a red alga, but is not related to a moss



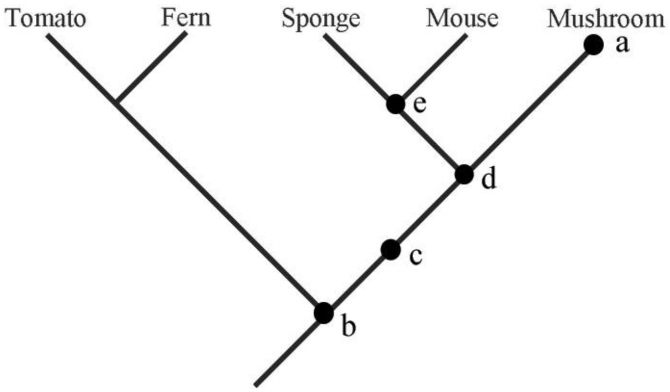
2) By reference to the tree above, which of the following is an accurate statement of relationships?

a) A crocodile is more closely related to a lizard than to a bird
b) A crocodile is more closely related to a bird than to a lizard
c) A crocodile is equally related to a lizard and a bird
d) A crocodile is related to a lizard, but is not related to a bird

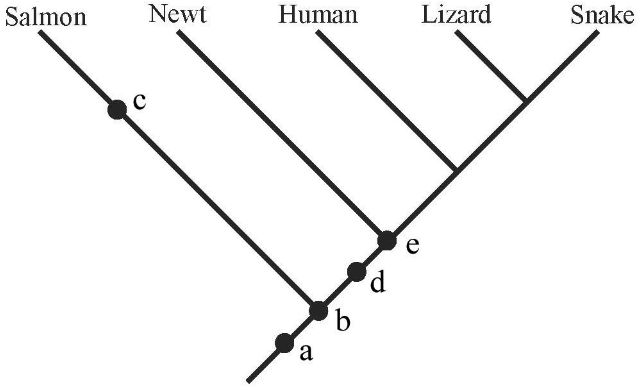


3) By reference to the tree above, which of the following is an accurate statement of relationships?

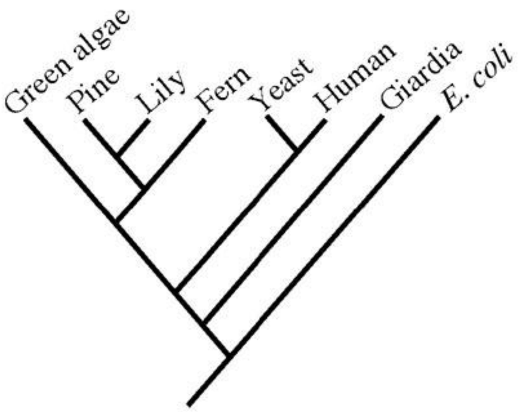
a) A seal is more closely related to a horse than to a whale
b) A seal is more closely related to a whale than to a horse
c) A seal is equally related to a horse and a whale
d) A seal is related to a whale, but is not related to a horse



4) Which of the five marks in the tree above corresponds to the most recent common ancestor of a mushroom and a sponge?

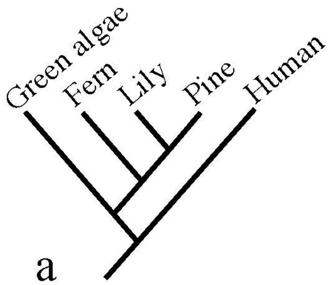


5) If you were to add a trout to the phylogeny shown above, where would its lineage attach to the rest of the tree?

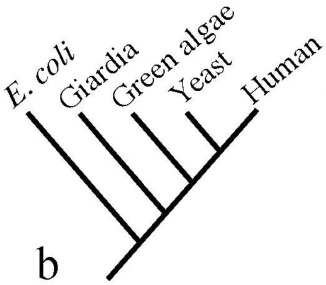


6) Which of trees below is false given the larger phylogeny above?

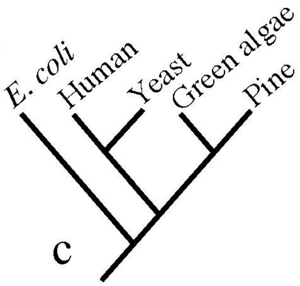
a



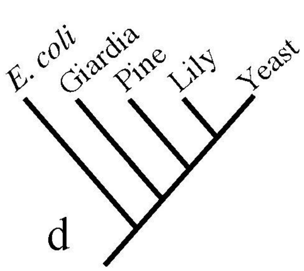
b

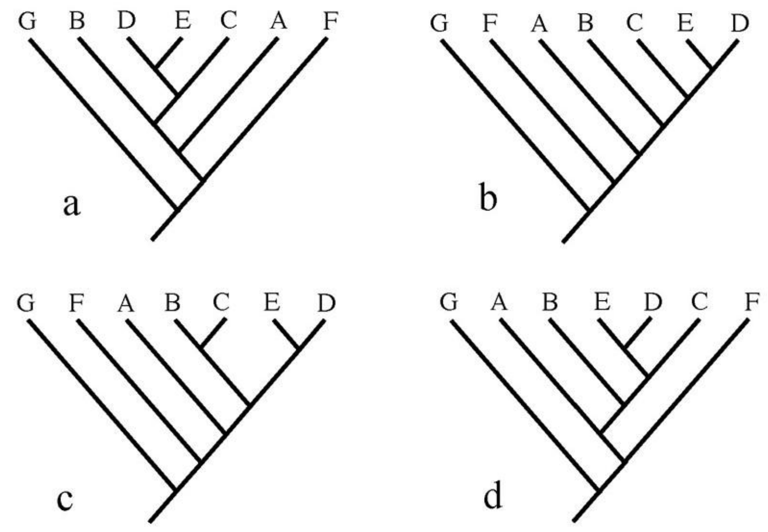


c

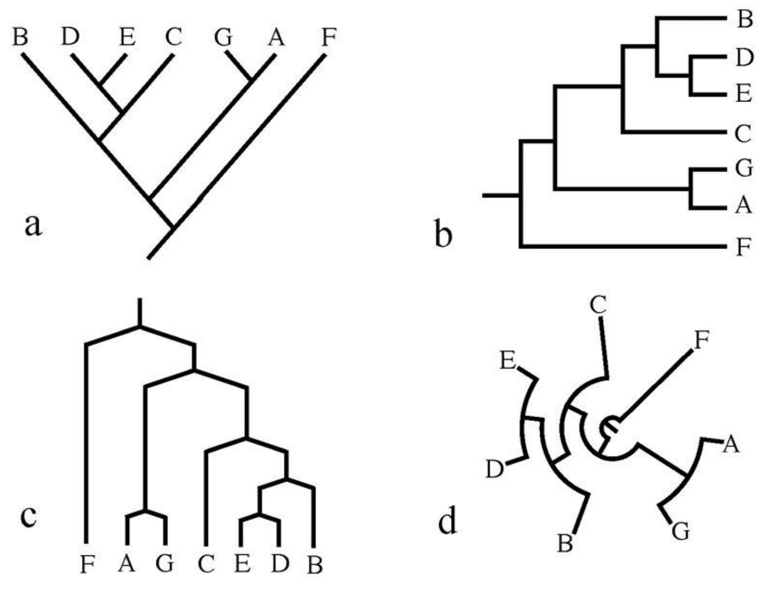


d

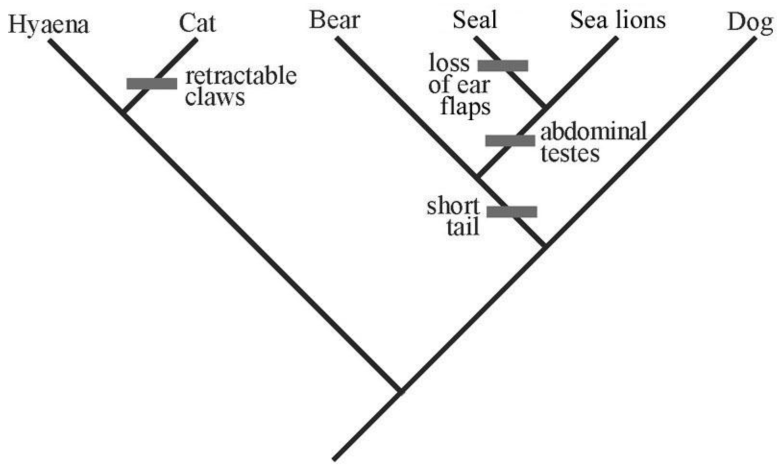




7) Which of the four trees above depicts a different pattern of relationships than the others?

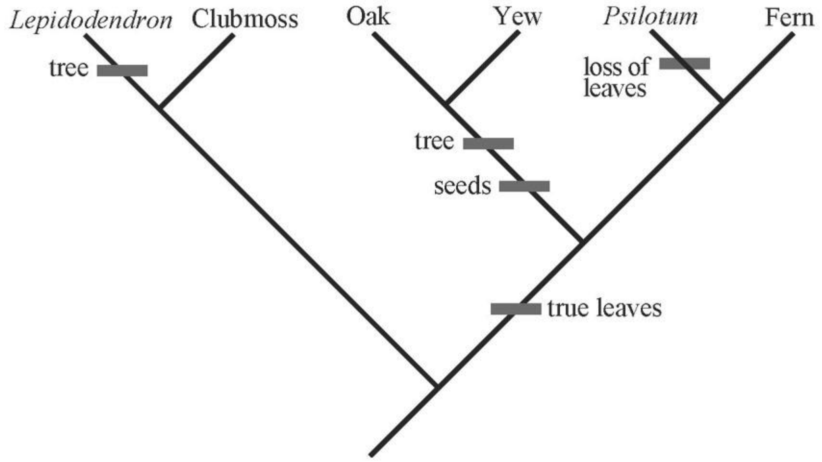


8) Which of the four trees above depicts a different pattern of relationships than the others?



9) In the above tree, assume that the ancestor had a long tail, ear flaps, external testes, and fixed claws. Based on the tree and assuming that all evolutionary changes in these traits are shown, what traits does a sea lion have?

- a) long tail, ear flaps, external testes, and fixed claws
- b) short tail, no ear flaps, external testes, and fixed claws
- c) short tail, no ear flaps, abdominal testes, and fixed claws
- d) short tail, ear flaps, abdominal testes, and fixed claws
- e) long tail, ear flaps, abdominal testes, and retractable claws



10) In the above tree, assume that the ancestor was a herb (not a tree) without leaves or seeds. Based on the tree and assuming that all evolutionary changes in these traits are shown, which of the tips has a tree habit and lacks true leaves?

- a) *Lepidodendron*
- b) Clubmoss
- c) Oak
- d) *Psilotum*
- e) Fern

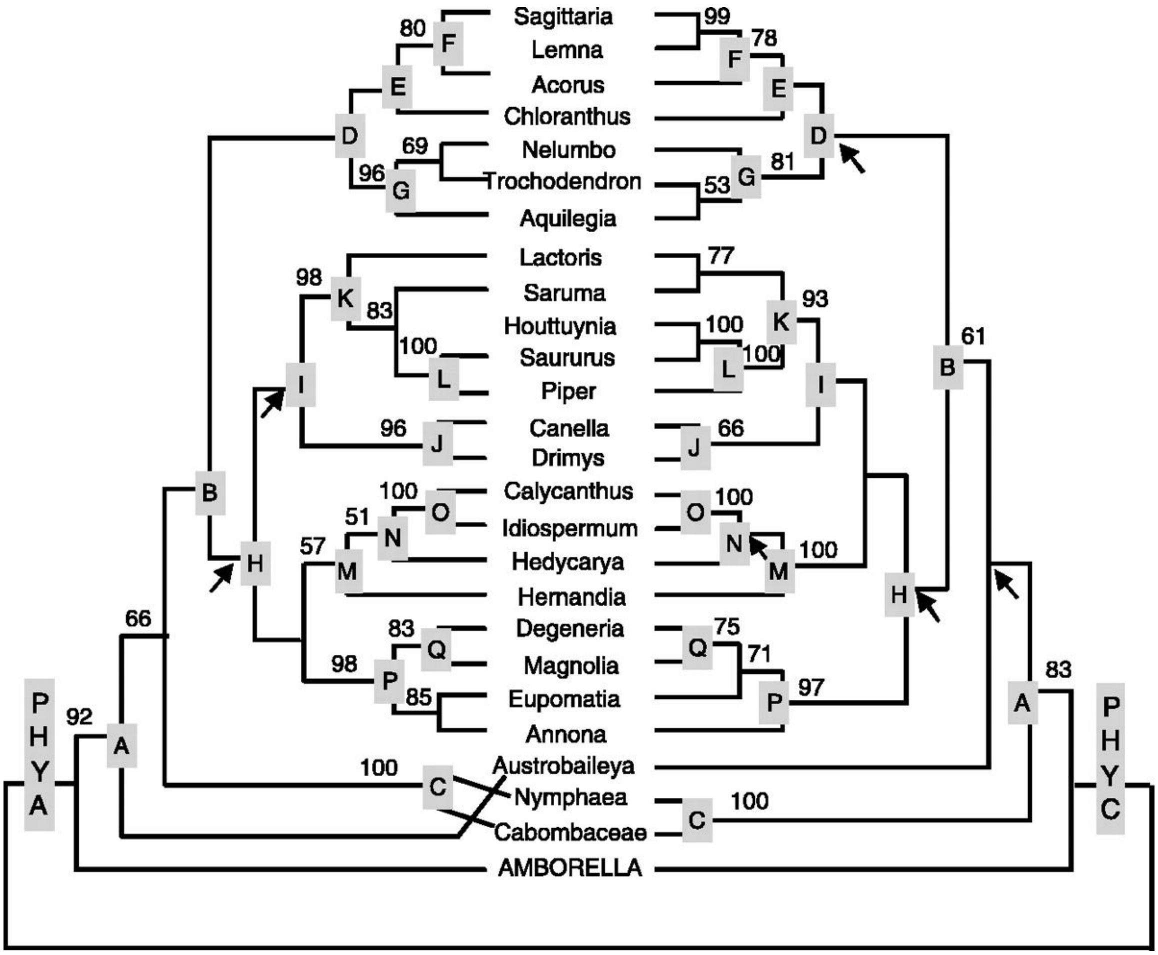
Bonus

1. On next pages, **answer advanced three thinking assignment** (20 pts, 2 points per answer).

Tree thinking quiz II

D. A. Baum, S. D. Smith, and S. D. Donovan

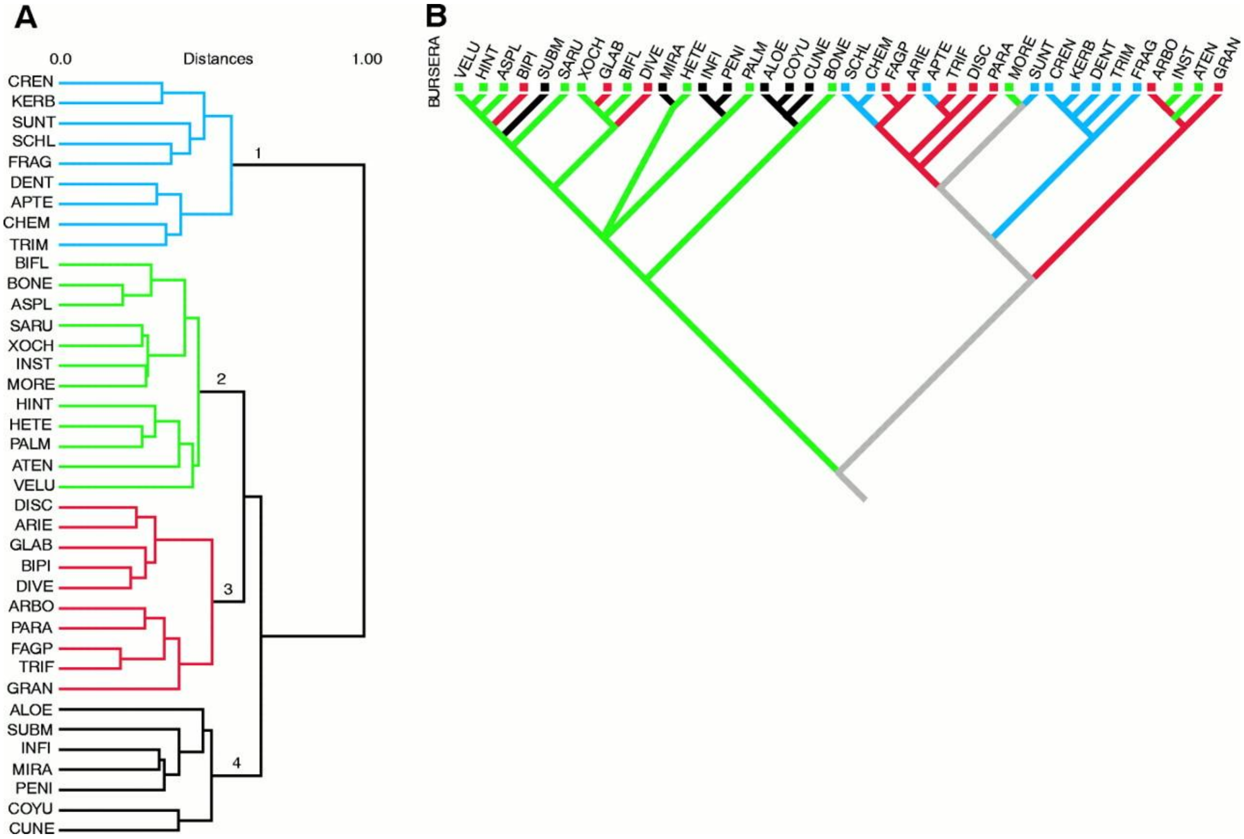
Each question in this quiz is built around a *Science* paper that depicted a phylogenetic tree. It should not be necessary to consult the original article to answer the questions.



S. Mathews, M. J. Donoghue. The root of angiosperm phylogeny inferred from duplicate phytochrome genes. *Science* **286**, 947 (1999).

1) The figure above shows the phylogeny estimated for a sample of flowering plants (angiosperms) from *PHYTOCHROME A* and *PHYTOCHROME C*, a pair of genes that duplicated prior to the origin of the angiosperms. Which of the following sets of taxa constitute a clade (=monophyletic group) on one gene tree but not on the other?

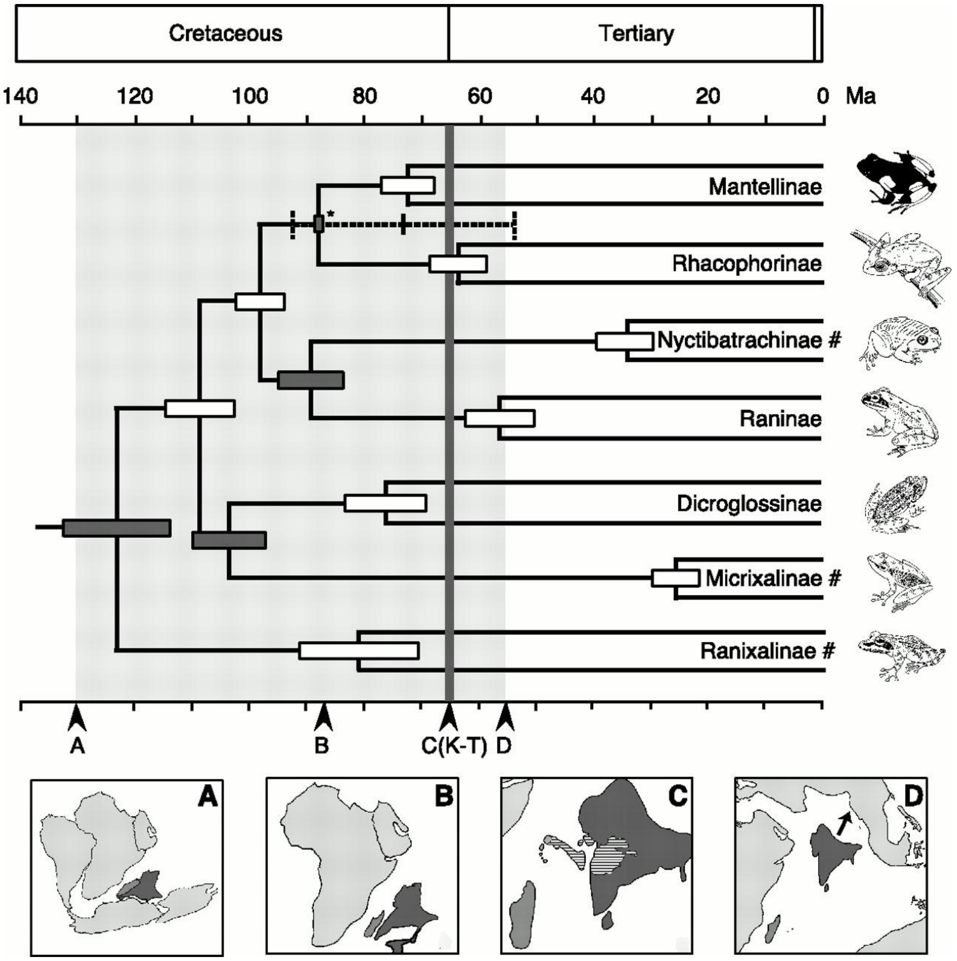
a) *Degeneria-Magnolia-Eupomatia*
b) All angiosperms except *Amborella*
c) *Austrobaileya-Nymphaea*-Cabombaceae
d) *Nelumbo-Trochodendron-Aquilegia*



J. X. Becerra. Insects on plants: macroevolutionary chemical trends in host use. *Science* **276**, 253 (1997).

2) The dendrogram on the left clusters plant species by chemical similarity; each of the four main chemical groups is indicated with a different color. This tree does not depict descent relationships, just degree of chemical similarity. On the right, the evolution of these chemical types is reconstructed on a phylogeny of the plants (this does depict inferred evolutionary relationships). The colors correspond to the chemical groups on the left, and the gray branches indicate uncertainty in character reconstruction. What does a comparison of these two figures tell us about the evolution of plant secondary chemistry?

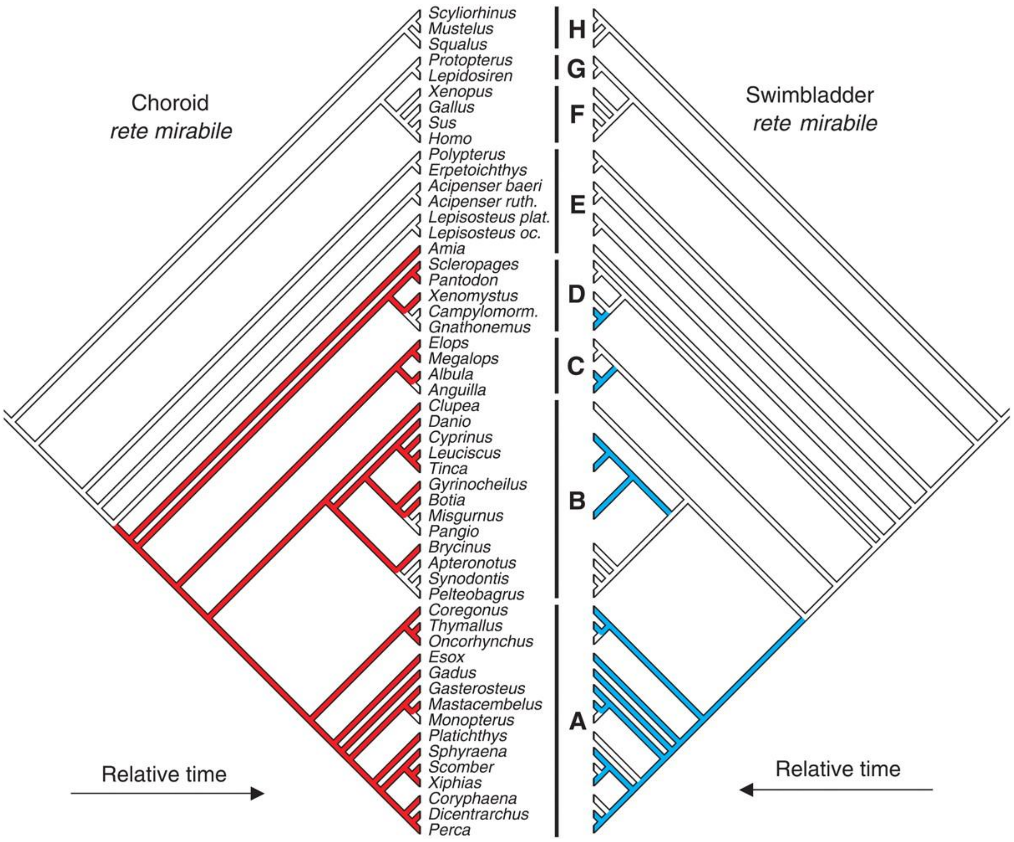
a) The four groups of chemically similar species each constitutes a distinct evolutionary lineage
b) The group colored “black” has the most advanced chemical defenses
c) The red (3) and blue (1) chemical groups are most distantly related
d) The chemical groups have each been gained and/or lost multiple times in evolution



F. Bossuyt, M. C. Milinkovitch. Amphibians as indicators of early tertiary "out-of-India" dispersal of vertebrates. *Science* **292**, 93 (2001).

3) This tree depicts inferred relationships among some major frog groups with branches drawn proportional to absolute time. Error bars on internal nodes depict confidence intervals on the dates of estimated nodes. Assuming this tree and the associated ages are correct which of the following statements is true?

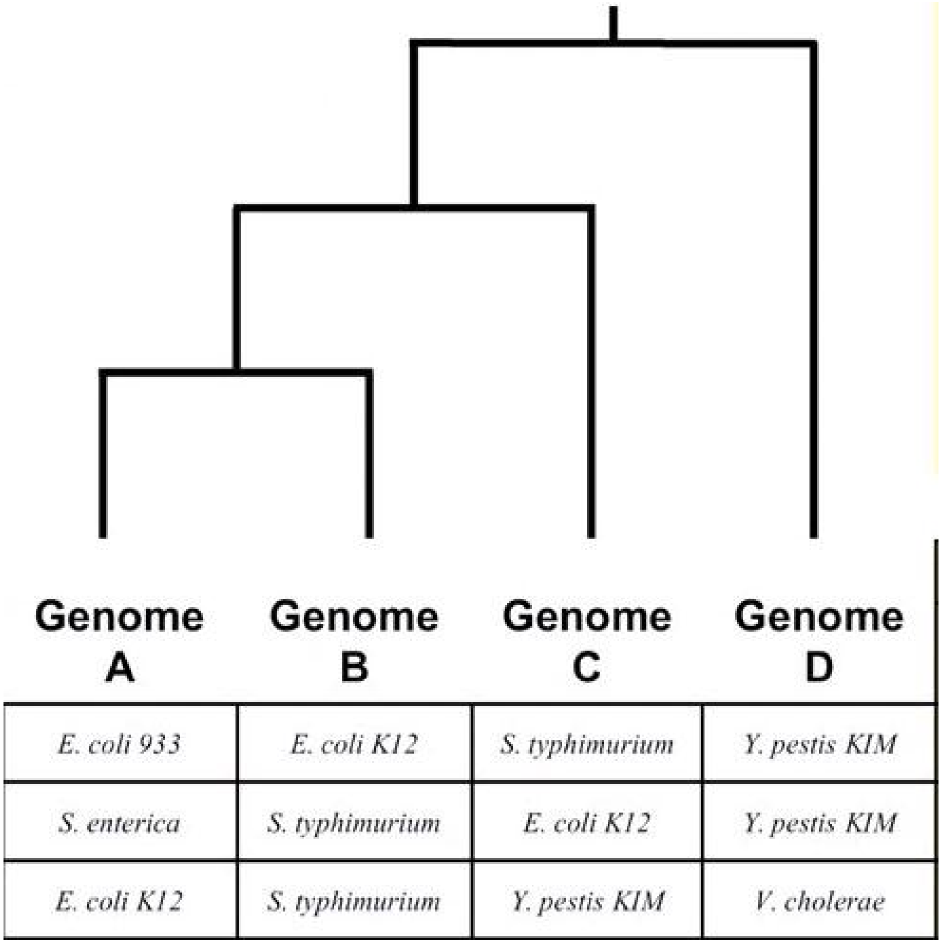
a) No individual living before 70 million years ago is an ancestor of Raninae
b) Raninae and Dicroglossinae shared a common ancestor about 75 million years ago
c) The divergence of Raninae and Nyctibatrachinae occurred more recently than the 85 million year old separation of India from Madagascar
d) The last common ancestor of Micrixalinae and Dicroglossinae lived before India and Madagascar separated (85 million years ago)



M. Berenbrink, P. Koldkjaer, O. Kepp, A. R. Cossins, Evolution of oxygen secretion in fishes and the emergence of a complex physiological system. *Science* **307**, 1752 (2005).

4) *Retia mirabilia* (sing. *rete mirabile*) are vascular bundles that allow fish to secrete O_2. In the above figure, red branches indicate lineages with choroid *retia*, blue branches indicate those with swimbladder *retia*, and white branches indicate absence of *retia*. Assuming the phylogeny and character evolution have been accurately inferred, we can see that:

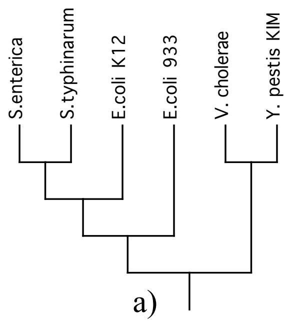
a) Swimbladder *retia* predate choroid *retia*
b) Gains of swimbladder *retia* primarily took place in lineages that already had choroid *retia*.
c) Loss of choroid *retia* causes gain of swimbladder *retia*
d) Choroid *retia* have been gained more often than swimbladder *retia*



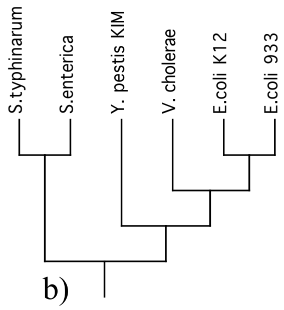
Genome A	Genome B	Genome C	Genome D
E. coli 933	*E. coli K12*	*S. typhimurium*	*Y. pestis KIM*
S. enterica	*S. typhimurium*	*E. coli K12*	*Y. pestis KIM*
E. coli K12	*S. typhimurium*	*Y. pestis KIM*	*V. cholerae*

V. Daubin, N. A. Moran, H. Ochman. Phylogenetics and the cohesion of bacterial genomes. *Science* **301,** 829 (2003).
[The above is only a portion of the figure].

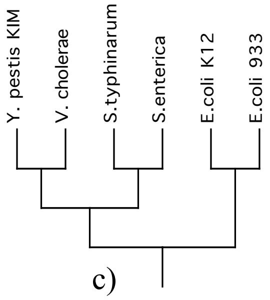
5) Each row in the table above lists a set of four bacterial taxa whose relationship follows the topology shown. Thus each row can be read as a four-taxon tree. Which of the four trees below is compatible with the information in the three rows of the table?



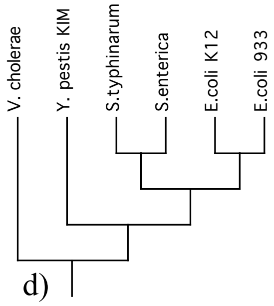
a)



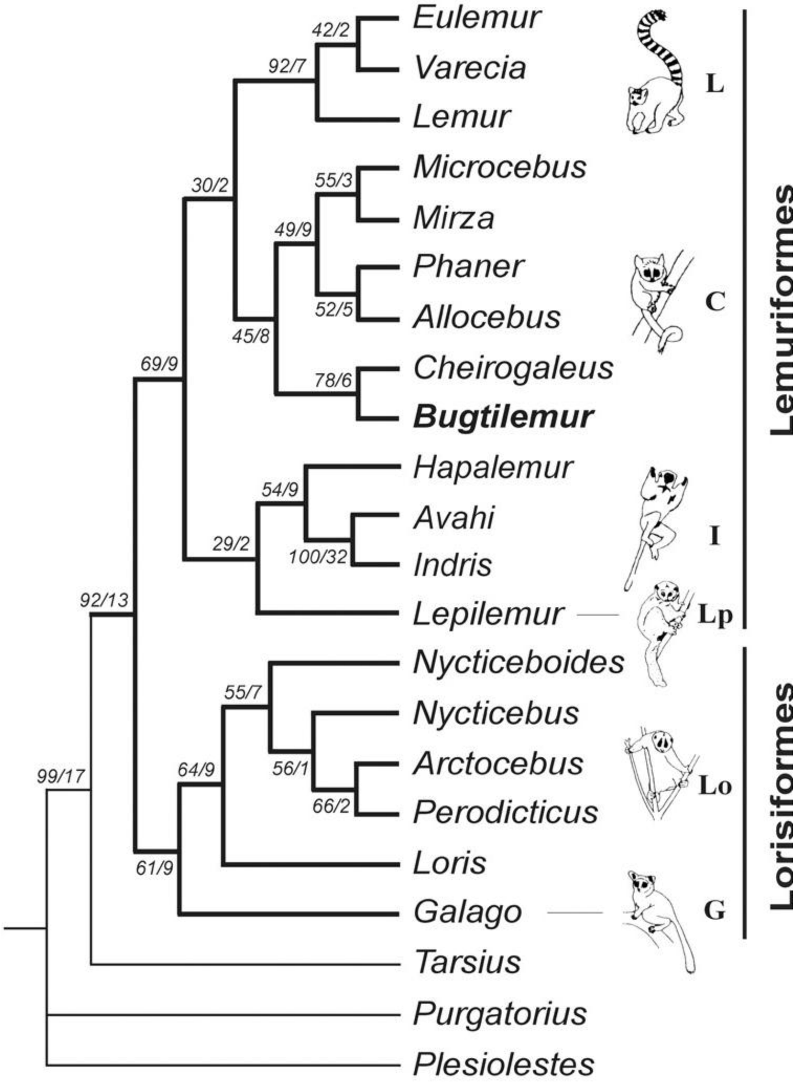
b)



c)



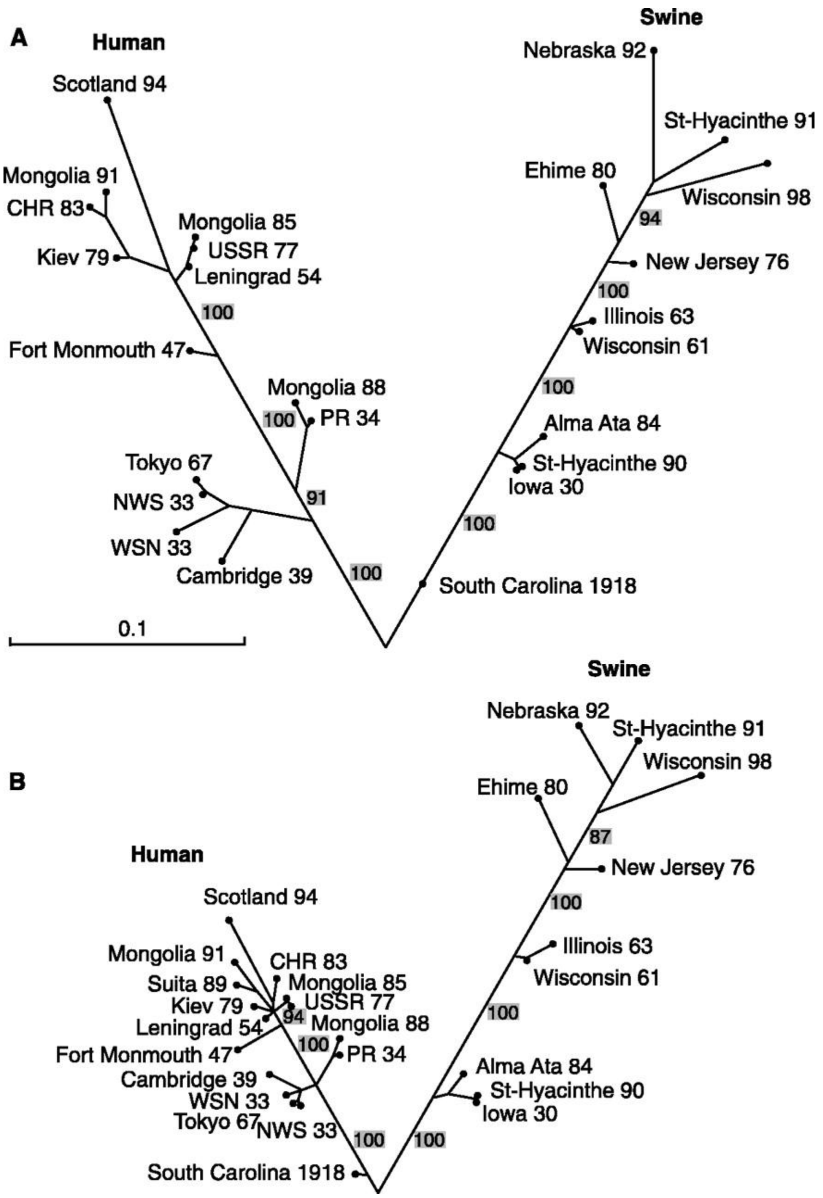
d)



L. Marivaux *et al.* A fossil lemur from the Oligocene of Pakistan. *Science* **294**, 587 (2001).
[The above is only a portion of the figure]

6) Lemuriformes are currently restricted to Madagascar, whereas Lorisiformes are found in Africa and Asia but not Madagascar, and *Tarsius* is Asian. The tree above was generated in order to assess the relationship of a fossil, *Bugtilemur*, found in 30 million year old deposits in Pakistan. Each branch of the tree has been annotated with two numbers, the first of which is the bootstrap percentage, a measure of support. In order to hold that *Bugtilemur* is more closely related to Lorisiformes than to Lemuriformes what is the minimum number of branches, with what bootstrap support, that would need to be incorrect?

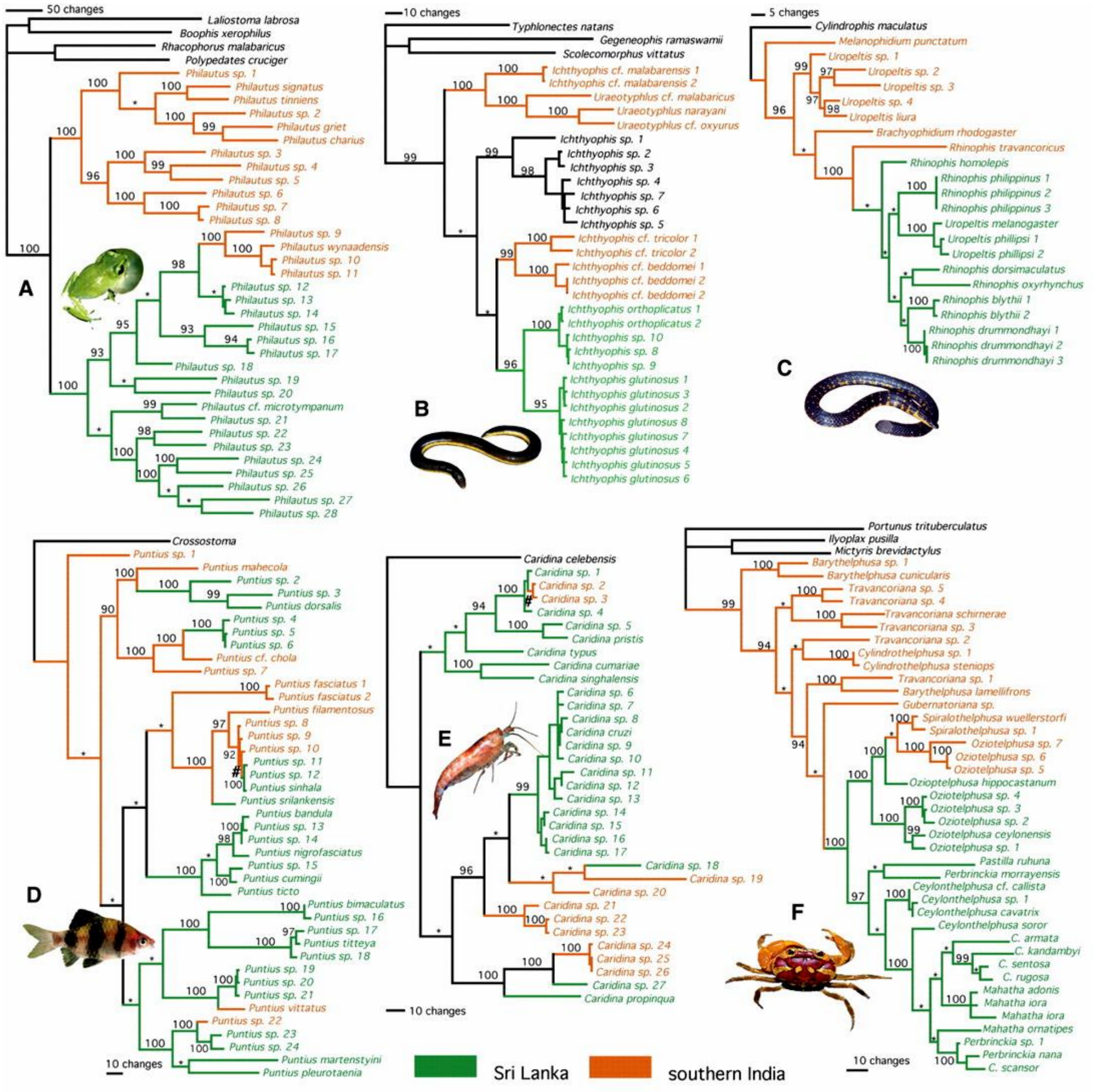
a) 1: 92%
b) 2: 78%, 69%
c) 4: 78%, 45%, 30%, 69%
d) 4: 78%, 45%, 30%, 29%



M. J. Gibbs, J. S. Armstrong, A. J. Gibbs. Recombination in the hemaglutinin gene of the 1918 "Spanish flu". *Science* **293**, 1842 (2001).

7) The trees show the phylogeny estimated for the same mammalian H1 influenzas based on portions of the haemaglutin genes. Tree A was estimated from bases 310-870 whereas tree B was estimated from bases 1070-1650. Under the assumption that these trees are correctly rooted, how does this result support the inference that the South Carolina 1918 strain (the only exemplar from the 1918 flu pandemic) arose from recombination between a human and a swine virus?

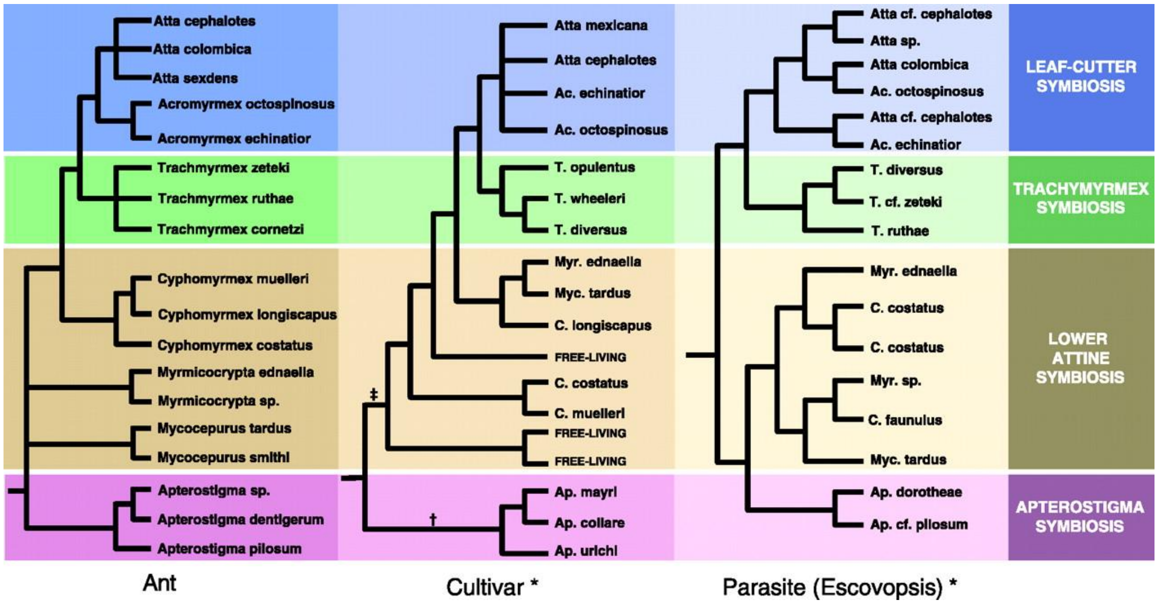
a) The fact that the South Carolina 1918 strain is near the root of both trees suggests that it is of mixed identity
b) The fact that the South Carolina 1918 strain is more closely related to the swine strains in tree A but to the human strains in tree B
c) The fact that the South Carolina 1918 strain is a direct ancestor of all the swine strains in tree A but was isolated from a human
d) The fact that the South Carolina 1918 strain is more closely related to the Iowa 30 swine strain than to the Scotland 94 human strain in both trees



F. Bossuyt *et al.* Local endemism within the Western Ghats-Sri Lanka biodiversity hotspot. *Science* **306**, 479 (2004).

8) These trees for six different groups of organisms for (**A**) tree frogs, (**B**) caecilians, (**C**) uropeltid snakes, (**D**) freshwater fishes, (**E**) freshwater shrimps, and (**F**) freshwater crabs are all colored based on whether the species come from mainland India or Sri Lanka. Assuming these trees and the reconstruction of ancestral geographic distributions are accurate, which of the trees includes a single well-supported migration from Sri Lanka back to southern India?

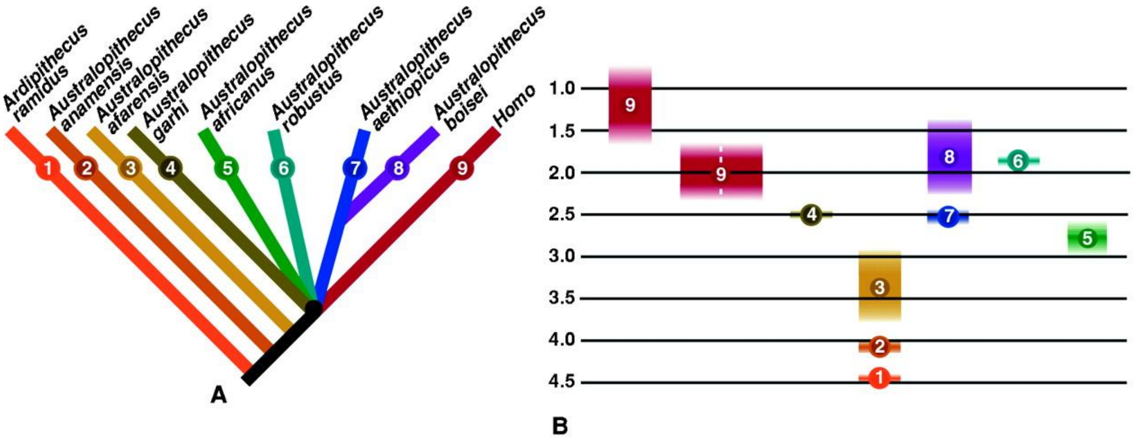
a) A, F
b) B, C
c) C
d) A, B



C. R. Currie *et al*. Ancient tripartite coevolution in the attine ant-microbe symbiosis. *Science* **299**, 325. (2003).

9) The three trees depict the relationships between leaf cutter ants (left), fungi cultivated by those ants (middle) and fungi parasitizing ant gardens (right). By comparing the phylogenies, we can distinguish whether fungi are usually passed on vertically (through generations of an ant lineage) or horizontally (e.g., among unrelated ants). Allowing that differences between trees could reflect error in phylogenetic inference, what does the similarity between the three phylogenies suggest?

a) Cultivars are passed horizontally and parasites vertically
b) Cultivars are passed vertically and parasites horizontally
c) Cultivars and parasites are passed vertically
d) Cultivars and parasites are passed horizontally



B. Asfaw *et al. Australopithecus garhi*: a new species of early hominid from Ethiopia. *Science* **284**, 629 (1999).
[The above is only a portion of the figure]

10) The above tree summarizes the inferred relationships among a number of hominid fossils as related to humans (*Homo*). The point where multiple lineage arise from a single node (a polytomy) is here intended to indicate uncertainty rather than evidence of a simultaneous divergence of an ancestral lineage into five descendant lineages. The fossil ages of each taxon are color-coded on the right. Note that fossils of species 7 predate those of its sister taxon, species 8. If the tree is accurate representation of the evolutionary history of these species, what is the minimum age we could infer for lineage 8?

a) 2.25
b) 2.5
c) 3.0
d) It cannot be determined from these data