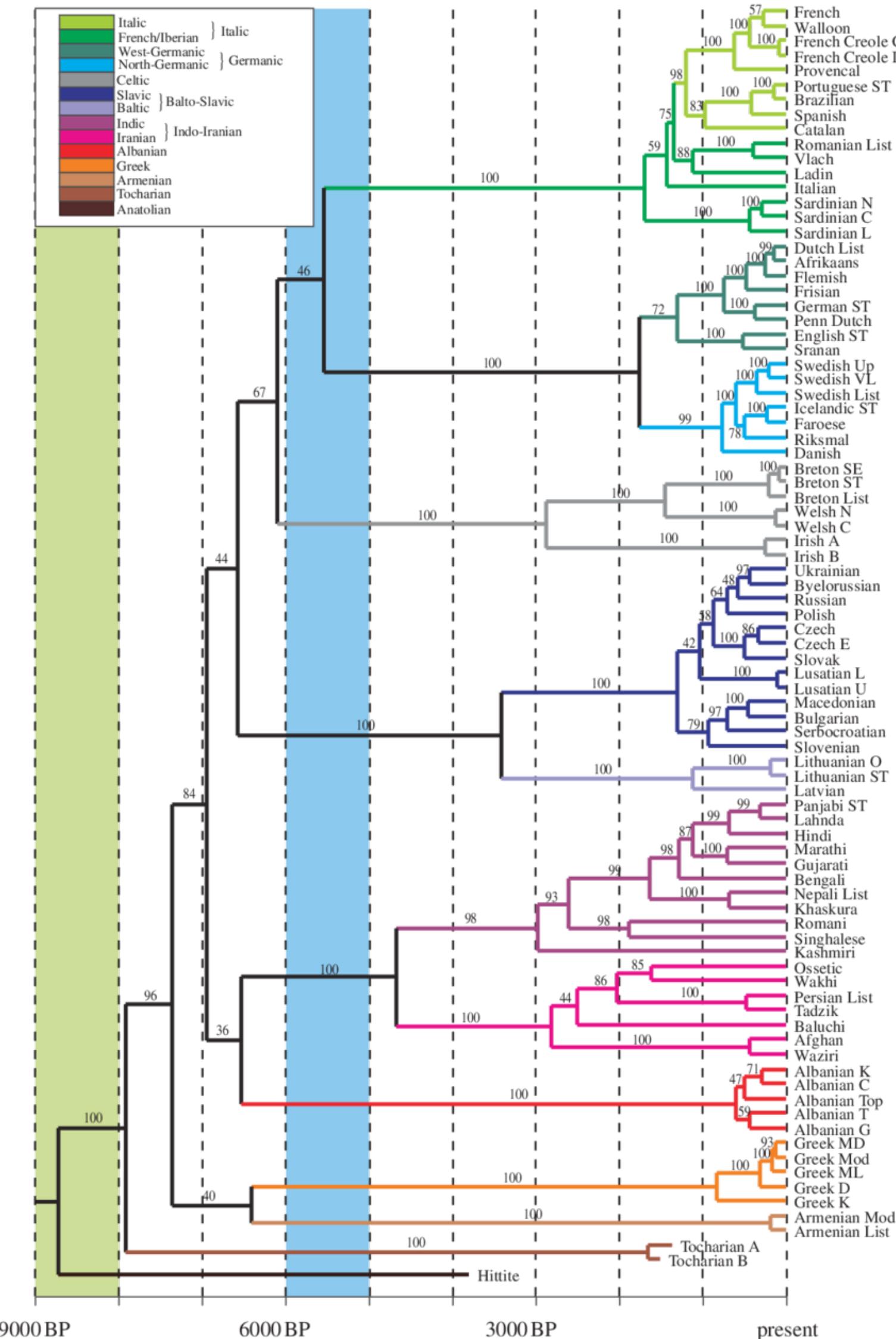


Phylogenetics 101

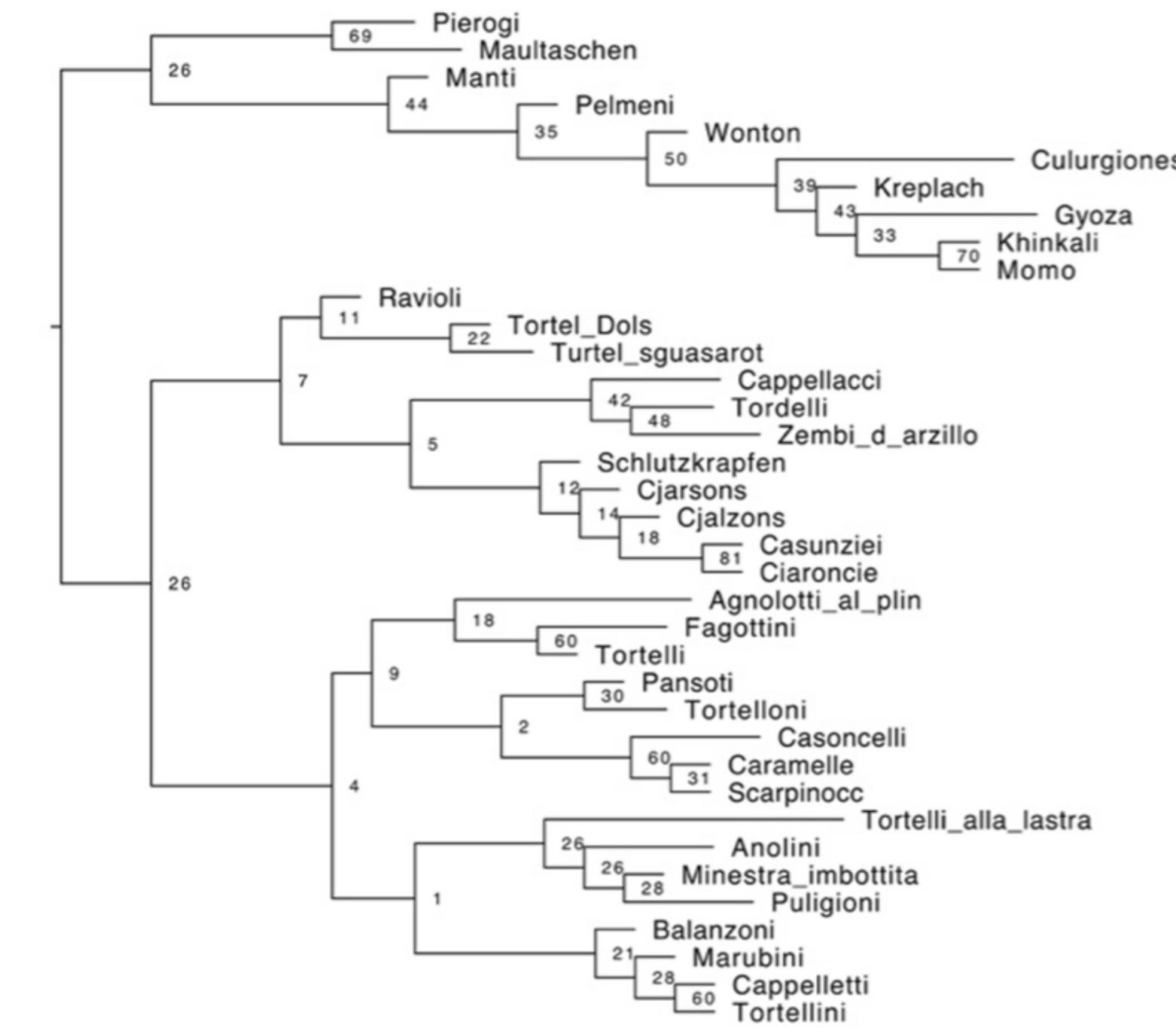
part B - molecular phylogenetics

Phylogenetics using any kind of character

- **Morphological characters:** bones, teeth, floral structures.
- **Molecular characters:** DNA, RNA, and protein sequences.
- **Behavioral traits:** mating behaviors, and ecological interactions.
- **Genomic structural features:** Gene order, synteny, and presence/absence of mobile elements.
- **Cultural and linguistic data:** Used in disciplines like anthropological phylogenetics.
... just to name a few ..



Gray et al. Language evolution and human history: what a difference a date makes. *PROC B* (2011).



- 1 Provenance (0: foreign, 1: Italian)
- 2 Italian region (0: North, 1: Central, 2: South, 3: Sardinia)
- 3 Flour (0: soft wheat, 1: durum)
- 4 Pasta (0: no eggs, 1: eggs)
- 5 Pasta size (0: small, 1: medium, 2: large, 3: very large)
- 6 Flatness (0: more or less flat, 1: 3-dimensional)
- 7 Folding (0: folded over and edges pressed, 1: two separate layers, 2: oval bundle, 3: square bundle, 4: hat-shaped bundle, 5: hat-shaped with a hole, 6: folded upright, 7: rolled and press-shaped, 8: round bundle)
- 8 Shape (0: half-moon, 1: rectangular, 2: triangular, 3: round, 4: roll, 5: bundles, 6: hat shaped)
- 9 Filling: vegetables as the main ingredient (0: absent, 1: present)
- 10 Filling: dairy (0: absent, 1: present)
- 11 Filling: meat (0: absent, 1: present)
- 12 Cooking method (0: boiled, 1: fried, 2: roasted, 3: steamed)
- 13 Sauce (0: no sauce, 1: meat-based, 2: nonmeat-based, 3: served in broth)
- 14 Serving sauce (0: none, 1: oil-based, 2: butter-based)
- 15 Sweet filling (0: absent, 1: present)

Nazari et al. Evolution of the Italian pasta ripiena: the first steps toward a scientific classification. *Discover Food* (2004).

Which criteria should characters posses?

Homologous: Characters among taxa must be due to shared ancestry.

Non-homoplasious: While some level of homoplasy is often unavoidable, characters that are less prone to these phenomena are preferred because they better reflect true evolutionary relationships.

Heritable: The trait or character must be genetically or developmentally inherited. This heritability is essential so that the character's distribution reflects historical descent.

Independent: Each character should vary independently from others. If characters are correlated or redundant, they may overemphasize certain signals and bias the analysis.

Informative: A good phylogenetic character should be neither completely invariant (providing no discriminatory power) nor so variable that the signal is overwhelmed by noise.

Clear: The character should be defined in a way that allows for unambiguous coding into discrete states (e.g., “present”/“absent” or multiple clearly defined categories).

Homoplasy:

Traits that appear similar but are not derived from a common ancestor are considered to result from homoplasy. This process attains similar outcomes **independent evolution**.

- Convergent evolution: independent evolution of similar trait
- Evolutionary reversals: reversal back to an ancestral character state

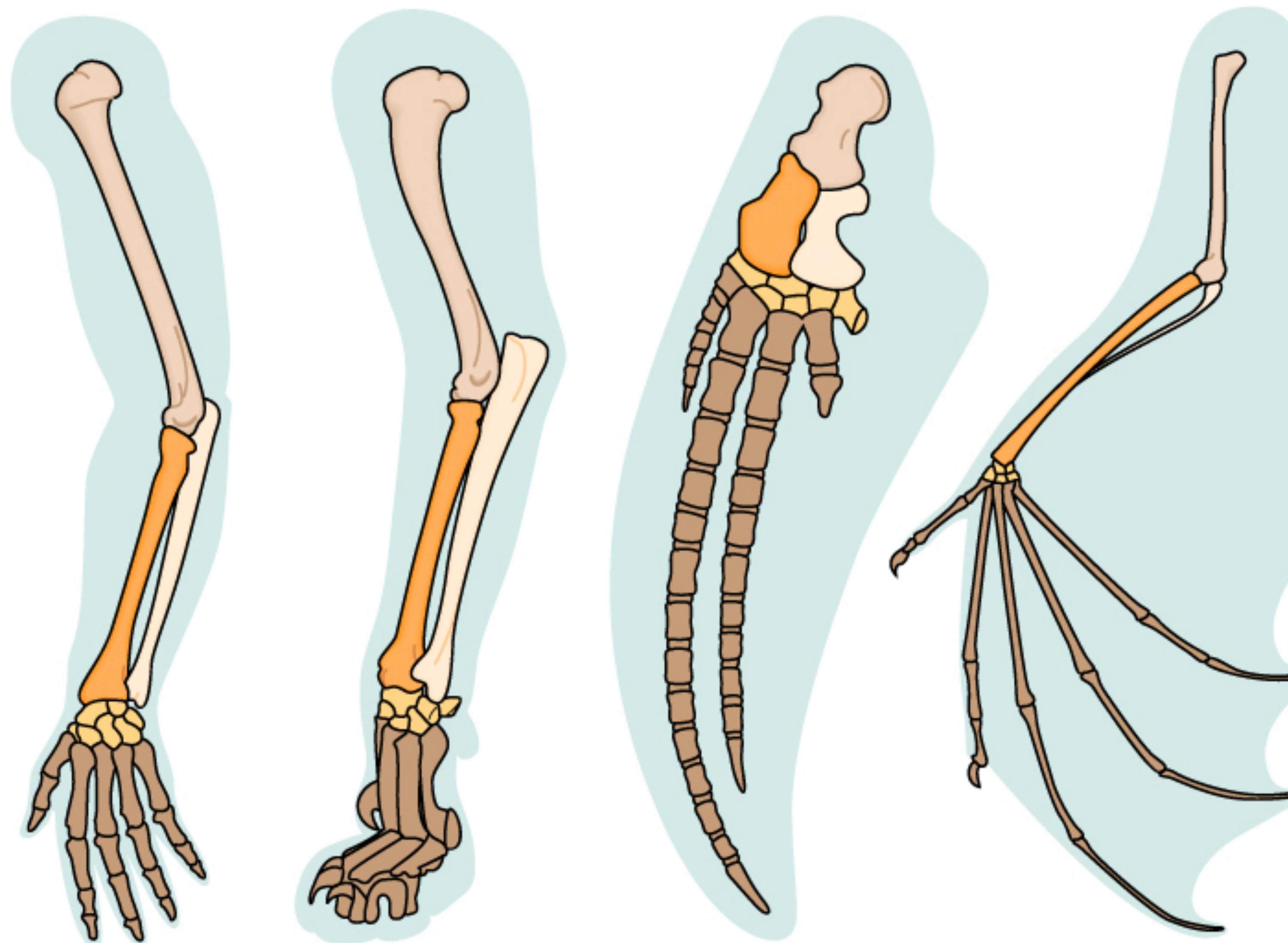
Impact on Phylogenetic Inference:

Homoplasious traits are usually **incongruent with each other**.

In the absence of homoplasy there would be no character conflict and inferring the phylogeny would be easy. Unluckily homoplasy can be common and **mislead the reconstruction of phylogenies**.

Homoplasious traits can be misleading in reconstruction of phylogenies.

- **Wings:** Birds, bats, and insects all have wings for flight, yet these structures evolved independently. While they serve the same function, the underlying anatomy and developmental origins differ.
- **Body streamlining in aquatic lifestyle:** Sharks (cartilaginous fish), dolphins (mammals) exhibit similar streamlined body shapes that help reduce drag in water. These similarities are convergent adaptations to an aquatic lifestyle.
- **Camera-Type Eyes:** Both cephalopods (like octopuses) and vertebrates have complex, image-forming eyes. Nonetheless these eyes have different anatomical structures.
- **Echolocation:** Although both bats and whales use echolocation for navigation and hunting, they evolved this ability independently.



Human

Cat

Whale

Bat

Why molecular phylogenetics?

Today, most phylogenetics is done using molecular sequence data because:

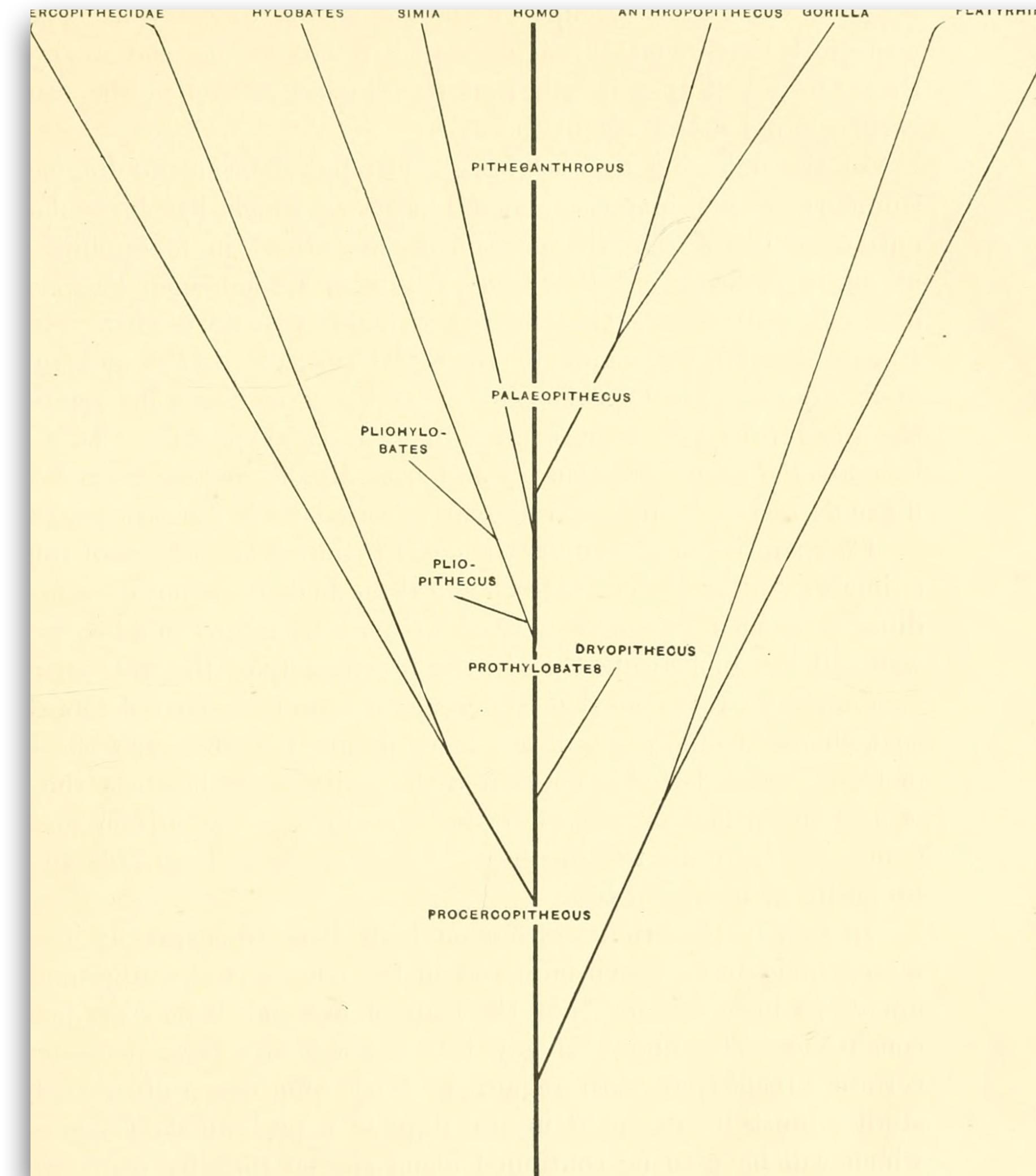
- **Abundant and inexpensive:** sequences provide thousands to millions independent characters at a cheap price.
- **Objective:** molecular data is less subjective than morphological assessment.
- **Multiple resolutions:** different approaches can resolve fine-scale and macroscopic differences relationships among closely related taxa.
- **Standardized:** computational tools and databases facilitate analyses and allow their reproducibility.

The first molecular phylogenies

Studies in immunochemistry revealed that **serological cross-reactions were stronger between more closely related organisms**, providing an approach to infer evolutionary relationships.

George Nuttall performed a study in **1904**, using immunological techniques to reconstruct the phylogenetic relationships among various animal groups.

His findings demonstrated that **humans' closest relatives were the great apes**, followed, in descending order of relatedness, by Old World monkeys, New World monkeys, and prosimians (such as lemurs and tarsiers). This study provided early molecular evidence supporting evolutionary theory.



George HF Nuttall. Blood Immunity and Blood Relationship.
Science (1904)

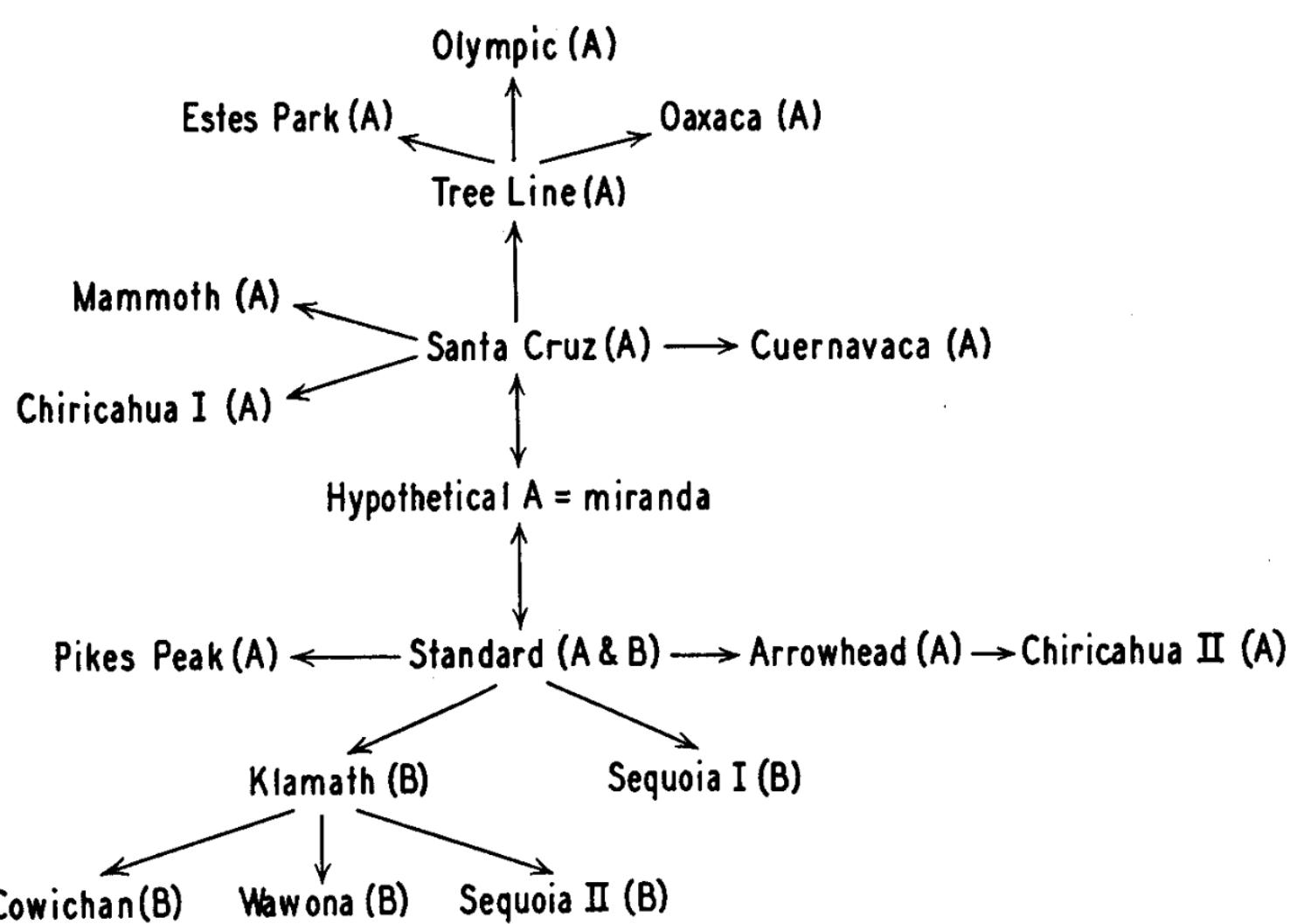


FIGURE 3.—Phylogeny of the gene arrangements in the third chromosome of *Drosophila pseudoobscura*. Any two arrangements connected by an arrow in the diagram differ by a single inversion. Further explanation in text.

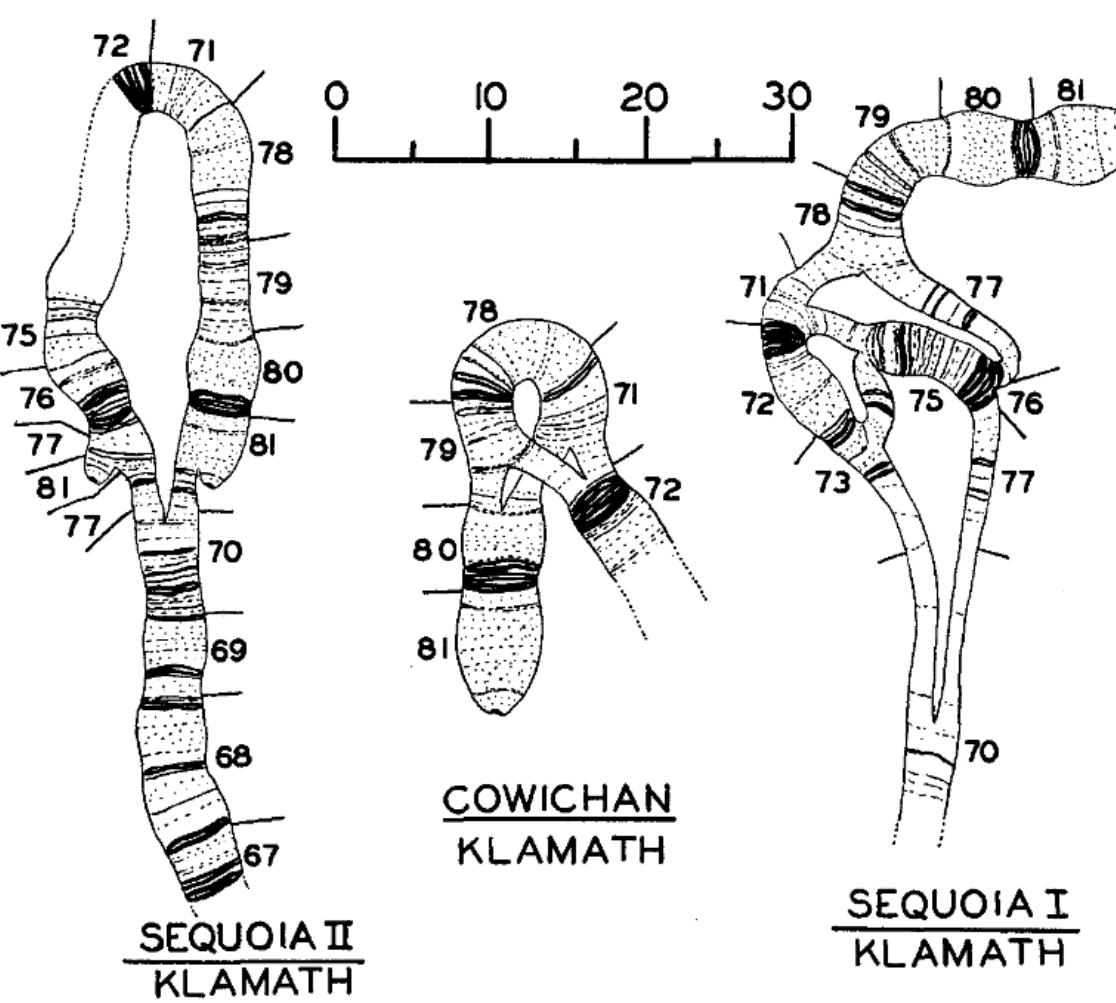


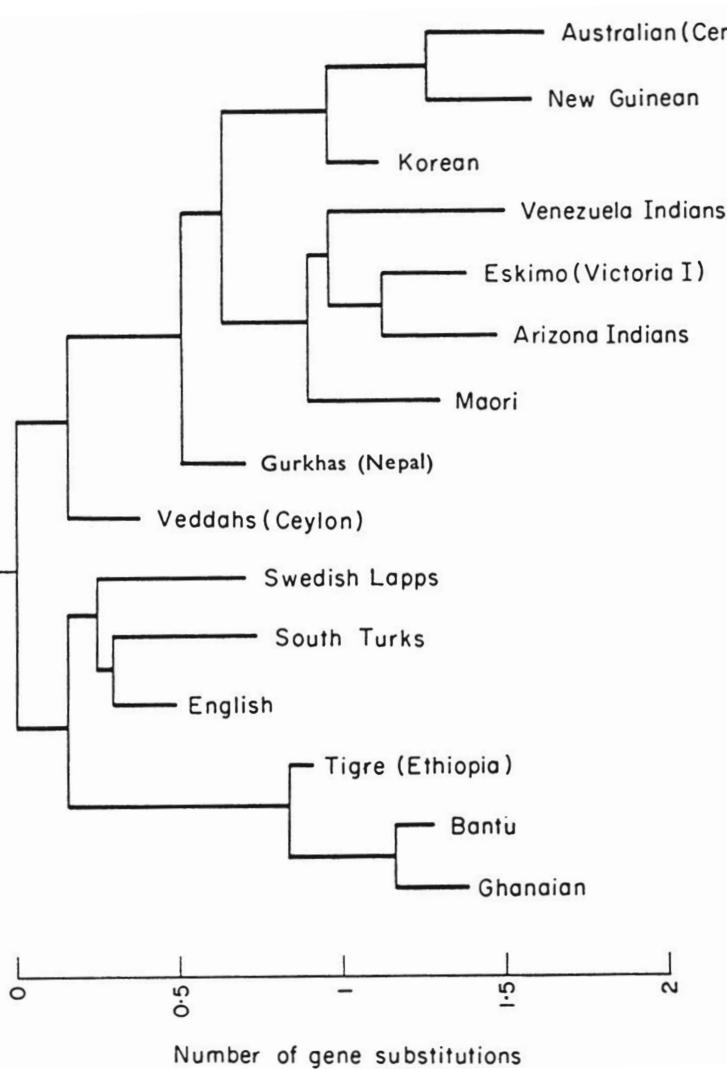
FIGURE 2.—Configurations in the third chromosome in hybrids between various strains of race B.



Dobzhansky and Sturtevant. Inversions in the Chromosomes of *Drosophila Pseudoobscura*. Genetics (1938).

"The Olivetti computer was not provided with a high-level language (FORTRAN II) until 1963, after which progress was rapid. The implementation of Cavalli's method was reported at the Eleventh International Congress of Genetics in September 1963, where my algorithmic alternative 'Method of minimum evolution' minimizing the total length of the tree was also mentioned, later to be called 'Parsimony'. The joint paper presented a computed tree for 15 populations based on blood-group gene frequencies..."

from A. W. F. Edwards FRS Memoirs on Luigi Cavalli Sforza



Cavalli Sforza and Edwards. Analysis of human evolution. Genet. Today, Proc. XI Int. Congr. Genet., vol. 2, 923–933.

Construction of Phylogenetic Trees

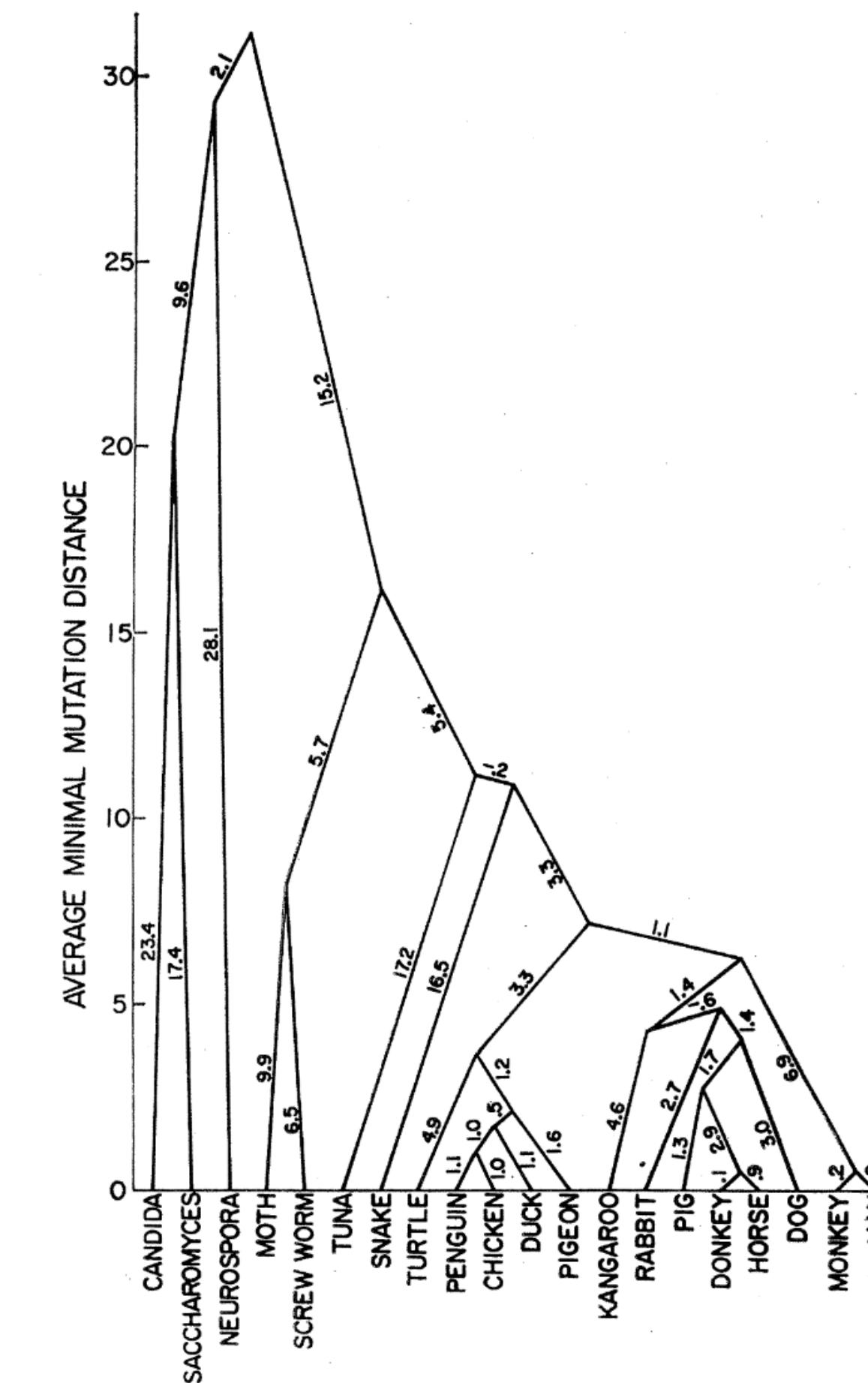
A method based on mutation distances as estimated from cytochrome *c* sequences is of general applicability.

Walter M. Fitch and Emanuel Margoliash

Biochemists have attempted to use quantitative estimates of variance between substances obtained from different species to construct phylogenetic trees. Examples of this approach include studies of the degree of interspecific hybridization of DNA (1), the degree of cross reactivity of antisera to purified proteins (2), the number of differences in the peptides from enzymic digests of purified homologous proteins, both as estimated by paper electrophoresis-chromatography or column chromatography and as estimated from the amino acid compositions of the proteins (3), and the number of amino acid replacements between homologous proteins whose complete primary structures had been determined (4). These methods have not been completely satisfactory because (i) the portion of the genome examined

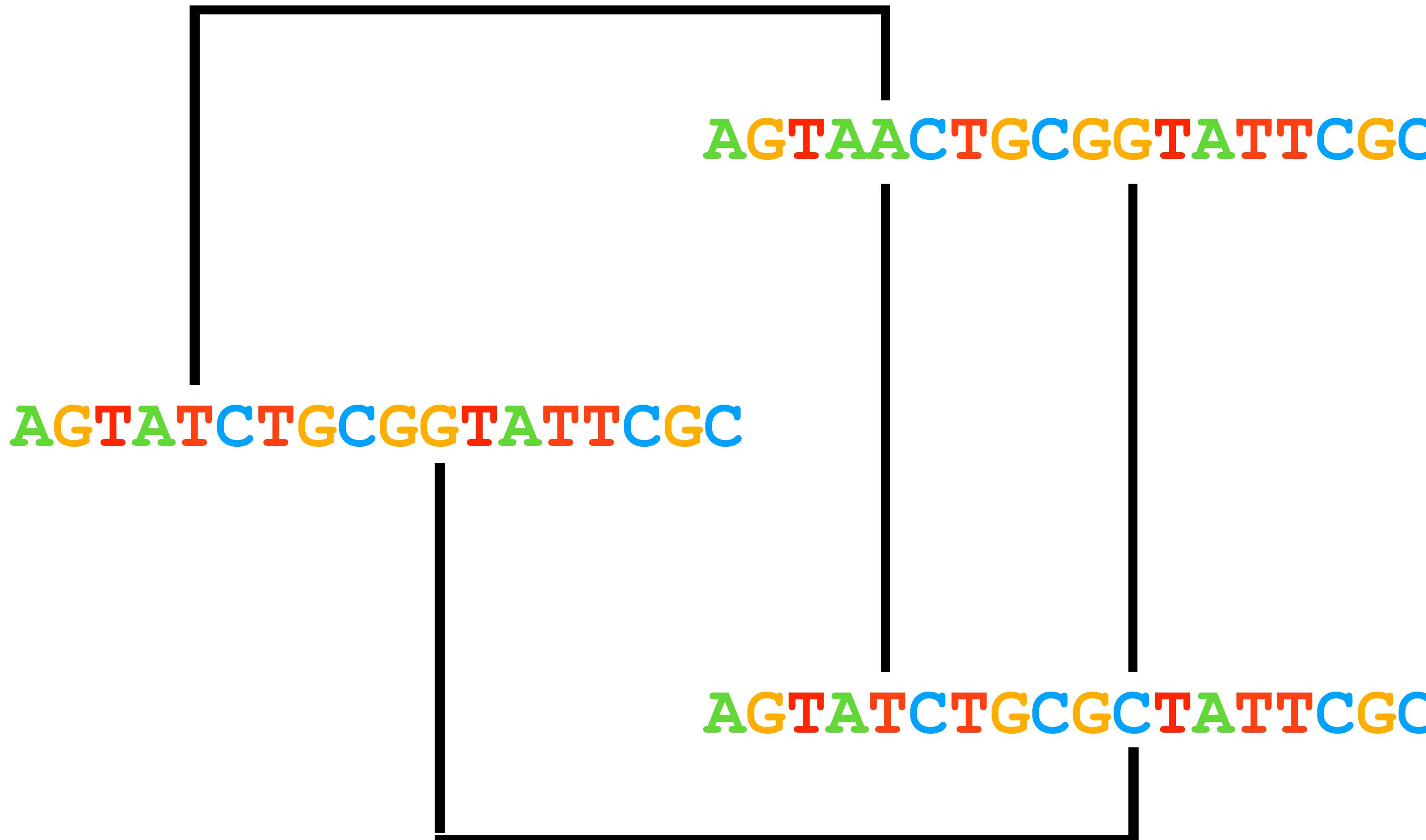
ogous proteins, both as estimated by paper electrophoresis-chromatography or column chromatography and as estimated from the amino acid compositions of the proteins (3), and the number of amino acid replacements between homologous proteins whose complete primary structures had been determined (4). These methods have not been completely satisfactory because (i) the portion of the genome examined

20 JANUARY 1967

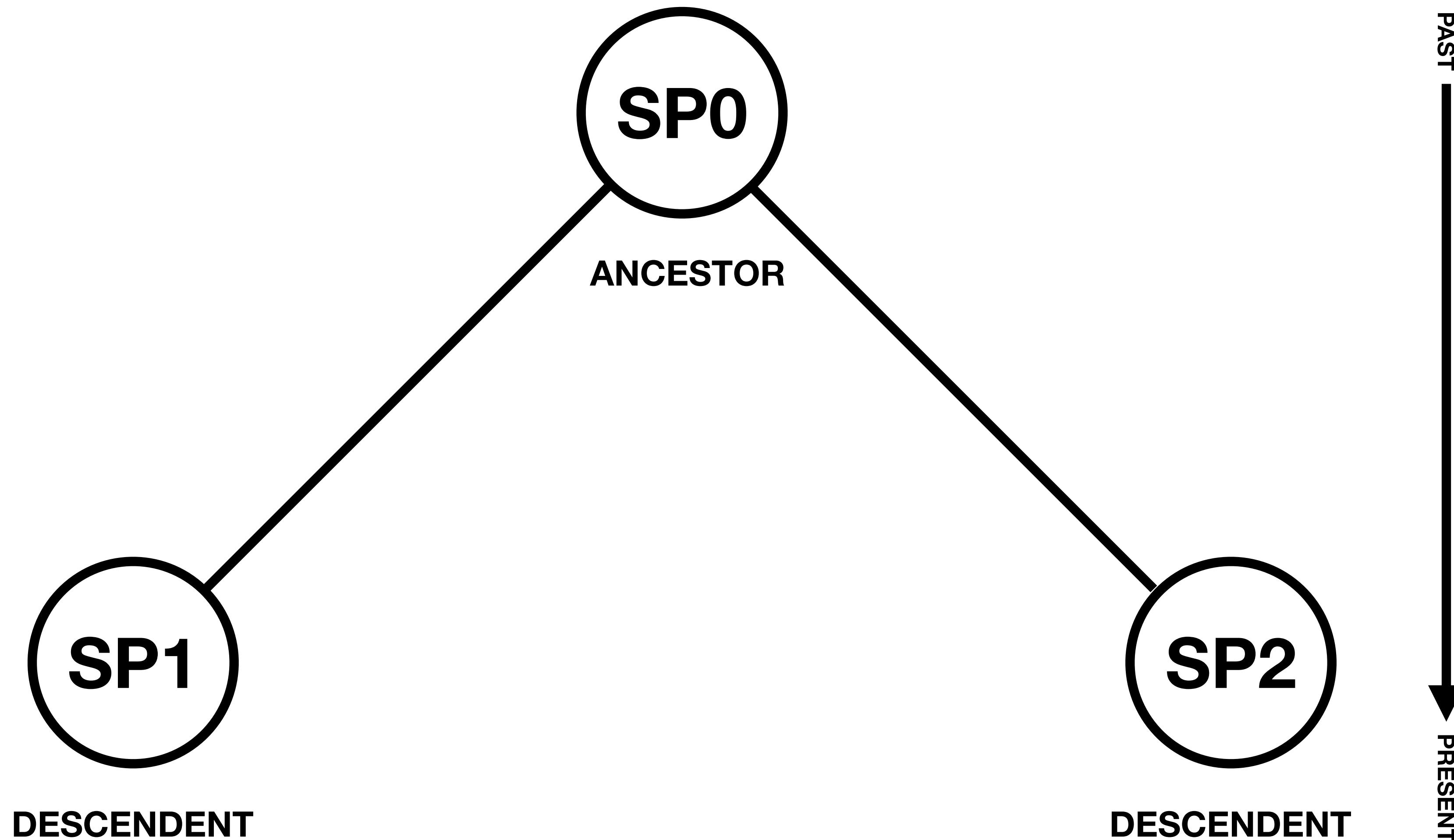


Fitch and Margoliash in 1967 published a [paper](#) where they provided the first phylogent based on a “modern” approach, which substantially consists of **sequence evolution and mutation distance** 😊.

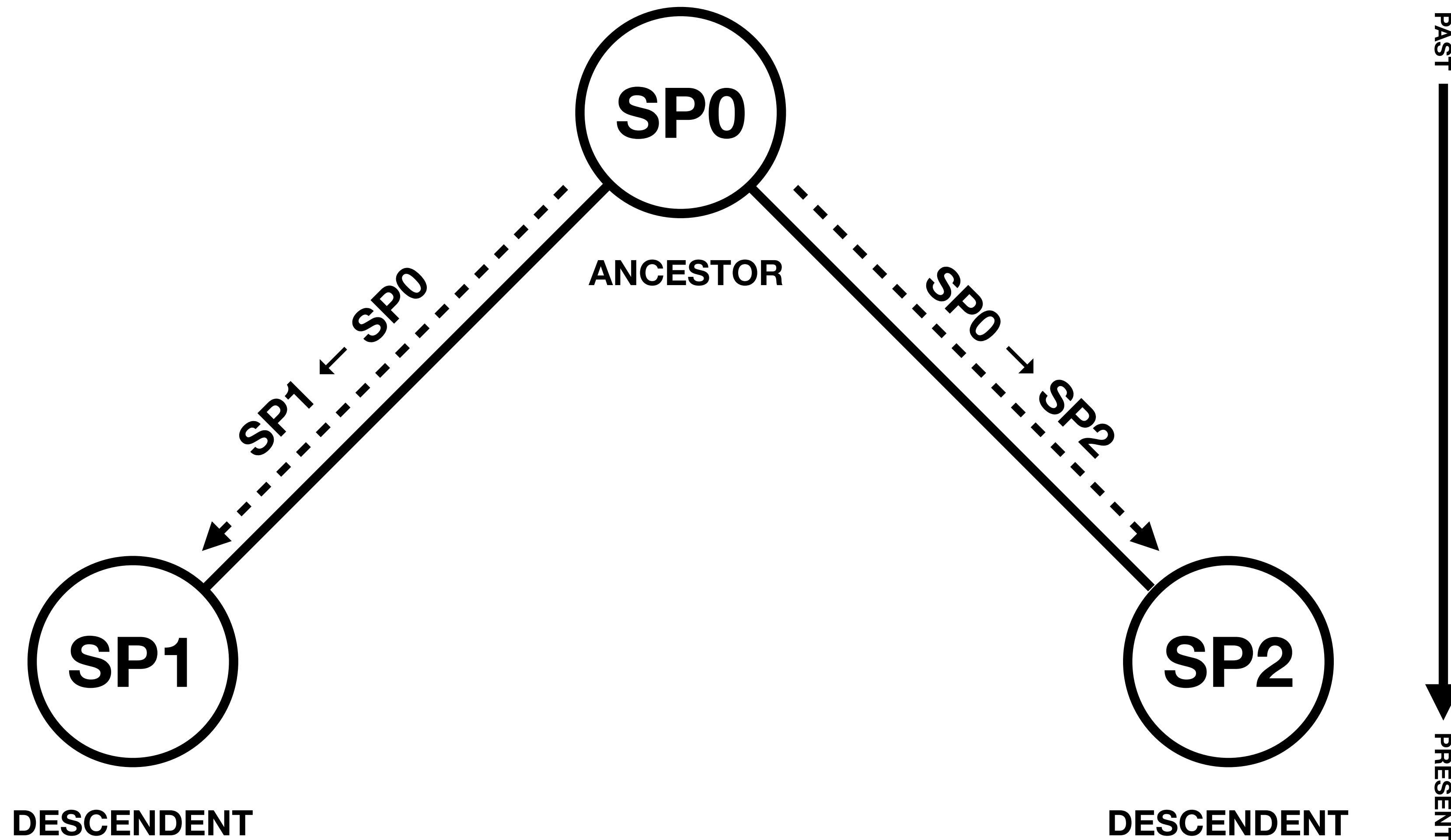
How to define the genetic distance between two taxa?



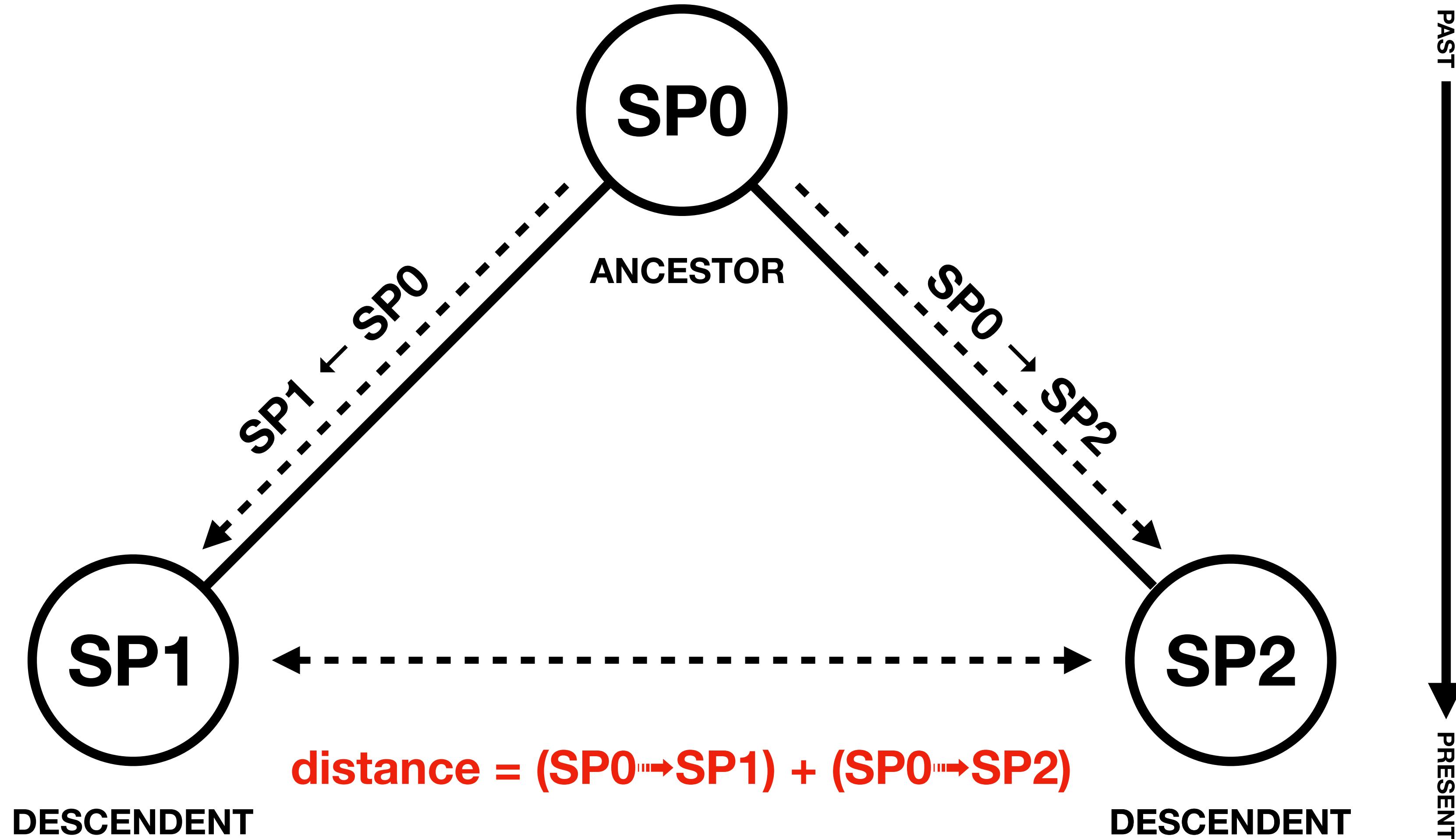
How to define the genetic distance between two taxa?

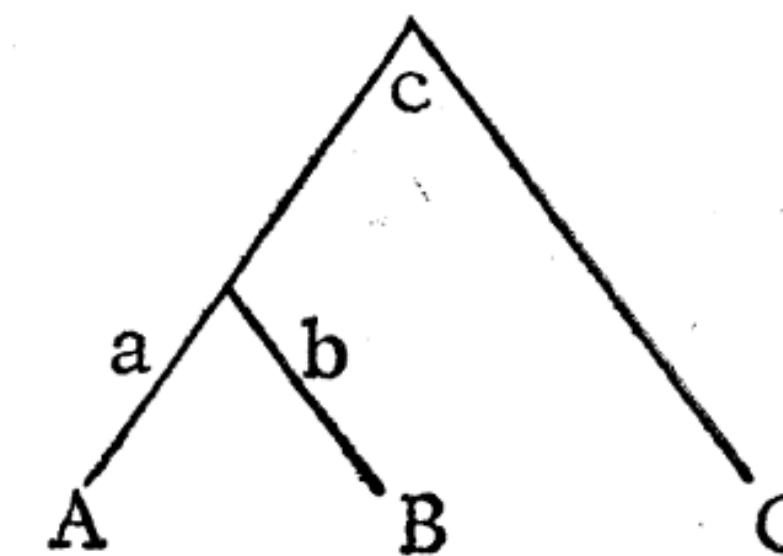


How to define the genetic distance between two taxa?



How to define the genetic distance between two taxa?



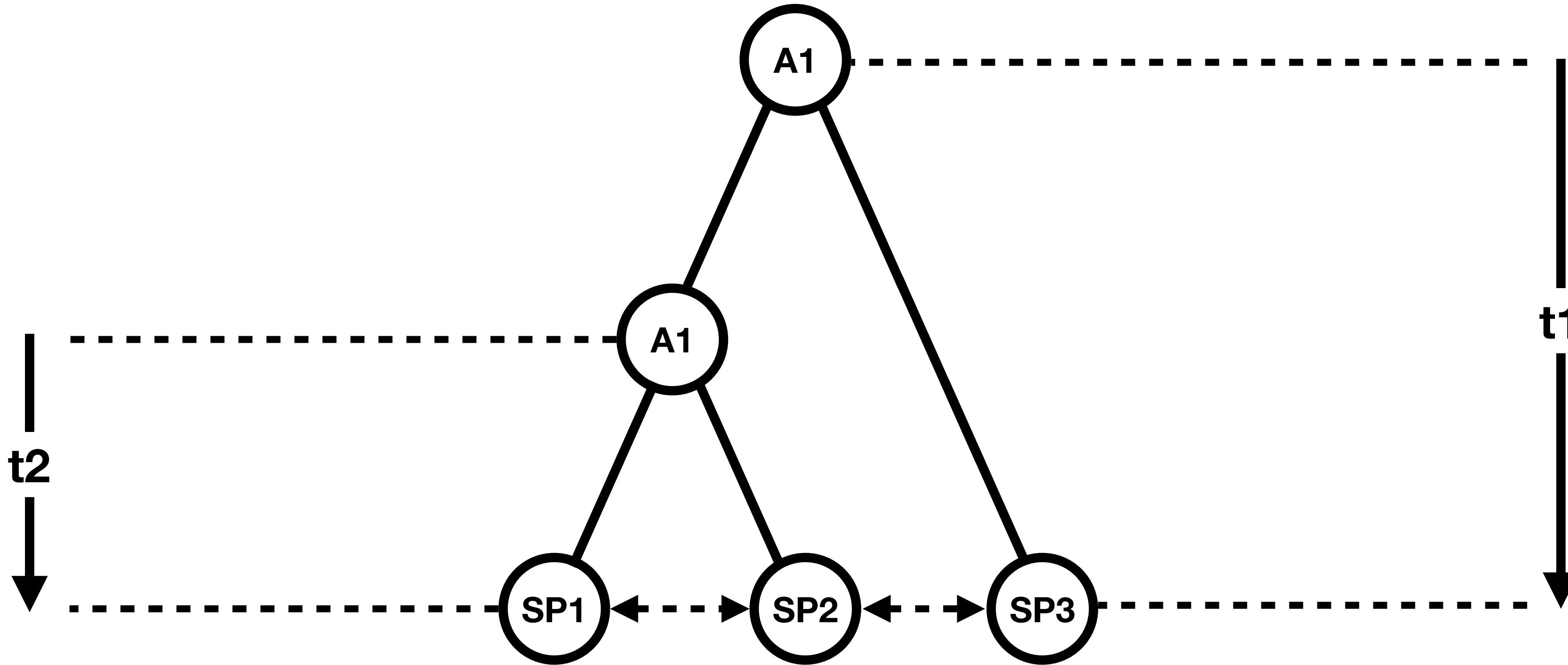


Mutation Distances

	B	C
A	24	28
B		32

Fig. 1. Calculation of observed mutation distances. The upper apex represents a hypothetical ancestral organism that divided into two descending lines, one of which subsequently also divided. Thus we have three present-day species, A, B, and C. The number of observable mutations that have occurred in a particular gene since the A and B lines of descent diverged are represented respectively by a and b . The number of mutations that separate the lower apex and C is represented by c . The sums of $a + b$, $a + c$, and $b + c$, then, are the mutation distances of the three species as currently observed.

AGTATCTGCGGTATT CGC
↓
AGCATCTGCGGTATT CGC
↓
AGTATTGCGGTATT CGC
↓
AGTATCTGCGGTATT AGC
↓
AGTATCTGCCGTATT CGC

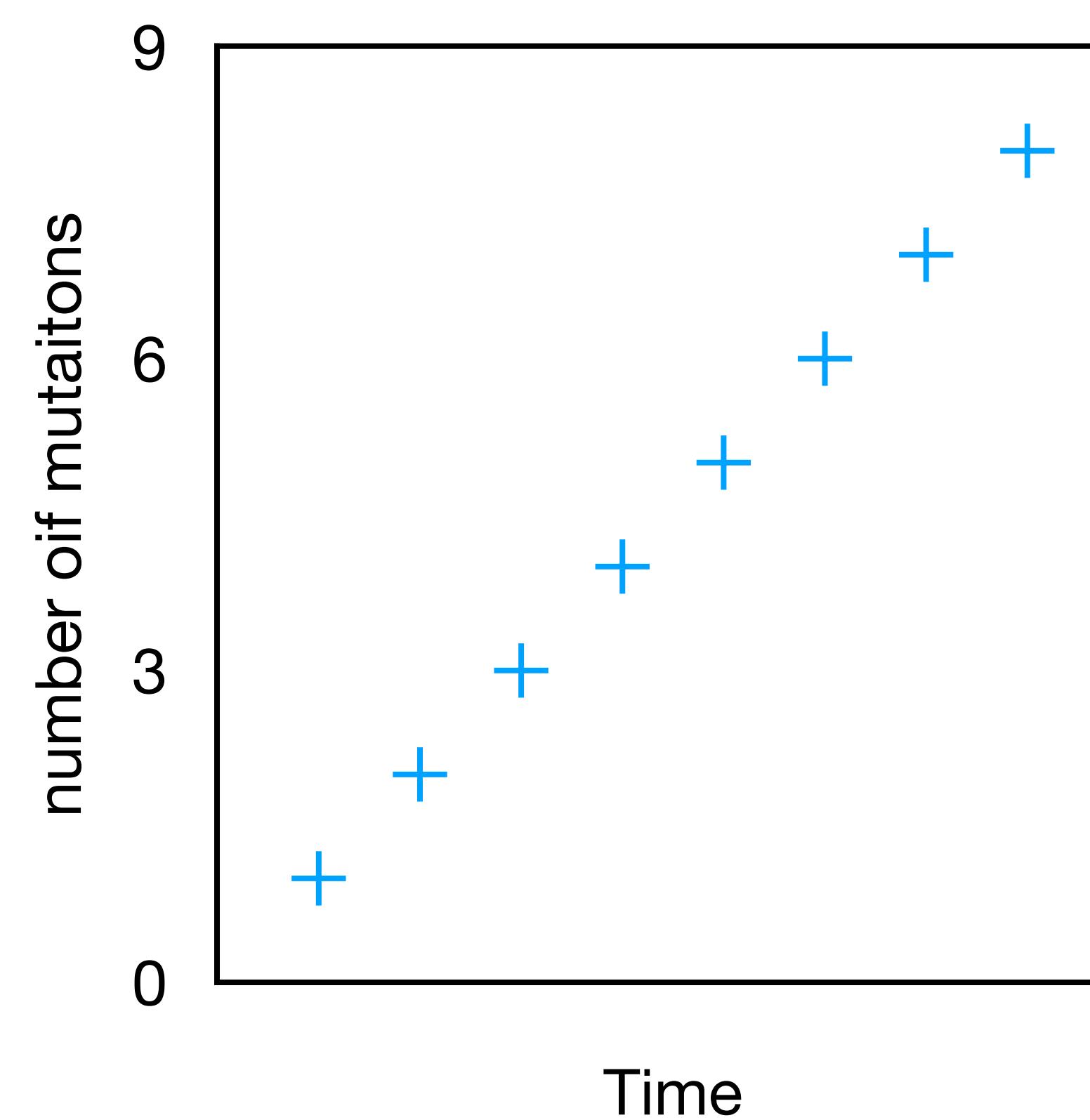


$$t_1 > t_2$$

$$d_1 > d_2$$

more ancient is the split
more time to accumulate changes
higher is the genetic distance

The molecular clock hypothesis suggests that **genetic changes** tend to accumulate at a relatively **constant rate** over long periods, which can help estimate divergence times between species.

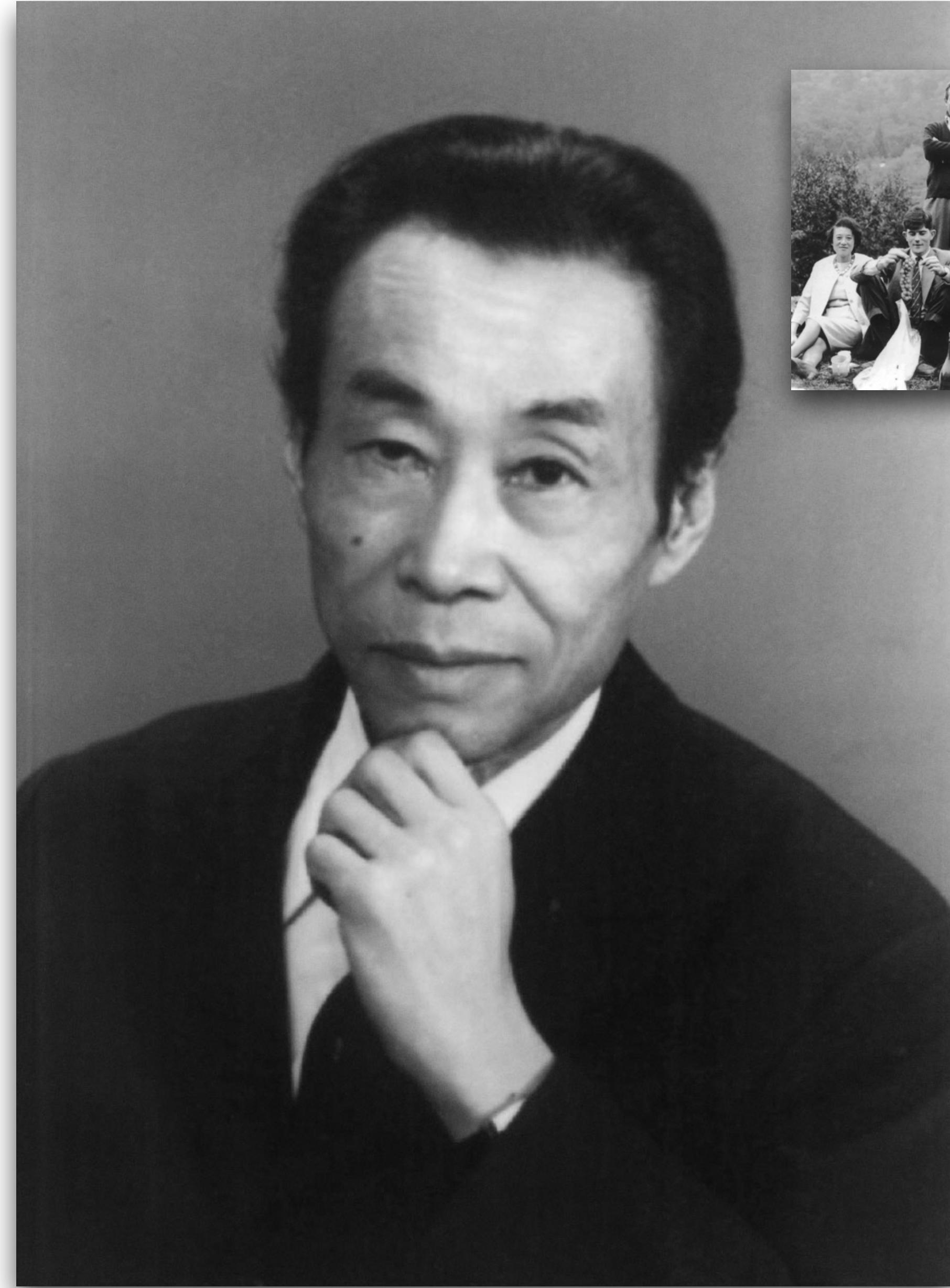


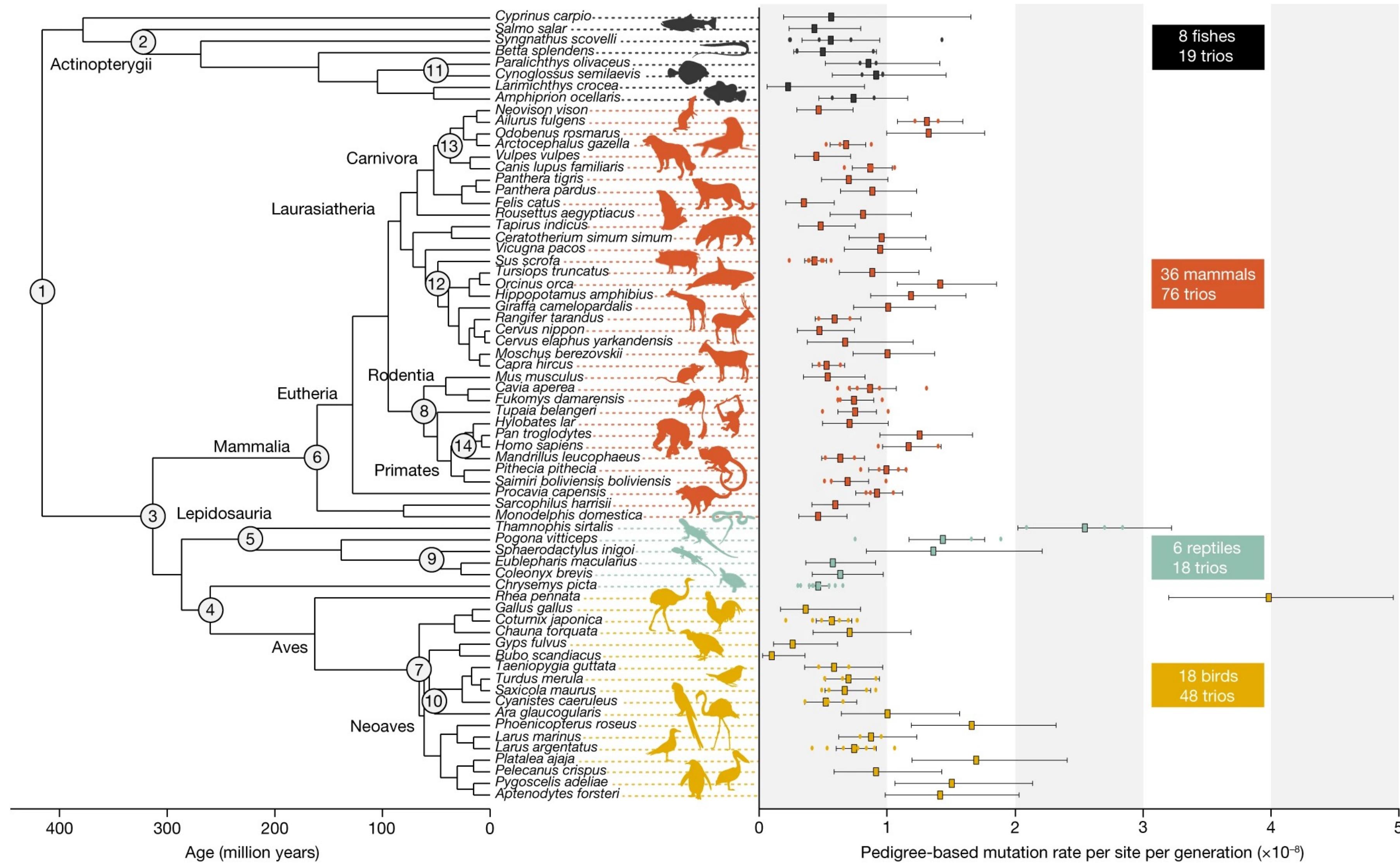
neutral theory & phylogenetics

Most molecular changes and within-species variability arise from random drift of neutral - or nearly neutral - mutant alleles, not from natural selection. Proposed by **Motoo Kimura**, this theory applies primarily to molecular evolution, while phenotypic traits are still shaped by natural selection (as per Darwin).

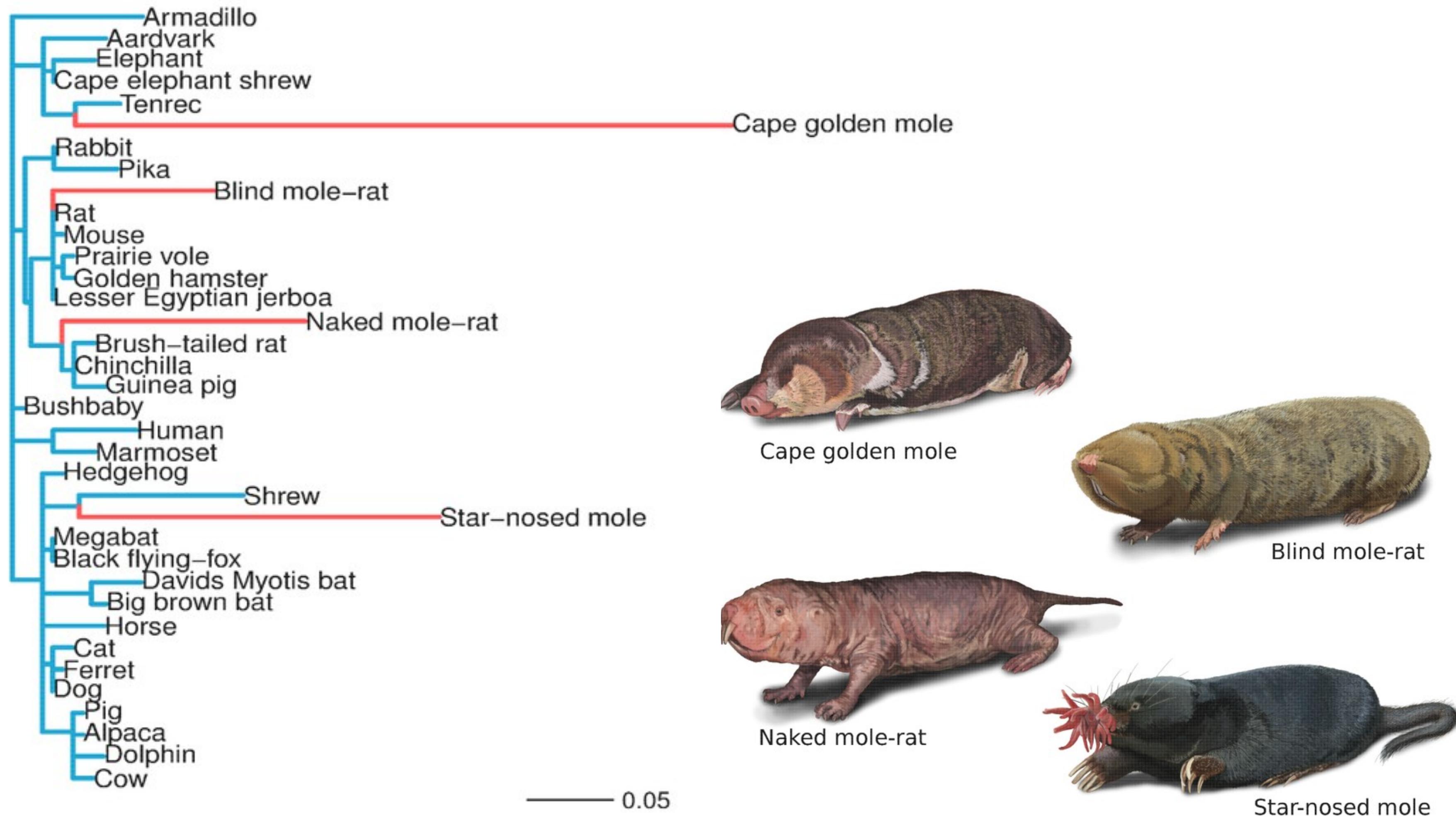
Implications for Phylogenetics:

- Since many molecular changes are neutral, they allow us to reconstruct evolutionary relationships.
- Since neutral mutations accumulate at a - relatively - constant rate over time, they allow us to estimate divergence times.





Lens intrinsic membrane 2 (Lim2) protein tree



... however ...

- **Mutation rates can vary:**

Mutation rates can differ among genes, species, and even different parts of the genome due to factors like generation time, metabolic rate, and DNA repair mechanisms.

- **Selection can vary (across genes and genomes):**

Some changes are subject to strong selective pressures and may be removed (purifying selection) or favored (positive selection), altering the rate of change.

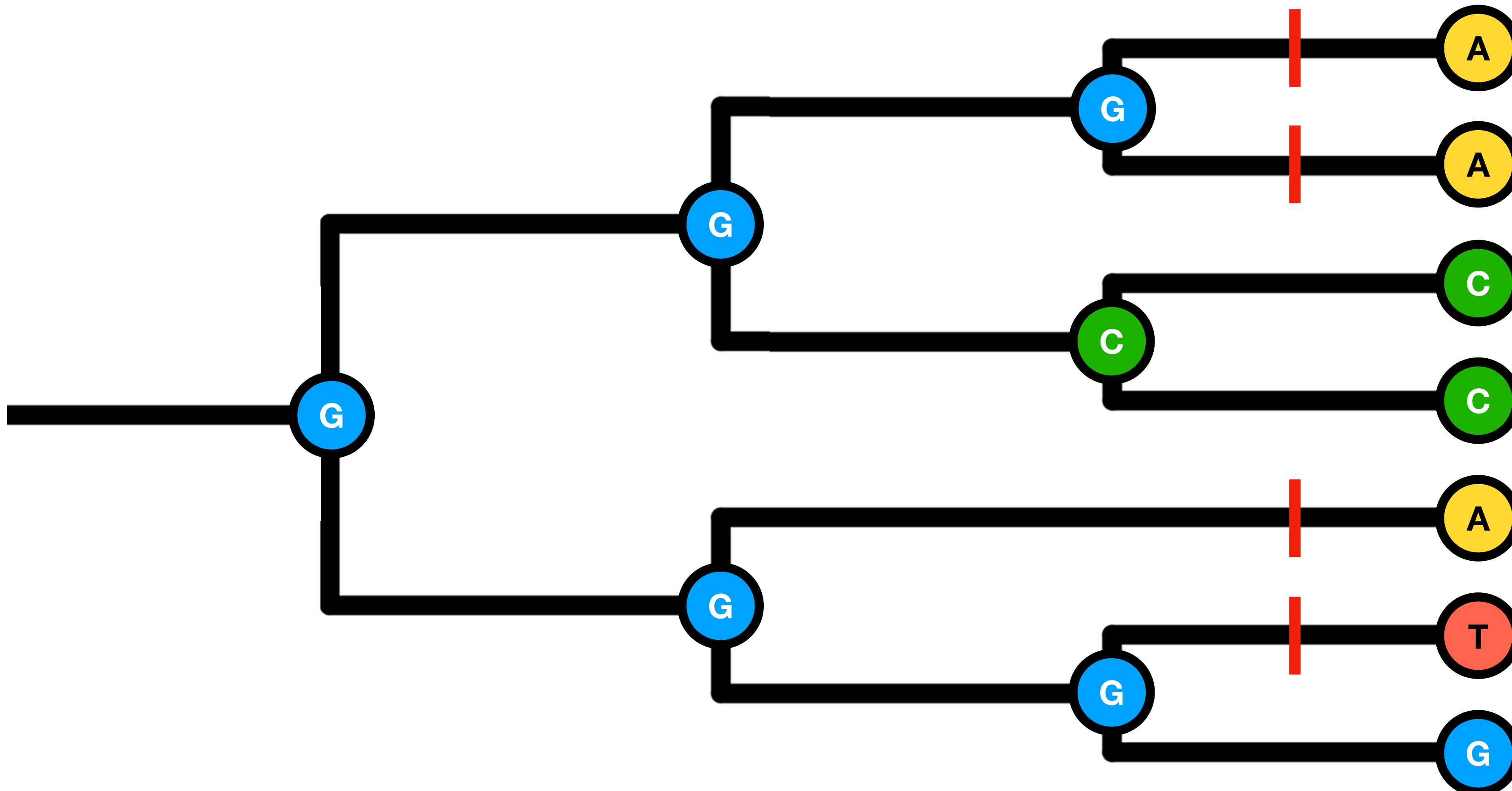
- **Environmental and Life-History Factors:**

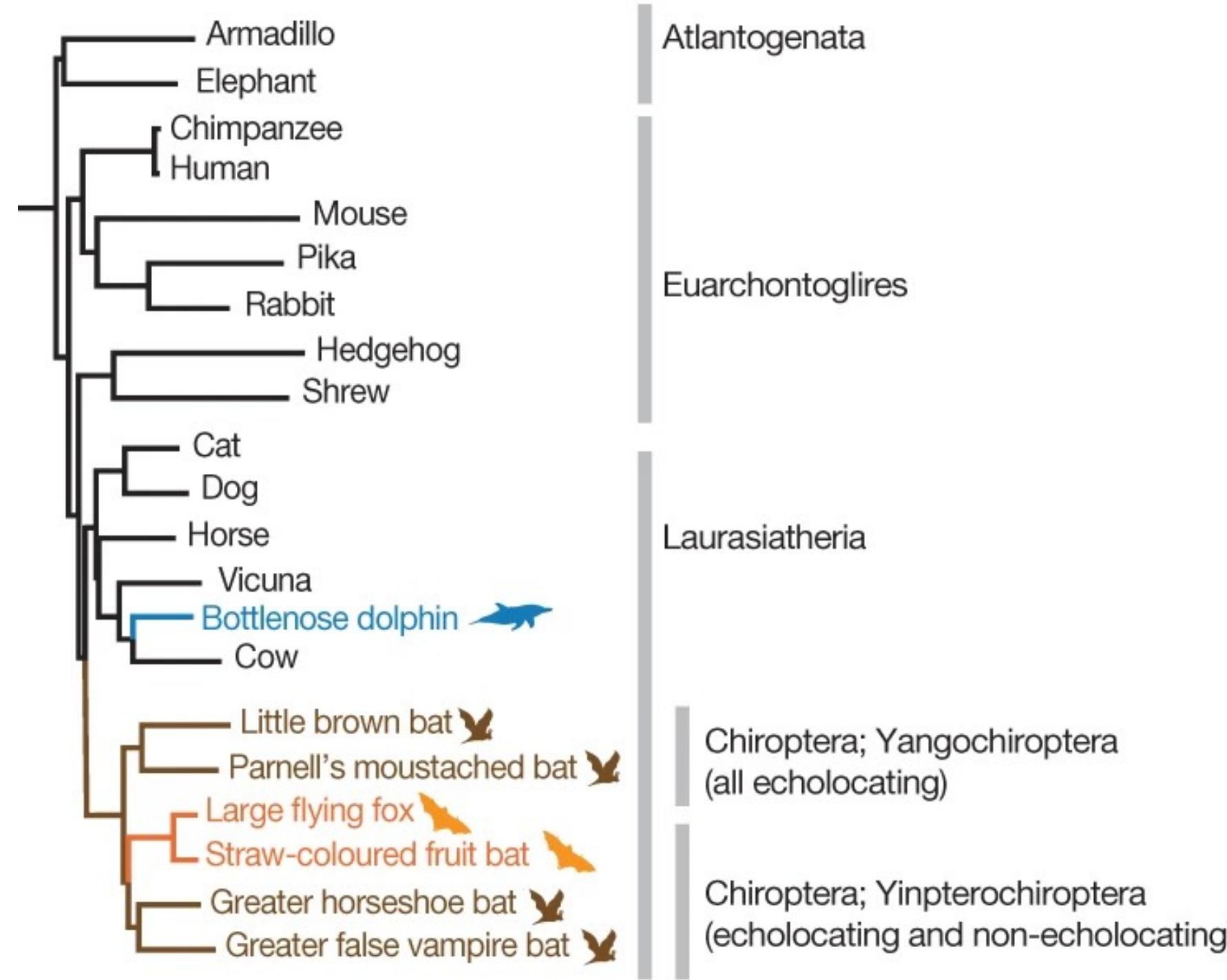
Organisms with shorter generation times or higher metabolic rates may accumulate mutations more quickly than those with longer generation times.

- **Practical Implications:**

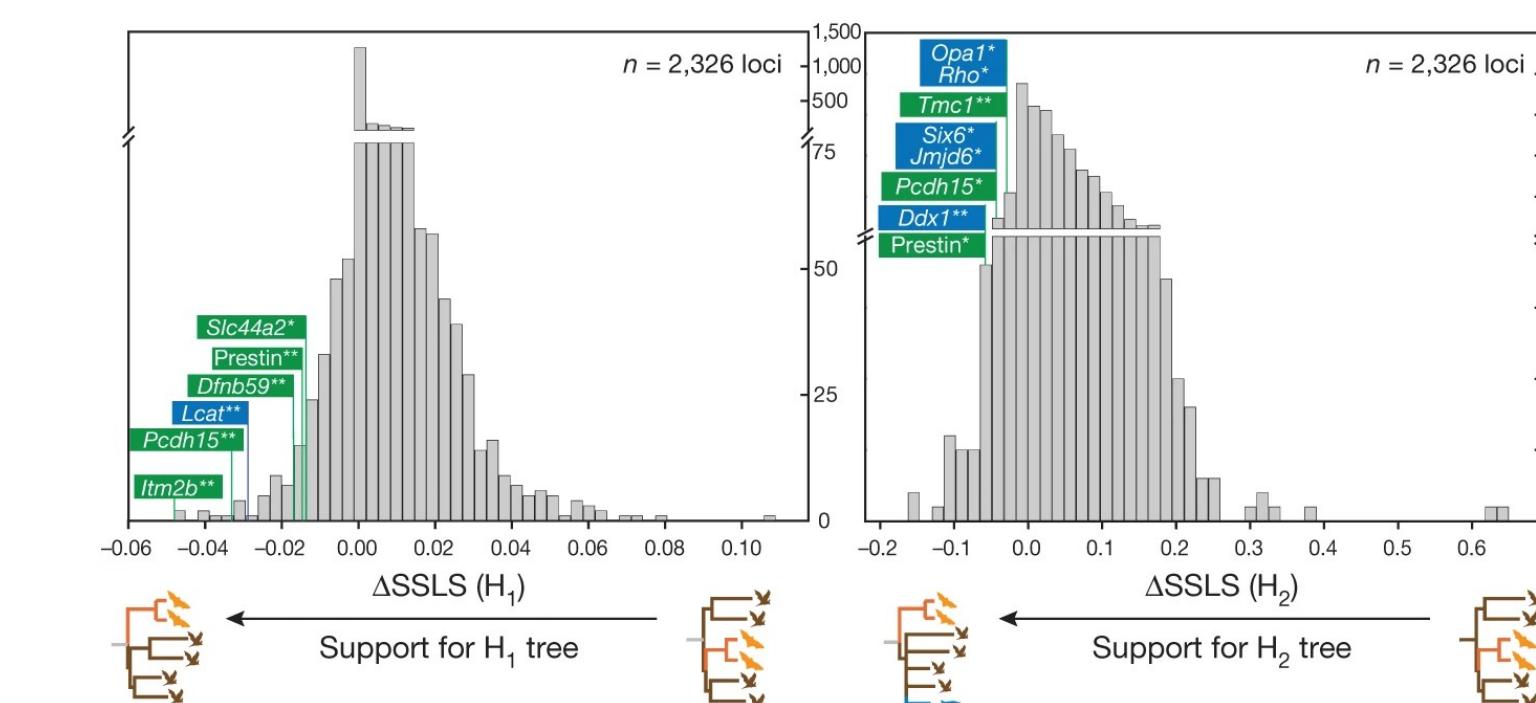
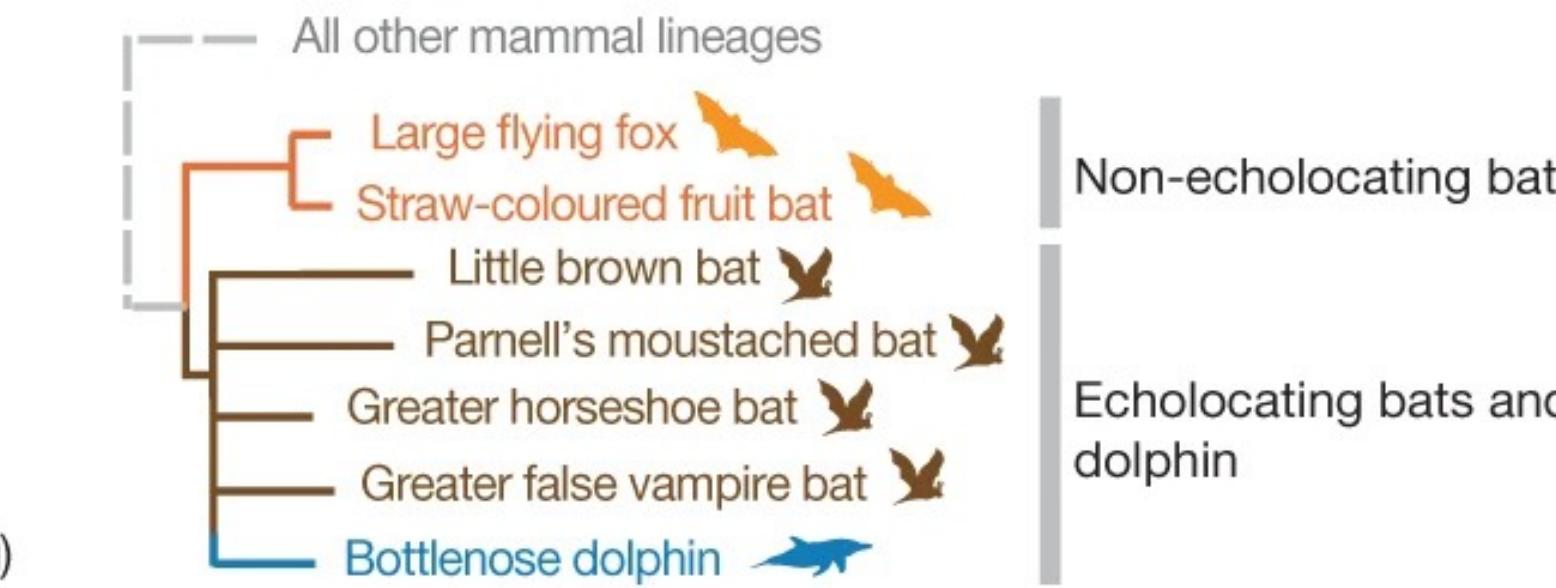
While the molecular clock provides a useful approximation for dating evolutionary events, calibrations using fossil records or known divergence times are essential to account for rate variability.

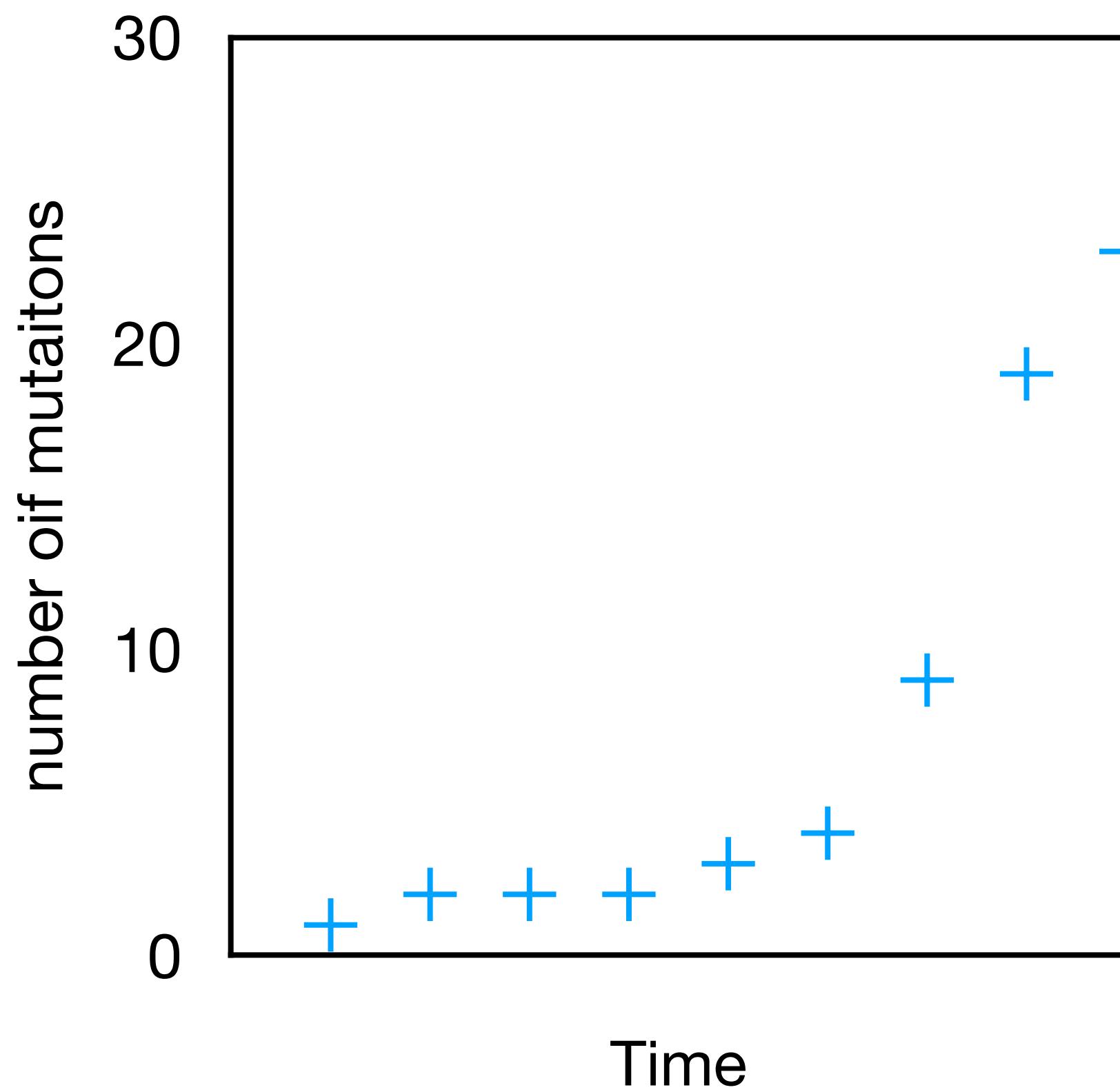
In summary, although genetic change can approximate a constant rate over evolutionary time, many factors cause deviations from a perfect molecular clock.

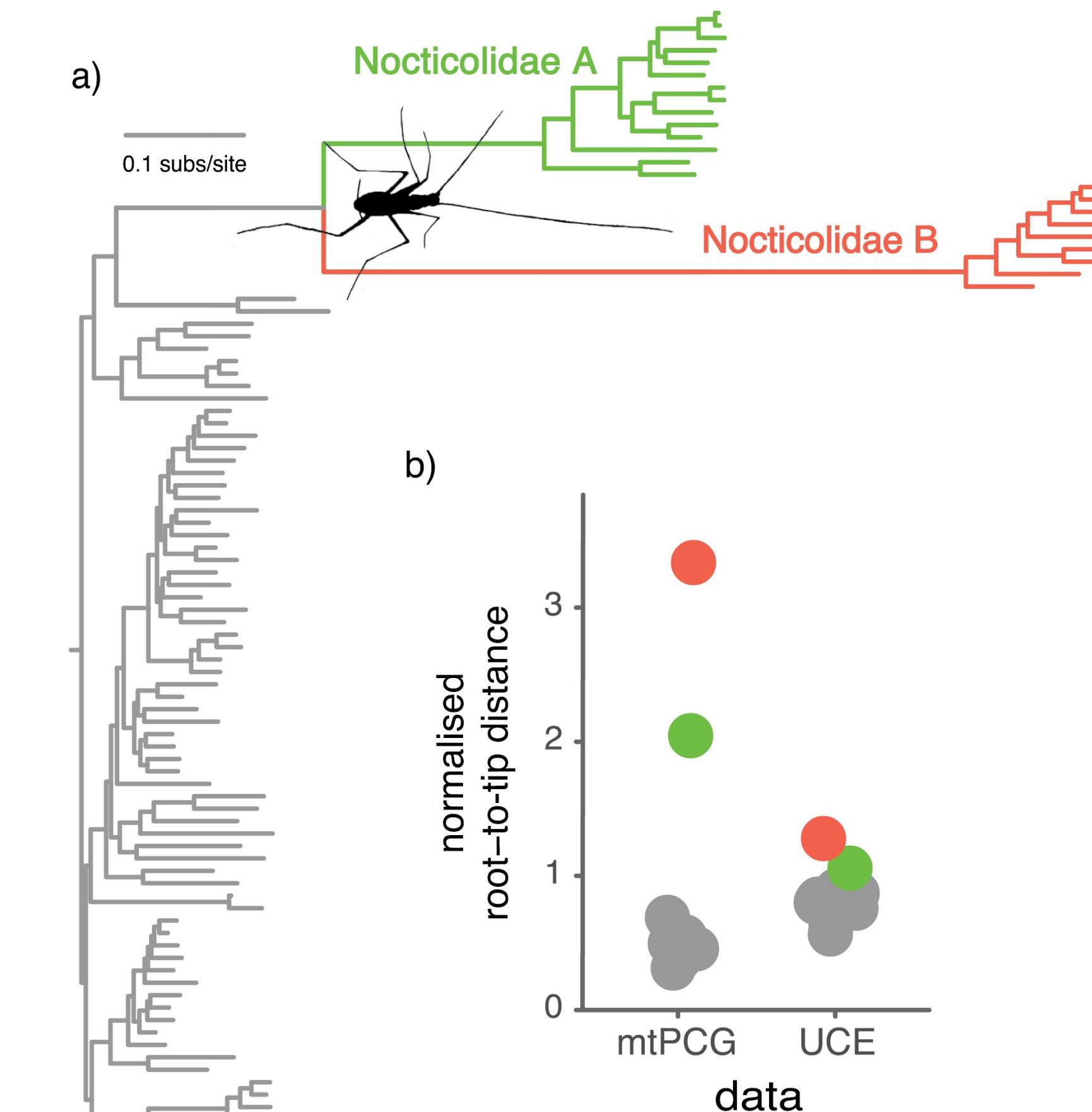
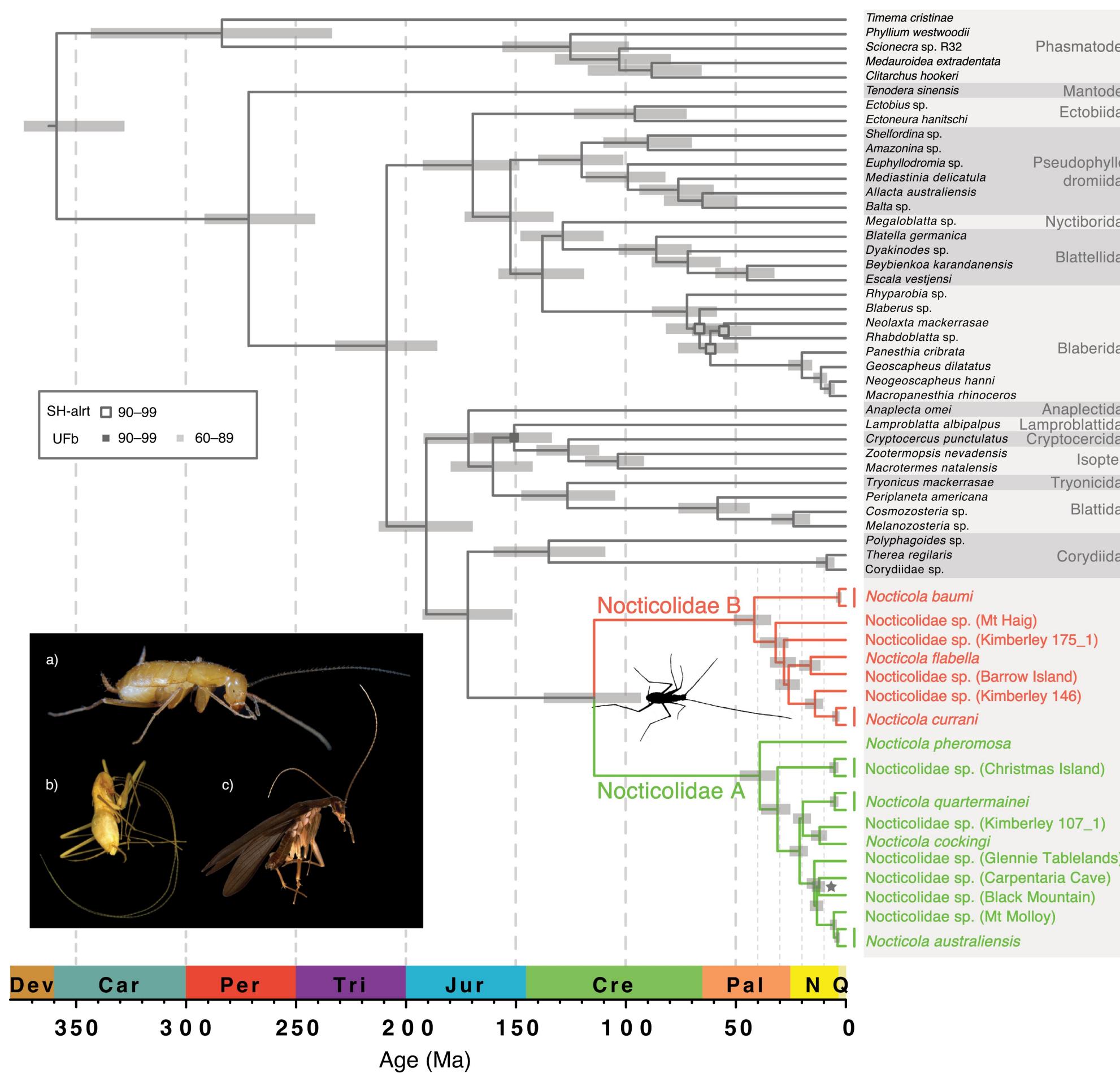


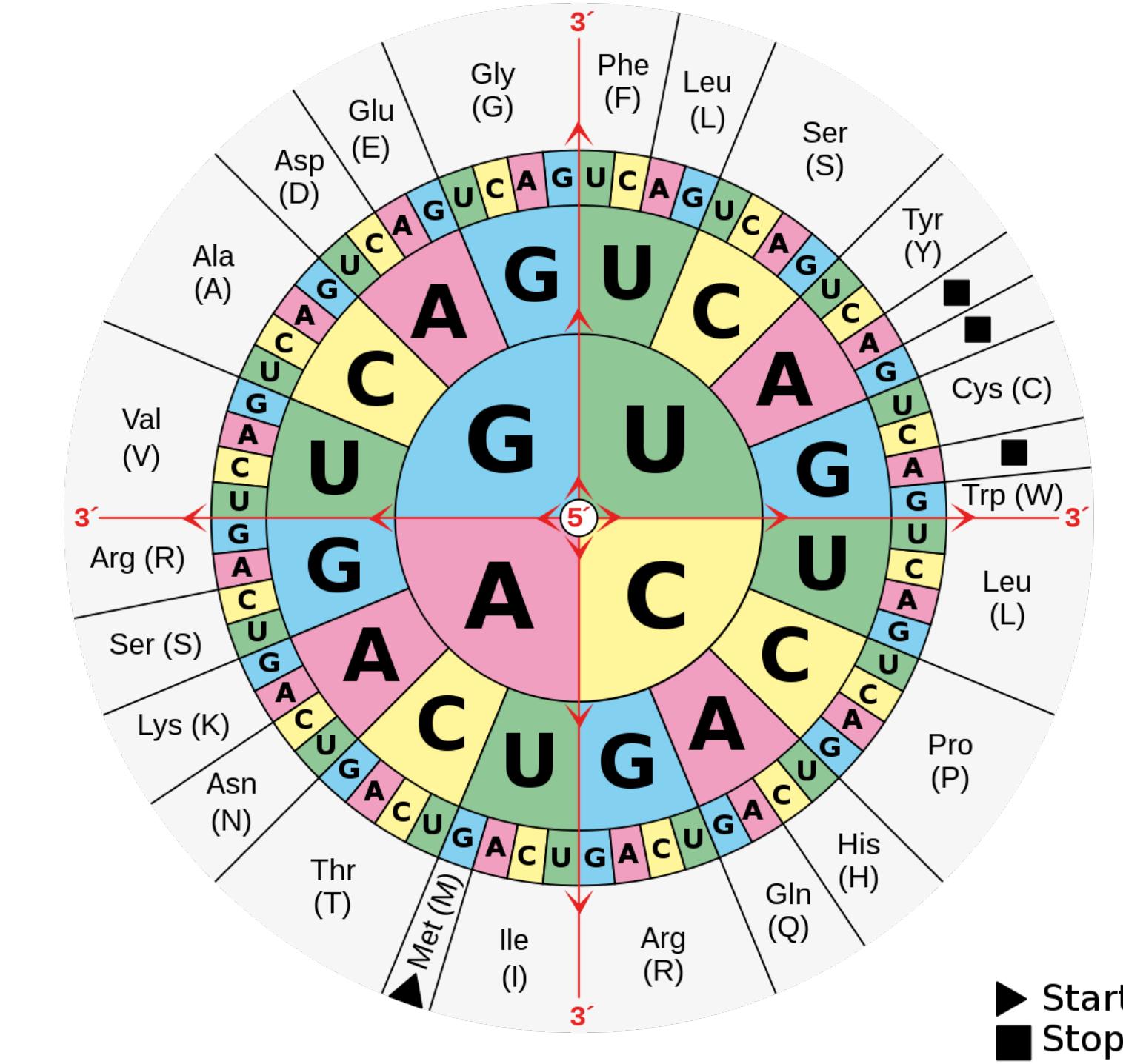
a Hypothesis H_0 (species tree)

Parker et al. Genome-wide signatures of convergent evolution in echolocating mammals. Nature (2013).

Hypothesis H_1 ('bat–bat convergence')**Hypothesis H_2 ('bat–dolphin convergence')**







Degeneracy of the genetic code: multiple codons encode the same amino acid, with variability mainly in the third codon position.

Differential evolutionary constraints:

- *1st & 2nd Positions:* stronger purifying selection; lower subst. rates.
- *Third Position:* synonymous substitutions, higher substitution rates.

FINISH