

Neutral Theory of Molecular Evolution

GCTGGT > GCUGGU > ALA-GLY
CGACCA

$$\int_a^b \epsilon \Theta^{\sqrt{17}} + \int \delta e^{i\pi} = -1$$

$\infty = \{2.7182818284\}$ λ χ^2 Σ \gg \approx λ $!$ $,$

μφερτυθιοπσδφγηξκλ

Neutral Theory of Molecular Evolution

Evolution is a two-step process:

1. Mutation (random)
2. Selection (non-random)

Detrimental mutation \Rightarrow negative selection \Rightarrow Mutation not seen

Beneficial mutation \Rightarrow positive selection \Rightarrow Mutation seen

Selectionist Views of What Drives Molecular Evolution

- Majority of all mutations are detrimental and not seen
 - Most observed substitutions have adaptive value
 - Classical school:
 - Single predominant version of gene (“wild type”) present in population
 - Natural selection rapidly fixates new, advantageous mutations
 - Balance school:
 - Appreciable amount of polymorphism in gene pool
 - Polymorphism maintained actively by natural selection (e.g., sickle cell anemia)
-

Neutralist Views of What Drives Molecular Evolution

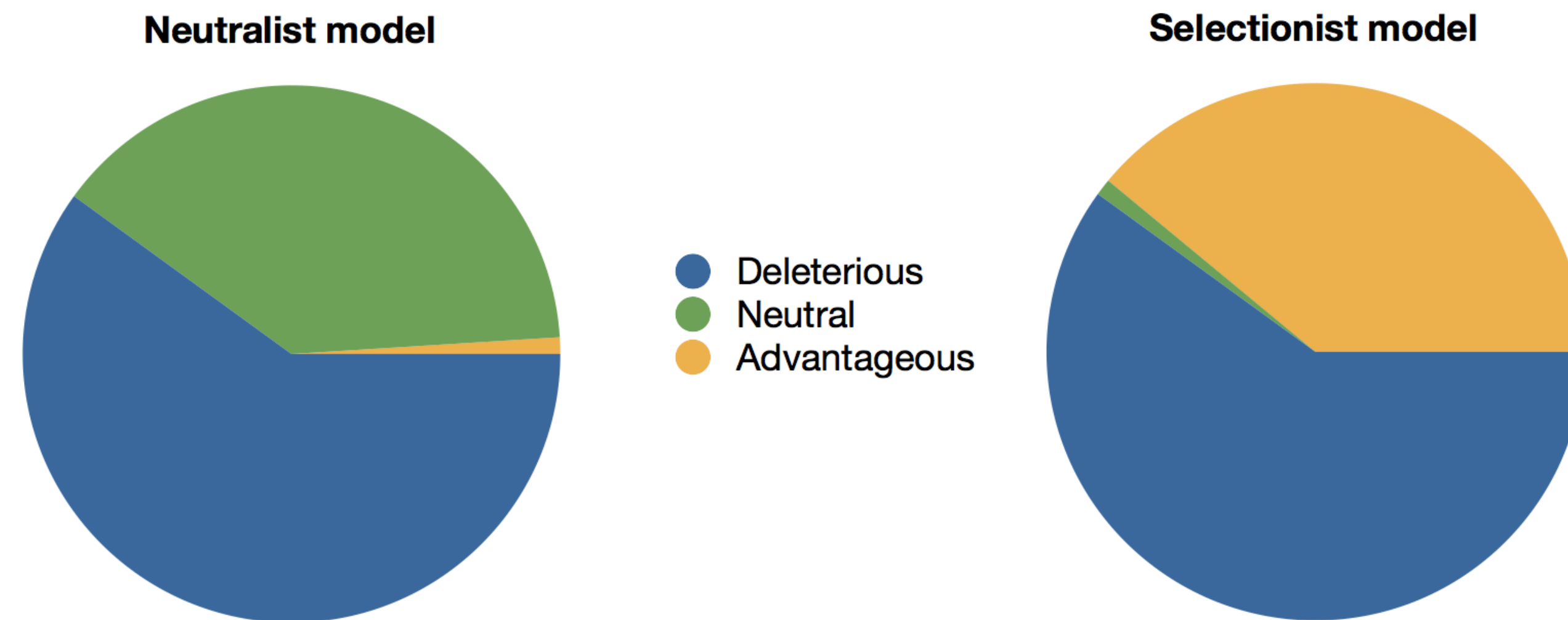
- Electrophoretic studies in 1960's showed much higher polymorphism than anticipated by either classical or balance school selectionists
- Kimura and others proposed the “Neutral Theory of Molecular Evolution”.

Detrimental mutation => negative selection => Mutation not seen

Neutral mutation => no selection => Mutation may be seen (genetic drift)

Beneficial mutation => positive selection => Mutation seen

Difference Between Selectionist and Neutralist Views of Evolution



Fraction of random mutations assumed to be deleterious, neutral, and advantageous

- Selectionist view:
 - Most observed mutations represent functional innovation
- Neutralist view:
 - Most observed mutations represent conservative changes, changes in unimportant regions

All Agree that Adaptations are Caused by Natural Selection



Gekko camouflaged on branch

[image source: wikimedia](#)

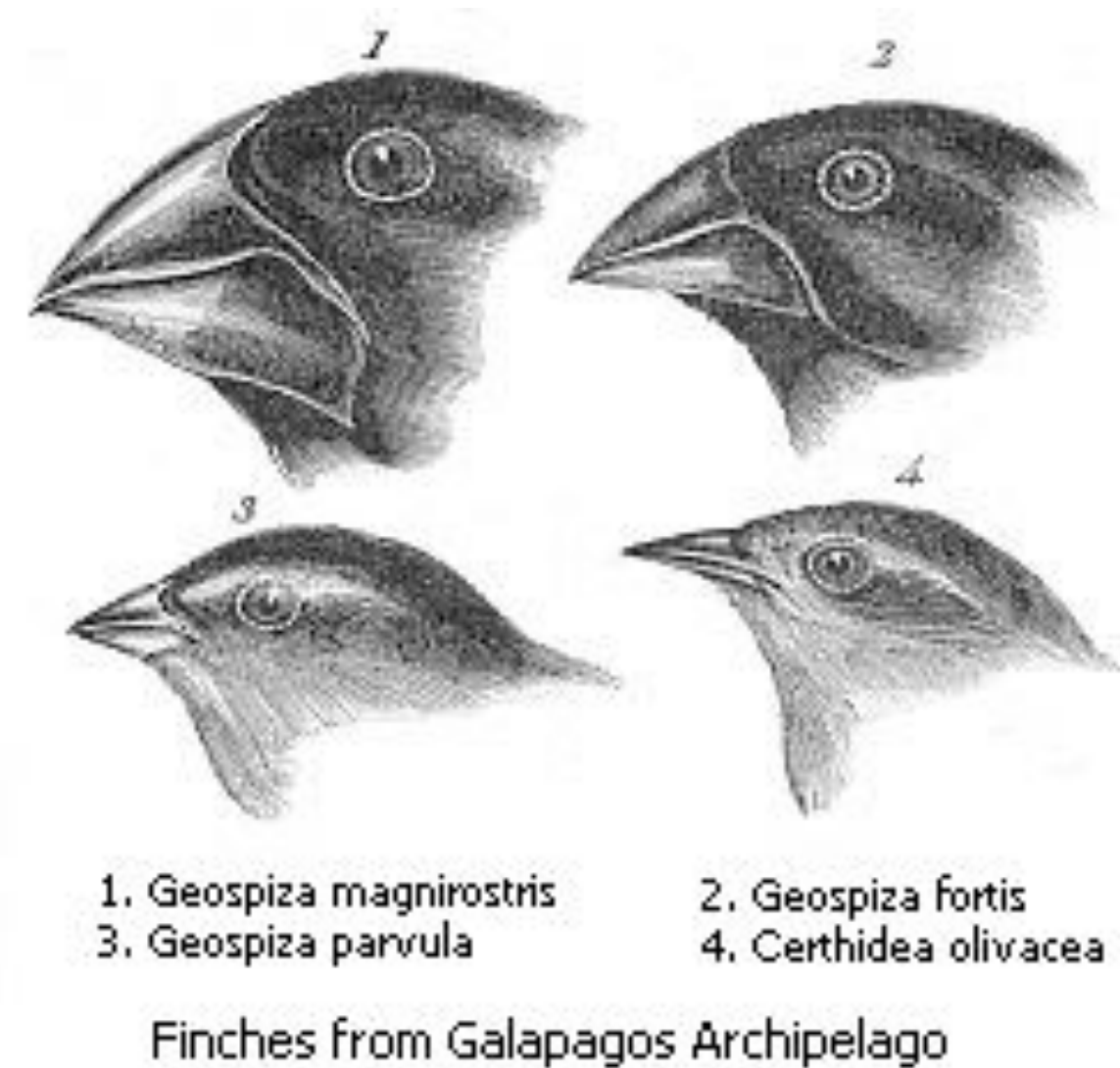
All Agree that Adaptations are Caused by Natural Selection



Gekko camouflaged on branch

[image source: wikimedia](#)

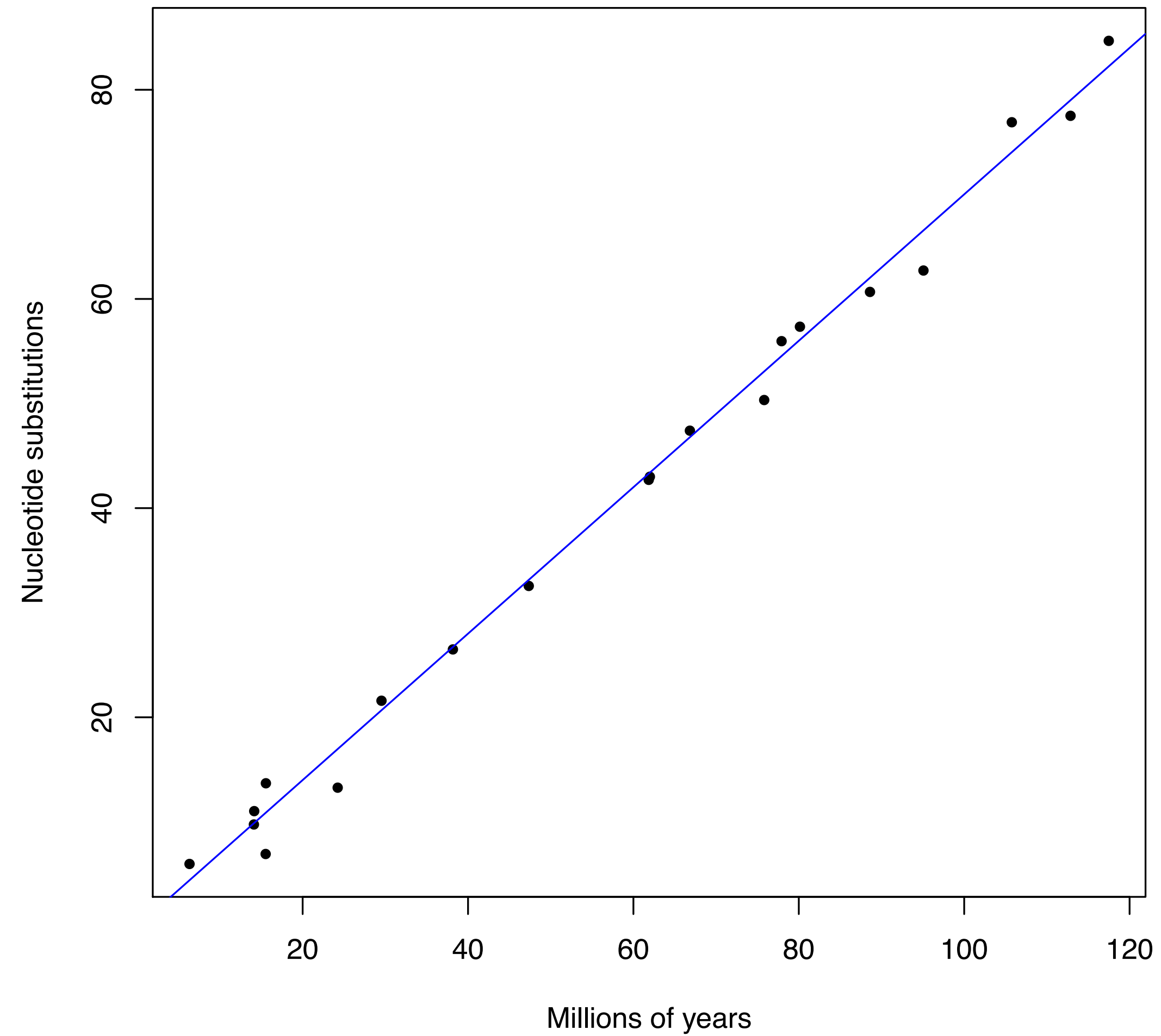
All Agree that Adaptations are Caused by Natural Selection



Galapagos finches with beak shapes suited to preferred food.

[image source: wikimedia](#)

The molecular clock

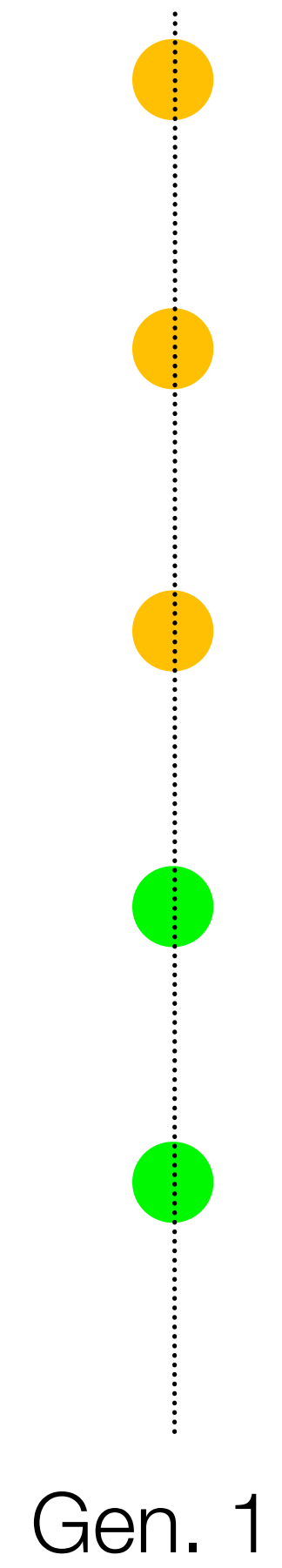


Genetic Drift

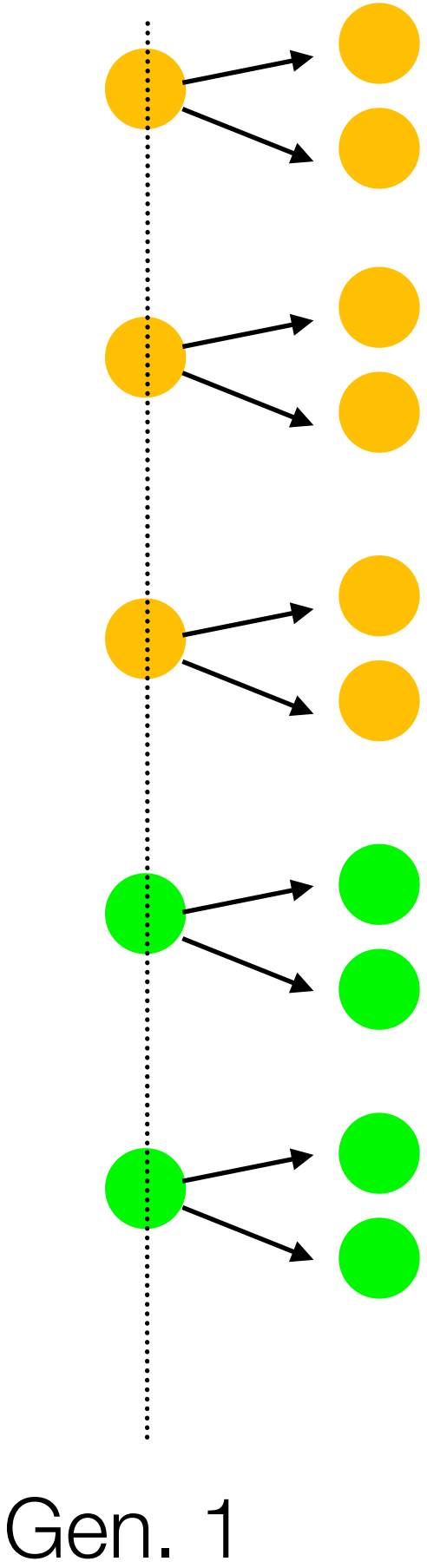
GCTGGT > GCUGGU > ALA-GLY
CGACCA

$$\int_a^b \epsilon \Theta^{\sqrt{17}} + \int \delta e^{i\pi} = -1$$
$$\infty = \{2.7182818284\}$$
$$\chi^2$$
$$\Sigma!$$
$$\lambda$$

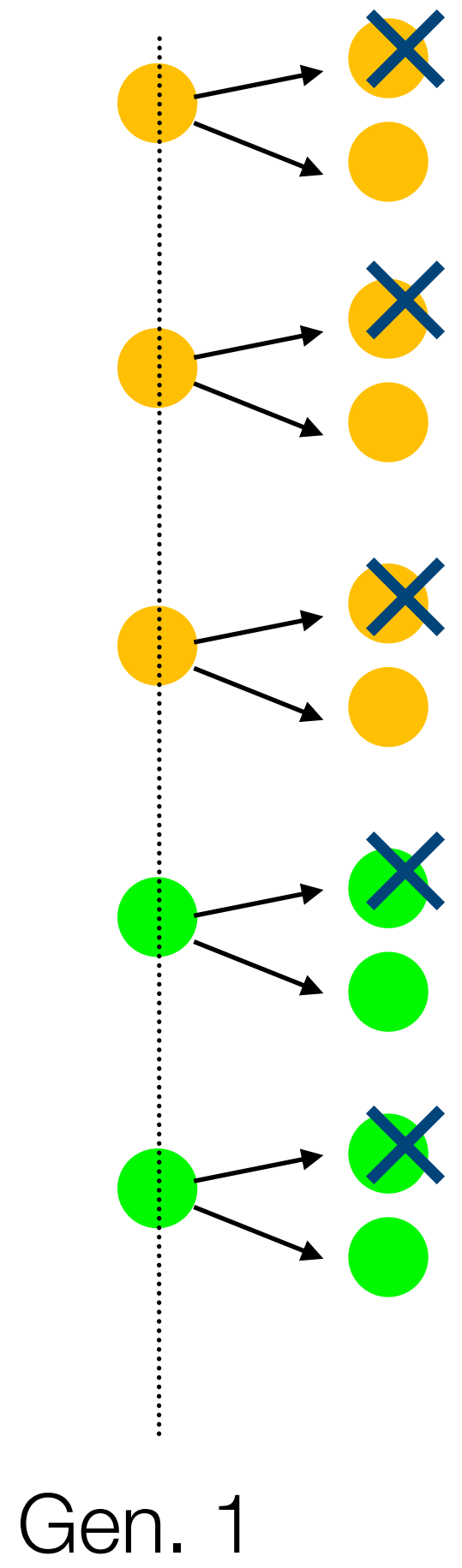
Genetic drift



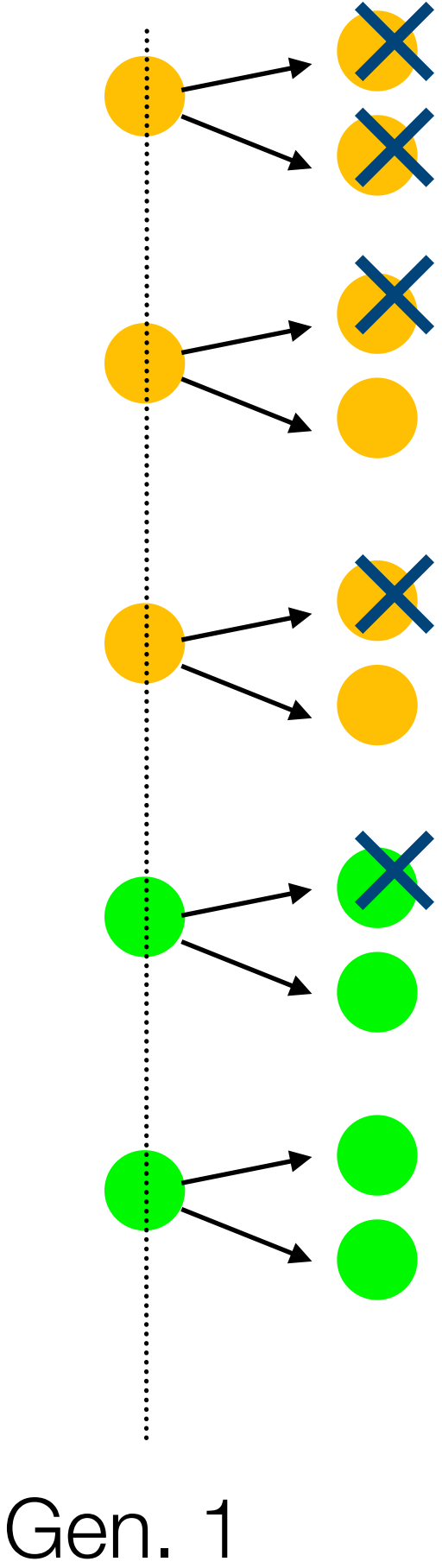
Genetic drift



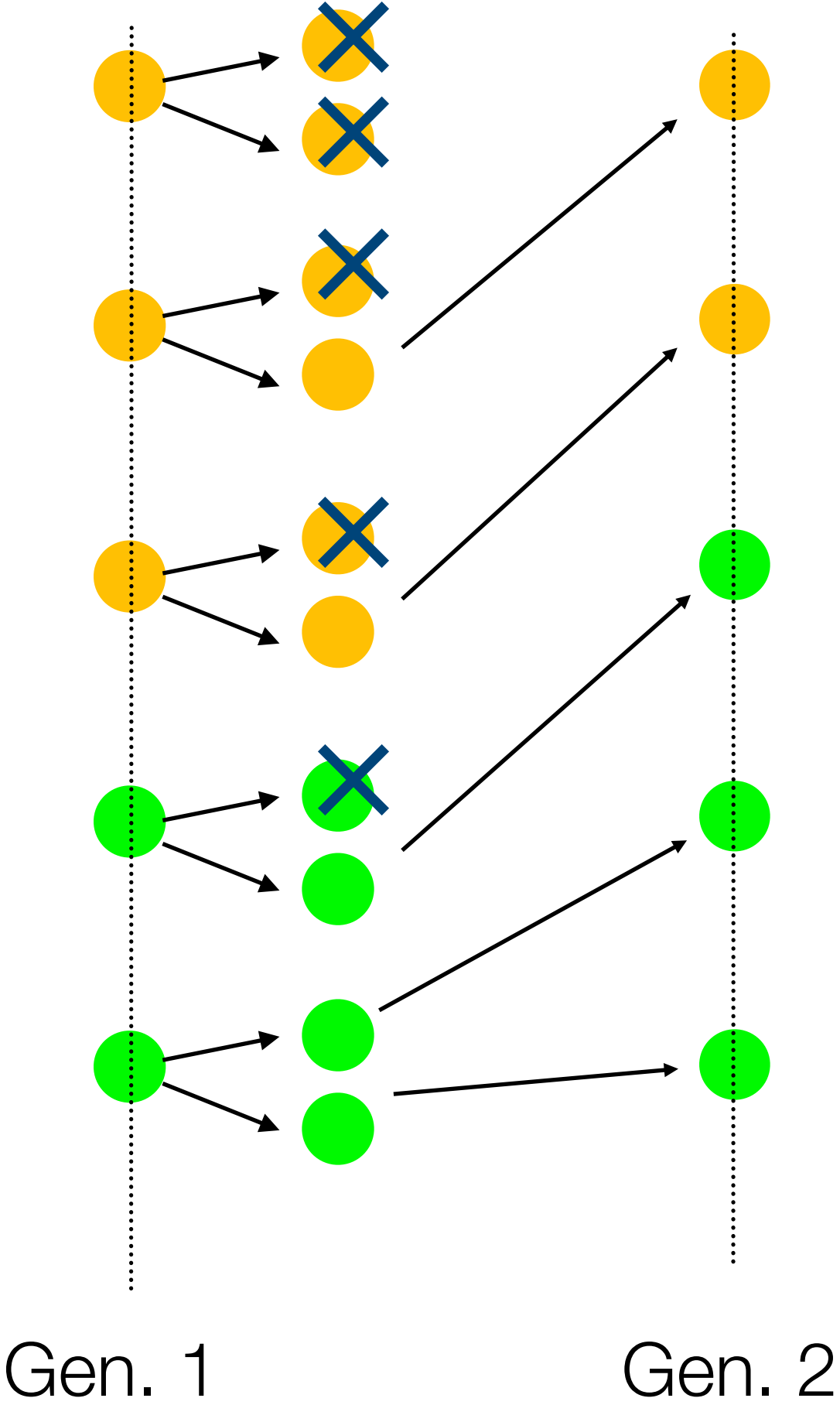
Genetic drift



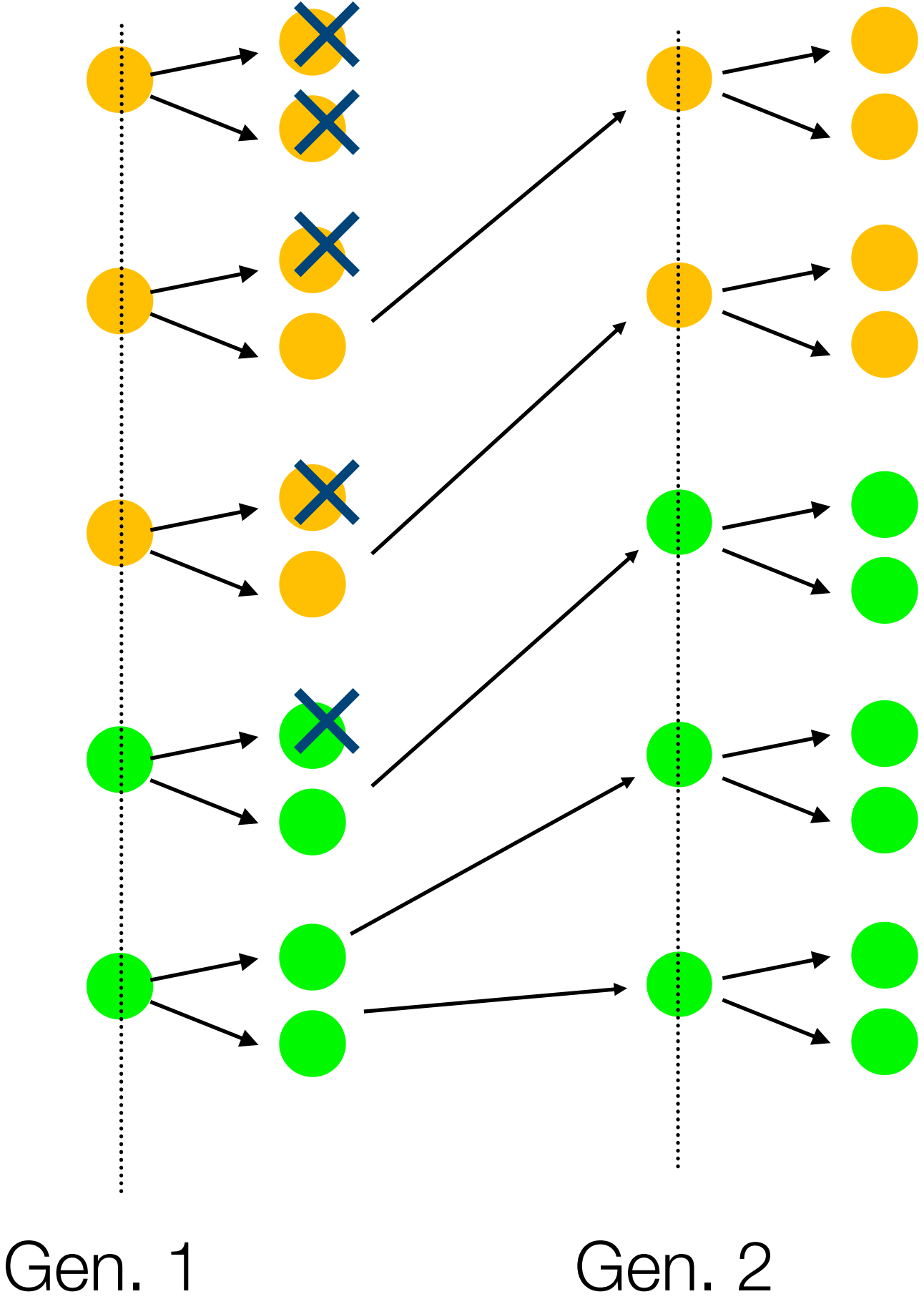
Genetic drift



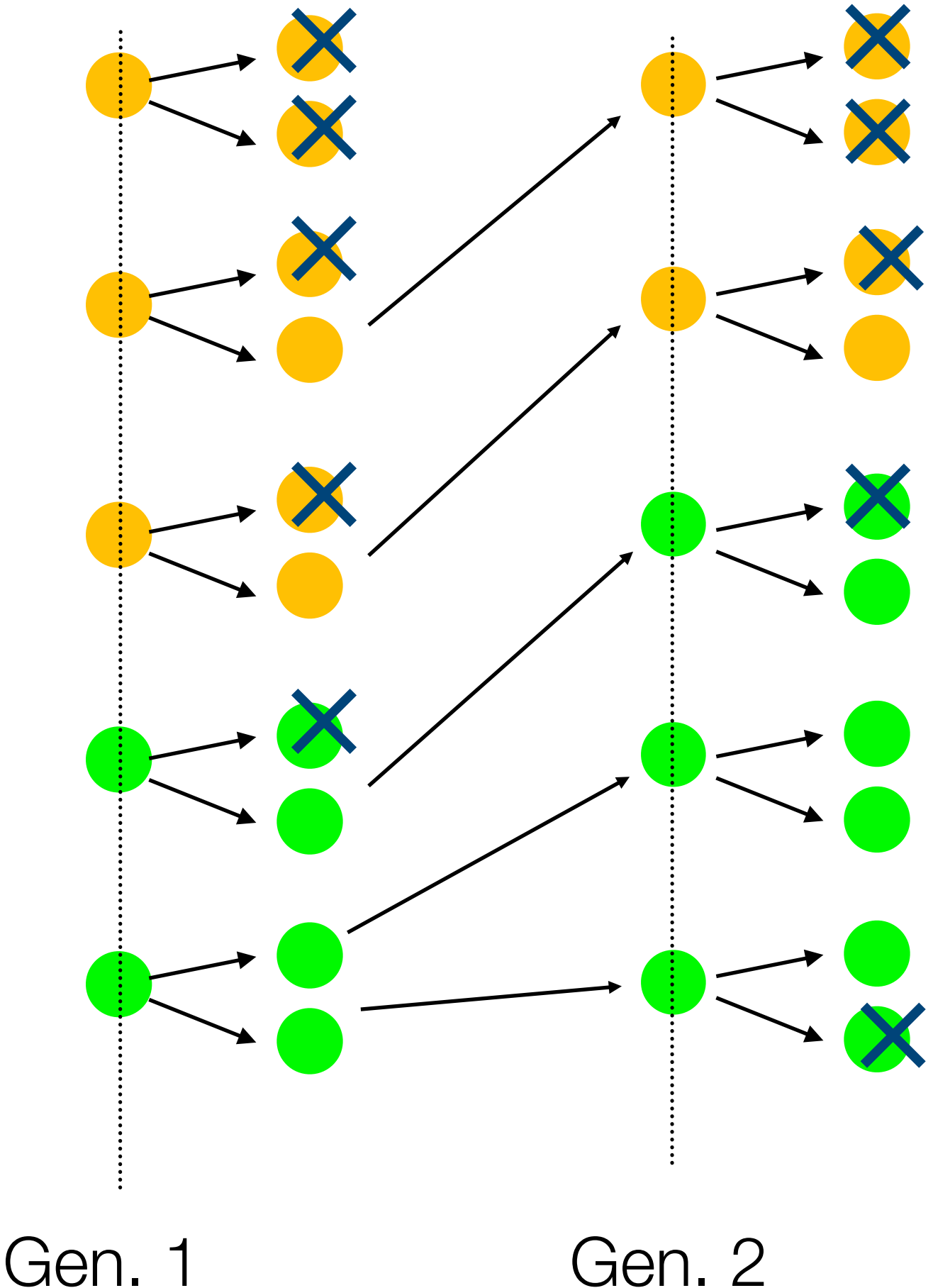
Genetic drift



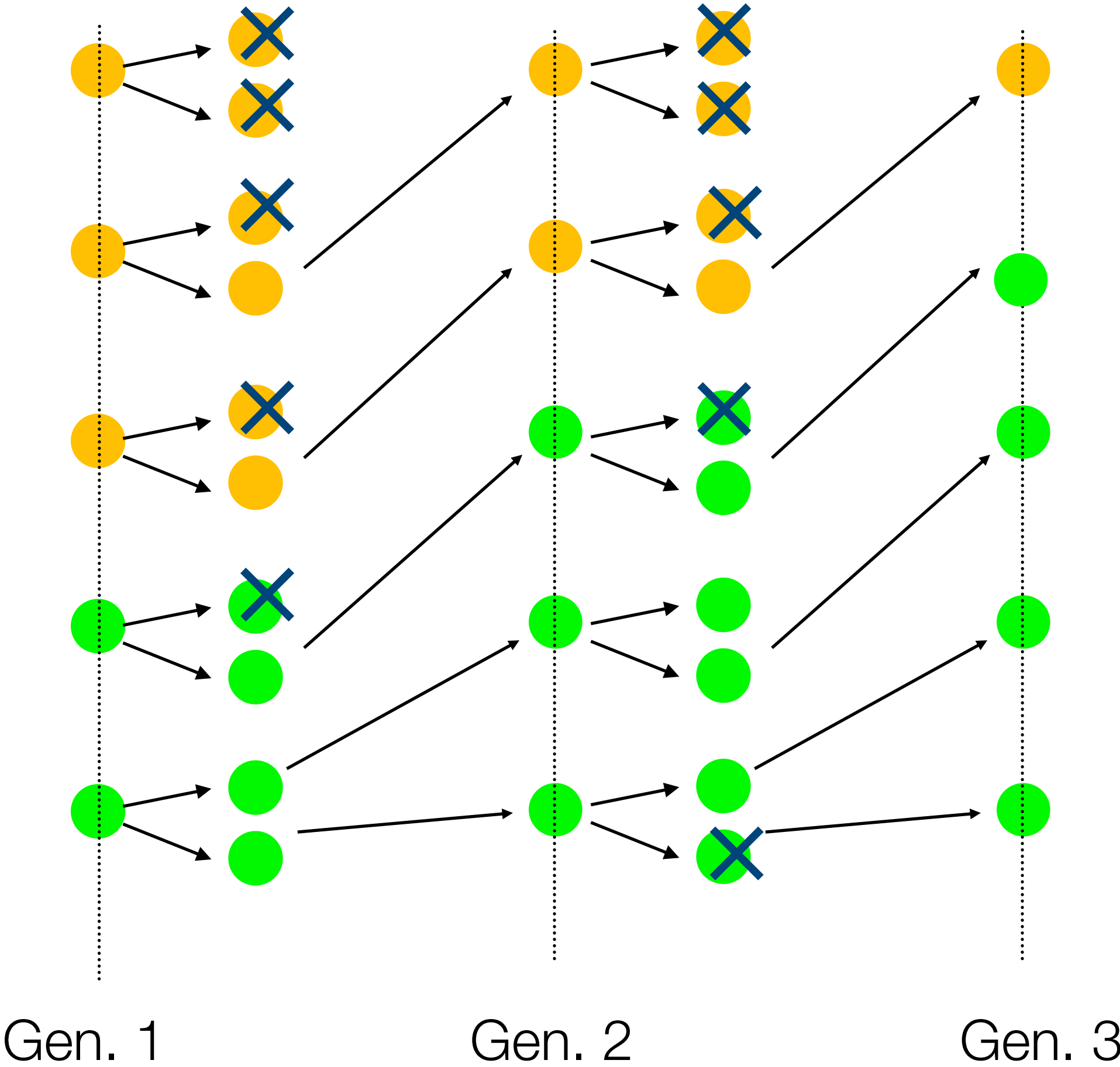
Genetic drift



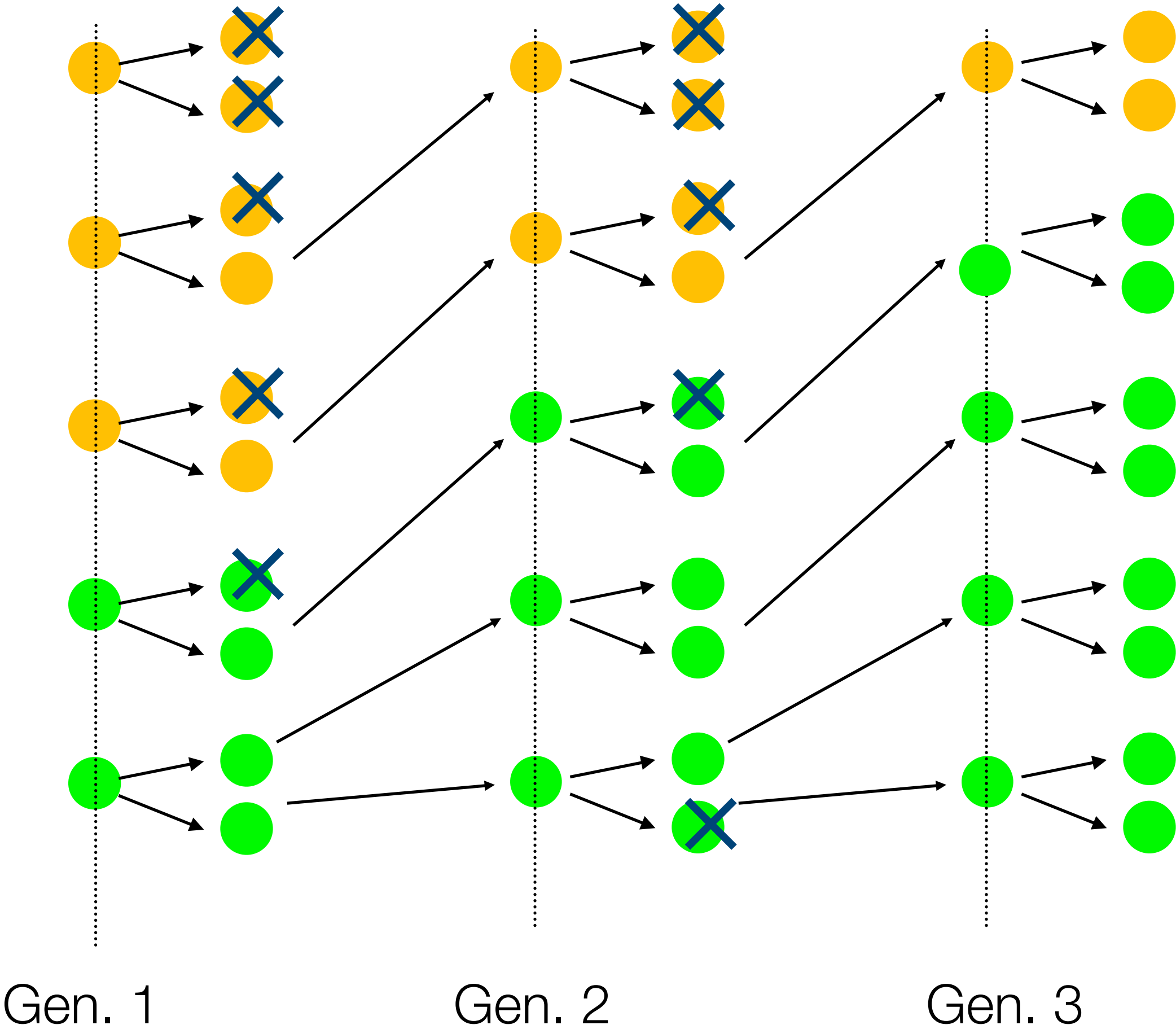
Genetic drift



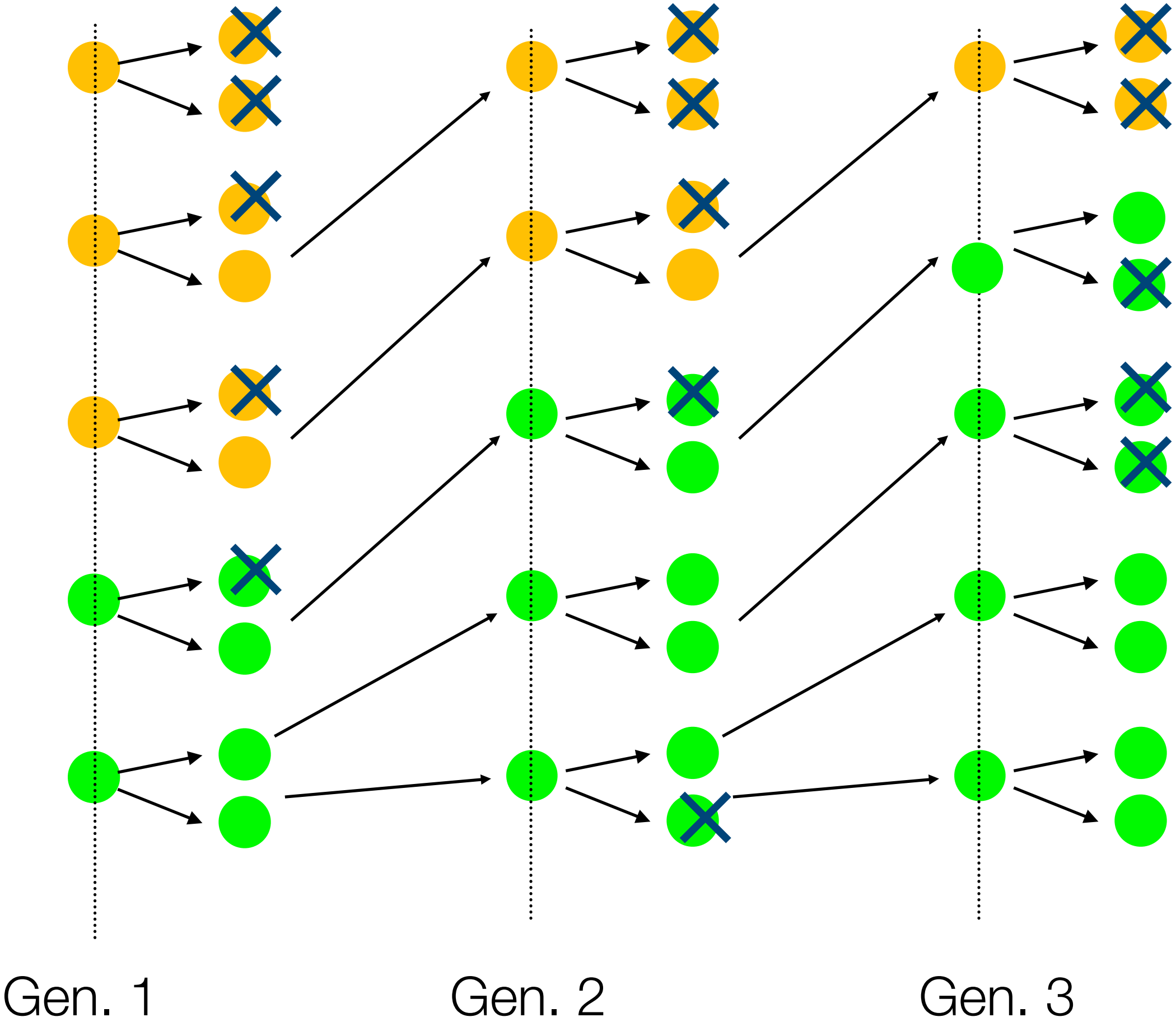
Genetic drift



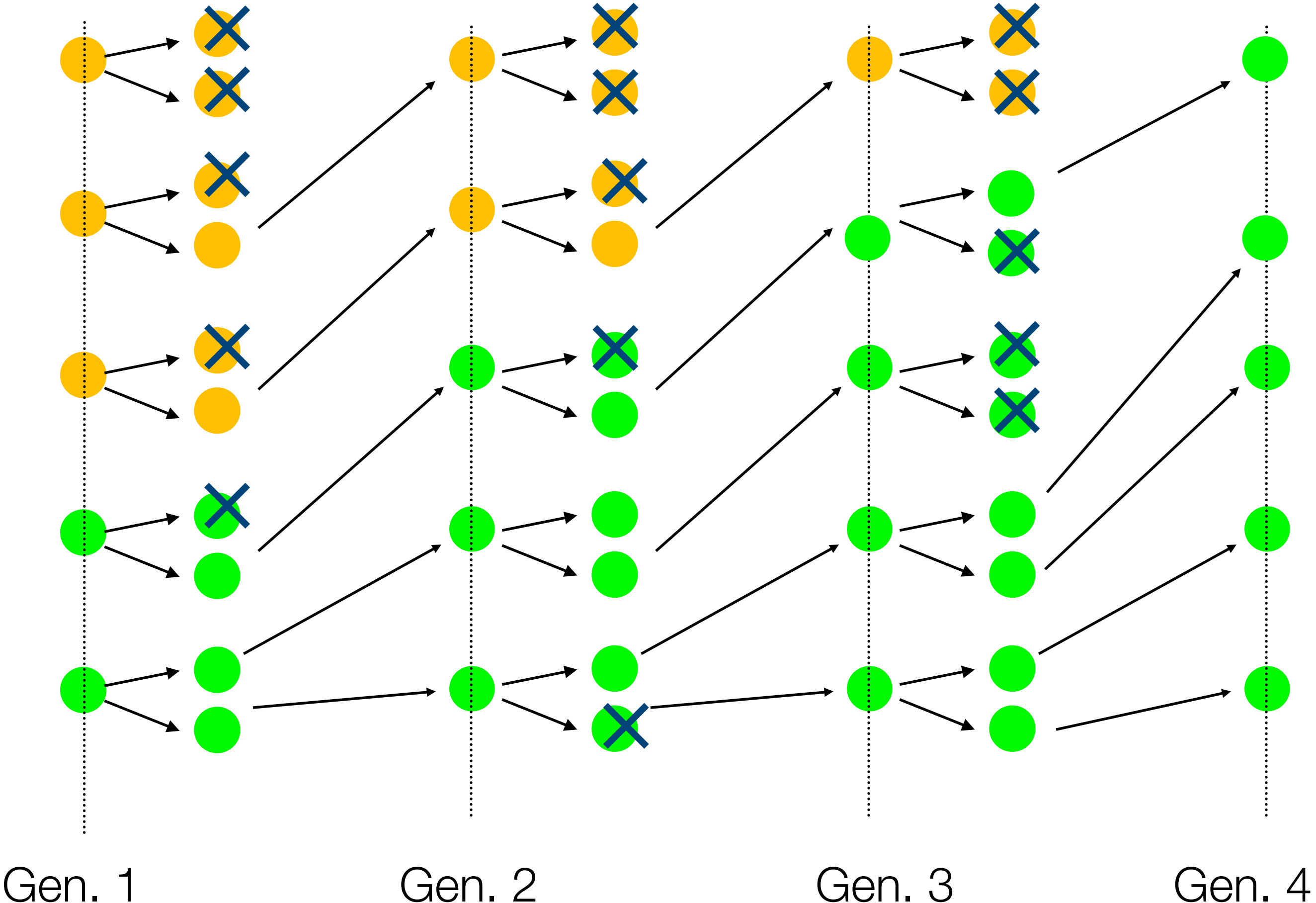
Genetic drift



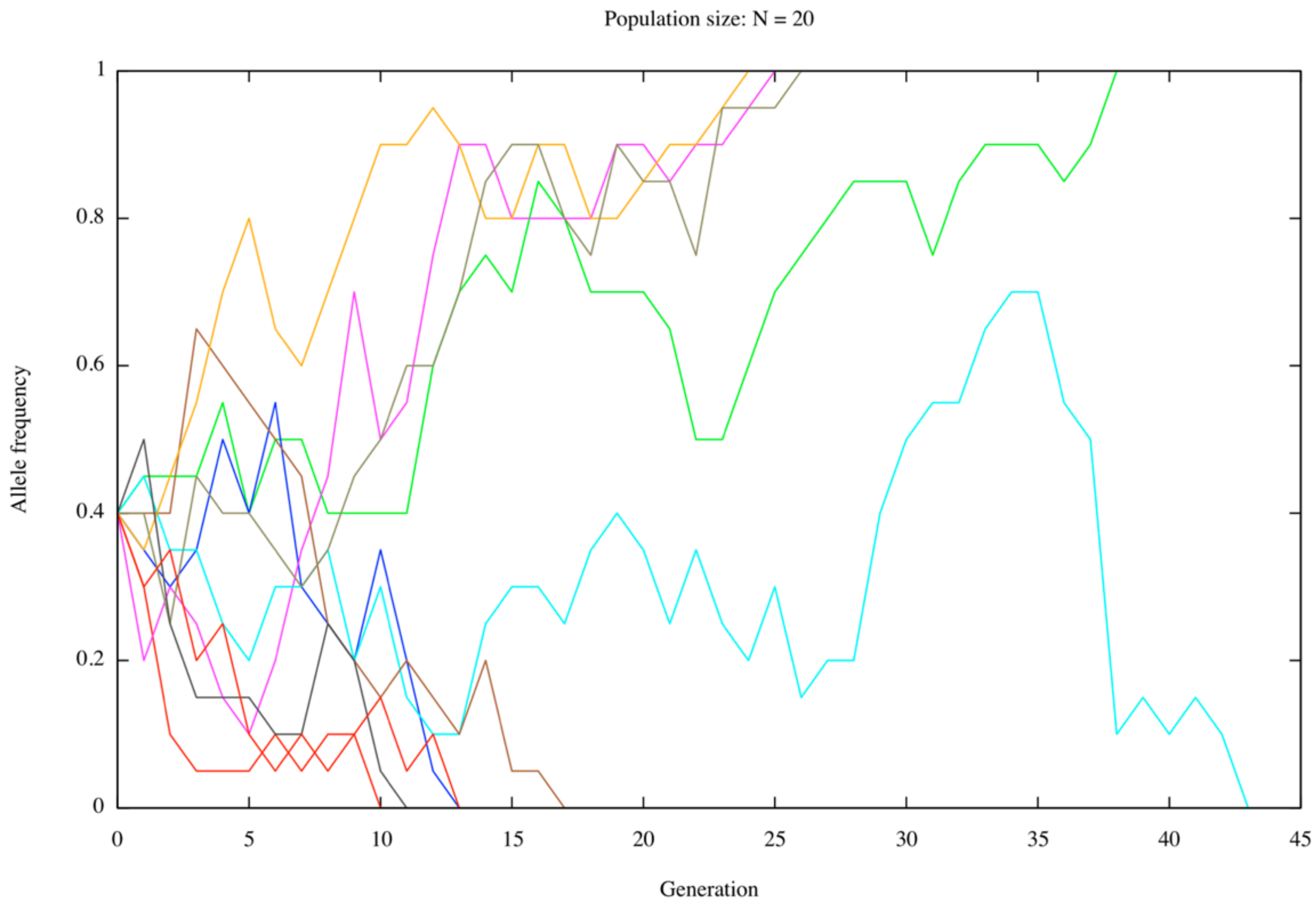
Genetic drift



Genetic drift



Genetic drift



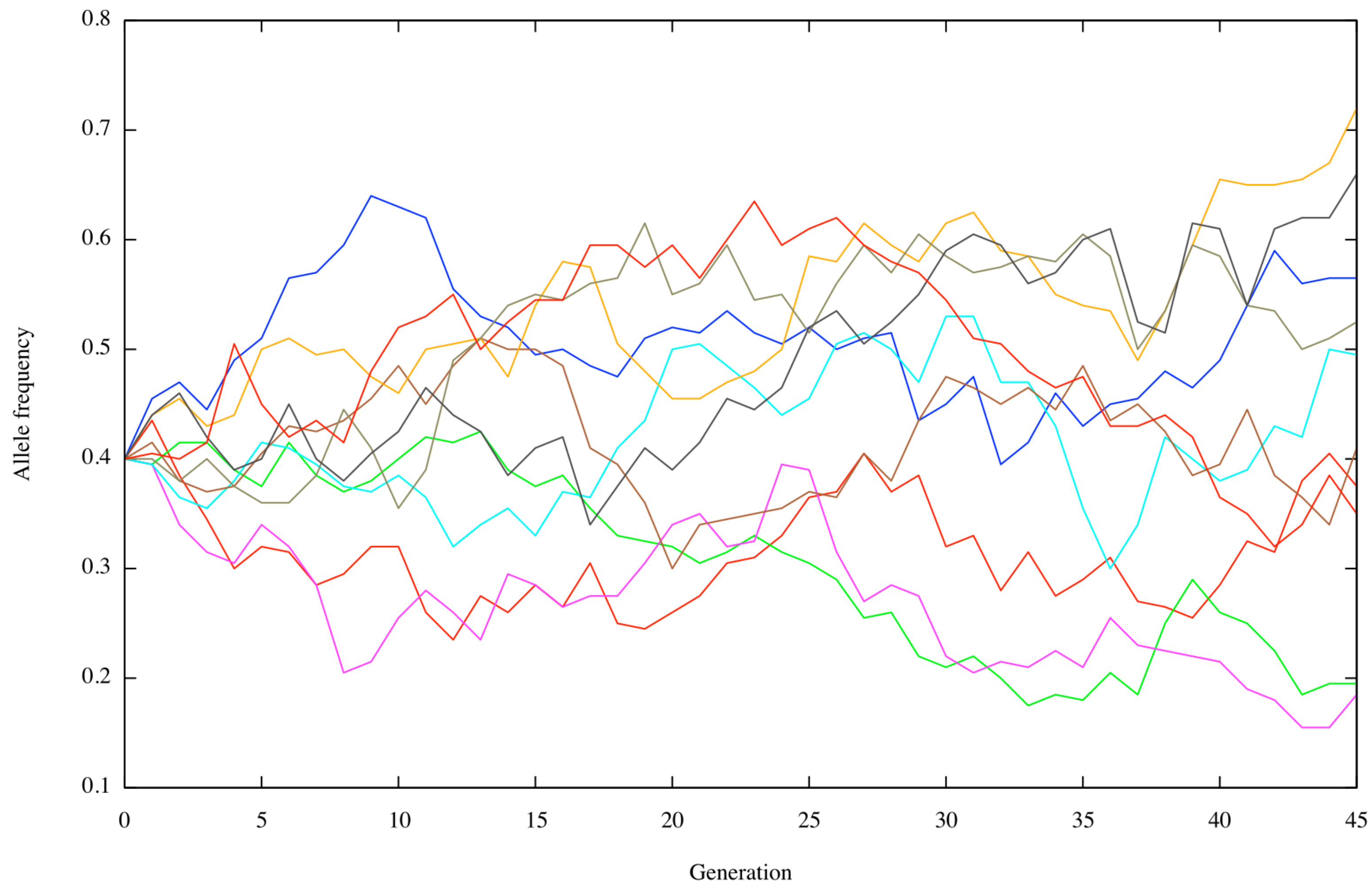
Alleles will eventually reach a frequency of 0 or 1

Genetic diversity decreases

Effect is more strongly felt in small populations

Genetic drift

Population size: $N = 200$



Alleles will eventually reach a frequency of 0 or 1

Genetic diversity decreases

Effect is more strongly felt in small populations

Drift and mutation

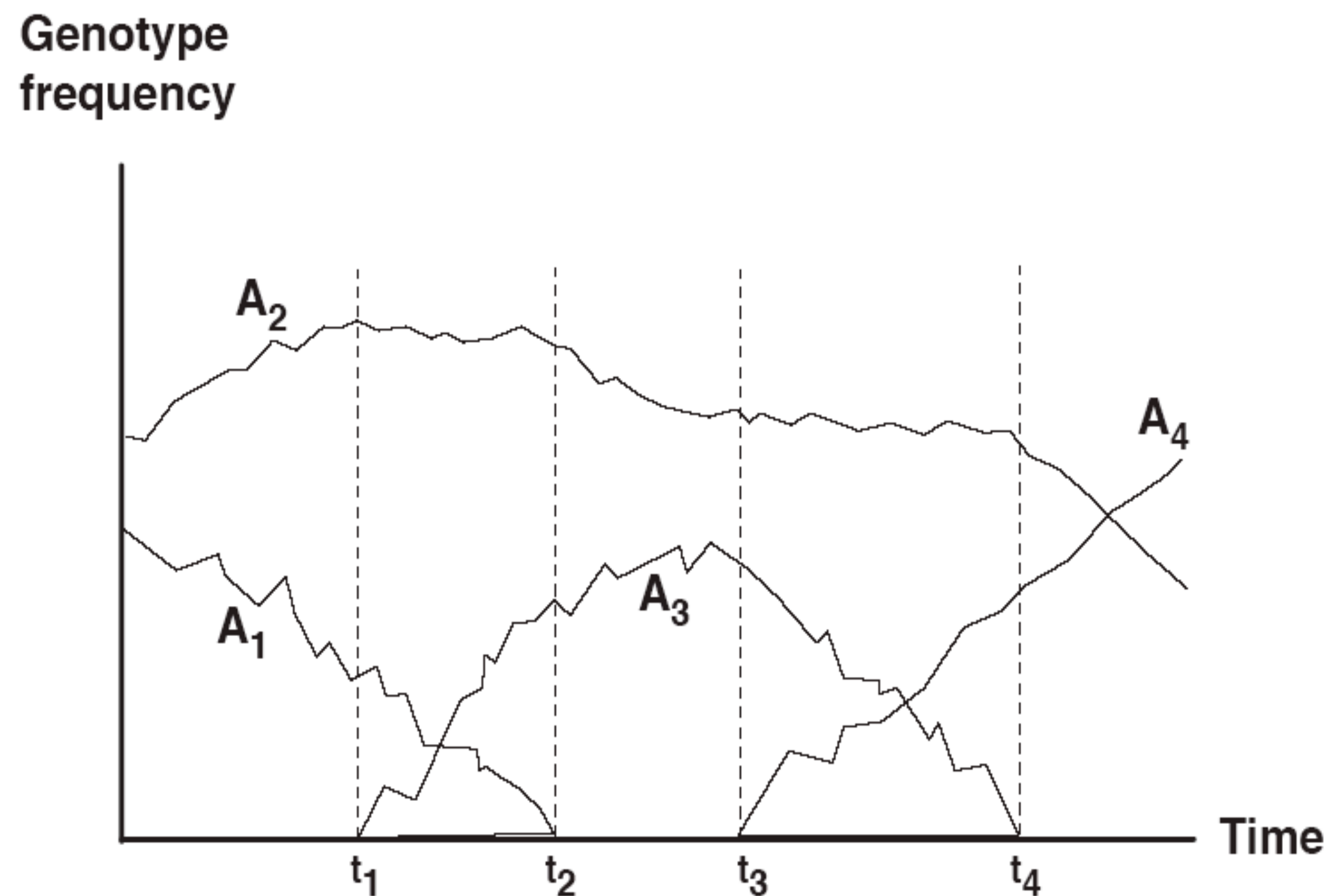
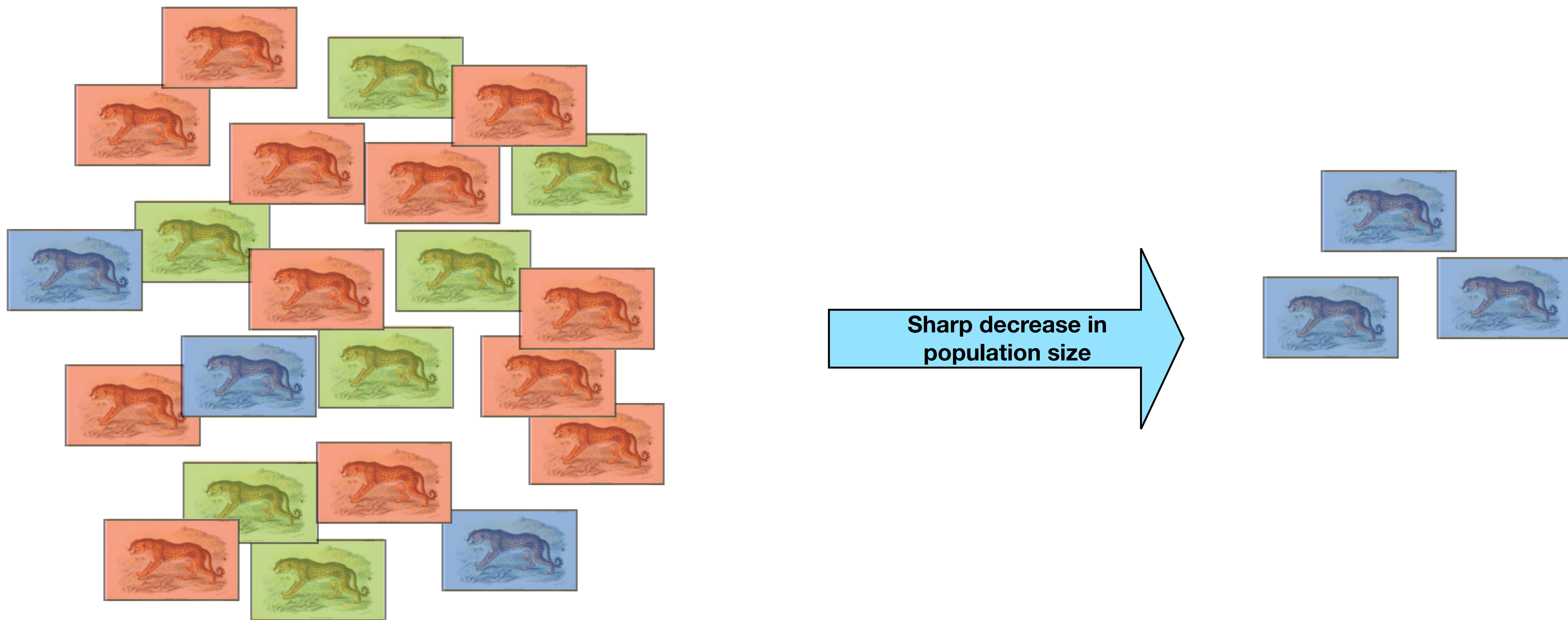


Figure 1.9: Genetic drift is counterbalanced by mutation resulting in the dynamic maintenance of genetic variation at a locus.

Bottleneck effect



- Change in allele frequencies when population size sharply decreases.
- e.g., due to natural disaster

Bottleneck effect



image source: [wikimedia](#)

- Cheetahs: Almost no genetic diversity
- Due to population bottleneck about 10,000 years ago

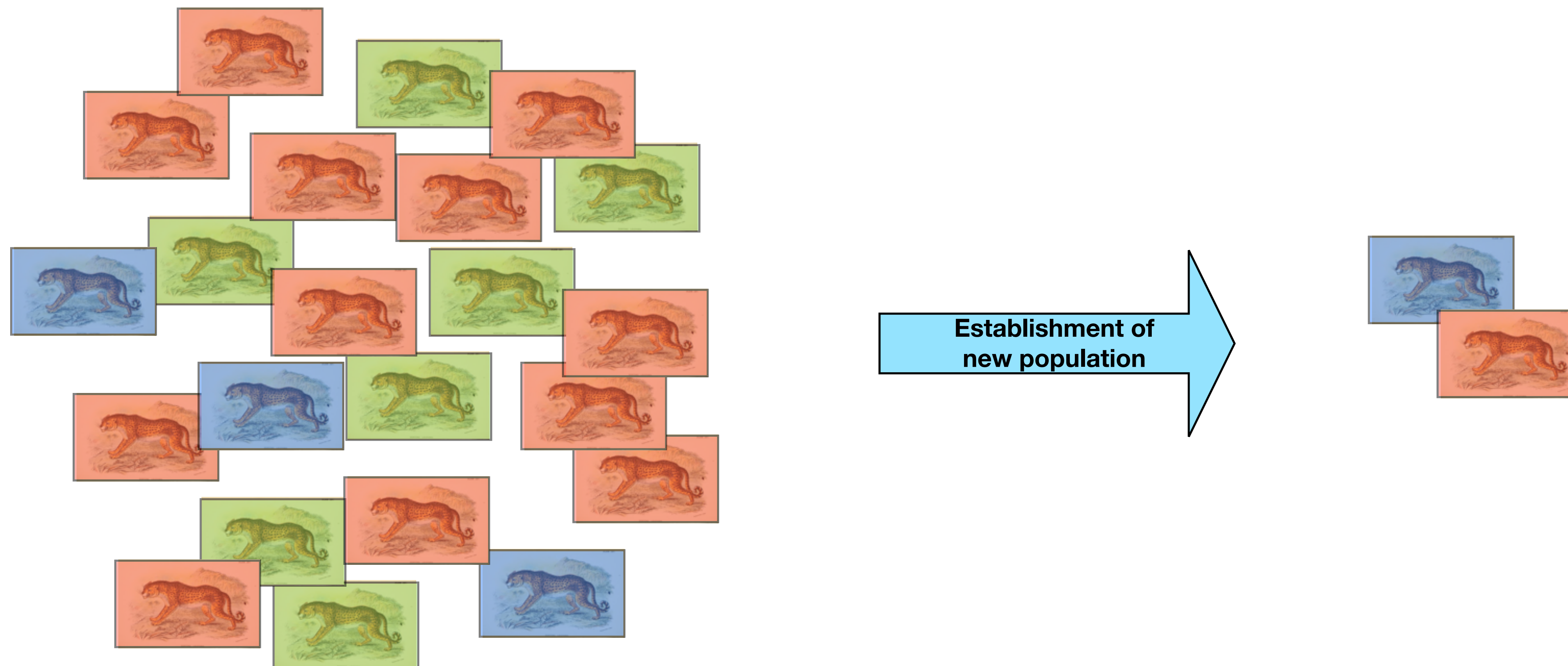
Bottleneck effect



[image source: wikimedia](#)

- Northern Elephant Seal
- Reduced to 20 individuals in 1896
- Now 30,000 individuals, with no detectable genetic diversity

Founder effect



- Change in allele frequencies when a new population arises from only a few individuals.
 - e.g., only a few fish are introduced into a lake.
 - e.g., only a few birds make it to an island.

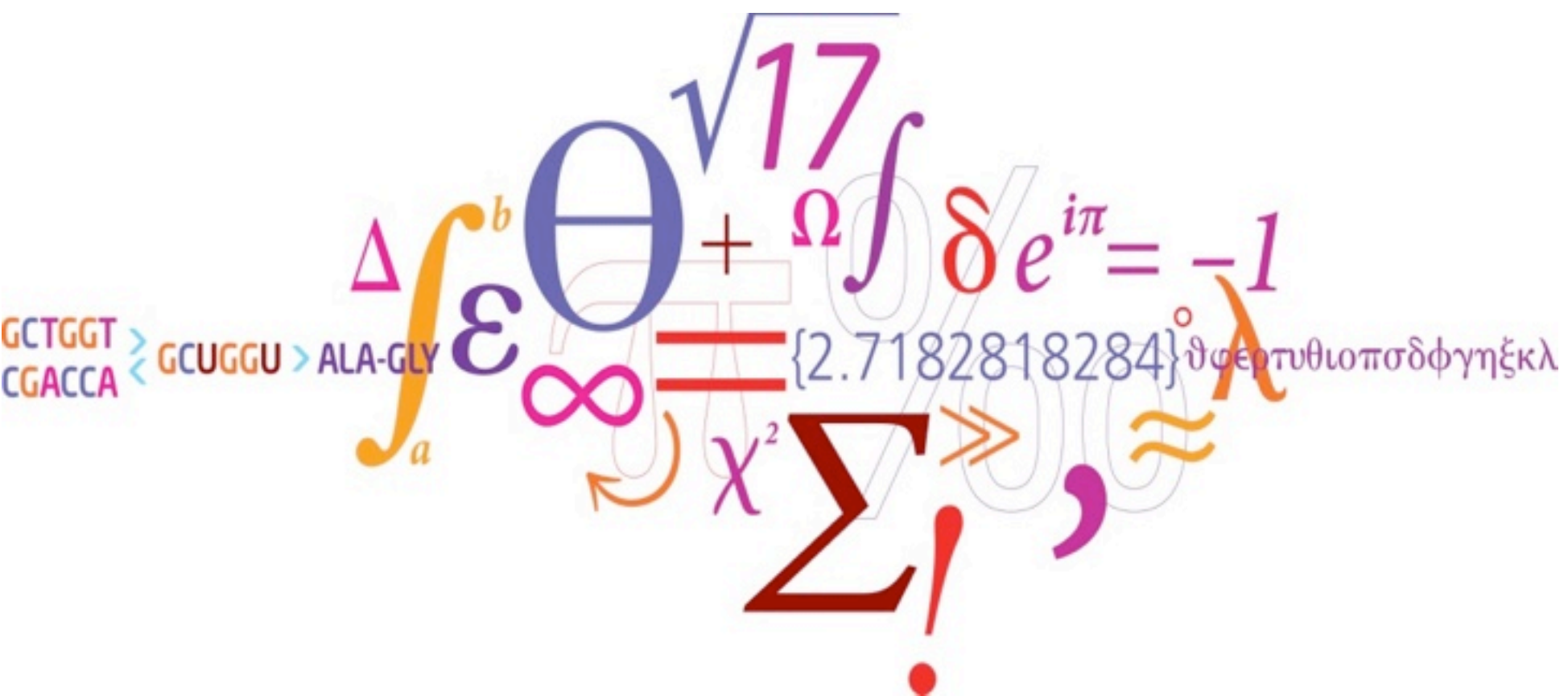
Founder effect



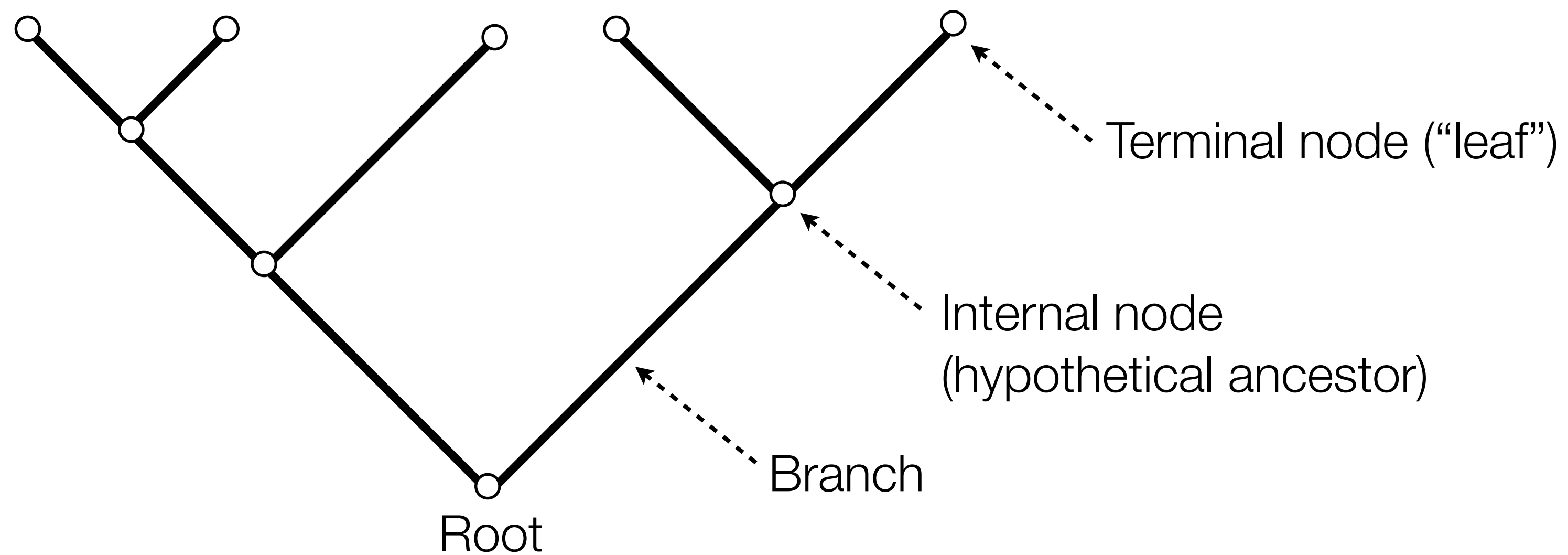
[image source: wikimedia](#)

- New Atlantic population, maybe from only 10 individuals

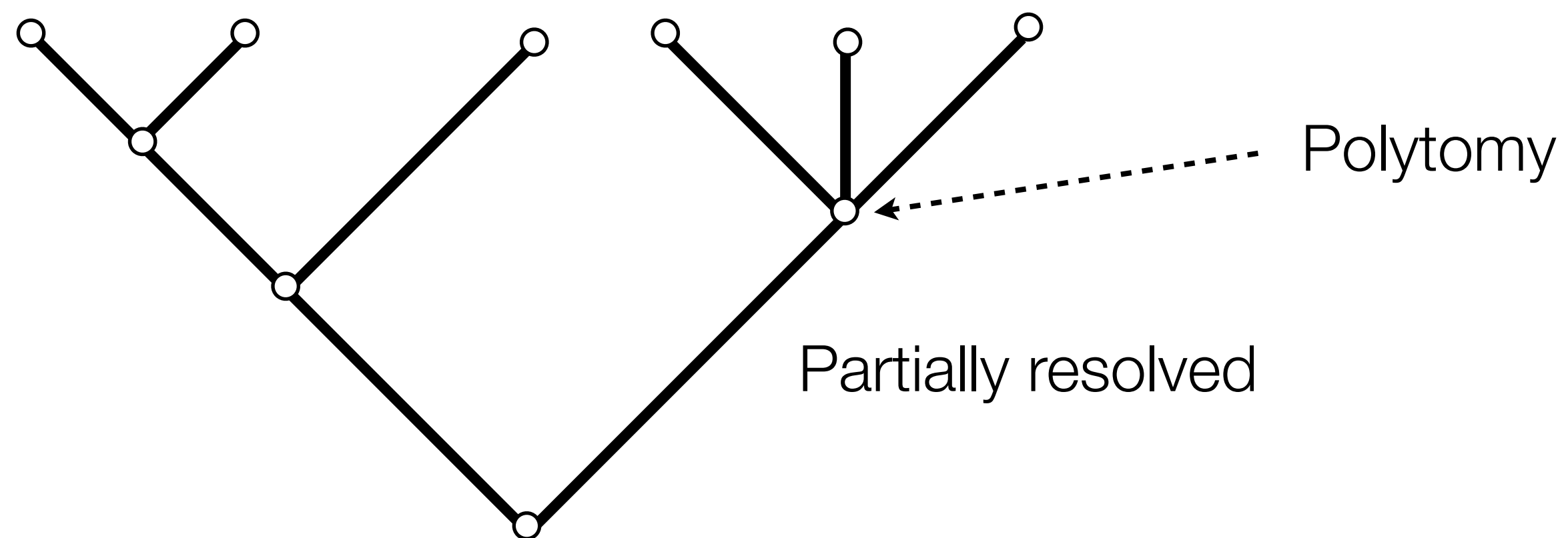
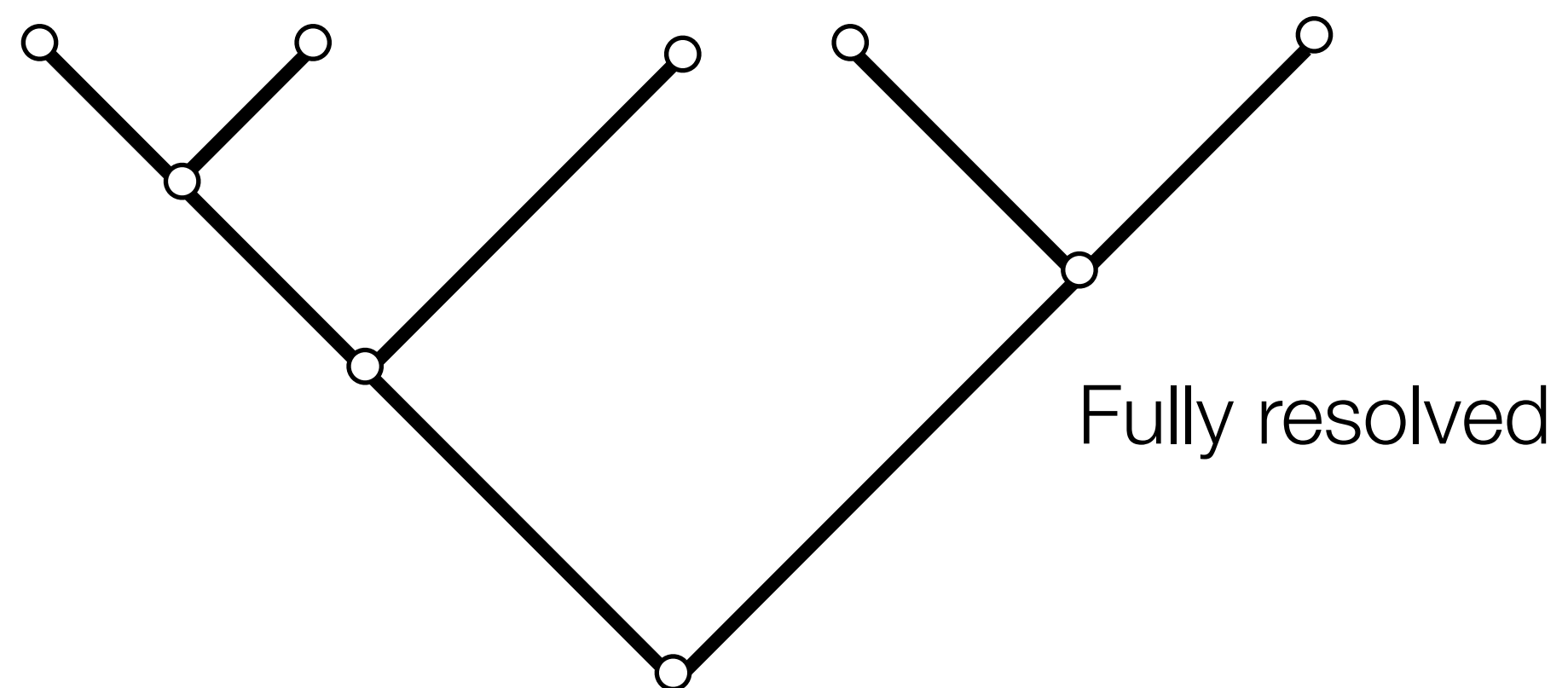
Phylogenetic Trees: Terminology and Representation



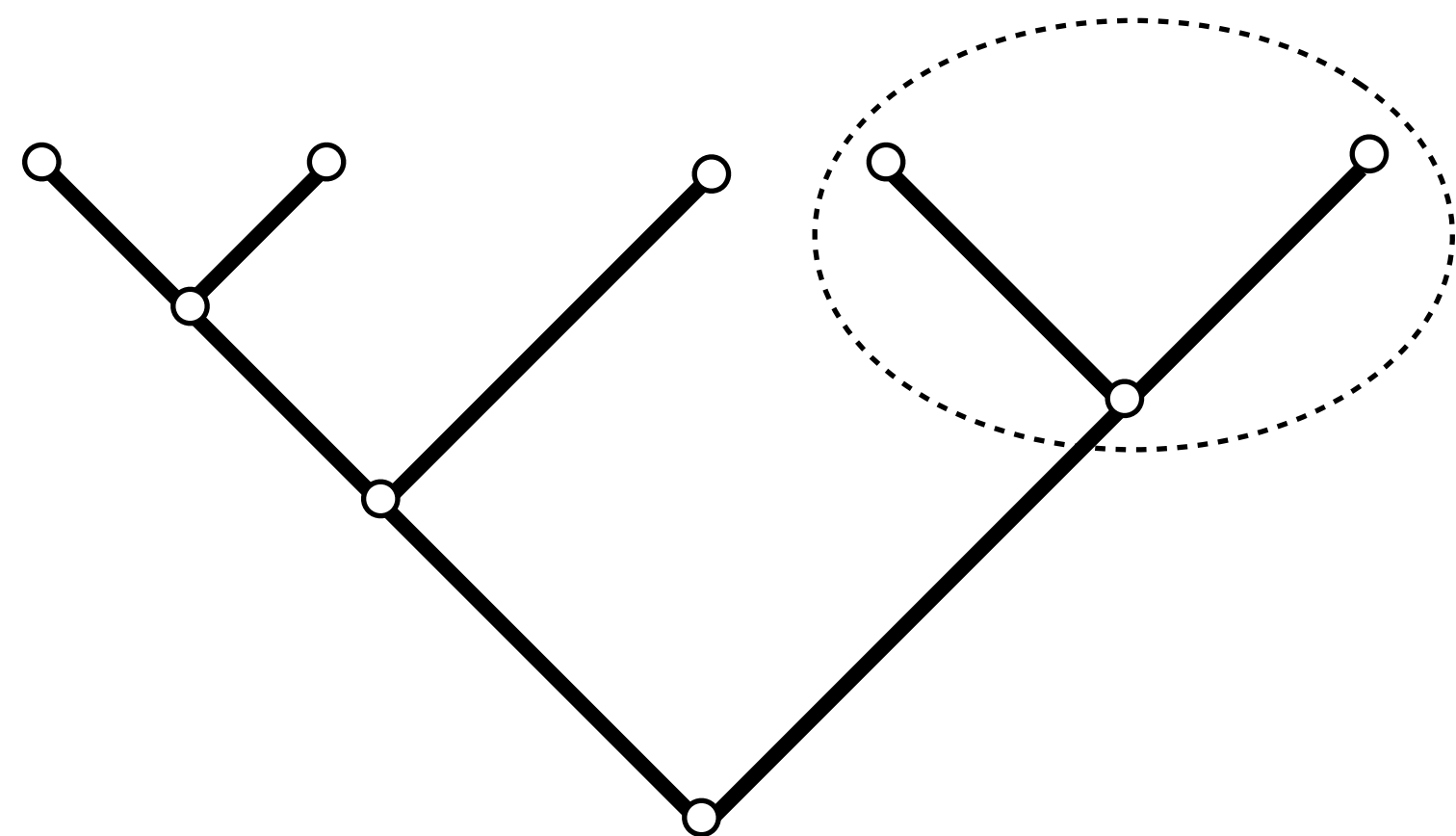
Trees: terminology



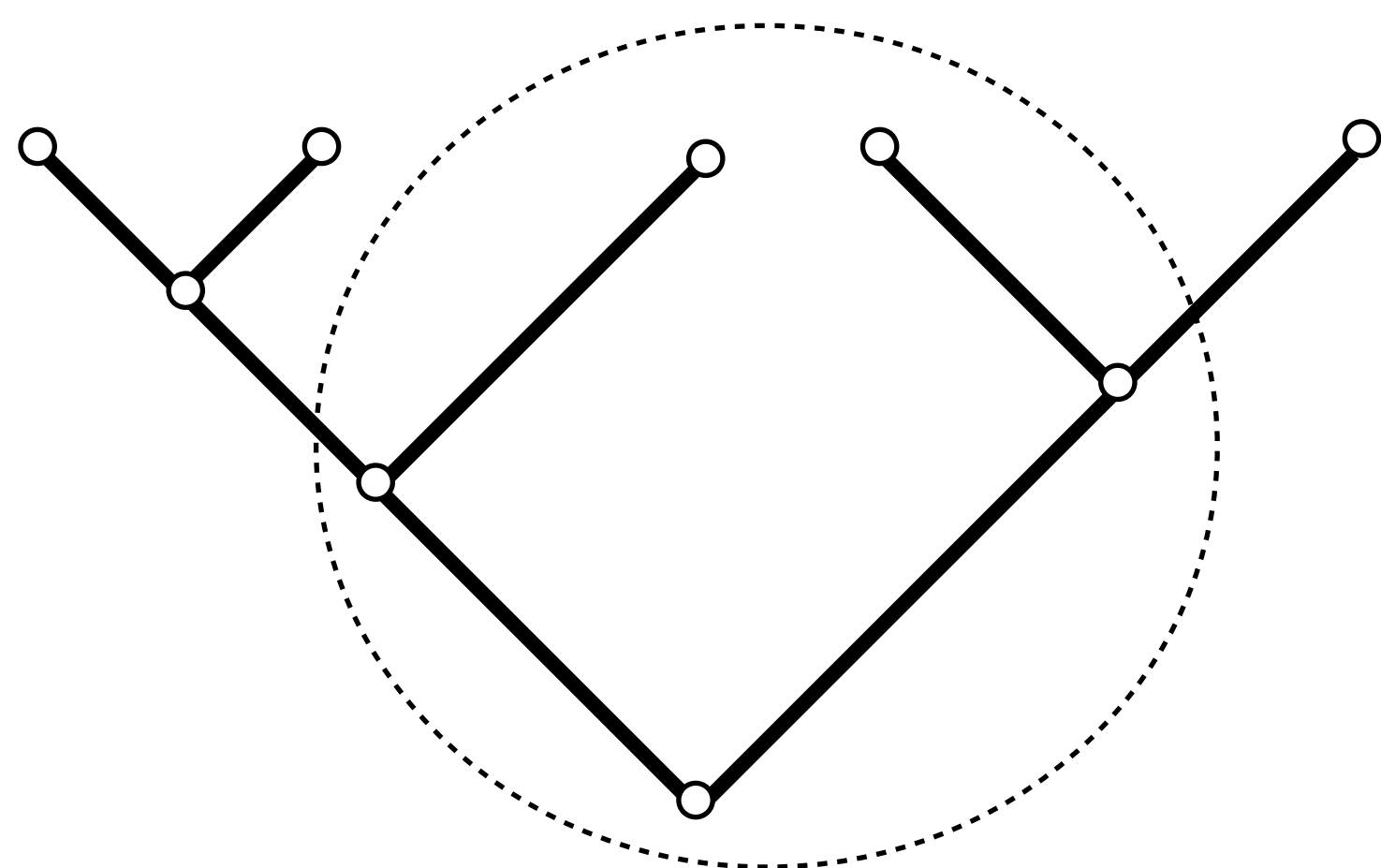
Trees: terminology



Trees: terminology

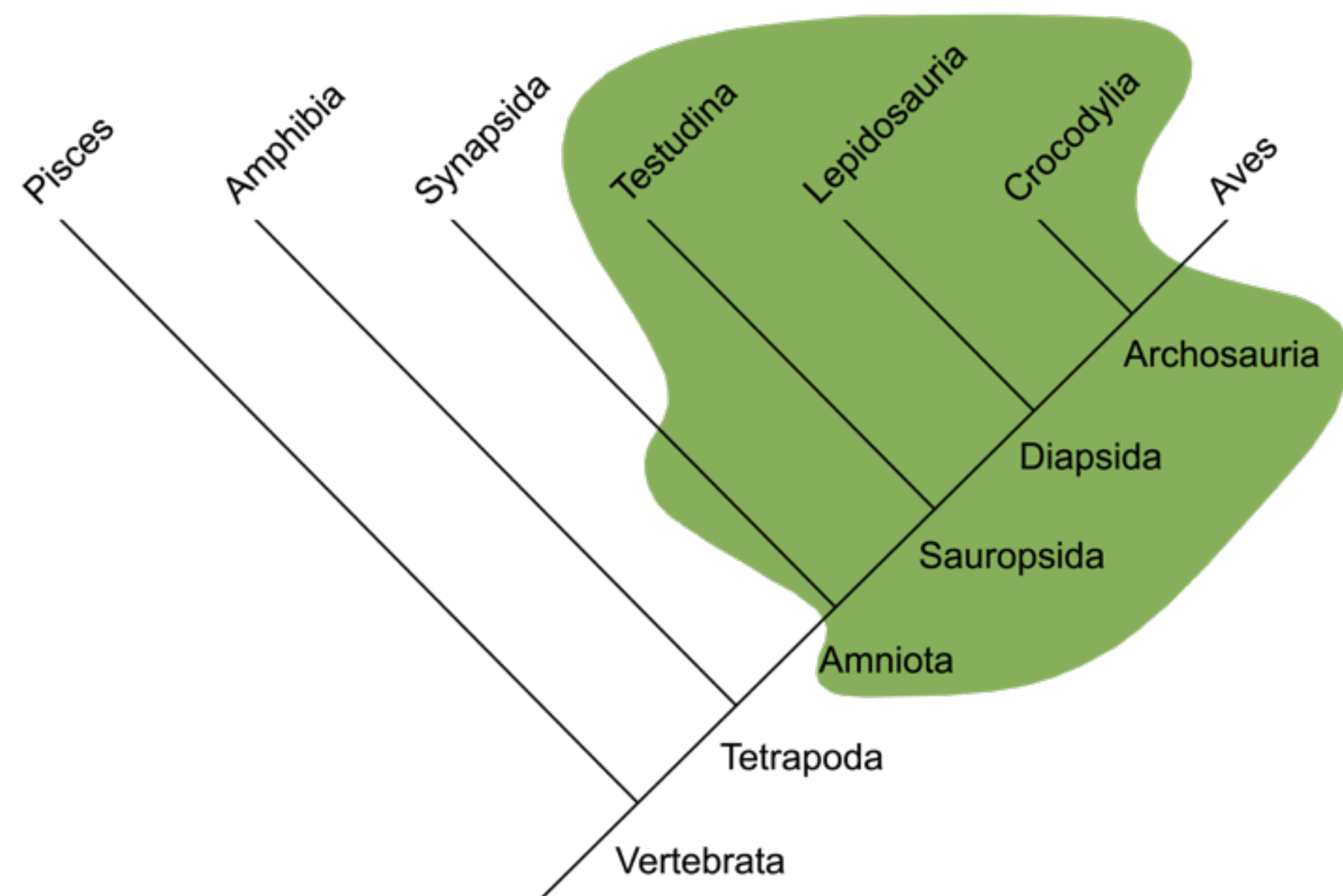


Monophyletic



Non-monophyletic
(paraphyletic)

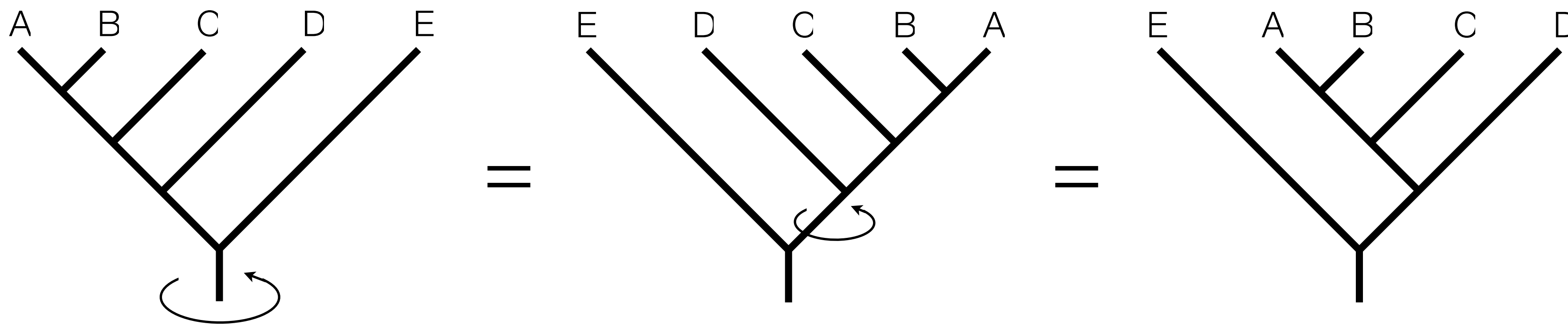
Trees: terminology



“Reptilia” is not a monophyletic group
(unless birds are included...)

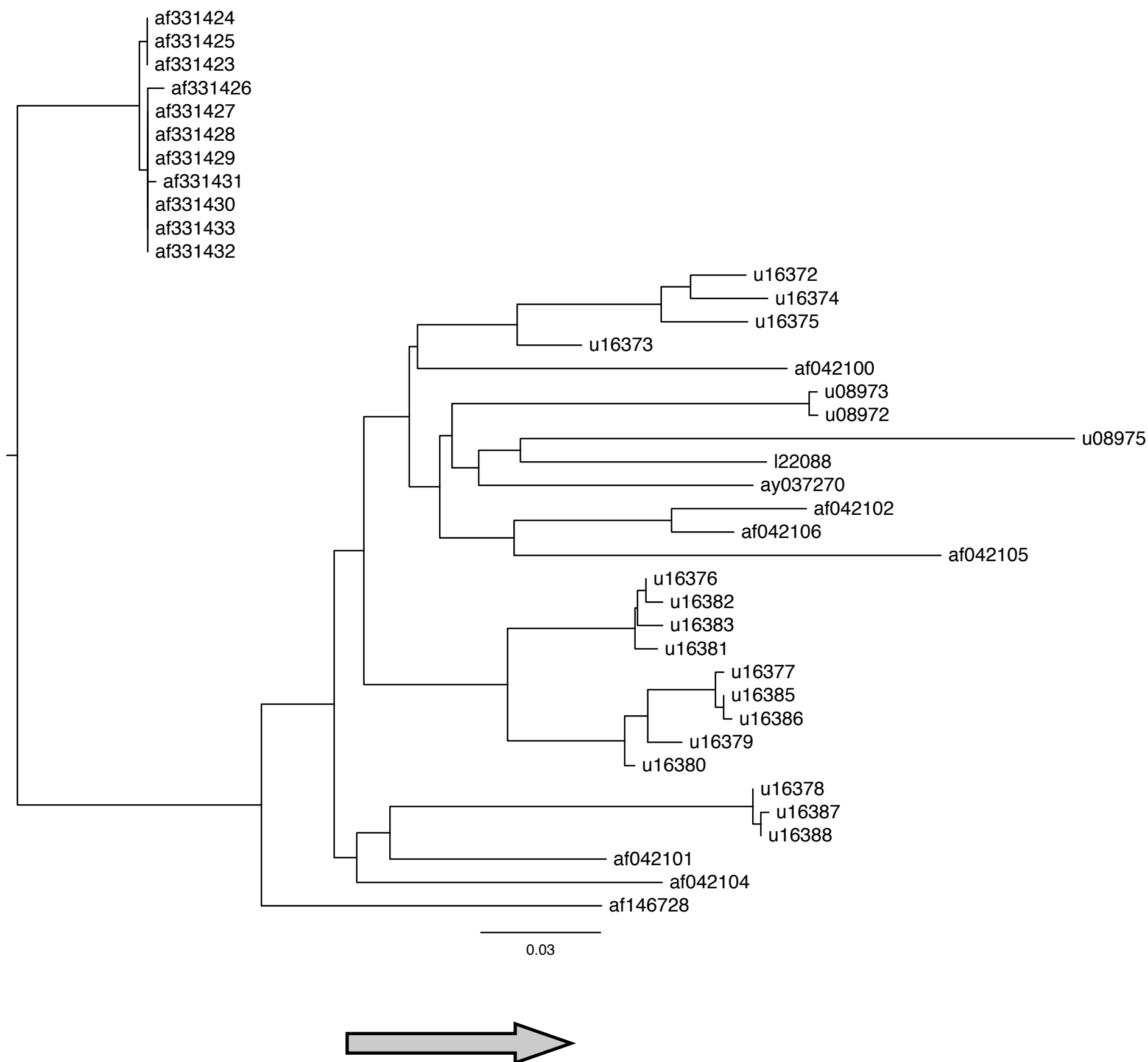
[image source: wikipedia](#)

Trees: representations



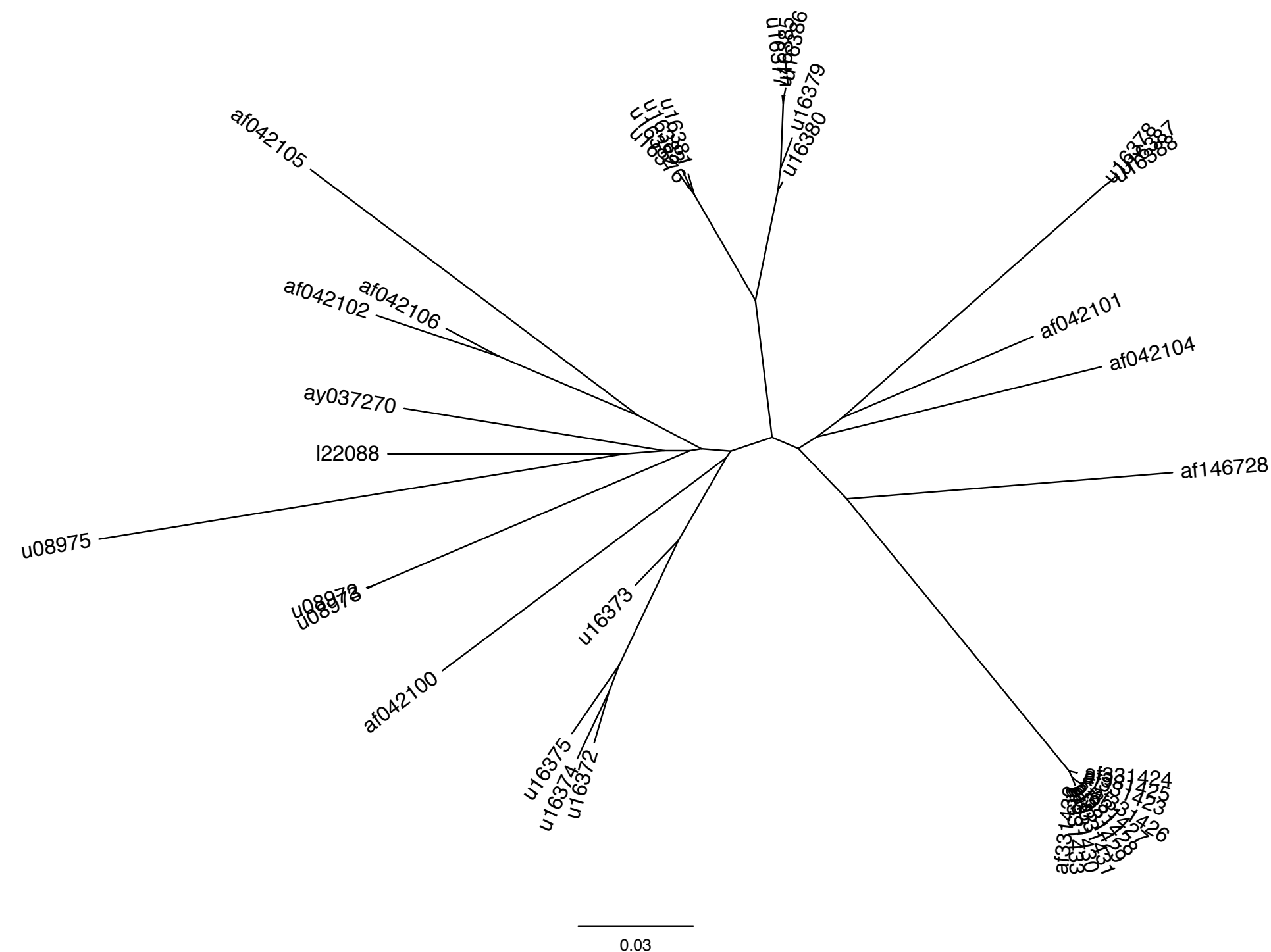
Three different representations of the same tree-topology

Trees: rooted vs. unrooted



- A rooted tree has a single node (the root) that represents a point in time that is earlier than any other node in the tree.
- A rooted tree has directionality (nodes can be ordered in terms of “earlier” or “later”).
- In the rooted tree, distance between two nodes is represented along the time-axis only (the second axis just helps spread out the leaves)

Trees: rooted vs. unrooted



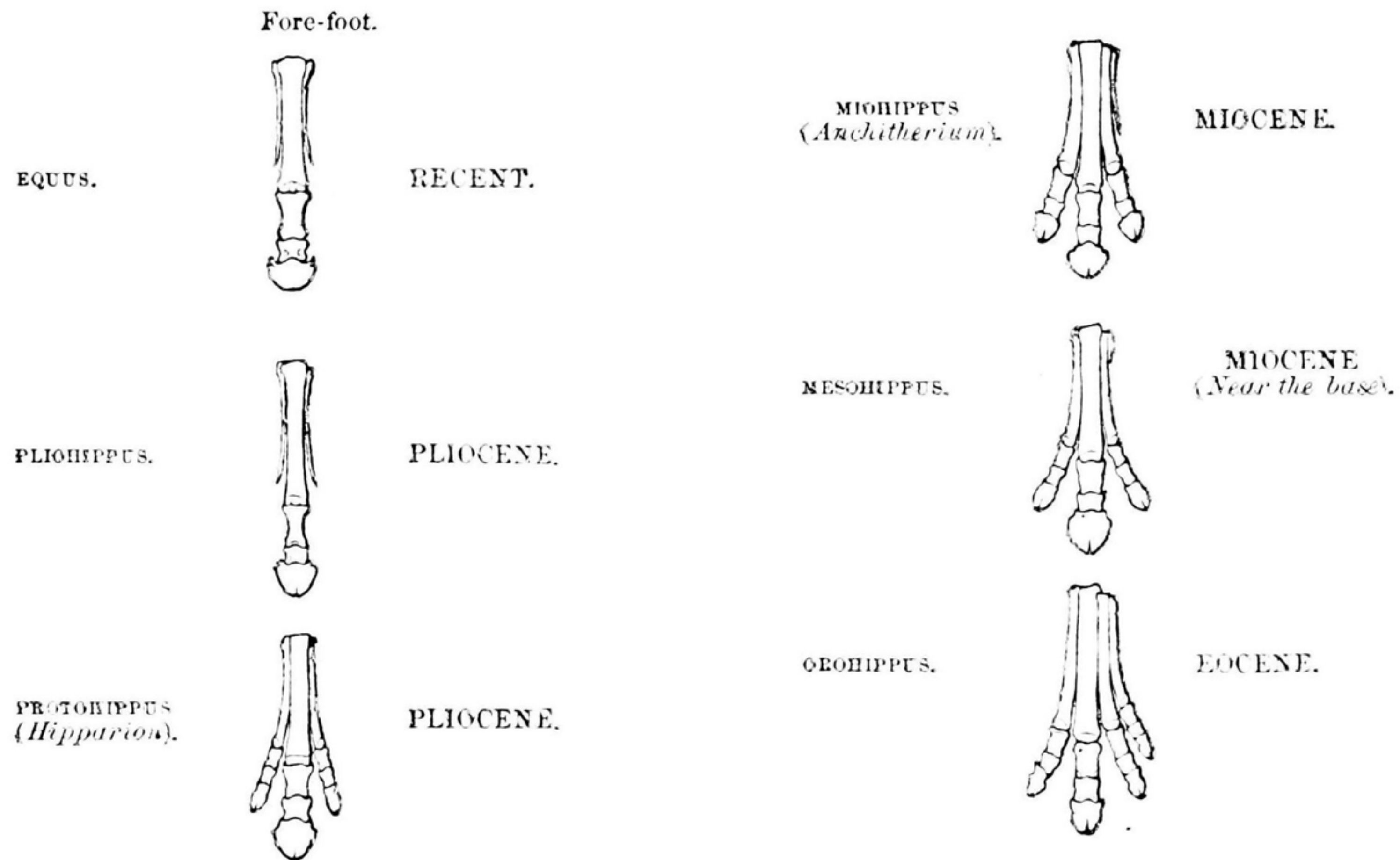
- In unrooted trees there is no directionality: we do not know if a node is earlier or later than another node
- Distance along branches directly represents node distance

Homology and Homoplasy

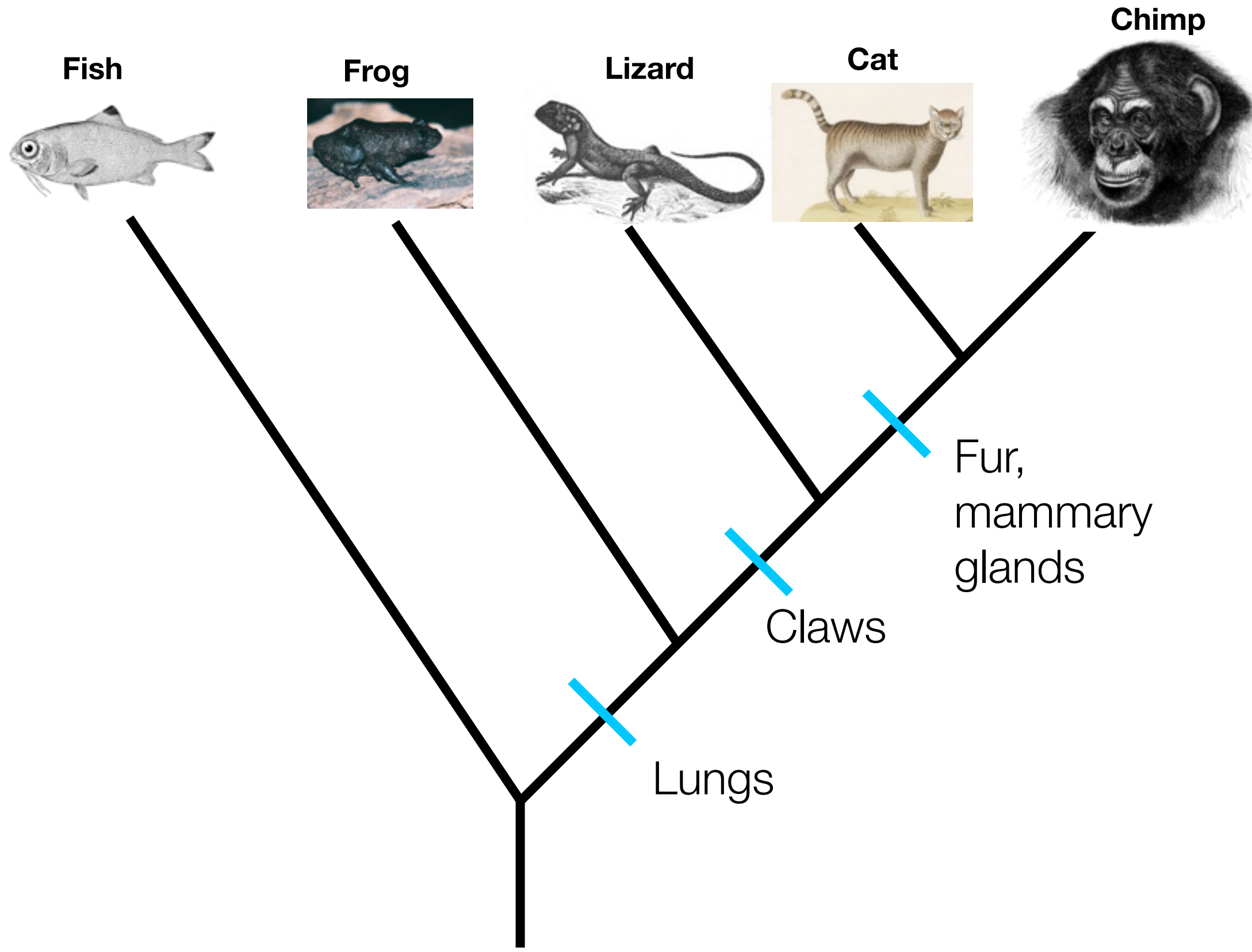
GCTGGT > GCUGGU > ALA-GLY
CGACCA

$\int_a^b \epsilon$ Θ $\sqrt{17}$ $+$ $\int \delta e^{i\pi} = -1$
 ∞ $\{2.7182818284\}$ λ
 χ^2 Σ \gg \approx λ
 $\Sigma!$ ϕ $,$

Reconstructing an evolutionary history using fossils



Reconstructing a tree using present-day data



Homology: limb structure

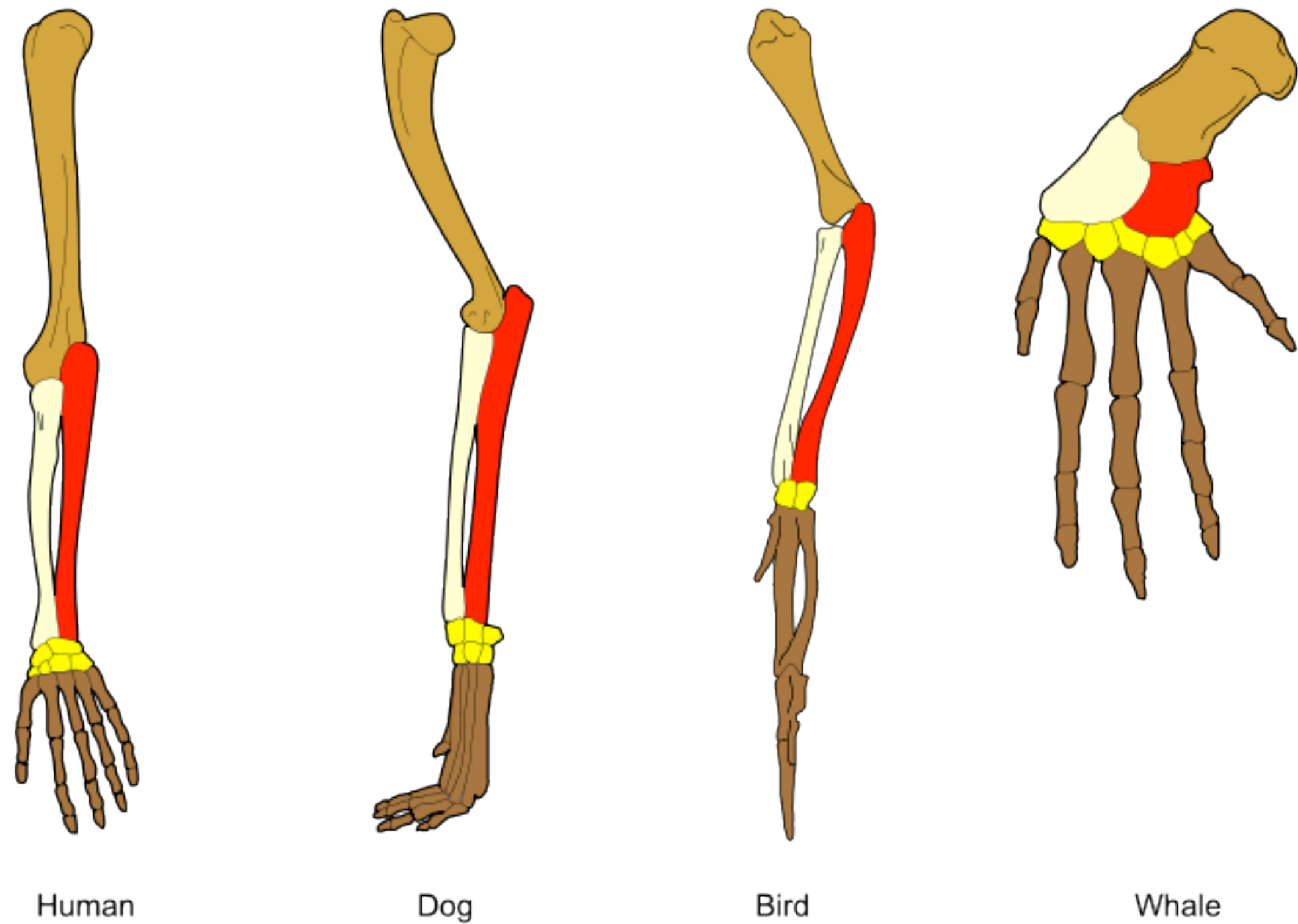


image source: [wikipedia](https://www.wikipedia.org)

Morphology vs. molecular data

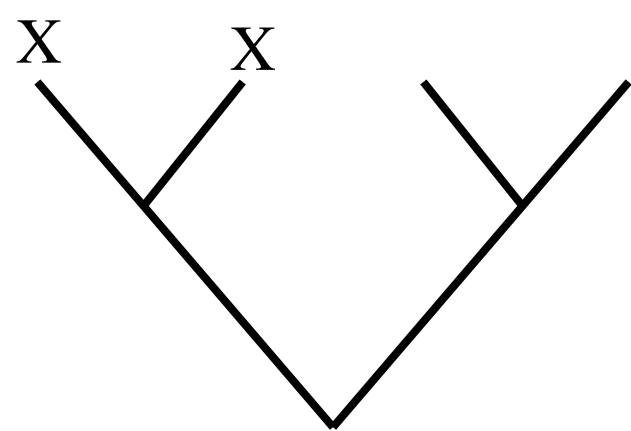


Red-headed Vulture (old world vulture)

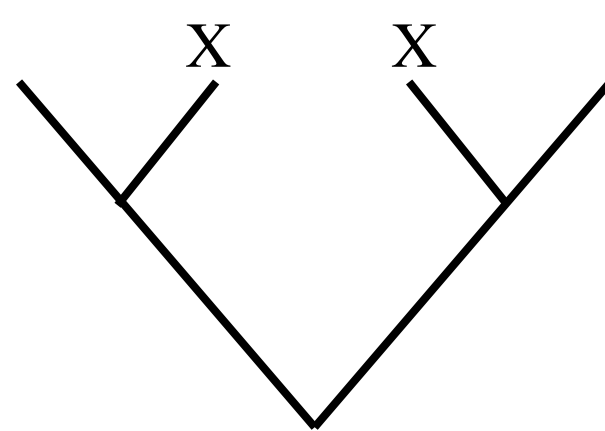
Turkey vulture (new world vulture)

- New and old world vultures seem to be closely related based on morphology.
- Molecular data indicates that old world vultures are related to birds of prey (falcons, hawks, etc.) while new world vultures are more closely related to storks
- Similar features presumably the result of convergent evolution

Homology *vs.* Homoplasy

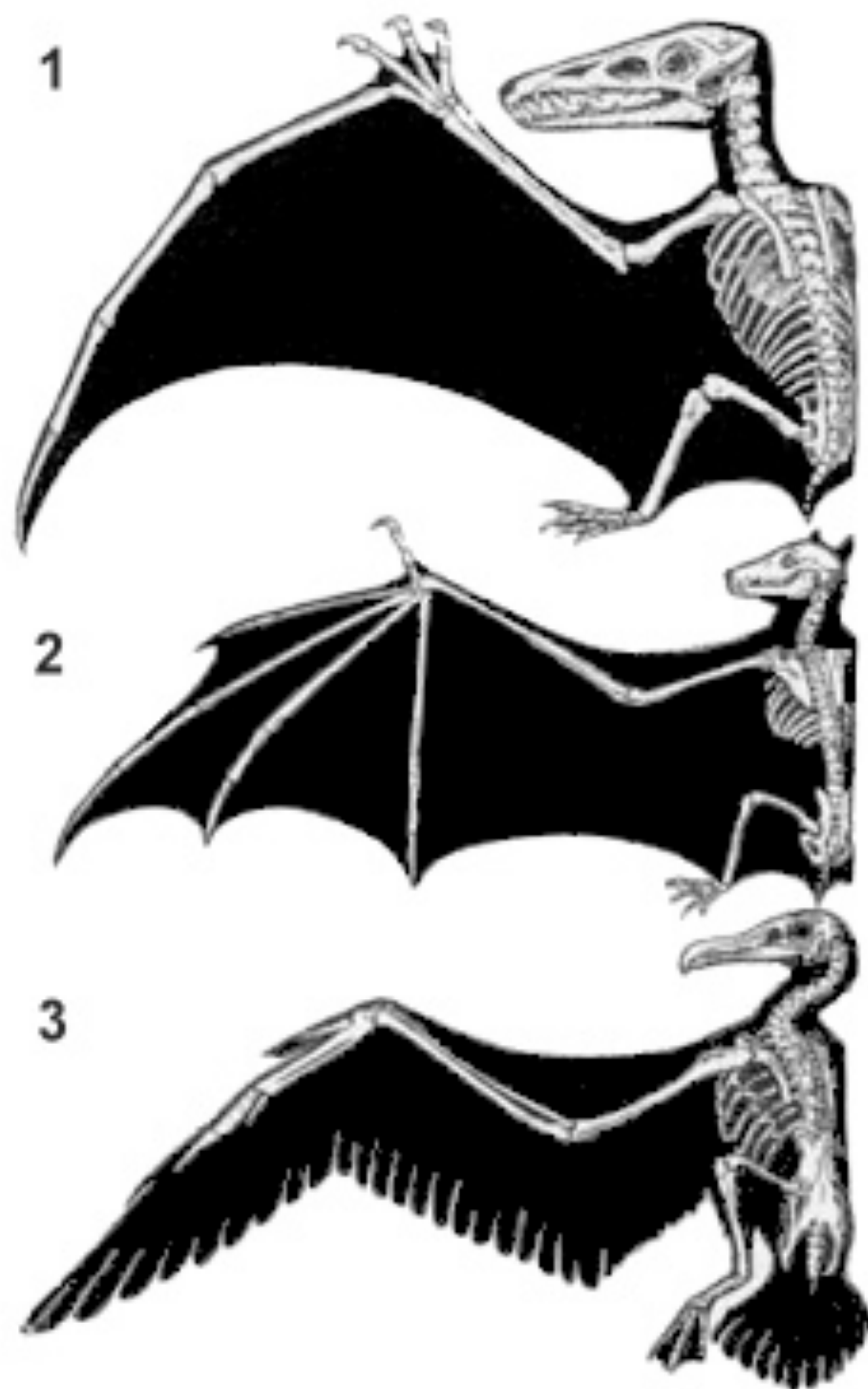


Homology: similar traits inherited from a common ancestor



Homoplasy: similar traits are not directly caused by common ancestry (convergent evolution).

Homoplasy: wings



Pterosaur

Bat

Bird

Molecular phylogeny

A	A	G	C	G	T	T	G	G	G	C	A	A
B	A	G	C	G	T	T	T	G	G	C	A	A
C	A	G	C	T	T	T	G	T	G	C	A	A
D	A	G	C	T	T	T	T	T	G	C	A	A
				1			2	3				

- DNA and protein sequences
 - Homologous characters inferred from alignment.
 - Other molecular data: absence/presence of restriction sites, DNA hybridization data, antibody cross-reactivity, etc. (but losing importance due to cheap, efficient sequencing).
-

Maximum Parsimony

$\int_a^b \epsilon \Theta^{\sqrt{17}} + \Omega \int \delta e^{i\pi} = -1$

GCTGGT > GCUGGU > ALA-GLY

CGACCA

$\{2.7182818284\}$

χ^2 Σ \gg \approx λ

υφερτυθιοποσδφγηξκλ



Phylogenetic reconstruction

Taxon	Nucleotide position		
	1	2	3
A	G	G	G
B	G	T	G
C	T	G	T
D	T	T	T

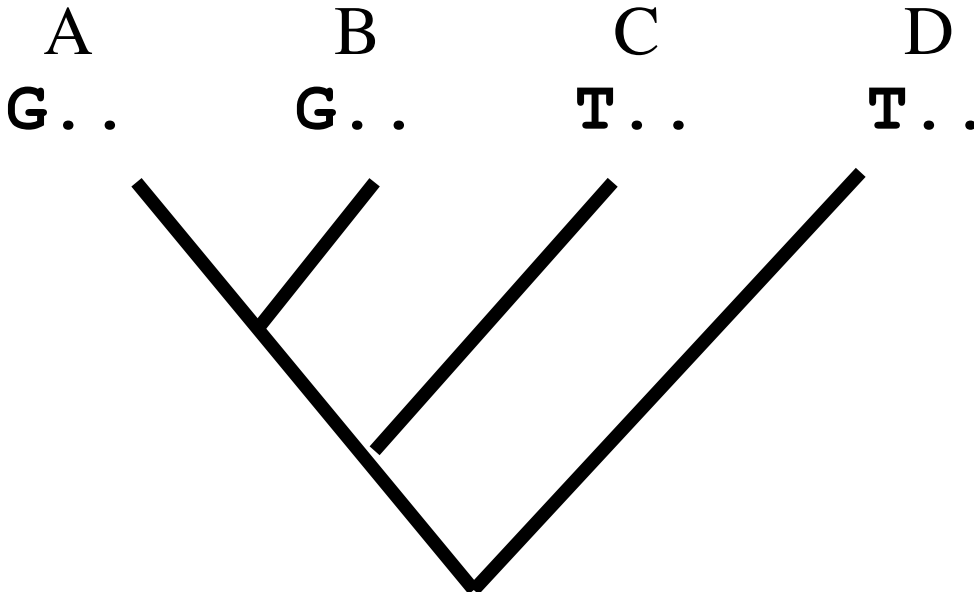
Phylogenetic reconstruction

Taxon	Nucleotide position		
	1	2	3
A	G	G	G
B	G	T	G
C	T	G	T
D	T	T	T

Parsimony criterion: choose simplest hypothesis

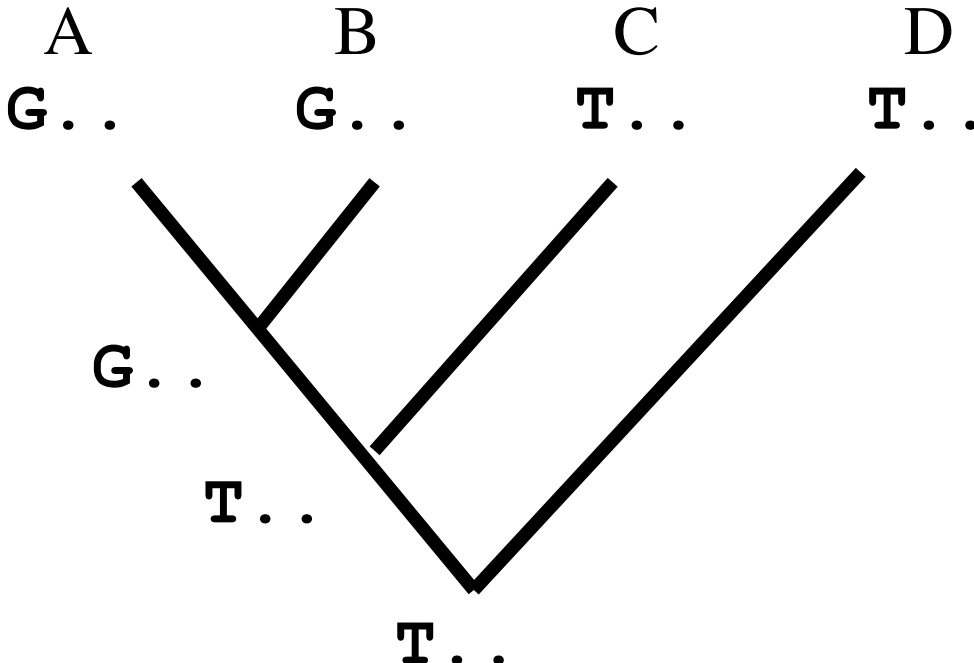
Taxon		Nucleotide position		
		1	2	3
	A	G	G	G
	B	G	T	G
	C	T	G	T
	D	T	T	T

Parsimony criterion: choose simplest hypothesis



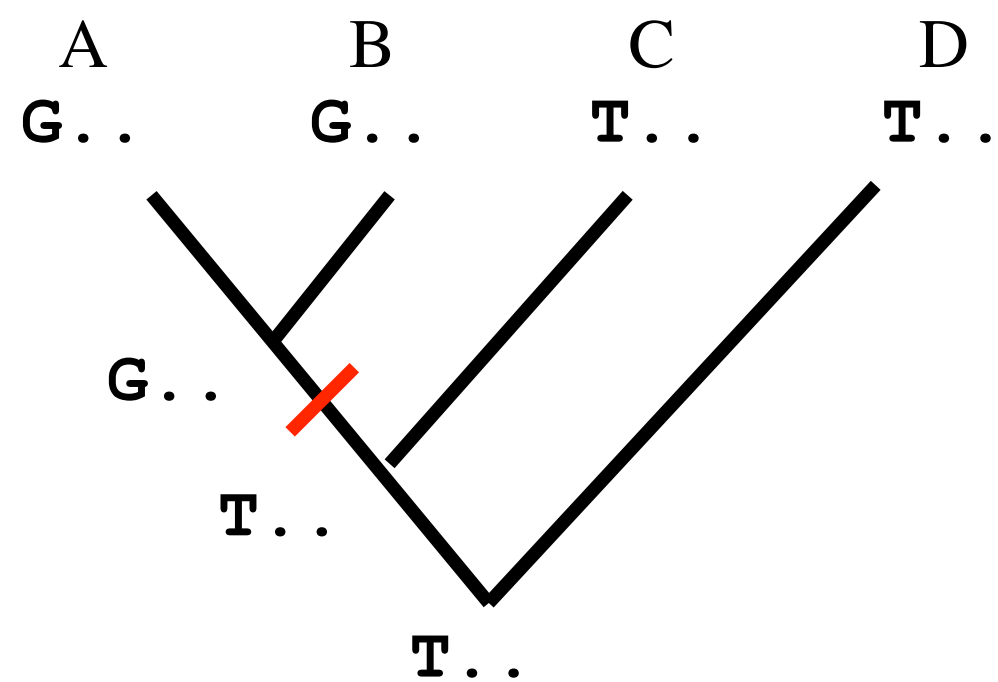
Taxon	Nucleotide position		
	1	2	3
A	G	G	G
B	G	T	G
C	T	G	T
D	T	T	T

Parsimonious reconstruction



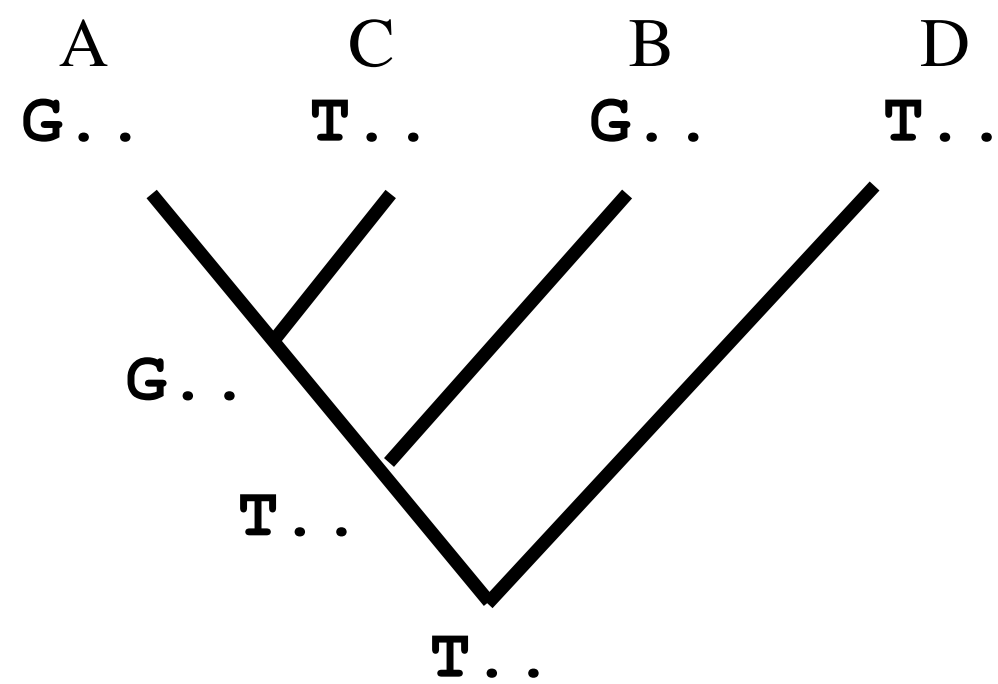
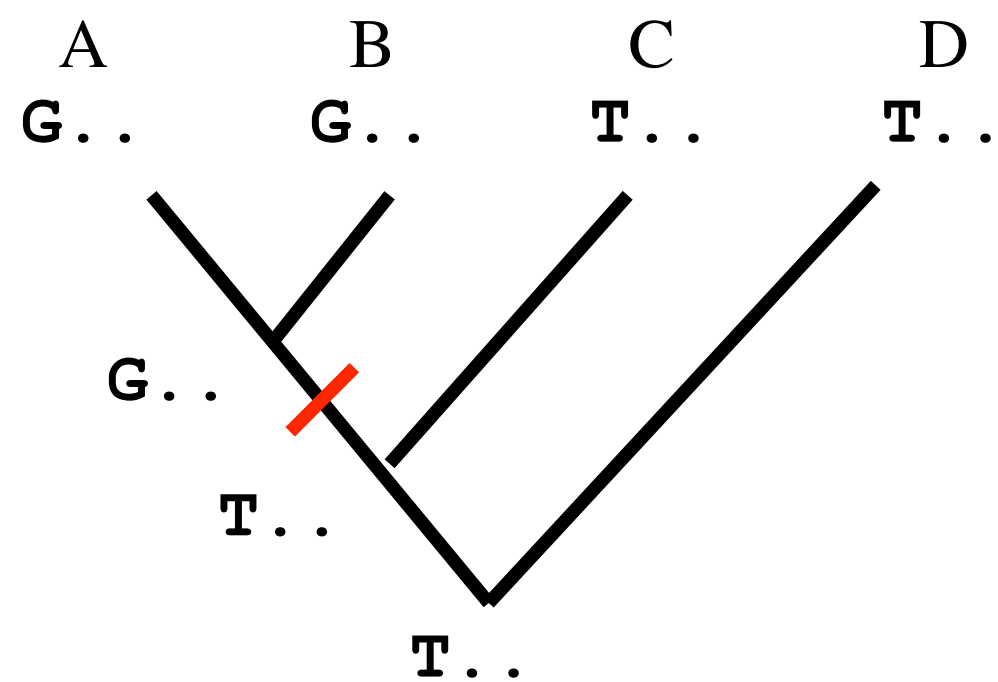
Taxon	Nucleotide position		
	1	2	3
→ A	G	G	G
→ B	G	T	G
C	T	G	T
D	T	T	T

Parsimonious reconstruction



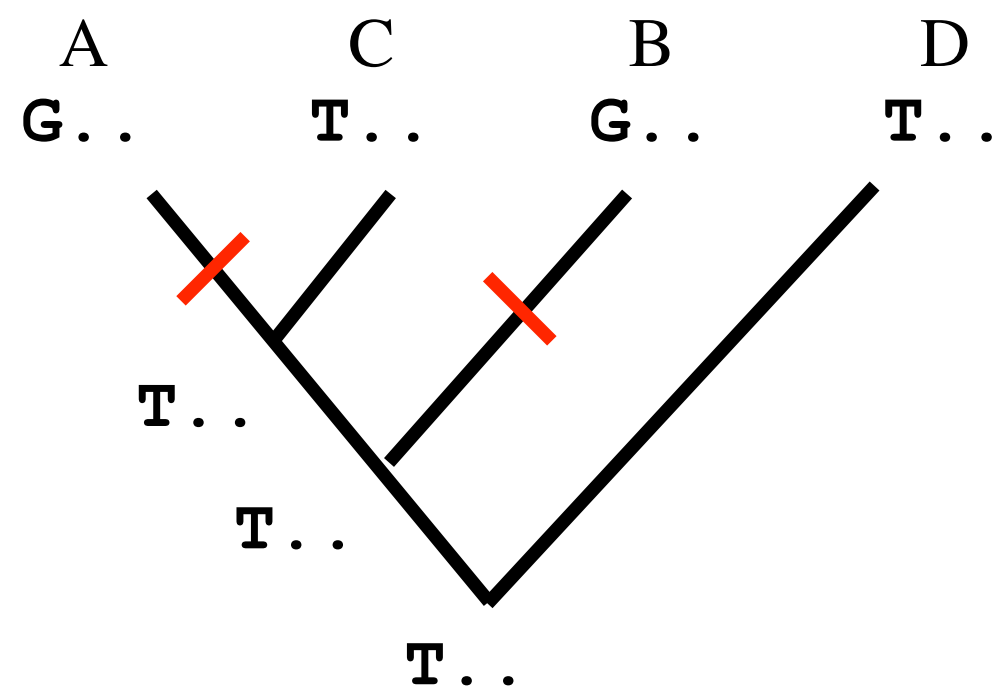
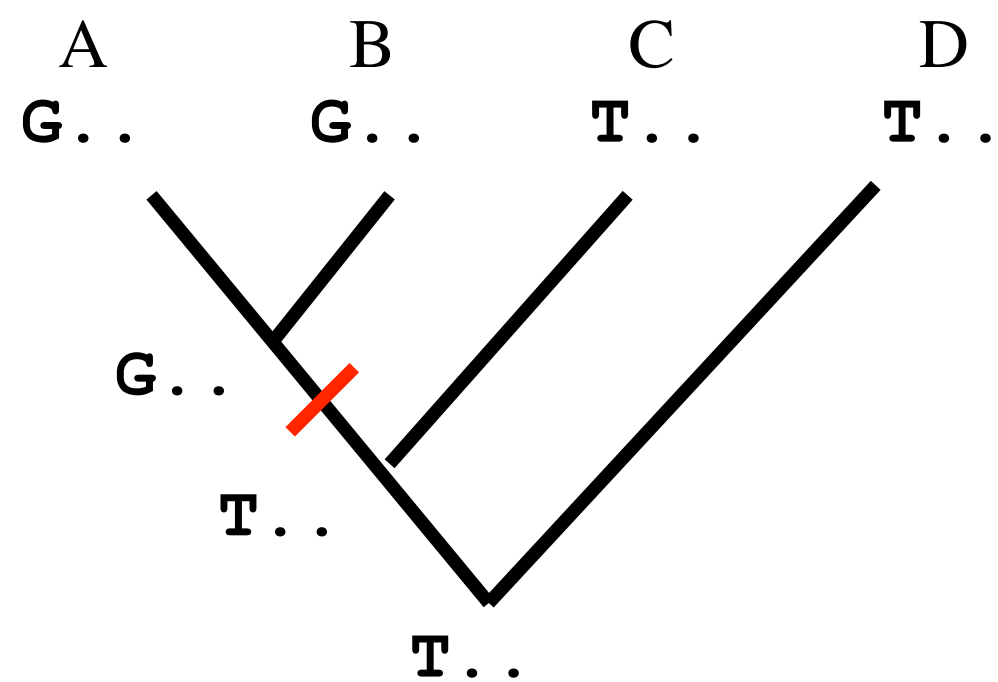
		Nucleotide position		
Taxon		1	2	3
→	A	G	G	G
→	B	G	T	G
	C	T	G	T
	D	T	T	T

Alternative tree: homoplasy



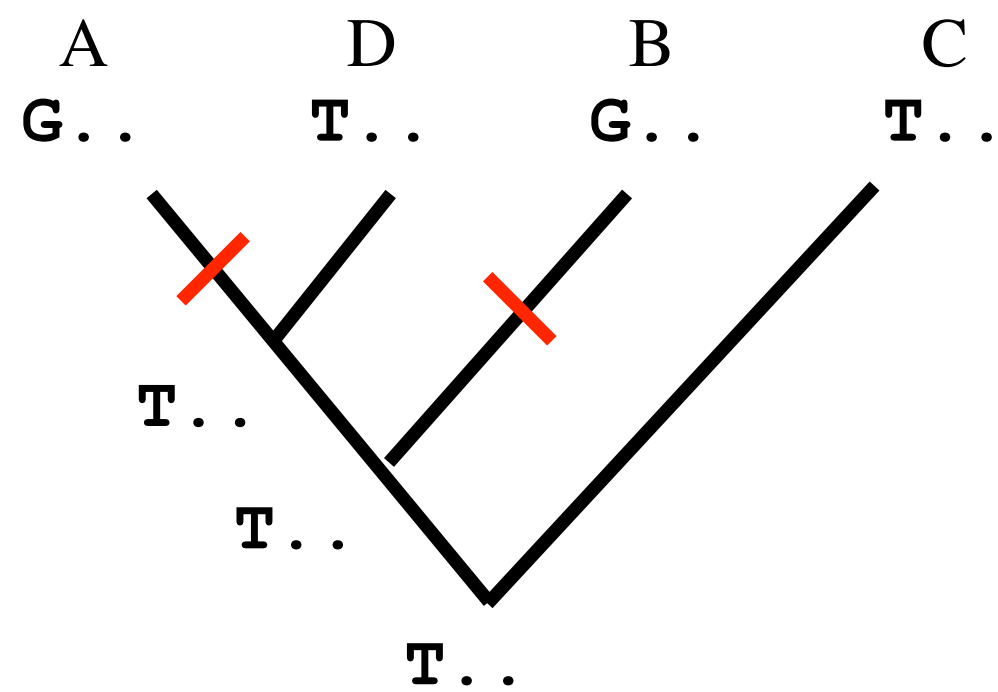
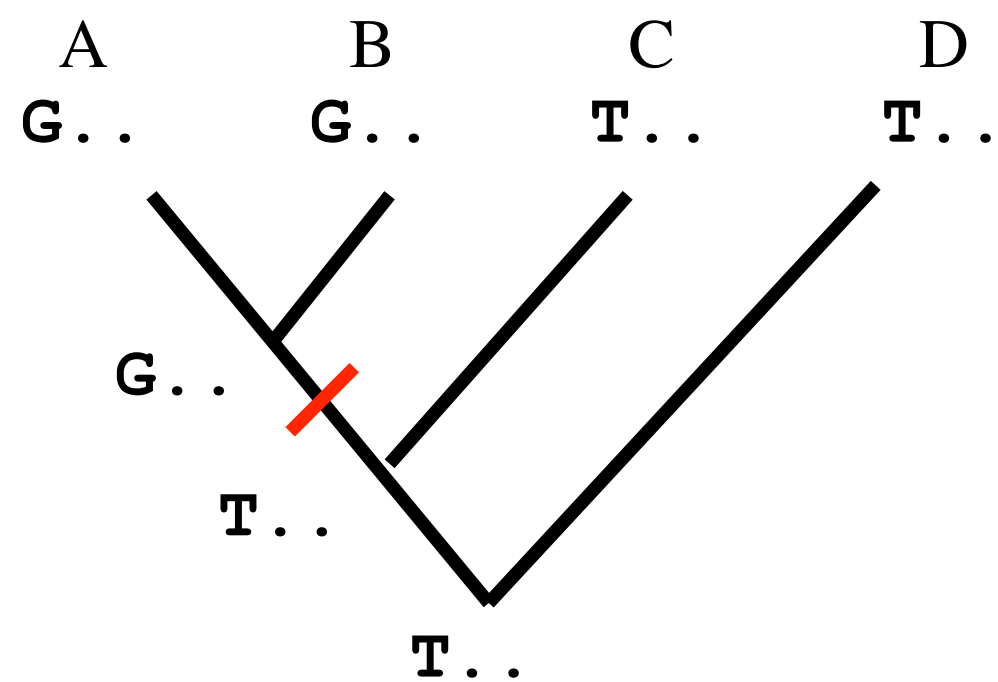
Taxon	Nucleotide position		
	1	2	3
A	G	G	G
B	G	T	G
C	T	G	T
D	T	T	T

Alternative tree: homoplasy



		Nucleotide position		
Taxon		1	2	3
→	A	G	G	G
	B	G	T	G
→	C	T	G	T
	D	T	T	T

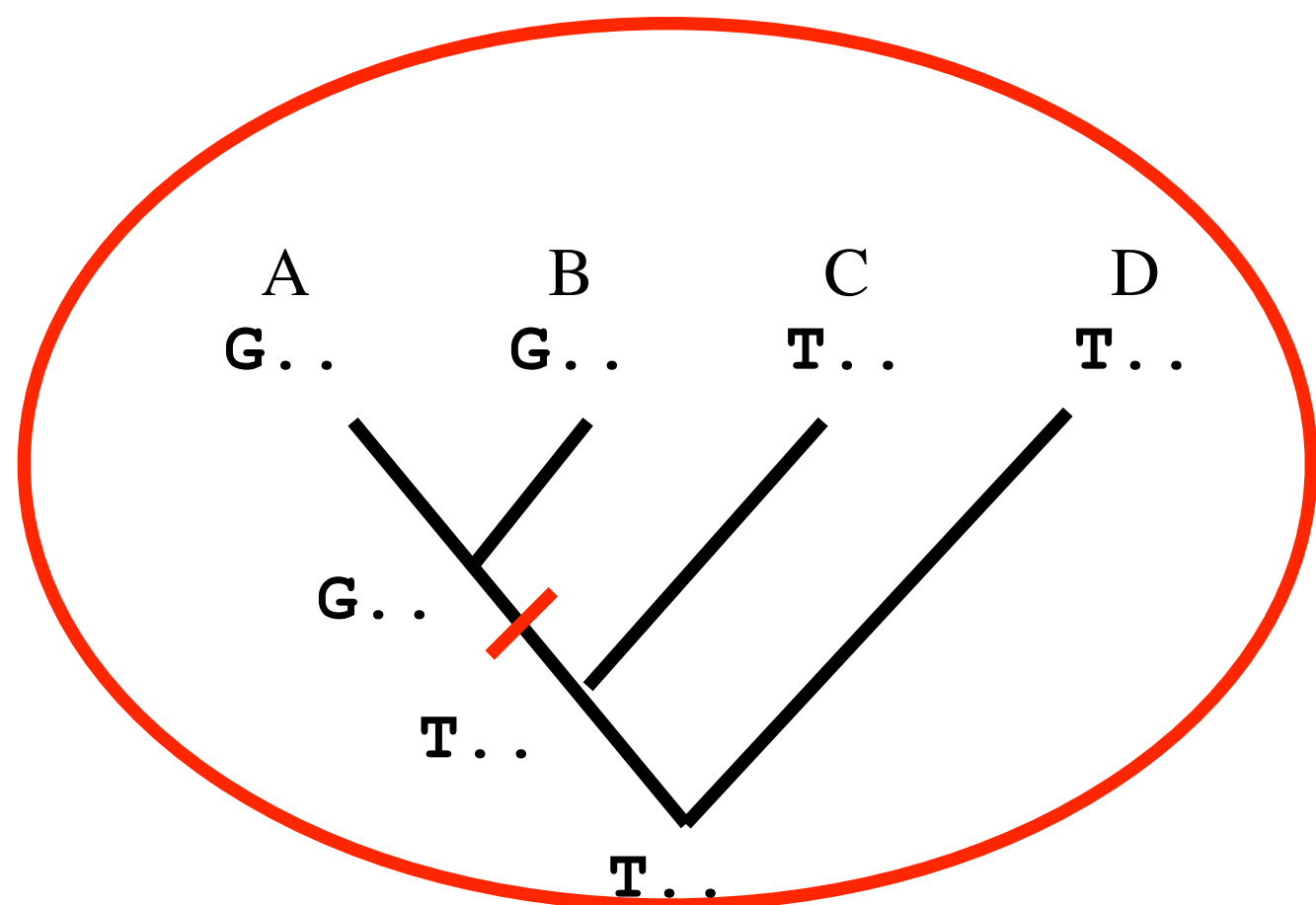
Alternative tree: homoplasy



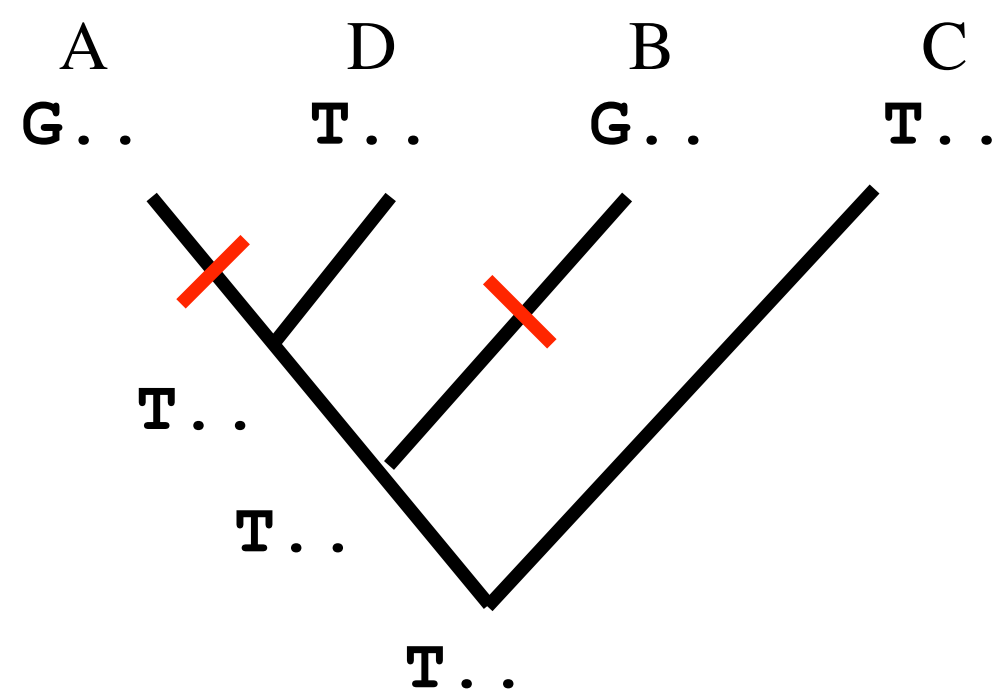
Taxon	Nucleotide position		
	1	2	3
A	G	G	G
B	G	T	G
C	T	G	T
D	T	T	T

One character:

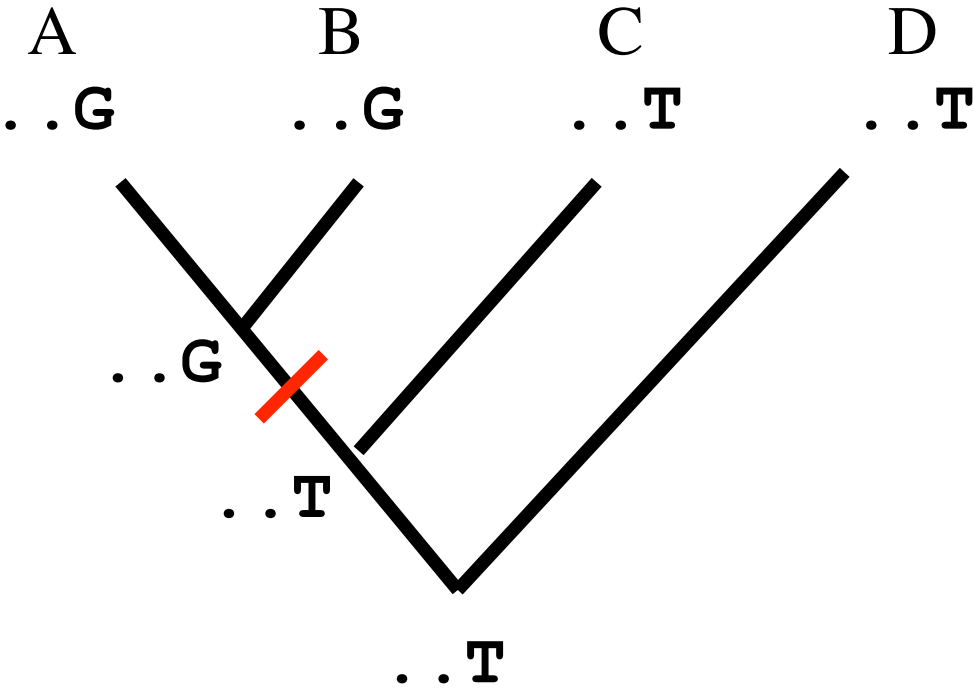
Assumption of no homoplasy is equivalent to finding shortest tree



Taxon	Nucleotide position		
	1	2	3
A	G	G	G
B	G	T	G
C	T	G	T
D	T	T	T

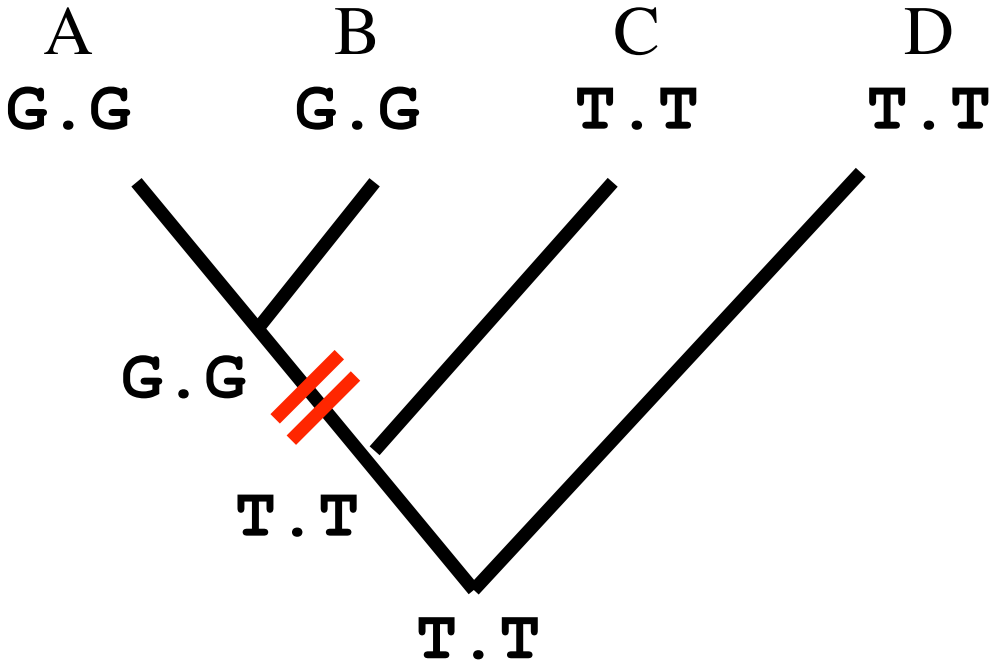


Phylogenetic reconstruction



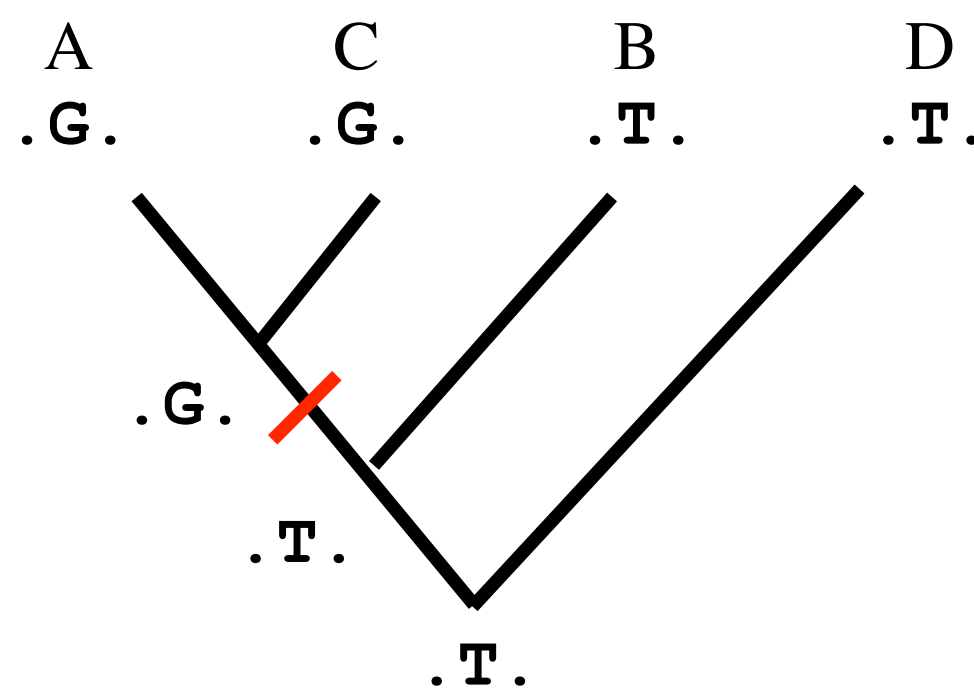
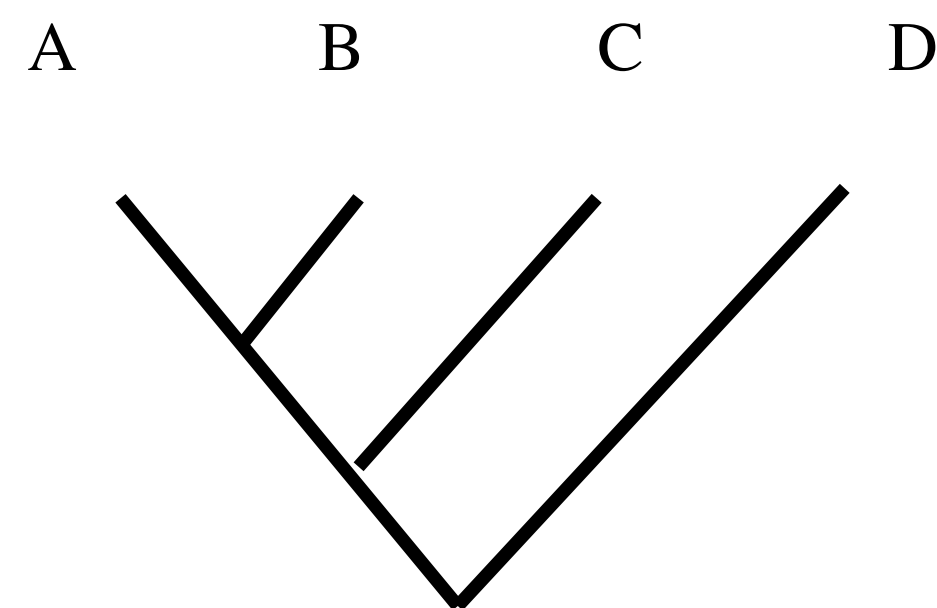
Taxon	Nucleotide position		
	1	2	3
→ A	G	G	G
→ B	G	T	G
C	T	G	T
D	T	T	T

Phylogenetic reconstruction



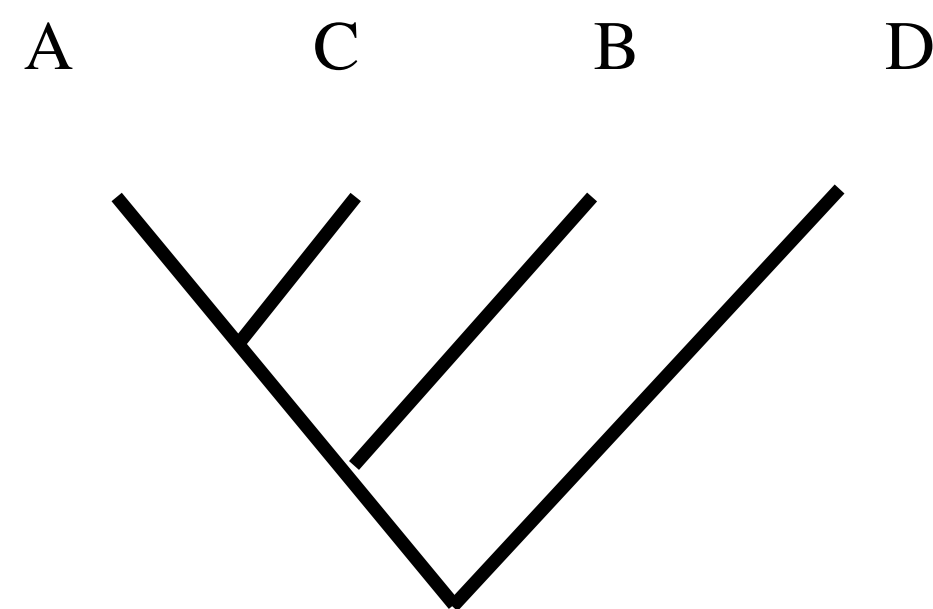
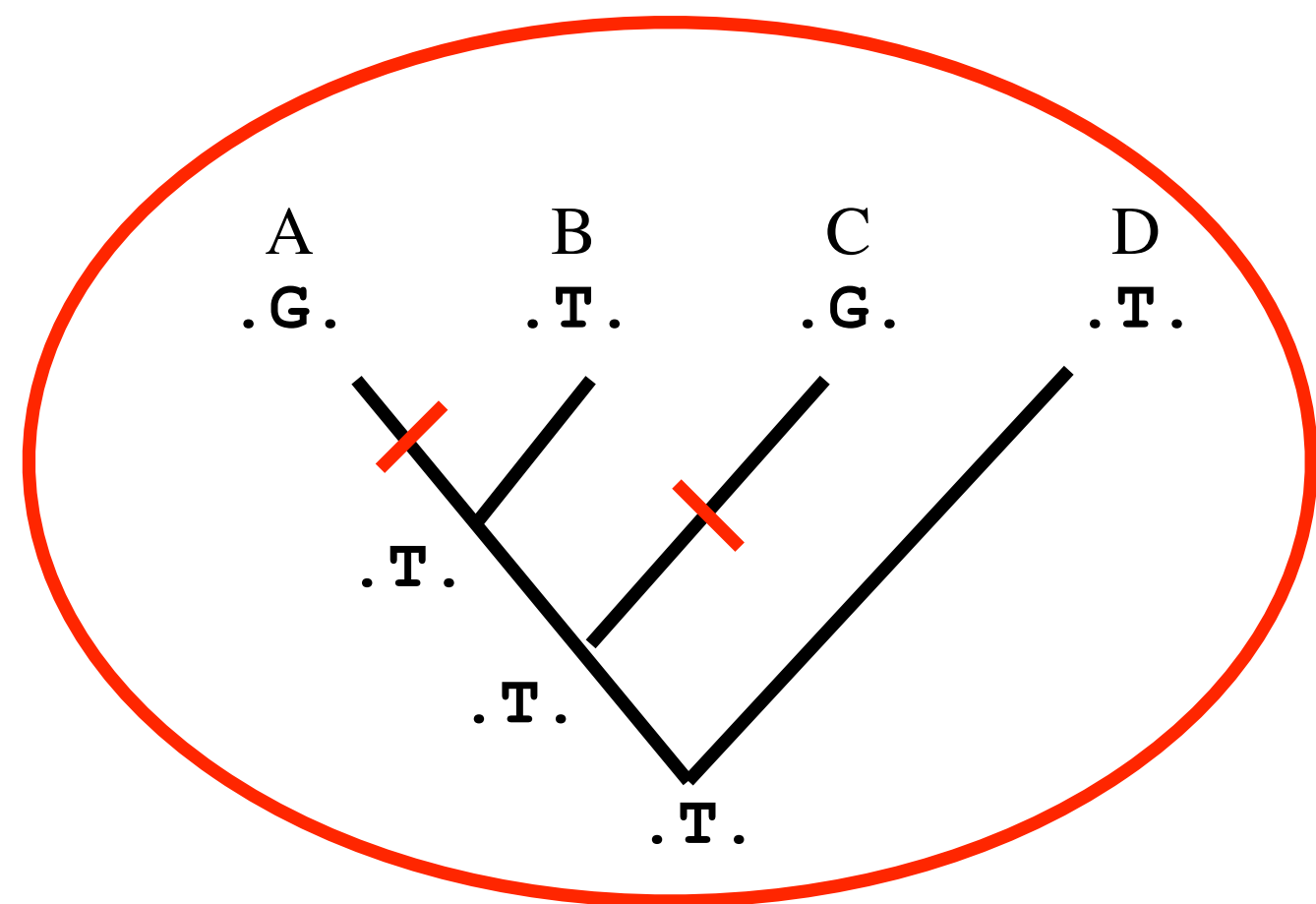
Taxon	Nucleotide position		
	1	2	3
A	G	G	G
B	G	T	G
C	T	G	T
D	T	T	T

Phylogenetic reconstruction



Taxon	Nucleotide position		
	1	2	3
A	G	G	G
B	G	T	G
C	T	G	T
D	T	T	T

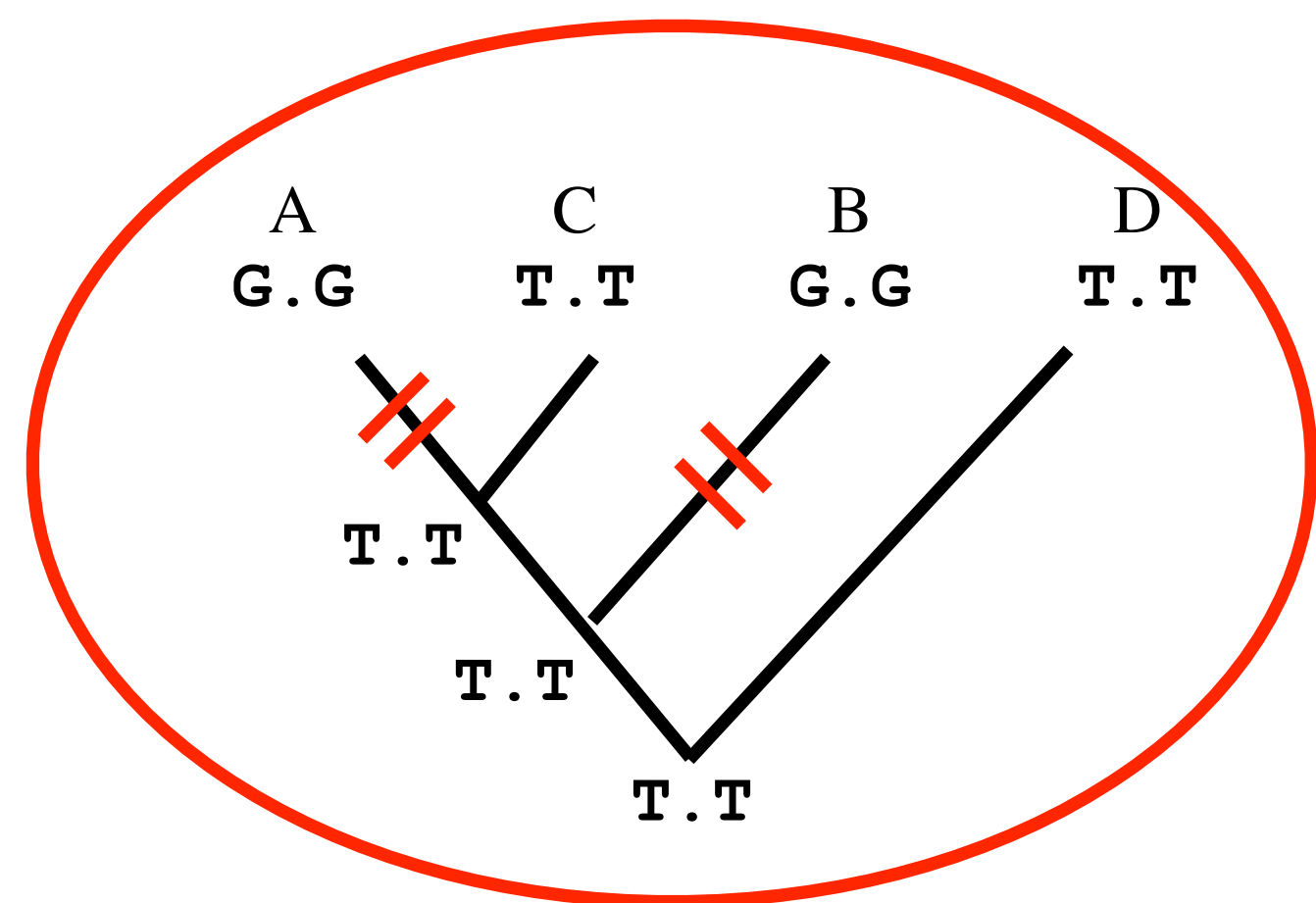
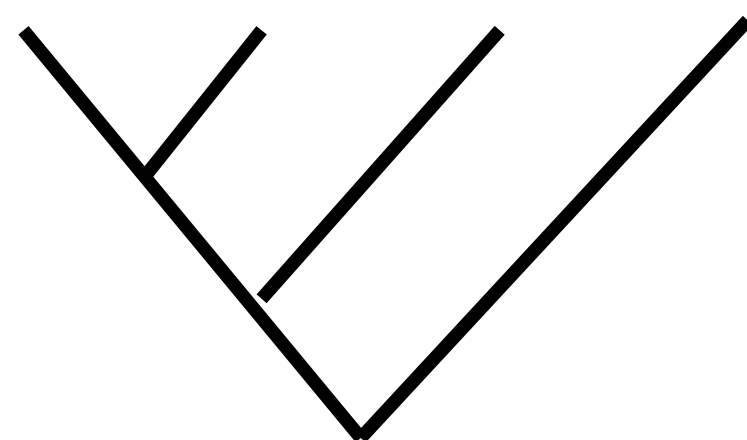
Phylogenetic reconstruction: conflicts



Taxon	Nucleotide position		
	1	2	3
A	G	G	G
B	G	T	G
C	T	G	T
D	T	T	T

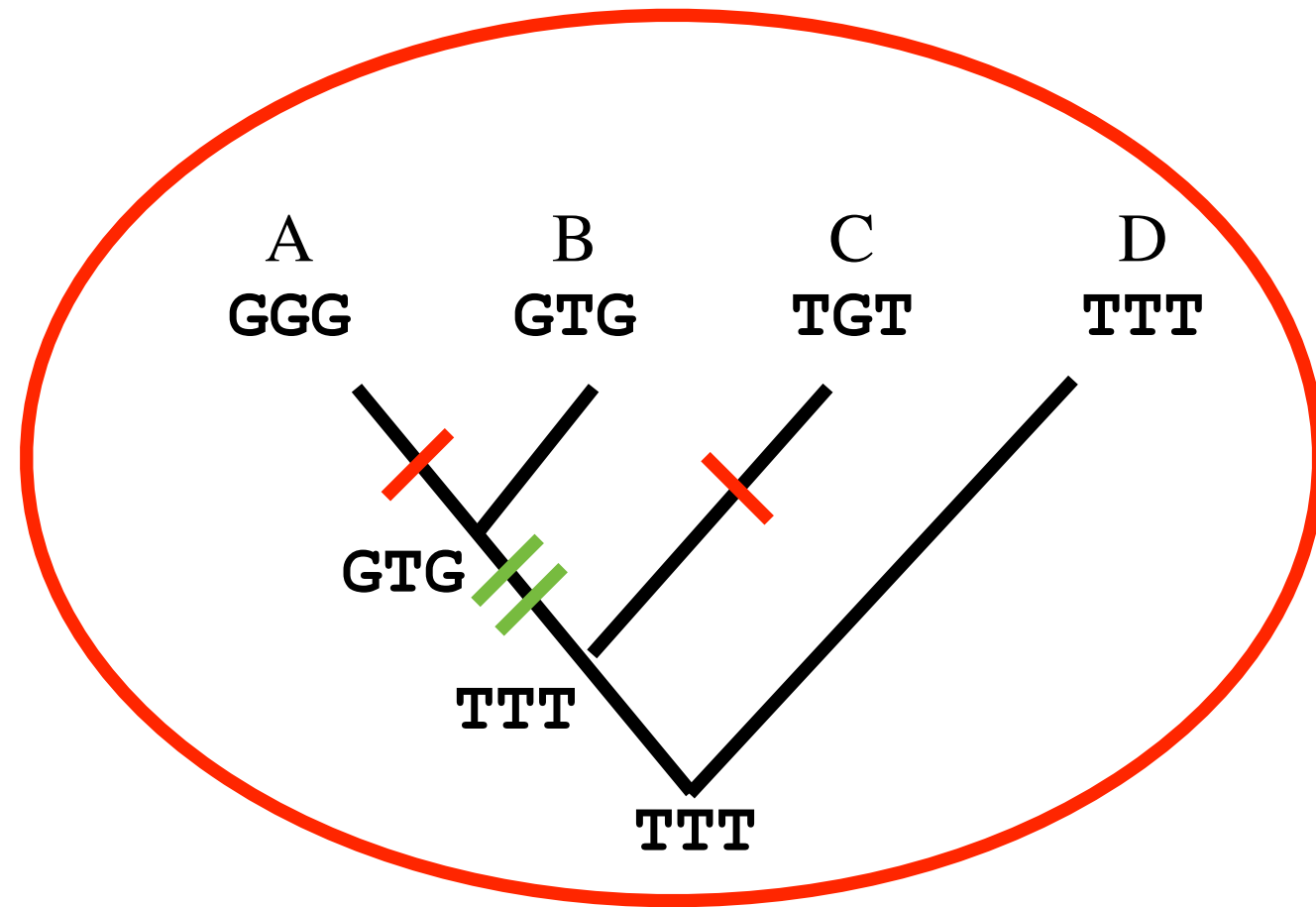
Phylogenetic reconstruction

A B C D

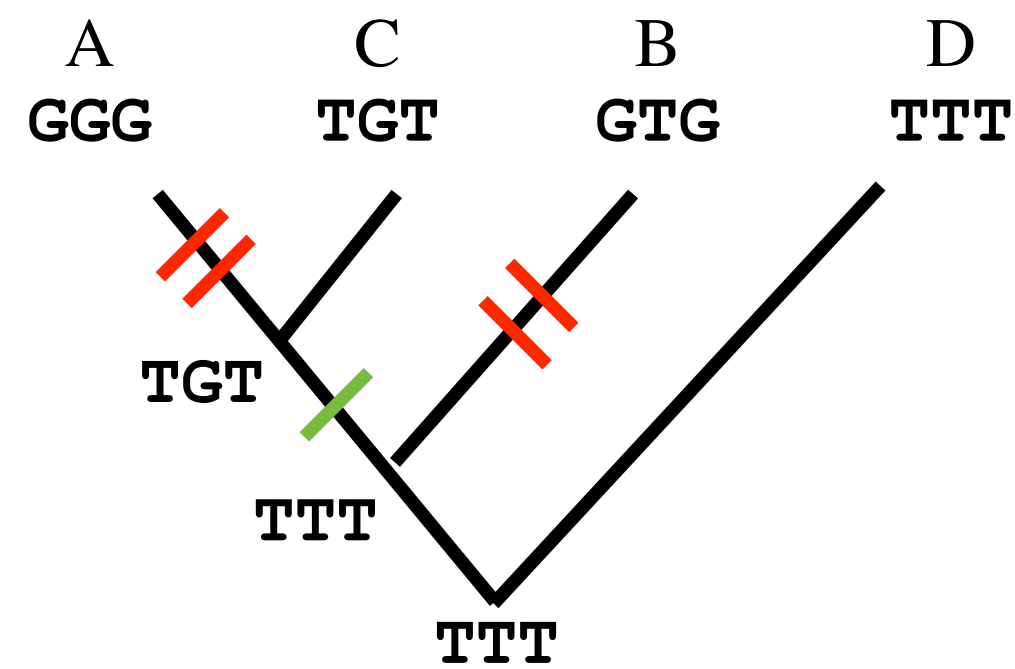


Taxon	Nucleotide position		
	1	2	3
A	G	G	G
B	G	T	G
C	T	G	T
D	T	T	T

Several characters: choose shortest tree
(equivalent to fewer assumptions of homoplasy)



Total length of tree: 4



Total length of tree: 5

Maximum Parsimony

- Maximum parsimony: the best tree is the shortest tree (the tree requiring the smallest number of mutational events)
 - This corresponds to the tree that implies the least amount of homoplasy (convergent evolution, reversals)
 - How do we find the best tree for a given data set?
-

The Fitch Algorithm

GCTGGT > GCUGGU > ALA-GLY
CGACCA

$\int_a^b \epsilon$ Θ $\sqrt{17}$ $+$ $\int \delta e^{i\pi}$ $= -1$
 ∞ $\{2.7182818284\}$ λ
 χ^2 Σ \gg \approx λ
 ϕ ϕ ϵ ρ τ υ θ ι \omicron π σ δ ϕ γ η ξ κ λ

Maximum Parsimony: Algorithms

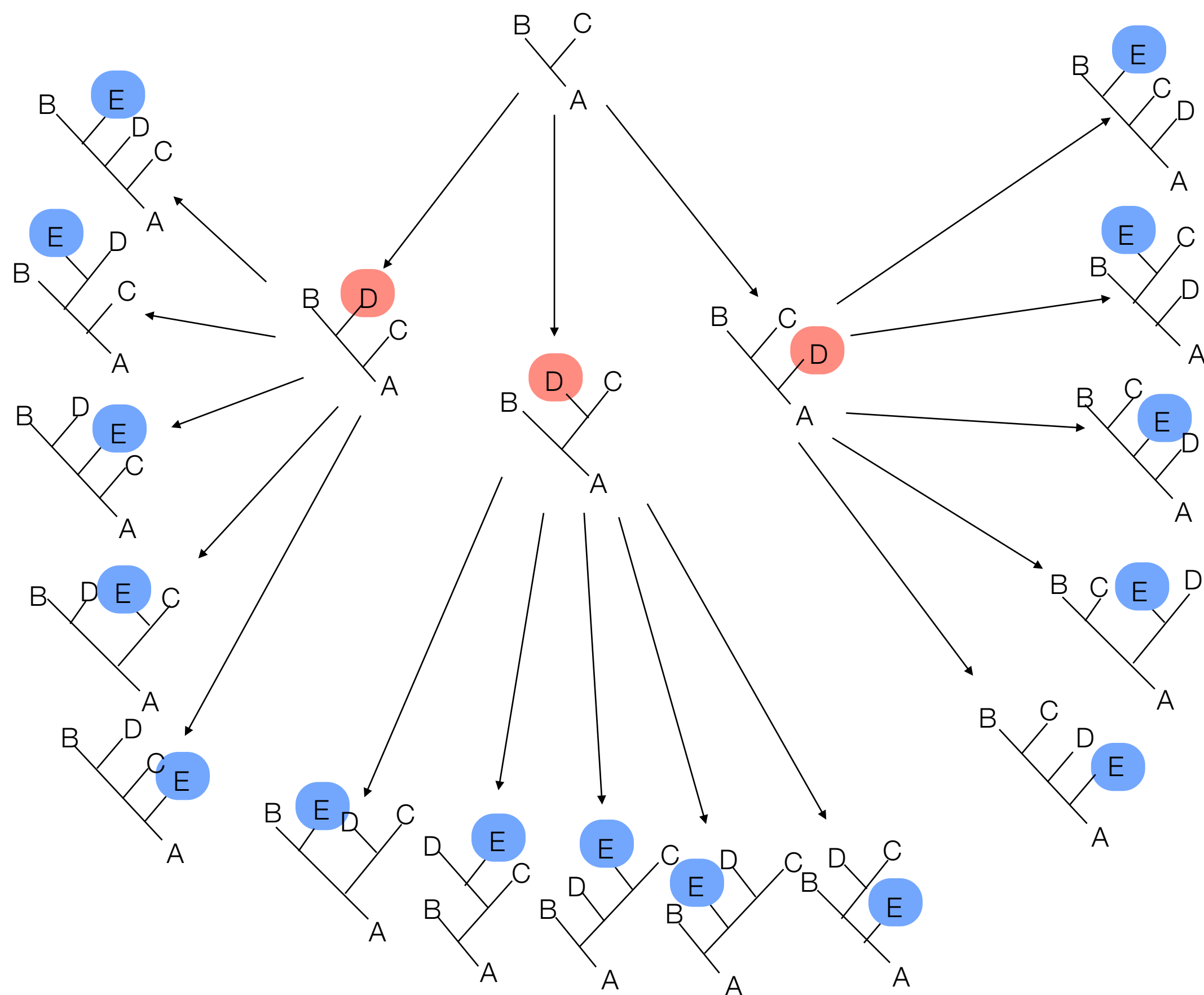
How do we find the maximum parsimony tree for a given data set?

1. Construct list of all possible trees for data set
 2. For each tree: determine length, add to list of lengths
 3. When finished: select shortest tree from list
 4. If several trees have the same length, then they are equally good (equally parsimonious)
-

Maximum Parsimony: Sub-problems

- We need algorithm for constructing list of all possible trees
- We need algorithm for determining length of given tree

Constructing list of all possible unrooted trees

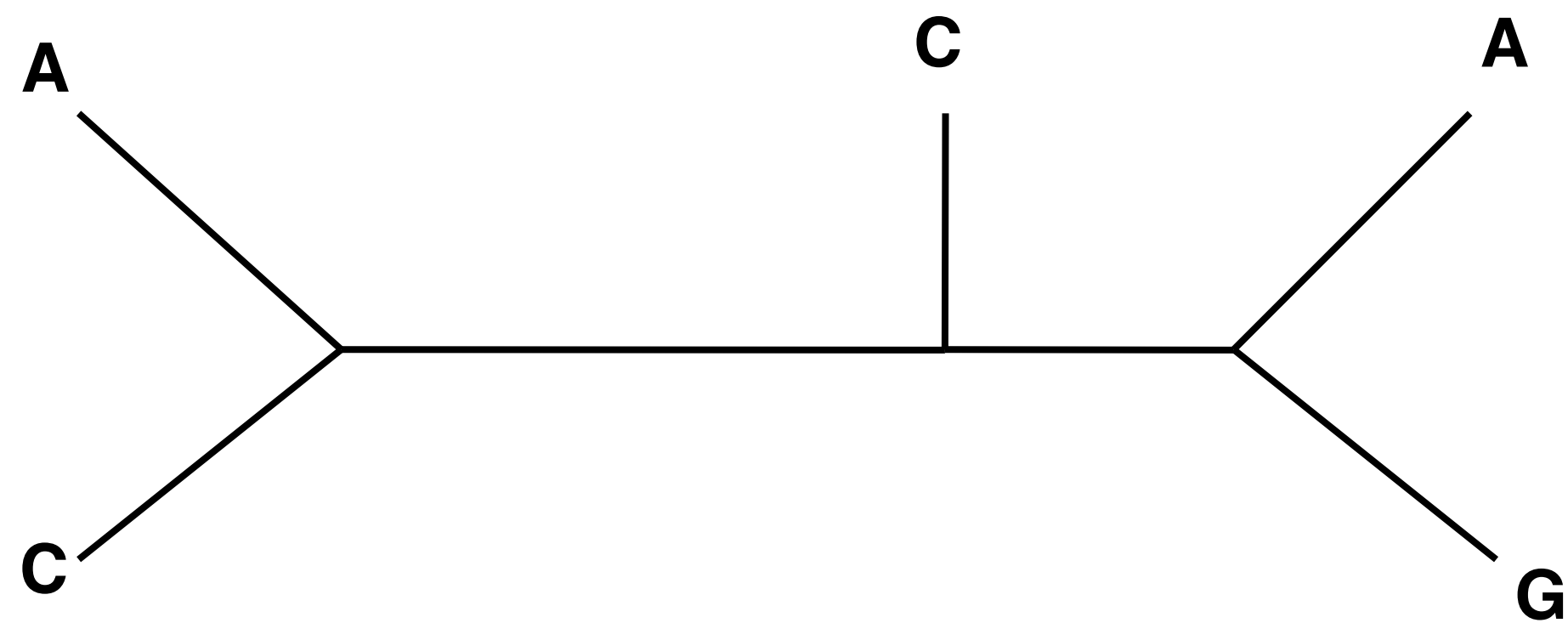


1. Construct unrooted tree from first three taxa.
There is only one way of doing this
2. Starting from (1), construct the three possible derived trees by adding taxon 4 to each internal branch
3. From each of the trees constructed in step (2), construct the five possible derived trees by adding taxon 5 to each internal branch.
4. Continue until all taxa have been added in all possible locations

Maximum Parsimony: problems

- We need algorithm for constructing list of all possible trees ✓
 - We need algorithm for determining length of given tree
-

Algorithm for determining length of given tree: Fitch

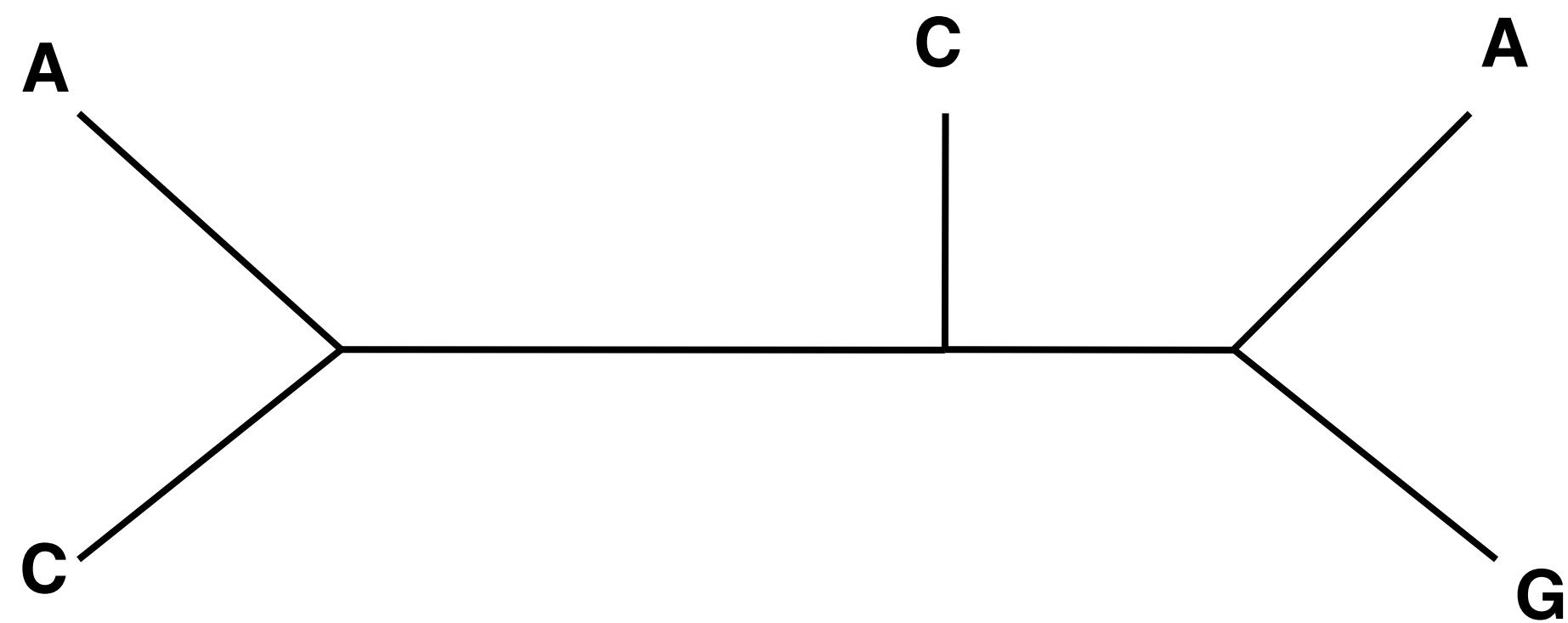


What is the length of this tree? (How many mutational steps are required?)

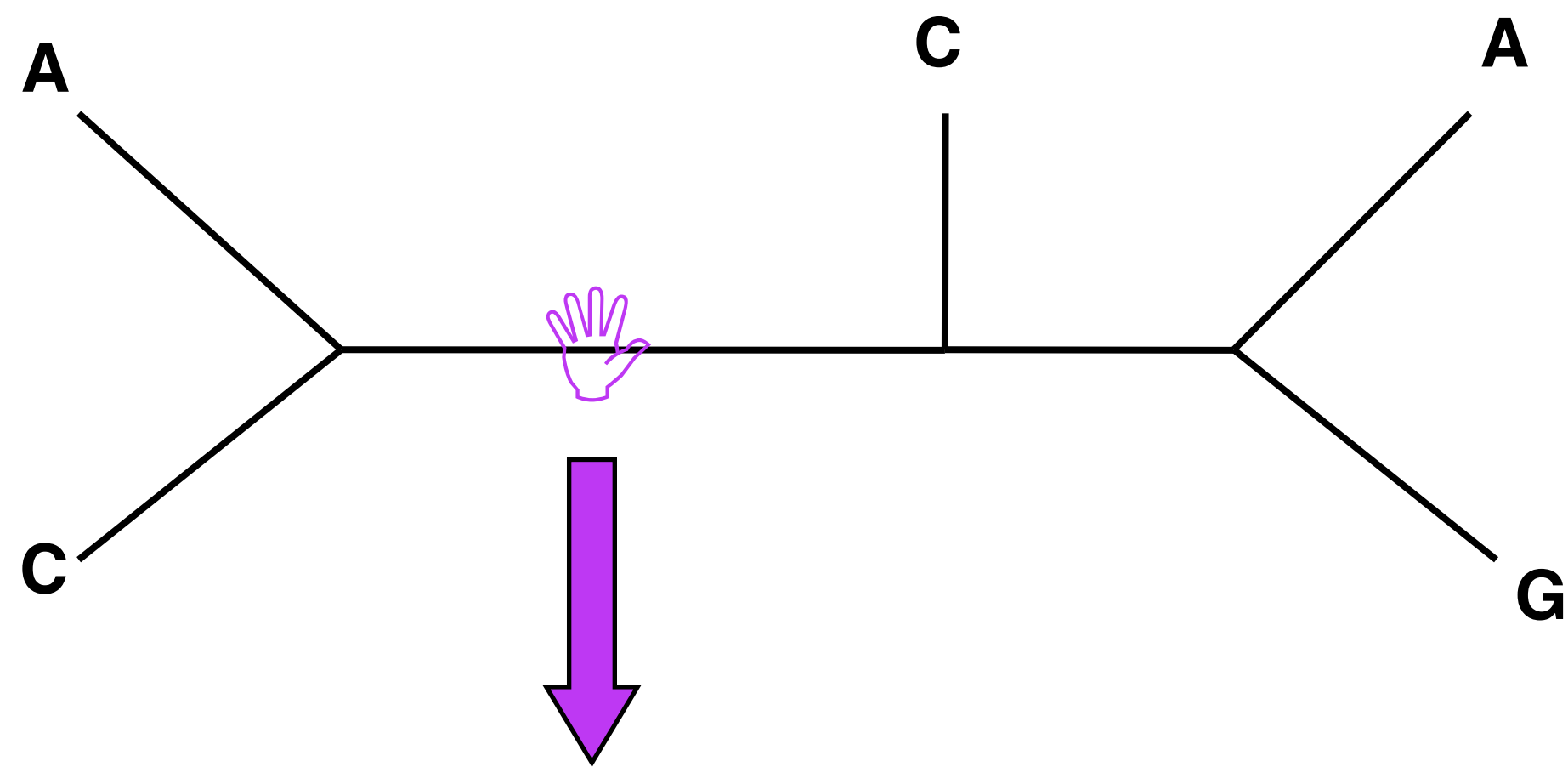
Algorithm for determining length of given tree: Fitch

- Root the tree at an arbitrary internal node (or internal branch)
 - Visit an internal node x for which no state set has been defined, but where the state sets of x 's immediate descendants (y,z) have been defined.
 - If the state sets of y,z have common states, then assign these to x .
 - If there are no common states, then assign the union of y,z to x , and increase tree length by one.
 - Repeat until all internal nodes have been visited. Note length of current tree.
-

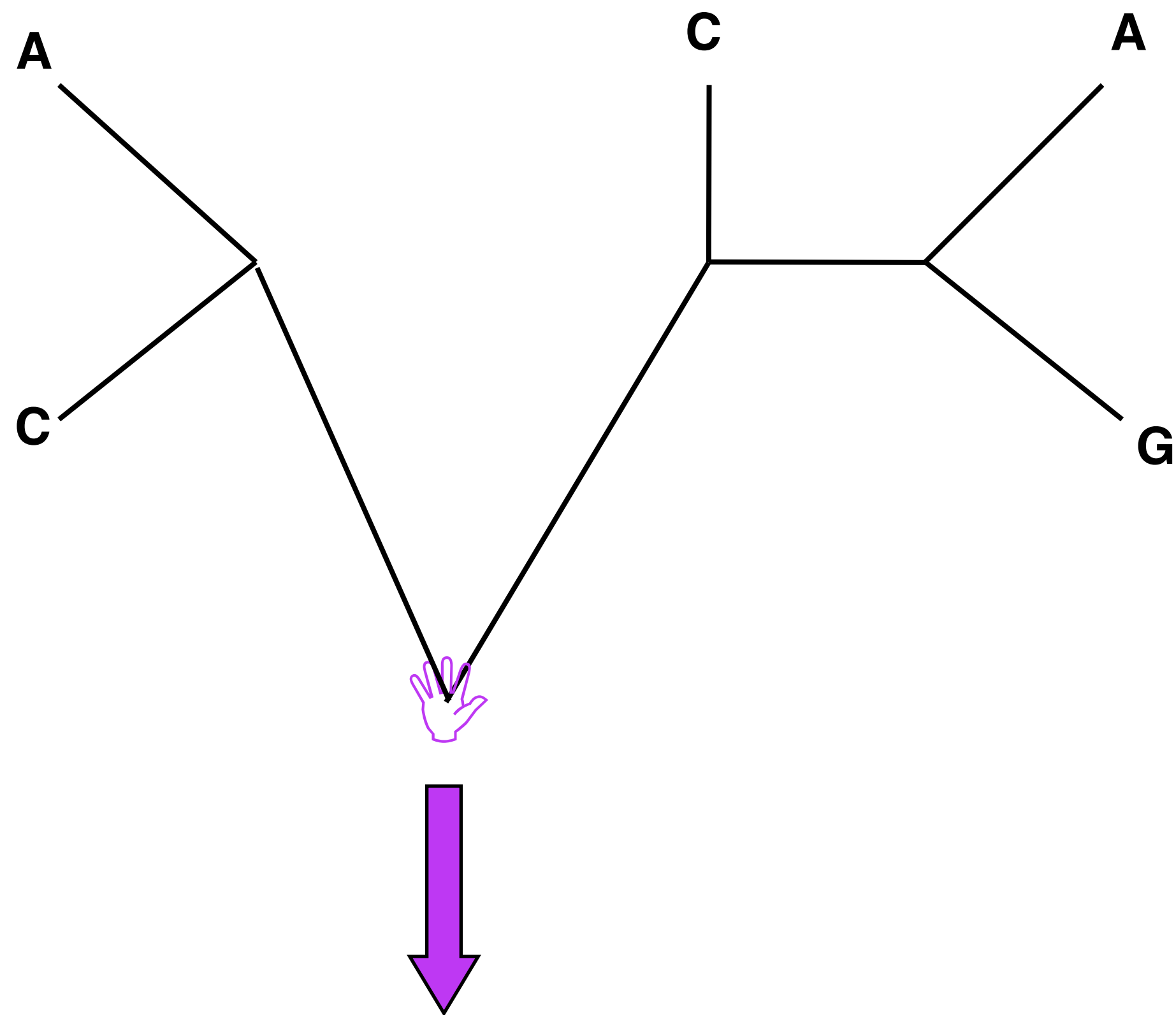
Algorithm for determining length of given tree: Fitch



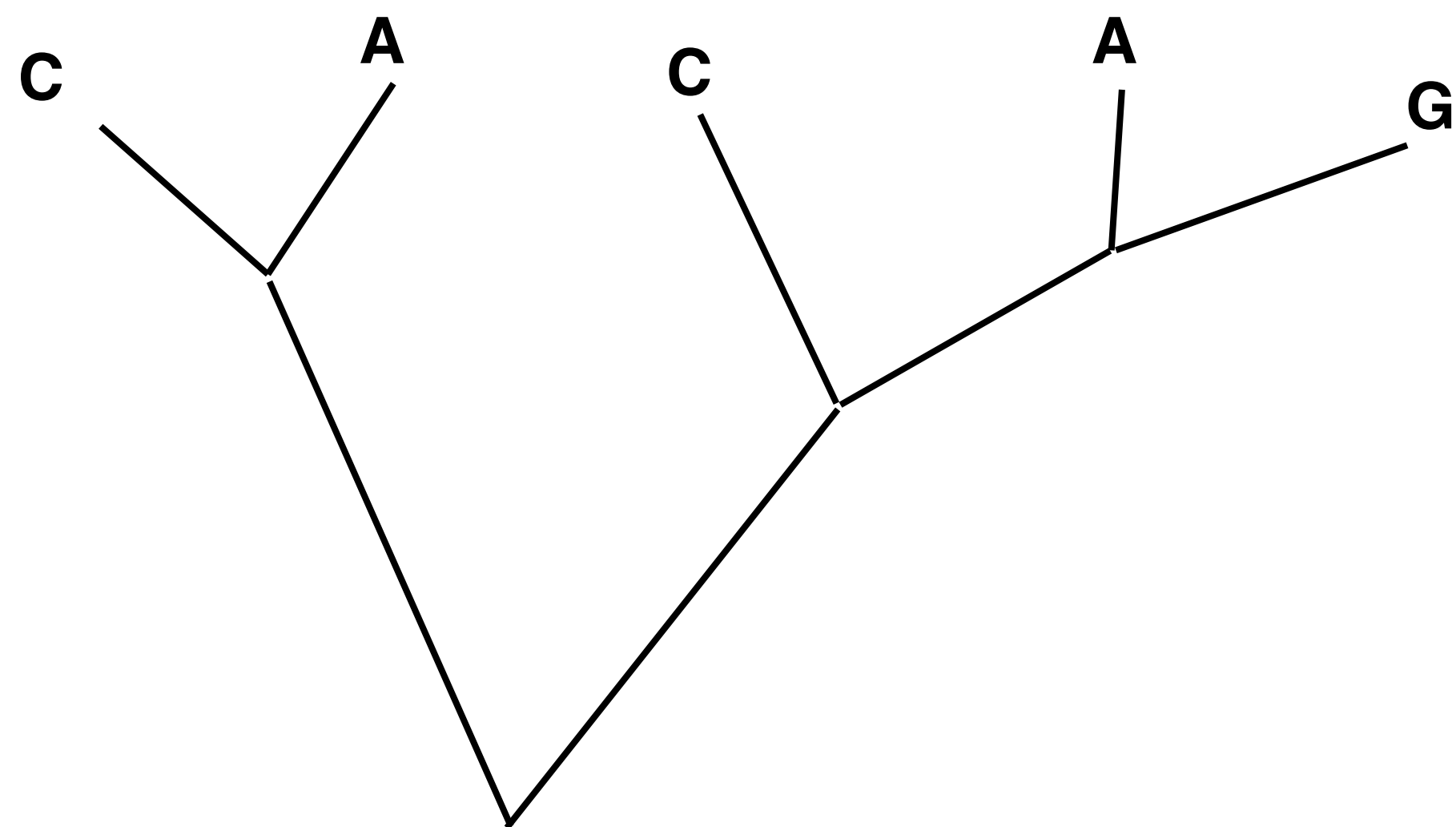
Algorithm for determining length of given tree: Fitch



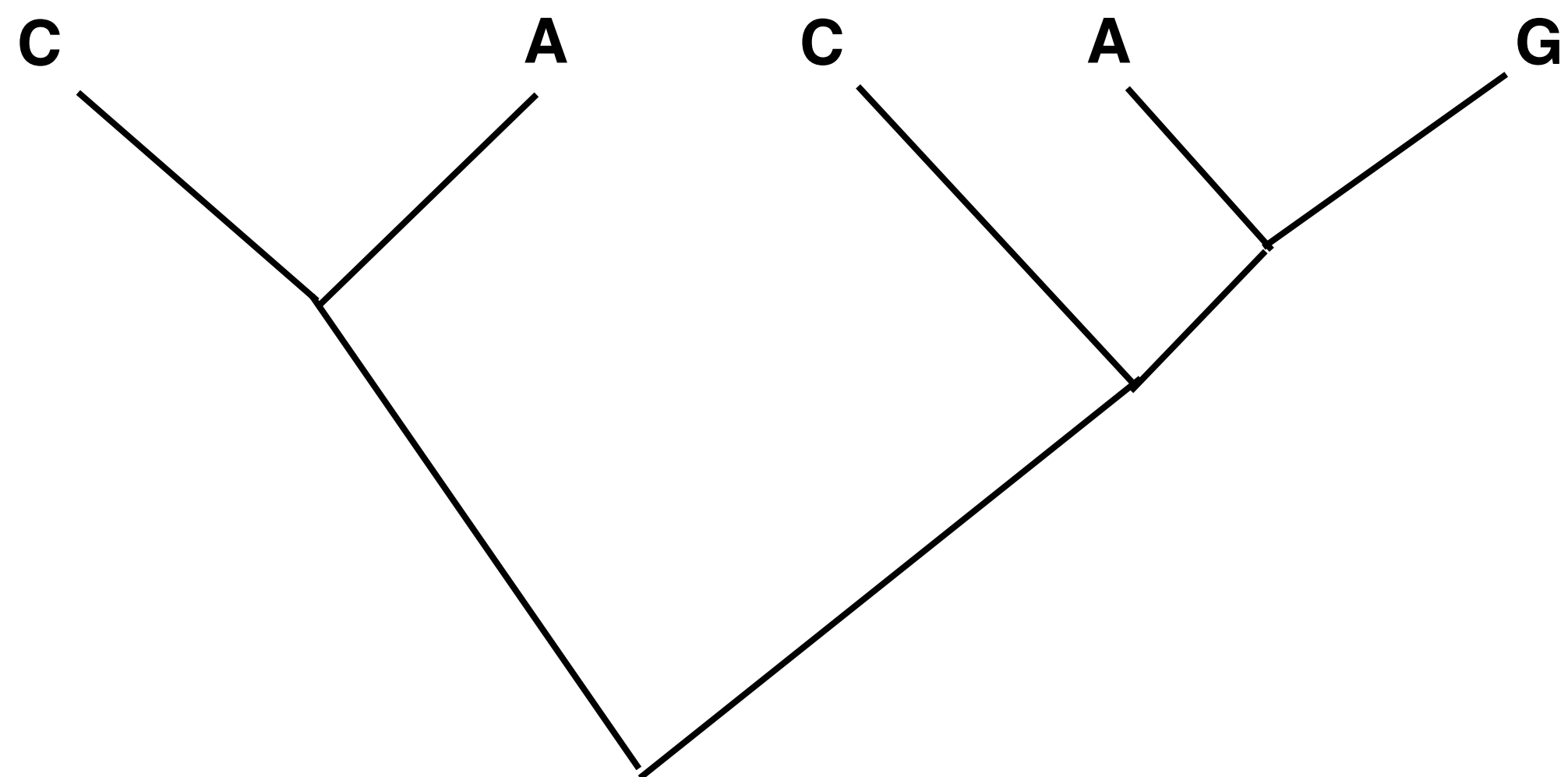
Algorithm for determining length of given tree: Fitch



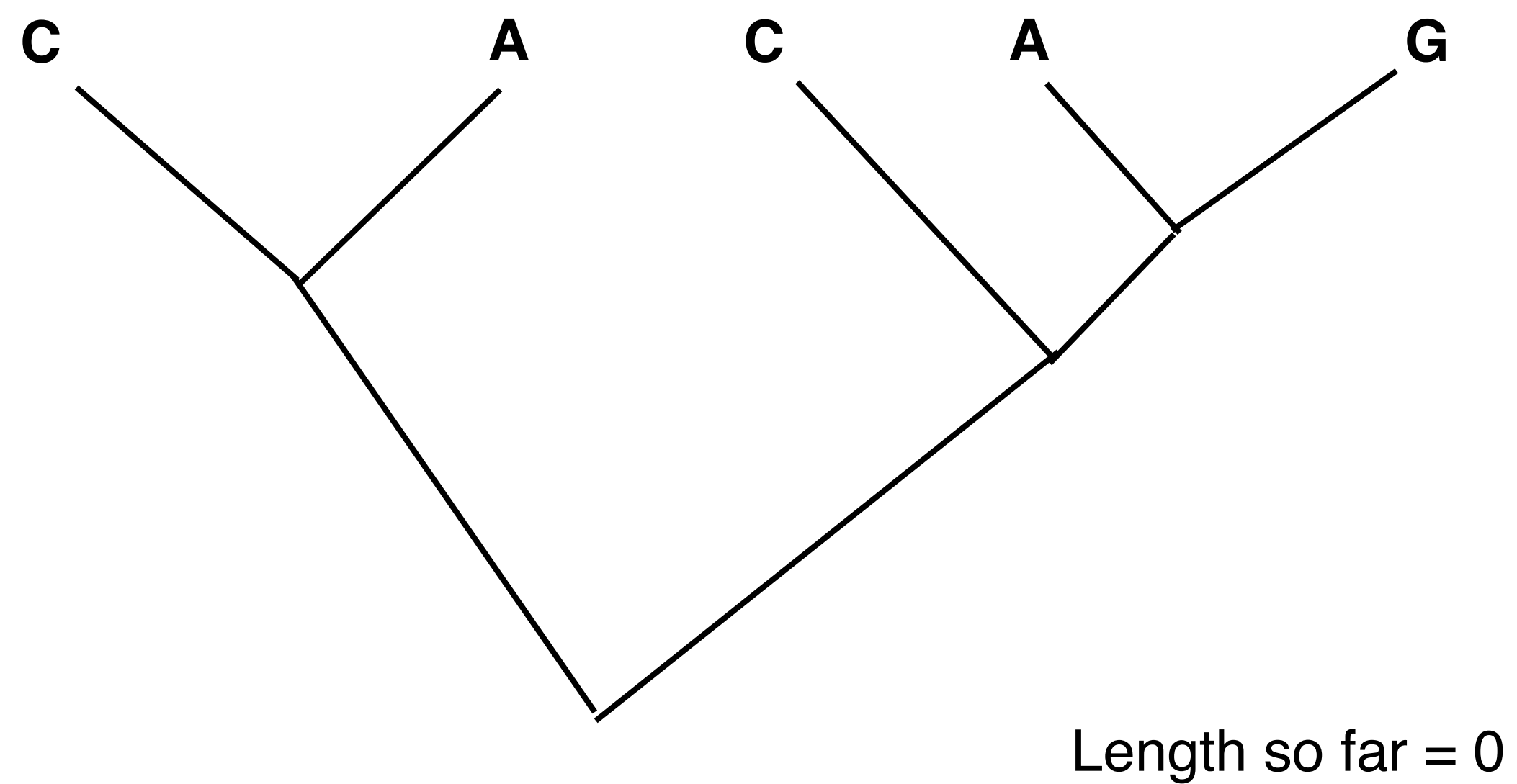
Algorithm for determining length of given tree: Fitch



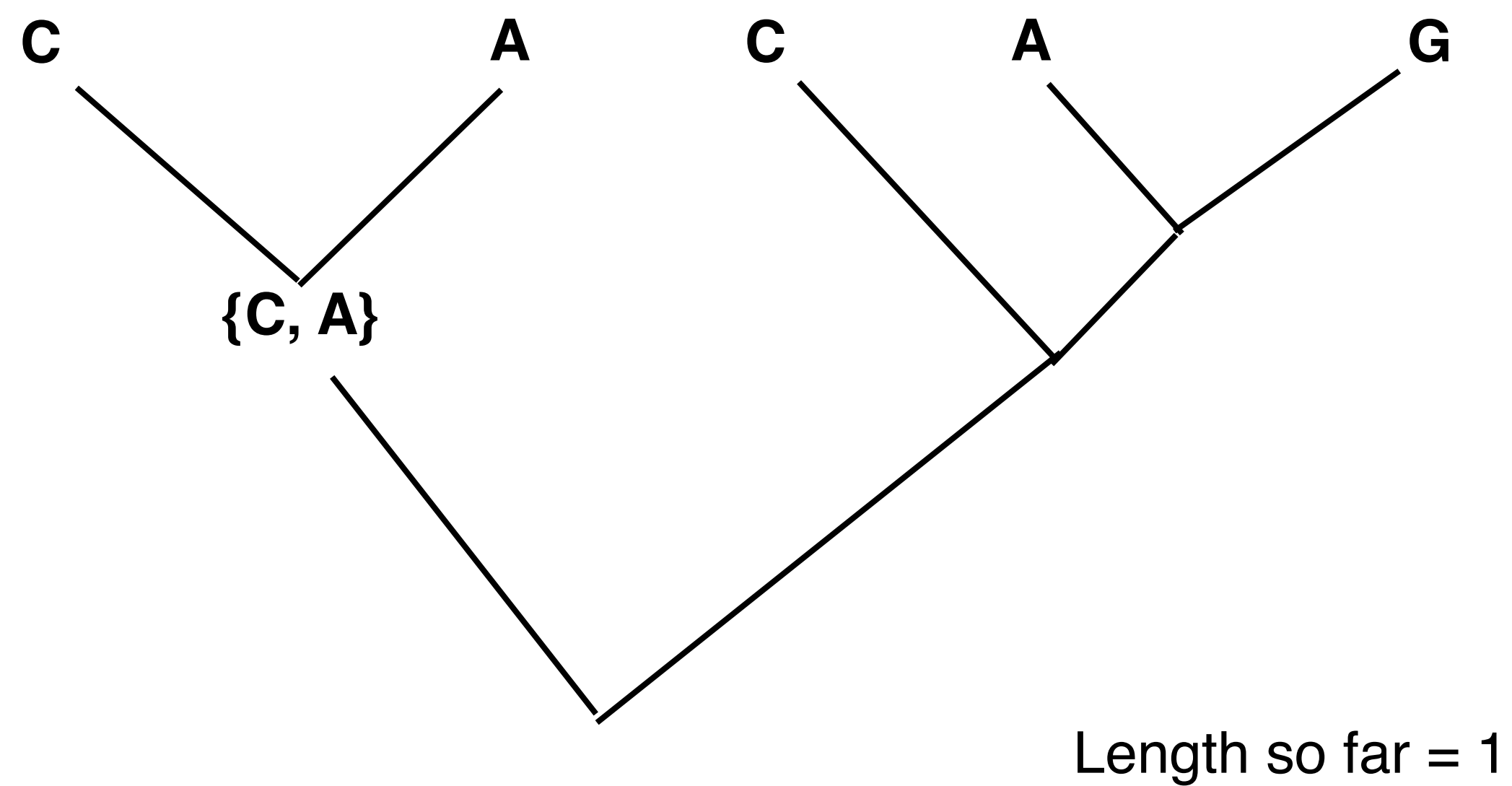
Algorithm for determining length of given tree: Fitch



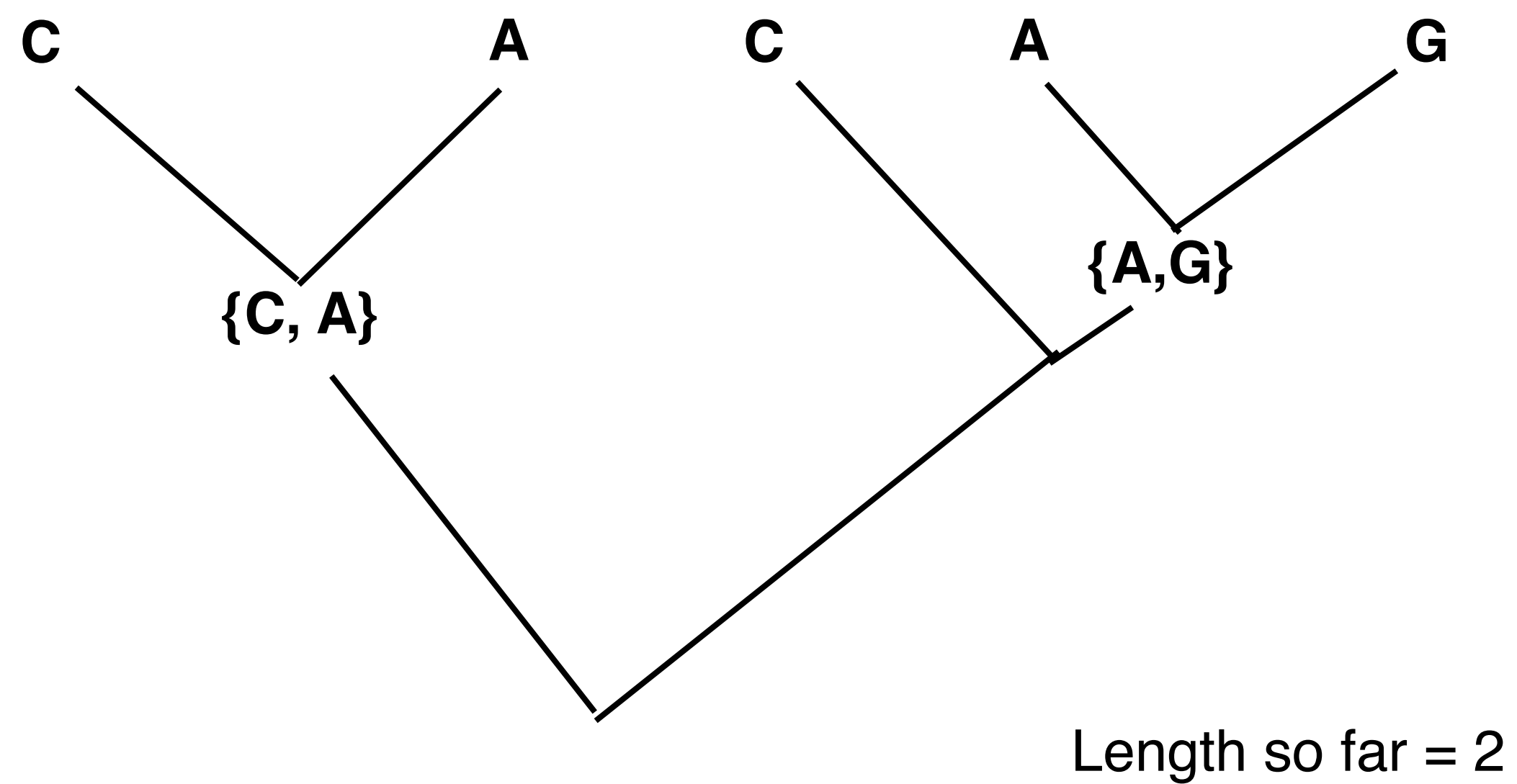
Algorithm for determining length of given tree: Fitch



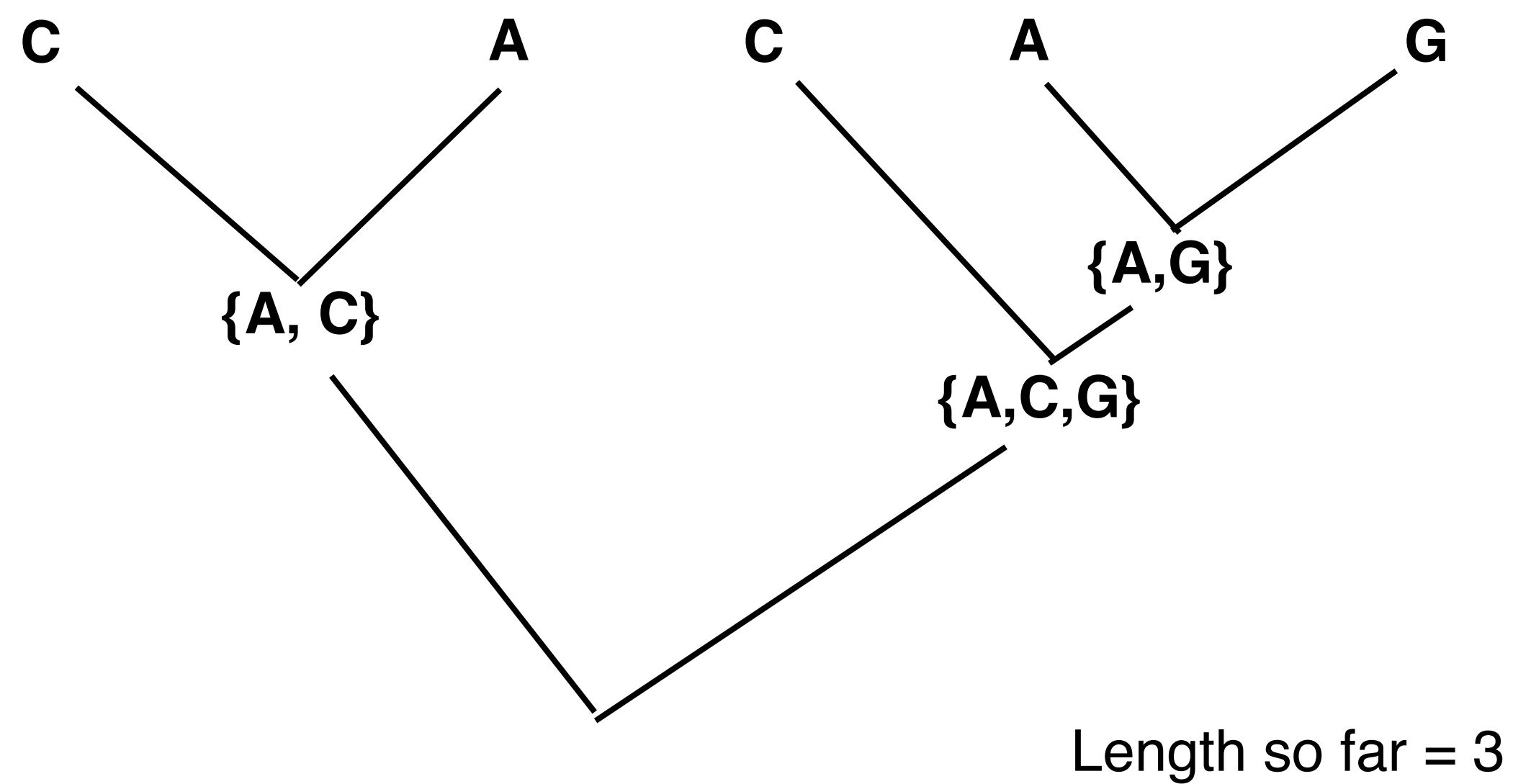
Algorithm for determining length of given tree: Fitch



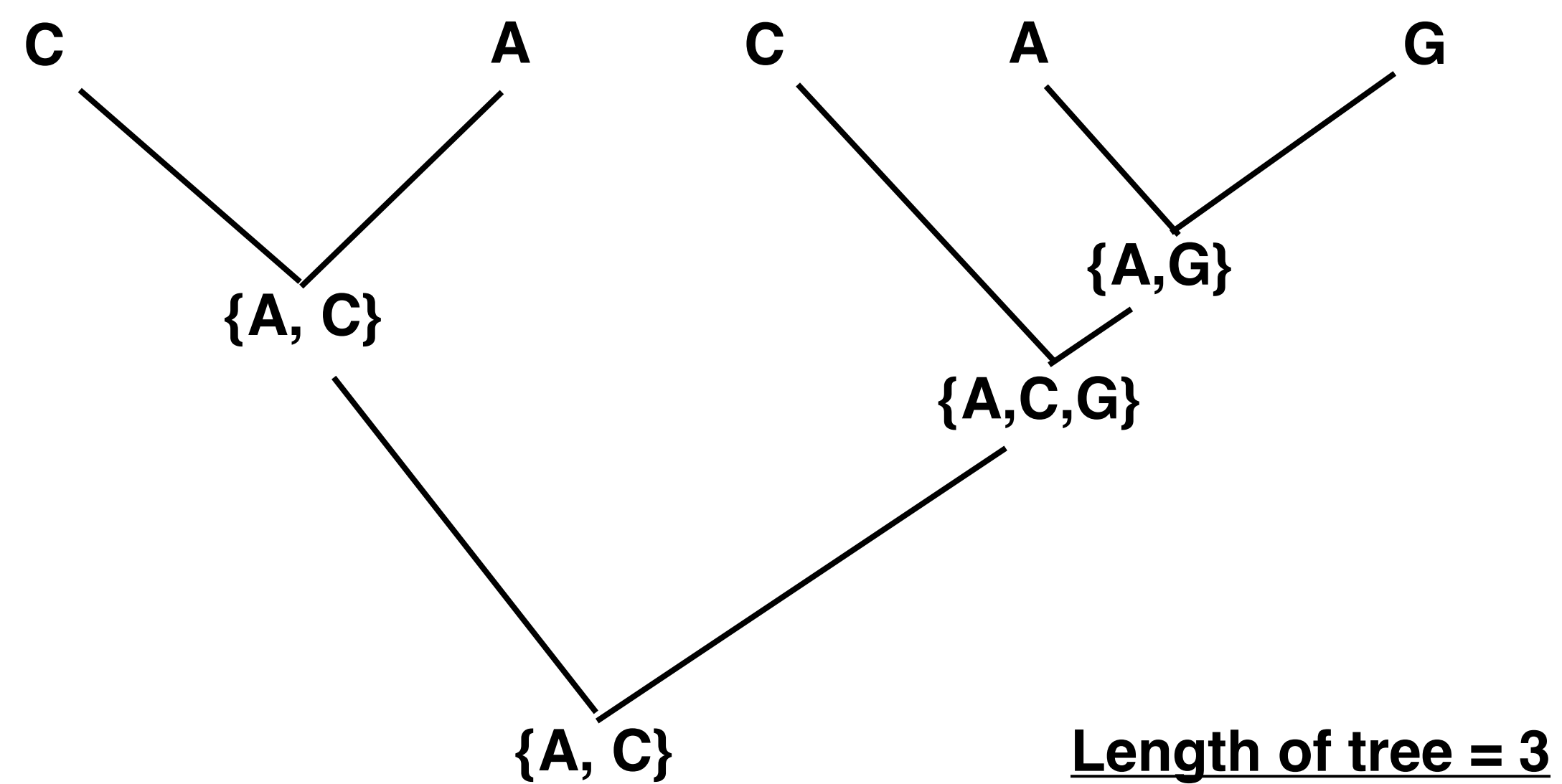
Algorithm for determining length of given tree: Fitch



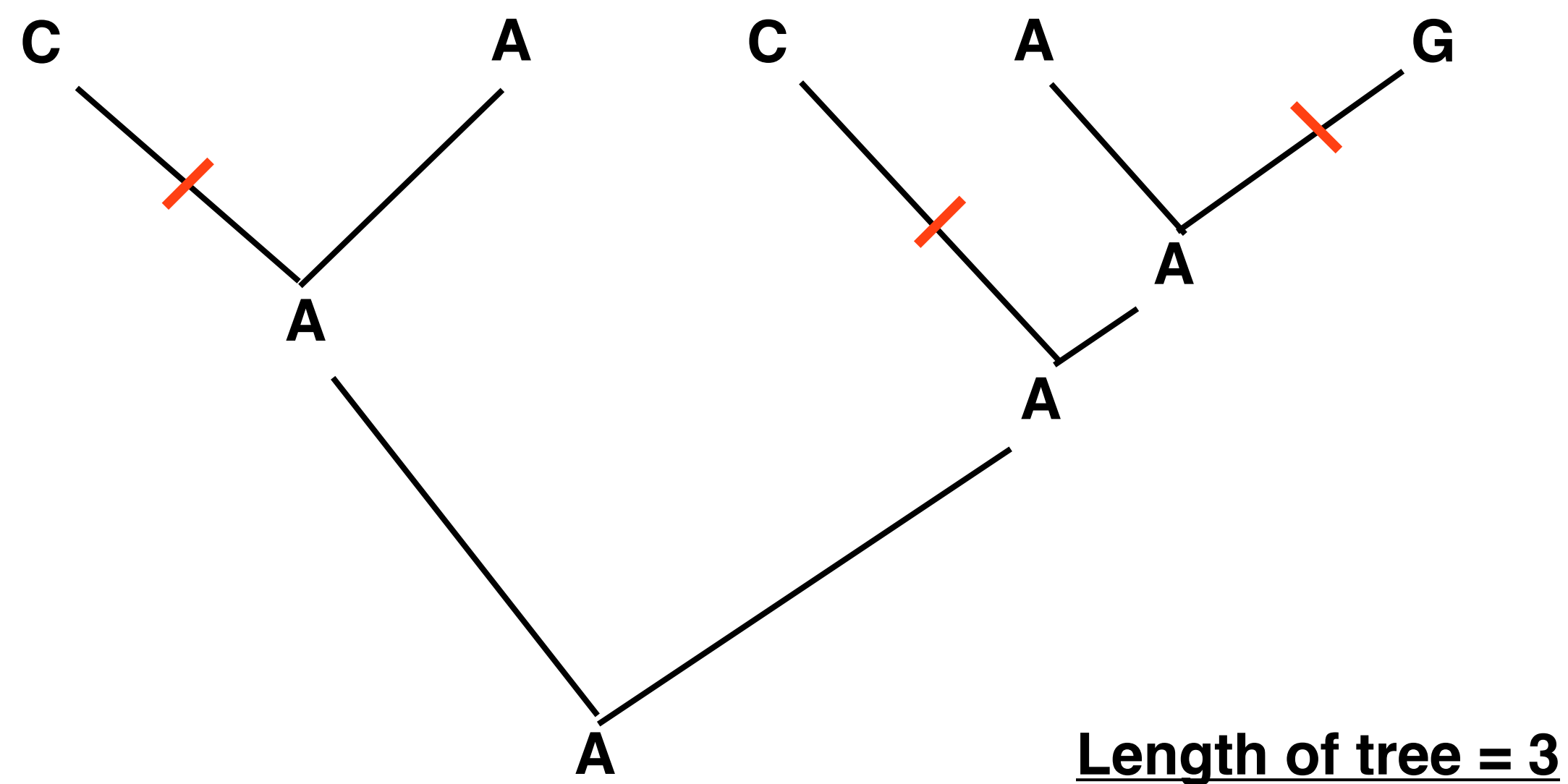
Algorithm for determining length of given tree: Fitch



Algorithm for determining length of given tree: Fitch



Algorithm for determining length of given tree: Fitch



One possible reconstruction (several others exist)

Maximum Parsimony: problems

- We need algorithm for constructing list of all possible trees ✓
 - We need algorithm for determining length of given tree ✓
-

Searching Tree Space

GCTGGT > GCUGGU > ALA-GLY
CGACCA <

$$\int_a^b \epsilon \Theta + \Omega \int \delta e^{i\pi} = -1$$

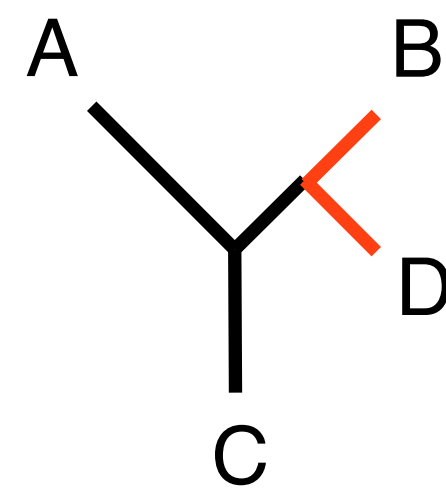
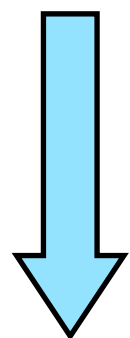
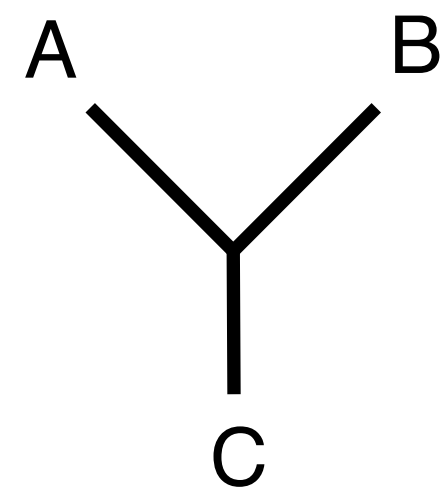
∞

$\{2.7182818284\}$ λ

χ^2 $\Sigma!$ \gg \approx

ϕ τ θ i o π σ δ ϕ γ ξ κ

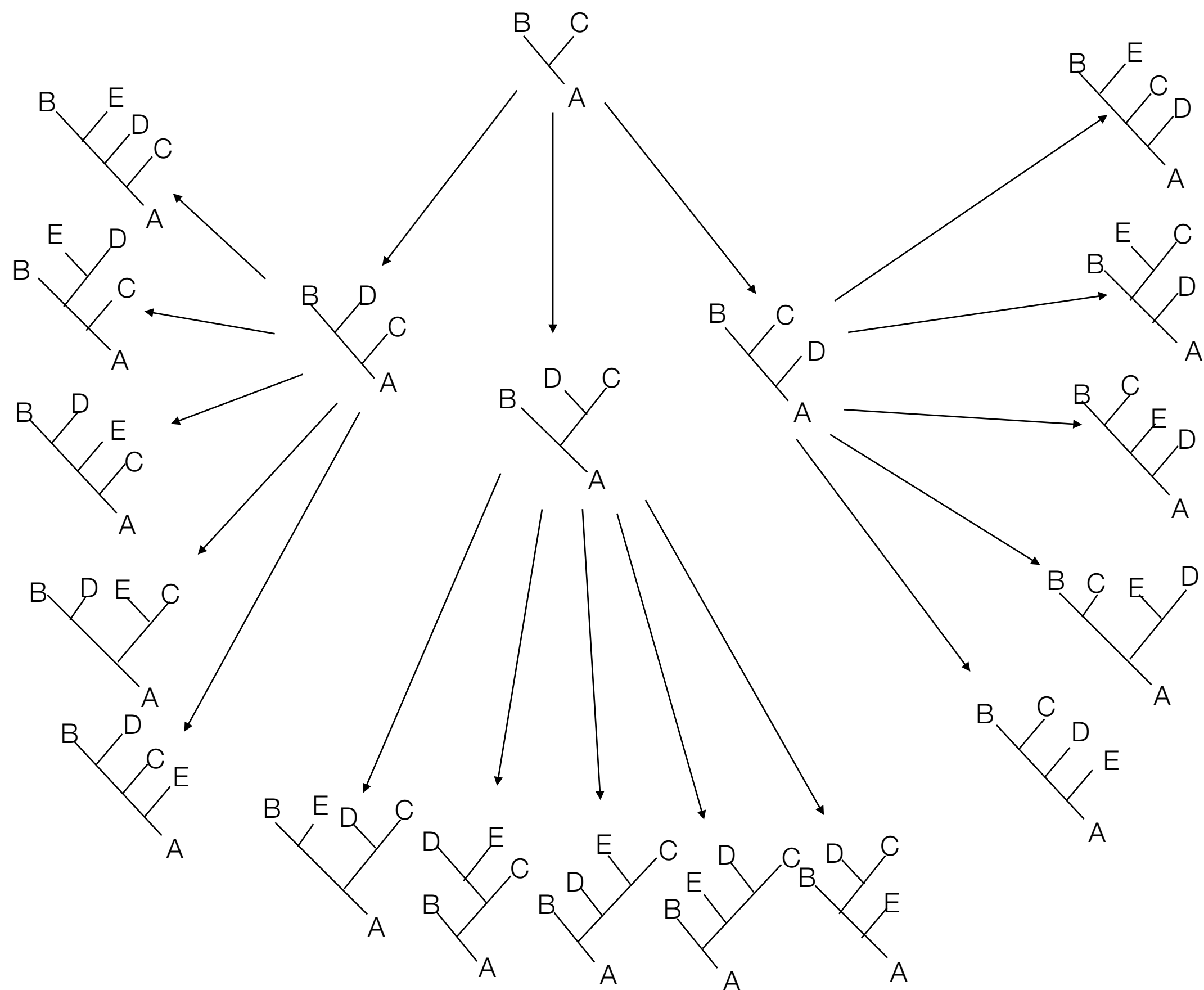
How many branches are there on an unrooted tree with n tips?



- There is only one way of constructing the first tree. This tree has 3 tips and 3 branches
- Each time an extra taxon is added, two branches are created.
- A tree with n tips will therefore have the following number of branches:

$$\begin{aligned}
 N_{\text{branches}} &= 3+(n-3)*2 \\
 &= 3+2n-6 \\
 &= 2n-3
 \end{aligned}$$

How many unrooted trees are there?



- A tree with n tips has $2n-3$ branches
- For each tree with n tips, we can therefore construct $2n-3$ derived trees (which each have $n+1$ tips).

How many unrooted trees are there?

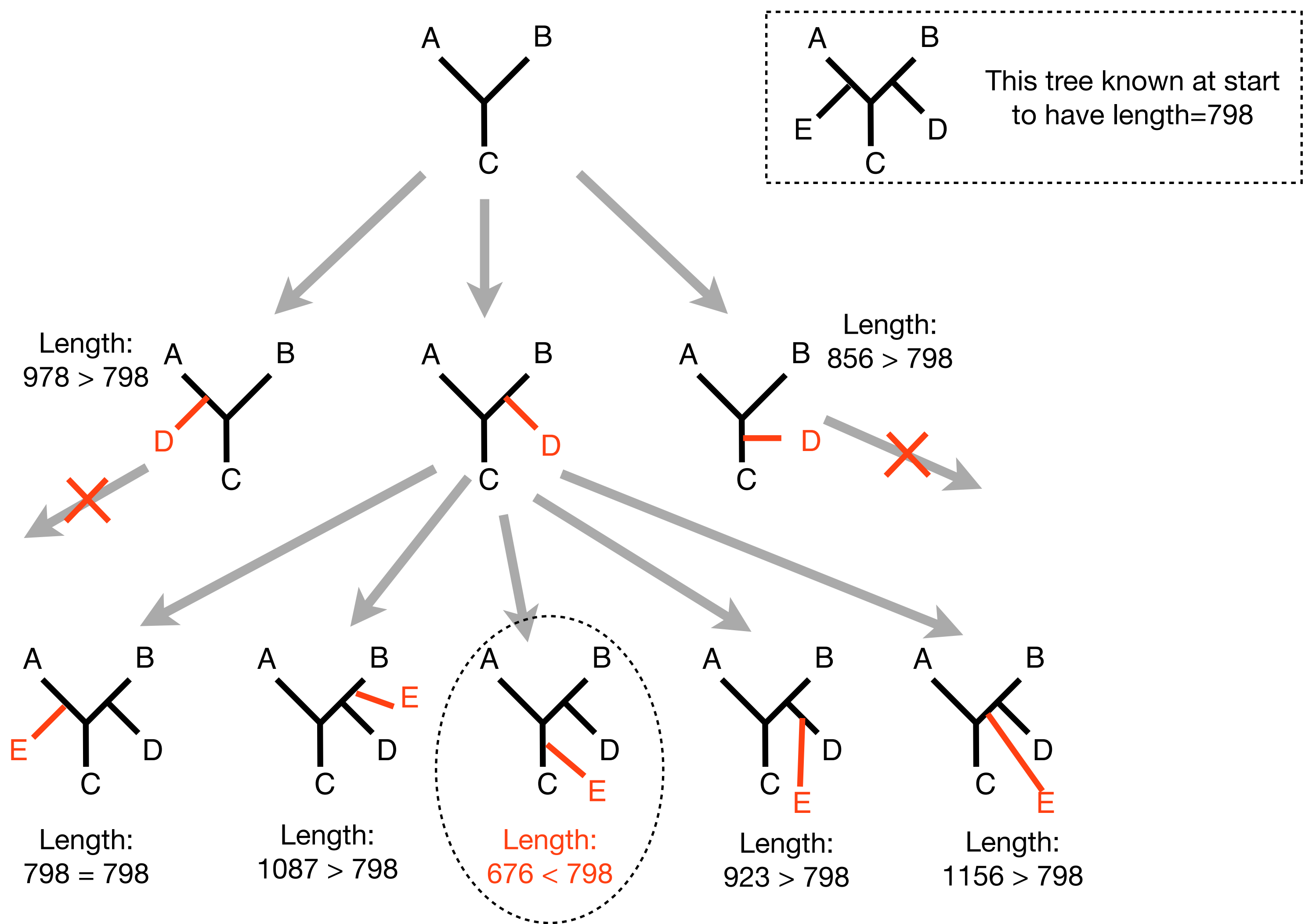
N_{tips}	N_{trees}	$N_{\text{branches}} = N_{\text{derived trees}}$
3	1	$2 \times 3 - 3 = 3$
4	1×3	$2 \times 4 - 3 = 5$
5	$1 \times 3 \times 5$	$2 \times 5 - 3 = 7$
6	$1 \times 3 \times 5 \times 7$	$2 \times 6 - 3 = 9$
7	$1 \times 3 \times 5 \times 7 \times 9$	$2 \times 7 - 3 = 11$
8	$1 \times 3 \times 5 \times 7 \times 9 \times 11$	$2 \times 8 - 3 = 13$
9	$1 \times 3 \times 5 \times 7 \times 9 \times 11 \times 13$...

$$N_{\text{trees with } n \text{ tips}} = \prod_{i=2}^{n-1} (2i - 3)$$

Exhaustive search impossible for large data sets

No. taxa	No. trees
3	1
4	3
5	15
6	105
7	945
8	10,395
9	135,135
10	2,027,025
11	34,459,425
12	654,729,075
13	13,749,310,575
14	316,234,143,225
15	7,905,853,580,625

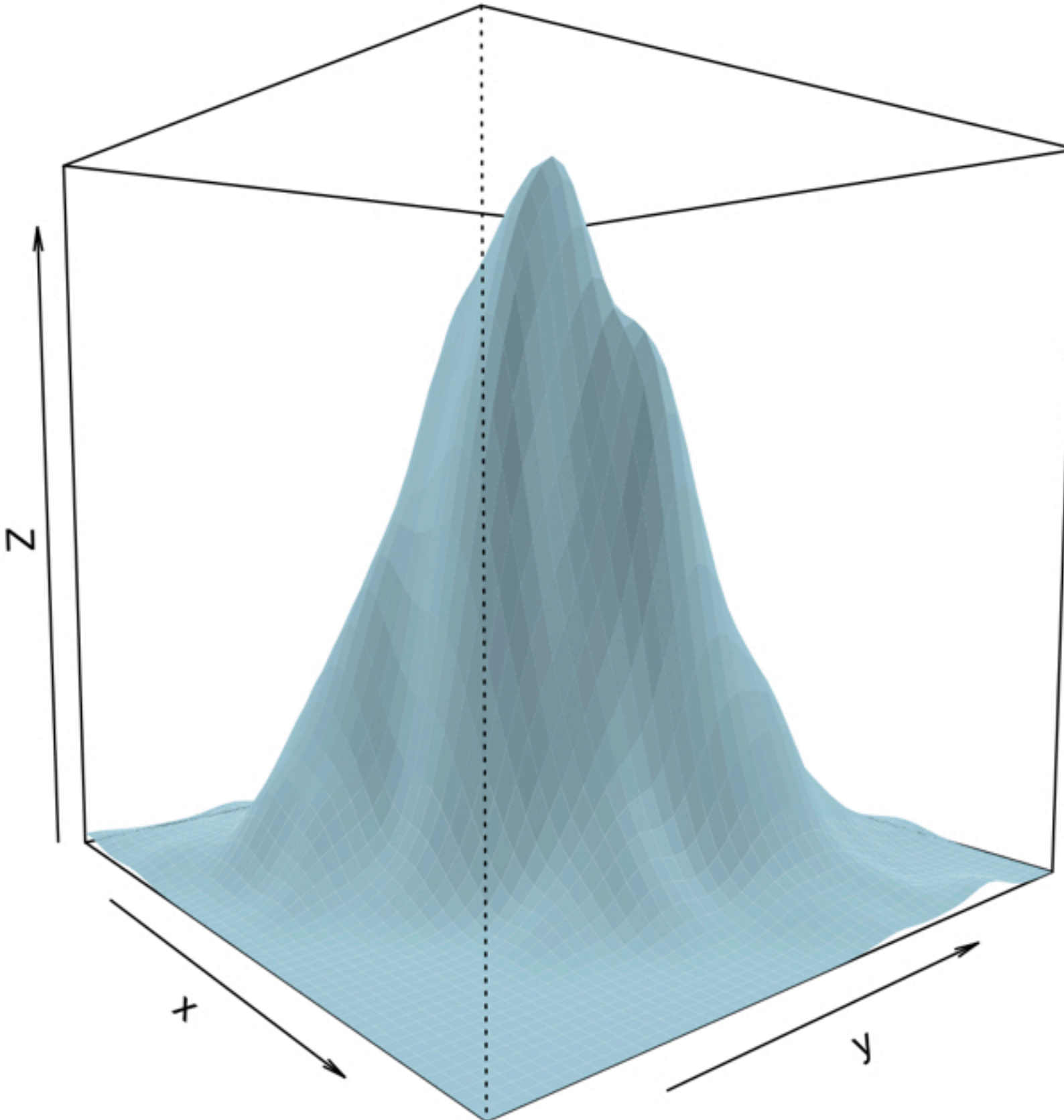
Branch and bound: shortcut to perfection



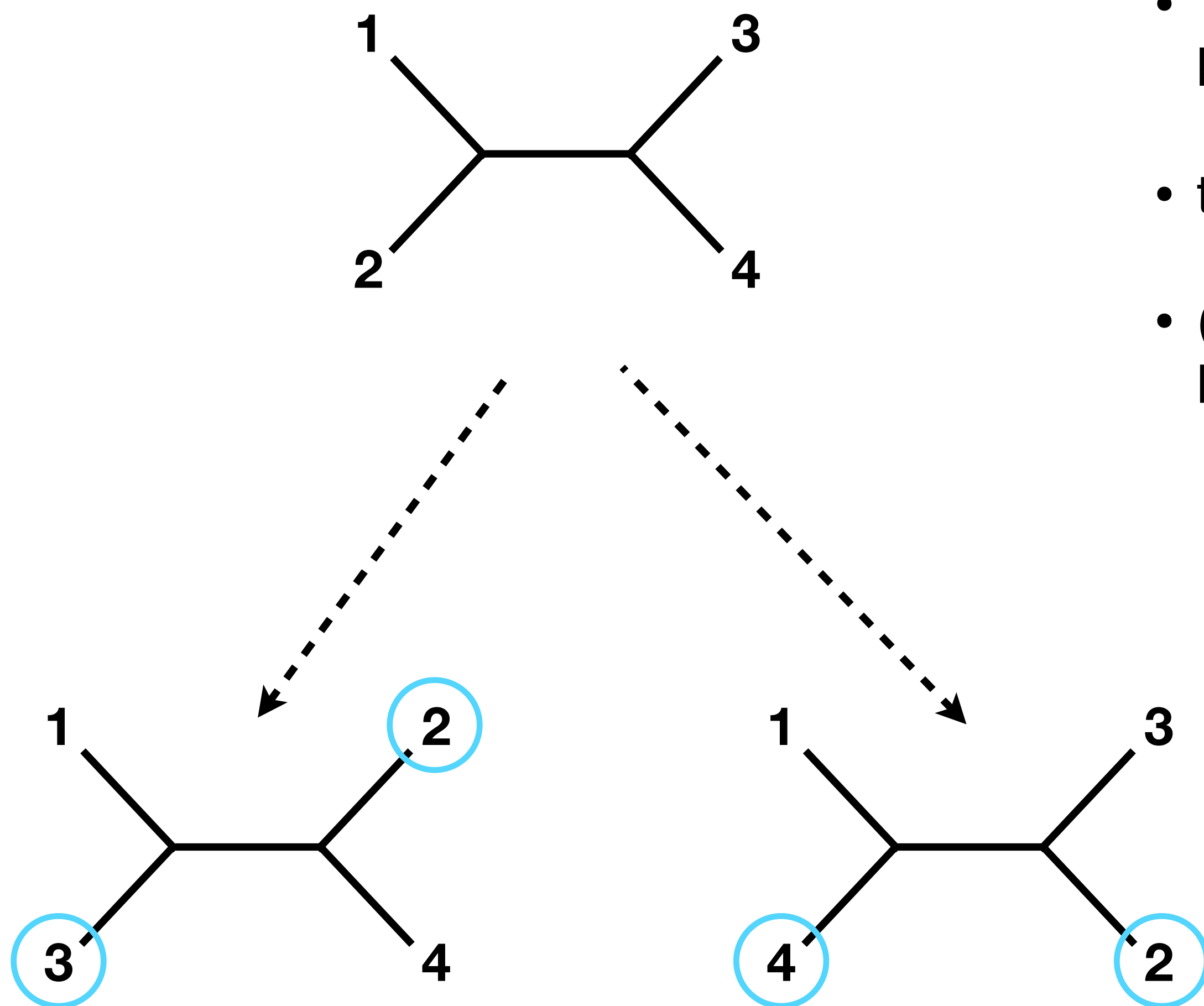
Heuristic search

1. Construct initial tree (e.g., sequential addition); determine length
 2. Construct set of “neighboring trees” by making small rearrangements of initial tree; determine lengths
 3. If any of the neighboring trees are better than the initial tree, then select it/ them and use as starting point for new round of rearrangements. (Possibly several neighbors are equally good)
 4. Repeat steps 2+3 until you have found a tree that is better than all of its neighbors.
 5. This tree is a “local optimum” (not necessarily a global optimum!)
-

Heuristic search: hill-climbing

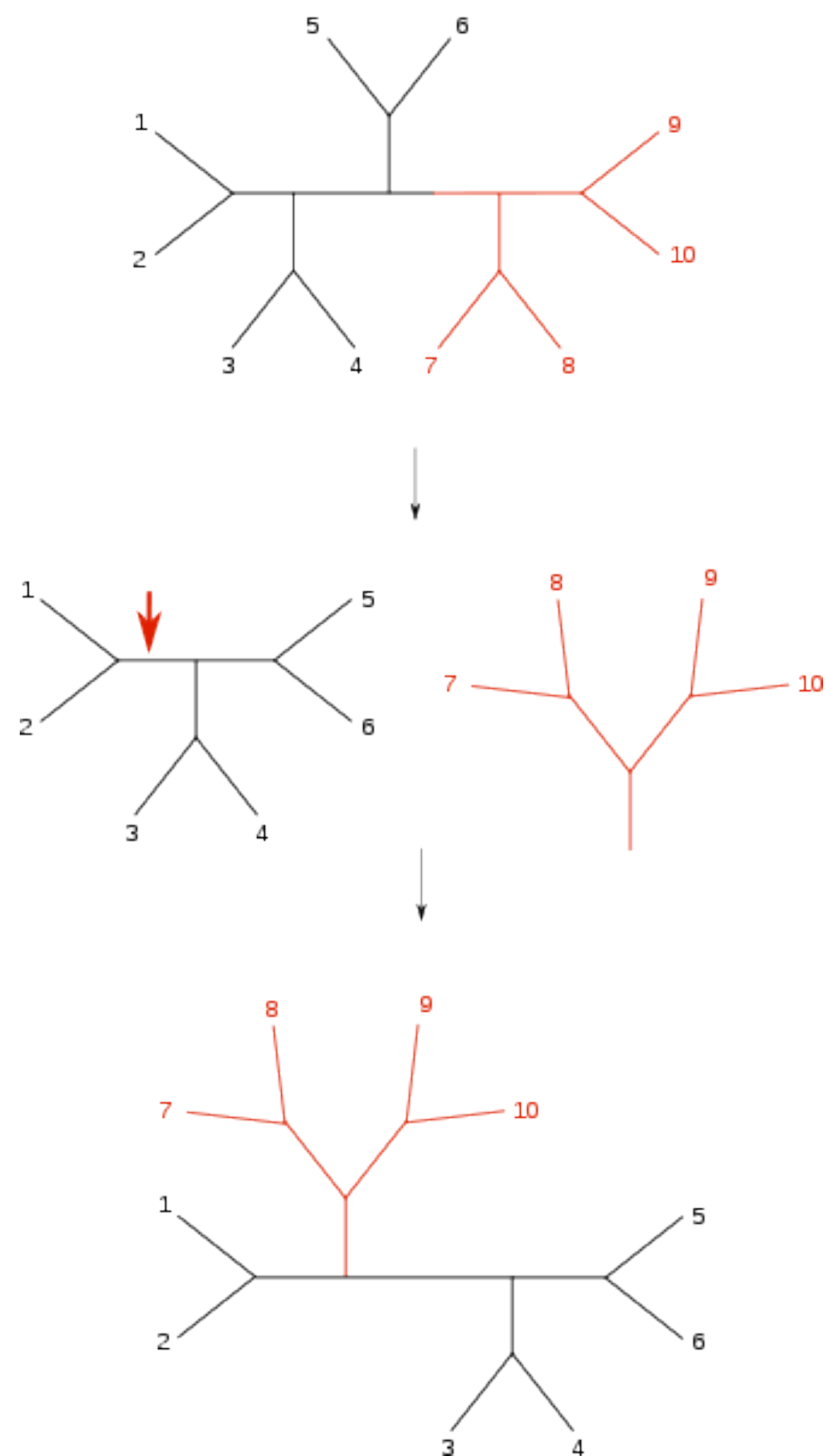


Types of rearrangement I: nearest neighbor interchange (NNI)



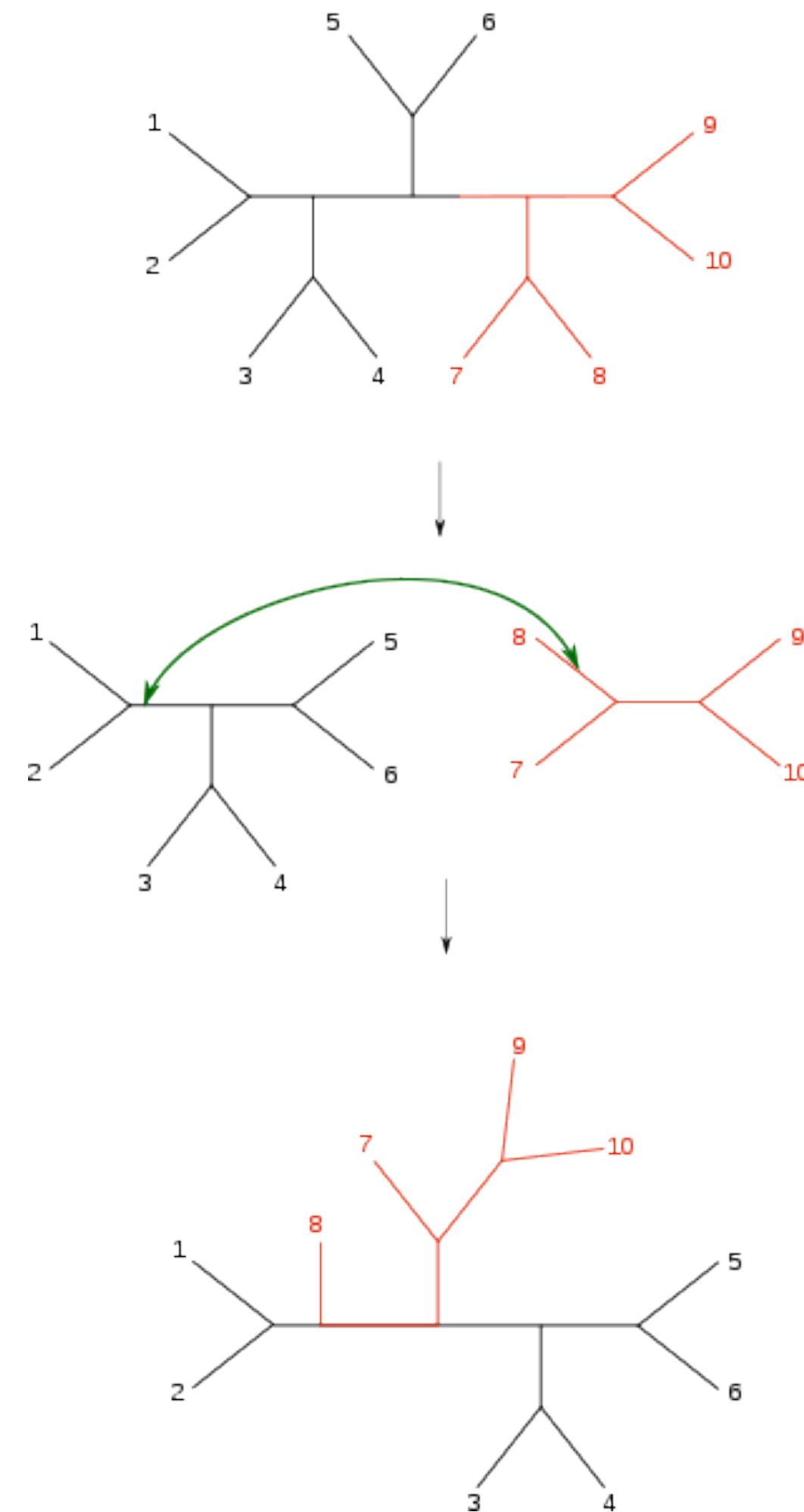
- Two neighboring trees per internal branch:
- tree with n tips has $2(n-3)$ neighbors
- (For example, a tree with 20 tips has 34 neighbors)

Types of rearrangement II: subtree pruning and regrafting (SPR)



- Detach subtree
- Re-attach subtree on all branches in other half of tree
- Use cut-point (root of detached subtree) for re-attachment
- NNI is a subset of SPR

Types of rearrangement III: tree bisection and reconnection (TBR)



- Divide tree into two parts.
- Reconnect subtrees using every possible pair of branches
- NNI and SPR are subsets of TBR