



SI-01: Phylogeny Construction Protocol

From Evidence to Tree — How to Build a Cladogram, Not Just Read One

S8 covered how to read trees. This module covers what AP actually tests under CED 7.9 and Skill 2.D — constructing trees from character data, DNA sequences, and distance matrices.

Unit 7

Skill 2.D

Q5 prep

Construction-focused

Sprint Review · Free at Sophriva

AP BIO STRATEGY · PHYLOGENY · SOPHRIVA.COM



Reading a tree (covered in S8.4)

You're given a finished tree. You identify nodes, MRCAs, sister taxa, and synapomorphies.



Constructing a tree (this module)

You're given raw data — a character matrix, DNA sequences, or a distance matrix. You build the tree from the data and justify your branching decisions.



Why it matters in 2025–2026:

- CED 7.9 explicitly states traits gained/lost can be used to construct phylogenetic trees and cladograms.
- Trees can be built from morphological similarities, DNA sequence similarities, or amino acid sequence similarities.
- Molecular data are generally more accurate and reliable than morphology alone — this is testable.
- Q5 can ask students to place a new taxon, build a tree from a character matrix or DNA sequence, or modify an existing tree based on new data.

2. Three Universal Construction Rules

1



Rule 1 — Outgroup first.

The outgroup is the taxon most distantly related — fewest derived traits, or explicitly stated. The outgroup defines the root and tells you which traits are ancestral.

2



Rule 2 — Use shared derived traits, not shared ancestral traits.

A trait inherited from a deep ancestor (symplesiomorphy) does NOT define a clade. Only traits that are new to a lineage (synapomorphies) define groupings.

3



Rule 3 — Place each derived trait at the branch where it FIRST appears.

If three taxa share a trait, the trait evolved on the branch leading to their






Get your AP Bio FRQ marked free at sophriva.com



3. Method A: Character Matrix -> Cladogram

5-Step protocol + worked vertebrate example

The 5-Step Protocol

- 1**  **Step 1 — Identify the outgroup.**
The outgroup has the fewest derived traits (most '0' entries) or is given.
- 2**  **Step 2 — Count derived traits per taxon.**
Rank taxa from fewest -> most derived traits.
- 3**  **Step 3 — Find shared derived traits.**
Group taxa that share the same derived traits.
- 4**  **Step 4 — Build branches outward from the outgroup.**
Each new derived trait creates a new branch point (node).
- 5**  **Step 5 — Mark each trait at its first-appearance branch.**
Each trait should appear once on the tree, at the earliest branch where it occurs.

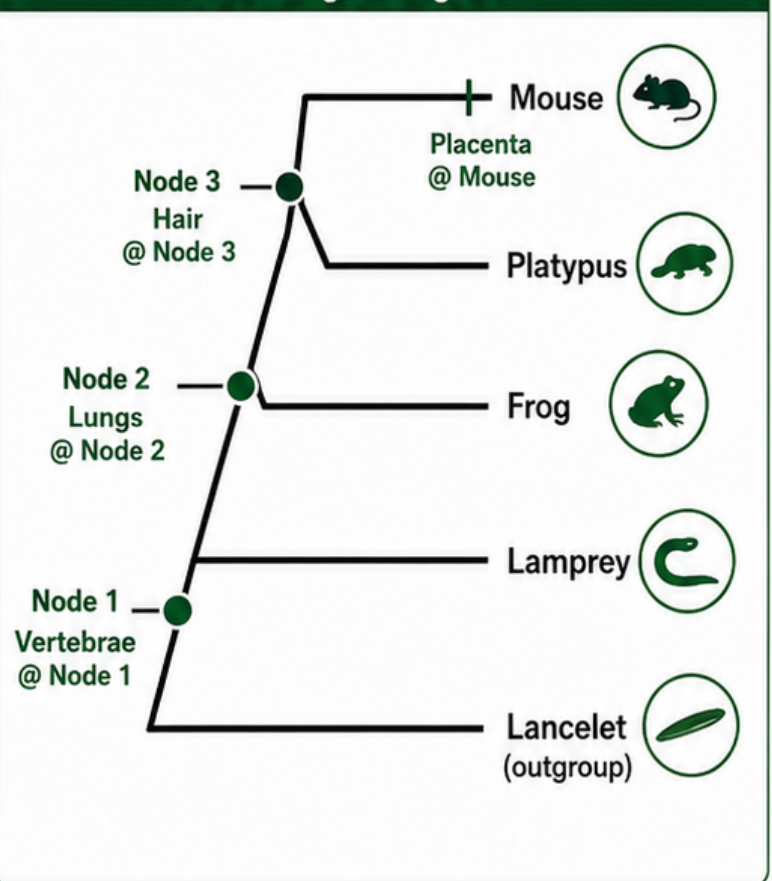
Worked Example — Vertebrate Cladogram

Character Matrix (1 = present, 0 = absent)

Taxon	Vertebrae	Lungs	Hair	Placenta
Lancelet (outgroup)	0	0	0	0
Lamprey	1	0	0	0
Frog	1	1	0	0
Platypus	1	1	1	0
Mouse	1	1	1	1

- 1. Outgroup:** Lancelet (zero derived traits — used to root the tree and define which states are ancestral).
- 2. Count derived traits:** Lancelet 0, Lamprey 1, Frog 2, Platypus 3, Mouse 4. Order: Lancelet -> Lamprey -> Frog -> Platypus -> Mouse.
- 3. Shared derived traits:**
 - **Vertebrae:** shared by Lamprey, Frog, Platypus, Mouse -> at Node 1.
 - **Lungs:** shared by Frog, Platypus, Mouse -> at Node 2.
 - **Hair:** shared by Platypus, Mouse -> at Node 3.
 - **Placenta:** Mouse only -> on Mouse's terminal branch.
- 4. Build:** (((Mouse, Platypus), Frog), Lamprey), Lancelet)
- 5. Mark traits at first appearance.**

Resulting Cladogram



Core idea: use the outgroup, order taxa by derived traits, group shared derived traits, and place each trait once at its first appearance.



4. Method B: DNA / Protein Sequence -> Tree

Compare sequences, find sister taxa, build outward

1 The 4-Step Protocol

1



Step 1 — Calculate pairwise differences.

Compare each pair of sequences and count positions where they differ.

2



Step 2 — Find the smallest pairwise difference.

That pair is the most closely related — sister taxa.

3



Step 3 — Build outward.

Group sister taxa, then add the next-closest taxon, and so on.

4



Step 4 — Apply the molecular-over-morphology rule.

If molecular and morphological data conflict, molecular data are generally considered more reliable because morphology can converge under similar selection pressures (analogous structures).

2 Worked Example — 4 Species, 8-Nucleotide Region

Sequences

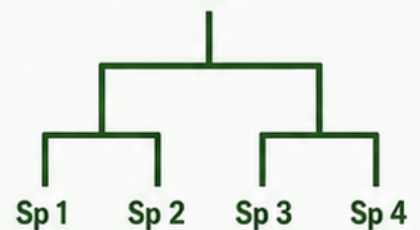
Sp 1	A	C	T	G	G	C	A	T
Sp 2	A	C	T	G	G	C	A	C
Sp 3	T	C	T	G	C	C	A	T
Sp 4	T	C	G	G	C	C	A	T

Pairwise differences (count of mismatches)

	Sp1	Sp2	Sp3	Sp4
Sp1	—	1	2	3
Sp2	1	—	3	4
Sp3	2	3	—	1
Sp4	3	4	1	—

Final tree

((Sp 1, Sp 2), (Sp 3, Sp 4))



- Smallest differences: Sp 1–Sp 2 = 1, and Sp 3–Sp 4 = 1. Two sister-pair candidates.
- Sp 1 and Sp 2 are sister taxa. Sp 3 and Sp 4 are sister taxa.
- Distance between groups: (Sp 1, Sp 2) to (Sp 3, Sp 4) averages ~3 — much larger than within-group distances, so the two pairs sit on separate branches with a deeper common ancestor.
- Final tree: ((Sp 1, Sp 2), (Sp 3, Sp 4))



Molecular vs. Morphological Conflict — How AP Tests It



Scenario

Species A and B look similar but their DNA differs at 50 sites. Species B and C look very different but their DNA differs at only 3 sites. Which two are most closely related?



Correct answer

B and C — fewer molecular differences = more recent common ancestor. Morphological similarity between A and B is likely convergent evolution (analogous structures), not common ancestry.



Justification language for FRQ

Molecular data provide more reliable evidence of evolutionary relatedness because morphological traits can evolve independently in unrelated lineages under similar selection pressures.



Rule to remember: fewer molecular differences usually mean a more recent common ancestor.



5. Method C: Distance Matrix -> Tree

Find the smallest distance, cluster, recalculate, repeat

1 The 4-Step Protocol

1 Step 1 — Find the smallest distance in the matrix.

Those two taxa are sister taxa.

2 Step 2 — Group them as a single cluster.

3 Step 3 — Recalculate distances from the new cluster to remaining taxa.

Use the average of the two distances.

4 Step 4 — Repeat until all taxa are placed.

★ AP does NOT require formal UPGMA notation — but you do need to show that smaller distance = more recent common ancestor.

2 Worked Example — 4 Taxa Distance Matrix

Initial matrix				
	A	B	C	D
A	—	2	8	9
B	2	—	7	8
C	8	7	—	3
D	9	8	3	—

- Smallest = 2 -> A and B are sister taxa. Group as (AB).
- Next smallest among remaining = 3 -> C and D are sister taxa. Group as (CD).
- Recompute (AB)-(CD): average of A-C, A-D, B-C, B-D = $(8 + 9 + 7 + 8) / 4 = 8.0$. Large distance -> deep split.
- Final tree: ((A, B), (C, D))

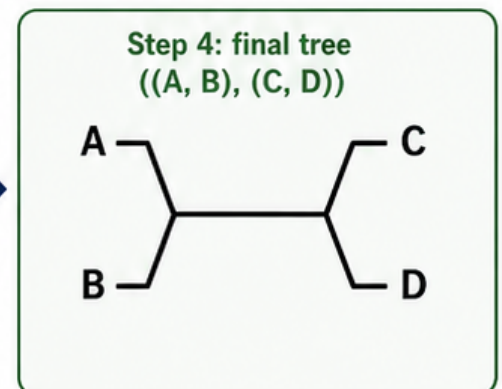
Step 1-2: cluster pairs (AB) and (CD)

	A	B	C	D
A	—	2	8	9
B	2	—	7	8
C	8	7	—	3
D	9	8	3	—

Step 3: recompute (AB) vs (CD)

	(AB)	(CD)
(AB)	—	8.0
(CD)	8.0	—

Average of A-C, A-D, B-C, B-D = 8.0



This page shows the same tree shape as the Method B example, illustrating that different data types can produce identical topology.

Key idea: the smaller the distance, the more recent the common ancestor.

6. Construction-Specific Traps



1. Trap 1 — Drawing the tree by table row order.

Tables in AP problems are often arranged alphabetically or randomly. The order of rows tells you nothing about evolutionary relationships.



2. Trap 2 — Using shared ancestral traits to define a clade.

“All four taxa have vertebrae, so vertebrae unites them as a clade.”
Wrong logic — vertebrae is the ancestral state for this group. Only new traits (synapomorphies) define clades.



3. Trap 3 — Letting morphology override molecular data.

“Bats and birds both have wings — they must be closely related.”
No — winged flight is convergent. DNA tells the real story.



4. Trap 4 — Drawing an ancestor as a living species.

Internal nodes are common ancestors that no longer exist. The taxa at the tips are modern descendants. Don't put 'Lamprey' or 'Frog' at a node.



5. Trap 5 — Using the words 'more evolved' or 'more primitive.'

All living tips are equally evolved — they've each had the same amount of time to evolve since the most recent common ancestor. Use derived and ancestral instead.

7. Practice Questions



MCQ Practice

A character matrix shows that taxa W, X, Y, Z share trait 1. Only X, Y, Z share trait 2. Only Y, Z share trait 3. Which tree topology is best supported?

- A (W, (X, (Y, Z)))
- B ((W, X), (Y, Z))
- C (((W, X), Y), Z)
- D ((W, Y), (X, Z))



Answer: A.

Trait 1 (shared by all four) is at the root. Trait 2 (X, Y, Z only) excludes W → next node groups (X, (Y, Z)). Trait 3 (Y, Z only) makes Y and Z sister taxa.



Q5-Style FRQ — Data Stem

A biologist sequences a 100-nucleotide region of a gene in four salamander species and reports the following pairwise differences:

	Sp1	Sp2	Sp3	Sp4
Sp1	—	4	18	17
Sp2	4	—	16	15
Sp3	18	16	—	5
Sp4	17	15	5	—

- (a) Construct a phylogenetic tree consistent with these data. [2 pts]
- (b) A second researcher claims that Sp1 and Sp3 are most closely related because they share a similar coloration pattern. Evaluate this claim. [2 pts]



Next page: model answer + final cheat card.

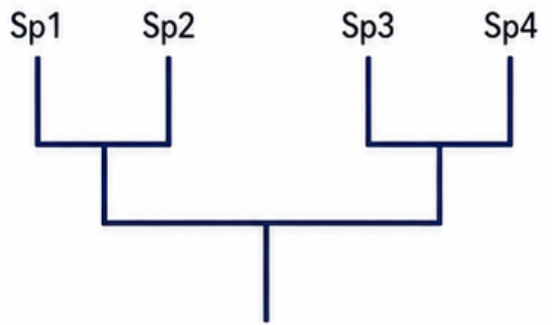
7. Practice Questions — Model Answer



Q5-Style FRQ — Model Answer

a (a) Construct a phylogenetic tree consistent with these data. [2 pts]

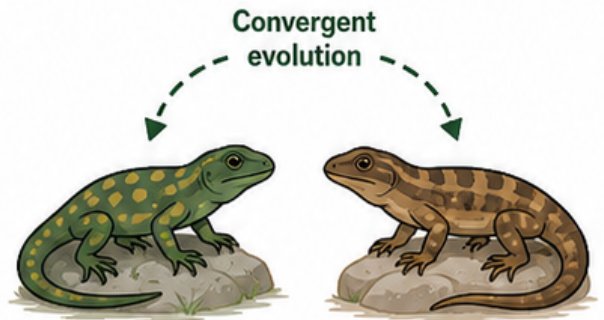
Smallest pairwise differences are Sp1–Sp2 (4) and Sp3–Sp4 (5). These are the two sister pairs. The between-group distances (~16) are much larger, indicating a deeper split.
Tree: ((Sp1, Sp2), (Sp3, Sp4)).



[1 pt: identifies correct sister pairs from molecular data;
1 pt: tree topology drawn or written correctly with two clades joining at a deeper node]

b (b) Evaluate the claim that Sp1 and Sp3 are most closely related because they share a similar coloration pattern. [2 pts]

The claim is not supported by the molecular data. Sp1 and Sp3 differ at 18 nucleotide positions, while Sp1–Sp2 differ at only 4. The molecular evidence strongly supports Sp1 and Sp2 as the closer pair. The shared coloration in Sp1 and Sp3 is most likely convergent evolution under similar selection pressures (e.g., camouflage in similar habitats), not a sign of common ancestry.



[1 pt: rejects the claim using molecular data; 1 pt: explains coloration as convergent evolution / molecular data more reliable than morphology]

8. One-Page Cheat Card

	Character Matrix	DNA Sequence	Distance Matrix
Step 1	Find outgroup	Calculate pairwise differences	Find smallest distance
Step 2	Rank by # of derived traits	Smallest difference = sister taxa	Group as sister taxa
Step 3	Group shared derived traits	Build outward from closest pair	Recalculate distances (average)
Step 4	Place traits at first-appearance branch	Apply molecular > morphology rule	Repeat until tree complete
Common trait		All trees built outward from the outgroup ; smaller difference / fewer changes = more recent common ancestor	



Universal rule (in AP-style simplified datasets):

Fewer sequence differences or smaller distances usually support a more recent common ancestor and a sister-taxon relationship.



End of SI-01 content — review complete.