

CONSTRUCTING A PHYLOGENETIC TREE FROM DNA SEQUENCE DATA

- Consider the following aligned DNA sequences from the **env** gene of the HIV virus from six taxa:

	1	2	3	4	5	6	7	8
	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890
dnt	CATTAGTAGA	GAAAAATGGA	ATAATACTTT	AAACCAGGTA	GTTACAGAAT	TAAGGGAACA	ATTTGGGAAT	AAAACAATAA
ptb	CATTAGTGGA	GCAAAATGGA	ATAATACTAT	AGAACAGGTA	AAGACAAAAT	TAAGAGAACA	ATTTGGGAAT	AAAACAATAA
ptc	CATTAGTAGA	GAAAAATGGA	ATAATACTTT	AAAACAGGTA	GTTACAAAAT	TAAGAGAACA	ATTTGTGAAT	AAAACAATAA
ptd	CCTTAGTGAA	GCAATATGGA	ATAACACGTT	AAAACAGATA	GTTAAAAAAT	TAAAAGAACA	ATTTAAGAAT	AAAACAATAG
lc1	CCTTAGTAAA	GCAGAAATGGA	ATAACACTTT	AAGACAGATA	GTTAAAAAAGT	TAAGAGAACA	ATTTAAGAAT	AAAACAATAG
lc5	CATTAGTAAA	GCAGAGTGGA	ATAACACTTT	AAAACAGATA	GTTCAAAAAT	TAAAAGAACA	ATTTAGGAAT	AAAACAATAG

- Give the phylogenetic tree that you think best relates these orthologous sequences. Explain how you arrived at this tree, and why your approach is reasonable.
- If you can, assign appropriate lengths to the edges of your tree.
- If you can, choose a root location for your tree.
- What other alternative methods might you use?

- Four taxa  $A$ ,  $B$ ,  $C$ , and  $D$  are related according to a certain metric tree. The total distances between taxa along the tree have been found to be as in the table below.

	$A$	$B$	$C$	$D$
$A$		.6	.6	.2
$B$			.4	.6
$C$				.6

- Using any approach you wish, determine the correct unrooted tree relating the taxa, as well as all edge lengths. Explain how you rule out other topological trees.
- Can you determine the root from this data? Explain why or why not.
- Repeat parts (a) and (b) for the table below.

	$A$	$B$	$C$	$D$
$A$		.55	.40	.40
$B$			.35	.25
$C$				.20