

Evolutionary Tree Construction

1. True or False: according to its definition, a distance matrix must satisfy the triangle inequality.

True

False

2. True or False: according to its definition, a distance matrix must have all positive values.

True

False

3. True or False: most distance matrices constructed from biological data are non-additive.

True

False

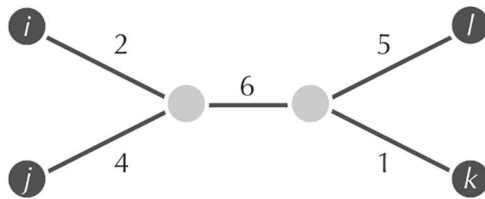
4. How many edges must a tree with 2349 nodes have?

2348 [edges = nodes - 1]

5. How many nodes must a tree with 1167 edges have?

1168 [nodes = edges + 1]

6. Which of the following matrices is fit by the tree shown below?



ijkl	ijkl	ijkl	ijkl
i 0 6 10 13	i 0 6 9 14	i 0 6 10 14	i 0 6 9 13
j 6 0 12 15	j 6 0 11 16	j 6 0 12 16	j 6 0 11 15
k 10 12 0 7	k 9 11 0 7	k 10 12 0 6	k 9 11 0 6
l 13 15 7 0	l 14 16 7 0	l 14 16 6 0	l 13 15 6 0

7. Compute $LimbLength(i)$ for the additive distance matrix shown below.

```

ijkl
i 0 13 16 10
j 13 0 21 15
k 16 21 0 18
l 10 15 18 0
4
  
```

8. Compute $LimbLength(k)$ for the additive distance matrix shown below.

```

ijkl
i 0 14 17 17
j 14 0 7 13
k 17 7 0 16
l 17 13 16 0
5
  
```

Constructing Tree from Distance Matrices

1. True or False: UPGMA does not necessarily identify neighboring leaves in the tree fitting an additive distance matrix D .

True

False

2. Compute the sum of squared errors $Discrepancy(T, D)$ for the tree T and distance matrix D given below.

i j k l

i 0 20 9 11

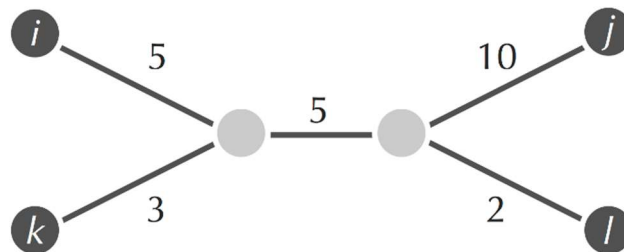
j 20 0 17 11

k 9 17 0 8

l 11 11 8 0

8

$D = \begin{bmatrix} 0 & 20 & 9 & 11 \\ 20 & 0 & 17 & 11 \\ 9 & 17 & 0 & 8 \\ 11 & 11 & 8 & 0 \end{bmatrix}$



$T = \begin{bmatrix} 0 & 20 & 8 & 12 \\ 20 & 0 & 18 & 12 \\ 8 & 18 & 0 & 10 \\ 12 & 12 & 10 & 0 \end{bmatrix}$

3. Below is a distance matrix D . If C_1 is the cluster containing i and j , and C_2 is the cluster containing k and l , compute $D(C_1, C_2)$.

i j k l

i 0 20 9 11

j 20 0 17 11

k 9 17 0 8

l 11 11 8 0

12

4. Below is a distance matrix D . Compute $D_{k,l}^*$, where D^* is the neighbor-joining matrix of D .

i j k l

i 0 13 16 10

j 13 0 21 15

k 16 21 0 18

l 10 15 18 0

-62

5. Below is a distance matrix D . After the neighbor-joining algorithm decides that j and l are neighbors, compute $LimbLength(l)$.

i j k l

i 0 20 9 11

j 20 0 17 11

k 9 17 0 8

l 11 11 8 0

1

Constructing Trees from Characters

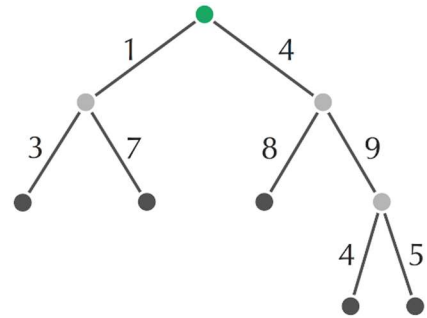
1. True or False: grouping all the species sharing the same character on one side of an evolutionary tree is not a reasonable way to construct the tree.

True

False

2. Compute the parsimony score of the following tree.

41 [1+3+7+4+8+9+4+5]



3. Below are the fifteen unrooted binary trees with five labeled leaves. Which trees are nearest neighbors of tree #1? (Select all that apply.)

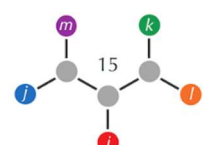
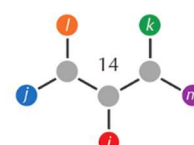
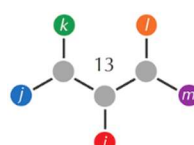
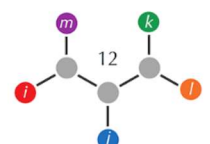
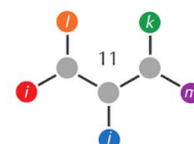
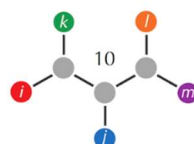
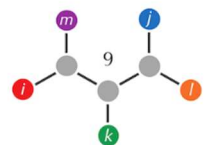
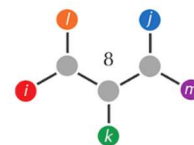
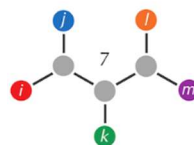
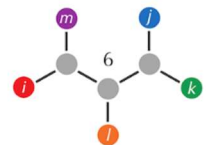
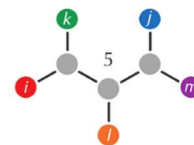
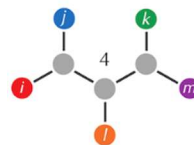
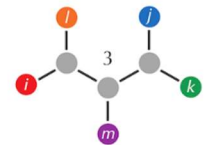
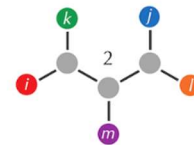
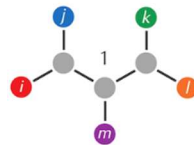
15 10 14 5 3 4 7 12

Swap j and k $\rightarrow ((i, k), (j, l), m)$ [Tree 12]

Swap j and l $\rightarrow ((i, l), (k, j), m)$ [Tree 15]

Swap m and i $\rightarrow ((m, j), (k, l), i)$ [Tree 4]

Swap m and j $\rightarrow ((i, m), (k, l), j)$ [Tree 7]



Peptide Sequencing

1. True or False: Whereas DNA sequencing has become much easier in the past few decades, protein sequencing remains difficult.

True

False

2. True or False: When biologists analyze protein samples, each peptide in the sample may break in its own way.

True

False

3. True or False: In practice, we will be able to infer a priori which peptide fragments correspond to prefixes and which fragments correspond to suffixes.

True

False

4. True or False: A mass spectrometer is able of measuring the mass-to-charge ratio of only short peptides.

True

False

5. True or False: A mass spectrometer is able of measuring the mass-to-charge ratio of very long peptides.

True

False

6. True or False: A weakness of summing the intensities of explained peaks for scoring peptides is that large peaks may dominate the score.

True

False

7. True or False: A weakness of summing the intensities of explained peaks for scoring peptides is that it ignores spectral intensities.

True

False

8. True or False: A weakness of using the shared peak count for scoring peptides is that it ignores spectral intensities.

True

False

9. Fill in the blank: we score a peptide vector against a spectral vector as the _____ of the two vectors.

dot product

10. The biologically correct peptide rarely scores highest among all peptides appearing in the proteome.

True

False

11. True or False: The highest scoring peptide against a spectrum (over all possible peptides) is rarely the biologically correct peptide for that spectrum.

True

False

Peptide Identification and Spectral Alignment

1. Fill in the blank: 90% of proteins making up animal bones are _____

collagen

2. Say that the probability of a collection of peptides *Dictionary* is equal to 0.00012. What is the expected approximate number of peptide matches from *Dictionary* that we expect to find in a decoy proteome of length 500,000?

60

3. True or False: Hemoglobins are more evolutionary conserved than collagens, implying that it is not very surprising that the hemoglobin peptide in the dinosaur sample indeed came from *T. rex*.

True

False

4. True or False: Contamination is uncommon in proteomics laboratories.

True

False

5. True or False: Most proteins are modified after translation.

True

False

6. True or False: Billions of peptides not occurring in Uniprot+ outscore *DinosaurPeptide* with respect to *DinosaurSpectrum*.

True

False

7. True or False: A modification of positive mass k to a protein corresponds to deleting a block of k zeroes from the corresponding peptide vector.

True

False