

Finding a Longest Path

1. There is a unique longest common subsequence of the strings CTCGAT and TACGTC.
What is it?

TCGT

2. True or False: There are some coin denominations for which **GreedyChange** does not solve the Change Problem (i.e., does not find a minimum number of coins making change).

True

False

3. Imagine a hypothetical world in which there are two amino acids, X and Z, having respective masses 2 and 3. How many linear peptides can be formed from these amino acids having mass equal to 25? (Remember that the order of amino acids matters.)

465

4. True or False: Not all recursive algorithms are inefficient.

True

False

5. Consider the following adjacency list of a DAG:

a -> b: 5 a -> c: 6 a -> d: 5

b -> c: 2 b -> f: 4

c -> e: 4 c -> f: 3 c -> g: 5

d -> e: 6 d -> f: 8

e -> g: 2

f -> g: 1

What is the longest path in this graph? Give your answer as a sequence of nodes separated by spaces. (Note: a, b, c, d, e, f, g is a topological order for this graph.)

a d f g

6. Here is the adjacency list of a graph with six nodes and ten edges:

a -> b, c, d, e, f b -> c, f c -> d d -> e -> d, f f ->

Which of the following are topological orderings of the nodes in this graph? (Select all that apply.)

a, b, c, e, d, f

a, e, b, c, f, d

a, b, e, c, f, d

a, b, e, d, c, f

a, b, e, f, d, c

a, b, c, d, e, f

7. True or False: The dynamic programming algorithm we introduced for finding a longest path in a DAG has runtime proportional to the number of edges in the graph.

True

False

Aligning DNA Strings

1. Consider the following alignment:

TCGAC--ATT

CC---GAA-T

What is the score of this alignment if the match score is 1, the mismatch penalty is 1, and the indel penalty is 2?

-10

2. Consider the following alignment:

TACTATTT**ACAGTAGACAC**GT

AACAGAC-**ATAC-AGATAC**CT

What is the score of the bold portion of this alignment as a *local* alignment if the match score is 1, the mismatch penalty is 3, and the indel penalty is 1?

-3

3. Say that the match score is 1, the mismatch penalty is 1, and the indel penalty is 1. What is the maximum score of a fitting alignment of GATACACT and ACGACCACAGATACCGCTATTCACTATATCGTT? (Hint: don't make this problem too difficult.)

5

4. Say that the match score is 1, the mismatch penalty is 0, and the indel penalty is 2.

Score the following overlap alignment.

AGTACATCAGAGGAGTT-ACATACTAACG

AGTTCACAGGCTA-CGTACAGATATTACGACAGGCAGA

8

5. What type of alignment should you use if you are interested in aligning short regions of two strings?

Global alignment

Local alignment

Fitting alignment

Overlap alignment

Sequence Alignment Advanced Topics

1. Consider the following alignment:

TCGAC--ATT

CC---GAA-T

What is the score of this alignment if the match score is 1, the mismatch penalty is 1, and the indel penalty is affine with a gap opening penalty of 4 and a gap extension penalty of 1?

-13

2. True or False: Using a three-level Manhattan is quicker than using a Manhattan with long indel edges to solve the Alignment with Affine Gap Penalties Problem.

True

False

3. True or False: Different longest paths in an alignment graph may have different middle nodes.

True

False

4. There is a unique multiple longest common subsequence of CGGAACTGGT, TGAGACGGTA, and TGCGACGGCT. What is it?

GGACGGT

5. True or False: Optimal pairwise alignments of three strings always can be assembled into an optimal 3-way alignment.

True

False

Genome Rearrangements and Fragility

1. True or False: genome rearrangements can be thought of as genomic earthquakes.

True

False

[Mutations are genomic erosion.]

2. Compute the number of steps required by **GreedySorting** to sort the following permutation (i.e., to transform it into the identity permutation.)

(-16 -20 +12 +18 -14 -17 -15 -6 -8 -19 -11 +13 -10 +4 -5 -2 +7 -3 +1 -9)

29

3. Compute the number of breakpoints in the following permutation.

(+6 -12 -9 +17 +18 -4 +5 -3 +11 +19 +20 +10 +8 +15 -14 -13 +2 +7 -16 -1)

18

4. What is the maximum number of breakpoints that a single reversal can eliminate?

2

Genome Rearrangements and Fragility Part 2

1. Given the genome containing the circular chromosomes $(+1 +2 +3 +4)(+5 +6)(+7+8+9)$, which of the following is a genome resulting from a 2-break on this genome? (Select all that apply.)

$(+1 +2 +3 +4)(+5 -9 -8 -7 +6)$

$(+1 +2 +3 +4 +5 +6 +7 +8 +9)$

$(+1 +2 +3 +4)(-5 +6) (+7 +8 +9)$

$(+1 +2)(+3 +4)(+5 +6)(+7 +8)(+9)$

2. What is the maximum number of cycles that the breakpoint graph of two genomes with 97 syntenic blocks can have?

97

3. Say that genomes P and Q share 200 syntenic blocks. P has 15 circular chromosomes, Q has 37 circular chromosomes, and the 2-break distance between P and Q is 57. How many red-blue alternating cycles are there in the breakpoint graph of P and Q?

143 $[d(P,Q) = n - c \rightarrow c = n - d(P,Q) = 200 - 57]$

4. Give the number of shared 3-mers of TGGCCTGCACGGTAG and GGACCTACAAATGGC. Please make sure that you consult the definition of shared k -mers from the epilogue in the interactive text (include reverse complements).

7

5. True or False: Evidence suggests that there are rearrangement hotspots in the human genome.

True

False

6. True or False: It is possible to prove a scientific theory in the same way that one would prove a mathematical theorem.

True

False