

1. The slow (sometimes called “brute force”) algorithm for finding the shortest common superstring of the strings in set S involves:

Trying all orderings of the strings in S

Concatenating the strings in S

Iteratively removing strings from S that don't belong in the superstring

Finding the longest common substring of the strings in S

2. Which of the following is **not** a true statement about the slow (brute force) shortest common superstring algorithm.

The superstring returned might be longer than the shortest possible one

The amount of time it takes grows with the factorial of the number of input strings

It might collapse repetitive portions of the genome

3. Which of the following is **not** a true statement about the greedy shortest common superstring formulation of the assembly problem?

It might collapse repetitive portions of the genome

The superstring returned might be longer than the shortest possible one

The amount of time it takes grows with the factorial of the number of input strings

4. True or false: an Eulerian walk is a way of moving through a graph such that each node is visited exactly once

False (each edge is Eulerian, each node is Hamiltonian)

True

5. If the genome is repetitive and we try to use the De Bruijn Graph/Eulerian Path method for assembling it, we might find that:

There is more than one Eulerian path

The genome “spelled out” along the Eulerian path is not a superstring of the reads

The De Bruijn graph breaks into pieces

6. In a De Bruijn assembly graph for given k , there is one edge per

$k-1$ -mer

k -mer

read

genome

7. Which of the following does not help with the problem of assembling repetitive genomes:

Paired-end reads

Increasing minimum required overlap length for the overlap graph

Longer reads