

Name _____

BIO/CMPSC 300

Chapter 8 Overlap in Sequence Assembly

Activity 3

Fall 2019

1. Suppose two sequence reads give GGCAGCC and GCCCCAGG.

a) What are the possible overlaps? Which one is the largest?

b) Describe an algorithm that finds the largest overlap. Remember that the algorithm is a step-by-step procedure, thus your algorithm should outline the sequence of steps required for finding the largest overlap given two sequences.

c) What would be the sequence of the contig produced using your algorithm?

2. Now suppose you would like your algorithm to account for the possibility that the sequences could come from either strand of the DNA. How would you modify the algorithm to accomplish this? Would the contig resulting from the two sequences in 1 change as a result?