

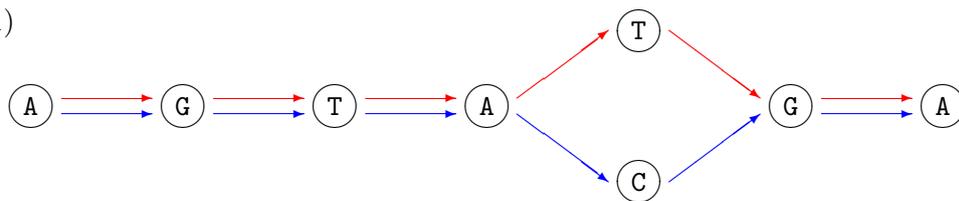
Algorithms in Genome Research
Winter 2020/2021

Exercises

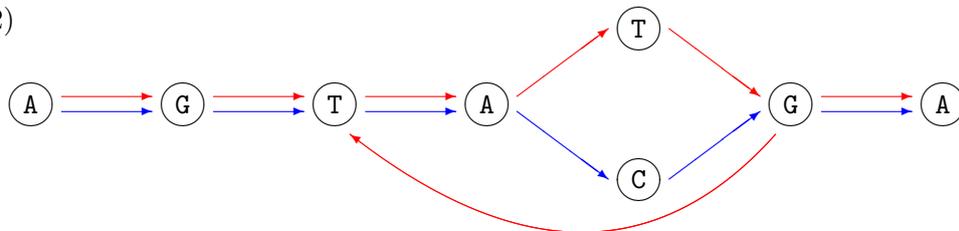
Number 9, Discussion: 2021 January 22

1. What is an *open* pangenome and what is a *closed* pangenome?
Schematically, how do an open and a closed pangenome look like
 - as a Venn diagram in gene-based pangenomics?
 - as a pangenome graph in genome-based pangenomics?
2. Two popular data structures to represent a genome-based pangenome are the variation graph and the colored de Bruijn graph.
 - (a) Given the following two variation graphs, find colored de Bruijn graphs of dimension $k = 3$ that contain the same sets of strings.

(a.1)



(a.2)



- (b) Given the following three “genome” sequences. Construct their compacted colored de Bruijn graph of dimension $k = 4$.

CAGGATCAGAACGGC
GGACCCAGGATAGA
AGGACCCATAGAACGGC

Find a variation graph that represents the same set of strings.

3. Develop the details of an algorithm that takes as input a variation graph G and a query sequence S , and finds a position in G where an optimal (unit-cost) semi-global alignment of S and any (sub)string represented in G ends.

Note: First consider that G is a directed acyclic graph (DAG). Then generalize your algorithm to the case where G may contain cycles.