

Algorithms for Comparative Genomics
Summer 2018

Exercises

Number 1, return 2018 May 11

1. Compute the breakpoint distance between all pairs of the following four circular genomes:

$$P = (1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12)$$

$$Q = (-1, -2, -3, -4, -5, -6, -7, -8, -9, -10, -11, -12)$$

$$R = (4, -3, 5, -6, 7, 1, 2, 9, -8, -10, 11, 12)$$

$$S = (3, 2, 1, 12, -11, -10, -8, -7, 9, -6, 5, 4)$$

How would the distances change if the chromosomes were linear?

2. Find matchings for the minimum (exemplar), intermediate and maximum model for the following two linear genomes with gene families

$$P = [1, 2, 3, 1, 4, 5, 5, 6, 7, 7, 7, 8, 4, 5, 9]$$

$$Q = [2, 1, 4, -5, 7, 7, 8, 5, 6, 7, 8, 10, 3, 1]$$

(a) such that the breakpoint distance $d_{bp} = n - a - t/2$ is minimized; or (b) such that the common adjacency similarity measure $s_{adj} = a$ is maximized.

3. Show that the breakpoint distance is a lower bound for two times the reversal distance,

$$d_{bp}(P, Q) \leq 2 \cdot d_{rev}(P, Q) \text{ for any two genomes } P \text{ and } Q.$$

For small genomes and large evolutionary times the underestimation might be considerable. Be prepared to discuss possibilities how one could correct for this in class.