Algorithms in Comparative Genomics

Universität Bielefeld, WS 2023/24
Prof. Dr. Jens Stoye · Dr. Marília D. V. Braga
https://gi.cebitec.uni-bielefeld.de/teaching/2023winter/alggr

Exercise sheet 11, 26.01.2024

Exercise 1 (Calculating Distances)
Given two linear, unichromosomal genomes

$$A = [\,1\;2\;-4\;-3\;5\;-6\,]$$

and

$$B = [\,1\;-3\;-2\;4\;5\;6\,]$$

1. What is the breakpoint distance between A and B?
2. What is the SCJ distance between A and B?
3. Give an SCJ sorting scenario from A to B.
4. Draw the adjacency graph of A and B.
5. What is the DCJ distance between A and B?
6. Give a DCJ sorting scenario from A to B.

Exercise 2 (Cooking Recipe) The *explode chromosome operation (ECO)*[1] transforms a chromosome $[g_1...g_k]$ or $(g_1...g_k)$ of arbitrary length k into k linear chromosomes $[g1], ..., [g_k]$ with one marker g_i each. The *assemble chromosome operation (ACO)*[2] transforms a set of linear chromosomes $\{[C_1], ..., [C_l]\}$ into a linear chromosome $[S]$ or circular chromosome (S) with $S = C_1 C_2 ... C_l$ for any orientation and order of C_1 to C_l.
For example: $[1\;2\;3], [4\;5] \stackrel{ECO}{\rightarrow} [1], [2], [3], [45] \equiv [1], [-5\;-4], [3], [-2] \stackrel{ACO}{\rightarrow} [1\;-5\;-4\;3\;-2]$.

1. Familiarize yourself with the operations. Give optimal sorting scenarios for the following:
   (a) Sort $[1\;2\;3], (4\;5)$ into $[1\;3\;2], (4\;5)$.
   (b) Sort $[1\;2\;3]$ into $[1], [2], [3]$.
   (c) Sort $[1], (2\;3\;4), [5]$ into $[1\;2\;3\;4], [5]$.
2. Use the "cooking recipe" from the lecture to derive a general formula for the minimum number of ECOs and ACOs needed to transform one genome into the other. You do not need to invent a data structure for this. The "quantity" you need for the cooking recipe can be derived directly from the genomes themselves.

Exercise 3 (Feather Median)

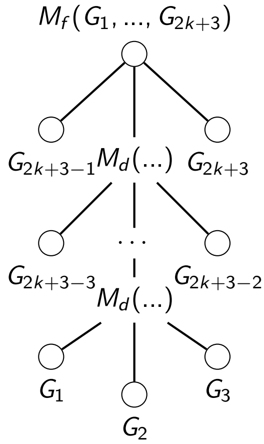
Given $2k+3$ genomes $G_1, G_2, G_3, ...G_{2k+3}$ and an algorithm to compute the median $M_d(A, B, C)$ of three genomes (A, B, C) under a distance model d. The *Feather Median*[3] is defined as

$$M_f(G_1, G_2, G_3, ...G_{2k+3-2}, G_{2k+3-1}, G_{2k+3}) = M_d(G_{2k+3-1}, M_f(G_1, G_2, G_3, ...G_{2k+3-2}), G_{2k+3}) \quad (1)$$

with recursion base

$$M_f(G_1, G_2, G_3) = M_d(G_1, G_2, G_3) \quad (2)$$

[1] which I made up.
[2] which I also made up :)
[3] which I also made up ;)



1. Disprove (e.g. via counter example): The Feather Median under the breakpoint distance is always a breakpoint median[4].
2. Prove or disprove: No metric d on a set with two or more distinct elements exists, under which the Feather Median is always a true Median. (Spoiler 1, 2, 3, 4, 5, 6).

[4] The true median of a set $K \subseteq S$ under metric d on space S being the element $M_d \in S$ that minimizes $m(M_d) = \sum_{k \in K} d(M_d, k)$.