

Theoretical advances on multichromosomal median computation

Ahmad Mahmoody Ghaidary¹, Cedric Chauve¹, and Ladislav Stacho¹

Department of Mathematics, Simon Fraser University

In the present work, we consider the classical problem of computing an optimal median for three multichromosomal genomes G_1 , G_2 , G_3 . More precisely, we consider the case where

- G_1 , G_2 , G_3 are mixed genomes as they can have (possibly) both linear and circular chromosomes;
- pairs G, G' of genomes on n genes are compared using the following dissimilarity measure: $d_C(G, G') = n - c(BG(G, G'))$ where $BG(G, G')$ is the breakpoint graph of G and G' and $c(BG(G, G'))$ the number of cycles of this graph;
- mixed median genomes (with both linear and circular chromosomes) are allowed.

Although d_C differs from the classical DCJ distance (and is not a distance for genomes that contain linear chromosomes), it is a widely used approximation of the DCJ distance. Our main result is the following.

Theorem 1. *If the breakpoint graph of three mixed genomes G_1 , G_2 and G_3 has maximum degree 2, then an optimal mixed median for the dissimilarity measure d_C can be computed in polynomial time.*

The proof is based on the fact that every breakpoint graph with maximum degree two is consist of paths and cycles. We show that there exist an optimal median such that: (1) vertices of cycles and paths with even number of edges, are matched between themselves, and (2) the set of all odd paths and odd cycles can be matched, such that no edge in the median connect two vertices in two different parts of the matching. From there, we prove that is is tractable to find optimal adjacencies within even paths and cycles, and that dealing with odd paths and cycles reduces to solving a maximum weight matching problem.