

Constructing Intermediate Breakpoint Medians of Random Genomes

Poly H. da Silva* (Columbia University), Arash Jamshidpey (University of California Berkeley), David Sankoff (University of Ottawa)

In the absence of duplications, unichromosomal genomes can be represented as permutations or their associated labeled path graphs, where vertices correspond to genes or syntenic blocks. The breakpoint distance quantifies genomic dissimilarity by counting differing gene adjacencies. Formally, for two genomes (paths) $G_1 = ([n], E_1)$ and $G_2 = ([n], E_2)$, with the same gene set $[n] = \{1, \dots, n\}$, the breakpoint distance is given by $d(G_1, G_2) = |E_1 \Delta E_2|/2$, where Δ denotes the symmetric difference of their edge sets. A *median genome* for a set of genomes A with the same gene set is a genome M that minimizes the total breakpoint distance: $x \mapsto \sum_{y \in A} d(x, y)$. In this talk, we analyze the structure of median genomes for k random genomes and, in particular, investigate the case $k = 3$, aiming to construct a median genome that is distant from all input genomes.

Keywords: Random Permutations, Breakpoint Medians, Genome Evolution.