Pangenomic FM-Indexes: Alignment and Beyond

Prof. Travis Gagie
Dalhousie University

DNA alignment has been a killer app for the FM-index, but aligning against a single reference sequence can bias research results and medical diagnoses. In the past few years we have found ways to FM-index pangenomes, which are already leading to more robust aligners and may soon result in new tools for comparing genomes to pangenomes. As an example, we will discuss the problem of indexing a pangenome T in very small space such that given a genome P, for each character $P[i]$ in P we can efficiently compute the length of the shortest substring of P that includes $P[i]$ and occurs fewer than k times in T.