

Abstract

Variation graphs, which represent genetic variation within a population, are replacing sequences as reference genomes. Path indexes are one of the most important tools for working with variation graphs. They generalize text indexes to graphs, allowing one to find the paths matching the query string. We propose using de Bruijn graphs as path indexes, compressing them by merging redundant subgraphs, and encoding them with the Burrows-Wheeler transform. The resulting fast, space-efficient, and versatile index is used in the variation graph toolkit `vg`.