Bacterial genome assembly in the presence of structural variation

Type of project: Bachelor End Project, Master End Project, Literature Review

Student background: CS, Nanobiology, LST

Daily supervisor: Jasmijn Baaijens, Pattern Recognition and Bioinformatics (j.a.baaijens@tudelft.nl)

Background
Bacterial genomes are very dynamic: DNA can be inserted and deleted from the genome, allowing bacterial genomes to be dramatically rearranged. Entire regions can be inserted, deleted, duplicated, inverted, or moved to a different position on the genome. These rearrangements are referred to as *structural variation*. Identifying structural variation is important, because these genomic changes can influence characteristics such as whether the bacterium causes disease (and how severe this disease is) and whether the bacterium is susceptible to antibiotics.

Bacterial genome sequencing then enables reading small fragments (*sequencing reads*) of the genome, from which the genome can be reconstructed (*genome assembly*) and structural variation can be identified (*variant calling*). However, genome assembly in the presence of structural variation is very challenging: current algorithms are unable to resolve repetitive sequences that are longer than the sequencing reads, causing incomplete or incorrect genome assemblies.

In this project we will develop an optimization algorithm to make bacterial genome assemblies more accurate and more complete. Then, we can apply this algorithm to several real sequencing datasets and evaluate the results.

Project components
Note that these are only suggestions, the scope and focus of the project will be adapted to suit your expertise and interests, as well as your program requirements.

- Review existing pipelines for bacterial genome assembly and evaluate performance on simulated and real benchmarking data
- Design and implement a new approach to improve genome assemblies