Examining bacterial variation with genome graphs and Nanopore sequencing

Michael Benjamin Hall

Due to their clonality, comparing bacterial genomes allows the inference of relatedness. In the context of bacterial pathogens, this has important implications for global public health, empowering epidemiology and enabling the prediction of antimicrobial drug resistance (AMR).

The vast majority of methods for examining genetic variation rely on a central idea: a single reference genome as a template to find variants. As an introduction to this thesis, we outline why this approach is inadequate for bacteria and describe a new approach using genome graphs. In the first chapter, we present algorithms for *de novo* variant discovery within such genome graphs.

The remaining chapters address a question relating to a critical bacterial pathogen: can Nanopore sequencing of *Mycobacterium tuberculosis* provide high-quality public health information? We collect data from Madagascar, South Africa, and England to help answer this question. First, we assess outbreaks identified using single-reference and genome graph methods. Second, we evaluate AMR predictions and introduce a framework for using genome graphs to improve current methods. Lastly, we train an *M. tuberculosis*-specific Nanopore basecalling model with considerable accuracy improvement.

Together, this thesis provides general methods for uncovering bacterial variation and applies them to an important global public health question.