Tools and algorithms to take us from routine pathogen surveillance data to biological insights

Routine genomic surveillance of pathogens has increased the amount of available pathogen data enormously. We are no longer in a regime where we can afford to analyse all samples as a batch, and we must work in a mode where new genomes are added into existing populations and exploit existing knowledge. One part of my talk will describe a recent method from my group, ggCaller, which aims to address this problem for bacterial annotation. By using genome graphs, ggCaller annotates an entire pangenome in a population in a single step, increasing accuracy and speed, and consistency of gene names between studies. The other part of my talk will focus on creating mathematical models of evolution using this data as input, with the aim of forecasting population changes after vaccination. I will also describe a general modelling framework, odin.dust, we developed for real-time modelling of COVID-19, and how this can be used to more rapidly develop, test and fit these models.