## Modelling and inferring recombination in bacterial genomic evolution

**Professor Xavier Didelot and Dr Rich Everitt**

### Background

Recombination happens frequently in most bacterial species. Traditional phylogenetic techniques do not account for this, which can greatly limit their usefulness for the analysis of genomic data. The coalescent with gene conversion accurately models the ancestry process of bacteria, and this can be used to simulate realistic data, but it is too complex to use in an inferential setting. Approximations have therefore been introduced, which are centred around the concept of the clonal genealogy, that is the phylogeny obtained by following the line of ancestry of the recipient of each recombination event. These mathematical models are the starting point of ongoing efforts to develop statistical methods and software to perform phylogenetic analysis in recombining bacteria, but current methods are either too simplistic or unable to scale to the large amounts of genomic data that are now available.

### Objectives

The aim of this project is to develop, implement and test a new statistical methodology to perform joint statistical inference of a phylogenetic tree and recombination events.

### Proposed research plan

1. Get familiar with phylogenetics and bacterial recombination (no prior knowledge on this is required)
2. Formulate a mathematical model under which to perform inference and establish a strategy to calculate or approximate the likelihood
3. Implement a method to perform inference (probably Markov Chain Monte Carlo, and possibly using approximate Bayesian Computation techniques)
4. Test the method thoroughly using simulated data and real data if time allows

### Starter References

1. Didelot X, Wilson DJ. ClonalFrameML: Efficient Inference of Recombination in Whole Bacterial Genomes. PLoS Comput Biol 2015; 11:e1004041.
2. Didelot X, Lawson DJ, Darling AE, Falush D. Inference of homologous recombination in bacteria using whole-genome sequences. Genetics 2010; 186:1435–49.
3. Collins C, Didelot X. A phylogenetic method to perform genome-wide association studies in microbes that accounts for population structure and recombination. PLoS Comput Biol 2018; 14:e1005958.
4. Didelot X, Croucher NJ, Bentley SD, Harris SR, Wilson DJ. Bayesian inference of ancestral dates on bacterial phylogenetic trees. Nucleic Acids Res 2018; 46:e134.