Population-level analysis of gut microbes prevalence, variation and adaptations in the human and animal gut microbiota

Recently, scientists discovered a link between health & disease and a previously unknown factor. This factor is the human microbiota, the collection of organisms that live in the human body. These organisms are microscopic and are at least as numerous as our human cells. Moreover, these organisms encode 150fold more genes (pieces of DNA carrying genetic information) than the humans that harbour them. While microbiota reside in all parts of our body, most inhabit the gut, where there is an abundance of nutrients. Gut microbiota have many functions in the human body. For example, they help with digestion of fiber, production of some vitamins (K and B) and they also provide protection from intestinal infections and other diseases. Scientists use information extracted from faecal samples (e.g. genomes, genes, metabolites) to study different aspects of the gut microbiota.

Using a combination of data and statistical analysis on various parameters and variables from already accumulated data (genomic and metabolomic), this project will provide fundamental information on the pathogenicity of certain microbes and their effect on the gut microbiome. As a result, this study will demonstrate which microbes could be associated with markers of a healthy gut microbiota in various populations. Subsequently, it will provide suggestions for further investigations to resolve the assumed pathogenic potential of various intestinal parasites.

For this project, the student will develop and employ hierarchical Bayesian models to deal with latent variables and to perform Bayesian model selection. The Markov Chain Monte Carlo algorithm required for inference will be performed either using existing software, such as STAN or NIMBLE, or by producing bespoke code, or a combination of the two. The methods to be developed will be motivated by the particular application and available data but will be generally applicable to cases where interest lies in modelling associations between random variables, with some of them potentially latent, at various levels.