Integrating bacterial population genomic analysis in the epidemiology of a human zoonotic infection

Campylobacter is the most common bacterial cause of gastroenteritis in the UK, causing in the order of 300,000-400,000 cases per year and worldwide. Most human infection in the UK comes from chicken and food industry interventions are seeking to reduce this risk. Population genetic models, developed based on 7 genes, allow attribution of human infection to source [1-3]. . Now datasets covering the whole genome (1,500 genes) are available but no standard approach has been developed to use this much richer data to improve attribution estimates. Some work on these whole genome datasets suggests that genes associated with particular biological pathways can predict the original host animal species [4]. The initial work showing that pathogen genome is predictive of host species indicated that increasing the number of genes in a model-based population genetic analysis should improve accuracy [1]. Development of quantitative methods to use whole genome data to improve estimation of sources of human infection will support control measures as well as improving our understanding of the biology of *Campylobacter* and forms one of the research options for this project.

A striking feature of *Campylobacter* infection in humans is that although many cases occur, detected outbreaks are relatively rare. Detection of outbreaks also supports identification of the sources of human infection to guide control. Bayesian analysis of human surveillance data along with other routine data may improve outbreak detection [5], but statistical techniques for outbreak detection are surprisingly underdeveloped. Integration of pathogen genomic data is likely to substantially improve detection of outbreaks. Additionally the joint analysis of other aspects of human epidemiology alongside pathogen genomic data can clarify the different patterns of risk across a human population [6].

The overall aim of this project is the development and application of statistical methods to use population genetic analysis on bacterial genome data and to combine population genomic data with other epidemiological data to answer questions of public health relevance and biological interest. The project will have access to sentinel surveillance data from projects funded by Public Health England (data collection complete) and the Food Standards Agency (project). Supervision will include a statistician [5] and epidemiologist [1] who have developed analytical approaches in these areas. Funding for a PhD from the Food Standards Agency is based on that PhD including research to develop source attribution using whole genome data from human and a range of animal and food origin isolates. The focus of the MSc project can be on testing approaches to source attribution but can also focus primarily on outbreak detection if preferred. Existing methodology in both of these areas relies heavily on computationally intensive statistical methods from Bayesian statistics, such as Markov chain Monte Carlo (MCMC).

1. McCarthy ND, Colles FM, Dingle KE, Bagnall MC, Manning G, Maiden MC, Falush D. Host-associated genetic import in Campylobacter jejuni. Emerg Infect Dis. 2007 Feb;13(2):267-72.
2. Sheppard SK, Dallas JF, Strachan NJ, MacRae M, McCarthy ND, Wilson DJ, Gormley FJ, Falush D, Ogden ID, Maiden MC, Forbes KJ. Campylobacter genotyping to determine the source of human infection. Clin Infect Dis. 2009 Apr 15;48(8):1072-8.
3. Sheppard SK, Didelot X, Meric G, Torralbo A, Jolley KA, Kelly DJ, Bentley SD, Maiden MC, Parkhill J, Falush D. Genome-wide association study identifies vitamin B5 biosynthesis as a host specificity factor in Campylobacter. Proc Natl Acad Sci U S A. 2013 Jul 16;110(29):11923-7.
4. Sheppard SK, McCarthy ND, Falush D, Maiden MC. Convergence of Campylobacter species: implications for bacterial evolution. Science. 2008 Apr 11;320(5873):237-9.
5. Spencer SE, Marshall J, Pirie R, Campbell D, French NP. The detection of spatially localised outbreaks in campylobacteriosis notification data. Spat Spatiotemporal Epidemiol. 2011 Sep;2(3):173-83.
6. Mullner P, Shadbolt T, Collins-Emerson JM, Midwinter AC, Spencer SE, Marshall J, Carter PE, Campbell DM, Wilson DJ, Hathaway S, Pirie R, French NP. Molecular and spatial epidemiology of human campylobacteriosis: source association and genotype-related risk factors. Epidemiol Infect. 2010 Oct;138(10):1372-83.