Generating tools for the molecular epidemiology of Campylobacter coli by next generation genome sequencing

Area of research interest: Foodborne pathogens
Study duration: 2013-07-01
Project code: FS101087
Conducted by: Institute of Food Research (IFR)

Background

The bacterium Campylobacter is the most common cause of bacterial food poisoning in the United Kingdom, with between a quarter and half a million cases annually. Similar figures are seen in other member states of the European Union, and reducing Campylobacter-related disease is one of the major goals of the Food Standards Agency and other funding bodies. There are two Campylobacter species which are primarily responsible for infections: Campylobacter jejuni, which is most commonly found on poultry meat and also in farm animals, and Campylobacter coli, which is also found on poultry meat but is thought to have more diverse set of environmental reservoirs. Approximately 10-15% of cases of Campylobacter illness is caused by C. coli, but despite this significant contribution to human disease, there has been relatively little attention given to C. coli, with most research focusing on C. jejuni. While these two bacterial species are very closely related, they are sufficiently different in their biology and transmission to warrant specific attention to be given to C. coli, and a major aim of this project was to generate the data required to allow for future C. coli-specific investigations, as well as generating new insights in the epidemiology of C. coli-caused disease.

Research Approach

Until recently, the most commonly used technology to compare and trace Campylobacter isolates has been multilocus sequence typing (MLST), where a fingerprint is generated from 7 small conserved parts of the bacterial chromosome. While powerful, this technology does not have the resolution required to work with C. coli, as most isolates cluster together in what is called a clonal complex, named ST-828 for C. coli. Recent work with genome sequencing has demonstrated that C. coli is divided into distinct lineages (Clades) representing agricultural and environmental isolates. However, that was still based on a relatively small sample, and needed to be tested in a larger study with independent isolates. Also, what was still lacking for C. coli is sufficient resolution to allow the development of typing tools for rapid diagnostic and epidemiological investigation.

DNA sequencing technology has changed very rapidly over the last 20 years, and sequencing of genomes has now become both feasible and cost-effective for the characterisation of large numbers of C. coli isolates, with the goal of identifying diagnostic markers and to use these for future epidemiological purposes.

The overall aim of this project was to use genome sequencing technology to determine and analyse the genetic diversity of Campylobacter coli, and to compare this diversity with information about where it was obtained from (agriculture, environment, clinical samples or food). Specific objectives were:

- to determine the genome sequence of 500 UK C. coli isolates from three different collections with a wide variety of isolation sources
- to define the genetic variation within the C. coli genome
- to compare the newly generated C. coli genomic information with existing typing methods
- to support development of a typing scheme specific for C. coli using the information obtained from the genomes.

Results

A total of 507 *C. coli* isolates were used for genome sequencing, with 497 used in further analyses. Almost all isolates originated from three independent UK collections. The isolates were subdivided in four categories: 1) 239 isolates from agricultural sources (farms, animals, poultry); 140 isolates from environmental sources (water, soil, ducks); 59 isolates from clinical sources (blood, stool); and 59 from food (chicken, lamb, other food). The genomes were assembled from the sequencing data, the genes identified and then analysed using a diverse set of tools looking at diversity, ancestry and phylogeny. From the analyses it was very clear that *C. coli* has a population structure which is very different from that of *C. jejuni*, and has three ancestral clades (major groups), with the largest group having three different sequence clusters, of which Cluster A (ST-828) and Cluster B (ST-1150) clusters have acquired sequences from *C. jejuni*. Clade 1-Clusters A and B consisted mostly of agricultural, food and clinical isolates, whereas Clade 1-Cluster C and Clades 2 and 3 consisted mostly of environmental isolates. These five groupings are very different, suggesting they do not exchange DNA readily, and hence may not share the same environmental/agricultural niches. Genes specific for each Clade/Cluster have been identified, and compared to other datasets with regard to their usability in the future development of epidemiological tools. Importantly, this study addressed the underrepresentation of the environmental Clades in the available *C. coli* genome sequences, thus strengthening downstream analyses. Finally, datasets from this study were used to predict the minimally required amount of sequence data to be able to determine the genome sequence of *Campylobacter*, thus assisting future studies.

Conclusions: *Campylobacter coli* has a very distinct population structure with five clearly defined groups, with two groups (Clade 1-Clusters A and B) primarily associated with agricultural sources and human disease, while the other three groups (Clade 1-Cluster C and Clades 2 and 3) were mostly associated with environmental sources. *C. coli* has a very distinct population structure, with separation of lineages. The genome sequences and analyses provided by this project will support future studies into the biological differences between these groups, and assist in determining their significance in human disease.

Research report

England, Northern Ireland and Wales

PDF
[View Generating tools for the molecular epidemiology of Campylobacter coli by next generation genome sequencing - Research report as PDF(Open in a new window)](#) (1.57 MB)

England, Northern Ireland and Wales

EXCEL
View FS 101087 supplementary tables as Excel(Open in a new window) (46.98 KB)

England, Northern Ireland and Wales

EXCEL
View FS 101087 supplementary tables sequencing information as Excel(Open in a new window) (93.86 KB)

England, Northern Ireland and Wales

EXCEL
View FS 101087 supplementary tables variable genes as Excel(Open in a new window) (44.81 KB)

England, Northern Ireland and Wales

EXCEL
View FS 101087 supplementary tables mobile elements as Excel(Open in a new window) (39.58 KB)

England, Northern Ireland and Wales

EXCEL
View FS 101087 supplementary tables ad specific as Excel(Open in a new window) (60.05 KB)

England, Northern Ireland and Wales

EXCEL
View FS 101087 supplementary tables amr as Excel(Open in a new window) (39.52 KB)

England, Northern Ireland and Wales

PDF
View FS 101087 supplementary tables 7 as PDF(Open in a new window) (375.87 KB)