'Molecular epidemiology study to investigate whether dairy cattle are a reservoir of STEC that may cause human clinical disease in the Waikato region of New Zealand'

Delphine Rapp[1], Colleen Ross[1], Paul MacLean[2], Adrian Cookson[1]

1 Food Assurance and Meat Quality
2 Bioinformatics & Statistics

October 2016

Report for Waikato Medical Research Foundation

REPORT FBP 67144. Contract A22459

WMRF 2015 Supplementary Grant Round; project entitled "investigation into the genome of *Escherichia coli* from the Waikato bovine reservoir"

Inquiries or requests to:
Delphine Rapp
Delphine.rapp@agresearch.co.nz
Food Assurance and Meat Quality Group, AgResearch Ltd
Ruakura New Zealand

Dr Gale Brightwell
Science Team Leader
Food Assurance and Meat Quality Team
Food & Bio-based Products Group

Contents

1. Background

Shiga toxin-producing *Escherichia coli* (STEC) O26 is among the leading *E. coli* serogroups responsible for severe cases of diarrhoea and haemolytic uraemic syndrome (HUS) in humans. STEC virulence factors include potent Shiga toxins encoded by bacteriophage-associated *stx* genes, and an outer membrane protein, intimin, encoded by the *eae* gene that lies in a DNA region responsible for the bacterial attachment to intestinal epithelial cells. Ruminants, with cattle in particular, have been recognized as an important reservoir of O26 strains. Some *E. coli* O26 isolated from cattle do not possess virulence characteristics including Shiga toxins (*stx*), however human isolates also lacking *stx* have been isolated on rare occasions from human cases of severe clinical infection, suggesting a potential role for *stx*-negative O26 variants in diarrhoeal disease.

The objective of the project was to genome sequence *E. coli* O26 strains originating from cattle and human clinical cases, and to compare their genetic characteristics to elucidate public health risks associated with bovine isolates.

2. Methods

Whole genome sequencing was performed on a panel of 20 *E. coli* O26, of which 17 were bovine isolates originating from dairy herds of the Waikato, two from Waikato clinical cases and one a reference STEC strain isolated from human case. The investigated isolates were *stx*+*eae*+ (i.e. STEC) (n=7), *stx*-*eae*+ (n=6) or *stx*-*eae*- (n=7). The genetic diversity of the investigated bovine isolates was ensured by selecting isolates that generated a different profile by enterobacterial repetitive intergenic consensus (ERIC)-PCR. Whole genome sequencing was done by New Zealand Genomics Limited. The expected read depth of sequencing ranged between 43 and 241-fold and averaged 104-fold across the isolates, indicating sufficient coverage to detect single nucleotide variation (SNP) with a high degree of confidence. Identification of genes was by alignment of the sequencing reads with the genome of a fully sequenced STEC O26 strain (*Escherichia coli* O26:H11 strain 11368) available in the NCBI data base. The genes with at least 25 matches were recognised.

3. Summary

Upon comparison of the assembled genomes, about 3,000 SNPs difference was observed between the bovine *stx*-*eae*+ strains and the STEC O26 strains, while >30,000 nucleotide variations were identified between the bovine *stx*-*eae*- strains and the STEC O26 strains. The bovine strains with the *stx*-*eae*+ virulence profile cluster were phylogenetically more closely related STEC O26 (*stx*+*eae*+) than to the *stx*-*eae*- strains (Figure1).

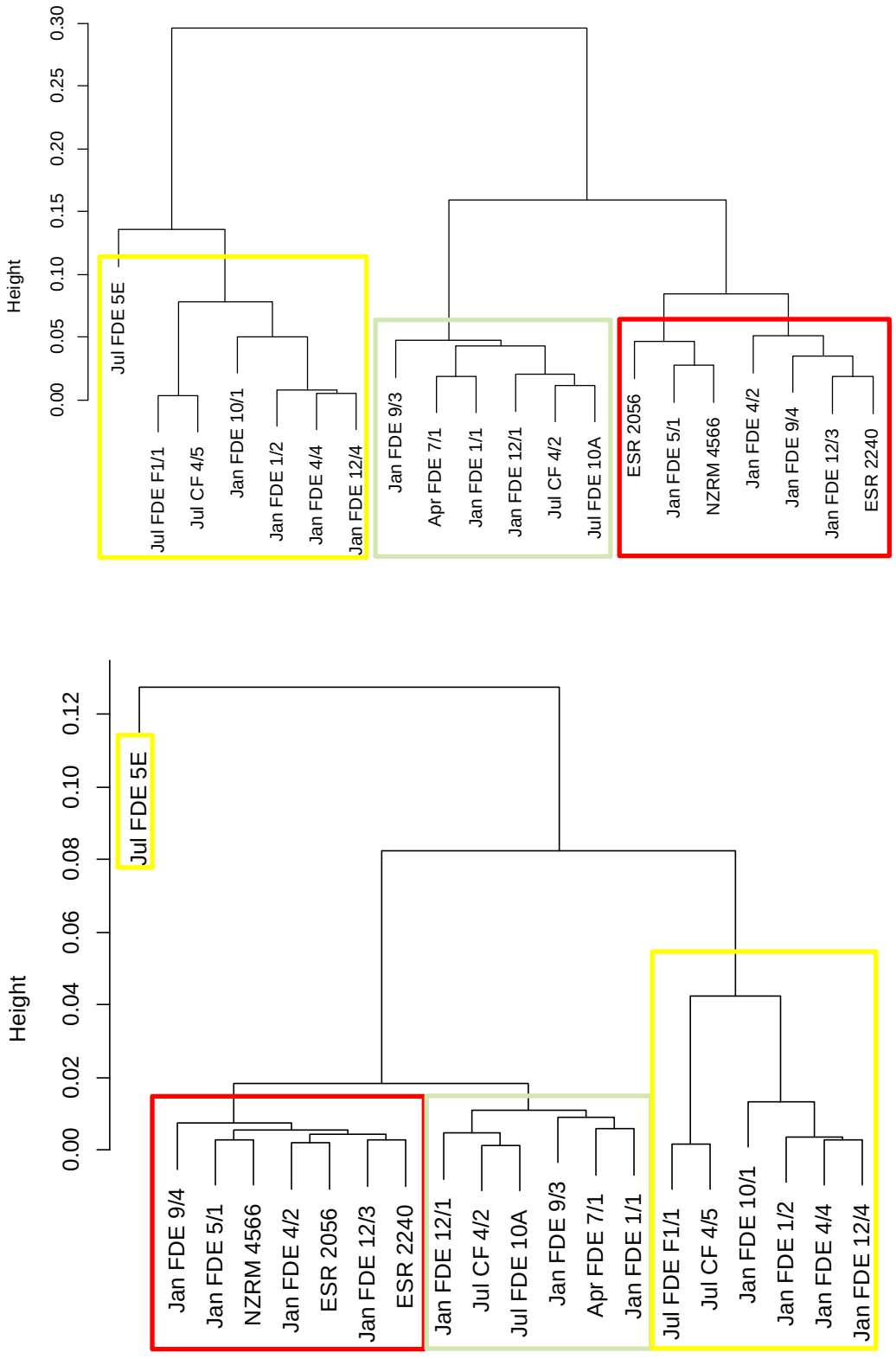

Fig 1. Clustering of the 20 investigated *E. coli* O26 isolates based on (top) 5,993 *E. coli* O26:H11 strain 11368 genes present in any of the samples, and (bottom) 4,384 *E. coli* O26:H11 strain 11368 genes present in any of the samples outside of phages, insertion elements and plasmids. STEC O26 strains are in the red box, *stx-eae*+ O26 strains are in the olive box and *stx-eae*- O26 strains are in the yellow box. CF: calf faeces; FDE: farm dairy effluent.

Over 90% of the genes associated with STEC O26 were present in the *stx-eae*+ strains of bovine origin. Pathogenesis and dioxygenase activity were the main gene-associated biological functions shared by STEC and *stx-eae*+ strains but not well represented in the *stx-eae*- strains.

As for STEC O26 strains, strains of the *stx-eae*+ O26 cluster carried the complete integrative DNA region responsible for the bacterial attachment to intestinal epithelial cells. They also carried two known and genetically conserved insertion sites of the *stx* genes. Additionally, of 96 genes encoding known virulence factors, 47 were sufficiently conserved among the *stx-eae*+ strains to permit identification, although they often harboured nucleotide variations compared to the STEC O26 strains (with for example between 4 and 22 variations found in the genes involved in fimbrial biosynthesis). Further comparison of the genome sequence data demonstrated several distinct genetic characteristics lacking in bovine *stx-eae*+ strains but present in STEC O26. For example, fewer "mobile" -often virulent- genetic elements (i.e. phages, plasmids, integrative elements); more nucleotide variations in both shared and the "mobile" genetic elements; a phenotype less able to integrate and recombine DNA and thus to acquire genes through horizontal transfer; and an inability to metabolise urea, or to bind nickel. Overall, these findings demonstrate the presence of an array of virulence factors in bovine *stx-eae*+ strains but suggest that *stx-eae*+ strains investigated in this study are unlikely to be associated with severe diarrhoeal disease in humans. Further investigation regarding the difference in protein sequences and functions are required to confirm virulence and pathogenicity associated with the nucleotide mutations and lack of virulence factors.

The bovine *stx-eae*- O26 strains were genetically very different from bovine or human STEC O26. A large proportion (31-35%; 1831-2125 genes) of the genes of *stx-eae*- strains were absent or highly divergent compared to the reference database. Compared to STEC O26 strains, fewer "mobile" elements and associated genes (between three and eleven phages for a total of 48-97 phage-encoded genes; one to four integrative elements with 1-37 genes in total; and up to two plasmid-encoded genes, with 1-46 genes in total) were identified. No factor specific for bovine colonisation were identified, as all the genes present in the bovine *stx-eae*+ and *stx-eae*-strains were found in bovine and human STEC O26 (*stx*+*eae*+).

To confirm these findings, we recommend further genomic studies to be carried out with a higher number of isolates, and with isolates selected to be closely related, i.e. that generate a similar profile by enterobacterial repetitive intergenic consensus (ERIC)-PCR.

4. List of key findings

- The three main clusters observed during phylogenetic analysis corresponded with virulence profile (*stx*+*eae*+, *stx*-*eae*+, or *stx*-*eae*-) independently of analysis of complete genome assembly or nucleotide variation (Single Nucleotide Polymorphism (SNP) analysis). The bovine *stx*-*eae*+ strains were phylogenetically more closely related to STEC O26 than to the *stx*-*eae*- strains.

- The lowest numbers of SNPs were obtained for the clinical and the bovine STEC isolates (100-800 SNPs difference relative to the reference genome AP010953), followed by *stx*-*eae*+ strains (3,000 SNPs). The highest number of SNPs (> 30,000) was observed for most of the stx-eae- strains.

- 60% of the genome (3,621 genes) from the O26:H11 reference strain 11368 (GenBank accession no. AP010953) were shared between all 19 genome sequenced strains; 332 genes present in STEC O26 were not detected in the bovine *stx*-*eae*+ isolates.

- Much of this genome variation was attributable to “mobile” genetic elements: phages were more abundant in the stx-eae+ isolates compared to STEC O26 (16-18 versus 7-19) but in contrast there were fewer phage-encoded genes in the *stx*-*eae*+ isolates compared to STEC O26 (314-418 versus >500). Genome sequence data from the *stx*-*eae*+ strains also indicated fewer integrative elements (IE) compared to STEC (6 to 7 versus 9). In the *stx*-*eae*+ strains there were fewer genes present in three IEs compared to STEC O26, however the number of genes in IE08 (locus for enterocyte effacement, LEE), which encodes *eae* and other factors responsible for attachment to intestinal epithelial cells, was identical in the bovine stx-eae+ and STEC strains. As noted in previous studies, STEC O26 contained the most plasmid-encoded genes (63-131 genes), compared to stx-eae+ isolates (28-62 genes).

- The *stx*-*eae*- strains had between three and eleven phages (48-97 phage-encoded genes in total); one to four integrative elements (1-37 genes) and up to two plasmid-encoded genes, (1-46 genes).

- Genome analysis demonstrated that there were fewer SNPs across the STEC O26 sequences compared to the *stx*-*eae*+ variants suggesting that the STEC O26 may form a tighter clonal type. The large virulence plasmid *pO26*_1 (85,167 bp; 93 genes) is stably maintained across all of the STEC and *stx*-*eae*+ isolates, suggesting it contains important virulence determinants for bovine/human colonisation.

- The presence of the *stx*-encoding bacteriophage insertion site in the *stx*-*eae*+ isolates indicated that bacteriophage insertion and stable (i.e. vertically

transmitted between successive generations) *stx*-phage acquisition is possible but its insertion frequency is unknown.

- The main biological functions that differentiated *eae*+ isolates (e.g. STEC and *stx*-*eae*+) and the *stx*-*eae*- isolates were pathogenesis (GO-term GO:0009405) and dioxygenase activity (GO:0051213). The most significant functions that differentiated STEC and *stx*-*eae*+ isolates were DNA integration and recombination; endo- and deoxyribonuclease activities; urea metabolic and catabolic processes, urease activity; and nickel ion binding. The GO terms for pathogenesis and iron uptake systems were also identified but with a low significance.

- Of the 96 genes encoding known virulence factors, 47 were sufficiently conserved between STEC and *stx*-*eae*+ strains, despite the presence of SNPs, to permit detection. Up to 26 SNPs were detected in the virulence genes from *stx*-*eae*+ strains compared to STEC O26. Between 4 and 22 SNPs were found in the genes involved in fimbrial biosynthesis (*fim*), which are important for initial bacterial attachment to the human intestine. However the impact of these SNPs on pathogenicity and attachment to contrasting bovine/human intestinal cells remains to be determined.

- The *stx* genes for the seven STEC isolates were all identified as *stx*1a. No *stx*2 genes were identified in any of the investigated bovine or human isolates.

- Whole genome phylogenetic comparison of the STEC O26 strains including phage content and a SNP (A->T, in nucleotide 1,579,682 of sequence AP010953.1) present in the *stx1a* gene suggests the presence of two closely related populations of STEC O26, each containing both bovine and human strains.

5. Dissemination of Project's Results

- Oral presentation at Waikato Hospital Journal club – 30th May 2016.

- Oral presentation at the 8th Annual New Zealand Next Generation Sequencing conference – 6-8th October 2016.

6. Statement of how the funds were used

The fund ($19,481) was used for the project accordingly to the application. Spending was as followed:

Project personnel: $8,767

Consumables: $3,688

Full genome sequencing service, outsourced to NZGL: $7,011

Postage: $15