8. Final discussion

Campylobacter is known to cause disease in humans. Symptoms of campylobacteriosis range from mild to severe with neurological sequelae. Different strains of *C. jejuni* have been shown to be genetically diverse and show a range of phenotypes that may relate to clinical outcome. In order to identify genetic differences between strains a similar differential hybridisation approach, using macroarrays, to that developed by Liang *et al.* [92] was used. This method was chosen above other comparative genomic methods that could be used, for example microarrays, which are relatively difficult to set up or subtractive hybridization, which requires a great deal of sequencing to achieve adequate coverage of novel regions [90] and in addition requires a subcloning step after PCR amplification and subtraction which can introduce biases. The method of subtractive hybridization has also been associated with a high level of false positive sequence identification [91].

The use of macroarrays has been shown to be sensitive for use with expression profiling so it should be a sensitive enough technique to identify clones carrying inserts that differ between strains, provided arraying and hybridization conditions have been optimised [226]. One limitation of the method is that some regions may not be picked out by hybridisation due to labelling difficulties or secondary structure of the DNA. There may be regions of the chromosome not represented in the clone libraries due to clone distribution biases [227] but generating a clone library with 5x coverage of the genome should ensure that as much of the genome as possible is represented. In addition some clones may be lost if they did not grow well or were lost during handling and membrane preparation but the effects of this should again be limited by using 5x coverage of the genome.

The analysis of the efficacy of the method from the differential hybridization of pUC libraries for strains 81-176 and M1 suggested that approximately 20% of novel sequence was

missed by the pUC library sequencing. When the pUC contigs were compared to the novel sequences from BAC clones 22% extra sequence was identified in strain 81-176; 23% extra sequence was identified in strain M1; 38% extra sequence was identified in strain 40671; and 35% extra sequence was identified in 52472 discounting bacteriophage associated sequence. Bacteriophage sequence was discounted from the coverage figures as the pUC sequences associated with bacteriophage are all highly similar and therefore true coverage cannot be determined. The extra sequence gained from the BAC clones also shows that the combination of two libraries constructed using different methods (i.e. pUC with sonication and BAC with enzyme digestion) is a powerful tool for identifying novel regions between different strains and may go some way to compensating for biases that occur using either method individually. However it is clear that this approach will never be as comprehensive as a full genome sequence and that there are likely to be regions that have not been identified using this method.

Other methods, such as microarrays, appear to show bias as well [170]. As strain 81-176 has been extensively studied many strain specific genes have already been characterized, some of which have been identified in this study. Using strain 81-176 as a benchmark it is clear that this method is not identifying all the novel regions when compared to strain 11168. This is likely to be due to both bias in the library construction and problems with the hybridization method. The pUC libraries identified 37 contigs out of 58 identified in a recent 81-176 microarray study [170]. A further 3 contigs were discovered in the BAC sequences, leaving 18 out of 58 which have not been identified in this study (31%) of the total number identified by Poly *et al*. [170]. However the microarray method failed to identify many regions that have been expanded in this study: 59% of pUC sequence and 40% of novel BAC sequence was not identified but this may in part be due to different hybridization stringencies and a 2.8x coverage of the genome sequence in the microarray study [170].

Not only do additional genes in the form of plasmids, bacteriophage and pathogenicity islands contribute to virulence but the loss of genes may also have a marked effect. For example *Shigella* spp. and enteroinvasive *Escherichia coli* have a large deletion around the lysine decarboxylase gene, *cadA*, when compared to the *Escherichia coli* laboratory strain K-12. When this gene was introduced to *Shigella,* attenuated virulence was seen and enterotoxin activity was also inhibited [228]. The method of differential hybridization, comparing pUC libraries of test strains to genomic DNA from strain 11168 can not detect regions that are absent from the test strains compared to 11168. Differential hybridization data can be interpreted in conjunction with microarray data comparing these strains against 11168 to give an overall picture of the gene content of these strains. In order to truly evaluate different methods it would be necessary to compare two sequenced strains as has been done for subtractive hybridization with two strains of *Helicobacter pylori* [90].

In this study a number of strains were compared using differential hybridization to represent a range of characteristics present within the *Campylobacter* species. Strain 81-176 represents a highly studied laboratory strain originally isolated from an outbreak thought to be associated with raw milk. Strain M1 represents a strain with a clear transmission link between poultry and humans as disease was developed after a visit to a poultry abattoir. Strain 40671 represents a strain from an outbreak thought to be associated with water. Strain 52472 was isolated from a patient with septicaemia. These strains were all compared to the sequenced strain 11168 using a differential hybridization method. This method led to the identification of 93 CDSs in strain 81-176 some of which were expanded using BAC libraries to give 8 novel regions from BAC clones. In strain M1 137 CDSs were identified some of which were expanded into 10 novel regions; 97 CDSs in 40671 were expanded into 6 novel regions; and 268 CDSs in strain 52472 were expanded into 7 regions.

The strain 81-176 has two plasmids; pVir and pTet. In order to discriminate between plasmid encoded and chromosomally encoded novel CDSs in strain 81-176 these plasmids were sequenced. Plasmids pVir and pTet were originally isolated from strain 81-176 but subsequently very similar plasmids have been isolated from other strains: 17% of clinical isolates from Canada contained pVir [130] and 50% of clinical isolates from another study were resistant to tetracycline, 67% of which harboured tetracycline resistance plasmids [207]. In addition the tetracycline resistance plasmid appears to be highly conserved between species with plasmids pCC31 from *C. coli* and pTet from *C. jejuni* showing 94.3% aa id to each other [128]. Interestingly predicted CDSs with homology to those from pTet have been discovered from the pUC screen in the strains 40671 and 52472. It is unclear whether these regions carrying the predicted CDSs may be integrated into the chromosome or present on plasmids, however these strains appear to contain a full complement of homologues necessary to form a type IV secretion system. In strain 40671 homologues of VirB5 and VirD2 were not identified but these are also missing from some type IV secretion systems involved in protein secretion [72;190].

An interesting feature apparent from the sequencing of pTet was the identification of a putative DNA invertase, flanked by 31bp inverted repeats, upstream of homologues of a type IV secretion system. This region is thought to hold a promoter which is moved upon inversion of the DNA segment to be located upstream of a putative type IV secretion system. The region flanked by inverted repeats has been shown to invert and the DNA invertase was shown to be responsible for this inversion. Further studies would be needed to confirm whether the type IV secretion system is under variable control and what effect this has on conjugation.

Although plasmids represent one method of horizontal transfer of genes into a strain and have been implicated in carrying virulence factors in other bacteria; such as YOPs of

Yersinia enterocolitica [78] and IPA of *Shigella flexneri* [77] as well as in *Campylobacter* [127], chromosomal determinants may also be important. In many other bacteria, pathogenicity islands, phage, IS elements and transposons have been implicated in the dissemination of virulence determinants [71]. Areas adjacent to tRNA genes have been shown to be a common site of insertion for such mobile elements, as is the case for pathogenicity islands of uropathogenic *Escherichia coli* and also for the integration of bacteriophage [71]. However no insertion sequence (IS) elements were found within the genome sequences of *C. jejuni* strains 11168 or RM1221, and only the genome of RM1221 contained genomic islands in the form of integrated phage and plasmid DNA [8;9].

A total of 595 partial and complete predicted CDSs were identified using the differential hybridization approach. There were some common themes between all the test strains with surface structure associated CDSs, transporters, restriction modification CDSs and hypothetical CDSs being identified which have all been shown to vary between strains in other studies [83-85;91]. In addition each of the strains used in this study had a unique profile of predicted CDSs.

Strain 81-176 contained the fewest chromosomal differences to 11168 of all the test strains as identified by the pUC library sequencing. The BAC libraries were used to locate and expand the sequence from a putative cytochrome C biogenesis operon and a putative dimethyl sulfoxide reductase operon which may aid survival under reduced oxygen tensions such as those found in the human and animal gut. Strain 81-176 also contains a novel putative serine protease and di-tripeptide transporters which may aid survival by providing nutrients. This strain also contained three putative TPS systems but further work will be needed to identify whether these are functional, as several other strains contain degenerate forms of these.

Strain M1 also contained a novel putative cytochrome C biogenesis operon and a *dmsABC* operon as identified in strain 81-176. The three putative TPS systems identified in this strain contain different putative secreted proteins to those identified in strain 81-176 and those of strain M1 show homology to adhesins. It is possible that these putative adhesins aid colonization of the chicken. Strain M1 also contains a novel putative autotransporter, and although a function for this cannot be extrapolated based on sequence homology, many autotransporters have been associated with virulence functions from toxins to adhesins [192]. Intriguingly a chromosomal *tetO* insert was discovered in this strain leading to the possibility that it may be associated with a transposable element. This possibility was explored further by identifying the location of *tetO* genes in clinical tetracycline resistant isolates from Canada that were not thought to contain tetracycline resistance plasmids. Out of 8 isolates studied 2 were found to contain a similar chromosomal insertion to that of strain M1 with a conserved *tetO* and downstream hypothetical CDS but different surrounding CDSs. This insertion also contained homologues of the IS607 transposable element of *Helicobacter pylori* [211]. This poses the intriguing possibility that *tetO* may be located on a transposable element. However more work will be needed to explore whether or not this region is mobile and whether it can be transmitted between strains. Previously studied tetracycline resistance determinants from *C. jejuni* have all been carried on plasmids, although not all tetracycline resistant strains appear to contain plasmids [207].

Many predicted CDSs from the pUC screen in strain 40671 were hypothetical with no predicted function. Further investigation identified many of these as being associated with a novel capsule locus. As this strain has been associated with water it may be that this capsule leads to increased survival of 40671 in the environment compared to other strains. A number of hypothetical CDSs that may be associated with metabolism and a novel oxidoreductase were identified. Oxidoreductases are used in many metabolic pathways so

further investigation would be needed to elucidate the precise function of these in strain 40671. A novel putative MCP chemotaxis CDS was identified which may also aid survival if this allows the bacterium to respond to environmental conditions. Components of a putative type IV secretion system were identified in strain 40671 but further work would be needed to examine whether these are colocalized with other homologues of pTet CDSs either on a plasmid or integrated into the chromosome.

Strain 52472 also has components of a type IV secretion system. This strain contains many regions of bacteriophage associated DNA some of which were identified as being inserted in the chromosome. There are homologues of an autotransporter and a TPS system, both represented by pseudogenes, suggesting that the intact equivalents performed no function which enhanced survival of this strain in the environmental niche it inhabited. There are many metabolism associated CDSs which showed limited identity to those from strain 11168, as identified by the pUC screen, and also an intact homologue of a PrpD-family protein as present in strain RM1221. These data suggest that strain 52472 may have different metabolic capabilities to strain 11168 although the implications of this for survival would require further study. Strain 52472 appears more similar to strain RM1221 than to strain 11168 with many of the CDSs identified as not being present in strain 11168 being found in strain RM1221. These similarities were mostly due to strains 52472 and RM1221 containing similar bacteriophage DNA. In strain 52472 a putative RM operon that also contained a putative protein kinase was identified. This arrangement is reminiscent of the phage limitation system of *Streptomyces coelicolor* [179]. It would be interesting to explore the implications of this putative operon in strain 52472.

There are many pseudogenes among the novel inserts suggesting that these are accessory genes only beneficial to the bacterium in a small subset of the environmental niches which it inhabits. Of the novel CDSs identified in this study 21% are inactivated in

one strain or another. This is compared to the number of pseudogenes in the chromosomal core of 1.3% in strain 11168. It should be noted that comparisons within this study identified many previously unrecognized pseudogenes on the chromosome of strain 11168. This suggests that if DNA present in one strain is found in another, as identified by DNA hybridization, this does not indicate that genes carried on this DNA are functional.

An important feature apparent from this study is that metabolic pathways may be variable and these may play a key role in adaptation and survival in different environments. It is thought that *C. jejuni* is extremely susceptible to a wide range of environmental stresses and does not grow below 30°C [2] but may persist in the environment for several weeks [16]. It appears likely that along with traditional features such as surface polysaccharide (LOS and capsule) that other features associated with accessory metabolic pathways, respiration, uptake of different nutrients and catabolism may be important for differential survival in the environment.

This study has identified many novel regions that could be involved in pathogenicity. Further work could be done to explore this possibility. It may be possible to examine the amount of respiratory divergence between these test strains by comparing growth rates in media supplemented with different respiratory substrates, such as formate, lactate and pyruvate [180]. Growth could also be compared in the presence of the alternative terminal electron acceptors fumarate, nitrate, dimethyl sulphoxide (DMSO) and trimethylamine-*N*-oxide (TMAO), under anaerobic conditions [229]. Strains 81-176 and M1 are predicted to encode a dimethyl sulfoxide reductase operon. In order to test this possibility alternative electron acceptor activity could be measured using methyl or benzyl viologen-linked reductase assays [18]. The role of the putative cytochrome C in strains 81-176 and M1 could be explored in cell-free preparations using physiological electron donor and acceptor systems [185].

A putative autotransporter and putative TPS systems were identified. Protein secretion could be analysed using a similar approach to Konkel *et al*. who used 2D gel electrophoresis and immunoblot analyses to identify proteins expressed on incubation with epithelial cells and then used fluorescence microscopy to visualize secreted proteins upon binding to fluorescently labelled antibodies to bacterial protein [46]. Adhesion assays could also be performed. Other regions of interest that warrant further investigation are the MCP chemotaxis genes from strains M1 and 40671, the novel capsule of strain 40671 and the protein kinase associated with RM system in strain 52472.

Pathogenicity islands are often associated with incorporation of large DNA segments with a different G+C content to the surrounding chromosomal DNA suggesting recent transfer from another organism [71]. With the exception of the capsule region in 40671 (24 Kb) and the bacteriophage (>24 Kb) and plasmid regions in strain 52472 the average insert size for all strains was small (4 Kb) with a range of 865 bp to 14146 bp. Very few multi-gene inserts were found, with the highest number of novel CDSs identified in BAC 8B4F10 (7). There were even examples of domains within a gene being novel, e.g. MCP chemotaxis CDS from strain 40671 and the putative autotransporter which contains a different passenger domain in strains M1 and 52472. There was also no marked difference in G+C content of the novel regions identified from BAC sequencing from the background chromosomal G+C content. Many of the inserts contained DNA with a G+C content +/- 2% of the 30% average of *C. jejuni*. Notable exceptions to this are the *tetO* insert in strain M1 which has a G+C content of 40%; the capsule region in 40671 which has G+C content of 26%; the small insert between *ceuE* and tRNA in strain M1 with a G+C content of 19%; a predicted CDSs downstream of cj0031 in strain 81-176 with a G+C content of 24%; two hypothetical proteins in 40671 4B2B1 with a G+C content of 22% and the RM inserts between *peb3* and *lpxB* in strain 40671 and 52472 which have a G+C content of 25%. In this study relatively

few inserts were found adjacent to tRNA genes, the only exceptions being the TraG-like insert in strains 81-176 and M1, and the insert between orthologues of *ceuE* and cj1356c in strain M1. Only the *tetO* insert of strain M1 seems to be associated with a transposon. The remainder of inserts are more likely to have occurred by homologous recombination with exogenously acquired DNA. This could be the case for novel regions located adjacent to rDNA regions as these will be conserved between similar species. A number of the novel inserts show high identity to CDSs from other delta-epsilon proteobacteria e.g. *Helicobacter pylori*, *Shewanella oniedensis* and *W. succinogenes.*

Even when strains have no detectable DNA differences there can still be marked differences in levels of transcription. It has been noted that the sequenced 11168 and the original clinical isolate of the same strain vary in colonization, gene expression and virulence phenotype even though no differences could be detected by multiple high resolution molecular genotyping techniques [230]. Changes in gene expression must be important to allow the bacterium to shift its metabolism and respiration to cope with changing environments. *C. jejuni* has been shown to survive in water and retail meats as well as poultry [230]. Studies have shown that passage under different conditions can also alter virulence phenotypes [16;91], and another study has identified the flagellar regulatory system as important for pathogenesis [231]. Reduced virulence was shown to be attributable to reduced expression of genes with σ^{28} or σ^{54} promoters. *flhA,* a component of the flagellar export apparatus, is important for expression of genes with σ^{28} or σ^{54} promoters and σ^{28} represses expression of σ^{54} [231].

This study has provided a comprehensive survey of differences between four strains with different characteristics when compared to strain 11168 for which the genome sequence is available. A range of novel DNA which may well be involved in virulence or environmental survival of these strains has been identified providing targets for further

research. Attempts to identify the global gene pool of *C. jejuni* coupled to transcription studies may help in attempts to elucidate the pathogenicity of this organism.