

**Thesis submitted for the degree of  
Doctor of Philosophy  
at the University of Leicester**

**by**

**Gemma Louise Marsden M.Sc , B.Sc. (Hons)**

**Department of Genetics**

**University of Leicester**

**July 2007**

UMI Number: U235849

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI U235849

Published by ProQuest LLC 2013. Copyright in the Dissertation held by the Author.  
Microform Edition © ProQuest LLC.

All rights reserved. This work is protected against  
unauthorized copying under Title 17, United States Code.



ProQuest LLC  
789 East Eisenhower Parkway  
P.O. Box 1346  
Ann Arbor, MI 48106-1346

## Characterisation of the genetic diversity in the lipooligosaccharide core biosynthesis region of *Campylobacter jejuni*

Gemma Louise Marsden

Lipooligosaccharides (LOS) are an important constituent of the bacterial outer membrane involved in maintaining cellular structural integrity. In *Campylobacter jejuni*, LOS is postulated to have a role in the virulence of this important gastrointestinal pathogen. The ability of *C. jejuni* to vary LOS structure and the molecular mimicry of host gangliosides by LOS epitopes suggests a role in the avoidance of host defence mechanisms. Molecular mimicry may form the immunopathological basis for an association with Guillain Barré syndrome. LOS structures differ between strains due to gene content, sequence and tract variation in the LOS biosynthesis gene cluster. Studies have examined the extent of gene content variation in the LOS gene cluster and showed gene content in this region can vary extensively. The aim of this work was to search for novel gene content and further investigate the role of LOS in *C. jejuni*.

Both PCR and microarray-based methods were developed to classify 50 clinical strains by LOS gene content. Several strains had gene content that did not appear to correlate with any known class and further detailed analysis highlighted that variation is greater than previously supposed. Sequencing revealed two strains that contained capsule polysaccharide biosynthesis genes within the LOS biosynthesis cluster. The presence of capsule genes within the LOS biosynthesis cluster highlights the potential for recombination between these variable polysaccharide loci and further extends the possibility of interaction between the capsule and LOS polysaccharides.

A LOS large deletion mutant was constructed in the genome strain, NCTC 11168 removing the genes between *wlaA* and *wlaT*. This mutant shared characteristics of other deep rough mutants being slow growing and sensitive to antibiotics and detergents. Construction of this mutant confirmed the minimal LOS core biosynthesis gene content for bacterial viability and showed the importance of LOS for this bacterium through a role for LOS in host cell invasion.

## Acknowledgements

I have many people to thank so I will do it chronologically! Firstly I would like to thank Phil Butcher at St. George's for encouraging me to do a PhD and helping all those potential supervisors to recognise my value by making me stand out from the crowd. I would also like to thank Sally Cooper, Claire Senner and the rest of BpG@S for their friendship and support at St. George's and beyond! To Jason Hinds, thank you for putting up with my constant emails and general harassment, as well as your microarray knowledge and letting me work at BpG@S as a visitor!

Thank you to Julian for recognising and drawing out my abilities (especially in the past few weeks!) Andy – thank you for your encouragement throughout and especially your dedication in the past few weeks – my thesis is now very well travelled and I am privileged to include the photo below!

Thank you to both past and present members of labs 145 and 121. Emma – thank you for your encouragement and support on a professional level and for introducing me to Bombay Sapphire! Julie – the lab hasn't been the same since you became a human geneticist! Thank you for being my bay, tea and lunch buddy, as well as cinema and shopping aficionado! Ollie – you are the best Spod I know! Carry on cooking! Leicester and Science should be sorry to lose you.....! (Their loss is Oxford's gain!?!);-). Claire, Randeep, Sue and the rest of 121, the department wouldn't be the same without you! Big thank you's go to those that have contributed to this work – Jason, Claire, Naomi, Jainjuin Li, Noel McCarthy and Paul Everest. Final thanks go to my family for their support and for their advice....I'm sorry if I bent your ears too much. Neil – how have you put up with me? I'm not sure how you've coped over the last 4 months, let alone the last 6 years!!

This work was funded by a Medical Research Council CASE studentship, with the Health Protection Agency, Colindale.





List of Figures .....	9
List of Tables .....	12
List of Abbreviations .....	14
<b>CHAPTER 1: INTRODUCTION .....</b>	<b>16</b>
1.1 Campylobacter jejuni .....	16
1.1.1 Significance.....	16
1.1.2 Campylobacter biology .....	17
1.1.4 Clinical Presentation .....	19
1.1.5 Pathogenesis.....	21
1.1.6 Campylobacter genomics.....	24
1.2 Lipopolysaccharides .....	32
1.2.1 Structure and function.....	32
1.2.2 Diversity in the core of bacterial lipopolysaccharides .....	35
1.2.3 Lipooligosaccharides in <i>C. jejuni</i> .....	46
1.3 Aims and Objectives .....	49
<b>CHAPTER 2: MATERIALS AND METHODS .....</b>	<b>50</b>
2.1. Bacterial Growth Conditions .....	50
2.2. Bacterial Strain Storage .....	50
2.3. Media .....	53
2.4. Antibiotic and Media Supplements.....	54
2.5. Miscellaneous buffers and solutions.....	54
2.6. Extraction of DNA from bacteria .....	56
2.6.1. Extraction of chromosomal DNA from <i>C. jejuni</i> .....	56
2.6.2. Small scale extraction of chromosomal DNA from <i>C. jejuni</i> .....	57
2.6.3. Extraction of plasmid DNA from <i>E. coli</i> and <i>C. jejuni</i> .....	57
2.7. Quantification of DNA .....	59
2.8. Electrophoresis of DNA.....	59
2.8.1. Electrophoresis of DNA using an agarose gel .....	59
2.9. Purification of DNA.....	60
2.9.1. Gel purification .....	60
2.9.2. Purification following PCR or other enzymatic reactions .....	60
2.9.3. Ethanol Precipitation.....	61
2.10. Enzymatic modification of DNA .....	61
2.10.1. Restriction endonuclease digestion of DNA.....	61
2.10.2. Ligation of DNA .....	61

2.10.3. Dephosphorylation of restriction digested DNA .....	62
2.11. Transformation of DNA.....	62
2.11.1. Preparation of electrocompetent <i>E. coli</i> .....	62
2.11.2. Preparation of chemically competent <i>E. coli</i> .....	62
2.11.3. Electroporation of Plasmid DNA into <i>E. coli</i> .....	63
2.11.4. Heat-shock transformation of chemically competent <i>E.coli</i> .....	63
2.11.5. Preparation of electrocompetant <i>C. jejuni</i> .....	63
2.11.6 Electroporation of Plasmid DNA into <i>C. jejuni</i> .....	64
2.12. Amplification of DNA by Polymerase Chain Reaction.....	64
2.12.1. Standard PCR protocol .....	65
2.12.2. Colony PCR .....	66
2.12.3. TripleMaster®Mix PCR system .....	66
2.12.4. Long range PCR.....	68
2.13. DNA Sequencing .....	69
2.15 Analysis of chromosomal DNA by Southern blotting.....	70
2.15.1 Transfer of DNA .....	70
2.15.2. Labelling of the probe.....	70
2.15.3. Hybridisation and stringency washes.....	71
2.15.4. Blocking, antibody incubation and detection.....	71
2.16. Analysis of LOS by microarray hybridisation .....	72
2.16.1 Microarray design and construction.....	72
2.16.2 Labelling, hybridisation and data acquisition .....	72
2.17. Analysis of Lipooligosaccharide (LOS) .....	73
2.17.1. Whole cell preparation of <i>C. jejuni</i> LOS .....	73
2.17.2. Electrophoresis of LOS using Tricine SDS-Polyacrylamide Gel Electrophoresis (PAGE).....	74
2.17.3. Silver staining of LOS Tricine SDS-page gels .....	75
2.17.4. Electrotransfer of LOS from Tricine SDS-PAGE gels to PVDF membrane .....	75
2.17.5. Hybridisation of antibodies and ligands to the membrane, and membrane development.....	76
2.18 Analysis of Outer Membrane preparations .....	77
2.18.1. Extraction of outer membrane proteins from <i>C. jejuni</i> .....	77
2.18.2. Electrophoresis of protein by SDS-PAGE.....	77
2.18.3. Staining of protein SDS-PAGE gels.....	78
2.19 Preparation of dehydrated cells for capillary electrophoresis mass spectroscopy .....	79
2.20 Phenotypic Assays .....	80
2.20.1 Growth assays to assess growth rate.....	80
2.20.2 Sensitivity to antibiotics and detergents .....	80
2.20.3 Growth assays to assess outer membrane function.....	81
2.20.4 Invasion and adherence assays.....	82

CHAPTER 3: GENETIC DIVERSITY IN THE LOS BIOSYNTHESIS REGION .....	83
3.1 Introduction.....	83
3.1. Variation in the LOS core biosynthesis gene cluster of <i>C. jejuni</i> .....	83
3.1.2 Variation in LOS gene content and Guillain Barré Syndrome. ....	93
3.1.3 Types of variation in LOS gene content. ....	94
3.1.2 Multi Locus Sequence Typing and LOS cluster gene content.....	98
3.1.3 Aims of study.....	99
3.2 Analysis of LOS gene content of 50 ‘Campylobacter Sentinel Surveillance Scheme’ strains .....	99
3.2.1 Design of PCR screening method and primers.....	99
3.2.2 Results of PCR screen of 50 clinical strains .....	105
3.2.3 Correlations between HS serotype and LOS PCR results. ....	114
3.2.4 Associations between LOS PCR results and phage type. ....	114
3.2.5 Correlations between MLST sequence type and LOS class. ....	115
3.3 Discussion.....	119
3.3.1 Validation of the PCR based assay .....	119
3.3.2 Distribution of LOS core gene loci.....	120
3.3.3 Correlations with LOS gene content, serotype and phage type .....	120
3.3.3 Correlations between MLST and LOS biosynthesis gene type. ....	121
3.3.4 Conclusions.....	122
CHAPTER 4: DETERMINING THE GENE CONTENT OF THE UNKNOWN AND NOVEL STRAINS.....	124
4.1 Introduction.....	124
4.1.1 Summary of Results from PCR screen .....	124
4.2 Primer walking.....	125
4.2.1 Introduction.....	125
4.3 Phenotypic analysis of the novel strain LOS. ....	152
4.3.1 Analysis of LOS by Tricine-SDS PAGE .....	152
4.3.2 Analysis of LOS by Western blot with cholera toxin. ....	152
4.3.3. Conclusions.....	155
4.4 Long range PCR.....	155
4.4.1 Strategy and optimisation .....	155
4.4.2 Results.....	158
4.4.3 Conclusion .....	158
4.5 Sequencing of long range PCR amplicons.....	159
4.5.1 Method .....	159
4.5.2 Results.....	159
4.5.3 Analysis of completed sequences. ....	162
4.5.4 Analysis of non-contiguous and almost completed sequences .....	168
4.5.5 Analysis of incomplete sequences .....	172
4.5.6 Conclusion .....	178

4.6 Discussion .....	181
4.6.1 Primer walking .....	181
4.6.4 Sequencing analysis .....	182
4.6.4 Correlations between LOS group and other typing methods .....	187
4.6.5 Conclusion .....	194

## CHAPTER 5: DESIGN AND VALIDATION OF A LOCUS SPECIFIC MICROARRAY..... 195

5.1 Introduction .....	195
5.1.1 Current typing methods .....	195
5.1.2 Microarrays .....	196
5.1.3 Genomotyping of <i>C.jejuni</i> .....	197
5.2 Microarray design and rationale .....	200
5.3 Validation and results from GMv1 locus specific microarray .....	202
5.3.1 Microarray Validation and results .....	202
5.3.2 Discussion .....	208
5.4 Results from the second generation locus-specific microarray, GMv2 .....	209
5.4.1. Increasing the signal intensity .....	209
5.4.2 Proof of principle and validation experiments .....	212
5.4.3 Microarray analysis of CSSS strains .....	225
5.5 Discussion .....	238
5.5.1 Improving signal intensity .....	238
5.5.2 Validation of the second generation microarray .....	239
5.5.3 Locus-specific Microarray analysis of clinical strains .....	239
5.5.4 Correlation of Locus-specific Microarray analysis with primer walking and sequencing analysis .....	240
5.5.5 Whole genome microarray analysis and the glycome .....	241
5.6 Conclusion .....	246

## CHAPTER 6: DETERMINING THE MINIMAL LOS CORE GENE CONTENT FOR VIABILITY..... 248

6.1 Introduction .....	248
6.1.1 LOS mutants in <i>Campylobacter jejuni</i> .....	248
6.1.2 Predicting the minimal LOS gene content for viability .....	254
6.1.3 Current mutagenesis strategies .....	256
6.2 Construction of a <i>wlaP</i> - <i>wlaSC</i> deletion mutant .....	260
6.2.1 Cloning of <i>wlaP</i> and <i>wlaSC</i> into pUC19 .....	260
6.2.2 Obtaining a streptomycin resistant strain. ....	260
6.2.3 Construction of the CAT- <i>rpsL</i> <sup>WT</sup> cassette. ....	263
6.2.4 Inserting the CAT- <i>rpsL</i> <sup>WT</sup> cassette into pGLM8 .....	263
6.2.5 Allelic exchange between pGLM9 and pGLM10 with <i>C. jejuni</i> NCTC 11168 <sup>StrepR</sup> .....	267

6.3 Phenotypic analysis of the <i>wlaP-wlaSC</i> deletion mutant .....	277
6.3.1 Analysis of LOS from the <i>wlaP-wlaSC</i> deletion mutant by Tricine SDS-PAGE and western blotting .....	277
6.4 Construction of a <i>wlaA-wlaT</i> large deletion mutant.....	282
6.4.1 Cloning of <i>wlaA</i> , <i>wlaT</i> and CAT- <i>rpsL</i> <sup>WT</sup> into pUC19.....	282
6.4.2 Allelic recombination between pGLM12 and pGLM13 with <i>C. jejuni</i> NCTC 11168 <sup>StrepR</sup> .....	282
6.4.3 Inserting a kanamycin resistance cassette into pGLM12.....	285
6.4.4 Allelic exchange between <i>C. jejuni</i> NCTC 11168:: <i>wlaP-wlaSC</i> , pGLM 14 and pGLM15.....	285
6.5 Phenotypic analysis of the <i>wlaA-wlaT</i> large deletion mutant .....	296
6.5.1 Analysis of LOS from the <i>wlaA-wlaT</i> deletion mutant by Tricine SDS-PAGE and western blotting.....	296
6.6 Phenotypic analysis of the <i>C.jejuni</i> <i>wlaP-SC</i> and <i>wlaA-T</i> deletion mutants.....	301
6.6.1 Growth patterns of the <i>C. jejuni</i> <i>wlaP-SC</i> and <i>wlaA-T</i> deletion mutants. ....	301
6.6.2 Sensitivity of the <i>C. jejuni</i> <i>wlaP-SC</i> and <i>wlaA-T</i> deletion mutants to SDS, novobiocin and polymyxin B.....	306
6.6.3 Invasive potential of the <i>C. jejuni</i> <i>wlaP-SC</i> and <i>wlaA-T</i> deletion mutants...	309
6.6.4 Outer membrane protein composition of the <i>C. jejuni</i> <i>wlaP-SC</i> and <i>wlaA-T</i> deletion mutants. ....	309
6.7 Analysis of the LOS from 11168:: <i>wlaP-SC</i> and 11168:: <i>wlaA-T</i> by Capillary Electrophoresis - Mass Spectroscopy .....	310
6.8 Restoration of LOS core region in the <i>wlaA-T</i> large deletion mutant.....	310
6.8.1 Natural transformations with RM1221, NCTC11168 $\Delta$ <i>wlaP-SC</i> and NCTC11168 $\Delta$ <i>waaF</i> .....	310
6.9 Discussion .....	324
6.9.1 Establishing the mutagenesis strategy .....	324
6.9.2 Biological effect of the <i>wlaP-wlaSC</i> deletion mutant.....	326
6.9.2 Biological effect of the <i>wlaP-wlaSC</i> deletion mutant.....	327
6.9.3 Biological effect of the <i>wlaA-wlaT</i> deletion mutant.....	327
6.9.4 Conclusion .....	332
CHAPTER 7: GENERAL DISCUSSION .....	333
7.1 Diversity in LOS biosynthesis gene cluster .....	334
7.1.1 Variation in LOS core biosynthesis gene content.....	334
7.1.2 Role of LOS in <i>C. jejuni</i> pathogenesis.....	337
7.1.3 LOS cluster gene content and typing methods. ....	339
7.2 Minimal LOS gene content for viability .....	340
7.2.1 Deletion of <i>wlaP-wlaSC</i> .....	340
7.2.2 Deletion of <i>wlaA-wlaT</i> .....	341
7.3 Final Conclusions.....	346

APPENDIX 1.....	347
BIBLIOGRAPHY .....	357
APPENDIX 2.....	CD-ROM



## List of Figures

Figure 1.1: The <i>C. jejuni</i> glycome	31
Figure 1.2 a) Diagram of Gram-negative bacteria cell envelope b) Schematic representation of lipopolysaccharide and lipooligosaccharides	34
Figure 1.3: The genetic organisation of genes involved in core synthesis of <i>E. coli</i> .	37
Figure 1.4: Structure and biosynthetic enzymes of the lipopolysaccharide core oligosaccharides from <i>E. coli</i> .	39
Figure 1.5: Core structures from the <i>N. meningitidis</i> immunotypes L1-L8	42
Figure 1.6: Schematic diagram showing the core structures of <i>H. influenzae</i> .	45
Figure 1.7: Representational diagram of the general protein glycosylation and lipooligosaccharide core biosynthesis gene cluster in NCTC 11168	48
Figure 3.1: Genetic organisation of the LOS biosynthesis loci of several <i>C. jejuni</i> strains	86
Figure 3.2: Schematic diagram of the <i>lpxL</i> – <i>waaV</i> regions of <i>C. jejuni</i> 81116, 2F68, C11, O3, 21F185, and 13156	88
Figure 3.3: Schematic diagram of the region between <i>lpxL</i> and <i>waaV</i> in <i>C. jejuni</i> O73, RM1221 and LIO87	90
Figure 3.4: Schematic diagram showing the different LOS gene content arranged into an organisational classification based on gene content similarity	92
Figure 3.5 showing a schematic diagram of LOS core structures	97
Figure 3.6: Schematic diagram of LOS gene content groups indicating primer positions	104
Figure 3.7 Gel photograph showing the typical results of the LOS subclass PCR screen	109
Figure 3.8. Distribution of LOS gene content subclass in 47 strains.	111
Figure 3.9 Gel Photograph showing the results for strain 44406	113
Figure 4.1 Analysis of LOS by Tricine-SDS PAGE	154
Figure 4.2: A gel photograph showing the long range PCR amplicons produced with the unclassified clinical strains	157
Figure 4.3 Schematic diagram of the <i>lpxL</i> - <i>waaV</i> region of <i>C.jejuni</i> RM1221, 34086 and 51566	164
Figure 4.4 Schematic diagram of the <i>lpxL</i> - <i>waaV</i> region of <i>C. jejuni</i> RM1221, 60319 and 54386	185
Figure 4.5 Schematic diagram showing the organisational classification based for group 1	189
Figure 4.6 Schematic diagram showing the organisational classification based for group 2.	191
Figure 4.7 Schematic diagram showing the organisational classification based for groups 3 and 4	193

Figure 5.1 showing the scanned images produced after testing the arrayed oligonucleotides with different hybridisation solutions.	211
Figure 5.2 a, b and c showing the hybridisation of the control strains to group 1 specific oligonucleotides.	213-16
Figure 5.3 showing the hybridisation of the control strains to group 2 specific oligonucleotides	218-20
Figure 5.4 showing the hybridisation of the control strains to group 3 and 4 specific oligonucleotides.	221-24
Figure 5.5 indicating the hybridisation of the CSSS strains to Group 1a, 1b and 1c specific oligonucleotides	226-9
Figure 5.6 indicating the hybridisation of the CSSS strains to Group 2A and 2B specific oligonucleotides.	231-3
Figure 5.7 indicating the hybridisation of the CSSS strains to Group 3A, 3B and 4A specific oligonucleotides	234-7
Figure 5.8 indicating comparisons of A) the surface polysaccharide loci and B) surface appendages using the whole genome microarray and selected CSSS strains.	244-5
Figure 6.1. Representational diagram of the general protein glycosylation and lipooligosaccharide core biosynthesis gene cluster in NCTC 11168.	250
Figure 6.2: Structure and biosynthetic enzymes of the lipooligosaccharide core from <i>C. jejuni</i> NCTC 11168	252
Figure 6.3 Diagrammatic illustration of the defined deletion mutagenesis strategy employed by Hendrixson <i>et al.</i> , (2001).	259
Figure 6.4 Schematic diagram of pGLM8	262
Figure 6.5 Schematic diagrams detailing the final constructs of pGLM6 and pGLM7.	265-6
Figure 6.6 Schematic diagrams detailing the final constructs of pGLM9 and pGLM10.	269-70
Figure 6.7 Gel photograph showing the results of the PCR using <i>wlaPF</i> (XbaI) and <i>wlaSCR</i> (KpnI).	272
Figure 6.8. Gel photograph showing primer walking results on <i>C. jejuni</i> NCTC 11168 and the <i>wlaP-SC</i> deletion mutant.	274
Figure 6.9a shows the results of the Southern blot probed with the <i>wlaP-wlaSC</i> fragment	276
Figure 6.9b shows the results of a Southern blot after probing with a <i>neuBI-neuCI</i> fragment	276
Figure 6.10a. Analysis of LOS by tricine-SDS PAGE	279
Figure 6.10b Analysis of LOS by western blot with cholera toxin subunit B	279
Figure 6.11. A schematic diagram indicating the proposed structure of the LOS core for the <i>wlaP-wlaSC</i> deletion mutant and the wild-type NCTC 11168	281
Figure 6.12 schematic diagram showing the four separate components of the four	284

way ligation used to construct pGLM12 and pGLM13.

Figure 6.13 Schematic diagrams representative of the final plasmid constructs for pGLM12 and pGLM13. 287-8

Figure 6.14: Schematic diagrams of the final constructs for pGLM14 and pGLM15 290-1

Figure 6.15. Gel photograph showing the amplicons produced by *cj1130F* and *cj1153R* primers. 293

Figure 6.16. showing Southern blots to analyse the *wlaA-wlaT* large deletion mutant. 295

Figure 6.17. Analysis of LOS by Tricine-SDS PAGE. 298

Figure 6.18 Analysis of LOS by western blot with cholera toxin subunit B 298

Figure 6.19. A schematic diagram indicating the proposed structure of the LOS core for the *wlaA-wlaT* deletion mutant and the wild-type NCTC 11168 300

Figure 6.20 comparing the growth of the parent and mutant strains at 42°C over 48 hours 303

Figure 6.21 comparing the growth of the *wlaA-T* mutant strain in various antibiotics at 42°C over 48 hours 305

Figure 6.22 comparing the sensitivity of the deletion mutants to A) SDS, B) Polymyxin B and C) Novobiocin 313-5

Figure 6.23 Graph showing the results of *in vitro* invasion experiments with the parent and mutant strains 317

Figure 6.24 Coomassie blue stained SDS-PAGE of outer membrane preparations from the parent and mutant strains 319

Figure 6.25: CE-MS analysis of LOS from the *wlaP-SC* mutant 321

Figure 6.26: CE-MS analysis of LOS from the *wlaA-T* mutant 323

Figure 6.27 A) Gel photograph showing the presence and absence of *wlaJ* in a single colony. B) Schematic diagram postulating the mechanism of recombination which results in the loss of *wlaJ*. C) Schematic diagram showing the steps involved in the mutagenesis strategy with the imperfect inverted repeats. 326

Figure 6.28 Schematic representation of the *C. jejuni* 81-176 and NCTC 11168 capsule structure 331

Figure 7.1 Graph comparing the growth of NCTC 11168 and the *wlaA-T* large deletion mutant in iron replete and iron limited conditions 344

## List of Tables

Table 1.1 Comparison of the genomes of <i>C. jejuni</i> NCTC 11168, RM1221 and 81-176 and <i>E. coli</i> K12	27
Table 2.1: Details of <i>C. jejuni</i> strains used in this study	52
Table 2.2, detailing the supplements used, their stock and working concentrations	54
Table 2.3 detailing the components of a standard PCR reaction and their final concentrations.	65
Table 2.4 showing a typical PCR cycling profile.	65
Table 2.5 showing the composition of a TripleMaster mix PCR reaction and the final concentration of the components	66
Table 2.6 detailing the cycling profile used for a TripleMaster mix PCR.	66
Table 2.7 detailing the composition of a Masteramp extra long PCR.	68
Table 2.8 detailing a typical long-range PCR cycling profile.	68
Table 2.9 showing the composition of Tricine - SDS PAGE gels	74
Table 2.10 showing the composition of SDS-PAGE for separation of proteins	78
Table 2.11 showing the concentrations used in assessing sensitivity to antibiotics and detergents	81
Table 3.1: Amplicon sizes (bp) expected for each LOS sub-class	102
Table 3.2: Summary of PCR screen results for each strain	107
Table 3.3 showing the MLST results for each CSSS strain	117-8
Table 4.1 a and b, summarising the primer walking results with primers corresponding to group 1C	127-8
Table 4.2 summarising the primer walking results with primers corresponding to group 1A	131
Table 4.3 summarises the primer walking results with primers corresponding to group 1B	134
Table 4.4 a and b, summarising the primer walking results with primers corresponding to group 2A	137-8
Table 4.5 a and b, summarising the primer walking results with primers corresponding to group 2B	139-40
Table 4.6 summarises the primer walking results with primers corresponding to group 3A,	144
Table 4.7 summarises the primer walking results with primers corresponding to group 3B	146
Table 4.8 summarises the primer walking results with primers corresponding to group 4A	148
Table 4.9 summarising the LOS group for each strain after primer walking.	151
Table 4.10 a) summarises the size of long range amplicons produced for each strain and b) summarises the expected size for each sub-group control	157

Table 4.11 Summary of the sequencing stages and fold coverage for each strain.	161
Table 4.12 a) and b). Summary of identity comparisons with RM1221 and ORF organisation in strains 34086 and 51566	166-7
Tables 4.13 A-D Summary of nucleotide identity for each strain	170-1
Tables 4.14 A-H Summary of nucleotide identity for each strain	175-6
Table 4.15 Summary of the LOS group assigned to each strain	180
Table 5.1 summarises the hybridisation results from the first generation locus specific microarray	206-7
Table 5.2 summarises the hybridisation of each CSSS strain to the group specific oligonucleotides	243
Table 6.1 detailing the final working concentrations of each treatment used in the assaying the sensitivity of the parent and mutant strains of <i>C. jejuni</i> .	308

## List of Abbreviations

A	Adenine
A (mA)	Ampère
Ac	O-acetyl substitution
ATP	Adenosine triphosphate
bp	base pairs
BSA	Bovine serum albumin
c	Centi ( $10^{-2}$ )
°C	Degrees Celsius
C	Cytosine
cfu	Colony forming unit
CDT	Cytolethal distending toxin
CE-MS	Capillary Electrophoresis Mass Spectroscopy
CTAB	Hexdecyltrimethyl ammonium bromide
Da	Dalton
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
EDTA	Disodium ethylenediaminetetraacetate
EtBr	Ethidium bromide
<i>g</i>	Angular velocity as multiples of the acceleration due to gravity ( $g = 9.8 \text{ m/s}^2$ )
g	Grams
G	Guanine
Gal	Galactose
GalNAc	N- acetylgalactosamine
Glc	Glucose
GlcNAc	N-acetylglucosamine
GBS	Guillain Barré syndrome
h	Hour(s)
H	Hexose
Hep	Heptose
IPTG	Isopropylthio- $\beta$ -D-galactoside
K	Kilo ( $10^3$ )
Kb	Kilobase pairs
kDa	Kilo Daltons
Kdo	3-deoxy- $\alpha$ -D-manno-oct-2-ulopyranosonic acid
l	Litres
LA	Luria Bertani agar



LB	Luria Bertani broth
LOS	Lipooligosaccharides
LPS	Lipopolysaccharides
$\mu$	Micro ( $10^{-6}$ )
m	Metre
m	Milli ( $10^{-3}$ )
M	Mole(s)
MHA	Mueller Hinton Agar
MHB	Mueller Hinton Broth
min	Minute(s)
MFS	Miller Fisher Syndrome
n	Nano ( $10^{-9}$ )
NeuNAc	N-acetylneuraminic acid
$\Omega$	Ohm (s)
OD <sub>600</sub>	Optical density at a wave length of 600 nm
%	Percent
p	Pico ( $10^{-12}$ )
P	Phosphate group
Pcho	Phosphocholine
PEtN	Phosphoethanolamine
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCR	polymerase chain reaction
s	Second (s)
SDS	Sodium lauryl sulphate
SNP	Single nucleotide polymorphism
SSC	Saline Sodium Citrate buffer
T	Thymine
T <sub>m</sub>	Melting temperature
UV	Ultraviolet
v/v	Volume per volume
V	Volts
VAIN	(Variable atmosphere incubator)
W	Watts
w/v	Weight per volume
X-Gal	5-Bromo-4-chloro-3-indol- $\beta$ -D-galactoside

## Chapter 1: Introduction

### 1.1 *Campylobacter jejuni*

*Campylobacter jejuni* is a causative agent of gastroenteritis in humans and animals, such as dogs and cats, and can also cause a range of disease including gastroenteritis, septicaemia, meningitis and abortion (Young and Mansfield, 2005). The most common clinical presentation in humans with campylobacteriosis is gastroenteritis and *Campylobacter* causes the majority of food-borne bacterial gastroenteritis in the developed world (Allos, 2001). *C. jejuni* enteritis has also been implicated as the most frequent antecedent infection prior to the development of the paralytic neuromuscular disorder, Guillain Barré syndrome. Although *Campylobacter* remains a significant cause of disease and presents a major health burden, public awareness of this bacterium remains limited and is not equivalent to that of other bacterial pathogens.

*Campylobacter jejuni* biology, significance and genomics will be examined in this chapter. Further discussion will be extended to lipopolysaccharides, their significance in other bacteria and the extent of the variation of these structures in other bacteria. A detailed examination of the *C. jejuni* lipooligosaccharide gene biosynthesis cluster and the extent of variation in this cluster will be examined at the beginning of Chapter 3 and Chapter 6. Molecular typing techniques and microarray technology will be examined in Chapter 4.

#### 1.1.1 Significance

As the most frequently reported cause of bacterial gastroenteritis in the developed world, infection by campylobacters exact a major public health and economic burden (Ketley, 1997). This is also the case in the developing world

were campylobacteriosis is a significant cause of infant mortality (Ketley, 1997). The majority of these infections are caused by the bacterium *C. jejuni*, and although most cases are thought to be unreported, the incidence of campylobacteriosis far outweighs that caused by *Salmonella* spp. (Frost, 2001). The number of cases of campylobacteriosis between 1992 and 2000 were estimated at over 450,000, with 43,000 cases being confirmed in the laboratory (Adak, 2002). This study also attributed over 134,000 GP visits and over 400,000 hospital bed days to infection by campylobacters. *C. jejuni* infection is the most frequent antecedent to Guillain-Barré syndrome, a form of neuromuscular paralysis (Allos, 1997). Other *Campylobacter* species such as *C. coli*, *C. hyoilei*, *C. helveticus*, *C. lari* and *C. upsaliensis* are also enteric organisms, but several other *Campylobacter* species appear to be niche specific, such as *C. fetus* which is associated with sporadic abortion in cattle and sheep and *C. concisus*, *C. curvus* and *C. rectus* which are associated with the human periodontal cavity (Gilbert *et al.*, 2005).

The mechanisms by which *Campylobacter* cause disease are poorly understood, partly due to the lack of an animal model, recalcitrance to traditional genetic manipulation and the lack of 'model' virulence mechanisms such as those found in *E. coli* and *Salmonella* spp.

### 1.1.2 *Campylobacter* biology

The genus *Campylobacter* contains 17 species, six subspecies and several biovars (On, 2005; [www.bacterio.cict.fr/c/campylobacter.html#r](http://www.bacterio.cict.fr/c/campylobacter.html#r)). The campylobacters are spiral Gram-negative rods, which possess polar flagella to aid motility and high motility is a characteristic of fresh clinical isolates. *C. jejuni* and *C. coli* are microaerophilic and 'thermophilic', requiring an oxygen concentration of 3 - 15%, a carbon dioxide concentration of 3-5% and growth temperatures between 34 - 44°C. Many *Campylobacter* species are present in the normal gut flora of wild and farmed birds and animals. Their optimal growth temperature of 42°C is thought to be an adaptation to their natural avian habitat.

### 1.1.3 Transmission and Epidemiology

The incidence of *Campylobacter* infection in the UK has been increasing since the early 1980's (Frost, 2001) and approximately 1% of the population of the USA contract campylobacteriosis per annum (Friedman *et al.*, 2000). The estimated number of laboratory reported cases of *Campylobacter* infection in England and Wales for 2006 currently stands at over 45,000 ([www.hpa.org.uk/infectious/topics\\_az/campy/data\\_uk2.htm](http://www.hpa.org.uk/infectious/topics_az/campy/data_uk2.htm)). The majority of cases are sporadic and show seasonality, with an increase in May and September and although this was postulated to be due to increased poultry consumption, no temporal association was found, indicating that the seasonal rise was due to a common but unidentified environmental source. Infection has been associated with contaminated water, untreated milk and contaminated food, especially undercooked or raw poultry. It has proven difficult to determine sources of infection as *Campylobacter* is ubiquitous in the environment and has a weakly clonal population structure (Frost, 2001). Skirrow (1977) and Adak *et al.* (2005) linked increasing foodborne disease, including campylobacteriosis to chicken contamination. Determining the source requires standard typing methods and markers (Frost, 2001).

Phenotypic methods such as the Penner and Lior serotyping schemes were developed in the 1980s to enable epidemiological characterisation of *Campylobacter* isolates. Penner serotypes are based on heat – stable antigens and the principal sero-determinant was shown to be the capsule polysaccharide (Karlyshev *et al.*, 2000). Lior serotypes are based on heat labile antigens, but to date the principal sero-determinant has not been isolated. Common serotypes exist among human infection - associated isolates such as HS1, HS2 and HS4 (Fitzgerald *et al.*, 2001). Over 65 Penner serotypes exist for *C. coli* and *C. jejuni* but some strains remain untypeable or are assigned multiple serotypes due to cross-reactivity (Miller and Mandrell, 2005). The Penner serotyping scheme was adapted to enable direct agglutination (Frost *et al.*, 1998) and although this method provides further discrimination, some strains remain untypeable and variation within serotypes is not detectable.

An EC-funded project, CampynetIII, has been established to ascertain the best practice for the commonly used molecular typing techniques and provide a database ([www.medvetnet.org/cms/templates/docphp?id-65](http://www.medvetnet.org/cms/templates/docphp?id-65)). These methods include Pulsed Field Gel Electrophoresis (PFGE), Amplified Fragment Length Polymorphism (AFLP), flagellin gene restriction fragment length polymorphism (*fla* typing; Ribot *et al.*, 2001; Ahmed *et al.*, 2002; Boer Pd *et al.*, 2002) and Multilocus Sequence Typing (MLST; Maiden *et al.*, 1998). MLST is a genotyping system based on the sequencing of 500 bp fragments of seven housekeeping genes. Distinct alleles are assigned a number and the seven numbers assigned to the housekeeping genes for each isolate form an allelic profile called a sequence type (Dingle *et al.*, 2005). MLST has a number of advantages over other methods as it is unaffected by changes in gene order, it is reproducible among laboratories and the data is easily stored and accessible (<http://pubmlst.org/campylobacter>).

The *Campylobacter* Sentinel Surveillance Scheme also served to gather information about laboratory reported cases of campylobacteriosis regarding exposure and illness onset (Gillespie *et al.*, 2002; CSSS 2005). In an earlier study this scheme also showed links between ethnic groups and propensity to develop campylobacteriosis (CSSS 2003)

### 1.1.4 Clinical Presentation

Ingestion of as few as 500 bacteria can cause campylobacteriosis and symptoms develop between 18 hours to 8 days post-ingestion. The clinical presentation of campylobacteriosis can range from severe, inflammatory diarrhoea to mild, non-inflammatory, watery diarrhoea in both western and developing countries. This range of symptoms is indistinguishable from disease caused by other enteric pathogens, such as *Salmonella* and *Shigella*. Usually symptoms are characterised by acute severe abdominal cramps, fever and diarrhoea. *C. jejuni* induced enteritis is usually self-limiting and lasts 5-7 days (Young and Mansfield, 2005).

Rare complications from *Campylobacter* infection may occur, including peritonitis, cholecystitis and gastrointestinal haemorrhage (Allos, 2001). These complications and patients that develop invasive disease such as

bacteraemia and sepsis are usually only seen in the very young, the very old and those with compromised immune systems. The neuromuscular disorder, Guillain-Barré syndrome (GBS) is a notable complication associated with *Campylobacter* infection. There are approximately 1.3 cases of GBS per 100,000 per year (Hughes and Rees, 1997) and two-thirds of these cases develop post-infection. The link between GBS and prior *C. jejuni* infection has been noted via subjective evidence, serological and culture data. Patients with GBS have been shown to exhibit high levels of antibodies to *C. jejuni*. This, together with culture studies, has shown that in 30-40% of GBS cases, patients have had a recent *C. jejuni* infection before the onset of their neurological symptoms (Allos, 2001; Nachamkin, 2002).

GBS is characterised by a rapid progressive ascending symmetrical weakness, starting in the lower limbs, and sensory disturbance. Two main forms of the disease occur; acute inflammatory demyelinating polyneuropathy (AIDP), which involves the destruction of Schwann cells and acute motor axonal neuropathy (AMAN), where antibodies bind to the nodal axonemmma, leading to degeneration of sensory and motor nerves (Hadden and Gregson, 2001). The more severe form, AMAN, is usually associated with *C. jejuni* infection (Koga *et al.*, 2001). Miller Fisher syndrome is a rare variant of GBS characterised by ophthalmoplegia, limb ataxia and areflexia without limb weakness.

GBS is thought to be an autoimmune disease and it is postulated that the immune response to *Campylobacter* infection generates antibodies that cross-react with neural antigens leading to disease development (Yuki *et al.*, 1993). Monoclonal antibodies raised against GBS-associated *C. jejuni* lipooligosaccharides were shown to be reactive with neuronal gangliosides, indicating that lipooligosaccharides are the bacterial epitope involved in the cross-reaction (Yuki *et al.*, 1992; Goodyear *et al.*, 1999). Cholera toxin binds to certain gangliosides and has also been shown to bind to *C. jejuni* lipooligosaccharides (Yuki, 1993). The prognosis of GBS patients is variable with one third of patients having fully recovered a year on from the disease onset, but between 2 - 18% of patients having died (Hadden and Gregson, 2001). Where *C. jejuni* has caused the antecedent infection, GBS patients have been shown to have a poorer prognosis and this is thought to be



because molecular mimicry results in greater axonal degeneration (Hadden and Gregson, 2001).

### 1.1.5 Pathogenesis

The mechanisms by which *C. jejuni* causes disease are unknown and undoubtedly multi-factorial. Upon ingestion *Campylobacter* pass through the stomach, colonise the mucus layer and adhere to cell surfaces in the ileum and colon. A number of factors are involved which cause inflammation, invasion and tissue damage, leading to the disruption of normal intestinal function and diarrhoea (Ketley, 1997). The nature of the interactions between the *Campylobacter* cell and the host remain unclear.

Bacterial pathogens such as *Salmonella* spp., *Shigella* spp. and enteropathogenic *E.coli* develop specific interactions with the host mucosal surface that exploit the host cell machinery and cause disease. In many enteric bacteria these virulence determinants can be either chromosomally located on pathogenicity islands (PAI) or present on virulence plasmids. In *Shigella* spp., the type III secretion system (TTSS), required for invasion in epithelial cells and translocation of effector proteins across the host cell membrane, is located on a large virulence plasmid (Schmidt and Hensel, 2004). Five pathogenicity islands have also been identified in *Shigella* spp. and these are involved in other aspects of disease, such as avoiding host defence systems, colonisation and producing toxin. Both plasmid genes and PAI are required for *Shigella* to cause all possible disease phenotypes. The genes required for invasion by *Salmonella enterica* (*S. enterica*) are clustered in a PAI named SPI-1 (*Salmonella* Pathogenicity Island), which encodes a TTSS, which is required for invasion of non-phagocytic cells. Another *Salmonella* pathogenicity island, SPI-2 is needed for the ability to cause systemic infections and proliferate within host organs.

Both plasmid and chromosomally-borne virulence genes involved in invasion are present in *C. jejuni*, but the sequencing of *C. jejuni* has revealed that 'traditional' PAIs do not exist in this bacterium. Plasmids are present in 19-53% of *C. jejuni* strains and these were generally thought not to be

associated with virulence (Bacon *et al.*, 2000). Yet a diarrhoeal outbreak strain, 81-176, which has been shown to invade at a higher level *in vitro* (Bacon *et al.*, 2000) contains two plasmids. One plasmid encodes a *tetO* gene and a second plasmid, pVir, has four open reading frames (ORFs) with sequence similarity to *Helicobacter pylori* proteins. One of these proteins is part of the *cag* pathogenicity island, which has been shown to induce host cell changes. The plasmid also contains an operon of three *comB* genes, similar to that seen in *H. pylori* isolates. Mutation of these genes changed the natural transformation ability of 81-176. The four genes also show homology to proteins of type IV secretion systems, and mutation of these genes results in decreased adherence and internalisation *in vitro* (Bacon *et al.*, 2000).

Adherence and invasion are known to cause the inflammatory response that leads to diarrhoea, but it is unknown in *Campylobacter* infections whether diarrhoea is caused because the bacterium interferes with the normal secretory or absorptive capacity of intestine. This alteration in the activity of the intestine may be due to the activity of bacterial toxins. Toxins have also been implicated in the cytopathic effects caused by bacterial infection. *E. coli*, *Shigella*, and some *C. jejuni* strains produce toxins that cause some *in vitro* cell lines to distend and die (Johnson and Lior, 1988). In *C. jejuni*, the only toxin genes shown by genome sequencing were *cdtA*, *cdtB* and *cdtC*, which encode cytolethal-distending toxin (CDT). CDT is postulated to be involved in disruption of intestinal epithelial cell function leading to diarrhoea and has been shown to cause cell cycle arrest, possibly due to nuclease activity (Whitehouse *et al.*, 1998; Pickett and Lee, 2005). At present it is unclear whether epithelial damage causes diarrhoea due to disruption of cell function or *vice versa*.

The ability of *C. jejuni* to invade human cell lines has been shown *in vitro* and involves *C. jejuni* proteins and host cell signal transduction (Szymanski *et al.*, 1995; Konkel *et al.*, 1999). The process of adherence and invasion is important in the development of campylobacteriosis and motility and chemotaxis have an important role in adherence. Chemotaxis is the ability to move up or down chemical gradients which aids penetration and colonisation of the mucus layer (Marchant *et al.*, 2002). Chemotaxis and motility are inextricably linked and aflagellate non-motile *Campylobacter* cells

are unable to invade and have impaired adherence (Wassenaar *et al.*, 1991; Yao *et al.*, 1994). Clinical isolates are also found to be highly motile. Other bacterial effectors are important in adhesion and invasion; and potential adhesins include PEB1, CadF, lipooligosaccharides, the capsule and JlpA. The PEB1, JlpA and CadF proteins, are involved in internalisation of the bacterium (Pei *et al.*, 1998; Jin *et al.*, 2001; van Vliet and Ketley 2001).

Further experiments that co-cultured *C. jejuni* with epithelial cells showed that the bacterium synthesised a set of proteins and those that were secreted were named *Campylobacter* invasion antigens or Cia proteins (Konkel *et al.*, 1999). These proteins have an unknown function, but insertional mutation of *ciaB* causes a significant reduction in the number of bacteria internalised, which is restored by *trans* complementation. Further work by Konkel *et al.*, has also shown that these proteins require a functional flagellar type III secretion system to be secreted (Konkel *et al.*, 2004).

The ability to regulate iron acquisition and acquire iron in areas of limited availability is also essential in the bacterial adaptation to the host environment. Unlike many other bacterial pathogens *C. jejuni* does not produce siderophores, but does commit siderophore piracy and uses those produced by other bacteria. Several iron uptake systems have been identified in *Campylobacter* (van Vliet *et al.*, 2002; Ridley *et al.*, 2006).

The importance of polysaccharide biosynthesis shown by the presence of three large clusters in the genome may indicate that these structures have an important role in the bacterium and they are also hypothesised to be involved in avoiding host defence mechanisms and the induction of an inflammatory response. In other bacteria, the capsule polysaccharide maybe involved in resistance to both specific and non-specific host immune response (Karlyshev *et al.*, 2001b).

The majority of animals that are colonised with *C. jejuni* remain disease free and acquired immunity against *C. jejuni* is developed by poultry workers and children in developing countries. This information suggests that *C. jejuni* may be recognised differently by immune system of humans and animals that develop disease. Another explanation could be that toxins and virulence determinants do not affect the animal host causing less disruption in the gut and no clinical disease. The innate ability of this organism to vary its surface

polysaccharides may contribute to its interactions with the host immune system.

### 1.1.6 *Campylobacter* genomics

To date three complete and five incomplete whole genome sequences of *C. jejuni* strains are available; those completed include NCTC 11168 (Parkhill *et al.*, 2000), RM1221 (Fouts *et al.*, 2005) and 81-176 (Hofreuter *et al.*, 2006); [www.tigr.org](http://www.tigr.org). Table 1.1 shows a comparison between the genome sequences of these 3 strains. Hofreuter *et al.* (2006) described synteny between the three strains with anomalies reflecting the insertion, deletion and duplication of sequences, especially the hypervariable loci, such as the capsule region and the LOS cluster.

The genome sequence of *C. jejuni* NCTC 11168 was published in 2000 and recently re-annotated ([www.lshtm.ac.uk/pmbu/cfu](http://www.lshtm.ac.uk/pmbu/cfu)), revealing 1654 predicted coding sequences and an average G+C content of 30.6% (Parkhill *et al.*, 2000). The genome is 1,641,481 base pairs in size, which is considerably smaller than other enteric organisms such as *E. coli* and *Salmonella* (Blattner *et al.*, 1997; Perna *et al.*, 2001). Approximately 94.3% of the genome encodes for proteins, making it a dense bacterial genome, and gene function is predicted for 77.8% for the coding sequences. Unlike other sequenced bacterial genomes, for example *E. coli*, there is an almost complete absence of repetitive DNA, such as insertion sequences and transposons, in the *C. jejuni* genome.

The NCTC 11168 genome sequence showed that the polysaccharide biosynthesis loci are arranged in clusters. Eight percent of the genome is involved in polysaccharide synthesis in this organism showing its importance. These include regions for flagella glycosylation, general protein glycosylation and lipooligosaccharide and capsule biosynthesis. The capsule biosynthesis and lipooligosaccharide biosynthesis loci were shown to have a percentage G+C content lower than the genome average suggesting these regions may be subject to genetic exchange. The presence of variable G/C

mononucleotide runs was also a notable discovery. The thirty homopolymeric tracts are associated with phase variable phenotypes, and are especially found in genes associated with the biosynthesis or modification of surface molecules, for example lipooligosaccharide biosynthesis, capsular biosynthesis and flagellin genes.

Linton *et al.* (2000b) has shown that homopolymeric tracts are directly involved in LOS expression. Other examples of phase variation in gram negative bacteria exist, such as Lewis antigens in *H. pylori* and lipooligosaccharides of *Neisseria* spp. (Appelmek *et al.*, 1998; De Bolle *et al.*, 2000). Although the availability of the genome sequence has provided invaluable information, further research is needed for conclusions to be drawn about the mechanisms of pathogenesis in *C. jejuni*.

**Table 1.1 Comparison of the genomes of *C. jejuni* NCTC 11168, RM1221 and 81-176 and *E. coli* K12.**(Blattner *et al.*,1997;Parkhill *et al.*, 2000; Fouts *et al.*, 2005; Hofreuter *et al.*, 2006; [www.tigr.org](http://www.tigr.org)). NA is not available.



Features	Strains			
	<i>E.coli</i> K12	<i>C. jejuni</i> NCTC 11168	<i>C. jejuni</i> RM1221	<i>C. jejuni</i> 81-176
Size/bp	4,639,221	1,641,481	1,777,831	1,616,554
%G+C content	50.8	30.6	30.3	30
No. of Coding sequences	4288	1654	1884	1653
% sequence encoding proteins	87.8	94	94	93
% function predicted	62	77.8	60	NA

### 1.1.6 The *Campylobacter* glycome.

The importance of polysaccharide biosynthesis shown by the presence of three large clusters in the genome may indicate that these structures have a vital role in the lifestyle of *C. jejuni*. These clusters form over 8% of the genome and are; the capsule gene cluster, O-linked (flagella) protein glycosylation, N-linked protein glycosylation and the lipooligosaccharide gene cluster. Figure 1.1 shows a schematic diagram of the interaction between the bacterial cell membrane and the glycome.

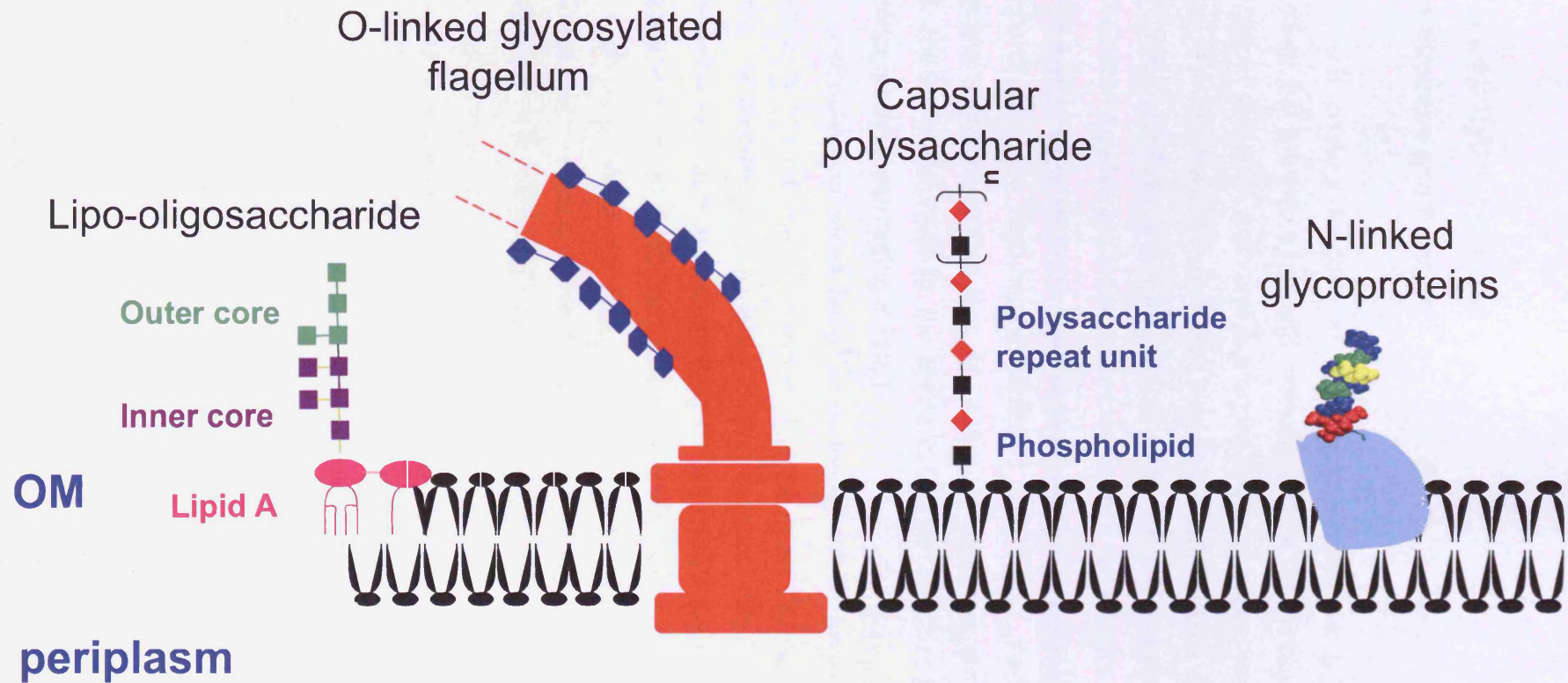
The post-translational modification of *C. jejuni* flagella was first reported in the late 1980s (reviewed in Karlyshev *et al.*, 2005b) and later confirmed by Doig *et al.* (1996). The flagellar glycosylation cluster consists of approximately 50 genes, including those encoding structural proteins and glycan biosynthesis enzymes. This region also contains multiple copies of the *neuA*, *neuC* and *neuB* genes which were thought to be involved in sialic acid biosynthesis (Parkhill *et al.*, 2000) but the flagellum structure has been shown to contain modified pseudaminic acid residues. The modification of these residues is important for flagellar assembly as mutations in the genes involved in pseudaminic acid biosynthesis result in aflagellate and non-motile cells (Linton *et al.*, 2000a; Goon *et al.*, 2003; Guerry *et al.*, 2006). Inter-strain variation occurs at this locus and can be caused by the insertion or deletion of genes and the presence of homopolymeric tracts (Karlyshev *et al.*, 2002; Guerry *et al.*, 2006).

The *wla* or *pgl* gene cluster was originally thought to be involved in lipopolysaccharide biosynthesis, which is located alongside (Fry *et al.*, 1998; Wood *et al.*, 1999), but was subsequently shown to be involved in protein glycosylation and mutations in these genes changed protein antigenicity (Szymanski *et al.*, 1999). This is the first example of a N-linked glycosylation in Gram-negative bacteria. The locus is approximately 17 kb in size and appears to be highly conserved among strains (Szymanski *et al.*, 2003b). These gene products are involved in the addition bacillosamine-containing residues on to at least 30 proteins e.g. PEB3 and CgpA (Linton *et al.*, 2002). The functions of various gene products have been elucidated for

example, PglB is a N-linked oligosaccharyltransferase (Szymanski *et al.*, 1999; Wacker *et al.*, 2002) and PglH is an  $\alpha$  1,4 GalNAc transferase. Mutations in *pglH* cause a decreased ability to adhere and invade *in vitro*, and a decreased ability to colonise chicks *in vivo* (Hendrixson and DiRita, 2004; Karlyshev *et al.*, 2004) it was originally proposed that the protein glycosylation system had a role in avoiding host defence mechanisms by masking primary amino acid structure (Szymanski *et al.*, 1999), but as this cluster appears to be highly conserved, it is now proposed that this locus has a role in adherence and pathogenesis via glycosylation of other proteins (Karlyshev *et al.*, 2004).

The sequencing of NCTC 11168 led to the discovery of a large gene cluster encoding genes with homology to type II/III capsular polysaccharide related genes (Parkhill *et al.*, 2000). Further work then detected and characterised the capsule (Karlyshev *et al.*, 2000; 2001a and 2001b; Bacon *et al.*, 2001) and determined the capsule structure (St. Michael *et al.*, 2002; Szymanski *et al.*, 2003a; McNally *et al.*, 2005). The capsule was found to be the major sero-determinant of the Penner serotyping scheme. Historically *Campylobacter* was thought to express lipooligosaccharides and lipopolysaccharides and the *kps* cluster was found to be responsible for biosynthesis of the higher molecular weight polysaccharide that was detected alongside LOS (Karlyshev, *et al.*, 2000; Oldfield *et al.*, 2002). The capsule (CPS) gene cluster in NCTC 11168 is composed of 38 genes and contains 3 major regions. The highly variable central region is comprised of the genes required to construct the repeating units of the CPS. The variability of this region is reflected by the wealth of different serotypes. The two flanking regions are more conserved and contain the genes required for the transport of the CPS to the cell surface and polymerisation of the repeating units (Karlyshev *et al.*, 2000; 2005a). The genes involved in heptose biosynthesis are especially conserved among strains (Dorrell *et al.*, 2001) but insertions, deletions and homopolymeric tracts all serve to cause variation in this locus. The ability to vary the CPS may be important in host immune evasion and virulence. A 81-176 CPS mutant was shown to be less adherent and invasive *in vitro* indicating the importance of the capsule in pathogenesis (Bacon *et al.*, 2001).

**Figure 1.1: The *C.jejuni* glycome** showing the four polysaccharide antigens that are produced by the *C. jejuni* cell and their relationship with the outer membrane (Karlyshev *et al.*, 2005).



## **1.2 Lipopolysaccharides**

### **1.2.1 Structure and function**

The primary function of the lipopolysaccharide (LPS) molecule is structural, but it is also a powerful stimulator of the host immune response and is a potent virulence toxin that is thought to play a critical role in the initiation of the proinflammatory events that lead to bacterial sepsis or infection (Tong *et al.*, 2001). LPS is a major surface constituent forming approximately 70% of most Gram-negative bacterial outer membranes (see figure 1.2a). A lipid A moiety anchors the molecule into the outer membrane, this is linked to a core oligosaccharide via keto-deoxyoctularonic acid (Kdo), and an O-chain of repeating oligosaccharides (see fig. 1.2b). In Gram-negative bacteria, such as *E. coli* and *Salmonella* sp, the core and O-chain residues provide the basis for antigen-based serotyping methods.

Lipopolysaccharides have an endotoxic effect, and lipid A has been shown to be the biologically active part of the molecule, causing the majority of the inflammation associated with bacterial infections. The core oligosaccharide maintains cell integrity by providing a physical barrier against substances such as antibiotics and detergents (Raetz and Whitfield, 2002). The O-chain provides the bacterium with resistance to the host's immune response, but it is not an essential part of the molecule and in species where the O-chain is missing, e.g. *Neisseria* spp. and *Haemophilus influenzae*, this truncated form is named lipooligosaccharide. Conversely lipid A has been shown to be an essential part of the cell for viability in the majority of Gram-negative bacteria (Schnaitman and Klena, 1993; Steeghs *et al.*, 1998).

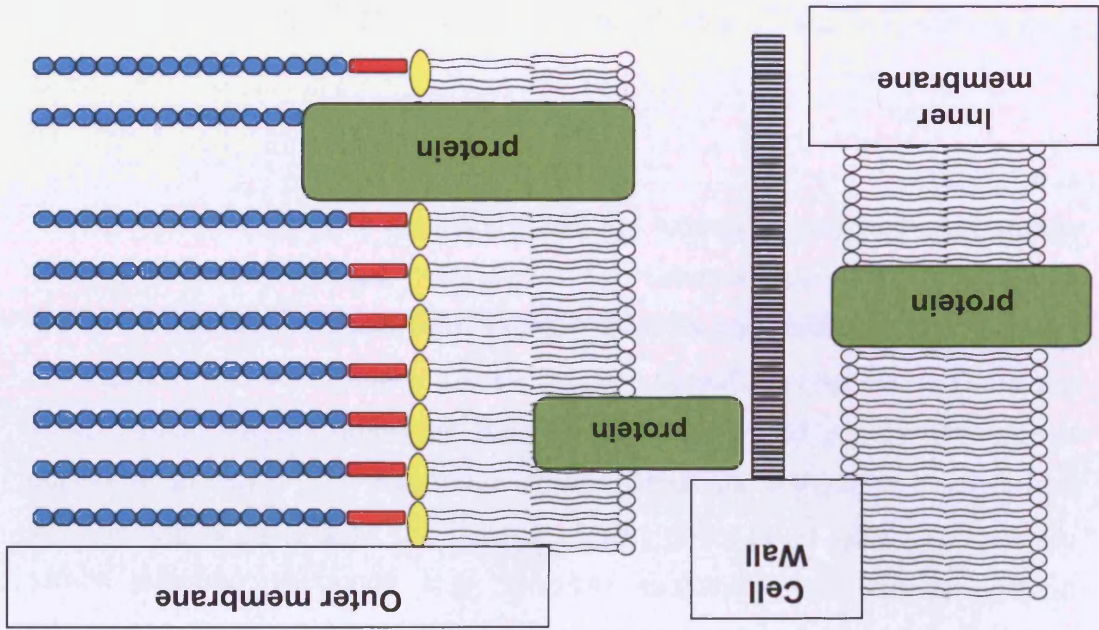
Lipopolysaccharides (LPS) are important components of many Gram-negative bacterial membranes and have been linked with a range of virulence associated roles. These epitopes also have an important role in stimulating the host immune defences such as Toll-like receptors. LPS acts as a toxin, triggering a cascade of immune responses that can lead to inflammation and

**Figure 1.2 a) Diagram of Gram-negative bacteria cell envelope**

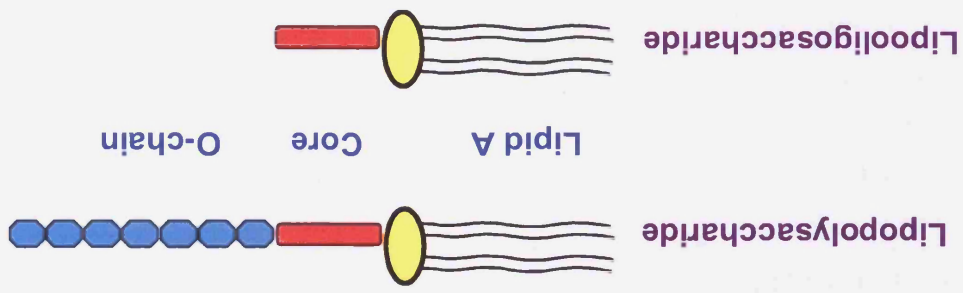
This model shows the location of lipopolysaccharide within the cell. LPS is present in the outer membrane of the bacterial cell envelope. It occurs in a leaflet of phospholipids with peripheral and integral proteins, and comprises of ~75% of the outermost membrane (Gronow *et al.*, 2001)

**b) Schematic representation of lipopolysaccharide and lipooligosaccharide**, composed of a lipid A moiety, core oligosaccharides and O-chain oligosaccharide.

A)



B)





sepsis (Raetz and Whitfield, 2002). In some bacteria, such as *Pseudomonas aeruginosa*, LPS has a more direct role in virulence, acting as a non-pilus adhesin, adhering to the cystic fibrosis transmembrane conductance regulator via the core polysaccharide, allowing invasion of the lung cells (Pier, 2000).

In *Neisseria* spp. LPS is used to enable the bacterium to avoid the host defence mechanisms as it removes host sialic acid from the blood stream and adds it to the LPS galactose residues, this enhances bacterial serum resistance (Vogel and Frosch 1999). In *H. pylori*, LPS plays a different role and does not induce an inflammatory response. The LPS of *H. pylori* expresses Lewis X and Y blood group antigens which are identical to those occurring in the human gastric mucosa as a possible example of molecular mimicry for the avoidance of host defence mechanisms (Appelmek *et al.*, 1997). *H. pylori* also express diverse antigens among strains which have been implicated in colonisation, persistent infection and pathogenesis (Wang *et al.*, 2000).

### 1.2.2 Diversity in the core of bacterial lipopolysaccharides

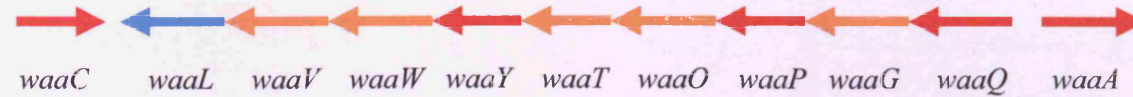
In the *Enterobacteriaceae* the *waa* locus contains all the genes required for the assembly of the core oligosaccharide in a sequential manner. The locus contains three operons; the *gmhD* operon, composed of the genes required for inner core assembly, the *waaQ* operon which encodes glycosyltransferases for the addition of sugar residues and an operon containing the *waaA* and *waaC* genes.

It is well established in *E. coli* and *S. enterica* that the core oligosaccharide structure reflects the genes present in individual strains (Heinrichs *et al.*, 1998b). The majority of structural variation in these bacteria occurs in the O-antigen, with ~170 distinct epitopes reported just for *E. coli*. *E. coli* expresses five different core oligosaccharide structures; K-12 and R1-R4 (see figure 1.3 for cluster gene content and figure 1.4 for LPS structure; Heinrichs *et al.*, 1998; Amor *et al.*, 2000). Core oligosaccharides structures

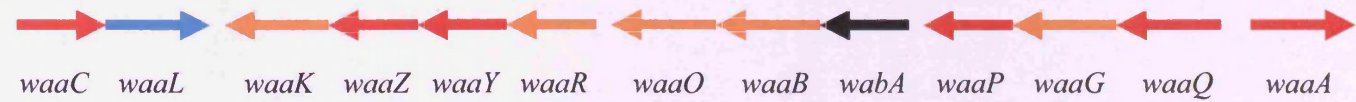
**Figure 1.3: The genetic organisation of genes involved in core synthesis of *E. coli*.**

Adapted from Heinrichs *et al.* (1998). Arrows represent open reading frames, and indicate the direction of transcription. The colour of each arrow indicates at which stage of the core biosynthesis pathway the gene product is involved; red, inner core assembly; orange, outer core assembly; blue, ligation of O-chain and black, unknown. Sequence data upstream of *waaC* in the *E. coli* R 1- 4 strains is unavailable. This diagram is not shown to scale.

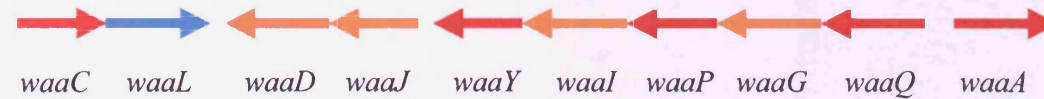
a) *E. coli* R1



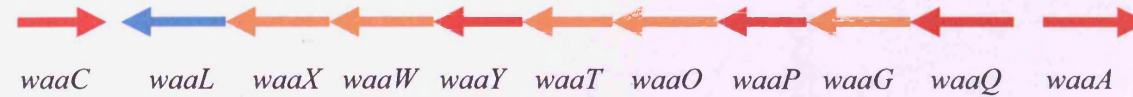
b) *E. coli* R2



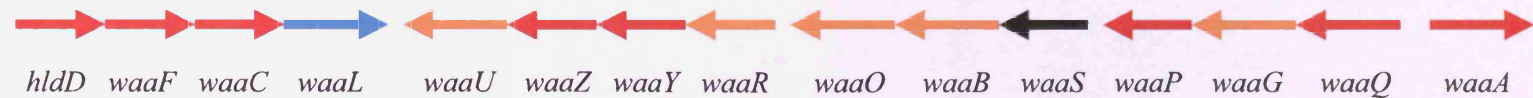
c) *E. coli* R3



d) *E. coli* R4



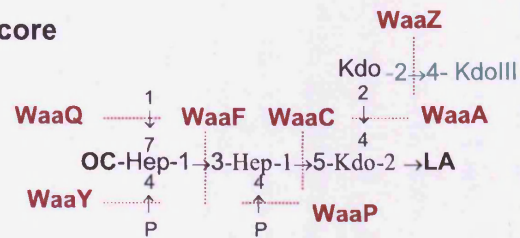
e) *E. coli* K12



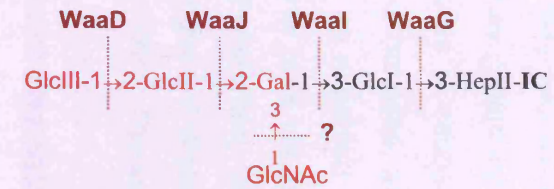
**Figure 1.4: Structure and biosynthetic enzymes of the lipopolysaccharide core oligosaccharides from *E. coli*.**

Figure adapted from Raetz and Whitfield (2002). The conserved core structure is shown in black with core-type specific additions in colour. The inner core structure is shown with core specific additions where known. The enzymes are represented in brown. Unless otherwise noted, all linkages are in  $\alpha$ -anomeric conformation. LA, lipid A; OC, outer core; IC, inner core; Glc, Glucose; Gal, Galactose; Hep, Heptose; GlcNac, N-acetylglucosamine; Kdo, 3-deoxy- $\alpha$ -D-manno-oct-2-ulopyranosonic acid; P, phosphate.

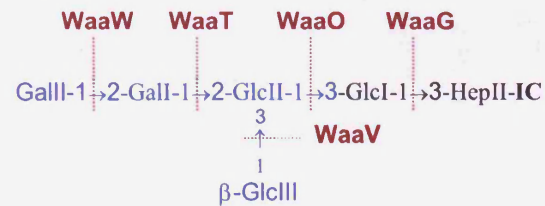
i) *E. coli* inner core



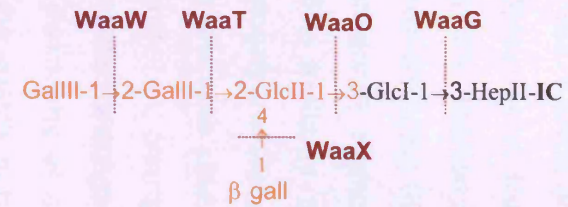
iv) *E. coli* R3 outer core



ii) *E. coli* R1 outer core



v) *E. coli* R4 outer core



iii) *E. coli* R2 outer core



vi) *E. coli* K12 outer core



similar to *E. coli* R1, R3 and R4 have been found in *Shigella* spp. *S. enterica* expresses two known core types represented by serovars Typhimurium and Arizonae (Heinrichs *et al.*, 1998; Kaniuk *et al.*, 2002). The variation in these structures is caused by differences in the carbohydrate backbone, i.e. the order of hexoses, and the nature, position and linkage of the side chains. As would be expected, the genes encoding the enzymes for the biosynthesis of the inner core are highly conserved. Homologues of the enzyme, WaaA, the Kdo transferase, share 96% similarity between all six core types and homologues of WaaC, a heptosyltransferase I, share 84% similarity (Heinrichs *et al.*, 1998). The variation in the core structure of the oligosaccharide can be linked to variation in gene content i.e. changes in the nature of the linkage site of the core oligosaccharide to the O-antigen are reflected by changes in the primary sequence of *waaL* (Heinrichs *et al.*, 1998b). The introduction of a glycosyltransferase causes the switch between the R1 and R4 core types. The distribution of the five *E. coli* core types within a genetically representative strain collection was measured using PCR and showed that while all five core types were present amongst commensal strains, the only core type found amongst the verotoxigenic isolates (serotypes O157, O111 and O26) was R3 (Amor *et al.*, 2000) suggesting some selective advantage in virulence or the acquisition of virulence genes (Whitfield *et al.*, 2003).

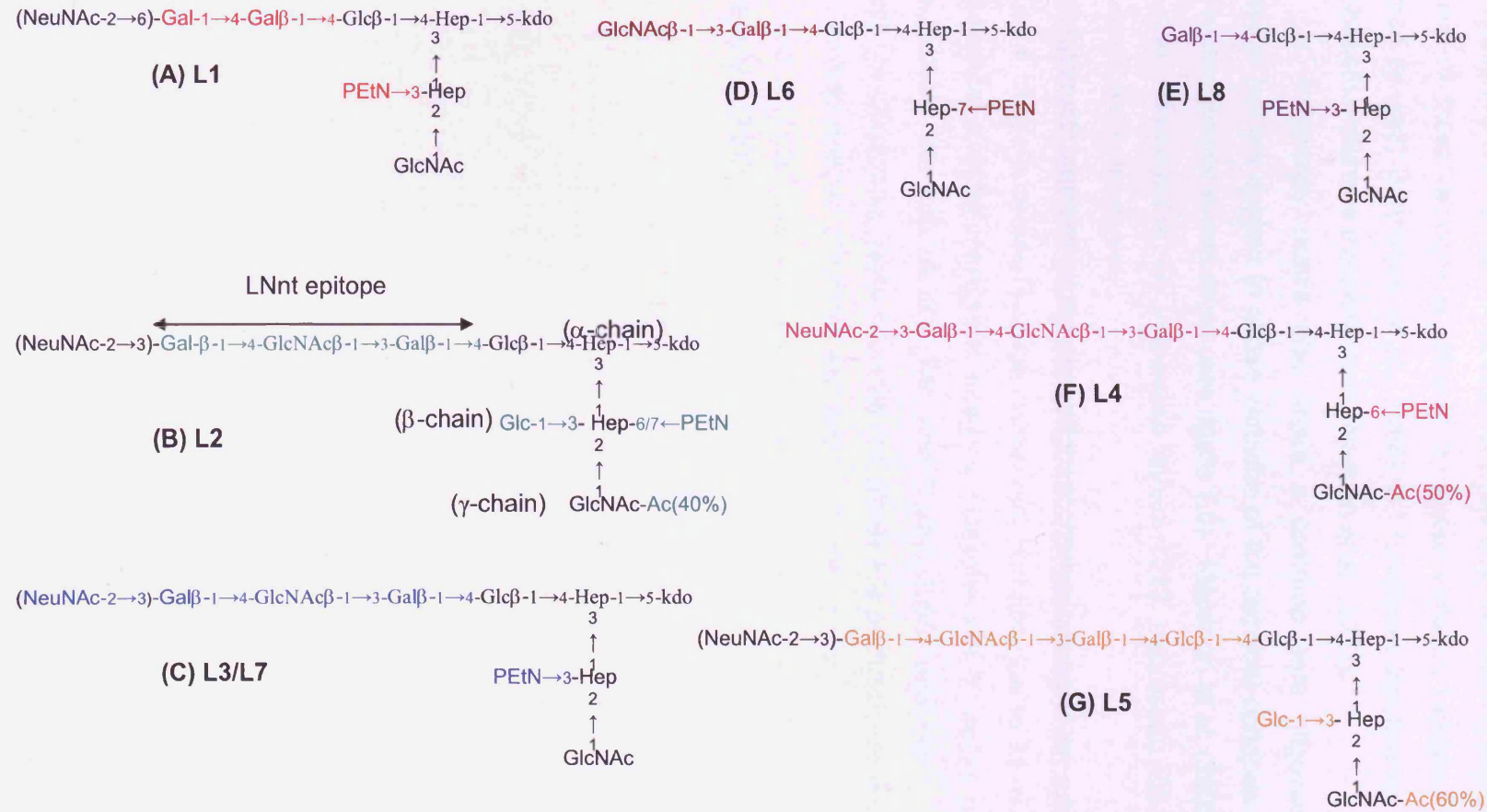
In *Salmonella* spp., the central operon of the *waa* cluster was highly conserved, except for the deletion of *waaK* in serovar Arizonae and the insertion of a novel glucosyltransferase, *waaH* (Kaniuk *et al.*, 2002). Other enterobacteriaceae such as *Yersinia* spp., *Klebsiella* spp. and *Proteus mirabilis* share the same common core backbone with *E. coli* and *Salmonella* (Regué *et al.*, 2001; Knirel *et al.*, 2005). Variation of the core within these bacteria is limited but *Y. pestis* exhibits temperature dependent alterations of lipid A.

In contrast to the enterobacteriaceae, *Neisseria meningitidis* expresses at least twelve immunologically different LOS epitopes (see Figure 1.5 for 7 examples). The inner core of the oligosaccharide is conserved consisting of two heptose residues and *N*-acetylglucosamine. LOS immunotype is determined by the length and composition of the chain extensions from the first heptose. The genes which encode the glycosyltransferases that

**Figure 1.5: Core structures from the *N. meningitidis* immunotypes L1-L8.** Adapted from Kahler and Stephens (1998) and Kogan *et al.* (1997). The conserved core structure is shown in black with core-type specific additions in colour. The inner core consists of the substituted Kdo and heptose sugars and the positions of the  $\alpha$ ,  $\beta$  and  $\gamma$  chains indicated on panel B. The brackets around the NeuNAc indicate the bacteria uses exogenous sources of CMP-NeuNAc to sialylate the LOS. The lacto-*N*-neotetarose (LNnt) epitope found in L2, L3, L4, L5 and L7 is indicated in L2. Unless otherwise noted, all linkages are in  $\alpha$ -anomeric conformation. Glc, Glucose; Gal, Galactose; Hep, Heptose; GlcNac, N-acetylglucosamine; Kdo, 3-deoxy- $\alpha$ -D-manno-oct-2-ulopyranosonic acid; PEtN, Phosphorylethanolamine; NeuNAc, N-acetylneuraminic acid; Ac, O-acetyl substitution.



## Chapter 1: Introduction



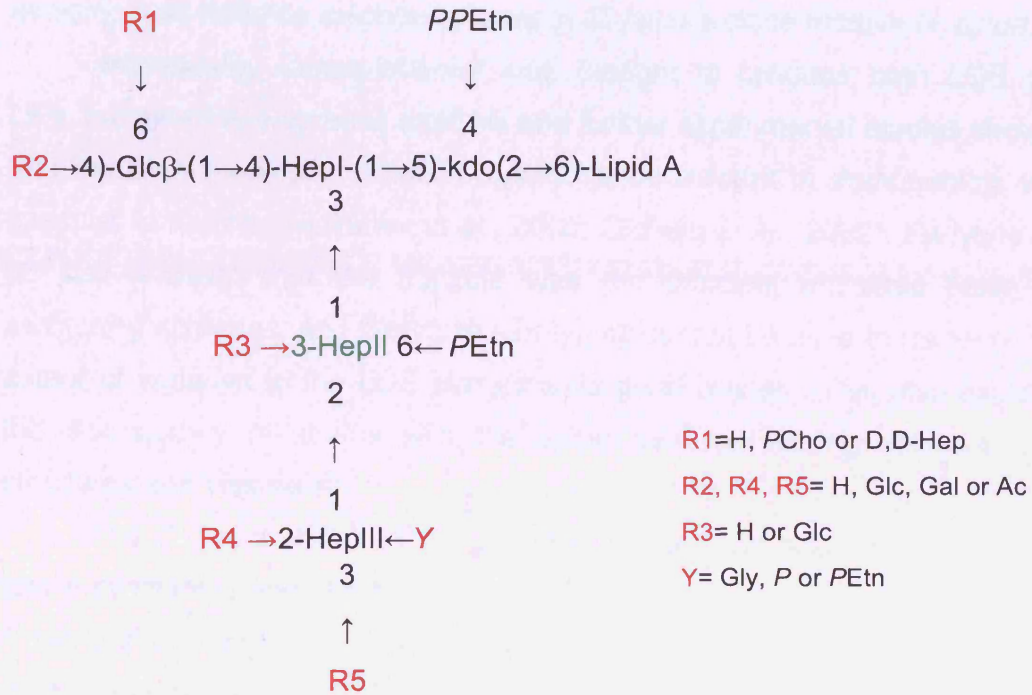


assemble the inner core are conserved and not subject to phase variation (Saunders *et al.*, 2000). The variability of the inner core oligosaccharide is also caused by the addition of various residues onto the second heptose. The addition of these residues is affected by phase variation (addition of  $\alpha$ 1,3-glucose by *lgt3*; Bannerjee *et al.*, 1998) or functional inactivation (*lpt3*, a phosphoethanolamine transferase; Mackinnon *et al.*, 2002).

*H. influenzae* strains also share a common core oligosaccharide backbone but are subject to phase variation of the terminal epitopes causing LOS heterogeneity in one strain (see figure 1.6). Månsson *et al.* (2003) found that one non-typeable *H. influenzae* strain 1233 expressed 38 variant oligosaccharide structures.

Although variation of the core oligosaccharide occurs to an extent, the backbone of the core seems to be conserved, perhaps due to its role in the maintenance of outer membrane integrity. Despite this *Y. pestis* cells are viable without the kdo moiety (Tan and Darby, 2005) and other bacterial species (*H. influenzae*, *Vibrio cholerae* and *Bordetella pertussis*) remain viable with one kdo residue (Gronow and Brade, 2000). *N. meningitidis* mutants have also been constructed with only lipid A and with no LPS present at all (Tzeng *et al.*, 2002).

**Figure 1.6: Schematic diagram showing the core structures of *H. influenzae*.** The conserved core structure is shown in black with the R groups representing the variable residues. The second heptose (in green) may also be replaced by L-*glycero*-D-manno-heptose. Adapted from Månsson *et al.* (2003). Glc, Glucose; Gal, Galactose; Hep, Heptose; GlcNac, N-acetylglucosamine; Kdo, 3-deoxy- $\alpha$ -D-manno-oct-2-ulopyranosonic acid; PEtN, Phosphorylethanolamine; Ac, O-acetyl substitution; Pcho, phosphocholine; H. hexose.



### 1.2.3 Lipooligosaccharides in *C. jejuni*

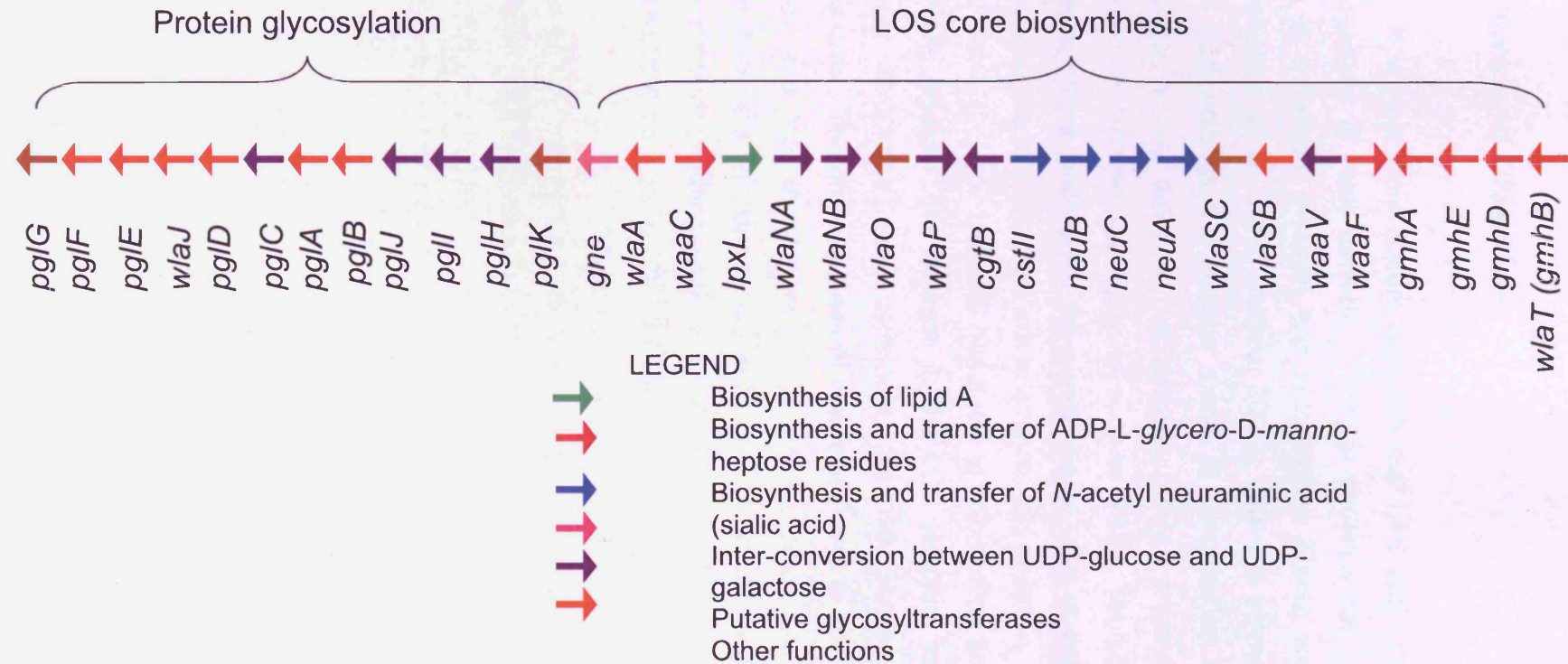
The importance of lipopolysaccharides in *C. jejuni* infection is still unclear but lipopolysaccharides and other adhesins have been shown to mediate adhesion (McSweeney and Walker, 1986; Oldfield, 2000). Although the role of lipopolysaccharides in enteric disease are still unknown, there is strong evidence for a role as an antecedent to GBS, which suggests a role in avoiding host defence mechanisms, as in *C. jejuni*'s close relative *H. pylori*.

Historically, *Campylobacter* was thought to produce both LOS and LPS, but genome sequence analysis and further experimental studies showed that the high molecular weight polysaccharide evident in immunoblots was capsular in nature (Karlyshev *et al.*, 2000; Oldfield *et al.*, 2002). Karlyshev *et al.* also showed that the capsule was the principal antigenic basis for serotyping schemes, and therefore serotyping cannot be used to measure the extent of variation in the LOS biosynthesis gene cluster. This also explains the discrepancy of strains with the same serotype having different LOS structures and *vice versa*.

The *C. jejuni* NCTC 11168 genome contains homologues to the *E. coli* lipid A synthesis genes *lpxA*, *lpxC*, *lpxD*, *lpxB* and *lpxK* indicating that the lipid A biosynthesis pathways are similar (Raetz and Whitfield, 2002). These genes are not linked in *C. jejuni* and a homologue to *lpxH* is absent showing potential differences in the way lipid A is synthesised. The gene encoding the Kdo transferase (*waaA*) is present alongside one late acetyltransferase gene, *lpxL*. LOS core biosynthesis in *C. jejuni* will be discussed in greater detail at the beginning of Chapter 6 (gene function) and Chapter 3 (genetic diversity).

The mechanisms by which *C. jejuni* causes enteric disease are still unclear, although certain aspects of the process are emerging. Lipopolysaccharides are obviously important toxins and virulence determinants in a wide range of Gram negative bacteria. Discovering their role in the enteric disease caused by *C. jejuni* will provide another step to determining the pathogenesis of campylobacteriosis.

**Figure 1.7: Representational diagram of the general protein glycosylation and lipooligosaccharide core biosynthesis gene cluster in NCTC 11168.** Brackets highlight genes involved in each system. Arrows show the direction of transcription and indicate putative or elucidated function. This diagram is not to scale. (Millar 2003)



### **1.3 Aims and Objectives**

It is clear from current research that LPS and LOS are important in pathogenesis of many bacteria. There is evidence that LPS and LOS have roles in adhesion, invasion and colonisation (Raetz and Whitfield, 2002; McSweeney, *et al.*, 1986, Oldfield, 2000). There is strong indication that *C. jejuni* LOS acts as a molecular mimic of gangliosides contributing to the avoidance of host defence mechanisms, rarely a cross-reaction may occur that leads to autoimmune diseases, such as GBS. Yet this focus has been a distraction from discovering the direct role of LOS in enteric infection, and certain aspects of this project aim to address this question.

As LOS has a role in colonisation and the development of enteric disease, the variation in the LOS structure indicates that these different structures may contribute the differences in host-pathogen interactions. Therefore one aim of this project was to continue from previous studies and look at the extent of gene content variation in the LOS biosynthesis cluster using a clinical disease strain collection. As the extent of gene content variation was unknown, it was postulated that novel gene combinations may be found in the clinical strain set.

Another aspect of this work involved characterising the role of LOS in interactions with the host and determining the functional role of individual LOS biosynthesis genes will involve the introduction of mutations in the LOS gene cluster. The construction of a deep rough mutant would confirm the minimal LOS gene content required for viability and potentially confirm the role of LOS in pathogenesis.

## **Chapter 2: Materials and Methods**

### **2.1. Bacterial Growth Conditions**

Cloning experiments were performed in two *E. coli* host strains, DH5 $\alpha$ E [*F*-80*dlacZ*M15 (*lacZYA-argF*) U169 *recA1 endA1 hsdR17(rk-, mk+)* *gal- phoA supE44 - thi-1 gyrA96 relA1*] (Donahue Jr and Bloom 1998) and *E.coli* DS941 (Summers and Withers 1990). *E. coli* was grown at 37°C in Luria-Bertani broth (LB; Oxoid) with shaking, or on Luria-Bertani agar (LA; Oxoid). Antibiotics were added to the media as necessary.

All *C. jejuni* strains used in this work are detailed in table 2.1. *C. jejuni* strains were maintained either on Mueller-Hinton agar (Oxoid) supplemented with 5% (w/v) horse blood (Oxoid) or in Mueller-Hinton broth. Vancomycin (Sigma) and trimethoprim (Sigma) were added to the media to aid the selection of *C. jejuni* and prevent contamination (see table 2.2 for concentrations). Cultures were grown in a Variable Atmosphere Incubator (VAIN, Don Whitley Scientific), which maintains an atmosphere of 5% O<sub>2</sub>, 10% CO<sub>2</sub> and 85% N<sub>2</sub>, and at a temperature of 37 or 42°C

### **2.2. Bacterial Strain Storage**

Bacterial strains were stored as glycerol stocks at -80 °C. To prepare the glycerol stocks, *E. coli* and *C. jejuni* were grown overnight in broth or on media plates, as required. After washing in PBS, the cells were resuspended in 500  $\mu$ l of fresh broth, with antibiotics if appropriate (see table 2.2), and transferred to a cryotube containing an equal volume of sterile, 70% (v/v) glycerol.



**Table 2.1: Details of *C. jejuni* strains used in this study.**

Table 2.1 shows the strains used in this project for mutagenesis and for the screening study, where UT stands for untypeable. CRU reference refers to the number assigned to the isolate on receipt by the *Campylobacter* reference unit.

## Chapter 2: Materials and Methods

<i>C.jejuni</i> strain (CRU reference)	HS serotype	Source
NCTC11168	2	National Collections of Type Cultures and Pathogenic Fungi, Colindale, London, UK. P. Everest, Glasgow T. Wassenaar, U of Mainz, Germany.
NCTC11351	23	
O:3	3	
O73	11	
NCTC11828 (81116)	6	
81-176	23/36	
480	UT	
34086	31	
34218	2	
34565	5	
34806	UT	J. Frost, Campylobacter Reference Unit, Laboratory of Enteric Pathogens, Specialist and Reference Microbiology Division, Health Protection Agency, United Kingdom
35303	18	
35451	4	
35503	12	
36670	11	
37531	50	
37895	5	
38577	19	
38608	37	
38625	50	
39271	6	
39864	21	
39893	2	
39902	19	
39918	21	
40973	50	
41803	4	
41999	50	
43771	13	
44253	13	
44406	5	
45283	2	
45385	12	
45600	13	
47185	12	
48643	37	
49068	4	
50612	5	
50702	11	
51566	31	
51585	37	
52831	12	
53305	UT	
54386	18	
54471	11	
54590	11	
57073	6	
57507	UT	
58526	50	
58728	6	
59653	19	
60087	31	
60238	18	
60319	18	
60584	50	
61666	2	

### **2.3. Media**

Autoclaving was used to sterilise media, at 121°C, 15 lb/inch for 15 min. The media was cooled to approximately 55°C before the addition of antibiotics, as required, and agar was poured into petri dishes (Sterilin).

#### **Luria-Bertani Broth and Luria-Bertani Agar**

LB broth (Roth, 1970) was prepared by dissolving 10 g of bacto-tryptone (Oxoid), 5 g of bacto-yeast (Oxoid) and 10 g NaCl in 950 ml distilled water. The pH of the media was adjusted to pH 7.0 with 10 M sodium hydroxide (NaOH) and the volume increased to 1 litre with water. To prepare LB agar, 1% (w/v) agar (Oxoid) was added prior to sterilisation.

#### **Mueller-Hinton Broth and Mueller-Hinton Agar**

Mueller-Hinton broth (MHB; Oxoid) was prepared by dissolving 21 g media (formula per 1 litre; 300 g beef dehydrate, 17.5 g caesin hydrolysate and 1.5 g starch) in 1 litre distilled water. Mueller Hinton Agar (MHA; Oxoid) contains 17g/l agar in addition to the other components of MHB, and was prepared as for MHB.

#### **SOC Media**

The recovery media used after bacterial transformation was SOC. The broth was prepared in a 180 ml volume with 0.5%(w/v) Yeast extract (Oxoid), 2.0%(w/v) Tryptone (Oxoid), 10 mM sodium chloride, 2.5 mM calcium chloride, 10 mM magnesium chloride and 20 mM magnesium sulphate. Following sterilisation by autoclave, 20 mM filter-sterilised glucose was added to a final volume of 200 ml. The broth was aliquoted and stored at -20°C.

## **2.4. Antibiotic and Media Supplements**

All antibiotics were purchased from Sigma and stock solutions were stored at 4°C in the dark. The table below details the solvent, stock concentration and the working concentration used. After preparation, all water-based antibiotic stock solutions were filter sterilised.

The chemicals, 5-Bromo-4-chloro-3-indol- $\beta$ -D-galactoside (X-Gal) and isopropylthio- $\beta$ -D-galactoside (IPTG) were used for blue-white screening of bacterial colonies that contain the plasmid, pUC19 and derivatives thereof, and these supplements were added to the media required. The solvent, stock concentration and working concentration are listed in the table below. Both solutions were stored at -20°C in the dark.

Supplement	Solvent	Stock conc. (mg/ml)	Working conc. ( $\mu$ g/ml)
Ampicillin	water	100	100
Kanamycin	water	50	50
Chloramphenicol	100% (v/v) ethanol	20	20
Vancomycin	water	10	10
Trimethoprim	50% (v/v) ethanol	5	5
Streptomycin	water	100	100
X-Gal	Dimethylformamide	20	20
IPTG	water	20	20
Novobiocin	water	2	0.0625 - 2

**Table 2.2**, detailing the supplements used, their stock and working concentrations

## **2.5. Miscellaneous buffers and solutions**

All chemicals were purchased from Fisher or Sigma unless otherwise stated, when required sterilisation was performed in an autoclave as previously stated.

### **3 M Sodium Acetate (pH 5.2)**

Sodium acetate was dissolved in 90 ml distilled water to a concentration of 3 M. The pH of the solution was adjusted to pH 5.2 using glacial acetic acid and the volume corrected to 100 ml before sterilisation.

### **40 x TAE buffer**

TAE buffer consisted of 1.6 M (Tris base) dissolved with 0.04 M EDTA. The pH was adjusted to 7.6 with glacial acetic acid and the solution made up to 10 l with distilled water.

### **Phenol/chloroform**

Identical volumes of liquefied phenol in Tris buffer and chloroform were mixed. The phenol/chloroform was stored at 4°C under 10mM Tris pH 7.5.

### **1 x Phosphate buffered saline (PBS)**

PBS tablets (ICN Biomedicals) were dissolved in distilled water, 1 tablet per 100 ml, and sterilised.

### **CTAB/NaCl solution**

To make CTAB/NaCl solution, 10% (w/v) CTAB (hexadecyltrimethyl ammonium bromide) was added to 0.7M sodium chloride prepared with distilled water and heated to 60°C to dissolve the CTAB.

### **5 x loading-buffer**

Loading buffer was prepared by dissolving 0.05%(w/v) Orange G (Sigma), in 1.25 ml 40 x TAE, 2.5 ml 50%(v/v) glycerol, and 6.25 ml distilled water.

### **Molecular-weight markers**

Molecular weight markers (Gibco-BRL) were prepared by using 50 µl of  $\lambda$  HindIII at 200 ng/µl, 20µl  $\phi$ X174 HaeIII at 200 ng/µl, 3 µl 5M NaCl, 60 µl loading-buffer, and 67 µl distilled water. This solution gives fragments of

known concentration and size ( $\lambda$  HindIII (kb): 23.1, 9.4, 6.5, 4.4, 2.3, 2.0, 0.56 and 0.12; and  $\phi\chi 174$  HaeIII (kb): 1.3, 1.1, 0.9, 0.6, 0.31, 0.28, 0.27, 0.23, 0.17, 0.11 and 0.07). A 1 Kb ladder (NEB; 10.0, 8.0, 6.0, 5.0, 4.0, 3.0, 2.0, 1.5, 1.0 and 0.5 Kb) was also used.

### **100 x Denhardt's Solution.**

Denhardt's solution consisted of 10 g Ficoll, 10 g polyvinylpyrrolidone and 10 g of bovine serum albumin were dissolved in 500 ml of distilled water. The solution was then aliquoted, and stored at -20°C.

## **2.6. Extraction of DNA from bacteria**

### **2.6.1. Extraction of chromosomal DNA from *C. jejuni***

*C. jejuni* cells were harvested from four swab plates with MHB and the pellets were then resuspended in 9.5 ml of PBS and lysed with 2 ml of 10% (w/v) SDS, 100  $\mu$ l of 20 mg/ml proteinase K and 20  $\mu$ l of 10 mg/ml RNase A. The suspension was incubated at 37°C for 1 h and then 1.8 ml of 5 M NaCl, and 1.5 ml of CTAB/NaCl were added and mixed thoroughly with the sample before a further incubation at 65°C for 20 min. An equal volume of 24:1 chloroform:isoamyl alcohol was added to extract the DNA and the two layers were separated by centrifugation at 3,020 *g* at 4°C for 20 min. The aqueous layer was pipetted into a clean tube and 0.7 volumes of isopropanol were added to precipitate the chromosomal DNA. The DNA was removed with a glass Pasteur pipette, washed in 70% (v/v) ethanol and left to dry before the addition of 1-2 ml of sterile, distilled water.

### **2.6.2. Small scale extraction of chromosomal DNA from *C. jejuni***

DNA extracted on a smaller scale using the Puregene Genomic DNA Purification kit (Gentra Systems). This method has been adapted from that described by Buffone and Darlington (1985). The protocol was followed as per manufacturer's instructions and adapted for plated bacteria. The composition of the buffers was not provided. Briefly, one plate of bacterial cells was harvested in 0.5 ml of MHB. This cell suspension was then spun at 13000 *g* for 10 s to pellet the bacteria. After removal of the supernatant, the cells were resuspended in 300 µl of Cell Lysis solution. The sample was then incubated at 80°C for 5 min. After the addition of 1.5 µl of RNase A solution, the cell suspension was then mixed and incubated at 37°C for 15 - 60 min. The sample was then cooled by placing on ice for 1 min before the addition of 100 µl of Protein Precipitation solution. After vortexing the sample to aid uniform mixing, the suspension was centrifuged for 3 min at 13,000 *g* to pellet the precipitated proteins. The supernatant was then poured into a microcentrifuge tube containing 300 µl of 100% (v/v) isopropanol and the solution mixed by inversion. The DNA was pelleted by centrifugation at 13,000 *g* for 1 min and the supernatant removed. The pellet was then washed with 70% (v/v) ethanol and centrifuged again. The supernatant was removed and the pellet allowed to air dry before resuspension overnight in 50 µl of DNA Hydration solution.

### **2.6.3. Extraction of plasmid DNA from *E. coli* and *C. jejuni***

The plasmid extraction methods used incorporate a modified version of the alkaline lysis method of Birnboim and Doly (1979). For small quantities of high-quality DNA, samples were prepared using the QIAprep Spin Miniprep kit (Qiagen), which was used according to manufacturer's instructions. All buffers were provided with the kit and all centrifugation steps were

performed at 13,000 *g*. The composition of all buffers was not available, and is stated where known. The basic method used is as follows: a 1 - 5 ml overnight culture was harvested and resuspended in 250  $\mu$ l of buffer P1 (100  $\mu$ g/ml RNase A, 50 mM Tris-HCl, 10 mM EDTA). 250  $\mu$ l of Buffer P2 (200 mM NaOH, 1% (w/v) SDS) was added causing cell lysis and 350  $\mu$ l of buffer N3 was added to precipitate the proteins, cell debris and chromosomal DNA. The precipitate was pelleted by centrifugation for 10 min and the supernatant transferred to the QIAprep spin-column. The plasmid DNA was bound by centrifugation for 60 sec and washed with 750  $\mu$ l of Buffer PE (contains ethanol). After elimination of residual wash-buffer, the plasmid DNA was eluted in 50  $\mu$ l of sterile distilled water or buffer EB.

Larger preparations of high-quality plasmid DNA were prepared using the HiSpeed Plasmid Midiprep kit (Qiagen). The kit was used according to manufacturer's instructions and all buffers were supplied with the kit. A 50 ml overnight culture was harvested and resuspended in 4 ml of buffer P1. 4 ml of buffer P2 was added and the mixture incubated at room temperature for 5 min. After the addition of 4 ml of buffer P3 (3 M potassium acetate, pH5.5), the lysate was immediately transferred into the QIAfilter Cartridge and incubated at room temperature for a further 10 min. The sample was filtered through the QIAfilter Cartridge into an equilibrated QIAGEN tip-100 (equilibrated with 4 ml of buffer QBT). The cell lysate passed through the tip by gravity flow, binding the plasmid DNA to tip-resin. The QIAGEN tip-100 was washed twice with 10 ml of Buffer QC before the DNA was eluted with 5 ml of Buffer QF. After precipitation of the DNA by the addition of 3.5 ml of isopropanol, the sample was then filtered through a syringe and QIAprecipitator midi module at constant pressure. The DNA was then washed by filtering 2 ml of 70% (v/v) ethanol through the syringe and QIAprecipitator module, which was then dried by passing air through it. One ml buffer EB or sterile distilled water was flushed through the syringe and precipitator to elute the DNA.



## **2.7. Quantification of DNA**

Chromosomal DNA was diluted in sterile, distilled water before absorbance was measured with a spectrophotometer (GeneFlow), and the concentration was calculated ( $A_{260} \times 50 \times \text{dilution factor}$ ). Genomic DNA used for microarray experiments was quantified using a Nanodrop ND-1000 spectrophotometer (Nanodrop Technologies, Inc.). A visual estimation of DNA concentration was made of plasmid and manipulated DNA following gel electrophoresis with ethidium bromide and visualisation with UV light of 290 nm. The sample DNA was compared to molecular weight markers of known concentration to estimate DNA concentration.

## **2.8. Electrophoresis of DNA**

### **2.8.1. Electrophoresis of DNA using an agarose gel**

DNA was separated by size using electrophoresis through a 0.8% -1.5% (w/v) agarose gel. The gels were prepared by dissolving SeaKem agarose (Cambrax) in 1 x TAE buffer and heated until completely molten. When cooled ethidium bromide (Sigma) was added to a final concentration of 0.5  $\mu\text{g/ml}$  and the agarose was poured into a gel-casting tray with a well comb and left to set. DNA samples were then prepared by the addition of 5 x loading-buffer. The gel was placed into the electrophoresis running tank, with 1 x TAE buffer, before the samples and molecular weight markers were loaded on a gel that had been placed in a gel running tank with 1 x TAE. Molecular weight markers were run along side the samples as an indicator of DNA size and concentration. Separation of the DNA occurred when approximately 5 V/cm was applied across the running tank until the loading-buffer front had travelled approximately three quarters of the way down the gel. The gel was visualised by UV light and photographed.

## **2.9. Purification of DNA**

The MinElute kit from Qiagen was used for the DNA purification and purifies up to 5 µg of DNA ranging from 70 bp – 4 kb in size. All buffers were provided with the kits and all centrifugation was carried out at 13,000 g for 1 min.

### **2.9.1. Gel purification**

To enable isolation of a particular sized DNA fragment, the sample was separated by agarose gel electrophoresis and the DNA - agarose fragment was excised from the gel. The *MinElute Gel Extraction protocol* from the MinElute purification kit was then followed. Briefly, the gel-fragment was weighed and incubated at 50°C with three volumes (100 mg = 100 µl) of buffer QG until the gel had dissolved. One volume of isopropanol was added to aid precipitation of DNA fragments, and the sample was transferred to the MinElute column. The DNA was bound to a silicon matrix in the presence of high ionic strength buffers. The column was then washed with 500 µl buffer QG to remove all traces of agarose. The column was then incubated for 1 min with 750 µl of buffer PE before the buffer was washed through by centrifugation. After the removal of excess buffer PE, the DNA was eluted in 10 µl of sterile, distilled water or buffer EB.

### **2.9.2. Purification following PCR or other enzymatic reactions**

For purification following PCR and restriction enzyme digestion, DNA was purified using the *MinElute PCR Purification Kit Protocol* from the MinElute purification kit. Briefly, 5 volumes of buffer PB were mixed with one volume of sample and transferred to the MinElute spin column. The DNA was bound to the silicon membrane in the presence of high ionic strength buffers and the sample was washed with 750 µl of buffer PE. After removal of

residual wash-buffer from the column, the DNA was eluted in 10  $\mu$ l of sterile, distilled water.

### **2.9.3. Ethanol Precipitation**

Ethanol precipitation was used to purify fragments above the size limit for the MinElute columns and ligation reactions. To one volume of sample, 0.1 volumes of 3 M sodium acetate (pH 5.2) and 2 volumes of 100% (v/v) ethanol were added. 1  $\mu$ l of glycogen (Roche Diagnostics) was also added to aid pellet formation in samples with a small amount of DNA. The mixture was incubated at -80°C for 1 h before centrifugation at 13,000 g for 20 min. The pellet was washed in 1 ml of 70% (v/v) ethanol for 5 min and dried at 37°C or room temperature, before resuspension in 10 – 20  $\mu$ l of sterile, distilled water.

## **2.10. Enzymatic modification of DNA**

### **2.10.1. Restriction endonuclease digestion of DNA**

New England Biolabs (NEB) supplied restriction endonucleases and buffers. Restriction digests were carried out according to manufacturer's recommendations.

### **2.10.2. Ligation of DNA**

T4 DNA Ligase was (NEB) used as recommended by the manufacturer. A typical reaction consisted of: distilled water, vector DNA (ng), insert DNA (ng), ligase buffer, 1 unit of ligase for sticky ends or 5 units of ligase for blunt ends. Ligation reactions were performed overnight at 4°C.

### **2.10.3. Dephosphorylation of restriction digested DNA**

The enzyme, Shrimp Alkaline Phosphatase (Roche) was used to prevent the self-ligation of vector molecules with complementary ends as the enzyme removes the 5' terminal phosphate. Reactions were carried out according to the manufacturer's recommendations and usually 1  $\mu$ l of enzyme was added to a restriction digestion reaction and incubated for 30 min at 37°C, and then deactivated by heating to 65°C for 20 min.

## **2.11. Transformation of DNA**

### **2.11.1. Preparation of electrocompetent *E. coli***

A 1 ml overnight culture of *E. coli* was used to inoculate 100 ml of LB. The cells were then agitated at 37°C until the OD<sub>600</sub> nm of the culture was between 0.4 - 0.6. The cells were harvested, and washed three times in decreasing volumes of ice-cold, sterile, distilled water (100 ml, 50 ml and 20 ml). The cells were then resuspended in 10 ml of ice-cold, 10% (v/v) sterile, glycerol and divided into 50  $\mu$ l aliquots. The electrocompetent cells were then used straightaway or frozen on dry ice and stored at -80°C.

### **2.11.2. Preparation of chemically competent *E. coli***

A 1 ml overnight culture of *E. coli* was used to inoculate 100 ml of LB and the cells were shaken to an OD<sub>600</sub> nm of 0.4 -0.6. After harvesting the cells were washed with 100 ml of 50mM calcium chloride before finally being resuspended in 10 ml of a 50 mM calcium chloride and 20% (v/v) glycerol solution. The cells were aliquoted on dry ice and stored at -80°C.

### **2.11.3. Electroporation of Plasmid DNA into *E. coli***

Plasmid DNA at a concentration of approximately 0.5 ng or ethanol precipitated ligation reactions were mixed with 50  $\mu$ l of electrocompetent cells and transferred to an ice-cold electroporation cuvette (Bio-Rad or GeneFlow). Cells were electroporated at 2.5 kV, 200  $\Omega$  and 25  $\mu$ F. One ml of SOC media was added to the cuvette and mixed with the cells. The transformation was then incubated for 1 h at 37°C whilst shaking. After centrifugation and resuspension in smaller amounts of media the transformation was plated on LA with the appropriate antibiotic selection.

### **2.11.4. Heat-shock transformation of chemically competent *E.coli***

Ethanol precipitated ligations or plasmid DNA of ~0.5 ng were mixed with 40  $\mu$ l of hand thawed chemically competent cells and chilled for 20 min. Heat shock was either performed at 42°C for 90 s or 37°C for 5 min, before placing the transformation on ice for 2 min. One ml of SOC media was then added to the transformation which was then incubated at 37°C with shaking for 1.5 h. After centrifugation and resuspension in smaller amounts of media the transformation was plated on LA with the appropriate antibiotic selection.

### **2.11.5. Preparation of electrocompetant *C. jejuni***

Four overnight swab plates of cells were harvested and washed three times in decreasing amounts of ice-cold CEB buffer (15% (v/v) glycerol, 272 mM sucrose). The cells were finally resuspended 50  $\mu$ l of CEB per

transformation, and used straight away or frozen on dry ice and stored at -80°C.

#### **2.11.6 Electroporation of Plasmid DNA into *C. jejuni***

Several micrograms of clean plasmid DNA were mixed with 50 µl of electrocompetent cells and transferred to an ice-cold electrocuvette. After 10 min incubation on ice, the sample was electroporated at 2.5 kV, 200Ω and 25µF, then 100 µl of SOC media was added to recover the cells from the cuvette and the suspension was plated on MHA. The plates were incubated in the VAIN, at 37°C overnight before the cells were transferred onto MHA supplemented with the appropriate antibiotic.

#### **2.11.7 Natural transformation of *C. jejuni***

The natural transformation of *C. jejuni* was carried out using a biphasic method. *C. jejuni* was recovered from a swab plate and resuspended to an OD<sub>600</sub> of 0.5. 500 µl of this suspension was then placed in a 15 ml tube containing 1 ml of MHA and this was incubated for 3 h with shaking under microaerobic conditions. The DNA to be transformed was added at a concentration of 1 - 5 µg and incubation was continued for a further 5 h. The suspension was then decanted from the tube and centrifuged before being resuspended in a smaller volume and plated on MHA containing the appropriate antibiotics.

### **2.12. Amplification of DNA by Polymerase Chain Reaction**

Polymerase chain reaction (PCR) was employed to amplify specific regions of DNA. Several different protocols were followed depending on the next use of the product. The reagents were obtained as from the following

companies: primers were obtained from either PNACL, University of Leicester (Protein and Nucleic Acid Laboratory) or Invitrogen; dNTPs (Amersham Pharmacia Biotech) were obtained as separate solutions, diluted and mixed; 25 mM MgCl<sub>2</sub> solution, *Taq* DNA polymerase and 10 x PCR buffer were purchased from Abgene (Buffer IV; 750mM Tris-HCl, pH 8.8, 200mM NH<sub>4</sub>SO<sub>4</sub>, 0.1% (v/v) Tween-20).

### 2.12.1. Standard PCR protocol

Where high fidelity DNA replication was not required, a standard PCR protocol was followed. Reactions were performed in 20 µl volumes, as set out in table 2.3. The reactions were placed in a thermocycler (Eppendorf or MJ Research) and table 2.4 below shows a typical cycling profile.

Component	Volume/µl	Final Concentration
Sterile, distilled water	Up to 20 µl	
Upstream primer, 2 pM/µl	1 .00	400 nM
Downstream primer, 2 pM/µl	1 .00	400 nM
dNTP mix, 10 mM	0.40	200 µM
10 x PCR buffer	2.00	1 x
MgCl <sub>2</sub> , 25mM	3.20	
Template DNA: Genomic, 10-50 ng/µl	1.00	10 – 50 ng
Plasmid, 10 ng/µl	1.00	10 ng
Taq DNA Polymerase	0.25	5 U

**Table 2.3** detailing the components of an standard PCR reaction and their final concentrations.

Cycle	Temperature (°C)	Time	Description
1x	96	2 min	Initial template denaturation
27x	96	30 s	Template denaturation
	40-65	30 s	Primer annealing
	72	30 s - 5 min	Primer extension/elongation
1x	72	5 min	Final extension/elongation

**Table 2.4** showing a typical PCR cycling profile.

The annealing temperature used was 5-10°C less than the lowest  $T_m$ °C of the primers, and extension times were related to the target size, with 1 min for every kb of DNA and extension was carried out at 72°C.

### **2.12.2. Colony PCR**

Colony PCR was used to directly screen transformants for recombinant plasmid DNA (*E. coli*) or mutations (*C. jejuni*). Each colony was picked and resuspended in 10 µl of water before boiling in a thermocycler for 5 min. The 'boilate' was then centrifuged to pellet the cell debris. The primer concentration in the reaction was increased to 4 pm/µl and 2 µl of the boilate supernatant was used as the template. The other conditions were the same as for the standard PCR protocol.

### **2.12.3. TripleMaster®Mix PCR system**

The TripleMaster® PCR system (Eppendorf) was used when products of high fidelity were required. The enzyme mix combines high extension rates and proofreading abilities, and two buffers are optimised for different lengths of amplification. Two master mixes were prepared on ice (see table 2.5) and mixed prior to cycling on a Hot Start programme (see table 2.6).



Components	High Fidelity Final Concentration
<b>Master Mix 1</b>	
Sterile distilled water	To a final volume of 10 $\mu$ l
Upstream Primer	400 nM
Downstream Primer	400 nM
g DNA (kb): 0.1-3	10-25 ng
3-10	25-50 ng
Plasmid DNA (kb): 0.1-3	0.1-1 ng
<b>Master Mix 2</b>	
Sterile, distilled water	To a final volume of 40 $\mu$ l
10 x High Fidelity buffer	1 x; 2.5 mM $Mg^{2+}$
dNTP mix	200 $\mu$ M
TripleMaster Mix Enzyme	0.5 –1.5 U

**Table 2.5** showing the composition of a TripleMaster mix PCR reaction and the final concentration of the components

An adapted hot-start program was used, where the thermocycler was pre-heated to the initial denaturation temperature before the reactions were loaded.

<b>TripleMaster mix - High Fidelity</b>		
Cycle	Temp(°C)	Time
1x	94	5 min
25-35x	94	20 s
	40-65	10-20 s
	72	20 s - 8 min
1x	72	1 min/ Kb

**Table 2.6** detailing the cycling profile used for a TripleMaster mix PCR.

#### 2.12.4. Long range PCR

The MasterAmp™ Extra-long DNA polymerase mix (Epicentre Biotechnologies) and 2x premix 6 (composition unknown) were used for the amplification of products over 10 Kb. Two mastermixes were prepared on ice and after mixing were placed in a pre-heated thermocycler. The components were added in the concentrations detailed in table 2.7. The recommended cycling parameters are shown in table 2.8.

Components	
<b>Master Mix 1</b>	
Sterile distilled water	To a final volume of 25 µl
Upstream Primer	400 nM
Downstream Primer	400 nM
Genomic DNA	100-150 ng
10-22 Kb	
DNA polymerase mix	1 µl (2.5 U)
<b>Master Mix 2</b>	
2x Premix	25 µl

**Table 2.7** detailing the composition of a Masteramp extra long PCR.

Cycle	Temperature (°C)	Time
1x	96	1 min
15x	96	1 min
	45-65	1 min
	68	1 – 21 min
10x	96	1 min
	40-65	1 min
	68	1– 21 min + 20 s per cycle
1x	68	1-21 min

**Table 2.8** detailing a typical long-range PCR cycling profile.

### **2.13. DNA Sequencing**

All cycle sequencing reactions were carried out using the BigDye v3.1 Terminator Cycle Sequencing Ready Reaction kit (Applied Biosystems). Briefly, approximately 100 – 500 ng of template was added to 4 pM of primer, 8 µl of diluted BigDye Version 3.1 (diluted 1 in 8 with sequencing buffer and water) and more water to 20 µl. The DNA was amplified by 30 cycles of 96°C for 30 s, primer annealing temperature for 30 s, and 60°C for 4 min. All sequencing reactions were cleaned by boiling with 2 µl of a 2.2% (w/v) solution of SDS to remove salts, before further cleaning using a DyeEx Spin column (Qiagen) and submitted to PNACL in the University of Leicester for sequence analysis using an ABI 3730 DNA sequencer (Applied Biosystems).

### **2.14. DNA Sequence analysis**

DNA sequencing profiles were received from PNACL as a chromatogram, which was viewed with Chromas v 1.45 (available at <http://www.technelysium.com.au/chromas14x.html>). Comparison with a known DNA sequence was carried out using Blast Two Sequences (<http://www.ncbi.nlm.nih.gov/blast/bl2seq/bl2.html>) and Clone Manager (Scientific & Education Software USA). Unknown DNA sequences were compared to other known DNA sequences using nucleotide-nucleotide BLAST (BlastN) (<http://www.ncbi.nlm.nih.gov/BLAST/>) and the *Campylobacter* database CampyDB, <http://campy.bham.ac.uk/>.

## ***2.15 Analysis of chromosomal DNA by Southern blotting***

### **2.15.1 Transfer of DNA**

Based on the protocol devised by Southern (1975) the genomic digest was separated by electrophoresis on a 1% (w/v) agarose gel at 4 V/cm in 1 x TAE. After visualisation under UV light, the gel was washed in depurinating solution (250mM HCl) for 10 min with agitation, rinsed in distilled water and then washed in denaturation solution (1.5M NaCl, 0.5M NaOH) for 25 min. After a final wash with distilled water, the gel was then washed in neutralisation solution (1.5M NaCl, 0.5M Tris pH 7.5) for 30 min. The gel was then placed in apparatus with 10 x SSC (Saline Sodium Citrate buffer) and Hybond N+ (GE Healthcare) to allow capillary transfer to occur overnight. After disassembling the apparatus, the sample lanes and markers were labelled prior to fixing the membrane by UV cross-linker.

### **2.15.2. Labelling of the probe**

The probe was labelled non-radioactively using the Gene Images Random Prime labelling module (GE Healthcare), as per manufacturer's instructions. Fluorescein-11-dUTP is used in place of dTTP to produce a fluorescein labelled probe. The DNA to be labelled was used at a concentration of between 2 - 25 ng/ $\mu$ l and was firstly denatured at 95°C for 5 min and then placed on ice. Ten  $\mu$ l of dNTPs, 5  $\mu$ l of random primer, 50 ng of DNA, and 1  $\mu$ l of Klenow fragment were mixed with water up to a volume of 50  $\mu$ l. The reaction was then incubated at 37°C for 1 h and used immediately in a hybridisation or stored at -20°C.

### **2.15.3. Hybridisation and stringency washes**

The Southern blot membrane was pre-hybridised with 0.125 ml/cm<sup>2</sup> of pre-warmed hybridisation solution (5 x SSC, 0.1% (w/v) SDS, 5% (w/v) dextran sulphate, 20 fold dilution of liquid block) for 2 h at 65 °C. The probe was denatured at 95 °C prior to addition to the hybridisation solution and added at a concentration of 10 ng/ml. After probe addition, hybridisation was allowed to continue at 65 °C overnight with agitation.

Stringency washes were carried out at 65 °C in pre-warmed buffers. The membrane was firstly washed in 1 x SSC, 0.1% (w/v) SDS for 15 min and this was followed by three washes lasting 15 min each in 0.5 x SSC, 0.1% (w/v) SDS.

### **2.15.4. Blocking, antibody incubation and detection**

The membranes were incubated at room temperature with constant agitation for 1 h in blocking solution. Blocking solution was made by diluting liquid blocking agent 10 fold in buffer A (300 mM NaCl, 100 mM Tris pH 9.5). The anti-fluorescein-AP was then diluted 5000 fold in buffer A with 0.5% (w/v) BSA. The membrane was incubated in this solution for 1 hr at a concentration of 0.3ml/cm<sup>2</sup>. The unbound antibody was removed by three washes of 10 min in 2-5 ml/cm<sup>2</sup> of buffer A and 0.3% (v/v) Tween 20.

Detection was carried out using the Gene Images CDP-Star detection kit (Amersham Biosciences). The membrane was removed from the wash solution and the excess drained before placing it DNA side up on a section of Saran wrap. The detection agent was added to the membrane at a concentration of 30 - 40 µl/cm<sup>2</sup> and allowed to incubate for 2 - 5 min. Excess detection reagent was drained from the membrane before it was enclosed in a fresh piece of Saran wrap. The membrane was then placed in a film cassette and X-ray film was placed on the blot in the dark. The film was exposed and then developed using a Compact X4 auto developer (Xograph).

## **2.16. Analysis of LOS by microarray hybridisation**

### **2.16.1 Microarray design and construction**

The 70-mer oligonucleotides were designed by Operon to the LOS sequences provided (appendix 2), with matched  $T_m$ s and amino-allyl modification. BlastN analysis of the oligonucleotides against the genome sequence established the absence of cross hybridisation. The 70-mers were also printed alongside a PCR product microarray based on the whole genome of *C. jejuni* NCTC 11168 as previously described (Moen *et al.*, 2005). The oligonucleotides were resuspended to a concentration of 50 pm/ $\mu$ l in 50% (v/v) DMSO (Dimethyl sulfoxide; Sigma).

### **2.16.2 Labelling, hybridisation and data acquisition**

Initially approximately 5  $\mu$ g of control strain (NCTC 11168) and test strain chromosomal DNA was labelled with Cy5 and Cy3 (GE Healthcare) with random priming, as previously described (Dorrell *et al.*, 2001). Briefly, microarray slides were prehybridised in 3.5x SSC, 0.1% (w/v) SDS and 10 mg/ml BSA at 65°C for 20 min. The slides were then washed in distilled water and isopropanol for 1 min at each stage. The Cy5 and Cy3 labelled DNA was mixed and purified using a MinElute spin column. Cy dye incorporation was measured using the Nanodrop spectrophotometer before hybridisation. Prior to hybridisation, the cleaned eluate was denatured and 4 x SSC, 0.3% (w/v) SDS was added to provide a hybridisation solution totalling 45  $\mu$ l in volume. The slides were hybridised overnight at 65°C under 22 x 22 mm LifterSlips (Erie Scientific) in a sealed, humidified hybridisation chamber (Telechem International) and in a 65°C water bath. The slides were washed for 2 min in 400 ml 1x SSC, 0.06% (w/v) SDS at 65°C and twice in 400ml of 0.06 x SSC with agitation. Slides were scanned using an Affymetrix 418 scanner (MWG Biotech) and the signal data was extracted using Imogene 5.2 (BioDiscovery). Further data acquisition and

data analysis was performed using BlueFuse (BlueGnome Ltd). The BlueFuse algorithm quantifies and normalises microarray data and also provides a 'PON' value that assesses binary changes between the test and control strains and the presence of the array element in both channels. The data from the whole genome microarray was also analysed using GeneSpring v7 (Agilent).

The hybridisation protocol was adapted to produce stronger signal from the spotted oligonucleotides. This involved the addition of deionised formamide (Qbiogene) to the hybridisation solution. The final working hybridisation solution consisted of: 30% (v/v) formamide, 3.75% (v/v) Denhardtts solution, 3.75 mM sodium pyrophosphate (Sigma), 0.75 mM Tris pH 7.4, 0.375% (w/v) SDS in sterile distilled water.

In later experiments, only one colour hybridisations were performed. In these experiments, SpotQC (Integrated DNA Technologies) was used as a universal control. This involved a second hybridisation step after the scanning of the Cy5 channel. Briefly, 45 µl of detector oligonucleotides and hybridisation solution was heated to 80°C for 5 mins, vortexed and cooled before adding to the microarray. The hybridisation was allowed to proceed in the dark at room temperature for 30 - 60 min. The slides were washed in buffer 1 (10x SSC, pH 7.0, 0.2% (w/v) Sarkosyl) for 3 min. They were then transferred to the second wash buffer (10x SSC) and washed for 2 min. The slides were finally washed in 2 x SSC for 2 - 3s and were then centrifuged to remove any residual wash solutions. The Cy3 channel was scanned for image analysis.

### **2.17. Analysis of Lipooligosaccharide (LOS)**

#### **2.17.1. Whole cell preparation of *C. jejuni* LOS**

Crude LOS/LPS extractions were prepared from one 0.6 OD<sub>600</sub> unit of an overnight culture. The bacterial cells were harvested and resuspended in 200 µl of 2x SDS-PAGE loading buffer (100 mM Tris-HCl pH 8.0, 2% (v/v) β-mercaptoethanol, 4% (w/v) SDS, 0.2% (w/v) bromophenol blue, 0.2% (w/v)

xylene cyanol, 20% (v/v) glycerol). After boiling for 10 min, 5 µl of 20 mg/ml proteinase K was added and the sample was incubated for 2 h at 65°C. Samples were stored at -20°C prior to analysis.

### **2.17.2. Electrophoresis of LOS using Tricine SDS-Polyacrylamide Gel Electrophoresis (PAGE)**

The gels were cast and run using Protean II small Electrophoresis cells (BioRad). All plates, combs and spacers were washed in distilled water and industrial methylated spirits (IMS) to remove impurities that could later inhibit silver staining of the gel. The stacking gel and separating gel were composed of: Protogel (30% (v/v) acrylamide/bisacrylamide, Flowgen), 4 x Gel Buffer (4M Tris-HCl, pH 8.45, 0.4% (w/v) SDS), glycerol, distilled water, 10% (w/v) Ammonium Persulphate (APS) and TEMED (N, N, N', N'-tetramethyl-ethylenediamine, Sigma) and were mixed according to the gel required. Table 2.9 details the composition for each percentage gel.

	Separating gel		Stacking gel
Size of gel	2x small		2x small
Component	12% (ml)	18% (ml)	4% (ml)
Protogel	4.00	6.00	0.750
4 x gel buffer	2.50	2.50	0.930
Glycerol	1.05	1.05	-
Distilled water	2.45	0.45	3.32
10% (w/v) APS	0.04	0.04	0.025
TEMED	0.004	0.004	0.01

**Table 2.9** showing the composition of Tricine - SDS PAGE gels

The separating gel was overlaid with isopropanol to aid uniform polymerisation, and the isopropanol was removed before the comb was



inserted and the stacking gel was poured. The running tank was assembled using SDS-PAGE running buffer (0.1M Tris, 0.1M Tricine and 0.1% w/v SDS, pH 8.25). The samples were boiled for 10 min prior to loading and 5  $\mu$ l of sample was loaded onto the gel. Biotinylated protein molecular-weight-markers (dissolved in SDS-page loading buffer, Vector Laboratories) of known molecular weights [(kDa) 220, 132, 95, 68, 55, 45, 31, 18] were run next to the sample as a guide to molecular weight. Small gels were run at a constant voltage of 150 V.

### **2.17.3. Silver staining of LOS Tricine SDS-page gels**

The LOS/LPS was stained using the silver stain method of Tsai and Frasch (1982). The SDS-page gel was fixed (40% (v/v) Ethanol, 5% v/v acetic acid) for 2 h - overnight before washing for 5 min with 0.7% (w/v) periodic acid, 40% (v/v) Ethanol, 5% (v/v) acetic acid. The gel was washed a further 4 x 10 min with distilled water before the stain is applied (28 ml 0.1 M NaOH, 2 ml  $\text{NH}_4\text{OH}$ , 115 ml distilled water). Five ml of fresh 20% (w/v) Silver nitrate was added drop-wise to the stain and incubated for 5 -10 min. To remove excess stain, the gel was washed 2 x 10 min with water before the developer was added (10 mg citric acid, 100  $\mu$ l formaldehyde, 200 ml distilled water). After 3 – 5 min the reaction was stopped with water. The gels were dried and photographed.

### **2.17.4. Electrotransfer of LOS from Tricine SDS-PAGE gels to PVDF membrane**

The LOS was also detected with antibodies or other ligands using the Western Blot procedure. After SDS-PAGE, the gel is equilibrated in transfer buffer (33 mM Tris, 0.19 M glycine, 20% (v/v) methanol) for 10 min to allow buffer exchange. The LPS was transferred to PVDF (flurotrans, Flowgen) membranes in ice-cold transfer buffer with constant stirring, for 2 h at 395

mA in a Trans-Blot cell (BioRad) or 1 h at 100 V in a Mini Trans-Blot cell (BioRad).

### **2.17.5. Hybridisation of antibodies and ligands to the membrane, and membrane development**

The BM Chemiluminescence Western Blotting Kit (Roche Applied Science) was used to detect the immunoblot probes according to the manufacturer's instructions. The following steps were carried out at room temperature with gentle agitation. Following transfer, the membrane was washed twice for 5 min in TBS (20 mM Tris, 500 mM NaCl, pH 7.5) and blocked for 1 h in blocking solution (0.2% w/v non-fat dry milk in TBS). After washing in TTBS (TBS and 0.1% v/v Tween-20) the primary ligand was added to antibody buffer (0.2% w/v non-fat dry milk in TTBS) and incubated with the membrane for 2 hours. The membrane was then washed twice with TTBS to remove unbound primary antibody/ligand before the addition of the second antibody/ligand diluted in antibody buffer. After incubation for 1 h the membrane was washed a further three times with TTBS before development. Detection agent B was diluted 1 in 100 with the substrate solution A and allowed to equilibrate prior to addition to the membrane. All excess liquid was drained from the membrane and the substrate was added to the membrane at a concentration of 125  $\mu\text{l}/\text{cm}^2$ . The substrate was drained from the membrane after 1 min and the membrane was exposing to X-ray film for approximately 1 - 2 min.

Two primary ligands were used *Riccinus communis* agglutinin (Sigma) and Cholera toxin subunit B conjugate (Sigma). Cholera Toxin B-Subunit biotin conjugate was diluted to 12.5  $\mu\text{g}/\text{ml}$  and detected with a streptavidin-Alkaline Phosphatase conjugate, Extravidin (Sigma), which was diluted 1:100,000.

## **2.18 Analysis of Outer Membrane preparations**

### **2.18.1. Extraction of outer membrane proteins from *C. jejuni***

Outer membrane proteins were prepared by differential solubility in Triton X-100 (Schnaitman 1971; Wooldridge *et al.*, 1992). Samples were prepared from one swab plate and the OD corrected to a  $A_{600}$  of 0.6. Cells were harvested, centrifugation at 10,000 x g for 10 min, resuspended in 0.5 ml envelope buffer (EB; 10 mM Tris-HCl, pH 7.5) and lysed by sonication. Unlysed cells were removed by centrifugation at 7,000 g for 5 min, and crude total membranes were isolated by ultracentrifugation at 100,000 g for 10 min in a Beckman TL-100 ultracentrifuge. Membrane pellets were resuspended in 1.0 ml of EB containing 2% v/v Triton X-100 (Sigma; EBT) and extracted at room temperature for 30 min. Triton insoluble pellets were collected by ultracentrifugation and re-extracted in 1.0 ml of EBT for a further 30 min at room temperature. Following ultracentrifugation outer membrane containing pellets were resuspended in 40  $\mu$ l EBT and 40  $\mu$ l 2x SDS-PAGE sample buffer.

### **2.18.2. Electrophoresis of protein by SDS-PAGE**

The gels were assembled and run using the Protean II small electrophoresis cells (BioRad). The stacking gel and separating gel were composed from: protogel (30% (v/v) acrylamide/bisacrylamide), 4 x Gel Buffer pH 8.8 (4 M Tris-HCl, pH 8.8; 0.4% (w/v) SDS), 1 x Gel buffer pH 6.8 (1 M Tris-HCl, pH 6.8; 0.1% (w/v) SDS), distilled water, 10% (w/v) Ammonium Persulphate (APS) and N, N, N', N'-tetramethyl-ethylenediamine (TEMED, Sigma) and were mixed according to the gel required (as detailed in table 2.10)

Size of gel	Separating Gel		Stacking gel
	2x small		2 x small
Component	8% (ml)	12% (ml)	4% (ml)
Protogel	4.00	6.00	0.65
4 x gel buffer pH8.8	3.75	3.75	-
1 x gel buffer pH6.8	-	-	1.25
Distilled water	7.25	5.25	3.05
10% (w/v) APS	0.05	0.05	0.025
TEMED	0.01	0.01	0.005

**Table 2.10** showing the composition of SDS-PAGE for separation of proteins

The separating gel was overlaid with isopropanol to aid uniform polymerisation, and the isopropanol is rinsed off before the comb was inserted and the stacking gel was poured. A 5x SDS electrophoresis buffer stock was prepared (0.125 M Tris base, 0.96 M glycine, 0.5% (w/v) SDS) and stored at 4°C. The Running tank was assembled and 1x buffer was added to the upper and lower chambers. All samples were boiled for 10 min prior to loading. Approximately 5 µl of sample was loaded. Biotinylated protein molecular-weight-markers (dissolved in SDS-page loading buffer, Vector Laboratories) of known molecular weights (kDa; 220, 132, 95, 68, 55, 45, 31, 18) were run next to the sample as a guide to molecular weight. Gels were run at a constant voltage of 150 V.

### 2.18.3. Staining of protein SDS-PAGE gels

Gels were stained for 30 min in Commasie blue stain (50% (v/v) methanol, 10% (v/v) acetic acid, 0.05% (w/v) Commasie Blue R-250) and destained in 12.5% (v/v) isopropanol, 10% (v/v) acetic acid. Destain solution was changed several times before the gels were dried and photographed.

### **2.19 Preparation of dehydrated cells for capillary electrophoresis mass spectroscopy**

Capillary electrophoresis mass spectroscopy was performed by Jianjuin Li (NISBC, Canada). Cells were prepared to the air-drying stage before being sent to Canada. *C. jejuni* cells from one swab plate were harvested using 0.3 ml of PBS and 100 % (v/v) ethanol was added to the mixture, which was then incubated for 1 - 2 h at room temperature. The cells were pelleted by centrifugation and the supernatant was discarded. The pellet was twice with 95% (v/v) ethanol and twice with acetone before being air-dried at room temperature. The dried cells were resuspended in 200 µl of 60 µg/ml proteinase K and incubated at 37°C for 4h. Digestion was stopped by increasing the temperature to 75°C for 10 min and the sample was then lyophilized. The cells are then resuspended in 200 µl of 20 mM ammonium acetate buffer (pH 7.5) containing 200 µg/ml RNase A and 100 µg/ml DNase I and incubated at 37°C for 6 h before being lyophilized. Cleavage of the O-linked fatty acids was achieved by mixing the digested cells with 200 µl of hydrazine under nitrogen at 37°C for 2 h. The sample was placed in an ice bath and the excess hydrazine was destroyed with 6 x 200 µl of cold acetone in dry ice added at 5 minutes interval, to prevent overheating of the solution. The deacylated LOS was isolated by centrifugation at 16,000 g for 15 min and the product washed again with 400 µl of acetone and centrifuged. The pellet was resuspended in 200 µl water, centrifuged and the LOS containing supernatant lyophilized.

## **2.20 Phenotypic Assays**

### **2.20.1 Growth assays to assess growth rate**

*C. jejuni* cells were initially harvested from MH agar plates grown for 24 h and washed in phosphate buffered saline. Cells were resuspended in MHB and used to inoculate 5 ml MHB overnight cultures to an initial optical density of 0.1 at 600 nm (OD<sub>600</sub>). Cultures were incubated at 42°C microaerobically for 12-16 h with agitation. Following this initial step, cells were harvested and used to inoculate fresh 10 ml MHB cultures to an initial OD<sub>600</sub> of 0.025. The test cultures incubated microaerobically with agitation over a 24 h period, with the OD<sub>600</sub> observed at 3 hourly intervals to 12 h, and then at 24, 36 and 48 h. Six biological replicates were performed for each strain at each time point. Viable counts were also performed. Growth assays were also performed in the presence of antibiotics. Standard deviation and error was calculated for each time point and Student's *t*-test was performed to assess significance. The formula used was

$$t = \frac{Y_1 - Y_2}{\sqrt{([SE_1]^2) + ([SE_2]^2)}}$$

where Y is the sample mean and SE is the sample standard error.

### **2.20.2 Sensitivity to antibiotics and detergents**

*C. jejuni* cells were initially harvested from MH agar plates grown for 24 h and washed in phosphate buffered saline. Cells were resuspended in MHB and used to inoculate 5 ml MHB overnight cultures to an initial optical density of 0.1 at 600 nm (OD<sub>600</sub>). Cultures were incubated microaerobically, for 12 -16 h with agitation. Following this initial step, cells were harvested and used to inoculate fresh 10 ml MHB cultures to an initial OD<sub>600</sub> of 0.05. The test cultures were supplemented with the antibiotics and detergents to various concentrations (see below) and incubated, microaerobically with

agitation over a 24 h period, with the OD<sub>600</sub> observed at determined intervals.

Sample No.	Treatment/final working concentration µg ml <sup>-1</sup>		
	SDS	Novobiocin	Polymyxin B
1	400	50	2
2	200	25	1
3	100	12.5	0.5
4	50	6.25	0.25
5	25	3.2	0.125
6	12.5	-	0.0625

**Table 2.11** showing the concentrations used in assessing sensitivity to antibiotics and detergents

### 2.20.3 Growth assays to assess outer membrane function

Growth assays to assess outer membrane function and growth with different iron sources were performed in triplicate by C. Miller (University of Leicester). *C. jejuni* cells were initially harvested from MH agar plates grown for 24 h and washed in phosphate buffered saline. Cells were resuspended in MEMα and used to inoculate 5 ml MEMα overnight cultures to an initial optical density of 0.1 at 600 nm (OD<sub>600</sub>). Cultures were incubated microaerobically, for 12-16 h with agitation. Following this initial iron-depletion step, cells were harvested and used to inoculate fresh 10 ml MEMα cultures to an initial OD<sub>600</sub> of 0.025. The test cultures were supplemented with 25 µM porcine haemin and incubated, microaerobically with agitation over a 24 h period, with the OD<sub>600</sub> observed at various intervals. Positive controls were performed in iron replete media using 10 µM iron sulphate and negative controls were performed using untreated media.

### **2.20.4 Invasion and adherence assays**

Adherence and invasion assays were performed with Caco-2 cells by Paul Everest (University of Glasgow; MacCallum *et al.*, 2006). CaCo-2 cells were seeded at  $1 \times 10^5$  cells per well in a 12 well plate (Costar) and allowed to grow until confluent. The confluent monolayers were washed and inoculated with 10  $\mu$ l of bacterial suspension of approximately  $1 \times 10^6$ , which was verified by optical density and viable counts. The infected monolayers were incubated for 3 h at 37°C in a 6% CO<sub>2</sub> humidified atmosphere to allow the bacteria to adhere and invade the cells. The cells were then washed three times and extracellular bacteria were killed with 250  $\mu$ g/ml gentamicin. The monolayers were then re-incubated for 2 h prior to cell lysis and viable plate counts.



## Chapter 3: Genetic diversity in the LOS biosynthesis region

### 3.1 Introduction

#### 3.1. Variation in the LOS core biosynthesis gene cluster of *C. jejuni*

The role of LOS in the pathogenesis of campylobacteriosis remains unclear although specific *C. jejuni* serotypes have been linked to the development of GBS, e.g. the clonally related serotype HS:19 (Nachamkin *et al.*, 2001). LOS plays a role in the development of GBS as the epitope is thought to be the molecular mimic of host gangliosides linked to the autoimmune reaction leading to GBS.

As previously discussed, current serotyping methods rely on the capsule antigen as the major sero-determinant and therefore do not provide a method for typing the LOS structure of the strain. It is well established that LPS/LOS core structure reflects the genes present in the strain genome. *E. coli* and *Salmonella*, for example, have five different sets of genes for the five different outer core structures (Heinrichs *et al.*, 1998). Molecular methods can be used to characterise the gene content at the LOS region of *C. jejuni* and therefore inform LOS structure. These molecular methods may also provide extra strain differentiation and provide a system for strains that cannot be differentiated by traditional serotyping methods. In *C. jejuni*, there is much greater inter-strain variation in LOS core structures than in *E. coli* and *Salmonella* spp. and several studies have been performed to elucidate the genetic basis of this variability. Initial studies examined the variation in the three regions of the protein glycosylation and LOS locus, and showed variation between *lpxL* and *waaV*, the absence or presence of *wlaJ* and insertions between *waaF* and *gmhA* (Wood *et al.*, 1999; Oldfield, 2000). Microarray studies using elements based on the first sequenced strain (NCTC 11168) showed that the genes involved in the biosynthesis of the LOS core were highly divergent or absent in a number of strains compared to the sequenced strain (Dorrell *et al.*, 2001).

Gilbert and colleagues characterised the outer LOS core region between *lpxL* and *waaF* in ten strains, including 8 Penner serostrains (Gilbert *et al.*, 2000; 2002). The ten strains were organised into three classes based on gene arrangement and content, the results are summarised in Figure 3.1, 3.2 and 3.3. The gene content of the strains they sequenced differed from NCTC 11168 in a number of ways: one novel gene, *orf11*, was identified; three genes were missing; and *neuA1*, which is a fusion of two genes in NCTC 11168, could be found as two separate ORFs. Another study used a PCR strategy to examine a collection of GBS/MFS isolates for the occurrence of the *cstII* sialyltransferase gene, with the aim of investigating the association between *cstII* and the expression of LOS epitopes that mimic human gangliosides (van Belkum *et al.*, 2001). There was no difference in gene frequency between GBS/MFS strains and the control strains, yet isolates displaying a GQ1B ganglioside epitope all possessed the *cst* gene, indicating this gene may be a prerequisite for the synthesis of this epitope.

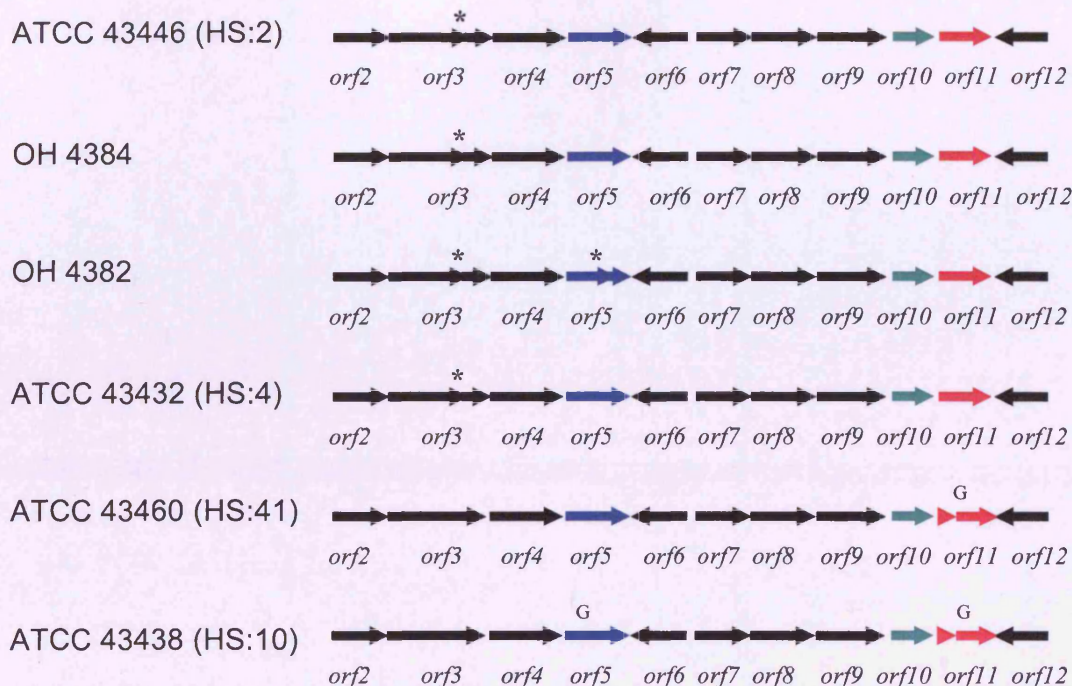
More extensive studies of over 50 laboratory strains revealed that the differences in gene content allowed the strains to be further divided into seven classes (Gilbert *et al.*, 2002; Poly *et al.*, 2004; Millar, 2003).

Parker *et al.* (2005) performed an extended study looking at the LOS gene content of over 100 strains by long range PCR and sequencing in an effort to correlate LOS gene type with Penner serotype, and LOS class and isolate source. This study found that all the GBS associated strains and the majority (64%) of other clinical and environmental isolates belonged to classes A, B and C. This work also found a strain with novel gene content and established a new class, Class H, which resulted from insertion and deletion events from class D or 81116-like strains.

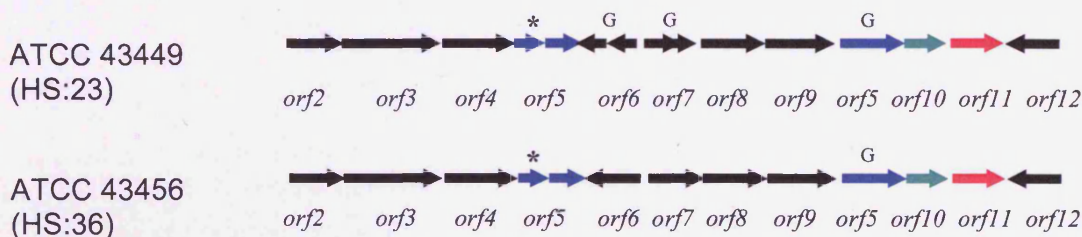
Figure 3.4 is a schematic diagram where the classes with similar gene content have been rearranged into groups, making it easier to highlight the differences in gene content (Karlyshev *et al.*, 2005). Group 1 includes classes A, B and C. Although these classes differ in cluster size, they are similar in gene content with repeated genes (class B has a second copy of *cgtA*, *cgtA-II*), gene insertions (class A and B have *orf11* inserted, class C *wlaSB*, *wlaO* and *wlaP*). Group 2, including *C. jejuni* NCTC 11828, was shown to have

**Figure 3.1: Genetic organisation of the LOS biosynthesis loci of several *C. jejuni* strains.** Reproduced from Gilbert *et al.* (2002) with arrows representing ORFs, and indicating the direction of transcription. Conserved genes are represented in black. Genes that are unique to class A and B are shown in pink, and the orange arrows represent genes that are unique to class C. Orf5, indicated by blue arrows, is found only once in class A gene arrangements, twice in class B gene arrangements and as an in-frame fusion with orf10 (represented by green arrows) in class C. Asterisks are shown where a premature translational stop was observed and G or C indicate the presence of homopolymeric tracts within the ORF. Phase-variable ORFs which were predominantly observed out of frame are shown as two ORFs. The gene nomenclature used by Gilbert and colleagues (2002) are as follows: orf2, *lpxL*; orf3, *cj1135*; orf4, *cj1136*; orf5, *cgtA*; orf6, *cgtB (wlaN)*; orf7, *cst-II*; orf8, *neuB1*; orf9, *neuC1*; orf10, *neuA*; orf12, *waaV*; orf14, *cj1137c*; orf15, *cj1138*; orf16, *cj1144c/cj1145c*. Penner sero-strains are indicated.

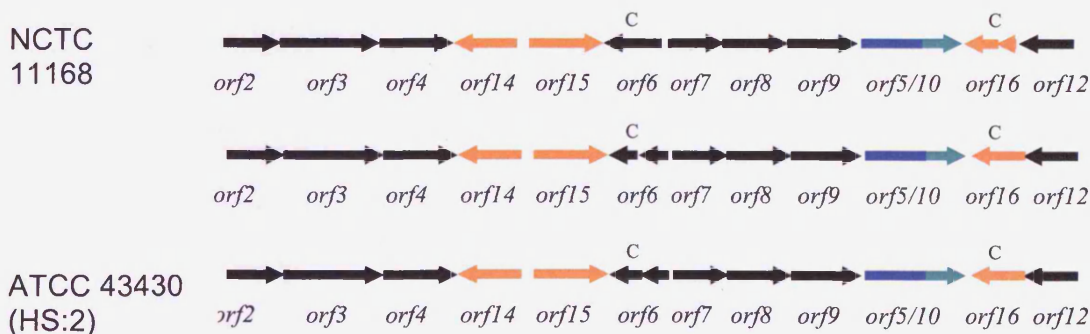
**'Class A'**



**'Class B'**

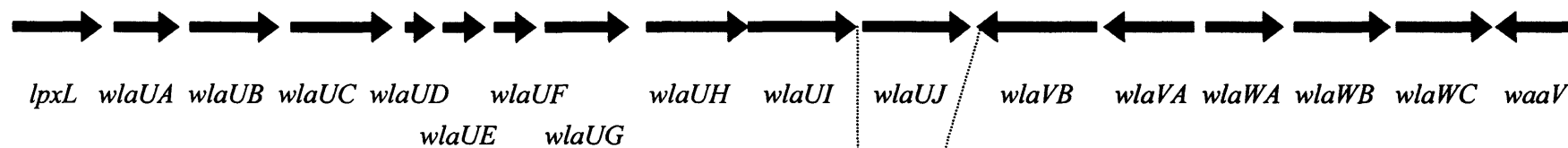


**'Class C'**

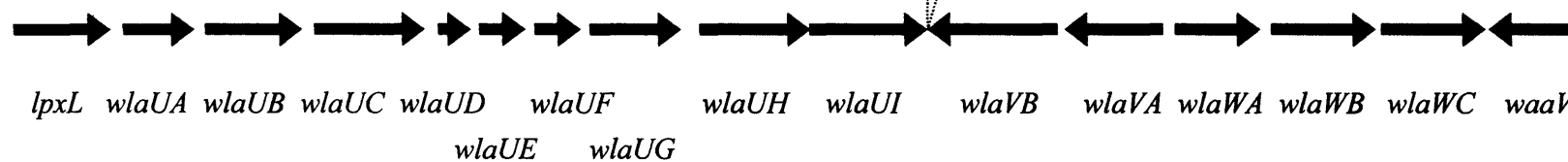


**Figure 3.2: Schematic diagram of the *lpxL* – *waaV* regions of *C. jejuni* 81116, 2F68, C11, O3, 21F185, and 13156**, reproduced from Millar (2003). Arrows indicate ORFs and the direction of their transcription. Gene names are suggested based on the existing *wla* nomenclature (Fry *et al.*, 1998; Fry *et al.*, 2000b).

*C. jejuni* 81116, 2F68, C11 (class D)



*C. jejuni* O3, 21F185, 13156 (class E)



**Figure 3.3: Schematic diagram of the region between *lpxL* and *waaV* in *C. jejuni* O73, RM1221 and LIO87**, reproduced from Millar (2003). To enable direct comparison, the notation used in the diagrams is based on the alphabetical notation given to the *C. jejuni* O73 loci. The notation follows that submitted to GenBank for *C. jejuni* LIO87 : *orfA*, *orf3*; *orfG*, *orf4*; *orfB*, *orf5*; *orfC*, *orf6*; *orfD*, *orf6*; *orfE*, *orf7*; *orfF*, *orf8*. The *C. jejuni* RM1221 LOS biosynthesis cluster contains genes similar to the O73 LOS gene cluster. The notation used by TIGR for the RM1221 genes are as follows: *orfA*, *orf346*; *orfB*, *orf345*; *orfC*, *orf344*; *orfD*, *orf344*; *orfE*, *orf343*; *orfF*, *orf341*. The genes *orfC* and *orfD* are found as a single ORF in strains LIO87 and are fused in RM1221. *Orf A* of LIO87 only contains the first ~700 bp of the 1550bp full-length gene. A novel open reading frame, *orf G*, represented in blue replaces the 3' end of *orfA* in this strain. Homopolymeric tracts, indicated by the yellow stars, are found in all three strains in the same places.

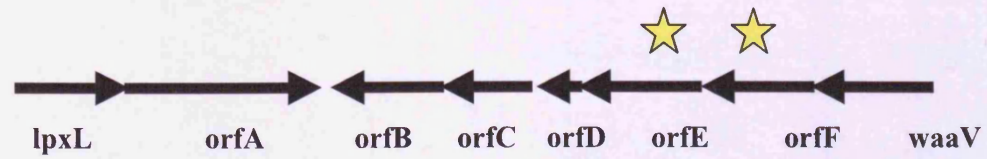
---

Chapter 3: Genetic diversity in the LOS biosynthesis region

---

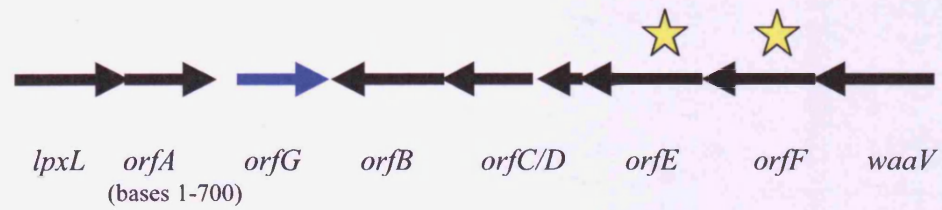
*C. jejuni* 073 and RM1221

(class F)



*C. jejuni* LIO87

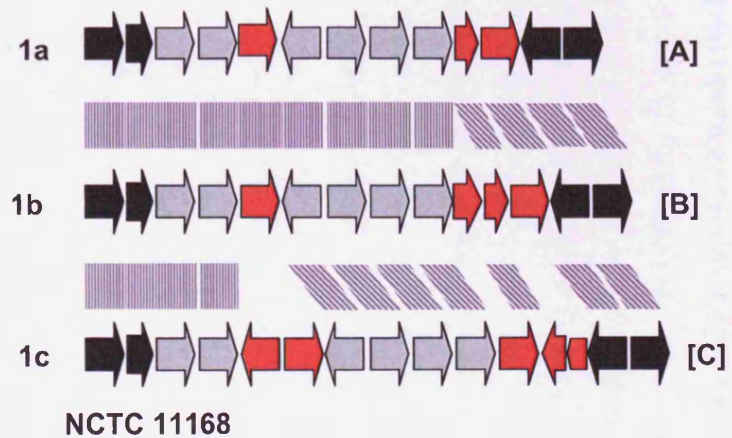
(class G)



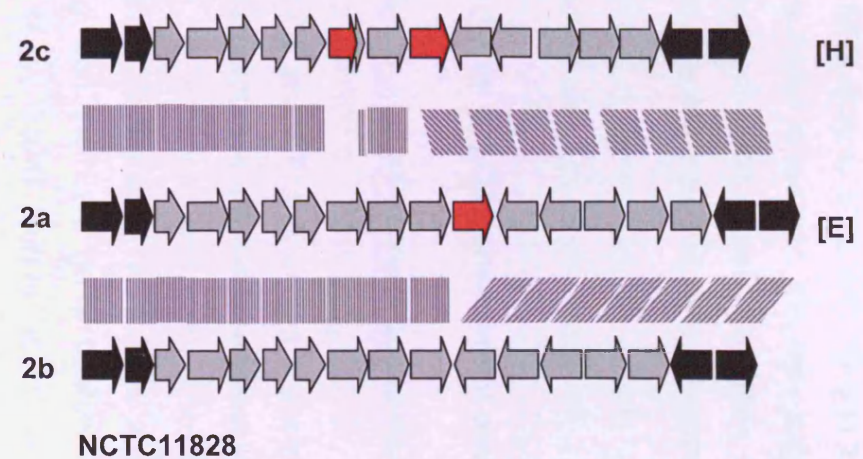


**Figure 3.4: Schematic diagram showing the different LOS gene content arranged into an organisational classification based on gene content similarity.** The black arrows represent genes conserved throughout the groups, such as *waaC* and *waaV*. Grey arrows represent genes that are shared by members of the same group. Genes which are divergent are highlighted by red arrows. Regions of similarity in the group are highlighted by the stripes. The boxes indicate the regions amplified by the primers. Letters in parentheses indicate nomenclature adopted by Gilbert *et al.* (2002), Poly *et al.* (2004), Millar (2003) and Parker *et al.* (2005). Example strains are indicated below each group. Adapted from Karlyshev *et al.*, 2005.

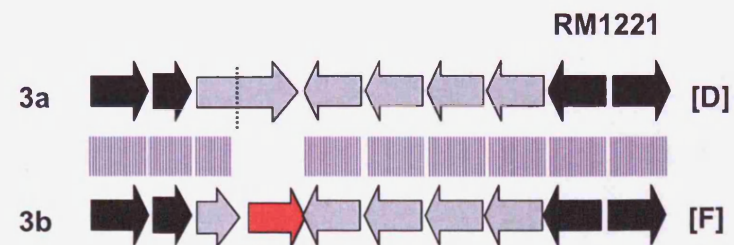
### Group 1



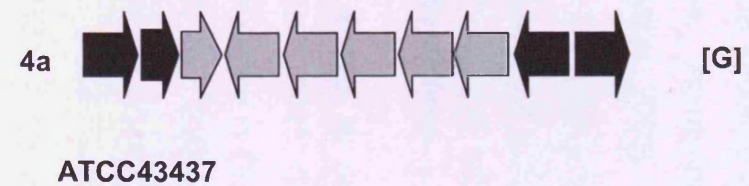
### Group 2



### Group 3



### Group 4



a completely different set of genes at the sequence level *in silico* (Oldfield, 2000) and PCR analysis showed that in some strains *wlaUJ* was missing (Millar, 2003, Poly *et al.*, 2004). Group 3 also contained novel genes and as these are fewer in number, this may suggest a simpler LOS structure or the involvement of another polysaccharide biosynthesis locus. Approximately 60% of the strains analysed were categorised as Group 1 (Millar, 2003).

### **3.1.2 Variation in LOS gene content and Guillain Barré Syndrome.**

Group 1 strain core oligosaccharides have been shown to be the molecular mimics of host gangliosides, such as GM1, GM2, GM3, GD1A, GD1C, GD3 and GT1A as shown overleaf in Figure 3.5. As stated in Chapter 1, several serotypes are over-represented amongst isolates from GBS cases in Japan, the Americas and South Africa including the clonal serotypes HS:19 and HS:41. Sub-group 1A strains are represented in these serotypes indicating that the LOS gene content is important in GBS development (Gilbert *et al.*, 2004). Sub-group 1B strains have also been linked to the neuropathy, MFS (Godschalk *et al.*, 2001b). These sub-classes are linked to specific neuropathies due to the presence of certain genes or polymorphisms. The presence of three genes; *cstII*, *cgtA* and *cgtB* in the LOS gene cluster was found to be strongly associated in GBS associated isolates (Nachamkin *et al.*, 2002b). The ability of *C. jejuni* to express many different ganglioside-like epitopes in the LOS outer core, within the same sub-class, indicates that mechanisms other than gene content variation exist to cause changes that are reflected in the LOS structure. These mechanisms include allelic variation and variation in homopolymeric tracts.

### 3.1.3 Types of variation in LOS gene content.

#### Sequence variation

Variation in the LOS gene content occurs on various levels. One of the most notable observations from the whole genome sequencing project of NCTC 11168 (Parkhill *et al.*, 2001) was the occurrence of short mononucleotide homopolymeric tracts predominantly in the carbohydrate biosynthesis loci. Slip strand mispairing of these tracts is thought to affect the expression of the associated gene. Linton *et al.* (2000b) showed that the phase variation of the homopolymeric tract in the  $\beta$ -1,3-galactosyltransferase, *wlaN* (*cgtB*) causes alternating production of GM<sub>1</sub> or GM<sub>2</sub> – like epitopes, as it affects the functionality of the two domain CgtB protein. In the sub-group 1B strain, the phase variation of two genes allows the production of four different LOS structures (Guerry *et al.*, 2002). This is important in virulence as mutants in one of these genes showed an increased ability to invade *in vitro* when compared to the wild type strain (Guerry *et al.*, 2002). In this strain, *cgtA* has also been shown to undergo phase variation *in vivo* causing production of GM<sub>2</sub>, GD2 or GD<sub>1B</sub> – like epitopes (Prendergast *et al.*, 2004).

Genes can also be inactivated by a single base mutation without phase variation such as in the case of the *cgtA* gene of strain OH4382 where a base deletion causes the production of a severely truncated protein (Gilbert *et al.*, 2005). This causes the outer core to be truncated at the first galactose residue. Some variants of *wlaNA* express a single domain  $\beta$ -galactosyltransferase in sub-groups 2A, 3B and 4A, but a two domain enzyme in groups 1 and 3A. In group 1 strains, a single base deletion leading to a truncation of the protein and the loss of the second domain which causes the  $\beta$ -1,2-glucose on the second heptose residue to be missing (Gilbert *et al.*, 2002).

Other examples of gene inactivation include that of *wlaN* and *wlaSC* in the strain ATCC 43430 (group 1C) where there are frameshift mutations. Other point mutations cause gene silencing in *wlaSC* and block the glycosyltransferase activity of the gene fusion *cgtAll-neuAI*. SNPS that lead to transcriptional differences might also lead to differences in structure

### Allelic variation










The group 1 strains contain four different  $\beta$ -1,4-*N*-acetylgalactosaminyltransferase genes that show allelic variation and encode CgtA, CgtA-I, CgtA-II and NeuAI/CgtA-II. Although these proteins have functional homology, there is only 34% residue conservation between the polypeptide sequences. The enzymes have different acceptor specificities; the NCTC 11186 CgtA will interact with mono- and di-sialylated acceptors, whereas ATCC 43438 will only interact with unsialylated acceptors (Gilbert *et al.*, 2002). The product of *cstII* is also an example of allelic variation. The enzymes vary in their activity having either  $\alpha$ -2,3-sialyltransferase activity or bifunctional  $\alpha$ -2,3-sialyltransferase and  $\alpha$ -2,8-sialyltransferase activities (Gilbert *et al.*, 2002).

### Mechanisms of variation

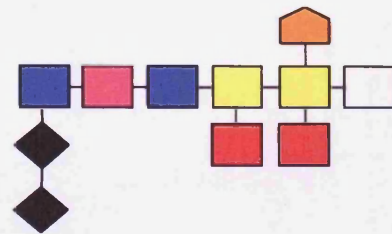
Lateral transfer and recombination also occurs within the LOS gene cluster. Gilbert *et al.* (2005) have proposed the evolutionary steps that resulted in the development of group 1B strains from group 1A strains, with the strain ATCC 43438 as an intermediary step. The repetition of gene *cgtA* is proposed to be due to gene duplication and the insertion of the duplicated gene is precise as there has been the addition of one base before the start codon.

The strain GB11, which was shown to be genetically related to NCTC 11168 by molecular typing and serotyping methods was shown to have a completely divergent LOS cluster compared to the sequenced strain. Sequencing analysis showed that the LOS cluster was nearly identical to that of the group 1A strain, ATCC 43446 indicating that horizontal transfer had potentially occurred between *gne* and *gmhA*. (Gilbert *et al.*, 2004).

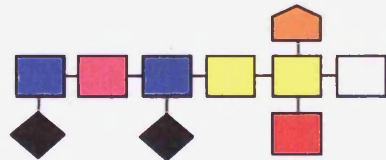
**Figure 3.5 A schematic diagram of LOS core structures** from various strains and detailing the ganglioside that the outer core mimics. The key is detailed below. Adapted from (Aspinall *et al.*, 1993; 1994; 1995, 1993; Oldfield *et al.*, 2002; Gilbert *et al.*, 2000; Millar, 2003)

 D-glucose	 D-galactose
 phosphorylethanolamine	 <i>N</i> -acetylneuraminic acid
 <i>L</i> -glycero- <i>D</i> -manno-heptose	 <i>N</i> -acetylgalactosamine
 3-deoxy- <i>D</i> -manno-octulosonic acid (Kdo)	 <i>N</i> -acetylglucosamine
	 3-amino-3,6-dideoxy- <i>D</i> -glucose

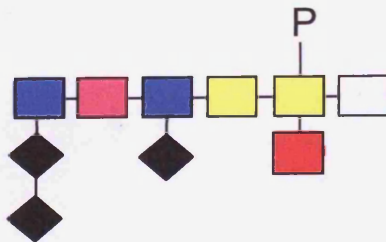




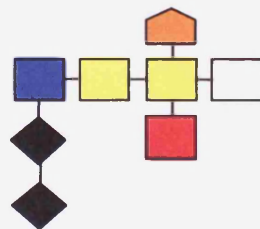
ATCC 43438:  
GD1c mimic



ATCC 43432:  
GD1a mimic

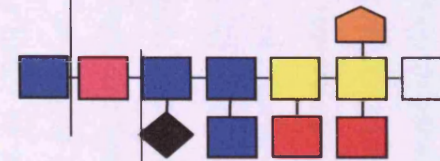


OH4384:  
GT1a mimic



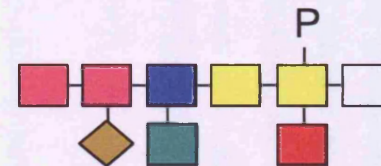
OH4382:  
GD3 mimic

ATCC 43429:  
GM2 mimic



NCTC 11168:  
GM1 mimic

ATCC 43430:  
GM3 mimic



HS:3

### 3.1.2 Multi Locus Sequence Typing and LOS cluster gene content

Phenotypic and molecular typing methods have been discussed previously in Chapter 1, section 1.1.3. Serotyping has a number of disadvantages that have been addressed to some extent by multi locus sequence typing (MLST). Primarily in relation to the analysis of LOS core gene content, serotyping has been shown to be reliant on the capsule as the main sero-determinant, although it was previously thought to be related to LOS and LPS (Karlyshev *et al.*, 2001a; 2001b). Other disadvantages exist, for example, serum can vary from batch to batch leading to potentially inconsistent results, both capsule and LOS are phase variable leading to unpredictable results and there are at least two different methods used for serotyping which are not directly comparable. In contrast, MLST relies on standard, reproducible molecular techniques and results can be compared across many laboratories.

MLST is one of the many techniques used to differentiate bacterial strains on the basis of genetic polymorphisms. This typing method is based on measuring the sequence variation in 500-700bp of seven housekeeping genes. Strains with the same DNA sequence are grouped into the same ST or sequence type, and related STs are grouped into clonal complexes. This method enables the examination of conserved genes and an insight into long-term epidemiology and the genetic structure of a population. Comparing MLST data together with gene content information can provide valuable data about the genetic context. MLST data shows that *C. jejuni* has a weakly clonal population structure, and strains are genetically diverse as lateral gene transfer is common (Dingle *et al.*, 2001a). Millar (2003) showed that of the strains typed into LOS gene content class, one third of the strains belonged to one clonal complex independent of LOS class. Strains that are seen to be genetically related by various typing methods may significantly diverge at the LOS biosynthesis locus and this is due to horizontal transfer (Gilbert *et al.*, 2004)

The sequence type data may also be used to relate the genetic background of a strain to its LOS gene content. The MLST and LOS data in combination may also provide context for other data, such as where the



isolate was first found and the gene content of other polysaccharide biosynthesis loci.

### **3.1.3 Aims of study**

The main aim of this study was to design and establish a PCR based screen to be used to determine the extent of variation at the LOS gene cluster. Initially the design of this PCR screen encompassed the LOS cluster variation discovered by Gilbert *et al.* (2002) and Millar (2003). The PCR screen was then tested against control strains and used to assess the variation of the LOS gene cluster in 50 clinical strains.

## **3.2 Analysis of LOS gene content of 50 'Campylobacter Sentinel Surveillance Scheme' strains**

The strains used to examine the extent of LOS gene cluster variation were selected from the Campylobacter Sentinel Surveillance Scheme collection (CSSS, 2005). These clinical strains were collated from laboratory confirmed cases of *C. jejuni* enteritis between 2000 and 2002, along with epidemiological data relating to exposure and illness onset.

### **3.2.1 Design of PCR screening method and primers.**

A PCR-based method was used as a preliminary screen for variation at the LOS locus. This approach has already been used in other studies (Oldfield, 2000, Millar, 2003) and presents an established method to provide an indication of LOS biosynthesis gene cluster content. This information can then be used to establish a new screening method which will allow more accurate characterisation of this locus. This preliminary approach involved using the class system, as described in Figures 3.1, 3.2 and 3.3 to determine variation at this locus (Gilbert *et al.*, 2002, Millar, 2003, Poly *et al.*, 2004). Sequences from GenBank (Accession numbers: Sub-Group 1A, AF400048, AF215659; Sub-Group 1B, AF401529, AF401528; Sub-Group 1C, AL139077;

Sub-Group 2A, AF411225, AF343914 and AJ131360; Sub-Group 2B, AY501976; Sub-Group 3A, RM1221 contigs 576681-585991; Sub-Group 3B, AF400669 and Sub-Group 4A, AY436358) were used to design class specific primers which were designed to provide unique amplicons, by band presence or amplicon size, for each class and sub-class, thereby providing an indication of gene content (see Figure 3.6, and table 3.1 which indicates the amplicon sizes expected). This strategy also decreased the number of PCR amplifications required per strain. Each strain was tested against a panel of 18 primers, as detailed in Appendix 1, in various combinations.

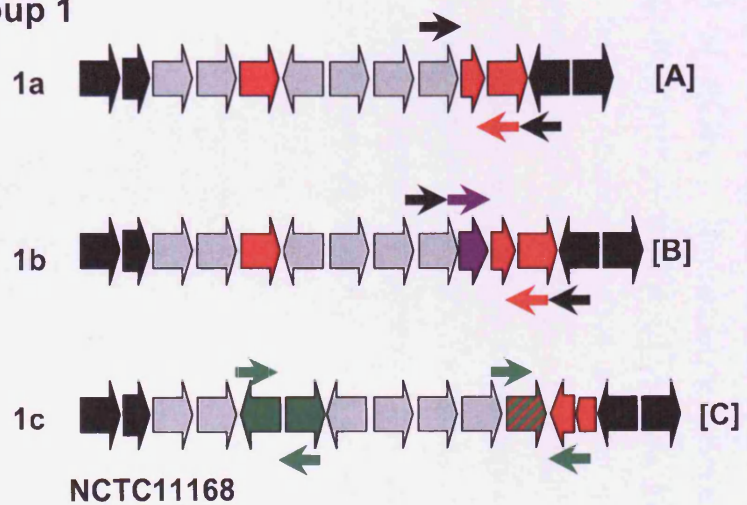
Positive controls were included in each set of PCRs and these include strains for which class has been fully determined by primer walking, long range PCR, sequencing and database comparisons (Millar, 2003). It was not possible to obtain a positive control for sub-class G or H. An internal positive control primer set was also included for each strain and were designed to include the gene, *gne*. Where a control strain was available, primers were tested and optimised to ensure they were working.

**Table 3.1: Amplicon sizes (bp) expected for each LOS sub-class.** The name indicated above the amplicon size represents the primer name and the sub-group is indicated in brackets.

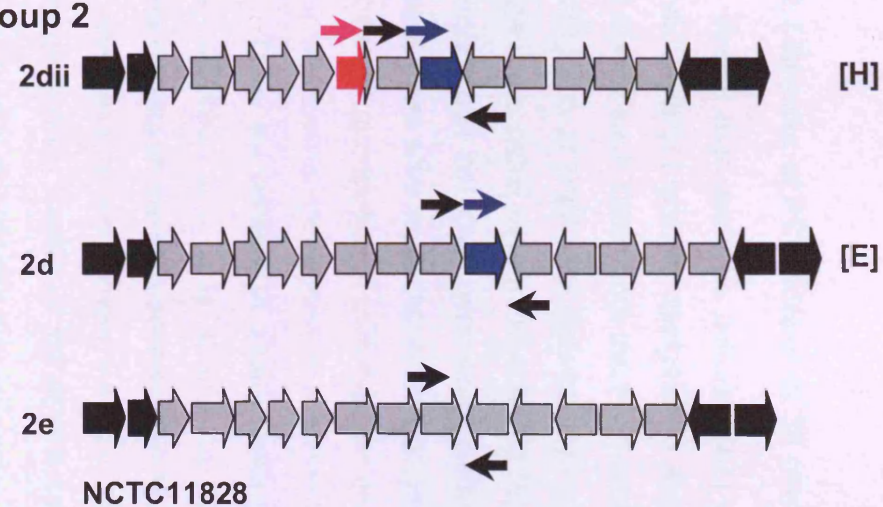
Sub-group	PCR amplicon (Size in bp)										<i>gne</i>
	1A	1BI	1BII	1CI	1CII	(2A)D	(2B)E	(3A)F	(3B)G	(4A)H	
1A	1977	832	0	0	0	0	0	0	0	0	723
1B	2923	1778	1332	0	0	0	0	0	0	0	723
1C	0	0	0	3233	1991	0	0	0	0	0	723
2A	0	0	0	0	0	1173	2107	0	0	0	723
2B	0	0	0	0	0	0	1087	0	0	0	723
3A	0	0	0	0	0	0	0	2383	1542	0	723
3B	0	0	0	0	0	0	0	0	1816	0	723
4A	0	0	0	0	0	0	0	0	0	829	723

**Figure 3.6: Schematic diagram of LOS gene content groups indicating primer positions.** Primers are indicated by the smaller arrows. Sub-class specific primers are indicated by coloured arrows and those that bind all strains in the group are coloured black. The black arrows represent genes conserved throughout the groups, such as *waaC* and *waaV*. Grey arrows represent genes that are shared by members of the same group. Genes which are divergent are highlighted by orange arrows. Regions of similarity in the group are highlighted by the stripes. The boxes indicate the regions amplified by the primers. Letters in parentheses indicate nomenclature adopted by Gilbert *et al.* (2002), Poly *et al.* (2004), Millar (2003) and Parker *et al.* (2005). Example strains are indicated below each group. Adapted from Karlyshev *et al.*, 2005.

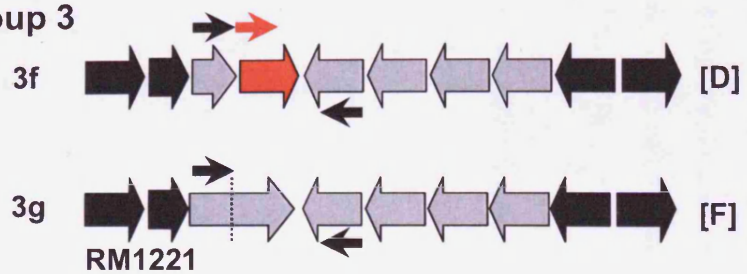
**Group 1**



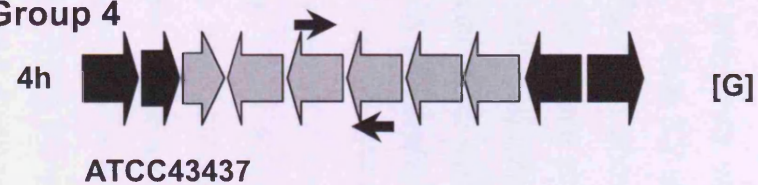
**Group 2**



**Group 3**



**Group 4**



### 3.2.2 Results of PCR screen of 50 clinical strains

All strains that could be cultured (47) produced amplicons with the internal positive-control primers designed to amplify, *gne*. Table 3.2 summarises the results for each strain from the PCR screen and Figure 3.7 shows a typical gel photograph of amplicons derived from a 96-well PCR. Each strain was tested against the panel of primers until the result was confirmed twice. Anomalous results, such as novel gene combinations or unclassified strains (discussed below) were also repeated with new genomic DNA preparations. Figure 3.8 shows the distribution of LOS type amongst the 47 strains. All the groups and their subclasses are present in the strain collection tested except for subclass 4A. Thirty six percent of strains belong to group 1 and sub-group C being most common and these strains are all variants with similarity to the LOS gene content of the sequenced strain NCTC 11168. There are also a number of new variants with amplicons from different groups, indicating new gene arrangements. Twelve of the strains could not be typed by group indicating potential new subclasses and although this could have been due to technical problems, this was unlikely as positive-control strains were included in each set of PCRs. Strain-specific primers were redesigned and the PCRs repeated, alongside positive controls, to check that the primer design was not at fault. Amplicons were still not produced highlighting that there are variations within the LOS core biosynthesis region.

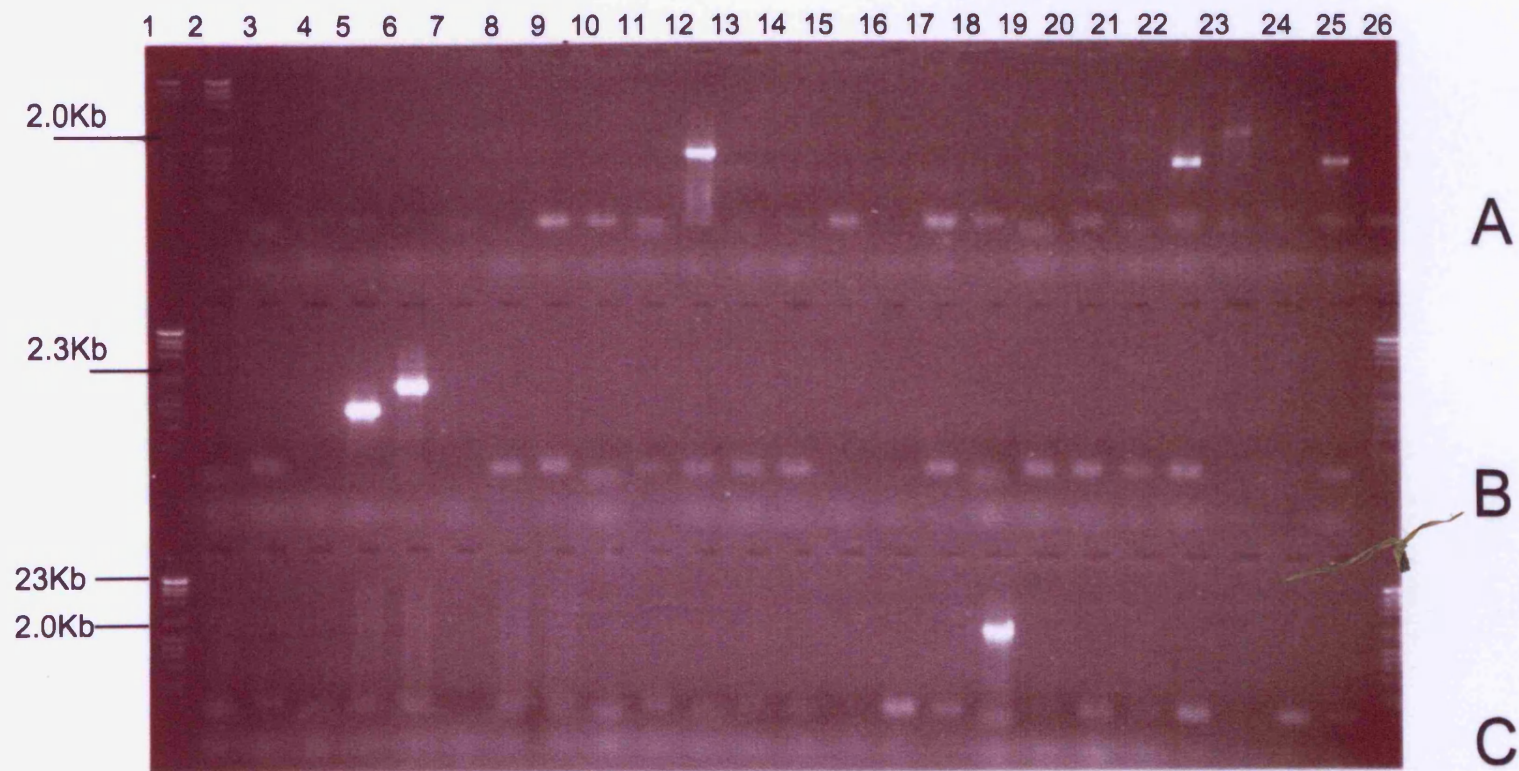
Seven strains were shown to produce amplicons corresponding to more than one sub-group and were therefore novel gene combinations. Figure 3.9 shows the amplicons produced by one of these strains, 44406. This strain produces amplicons with primers corresponding to sub-group 1B, 1C and 2B.

**Table 3.2: Summary of PCR screen results for each strain.** The table also details the serotype of each strain. UK; unknown, and the strains 39902, 45908 and 46036 are not listed as they were uncultureable.



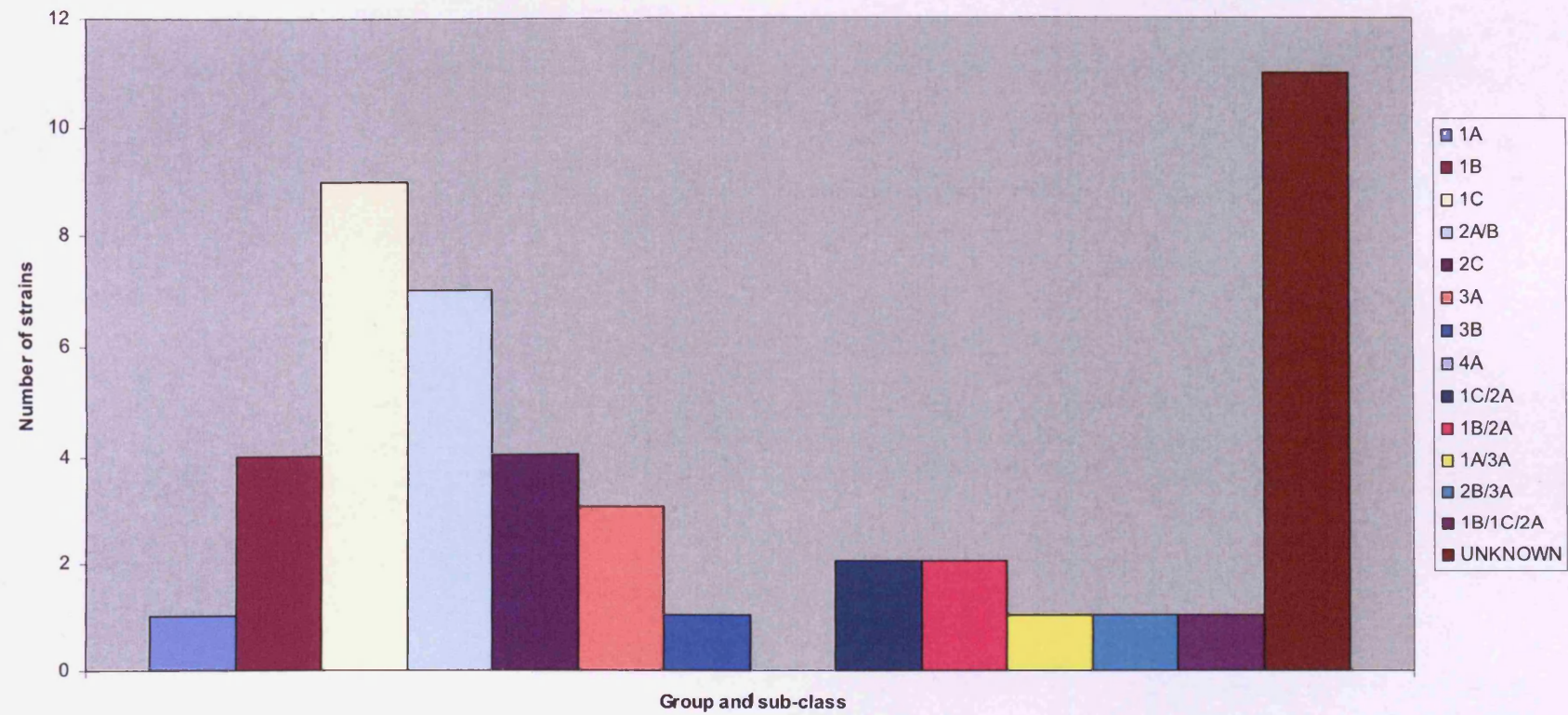
Strain Number	Heat Stable Serotype	Phage type	LOS Sub-group
34086	31	1	UK
34218	2	1	UK
34565	5	34	UK
34806	50	34	1B/3A
35303	UK	UK	1C
35451	4	1	2A
35503	12	2	1B
36670	11	1	3A
37531	50	5	2B
37895	5	1	3A
38577	19	33	1C
38608	37	1	2A
38625	50	1	UK
39271	6	1	1C
39864	21	44	2B/3A
39893	2	36	1C
39918	21	1	2B
40973	50	34	1B/2A
41803	4	1	1C
41999	50	6	1B
43771	13	2	2B
44253	13	33	1C
44406	5	34	1B/1C/2A
45283	2	33	1C
45385	12	1	2A
45600	13	1	2A
47185	12	2	UK
48643	37	1	1C
49068	4	34	1B/2A
50612	5	1	2A
50702	11	1	UK
51566	31	39	UK
51585	37	1	1C/2A
52831	12	2	2A
53305	21	1	UK
54386	18	2	UK
54471	11	2	2B
54590	11	1	1B
57073	6	39	1C/2A
57507	13	1	1C
58526	50	34	1B
58728	6	2	2A
59653	19	2	1A
60087	31	1	3A
60238	18	1	3B
60319	18	2	UK
60584	50	5	1B
61666	2	1	UK

**Figure 3.7 Gel photograph showing the typical results of the LOS sub-group PCR screen.** Panel A: Lanes 3-10, 34086: unknown, lanes 11-18: 34565: group 1B, lanes 19-26: 34806: 1C/3B. Panel B: Lanes 2-9, 41803: 1C, lanes 10-17: 45385: unknown, lanes 18-25: 48643: unknown. Panel C: Lanes 2-9, 45600, unknown; 10-17 52831, unknown; 18-25, 59653, 1A. Positive control strains were also included (data not shown).



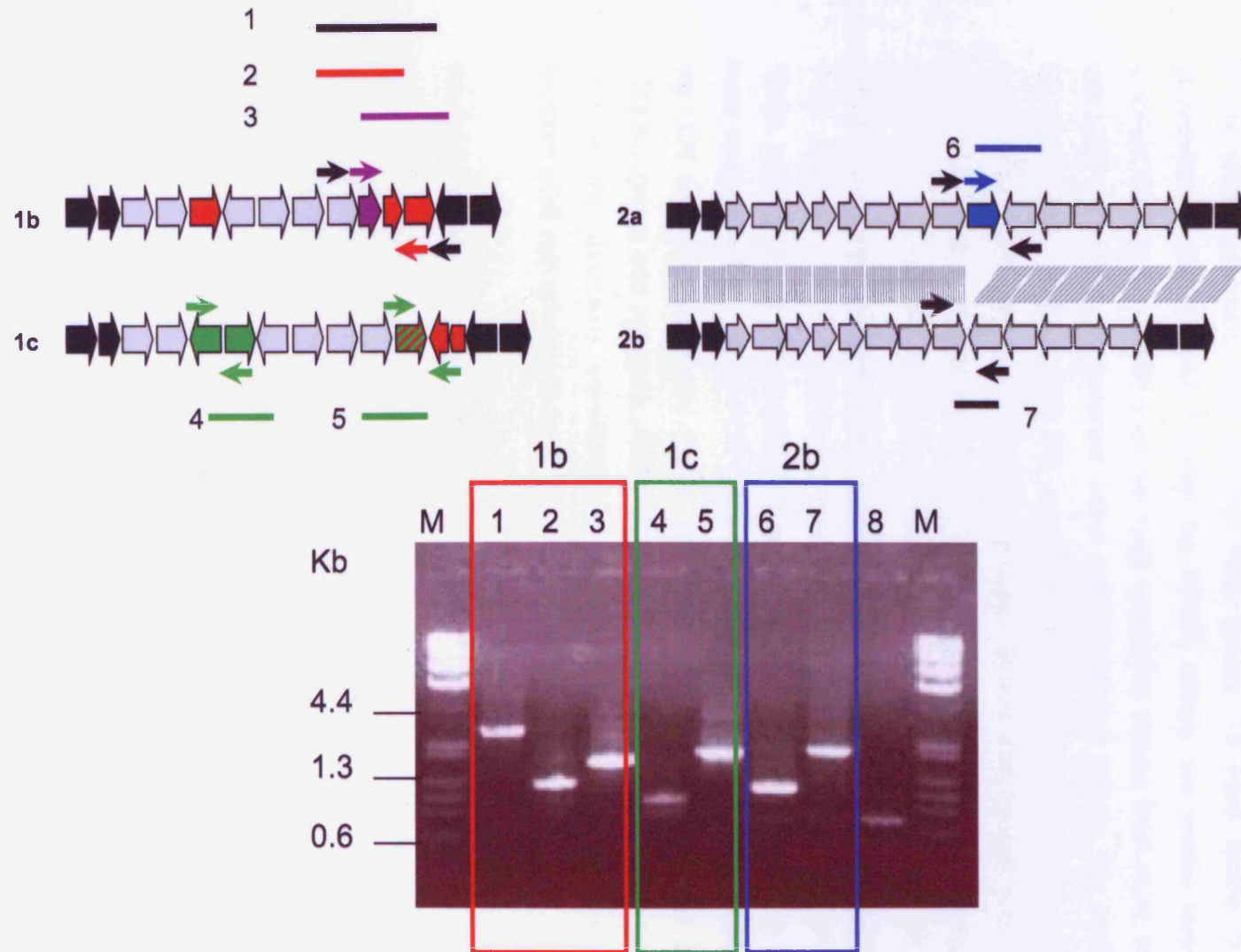
**Figure 3.8 Distribution of LOS gene content sub-group in 47 strains.**

Distribution of LOS gene content type amongst 47 strains



**Figure 3.9 Gel Photograph showing the results for strain 44406**, where the group specific primers annealing to sub-groups 1B, 1C and 2B produce amplicons. Lanes 1, 2 & 3 correspond to sub-group 1B, lanes 4 & 5 correspond to sub-group 1C, lanes 6 & 7 indicate positive results with primers for sub-group 2B and lane 8 is the positive control, *gne*. Schematic diagrams showing the primer placement on the corresponding sub-class is shown above the gel photograph.





### **3.2.3 Correlations between Heat Stable serotype and LOS PCR results.**

In the strain collection used for this PCR screen, 13 Heat Stable (HS) serotypes are represented. Except for HS:50, strains are evenly spread amongst the serotypes with each serotype containing strains from more than one LOS sub-group, for instance HS:4 contains strains: 35451, 2A; 41803, 1C and 49068; 1B/2A. With the exception of HS:6 and HS:11, the other serotypes contain only one or two LOS groups. Strains that contain potential novel LOS gene combinations, except for 39271 (2B/3A) appear to belong to serogroups with strains belonging to either LOS sub-group, for instance strain 51585 (1C/2A) is serotype HS:37, along with strains 38608 (2A) and 48643 (1C). Serotype HS:50 contains 4 strains with LOS type 1B, one strain with a 1B/3A LOS type and an unclassified strain. The serotypes containing the most unclassified strains by PCR were HS:2 and HS:31 and determination of the LOS sub-group of these strains will allow confirmation of any link with LOS sub-group and serotype. Although the strain collection may be too small to draw any conclusive correlations, there may be a potential relationship between LOS sub-group and serotype.

### **3.2.4 Associations between LOS PCR results and phage type.**

Nine phage types were represented in this strain collection. The majority of strains belonged to phage type 1 (22 strains). Perhaps unsurprisingly this phage type also contained the majority of unclassified strains and contained strains with all the LOS sub-groups represented. Certain phage types, for instance type 5 and 33, appeared to contain strains with only one LOS sub-group. Phage type 34 appeared to contain strains that were classified as a 1B LOS sub-group either on its own or in novel combinations with other sub-groups. Determination of the LOS type of the unclassified strains will allow further interrogation of the relationship between LOS gene content and phage resistance.



### 3.2.5 Correlations between MLST sequence type and LOS class.

Sequence Types for the 37 strains shown in table 3.3 were compiled by N. McCarthy at the University of Oxford according to the methods previously described (Dingle *et al.*, 2001). The results have also been deposited in the *C. jejuni* PubMLST Isolate Database found at <http://campylobacter.mlst.net/>. Over 11 clonal complexes were represented in this strain collection and 44% of the CSSS strains belonged to either clonal complex ST21 or ST45, which are the two most common clonal complexes associated with human isolates in the UK (Dingle *et al.*, 2002). In both of these clonal complexes over 50% of the strains in the group are related by LOS sub-group, for instance ST21 includes strains 35303, 38577, 41803, 44253, 45283, and 57507 which are all LOS group 1C and ST45 includes strains 38608, 45385, 50612, 52381 and 58728 which all belong to LOS group 2A. There are five strains from the ST48 complex in this study and they all contain genes relating to the sub-group 1B LOS gene content type. Eight more ST complexes are represented in this strains set with the only other commonly disease associated ST complex ST22 being represented by one strain, 59653, which is also the only group 1A strain in the study. Strain 58526 was also discovered to belong to a new ST as it had a novel combination of alleles at the seven loci tested. Establishing a definitive correlation between ST complex and LOS gene content would require an extension of this study.

**Table 3.3 MLST results for each CSSS strain**, with numbers for each allele as allotted by the MLST database ([www.pubmlst.org](http://www.pubmlst.org)). The LOS biosynthesis gene classes are indicated and the LOS gene clusters that cannot be identified are shown as UK (unknown). The clonal complex and sequence type (ST) are shown with the individual genes variations: *asp*, aspartase A; *gln*, glutamine synthase; *glt*, citrate synthase; *gly*, serine hydroxymethyltranserase; *pgm*, phosphglucomutase; *tkt*, transketolase; *unc*, ATP synthase  $\alpha$  subunit. HS serotype is also shown were known. UT, untypable. Five strains were not assigned either a ST or a clonal complex as the sequencing data was not available at all loci.

CRU Ref	LOS GROUP	Serotype	Phage type	<i>aspA</i>	<i>glnA</i>	<i>gltA</i>	<i>glyA</i>	<i>pgm</i>	<i>tkt</i>	<i>uncA</i>	ST	CLONAL COMPLEX
34086	UK	31	1	9	2	4	62	4	5	6	257	ST257
34218	UK	2	1	2	1	1	3	2	1	5	21	ST21
34565	UK	5	34	UK	UK	UK	UK	UK	UK	UK	UK	UK
34806	1B/3A	50	34	33	39	30	79	113	47	17	860	ST828
35303	1C	4	1	2	1	12	3	2	1	5	50	ST21
35451	2A	12	2	2	1	1	3	2	1	5	21	ST21
35503	1B	11	1	4	7	10	4	1	7	1	45	ST45
36670	3A	50	5	9	2	4	62	4	5	6	257	ST257
37531	2B	5	1	2	4	1	2	7	51	5	205	ST48
37895	3A	19	33	9	2	4	62	4	5	6	257	ST257
38577	1C	37	1	2	1	1	3	2	1	5	21	ST21
38608	2A	50	1	4	7	10	4	UK	7	1	UK	ST45
38625	UK	6	1	4	7	10	4	1	7	1	45	ST45
39271	2B/3A	21	44	4	7	40	4	42	51	1	267	ST283
39864	1C	2	36	UK	7	10	4	42	25	1	UK	ST45
39893	UK	60	2	2	1	21	3	2	1	5	53	ST21
39918	2B	21	1	4	7	10	4	42	7	1	137	ST45
40973	1B	50	34	2	4	1	2	7	1	5	48	ST48
41803	1C	4	1	2	1	1	3	2	1	5	21	ST21
41999	1B	50	6	1	4	2	2	6	3	17	61	ST61
43771	2B	13	2	7	21	5	62	4	61	44	436	Unassigned
44253	1C	13	33	2	1	1	3	2	1	5	21	ST21
44406	1B/1C/2A	5	34	UK	UK	UK	UK	UK	UK	UK	UK	UK
45283	1C	2	33	2	UK	21	3	2	123	5	UK	ST21
45385	2A	12	1	4	7	10	4	1	7	1	45	ST45
45600	2A	13	1	7	2	5	2	10	3	6	5	ST353
47185	UK	12	2	4	7	10	4	1	7	1	45	ST45

CRU Ref	LOS GROUP	Serotype	Phage type	<i>aspA</i>	<i>glnA</i>	<i>gltA</i>	<i>glyA</i>	<i>pgm</i>	<i>tkt</i>	<i>uncA</i>	ST	CLONAL COMPLEX
48643	1C	37	1	7	17	2	15	23	3	12	51	ST443
49068	1B/2A	4	34	2	4	1	2	7	51	5	205	ST48
50612	2A	5	1	4	7	10	4	42	51	1	583	ST45
50702	UK	11	1	7	2	5	2	10	3	6	5	ST353
51566	UK	31	39	9	2	4	62	4	5	6	257	ST257
51585	1C/2A	37	1	7	2	5	2	10	3	6	5	ST353
52831	2A	12	2	4	7	10	4	1	7	1	45	ST45
53305	UK	21	1	4	7	10	4	42	7	1	137	ST45
54386	UK	18	2	8	10	2	2	11	12	6	354	ST354
54471	2B	11	2	9	2	4	62	4	5	6	257	ST257
54590	1B	11	1	9	2	4	62	4	5	6	257	ST257
57073	1C/2A	6	39	4	7	40	4	42	51	1	267	ST283
57507	1C	13	1	2	1	1	3	2	1	5	21	ST21
58526	1B	50	34	2	4	221	2	7	1	5	NEW**	ST48
58728	2A	6	2	4	7	10	4	42	51	1	583	ST45
59653	1A	19	2	1	3	6	4	3	3	3	22	ST22
60087	3A	31	1	9	2	4	62	4	5	6	257	ST257
60238	3B	18	1	3	1	5	17	11	11	6	49	ST49
60319	UK	18	2	8	10	2	2	11	12	6	354	ST354
60584	1B	50	5	2	4	1	2	7	1	5	48	ST48
61666	UK	2	1	2	1	1	3	2	1	5	21	ST21

### **3.3 Discussion**

#### **3.3.1 Validation of the PCR based assay**

Initial problems were encountered with the DNA extraction from this clinical strain set, with low concentration and poor quality DNA being extracted from a number of strains. The inclusion of an internal positive control in the PCR assay helped to assess the DNA quality for use as a PCR target.

A total of 20 primers were designed and evaluated for this PCR-based protocol. The primers designed to amplify within the *gne* open reading frame produced an amplicon with genomic DNA from all of the strains. The sub-group specific primers were tested with the control strains available and found to produce amplicons from the desired class with no cross-reactivity (data not shown).

The initial primers used to determine if the strain belonged to sub-group 2B worked sporadically and were therefore re-designed and re-evaluated prior to commencement of the screen. Following the publication of Parker *et al.* (2005), primers were designed to amplify a specific product to recognise sub-group 4A. No control strain was available to evaluate the sub-group 3B or 4A primers and despite optimising the 4A primer pair with temperature and salt gradients, no amplicons were produced with this strain set.

Parker *et al.* (2005) utilised a panel of 42 primer pairs to determine the LOS gene content of 123 strains. Eighty percent of the strains tested could be classified and two LOS sub-group variants were identified. This may indicate that a larger panel of primers would yield greater differentiation of LOS type in the initial screening PCR.

### 3.3.2 Distribution of LOS core gene loci

All the groups and sub-classes were represented in this clinical strain collection as predicted, with the exception of sub-group 4A. A total of 36 % of strains belonged to group 1 and sub-group 1C being most common. These strains are all variants with similarity to the LOS gene content of the sequenced strain NCTC 11168. There are also a number of new variants with amplicons from more than one different group suggesting potential new gene arrangements and new gene combinations, which has not been seen by other studies on LOS gene content (Parker *et al.*, 2006). This suggests that variation may also occur on a gene by gene level, reflecting the need of *C. jejuni* for the particular function of that gene.

Twelve of the strains could not be typed by group indicating potential new sub-groups and although this could have been due to technical problems, this was unlikely as positive control strains were included in each set of PCRs. The inability to detect a sub-group 4A strain may be unsurprising as only one strain was reported to have this gene content in the study of over 100 strains as reported by Parker *et al.* (2005). This strain was the only isolate examined that had been recovered from goats and although further work would need to be performed, this may be an indication of LOS gene content relating to ecological niche.

### 3.3.3 Correlations with LOS gene content, serotype and phage type

HS serotype information was available for all of the strains in this study. The HS serotype is thought to be based mainly on the capsular antigen and therefore should mainly reflect differences in capsule biosynthesis gene content. Perhaps as expected serotype does not match particularly well with LOS gene content, but a few trends can be observed. The strongest potential correlation exists with serotype and LOS sub-group, with all but one serotype only containing strains belonging to one or two LOS sub-groups. This may indicate that either LOS epitopes contribute to serotype, or that there is an association between LOS type and the capsule type that exerts influence on the serotype. This strain set is too small to draw any definitive conclusions, especially as confusion may be caused by the unclassified strains. A larger

strain set is needed to examine any correlation between LOS and serotype. A comparison of the gene content of the capsule biosynthesis cluster would also have to be performed to draw any firm conclusions.

The majority of strains in this collection were associated with phage type 1 (22 strains). This group also contained the majority of unclassified strains and all the LOS groups represented. Certain phage types, for instance type 5 and 33, appeared to contain strains with only one LOS sub-group. Phage type 34 appeared to group strains that were classified as a 1B LOS sub-group either on its own or in novel combinations with other sub-groups. Again analysis of further strains would need to be completed before any strong correlation could be drawn between LOS gene content and phage resistance. The mechanisms of interactions between bacteriophages and *C. jejuni* are unclear, although the capsule and flagella have been implicated in this process in a recent study (Coward *et al.*, 2006). This may further indicate that other surface loci that are subject to variation may also be potentially involved in this process.

### **3.3.3 Correlations between MLST and LOS biosynthesis gene type.**

In order to view the distribution of the different LOS biosynthesis gene clusters from an epidemiological context, MLST typing was performed on the majority of the isolates. The STs that have been identified thus far have been grouped into clonal-complexes of genetically related strains. A total of twenty percent of the isolates typed belonged to the ST21 complex, which was the largest clonal complex identified from 800 isolates (Dingle *et al.*, 2002). Perhaps due to the large nature of this complex, it is also the most genetically heterogeneous identified to date as it is composed of many sequence types. In partial agreement with the MLST results of Millar (2003), the majority of sub-group 1C strains appear to cluster in this complex. This could indicate that either sub-group 1C gene content is only acquired through recombination with strains from the ST21 complex, or LOS structures synthesised from the genes in the group1C loci provides a selective advantage for colonisation of the niches inhabited by the ST21 strains. The clonal complex ST21 is the most common complex associated with human gastroenteritis and is also

associated with GBS and MFS antecedent strains (Dingle *et al.*, 2001; Dingle *et al.*, 2002). The second most commonly associated human isolate clonal complex is ST45 and 20% of the CSSS strains belong to this ST complex. There are five strains from the ST48 complex in this study and they all contain genes relating to the sub-roup 1B LOS gene content type.

Complexes ST21, ST48 and ST206 are thought to form a supercomplex of closely related genotypes that undergo frequent horizontal exchange (Dingle *et al.*, 2002). The strains in this supercomplex exhibit a high level of antigenic diversity in relation to their Penner serotype and variation at the *flaA* short variable region. The results of the present study infer that either this is not the case at the LOS gene cluster or that exchange at the LOS cluster is disadvantageous. This may indicate that LOS is important in virulence but that variation in LOS does not contribute to virulence for *C. jejuni*. There are no strains representing the ST206 complex, but the majority of strains from the other two complexes have group 1 LOS gene content. To determine the relationship between LOS gene content and ST complex, an extended strain collection needs to be examined.

### 3.3.4 Conclusions

This study shows that all the LOS sub-groups (except sub-group 4A) are present in strains isolated from cases of gastroenteritis in the UK. As with previous studies, the results from this PCR screen infer that the majority of strains can be categorised as group 1. Unlike the previous studies by Millar (2003) and Parker *et al.* (2005), the percentage of strains belonging to this sub-group is much lower; 36% compared to ~60% and 64% respectively. This present study found that 24% of the strains examined could not be assigned a group or sub-class. The results of the study by Parker *et al.* (2005) are in agreement with this finding, with 22% of their strains being unclassified. Unlike the previous two studies, the current studies also suggested the presence of potential novel gene combinations within the LOS core gene locus and due to primer design this is unlikely to be due to mis-priming or the occurrence of the gene elsewhere in the genome.



Unlike the other studies, no correlations can be drawn with environmental niche as these strains were all clinical isolates related to gastroenteritis. For further correlations to be drawn between LOS type and serotype, the strain collection would need to be extended and the PCR-based assay should also be extended to include analysis of capsule type.

Future studies of this nature could also include a survey of capsule and flagellar modification genes as it may provide more information about the combinations of antigens required to produce the different HS serotypes. It is likely that the antisera is based on a mixture of antigens and not only based on reactions to the capsule. Information about all three loci would also indicate if the variable carbohydrate biosynthesis genes recombine freely or if certain combinations are not feasible or are selectively disadvantageous, limiting the combinations of antigens that can be expressed.

General predictions of LOS structure can also be made based on gene content information and the information from this study only provides a guide to LOS group type. The unclassified strains and those with novel gene combinations were further examined by primer walking, long range PCR and sequencing which will be discussed in Chapter 4.

## **Chapter 4: Determining the gene content of the unknown and novel strains**

### **4.1 Introduction**

#### **4.1.1 Summary of Results from PCR screen**

As described in Chapter 3, a PCR screen was developed to enable determination of LOS gene content group of 50 clinical strains. All of the cultured strains produced amplicons with the positive control primers for *gne*. All the groups and their subgroups are present in the strain collection tested except for sub-group 4A, with 36% of strains belonging to group 1 and sub-group C being most common. Previous studies (Millar, 2003; Parker *et al.*, 2005) have found that a greater proportion of their strain collection can be classified as group 1 strains. This may reflect the diversity of their strain collections, whereas the strain collection in this study has been compiled from clinical strains. This present study also found that 24% of the strains examined did not conform to the groups previously described, indicating the potential presence of new sub-groups. Novel combinations of genes were also found as some strains produced amplicons with primers corresponding to more than one sub-group.

The strains containing potentially novel gene combinations and new gene content were examined by primer walking, long range PCR and DNA sequencing. Primer walking entails the use of a forward 'anchor' primer in combination with reverse primers designed to each open reading frame. The primers for this experiment were designed by the Bacterial Microarray Group (BμG@S, St. George's, London) for the construction of their *C.jejuni* v2 microarray as described in Möen *et al.* (2005). This method was used

successfully by Miller (2003) to characterise *C.jejuni* O73 as a sub-group 3A strain. Although this method can be used as an indication of the change in gene content, it will not provide sequence level detail of the differences between strains. Long range PCR was utilised to amplify the LOS core biosynthesis gene cluster between *lpxL* and *waaV*. The long range PCR products were examined for size and then sequenced in multiple stages. To complete the characterisation of these strains, the LOS was extracted and examined by SDS-PAGE and Western blotting.

## **4.2 Primer walking**

### **4.2.1 Introduction**

Firstly the primers were tested and optimised against the control strains that were available. Primers were designed for any ORFs or sub-group that were not covered by the BµG@S microarray design as listed in Appendix 1.

Primer walking was carried out for each strain and each sub-group and repeated to confirm the results. The reactions were also carried out with an annealing temperature of 45°C and at 55°C. The first panel of primers used corresponded to sub-group 1C and extended from *gne* (Cj1131) to *wlaT* (Cj1152), therefore including the conserved genes; *gne*, *wlaA* and *waaC*, the variable genes between *lpxL* and *waaV* and the second variable region between *waaF* and *gmhA*. These regions were also completely covered by the Group 2 primers. For the other sub-group, the region covered extended from *lpxL* to *waaV*. The results for each strain with sub-group 1C primers will be described first and then the sub-group will be examined in turn.

### **4.2.2 Group1: Sub-group 1C**

Table 4.1 a and b summarises the primer walking results for all strains. Amplicons were produced for *gne* (Cj1131) with all the strains except 38625

**Table 4.1 a and b, Summary of primer walking results with primers corresponding to group 1C**, where A) shows the results from *gne* (1131) to *cst-II* (1140) and B) shows the results from *cst-II* (1140) to *wlaT* (1152). Product size is indicated in bp. A + indicates the presence of the correct size amplicon, - indicates the absence of the amplicon and < or > indicates an amplicon of a different size. The size of this larger or smaller amplicon is indicated alongside the greater than or less than sign. 'd' indicates that the primers produced double bands. Primer names indicate the 'Cj' ORF as allocated by the NCTC 11168 sequencing project. The genes to which these primers anneal are detailed in Appendix 1.

A)

Strains	Primer Combinations										
	1131cr					1135f					
	1131cf	1132cf	1133r	1134r	1135r	1135r	1136f	1137cf	1138r	1139cf	1140r
Product/bp	580	1802	2843	3564	4703	730	2275	2805	4544	5353	6309
Control	+	+	+	+	+	+	+	+	+	+	+
34086	+	+	-	+	+	+	+	-	-	-	-
34218	+	+	-	-	-	+	+	-	-	<0.5	-
34565	+	+	-	+	-	+	-	-	-	-	-
34806	+	-	-	-	-	+	-	-	-	-	-
38625	-	+	-	+	-	-	-	-	-	-	-
39864	+	+	-	+	-	+	+	-	-	-	-
40973	+	+	-	+	+	+	-	-	-	-	-
44406	-	+	-	+	+d	-	+	-	-	-	-
47185	+	-	-	+	+	+	-	-	-	-	-
50702	+	-	-	+	+	+	-	-	-	-	-
51566	+	-	-	+	+	+	-	-	-	-	-
51585	+	-	-	+	+	+	-	-	-	-	-
53305	+	+	+	+	-	+	-	-	-	-	-
54386	+	+	+	+	-	+	-	-	-	-	-
57073	+	+	+	+	+	+	-	-	-	-	-
60319	+	+	+	+	+	-	+	-	-	-	-
61666	+	+	+	+	+	+	<1.5	-	-	<0.6	<0.8

B)

Strains	Primer combinations												
	1140f							1146cr					
	1140r	1141r	1142r	1143r	1144r	1145cf	1146cf	1146cf	1148r	1149cf	1150cf	1151cf	1152cf
Product/bp	166	1407	1845	3199	4532	4987	5648	293	1259	2032	3001	4371	4755
Control	+	+	+	+	+	+	+	+	+	+	+	+	+
34086	-	-	-	-	-	-	-	+	+	+	+	+	+
34218	+	-	-	-	-	-	-	+	+	+	+	-	-
34565	-	-	-	-	-	-	-	+	+	+	<1.0	+	+
34806	+	-	-	-	<3.0	-	-	+	+	-	<2.3	-	-
38625	-	-	-	-	-	-	-	+	+	-	-	-	-
39864	-	-	-	-	-	-	-	+	+	+	+	+	-
40973	-	-	-	-	-	-	-	+	+	+	+	+	+
44406	d>	-	-	-	-	-	-	+d	+	+	+	+	+
47185	+	-	-	-	-	-	<1.5	+	+	+	-	>6.0	>6.6
50702	-	-	-	-	-	-	<1.8	+	+	>2.5	>1.0	>6.0	>6.6
51566	+	-	-	-	-	-	<	+	+	+	+	+	+
51585	+	-	-	-	-	-	<	+	+	+	+	+	+
53305	-	-	-	-	-	-	-	-	-	>6.7	>7.2	>8.1	>9.0
54386	-	-	-	-	-	-	-	-	-	>4.3	>5.0	>6.2	7.0
57073	-	-	-	-	-	-	-	+d	+d	+d	+d	+d	+d
60319	-	+	-	-	+	-	+	+	+	+	+	+	+
61666	-	-	-	-	-	-	+	+	+	+d	+d	+d	+d

and 44406. This was unexpected as this gene is conserved amongst the sub-group. These strains did not produce an amplicon for this gene as the reverse primer did not anneal. The forward primer did not fail as other primer combinations produced amplicons with these strains. The majority of the strains also produce an amplicon of expected size with reverse primers corresponding to other conserved genes *wlaA* (Cj1132) and *lpxL* (Cj1134), but only 5 strains produced a PCR product with the reverse primer for *waaC* (Cj1133), another conserved gene. Although the majority of strains produced amplicons corresponding to *wlaNB* (Cj1135), amplicons were not produced for the majority of strains with reverse primers corresponding to *wlaNB* (Cj1136), *wlaO* (Cj1137), *wlaP* (Cj1138), *cgtB* (Cj1139) and *cstII* (Cj1140). Even though the control strain, NCTC 11168, produced amplicons with primers for *cstII* – *waaV* (Cj1146), only a minority of strains produced amplicons with these primers. Some strains produced smaller amplicons than expected with the primers corresponding to *cstII* and *waaV* indicating sequence deletions or rearrangements as expected in this highly variable region.

The strain 44406, which produced multiple amplicons in the PCR screen (1B/1C/2A), does not appear to produce amplicons with the primers that cover the variable gene content of group 1C. Amplicons would have been expected with primers designed to *wlaO* (1137), *wlaP* (1138), *cgtAll/neuA* (1143) and *wlaSB* (1144). The same can be said for strains, 51585 and 57073 (both 1C/2A), which again do not produce amplicons with primers corresponding to these genes.

The region between *waaF* and *gmhA* (Cj1149) varies by gene insertion. It is not surprising that the majority of strains produce amplicons concurrent with the NCTC 11168 gene arrangement in this region. Notably strains 50702, 53305 and 54386 produced amplicons of increased size indicating gene insertions in this region.

**Table 4.2 Summary of the primer walking results with primers corresponding to group 1A**, showing the results from *lpxL* (1134) to *waaV*. Product size is indicated in bp. A + indicates the presence of the correct size amplicon, - indicates the absence of the amplicon and < or > indicates an amplicon of a different size. The size of this larger or smaller amplicon is indicated alongside the greater than or less than sign. 'd' indicates that the primers produced double bands. The ORFs referred to by the primer name are detailed in Appendix 1.



# Chapter 4: Determining the gene content of the unknown and novel strains

Strains	Primer Combinations													
	1134f					AcgtAf			43456-1f					
	1134r	1135r	wlaNBr	A1cgtAr	A2cgtAr	A1cgtAr	A2cgtAr	43456-1r	43456-1r	BneuBr1	neuCr1	neuAr1	GroupABr	waaVr
Product/bp	587	1724	3067	4331	4331	479	479	2197	403	993	2309	3479	4188	5115
Control	+	+	+	+	-	+	-	+	+	+	+	+	+	+
34086	+	-	-	-	-	-	-	-	-	-	-	-	-	-
34218	+	-	-	-	-	-	-	-	-	-	-	-	-	-
34565	-	-	-	-	-	-	+	-	-	-	-	-	-	-
34806	+	+	-	-	-	-	-	-	-	-	-	-	-	-
38625	+	-	-	-	+	-	+	+	+	+	+	+	-	-
39864	-	+	-	-	>6.4	-	+	+	-	-	-	-	-	-
40973	+	+	-	-	>6.4	-	+	-	+	+	+	+	-	-
44406	+	+	+	-	-	-	+	+	+	+	-	-	-	-
47185	+	-	-	-	>6.4	-	-	-	+	+	+	-	-	-
50702	+	+	+	-	+	-	+	+	+	+	+	+	+	-
51566	-	-	-	-	>6.4	-	-	-	-	-	-	-	-	-
51585	+	+	-	-	-	-	+	-	+	+	-	-	-	-
53305	+	+	+	-	-	-	+	<1.9	+	+	-	-	-	-
54386	+	+	+	+	-	+	-	<1.9	+	+	-	-	-	-
57073	+	-	-	-	-	+	-	<1.9	-	-	-	-	-	-
60319	+	+	+	-	-	+	-	+	+	-	-	-	-	-
61666	+	+	-	-	-	-	-	-	-	-	-	-	-	-

### **Sub-group 1A**

The primer walking results for the strains with group 1A primers are summarised in Table 4.2. As expected, amplicons were produced with the 1134f, 1134r and 1135r primers for the majority of strains. Four strains produced an amplicon with 1134f and A2cgtAr that was 2 Kb larger than expected, suggesting an alternative gene arrangement or the primer annealing non-specifically. Two of these strains; 39864 (UC) and 40973 (1B/2A), also produce amplicons with the majority of the primers after this set suggesting that these strains contain gene content homology with group 1A strains.

### **Sub-group 1B**

Again as can be seen in Table 4.3, the majority of strains produced an amplicon with the primers corresponding to *lpxL* (1134). The four strains which produced amplicons greater than those expected with 1134f and A2cgtAr in the group 1A primer walking did so again in this repeat experiment. Seven strains produced amplicons of decreased size with primers corresponding to *cstII* (43438-1) indicating possible gene rearrangements or mis-priming. Strain 50702 (UC) produced amplicons of the correct size with 11 out of 16 primer pairs, indicating that this strain contains group 1B like gene content. Strains 34806 (1B/3A), 40973 (1B/2A) and 44406 (1B/1C/2A) do not produce the amplicons expected with the sub-group 1B specific primers indicating that mis-priming may have produced these extra amplicons.

**Table 4.3 Summary of the primer walking results with primers corresponding to group 1B**, showing the results from *lpxL* (1134) to *waaVI* (1146). Product size is indicated in bp. A + indicates the presence of the correct size amplicon, - indicates the absence of the amplicon and < or > indicates an amplicon of a different size. The size of this larger or smaller amplicon is indicated alongside the greater than or less than sign. 'd' indicates that the primers produced double bands. The ORFs referred to by the primer name are detailed in Appendix 1

# Chapter 4: Determining the gene content of the unknown and novel strains

Strains	Primer combinations														
	1134f					AcgtAf			43438-1f						
	1134r	1135r	wlaNBr	A1cgtAr	A2cgtAr	A1cgtAr	A2cgtAr	43438-1r	43438-1r	BneuB r1	neuC r1	cgtA1lr	neuA r1	GroupABr	waaVr
Product/ bp	587	1724	3067	4331	4331	479	479	2093	305	990	2309	3866	4422	5131	6037
Control	+	+	+	-	+	-	+	+	+	+	+	+	+	+	+
34086	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-
34218	-	-	-	-	-	-	-	-	-	-	-	-	-	-	<1.8
34565	+	-	-	-	-	-	+	<1.5	-	-	<0.9	-	-	-	-
34806	+	+	+	-	-	-	-	<1.5	-	>1.3	-	-	-	-	-
38625	-	-	-	-	+	-	+	+	+	+	+	-	-	-	-
39864	+	-	-	-	>6.4	-	+	<1.5	-	-	-	-	-	-	-
40973	+	-	-	-	>6.4	-	+	-	-	-	-	-	-	-	-
44406	+	+	+	-	-	-	+	+	+	+	<2.0	-	-	-	+
47185	+	-	-	-	>6.4	-	-	<1.5	-	+	+	-	-	-	-
50702	+	+	+	+	-	-	+	+	+	-	+	+	+	+	-
51566	+	+	-	-	>6.4	-	+	<1.5	>0.8	-	-	-	-	-	-
51585	+	-	+	-	-	-	+	-	+	-	-	-	-	-	-
53305	+	+	+	-	-	-	+	+	-	-	-	-	-	-	-
54386	+	-	-	-	-	+	-	+	-	-	-	-	-	-	-
57073	-	+	-	-	-	+	-	+	-	>2.0	<1.1	-	-	-	-
60319	+	-	-	-	-	+	-	+	-	-	-	-	-	-	-
61666	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

### 4.2.3 Group 2: Sub-group 2A

Table 4.4 shows the summary of primer walking results with sub-group 2A primers. The primers corresponding to 1131cf (*gne*) and *wlaXr* are predicted *in silico* to only produce an amplicon with group 2 strains, but all the strains produce an amplicon with these primers suggesting mis-priming or homology to a conserved region of the LOS gene cluster. Notably strain 57073 (1C/2A) produces amplicons with all the primer combinations to 11828-4 (*wlaUG*), suggesting that any potential gene re-arrangement has taken place after this region. The other strain that produced this amplicon combination, 51585, does not produce amplicons after the primer combination 1131cr (*gne*) and 1134r (*lpxL*) suggesting that this LOS cluster does not contain sequence with homology to group 2 LOS genes. Similar information can be inferred from the primer walking results for strains 40973 and 44406.

### Sub-group 2B

Table 4.5 a and b summarise the results of the primer walking with the group 2B primers. Strain 39864, 2B/3A, produced amplicons with the 2B specific primers only to *wlaUA* (81116-24r) and two of these products were smaller in size than expected. As may be expected due to high sequence homology strain 57073 (1C/2A) produces amplicons to ORF tgh160, although this is not the case with strain 51585 which has been assigned to the same LOS groups by PCR. Notably three unclassified strains, 38625, 53305 and 54386 produce a range of amplicons with these primers, but there is an indication that their LOS clusters may have some homology to group 2B.

**Table 4.4 a and b. Summary of the primer walking results with primers corresponding to group 2A,** where A) shows the results from *gne* (1131) to *wlaUG* (11828-4) and B) shows the results from *wlaUG* (11828-4) to 11828-12. Product size is indicated in bp. A + indicates the presence of the correct size amplicon, - indicates the absence of the amplicon and < or > indicates an amplicon of a different size. The size of this larger or smaller amplicon is indicated alongside the greater than or less than sign. D indicates that the primers produced double bands; 'M' indicates the presence of multiple bands. The ORFs referred to by the primer name are detailed in Appendix 1.

# Chapter 4: Determining the gene content of the unknown and novel strains

A)

Strains	Primer Combinations										
	1131cr						81116-26f				
	1131cf	wlaXr	1134r	81116-24r	81116-25r	81116-26r	81116-26r	129600r	11828-2f	11828-3A	11828-4f
Product/bp	580	1794	3565	4185	5272	5877	492	1175	1805	2265	2760
Control	+	+	+	+	+	+	+	+	+	+	+
34086	+	+	+	-	-	-	-	-	-	-	-
34218	+	+	+	-	-	-	-	-	-	-	-
34565	+	+	+	-	-	-	>0.8	-	-	-	-
34806	+	+	-	+	+	-	-	+	+	+	-
38625	+	+	<2.3	+	+	-	+	-	-	-	-
39864	+	+	<2.3	-	-	-	-	-	-	-	-
40973	+	+	+	-	-	-	+	-	-	-	-
44406	+	+	+	-	-	-	-	-	-	+	-
47185	+	+	+	+	-	-	+	-	-	-	-
50702	+	+	<2.3	<3.0	<3.3	-	+	+	-	<1.5	-
51566	+	+	+	+	+	-	-	+	+	+	+
51585	+	+	+	-	-	-	-	-	-	-	-
53305	+	+	+	+	+	+	-	+	+	+	-
54386	+	+	+	-	-	-	+	-	-	-	-
57073	+	+	+	+	+	+	+	+	+	+	+
60319	+	+	+	-	+	+	+	+	-	+	-
61666	+	+	+	-	-	-	-	-	<0.3	-	-

# Chapter 4: Determining the gene content of the unknown and novel strains

B)

Strains	Primer combinations													
	11828-4r					11828-1r								
	11828-4f	11828-5f	11828-6f	11828-7f	11828-1f	11828-1f	129592r	11828-8f	11828-9f	11828-10f	1146cf	1148r	11828-11f	11828-12f
Product/bp	355	1705	2537	3773	4523	109	1149	2519	3207	4287	4858	6145	7755	8864
Control	+	+	+	+	+	+	+	+	+	+	+	+	+	+
34086	-	+	+	-	-	-	-	-	-	-	-	-	-	-
34218	-	-	-	-	-	+	-	-	-	-	-	-	-	-
34565	-	-	-	-	-	-	+	+	-	-	-	-	-	-
34806	-	-	-	-	-	+	+	-	-	-	-	-	-	-
38625	-	-	-	-	-	+	+	+	-	-	-	-	-	-
39864	-	+	+	+	+	+	+	+	+	+	-	-	-	-
40973	-	-	-	-	-	-	-	-	-	-	-	-	-	-
44406	-	-	-	-	-	+	-	-	+	-	-	-	-	-
47185	-	-	-	-	d<1.3	d>1.0	M	-	+	-	-	-	-	-
50702	-	-	-	d<1.0	M	-	+	-	+	-	+	-	-	-
51566	+	+	-	-	-	-	-	-	-	-	-	-	-	-
51585	-	-	-	-	>	-	-	-	-	-	-	-	-	-
53305	-	-	-	-	>	-	-	-	-	-	-	-	-	-
54386	+	+	+	<2.0	<2.9	-	-	-	-	-	-	-	-	-
57073	+	-	+	-	-	-	-	-	-	-	-	-	-	-
60319	-	<1.3	-	-	-	+	>4.4	-	-	-	-	-	-	-
61666	-	-	-	-	-	-	-	-	-	-	-	-	-	-



**Table 4.5 a and b, Summary of the primer walking results with primers corresponding to group 2B**, where A) shows the results from *gne* (1131) to *wlaUH* (11828-5) and B) shows the results from *wlaUH* (11828-5) to *waaV*. Product size is indicated in bp. A + indicates the presence of the correct size amplicon, - indicates the absence of the amplicon and <> indicates an amplicon of a different size. The size of this larger or smaller amplicon is indicated alongside the greater than or less than sign. 'd' indicates that the primers produced double bands. 'M' indicates the presence of multiple bands. The ORFs referred to by the primer name are detailed in Appendix 1.

# Chapter 4: Determining the gene content of the unknown and novel strains

A)

Strains	Primer combinations											
	1131cr						81116-26f					
	1131cf	wlaXr	1134r	81116-24r	81116-25r	81116-26r	81116-26r	tgh020r	11828-2f	11828-3A	tgh0160r	11828-5f
Product/bp	580	1794	3565	4158	5272	5877	492	1177	1804	2263	2754	4125
Control	+	+	+	+	+	+	+	+	+	+	+	+
34086	+	+	+	-	-	-	-	-	-	-	-	-
34218	+	+	-	-	-	-	-	-	-	-	-	-
34565	+	+	+	+	+	-	>0.8	>2.0	-	>3.0	-	-
34806	+	+	-	-	-	-	-	-	+	+	-	-
38625	+	+	+	+	+	-	+	+	+	-	-	-
39864	+	+	<2.3	<3.0	-	-	-	-	-	-	-	-
40973	+	+	<2.3	-	-	-	+	-	-	-	-	-
44406	+	+	+	-	-	-	-	-	-	-	-	-
47185	+	+	+	+	-	-	+	-	-	-	-	-
50702	+	+	+	-	-	-	+	+	+	+	-	-
51566	+	+	<2.3	<3.0	<3.3	-	-	+	+	+	+	+
51585	+	+	+	-	-	-	-	-	-	-	-	-
53305	+	+	+	+	+	+	-	+	+	+	+	+
54386	+	+	+	-	-	-	+	+	-	+	+	-
57073	+	+	+	+	+	+	+	-	+	+	+	-
60319	+	+	+	-	+	+	+	+	-	-	-	-
61666	+	+	+	-	-	-	-	+	-	-	-	-

# Chapter 4: Determining the gene content of the unknown and novel strains

B)

Strains	Primer combinations							
	81116-1f				11828-8r			
Product/bp	11828-5f 460	11828-7f 1507	11828-1f 2255	tgh004r 3615	11828-8f 4673	11828-8f 452	11828-9f 1140	waaV 1818
Control	+	+	+	+	+	+	+	+
34086	-	-	-	-	-	-	-	-
34218	-	-	-	-	-	-	-	-
34565	>1.3	-	-	<0.9	-	-	-	-
34806	-	-	-	-	-	-	-	-
38625	+	+	+	+	-	-	-	-
39864	+	-	+	-	-	-	-	-
40973	-	-	-	-	-	-	-	-
44406	-	-	-	-	-	-	-	-
47185	-	-	-	-	-	-	-	-
50702	-	-	-	-	-	-	-	-
51566	-	-	-	-	-	-	-	+
51585	-	-	-	-	-	-	-	-
53305	+	+	+	+	-	-	-	+
54386	>2.3	>2.8	>3.5	>4.0	-	-	-	+
57073	>2.3	-	>3.5	-	-	-	-	-
60319	-	-	+	-	+	-	-	-
61666	-	-	-	-	+	-	-	-

#### 4.2.4 Group 3: sub-groups 3A and 3B

As can be seen in Table 4.6, four strains produced amplicons with primers designed to ORFs other than 1134 (*lpxL*), *orfA* and 1135(*wlaNA*). None of the strains, except the control strain produced an amplicon with the primer pair 1134f (*lpxL*) and *waaV*, indicating PCR failure. Strain 44406 (1B/1C/2A) produced amplicons that were a different size than those expected with three primer pairs indicating that there may be rearrangement in the LOS cluster or mis-priming with these primers. Strains 34806 (1B/3A) and 39864 (2B/3A) do not produce amplicons with all of the group 3A specific primers. The primer *GorfAr* also anneals to group 1 strains indicating that there may be some false-positive results with this primer.

Table 4.7 summarises the results for the primer walking of all of the strains with group 3B primers. No control strain was available to test these primers, although many of them are designed to ORFs that share homology with group 3B. The group 3B specific primer, GF2 was previously used in the PCR screen and works successfully. Only four strains, 34218, 34565, 38625 and 51585 produced amplicons with the 3B specific primer GF2 in combination with other primers. Although these primers have been used successfully in other combinations, it can not be ruled out that some of the primers did not work in these particular combinations.

#### 4.2.5 Group 4A

Primer walking results with this primer set are summarised in Table 4.8. Again due to the lack of a positive control strain, the primers could not be tested and optimised. The only strain which was successfully amplified by group 4A specific primers was 38625. Four strains also produced amplicons with group 4A primers, although they were less than the predicted size, indicating gene re-arrangements or non-specific priming.

**Table 4.6 Summary of the primer walking results with primers corresponding to group 3A**, showing the results from *lpxL* (1134) to *waaV* (1146). Product size is indicated in bp. A + indicates the presence of the correct size amplicon, - indicates the absence of the amplicon and <> indicates an amplicon of a different size. The size of this larger or smaller amplicon is indicated alongside the greater than or less than sign. 'd' indicates that the primers produced double bands. . 'M' indicates the presence of multiple bands. The ORFs referred to by the primer name are detailed in Appendix 1.

Chapter 4: Determining the gene content of the unknown and novel strains

Strains	Primer combinations							
	1134f							
	1134r	GorfAr	1135r	LIO87-orfBr	LIO87-orfC/ Dr	LIO87-orfEr	LIO87-orfFr	waaV
Product/bp	587	1223	1724	2714	4217	5210	6286	7028
Control	+	+	+	+	+	+	+	+
34086	+	+	-	-	+	-	<	-
34218	-	-	-	-	-	-	<	-
34565	+	-	-	-	-	-	-	-
34806	+	+	+	-	-	-	-	-
38625	-	-	-	-	-	-	-	-
39864	+	-	-	-	-	-	-	-
40973	+	+	-	-	-	-	<0.5	-
44406	+	>1.5	>2.0	-	<2.3	-	-	-
47185	+	-	-	-	-	-	-	-
50702	+	-	+	-	-	-	-	-
51566	+	-	-	-	-	-	<1.9	-
51585	+	-	+	-	-	-	-	-
53305	+	-	-	-	-	-	-	-
54386	+	+	+	-	-	-	-	-
57073	-	-	-	-	-	-	-	-
60319	+	-	+	-	-	-	-	-
61666	-	-	+	-	-	-	-	-

**Table 4.7 Summary of the primer walking results with primers corresponding to group 3B**, showing the results from *lpxL* (1134) to *waaV* (1146). Product size is indicated in bp. A + indicates the presence of the correct size amplicon, - indicates the absence of the amplicon and <> indicates an amplicon of a different size. The size of this larger or smaller amplicon is indicated alongside the greater than or less than sign. D indicates that the primers produced double bands and na indicates that a control strain was unavailable. 'M' indicates the presence of multiple bands. The ORFs referred to by the primer name are detailed in Appendix 1.

Strains	Primer combinations							
	1134f		GF2					
	1134r	GorfAr	LIO87-orfGr	LIO87-orfBr	LIO87-orfC/ Dr	LIO87-orfEr	LIO87-orfFr	waaV
Product/bp	587	1223	509	1365	2868	3861	4937	5869
Control	na	na	na	na	na	na	na	na
34086	+	+	-	-	-	-	-	-
34218	-	-	-	-	-	+	-	+
34565	+	-	-	-	-	+	-	-
34806	+	+	-	-	-	-	-	-
38625	-	-	-	-	+	-	-	-
39864	+	-	-	-	-	-	-	-
40973	+	-	-	-	-	-	-	-
44406	+	+	-	-	-	-	-	-
47185	+	-	-	-	-	-	-	-
50702	+	+	-	-	-	-	-	-
51566	+	-	-	-	-	-	-	-
51585	+	+	-	-	M	M	-	-
53305	+	+	-	-	-	-	-	-
54386	+	+	-	-	-	-	-	-
57073	-	-	-	-	-	-	-	-
60319	+	+	-	-	-	-	-	-
61666	-	+	-	-	-	-	-	-



**Table 4.8 Summary of the primer walking results with primers corresponding to group 4A**, showing the results from *lpxL* (1134) to *waaV* (1146). Product size is indicated in bp. A + indicates the presence of the correct size amplicon, - indicates the absence of the amplicon and <> indicates an amplicon of a different size. The size of this larger or smaller amplicon is indicated alongside the greater than or less than sign. 'd' indicates that the primers produced double bands and na indicates that a control strain was unavailable. 'M' indicates the presence of multiple bands. The ORFs referred to by the primer name are detailed in Appendix 1.

Chapter 4: Determining the gene content of the unknown and novel strains

Strains	Primers							
	1134f							
	1134r	81116-24r	43437-orf4cf	43437-orf5cf	Lio87-orfFr	43437-orf7cf	D1146r	1148r
Product/bp	587	1180	2088	4213	5064	5617	6634	
Control	na	na	na	na	na	na	na	na
34086	+	-	-	-	-	-	-	-
34218	+	-	-	-	-	-	-	-
34565	+	-	-	-	-	-	-	-
34806	+	-	-	-	-	-	-	-
38625	+	+	+	+	-	-	-	-
39864	+	-	-	-	-	<0.8	<1.9	-
40973	+	-	-	-	-	-	-	-
44406	+	-	-	-	<0.5	<1.2	<1.5	+
47185	+	-	-	-	-	-	-	-
50702	+	-	-	-	-	-	-	-
51566	-	-	-	-	-	-	-	-
51585	+	-	-	-	<1.9	-	<1.9	-
53305	+	-	-	-	-	-	-	-
54386	+	-	-	-	-	-	<1.9	-
57073	-	-	-	-	-	-	-	-
60319	+	-	-	-	-	-	-	-
61666	-	-	-	-	-	-	-	-

#### **4.2.6 Conclusion**

Primer walking provides an overview of LOS gene content. Although this can provide useful information in this instance no definitive conclusions about the LOS group could be assigned despite repetition of the experiment at 45°C and 55°C. Table 4.9 summarises the LOS group assignment for each strain following primer walking and despite some strains being assigned one sub-group type, no LOS cluster was amplified by all the primer combinations for one sub-group, perhaps indicating different gene content combinations. Other indicators of gene re-arrangements were changes in amplicon size. Primer walking relies on prior knowledge of LOS gene content and therefore perhaps is not the ideal way to examine novel gene content. It also does not provide detail of SNP and other changes at the sequence level.

**Table 4.9 Summary of the LOS sub-group for each strain after primer walking.**

## Chapter 4: Determining the gene content of the unknown and novel strains

---

Strain	PCR sub-grouping	Possible Primer walking grouping
34086	UC	3A
34218	UC	1C
34565	UC	2B
34806	1B/3A	2A
38625	UC	2A; 4A
39864	2B/3A	2A
40973	1B/2A	1A
44406	1B/1C/2A	1B
47185	UC	UC
50702	UC	1A;1B; 2A
51566	UC	2A
51585	1C/2A	1C
53305	UC	1A; 2A; 2B
54386	UC	UC
57073	1C/2A	2A; 2B
60319	UC	2A; 2B
61666	UC	UC

### **4.3 Phenotypic analysis of the novel strain LOS.**

The LOS from each strain was prepared and compared phenotypically to determine if any differences could be seen on a structural level.

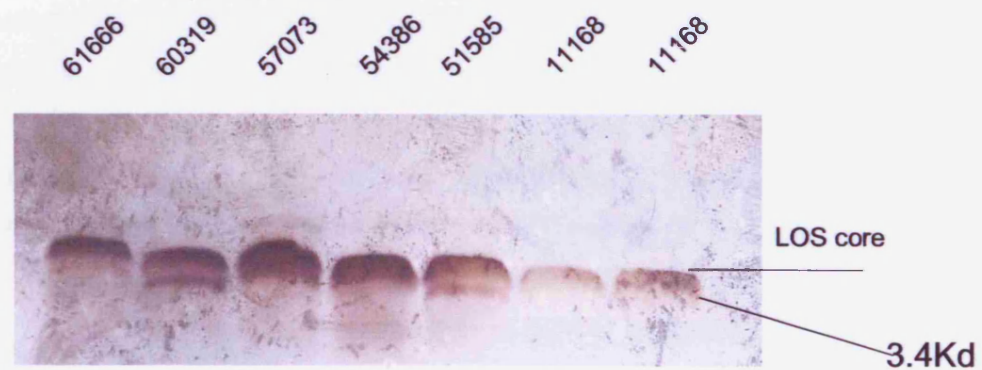
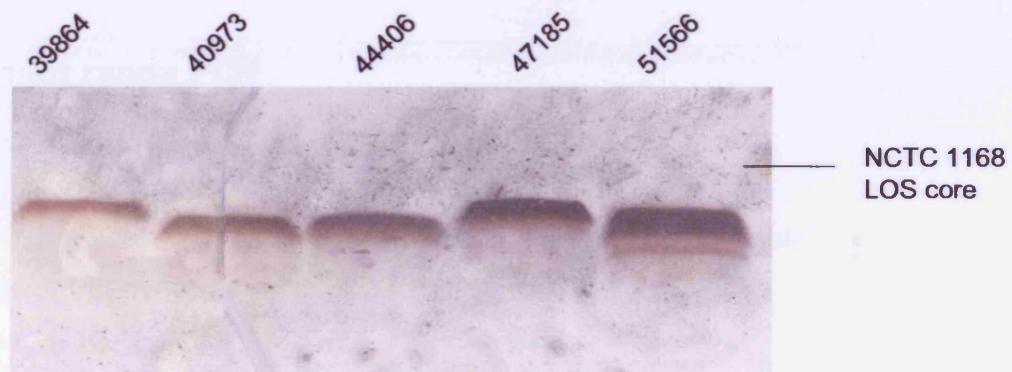
#### **4.3.1 Analysis of LOS by Tricine-SDS PAGE**

LOS samples were prepared as detailed in Chapter 2. Samples from each strain were examined by 18% (v/v) Tricine-SDS PAGE as can be seen in Figure 4.1 a, b and c. A LOS extract from NCTC 11168 was also included on each gel for comparison. The sizes of the LOS preparations were all variable and several strains, 51566, 51585 and 60319 also produce a truncated core suggesting the presence of homopolymeric tracts in the LOS cluster or the ability to produce two differing LOS structures. Strains 51585 and 57073 which were both classified as 1C/2A following the initial screening PCR do not have LOS cores of the same length, but strains 40973 (1B/2A) and 44406 (1B/1C/2A) appear to have LOS cores of similar size.

#### **4.3.2 Analysis of LOS by Western blot with cholera toxin.**

All of the LOS samples were analysed for their ability to bind cholera toxin subunit B by dot-blotting. NCTC 11168 was included as a positive control and 11168::*wlaPSC* (see Chapter 6) was included as a negative control. None of the clinical strains bound cholera toxin, indicating that the LOS outer core did not mimic a GM1-like epitope (data not shown), although signal was produced with NCTC 11168. The LOS from each strain was also probed with the lectin, RCA, but no signal was produced as with 11168::*wlaPSC* and 11168::*wlaAT* in Chapter 6.

**Figure 4.1. Analysis of LOS by Tricine-SDS PAGE.** The figures show 18% (v/v) SDS-PAGE gels with LOS extracts from the clinical strains stained with silver nitrate using the method of (Tsai and Frasch, 1982). The size of the NCTC 11168 WT core is indicated.





### **4.3.3. Conclusions**

Phenotypic analysis of the LOS preparations from all the strains provides basic information about the size of the LOS core but does not provide information about gene content. The presence of two bands in some of the samples suggests that the locus may contain homopolymeric tracts which result in the production of truncated LOS moieties. The probing of dot-blotted LOS samples with cholera toxin confirmed that they did not express GM1 like outer core structures.

## **4.4 Long range PCR**

### **4.4.1 Strategy and optimisation**

Long range PCR was used to amplify the LOS cluster, both to compare the size of the clusters, which may highlight changes in gene content and also to use as a template for sequencing the whole cluster. As long range PCR is a much less robust technique than standard PCR, the decision was taken to use previously published primers designed to LOSXLF (*lpxL*) and LOSXLR (*waaV*; Parker *et al.*, 2005). This experiment was tried with at least four kits; including the TripleMaster™ mix system (Eppendorf), Expand Long Template PCR kit (Roche), KOD XL (Novagen) and MasterAmp™ Extra-Long PCR kit (Epicentre). Initial optimisation of the kits failed to produce amplicons with any control strain, although an amplicon was intermittently produced with NCTC 11168. In order to improve this situation further optimisation was performed with each kit in turn, involving magnesium titrations, primer concentration gradients, annealing temperature gradients and template concentration gradients. It was finally established that the most robust and reproducible amplification was produced with the MasterAmp™ Extra-Long PCR kit. Despite this it was still impossible to amplify the LOS clusters from any of the clinical strains. The purity and integrity of the template DNA was known to be extremely critical to the success of long range PCR amplification,

**Figure 4.2: A gel photograph showing the long range PCR amplicons produced with the unclassified clinical strains.** Amplicon size is indicated in Kb and this was determined after comparison against a standard curve of the DNA marker size against distance migrated. Sizes shown in brackets are estimated from other gel photographs (data not shown).

**Table 4.10 a) summarises the size of long range amplicons produced for each strain and b) summarises the expected size for each sub-group control.** Amplicon size is indicated in Kb, ND indicates that the size of the PCR product was not determined and NP indicates no product could be produced.

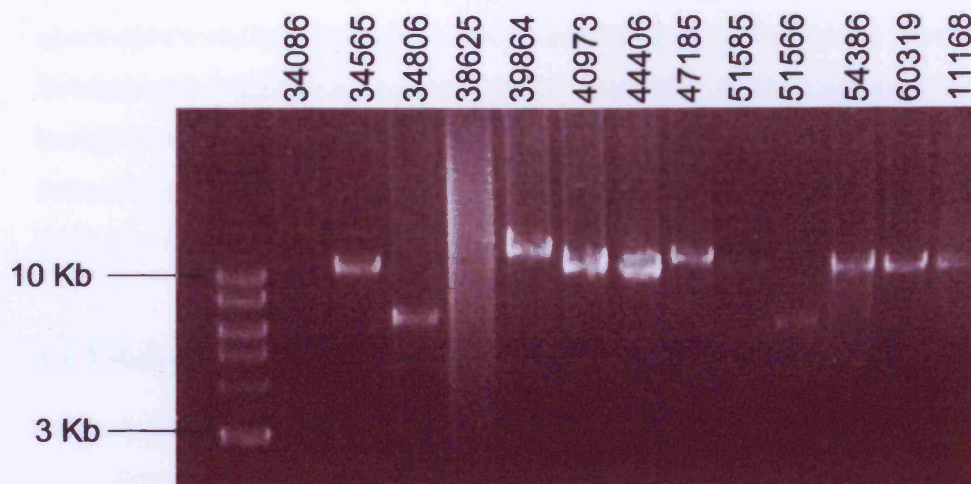


Figure 4.2

Strain	Long range product size/Kb
34086	ND (6.9)
34218	ND (13.75)
34565	12.50
34806	7.20
38625	ND (12.5)
39864	13.75
40973	11.20
44406	10.90
47185	12.80
49068	NP
50702	ND (10.9)
51566	6.90
51585	11.20
54386	12.50
57073	NP
60319	12.80
61666	NP

Table 4.10 a)

Group	Strain	Long range product size/Kb
1A	11351	10.7
1B	81-176	11.6
1C	11168	12.7
2A	81116	14.4
2B	Tgh09011	12.3
3A	RM1221	6.9
3B	LIO87	7.2
4A	ATCC43437	6.9

Table 4.10 b)

and it was noted during examination of the DNA samples with a Nanodrop™ spectrophotometer that the DNA quality was poor (data not shown). Subsequent microarray hybridisations with this DNA also produced high background and poor quality microarray data (data not shown). DNA extractions were repeated using the Puregene kit (Gentra) which produced DNA of a higher quality and increased yield (data not shown).

### 4.4.2 Results

Long range PCR was performed as discussed in Chapter 2 at three annealing temperatures; 45, 50 and 55 °C. Despite numerous attempts at optimisation and re-extraction of genomic DNA, no product was produced with strains 49068, 57073 and 61666. Figure 4.2 shows a gel photograph of the products amplified for the majority of the strains and Table 4.10 indicates the size of the LOS cluster after amplification by PCR, NP indicates no product and ND indicates that the product size was not determined. Products ranged from 6.9 Kb to 13.75 Kb and may provide an indication of LOS group, for instance 40973 (1B/2A) and 44406 (1B/1C/2A) produce amplicons that are 300 bp different in size. On examination of the size of the control strain long range products, strains such as 51585 that produce a smaller amplicons may be predicted to belong to group 3 or group 4 and those that produce larger amplicons are predicted to belong to group 1 or 2.

### 4.4.3 Conclusion

Long range PCR amplification proved to be difficult to optimise for these clinical strains and success after DNA extraction using a non-organic solvent based method emphasises the importance of good quality template DNA for this technique. Long range PCR is less robust and variables that may not be an issue with standard PCR may cause a problem with this technique. As such DNA samples were never subjected to freeze-thawing cycles and steps such as using hot-start PCR programs were used. For the strains where no product could be recovered, the cycling program may need a

longer extension time or additional cycles, but it is possible that the primer pair may not anneal to these strains. The PCR primers are used at between 8 and 18°C below their ideal annealing temperature and should still function despite minor mismatches in the DNA sequence. Amplicon size may provide an initial method of classifying amplicons into LOS gene content group.

## **4.5 Sequencing of long range PCR amplicons**

### **4.5.1 Method**

Long range PCR amplicons were cleaned and concentrated by ethanol precipitation. After estimating the DNA concentration by visualisation on an agarose gel, sequencing was carried out using Big Dye version 3.1 as described in Chapter 2. The initial round of sequencing was performed using the long range PCR primers, LOSXLF (*lpxL*) and LOSXLR (*waaV*) at 45°C. After the analysis of this first round of sequencing, the sequences of the amplicons diverged as expected and individual primers were designed to each strain, as detailed in Appendix 1.

### **4.5.2 Results**

Sequencing analysis was completed using Chromas, Clone manager (Science and Education Software Inc.) and the NCBI BLAST database. Contigs were assembled using the sequence assemble function in Clone Manager. Analysis of the sequences was completed using the nucleotide and protein databases compiled by NCBI. Table 4.11 summarises the extent to which the sequencing for LOS cluster of each strain was completed. Several

**Table 4.11 Summary of the sequencing stages and fold coverage for each strain**, where NP indicates no amplicon, F indicates the forward strand and R indicates the complementary strand.

Strain	Forward/bp	Reverse/bp	approx size/bp	total bp sequenced	Fold coverage	notes
34086	5000	4350	6900	9350	1.36	complete
34218	6050	5900	13750	11950	0.86	
34565	5427	1700	12500	7127	0.57	
34806	5600	4100	7200	9270	1.28	contigs need joining
38625	3900	2900	12500	6800	0.54	
39864	1327	5049	13750	13477	0.46	
40973	2000	2500	11200	4500	0.40	homopolymeric tract (F)
44406	6200	1400	10900	7610	0.70	homopolymeric tract (R)
47185	5100	2200	12800	7300	0.57	homopolymeric tract (R)
49068	0	0	0	0	0	NP
53305	0	0	0	0	0	NP
50702	5720	4500	10900	10220	0.94	
51566	4000	3000	6900	7000	1.01	complete
51585	3670	5597	11200	9267	0.83	
54386	6485	4850	12500	11335	0.90	homopolymeric tract (F)
57073	0	0	0	0	0	NP
60319	6080	5500	12800	11580	0.90	
61666	0	0	0	0	0	NP

issues arose during this exercise. Technical problems were encountered where the long range PCR started to fail, especially with those strains that required a higher annealing temperature. This PCR failure coincided with a change in the buffer batch and the effect of this was confirmed when a new batch of buffer was tested. Several of the strains, 40973, 44406 and 47185 contained homopolymeric tracts early in the forward or reverse sequence. The sequence for these strains was mixed after these tracts and although numerous attempts were made to cover the homopolymeric tracts and tease apart the mixed sequencing using different primers, the ability to do this proved to be beyond the scope of this study. As can be seen from Table 4.11, not all the strains were sequenced to completion, five strains were over 0.9 fold complete at the end of this study. Two strains were completed with the contigs joined but a further strain needed the contigs to be joined for completion. This would require the design of further primers to complete the process. Although these strains are incomplete, some conclusions can still be drawn about their LOS clusters.

#### **4.5.3 Analysis of completed sequences.**

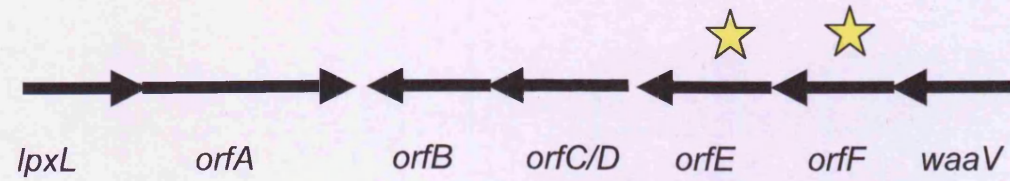
Two strains had their LOS cluster completely sequenced; 34086 and 51566 where both 6.9 Kb in length. Initial analysis of these strains by primer walking indicated that strain 34086 belonged to sub-group 3A and that strain 51566 could be classified as a 2A strain. Initial analysis at the sequence level showed that both of strains had LOS clusters with homology to group 3A. Figure 4.3 is a schematic diagram representing the gene arrangement for both of these strains. Further comparisons of both of these strains revealed variations in the LOS cluster from RM1221 with regards to ORF arrangement. Table 4.12 summarises the percentage identities of each ORF to the respective RM1221 orthologue and any differences in the expressed protein. Unlike RM1221 and O73, both of these strains contain two ORFs in place of the single *orfA*, albeit in a different arrangement. This perhaps is unsurprising as RM1221 *orfA* encodes a putative two domain glycosyltransferase and part



**Figure 4.3 Schematic diagram of the *lpxL-waaV* region of *C.jejuni* RM1221, 34086 and 51566.** Arrows indicate ORFs and the direction of their transcription. Both strain 34086 and 51566 show new arrangements of *orfA* and *orfE*. These strains also do not have homopolymeric tracts in (indicated by ✕) *orfE*, as in examples of other group 3 strains; this is due to the re-arrangement of this ORF in the CSSS strains.

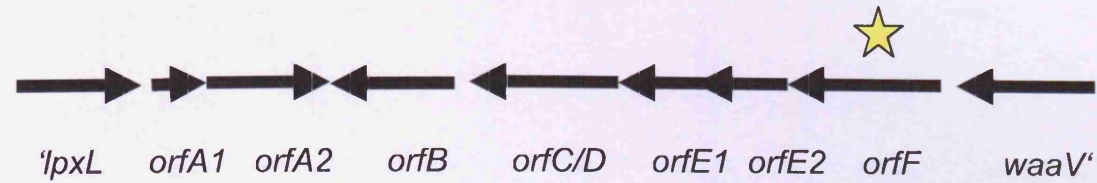
RM1221

(Long range product  
= 6948 bp)



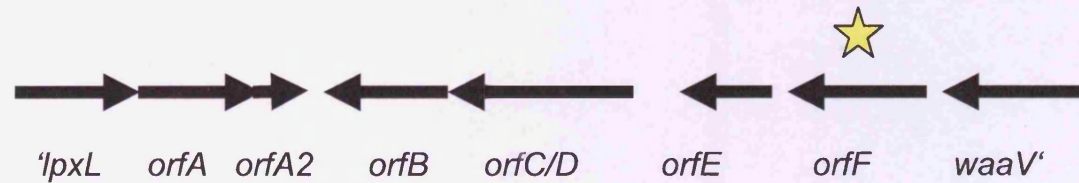
34086

6877 bp



51566

6827 bp



**Table 4.12 a) and b). Summary of identity comparisons with RM1221 and ORF organisation in strains 34086 and 51566.**

A)

Open Reading Frame	RM1221 orthologue	Function	% nucleotide identity (bases)	% amino acid similarity (aa)	Comments
34086 -A1 34086 - A2	A	putative 2 domain glycosyltransferase	99 (1-2045)	98 (1-101) 98 (132-516)	insertion of 2 'a' bases at 791 & 792 causes missense mutation
34086 - B	B	putative 2 domain glycosyltransferase (WcaA-like)	98 2136-6846	96 (1-295)	NA
34086 - C/D	C/D	putative galactosyltransferase		100 (1-299)	base change at 3230 & 3231 causes missense mutation
34086 - E1 34806 - E2	E	putative galactosyltransferase		100 (189-323) 93 (1-120)	insertion of 'ag' at 4802 & 4802 leads to missense mutation
34086 - F	F	Unknown		90 (1-295)	NA

B)

Open Reading Frame	RM1221 orthologue	Function	% nucleotide identity (bases)	% amino acid similarity (aa)	Comments
51566 - A1 51566 - A2	A	putative 2 domain glycosyltransferase	99 (1-4523)  98 (4523-6827)	100 (1-355) 100 (374-516)	insertion of 't' at base 1532 causes missense mutation
51566 - B	B	putative 2 domain glycosyltransferase (WcaA-like)		95 (1-93)	Deletion of 'a' at 2605 causes missense mutation
51566 - C/D	C/D	putative galactosyltransferase		98 (1-402)	NA
51566 - E	E	putative galactosyltransferase		97 (78-266)	Missing first 78 aa as homopolymeric tract deleted and inserted bases cause frameshift. Insertion of 'g' at 4343 causes missense mutation
51566 - F	F	Unknown		98 (1-196)	Early termination due to base change at 4321 from a to t

of this gene is deleted and replaced by *orfG* in group 3B. In contrast to other group 3 strains, strains 34086 or 51566 appear to have homopolymeric tracts in *orfE* unlike strains O73 and RM1221 but this may be because the expressed protein is already shorter in these strains than it is in RM1221 due to missense and frame shift mutations. In 34086, the *orfE* orthologue also is split into two ORFs. In both these strains, *orf C/D* are also combined such as in RM1221, but in 34086 the protein is shortened due to a missense mutation.

#### 4.5.4 Analysis of non-contiguous and almost completed sequences

Table 4.13 A, B, C and D summarise the sequence identity of each strain. Where partial contigs have been constructed the homologous orthologues have been displayed based on percentage identity. For instance with strain 54386 where the percentage identities change over contig, the contig is broken down into the corresponding ORFs.

##### Strain 34806

The size of the long range PCR product from this strain, together with the primer walking and initial PCR results indicate that this strain has a group 3A LOS gene content. Examination of the sequencing confirms that this is correct. Deviations in the sequence from RM1221 may have allowed primers that were designed specifically for sub-group 1B to also anneal to this LOS cluster.

##### Strain 50702

Primer walking analysis classified this previously unclassified strain as belonging to sub-groups 1A, 1B or 2A. Sequencing analysis has indicated that this strain is similar to RM3423, classified as a class 'O' strain (Parker and Horn, unpublished, accession number EF143352). Examination of strain RM3423 indicates this strain is similar to but divergent from the other strains in group 2, indicating a new sub-group. On comparing strain 50702 and

**Table 4.13 a-d Summary of nucleotide identity for each strain.**  
Sequenced fragments and contigs are shown with their orthologous strains and ORFs together with percentage identity.

## Chapter 4: Determining the gene content of the unknown and novel strains

<b>a) 34806 7.2 Kb</b>		<b>1.28 x coverage</b>
Sequencing stage or contig (size/bp)	Sequence homology to	% identity
F1-2 contig (1858)	RM1221 <i>lpxL</i> , <i>orfA</i>	99
F3 contig (990)	RM1221 <i>orfA</i> , <i>orfB</i>	96
F4 (1113)	RM1221 <i>orfB</i> , <i>orfC/D</i>	96
F5 (1276)	RM1221 <i>orfE</i> , <i>orfF</i>	95
F6 (1204)	RM1221 <i>orfE</i> , <i>orfF</i>	96
R1-4 contig (3003)	RM1221 <i>orfC/D</i> , <i>orfE</i> , <i>orfF</i> & <i>waaV</i>	97

<b>b) 50702 10.9 Kb</b>		<b>0.94 x coverage</b>
Sequencing stage or contig (size/bp)	Sequence homology to	% identity
F1 (959)	RM3423 <i>lpxL</i> , <i>wlaX</i>	95
F2 contig (860)	RM3423 <i>rmlA</i>	97
F3 (1067)	RM3423 <i>rmlB</i>	95
F4 (1079)	RM3423 <i>rmlB</i>	98
F5 (1222)	RM3423 <i>rmlB</i> , <i>wlaUC</i> , <i>wlaUD</i>	98
F6 (1228)	RM3423 <i>wlaUD</i> , <i>wlaUE</i> , <i>wlaUF</i>	95
F7 (1240)	RM3423 <i>wlaWA</i> , <i>wlaWB</i>	97
R1-2 contig (1556)	RM3423 <i>waaV</i> , <i>wlaWC</i>	97
R3-4 contig (1772)	RM3423 <i>wlaWA</i> , <i>wlaWB</i> , <i>wlaWC</i>	99
R6 (1216)	RM3423 <i>wlaWB</i> , <i>wlaWC</i>	98
R7 (1278)	RM3423 <i>wlaWB</i> , <i>wlaWC</i>	97



## Chapter 4: Determining the gene content of the unknown and novel strains

<b>c) 54386 12.5 Kb</b>		<b>0.9 x coverage</b>
Sequencing stage or contig (size/bp)	Sequence homology to	% identity
F1-7 (5300)	RM1221 <i>lpxL</i> , <i>orfA</i> , <i>orfB</i> , <i>orfC/D</i> , <i>orfE</i>	98
R1-9 (4850) Bases 1- 641  Bases 835 - 1737  Bases 1797-4055  Bases 4070-4799	Strain	Function
	RM1221 <i>waaV</i>	Putative glycosyltransferase
	176.83 CPS ORF <i>HS41:27(galE2)</i>	Nucleotidyl sugar pyranose mutase
	176.83 CPS <i>HS41:25 (galE1)</i> <i>HS41:26 (udg)</i>	Putative sugar epimerase UDP-glucose 6-dehydrogenase
	176.83 CPS <i>glf3</i>	Nucleotidyl sugar pyranose mutase

<b>d) 60319 12.8 Kb</b>		<b>0.9 x coverage</b>
Sequencing stage or contig (size/bp)	Sequence homology to	% identity
F1-7 (5082)	RM1221 <i>lpxL</i> , <i>orfA</i> , <i>orfB</i> , <i>orfC/D</i> , <i>orfE</i>	98
	Strain	Function
F8 (1000)	176.86 CPS <i>HS41.29</i>	Putative sugar transferase
R1 (429)	ATCC 43437 <i>waaV</i>	Putative glycosyltransferase
R2-9 (5118) Bases 225-942	176.86 CPS <i>HS41:25 (galE1)</i>	Putative sugar epimerase
Bases 1185-3445	176.86 CPS <i>HS41:25 (galE1)</i> <i>HS41:26 (udg)</i>	Putative sugar epimerase UDP-glucose 6-dehydrogenase
Bases 3460-4110	176.83 CPS <i>glf3</i>	Nucleotidyl sugar pyranose mutase
Bases 4912-5032	176.86 CPS <i>HS41.29</i>	Putative sugar transferase

## Chapter 4: Determining the gene content of the unknown and novel strains

---

RM3423, the long range product of strain 50702 is approximately 2.3 Kb shorter (10.9 Kb compared to 13.2 Kb) indicating that further sequence divergence has occurred. Perusal of Table 4.13 B shows that the genes *wlaUG*, *wlaUH* and *wlaUI* may be missing.

### Strains 54386 and 60319

These strains were similarly unclassified after initial examination by PCR screening and primer walking. Primer walking showed that primers from each group designed to the more conserved regions of the LOS cluster produced amplicons with these strains. Long range PCR with these strains produced amplicons of approximately the same size; 12.5 and 12.8 Kb. Initial sequencing analysis showed that forward from *lpxL* both strains appeared to have a group 3F LOS gene content. Examination of the sequencing in the reverse direction showed that this region had high sequence homology to the capsular polysaccharide region of strain 176.86 (Accession number BX54857). Comparisons against the NCBI databases showed that after base 225 this region produced no matches with any LOS region and only had homology to the capsule polysaccharide region.

### 4.5.5 Analysis of incomplete sequences

Incomplete sequences will be examined in order of fold coverage completed and involves a basic comparison on the basis of sequence homology which is summarised in Table 4.14 A-H.

#### Strain 38625

After primer walking this strain was classified as a group 2A/4A strain but sequencing has shown that this strain has homology to the group 2, class 'O' strain RM3423, a new group 2 LOS variant. Any further variation in this strain will have to be examined on completion of the sequencing.

### **Strain 51585**

The initial PCR screen indicated that this strain had a novel combination of ORFs but later analysis by primer walking indicated that this strain had a sub-group 1C type LOS gene content. Sequencing the long range product of this strain has indicated that it has significant homology to ATCC 43431 which has been classified as a group 2C strain. The long range product from strain 50702 is 2 kb smaller than that of the control strain indicating further divergence from ATCC 43431. Genes between *wlaUE* and *wlaVB* may be missing from this strain.

### **Strain 44406**

Strain 44406 was classified as a strain containing a novel combination of genes (1B/1C/2A) after PCR screening, but primer walking further classified this strain as having a sub-group 1C type LOS gene content. Sequencing analysis revealed this strain's LOS gene content to have sequence homology to sub-group 1B, but as the amplicon produced by long range PCR is smaller than that produced by the sub-group 1B strain 81-176 and larger than that produced by the sub-group 1A strain 11351, it may be an intermediate between these classes. Sequencing should be completed to confirm this LOS of this strain as an intermediary between these sub-groups.

### **Strain 47185**

Strain 47185 remained unclassified after primer walking as no conclusions could be drawn to its LOS type. Initial sequencing analysis has shown that this strain has a LOS gene content homologous to strain Tgh09011, a group 2 strain. The amplicon produced by long range PCR is larger than that expected from Tgh09011 by 500 bp, indicating that this is another group 2 LOS variant.

**Tables 4.14 a-h Summary of nucleotide identity for each strain.** Sequenced fragments and contigs are shown with their orthologous strains and ORFs together with percentage identity.

## Chapter 4: Determining the gene content of the unknown and novel strains

<b>a) 34218 13.75 Kb</b>		<b>0.86 x coverage</b>
Sequencing stage or contig (size/bp)	Sequence homology to	% identity
F1-3 (2555)	11168 <i>lpxL</i> , 1135, 1136, 1137, 1138	98
R1 (1040)	11168 1145, <i>waaV</i>	98
R2 (580)	11168 1145, <i>waaV</i>	98
R4-8 (3521)	11168 1136, 1137, 1138, 1139	98

<b>b) 51585 11.2Kb</b>		<b>0.83 x coverage</b>
Sequencing stage or contig (size/bp)	Sequence homology to	% identity
F1-2 (1760)	ATCC 43431- <i>lpxL</i> , <i>wlaX</i> , <i>wlaUA</i>	99
F3 (970)	ATCC 43431- <i>wlaUA</i> , <i>wlaUB</i>	99
F4 (1032)	NCTC 81116 - <i>wlaUC</i> , <i>wlaUD</i>	99
R1-2 (1578)	NCTC 81116- <i>wlaWC</i> , <i>waaV</i>	96
R3 (1250)	ATCC 43431- <i>wlaWB</i> , <i>wlaWC</i>	99
R4 (986)	ATCC 43431- <i>wlaVA</i> , <i>wlaWA</i> , <i>wlaWB</i>	94
R5 (870)	ATCC 43431- <i>wlaVA</i> , <i>wlaWA</i>	94
R6 (913)	ATCC 43431- <i>wlaVA</i> , <i>wlaWA</i>	94

<b>c) 44406 10.9 Kb</b>		<b>0.7 x coverage</b>
Sequencing stage or contig (size/bp)	Sequence homology to	% identity
F1 (578)	81-176 <i>lpxL</i>	88
F2-4 (2786)	11351 <i>lpxL</i> , <i>wlaNA</i> , <i>wlaNB</i>	95
F5 (1000)	81-176 <i>wlaNB</i> , <i>cgtA1</i>	94
F7-9 (1850)	81-176 <i>cgtB</i> , <i>cstII</i> , <i>neuB1</i>	90
R1-2 (1400)	81-176 <i>orf11</i> , <i>waaV</i>	93

<b>d) 47185 12.8 Kb</b>		<b>0.57 x coverage</b>
Sequencing stage or contig (size/bp)	Sequence homology to	% identity
F1-3 (2809)	Tgh9011 <i>lpxL</i> , <i>wlaNA</i> , <i>wlaUA</i> , <i>wlaUB</i>	98
F4-6 (2297)	Tgh9011 <i>wlaVB</i> , <i>wlaVA</i> , <i>wlaWA</i> , <i>wlaWB</i>	98
R1 (523)	Tgh9011 <i>waaV</i>	93
R2-3 (1695)	Tgh9011 <i>wlaWB</i> , <i>wlaWC</i> , <i>waaV</i>	96

## Chapter 4: Determining the gene content of the unknown and novel strains

<b>e) 34565 12.5 Kb</b>		<b>0.57 x coverage</b>
Sequencing stage or contig (size/bp)	Sequence homology to	% identity
F1-9 (5427)	81-176 <i>lpxL</i> , <i>wlaNA</i> , <i>wlaNB</i> , <i>cgtA</i> , <i>cgtB</i> , <i>cstII</i> , <i>neuB1</i>	97
R1-3 (1706)	81-176 <i>neuA1</i> , <i>orf11</i> , <i>waaV</i>	96

<b>f) 38625 12.5 Kb</b>		<b>0.54 x coverage</b>
Sequencing stage or contig (size/bp)	Sequence homology to	% identity
F1-2 (1302)	RM3423 <i>lpxL</i> , <i>wlaX</i> , <i>rmlA</i>	96
F3 (840)	RM3423 <i>rmlA</i>	98
F4 (951)	RM3423 <i>rmlA</i> , <i>rmlB</i>	95
F5 (1000)	RM3423 <i>rmlB</i> , <i>wlaUC</i> , <i>wlaUD</i> , <i>wlaUE</i>	91
R1-2 (1360)	RM3423 <i>waaV</i> , <i>wlaWC</i>	99
R3 (1196)	RM3423 <i>wlaWB</i> , <i>wlaWC</i>	95

<b>g) 39864 13.75 Kb</b>		<b>0.46 x coverage</b>
Sequencing stage or contig (size/bp)	Sequence homology to	% identity
F1-2 (1327)	NCTC 81116 <i>lpxL</i> , <i>wlaUA</i> , <i>wlaUB</i>	96
R1-3 (2323)	NCTC 81116 <i>wlaWB</i> , <i>wlaWC</i> , <i>waaV</i>	97
R4-8 (2726)	NCTC 81116 <i>wlaVB</i> , <i>wlaVA</i> , <i>wlaWA</i> , <i>wlaWB</i> , <i>wlaWC</i>	99

<b>h) 40973 11.2 Kb</b>		<b>0.4 x coverage</b>
Sequencing stage or contig (size/bp)	Sequence homology to	% identity
F1 (840)	MF6 <i>lpxL</i> , <i>wlaNA</i>	99
F2 (1040)	MF6 <i>wlaNA</i>	98
R1-4 (2566)	MF6 <i>waaV</i> , <i>orf11</i> , <i>neuA1</i>	98

### **Strain 34218**

Primer walking classified this strain as having LOS gene content with homology to NCTC 11168, group 1C. To date sequencing of this strain has confirmed this classification. This analysis may have also revealed that some of the genes between *Cj1140-1145* may be missing. The long range amplicon size is larger than the NCTC 11168 amplicon, perhaps indicating a change in gene content. The sequencing of this strain will need to be completed before any conclusions can be drawn.

### **Strain 34565**

This strain was classified as a sub-group 2B strain after primer walking but sequencing analysis has indicated that the LOS of this strain has a high degree of homology to the group 1B strain 81-176. Again strain 34565 produces a 900 bp larger amplicon during long range PCR than that expected of 81-176, perhaps indicating the presence of new ORF.

### **Strain 39864**

Strain 39864 was previously classified as a sub-group 2B/3A strain by PCR screening and a sub-group 2A strain after primer walking analysis. Sequencing analysis revealed homology to sub-group 2A strains but the long range PCR amplicon is smaller than that expected for sub-group 2A strains and larger than that expected for other group 2 strains indicating a further LOS variation in this group.

### **Strain 40973**

This strain was initially classified as a sub-group 1B/2A strain by PCR screening and further classified as a sub-group 1A strain after primer walking. The limited sequencing and analysis that has performed on this strain indicates that its LOS region has sequence homology to the LOS cluster of a sub-group 1B strain, MF6 (Accession number: AY422196). None of the sub-group 1B specific genes have been sequenced so no definitive conclusions can be drawn about the LOS sub-group this strain belongs to. Again the long range amplicon from this strain is smaller than that of group 1 strains (11.2 Kb

compared to 11.6 Kb) and larger than that of group 1A strains (10.9 Kb) indicating a further variant to this LOS group.

#### **4.5.6 Conclusion**

Sequencing analysis showed that any strains that had been highlighted as having a possible combination of LOS groups by PCR screen, did in fact only have one type of LOS gene content. Possible reasons for the initially misleading results include changes at the nucleotide level which provide homology to primers specific to another group or sub-group, or the primers annealing to another region in the genome. Table 4.15 shows a summary of LOS group after sequence analysis.

This sequence analysis also showed that in the cases of strains 34086 and 51566 which both have homology to sub-group 3A strains, SNPs in the sequence can cause changes in the proteins expressed. Further possible variation was seen in both group 1 and group 2 strains, with strains that have been incompletely sequenced showing possible variation based on the sequencing and the length of the long range amplicon produced compared to those that currently belong to the group. Interestingly this research also highlighted the interaction between LOS and the capsule polysaccharide and shown that recombination may occur between these loci.



**Table 4.15 Summary of the LOS group assigned to each strain after sequencing and any possible deviations from the sub-group.**

## Chapter 4: Determining the gene content of the unknown and novel strains

Strain	PCR	Primer walking	Sequencing	Differences from current sub-groups
34086	UC	3A	3A	Premature translation stops in <i>orfA</i> , <i>orfC/D</i> & <i>orfE</i>
34218	UC	1C	1C	Part of cluster missing?
34565	UC	2B	1	Similar to new variant in database
34806	1B/3A	2A	3A	
38625	UC	2A; 4A	2	Similar to new variant in database
39864	2B/3A	2A	2A	Smaller than 81116 and larger than other group 2 strains.
40973	1B/2A	1A	1B	Smaller than group 1B LOS
44406	1B/1C/2A	1B	1B	Smaller than group 1B LOS
47185	UC	UC	2	Larger than Tgh09011, smaller than 81116
49068	1B/2A	NA	NA	
50702	UC	1A; 1B; 2A	2	Similar to new variant in database
51566	UC	2A	3A	Premature translation stops in <i>orfA</i> , <i>orfB</i> & <i>orfE</i>
51585	1C/2A	1C	2	ATCC 43431 genes between <i>wlaUE</i> and <i>wlaVB</i> missing?
53305	UC	1A; 2A; 2B	NA	
54386	UC	UC	3	Group 3A and capsule genes
57073	1C/2A	2A; 2B	NA	
60319	UC	2A; 2B	3	Group 3A and capsule genes
61666	UC	UC	NA	

## **4.6 Discussion**

### **4.6.1 Primer walking**

Primer walking experiments proved inconclusive for many strains. PCR was repeated at 45°C and 55°C which should have allowed for primer mismatches at the lower temperature and also high specificity at the higher temperature. The inconclusive nature of primer walking for these unclassified and novel strains is perhaps unsurprising as the primers are based on known sequences. Primer walking provides an overview of the LOS gene content and an implication of changes in gene content by changes in amplicon size but cannot provide information about changes in the sequence, SNPs and other mechanisms of variation. Other studies utilising primer walking (Millar, 2003) have used the process to imply LOS group and have used sequencing to fully describe novel LOS gene content.

### **4.6.2 Phenotyping**

Examination of LOS composition by SDS-PAGE and silver staining provides information about the size of the molecule expressed by each strain and may provide information about the presence of homopolymeric tracts and antigenic switching, but no information can be gained about the gene content of the LOS biosynthesis locus. Although in general strains of the same subgroup seem to have a similar size LOS, but the LOS samples from these strains would have to be run on a larger gel to confirm this. Dot blotting with cholera toxin also confirms the presence or absence of GM1 like epitopes which were not present in these strains.

Examination of LOS extracts by SDS-PAGE provides no information about LOS structure and therefore mass spectroscopy should be used to determine the LOS core structures of these strains. Resolution of LOS structure is important as some studies have found that determination of LOS cluster gene content may not directly inform LOS structure. LOS structure is also affected by mechanisms of variation such homopolymeric tracts,

frameshift and missense mutations (Godschalk *et al.*, 2007). This study also found that within the group 1A strains, five different LOS structures were found. Of the twenty six GBS/MFS associated strains examined, fourteen strains expressed a mixture of at least two differing outer core structures.

### 4.6.3 Long range PCR

After optimising many kits and trying DNA extracted using non-organic methods, long range PCR was successfully applied for the amplification of the majority of LOS clusters from the CSSS strains. The discovery that the DNA quality and purity influenced the success of the long range reaction highlights the importance of component quality for this process.

The size of the amplicon may provide a rough gauge of LOS group as strains with smaller LOS cluster are more likely to be members of groups 3 or 4. Parker *et al.* (2005) reported using long range amplicon size as an indicator of LOS gene content but failed to provide the data to re-enforce this statement. Examination of the sequencing data also shows that changes in gene content may cause assumption of LOS group by long range amplicon size to be misleading and lead to erroneous results.

The LOS from four strains could not be amplified despite optimisation of the PCR and re-extraction of the DNA. This may be due SNPs in the primer sites or they may require the use of a lower annealing temperature. Amplicon size is also useful when comparing sequencing to known strains as differences in size may imply change in gene content.

### 4.6.4 Sequencing analysis

Sequencing of the long range amplicons enabled further classification of the novel and unclassified strains. Eight of the strains shown were over 70% complete allowing greater conclusions to be made about their LOS gene clusters.

Strains 34086 and 51566, although both classed as sub-group 3A demonstrate that methods based on DNA homology do not always reflect the nature of the molecule produced, as full sequencing and analysis revealed

truncated proteins due to missense mutations and premature translation stops. The inactivation of *orfA* (which has homology to *wlaNA* from NCTC 11168) by the insertion of a single base is seen in group 3A, but the effect of this gene inactivation remains unknown as the LOS structure for this group remains to be fully elucidated (Gilbert *et al.*, 2005). This glucosyltransferase gene, *orfA* is also subject to non-phase variable frameshift mutations which lead to the expression of one or two-domain glucosyltransferases.

Sequencing analysis has also revealed the presence of homopolymeric tracts, which cause variation by slip strand mispairing in many of the strains. In four of the strains, this has caused technical issues with the sequencing as these strains have produced mixed signals. Nine ORFs in the LOS core biosynthesis cluster have been reported to contain phase variable poly G tracts including; *cgtA*, *wlaN*, *cstII*, *orf11*, *Cj1144/45*, *orfE*, *wlaUD* and *wlaUF* and there is also one example of a poly A tract in the *cstII* gene of the MF6 LOS gene cluster (Gilbert *et al.*, 2002; (Gilbert *et al.*, 2005). Phase variation is often associated with virulence genes in bacteria such as *H. influenzae*, *H. pylori* and *Neisseria* spp. and in *C. jejuni* homopolymeric tracts are commonly associated with the genes involved in carbohydrate biosynthesis.

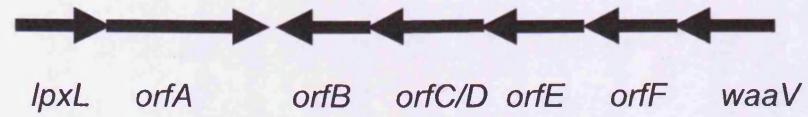
Importantly strains 54386 and 60319 highlighted the interaction between the LOS and capsule polysaccharide loci, as the LOS gene cluster of these strains contain 4 genes from the CPS cluster of the HS: 41 strain 176.83 sequenced by Karlyshev *et al.* (2005a). Figure 4.4 shows a schematic representation of the known ORF arrangement in these strains. The genes include *galE1*, *galE2* and *udg*. The protein UDG is required for the biosynthesis of glucuronic acid and it would be interesting to see if this residue was present in the LOS structure of these strains.

Further information about the amount of variation in groups has been discovered, implying that although LOS group may provide an idea of general gene content it may not provide the full picture. In both groups 1 and 2, new strains have been seen to have different long range amplicons than their

**Figure 4.4 Schematic diagram of the *lpxL-waaV* region of *C. jejuni* RM1221, 60319 and 54386.** Arrows indicate ORFs and direction of transcription. Gene names are based on existing nomenclature (Miller, 2003). Red arrows indicate addition of capsule gene from strain 176.83 (Karlyshev *et al.*, 2005a) and the yellow arrow indicates a predicted ORF of unknown identity. Diagrams are not to scale.

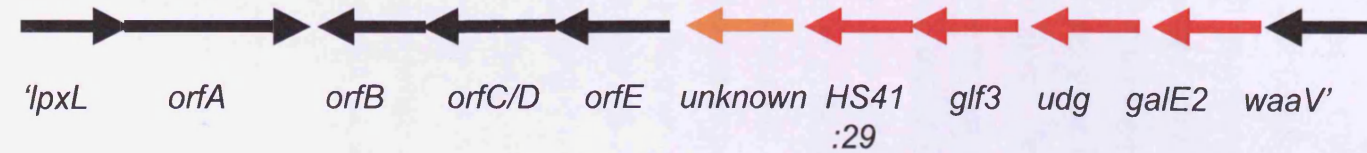
RM1221

6.9 Kb



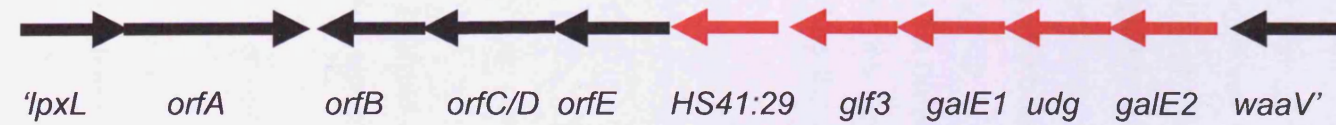
60319

12.8 Kb



54386

12.5 Kb



## Chapter 4: Determining the gene content of the unknown and novel strains

---

founder members, implying changes in gene content and the presence of LOS clusters that are intermediaries in the sub-group system. Evidence of recombination and lateral gene transfer has previously been seen amongst the group 1 LOS type members. Gilbert *et al.* (2002; 2005) observed that due to sequence identity group 1B strains evolved from group 1A strains with strains such as ATCC 43438 representing an evolutionary intermediate. The events that occurred to produce group 1B strains involved the duplication of *cgtA*, lateral transfer to alter the sequence from *cgtA* to *cgtB*, and finally the insertion of a 'C' terminus cassette in this region which leads to a large number of amino acid substitutions. Group 1B has also been proposed as the evolutionary intermediary between group 1A and group 1C strains, as group 1C strains also contain the duplicated *cgtA* gene as a gene fusion with *neuA*. Other recombination events have occurred to cause the deletion of *cgtA* and the insertion of *wlaO* and *wlaP* in its place, as well as the insertion of *wlaSB* between *neuA* and *waaV* (Gilbert *et al.*, 2005).

There have been various reports of gene deletions occurring in group 2 strains. The strain O:3 was reported to be missing the gene *wlaUJ* (Millar, 2003). There has only been one observation of insertion and deletion events occurring in this group as highlighted by the strains ATCC 43431 where genes *wlaUG* and *wlaUI* have been deleted, and a gene encoding a butyryltransferase has been inserted in between *wlaUF* and *wlaUH* (Parker *et al.*, 2005). Variation in this group appears to occur between genes *wlaUF* and *wlaVB* indicating that the deletion of these genes in strain 50702 may be a possibility.

Sequencing has also shown that there are as yet unclassified strains in the public databases. The information gained in this research can be used to update and refine the group system as described in Chapter 3. It will also enable the design of more specific primers and new primer sets. Figures 4.5, 4.6 and 4.7 are schematic diagrams representing all the LOS groups to date and any strains with new gene arrangements.

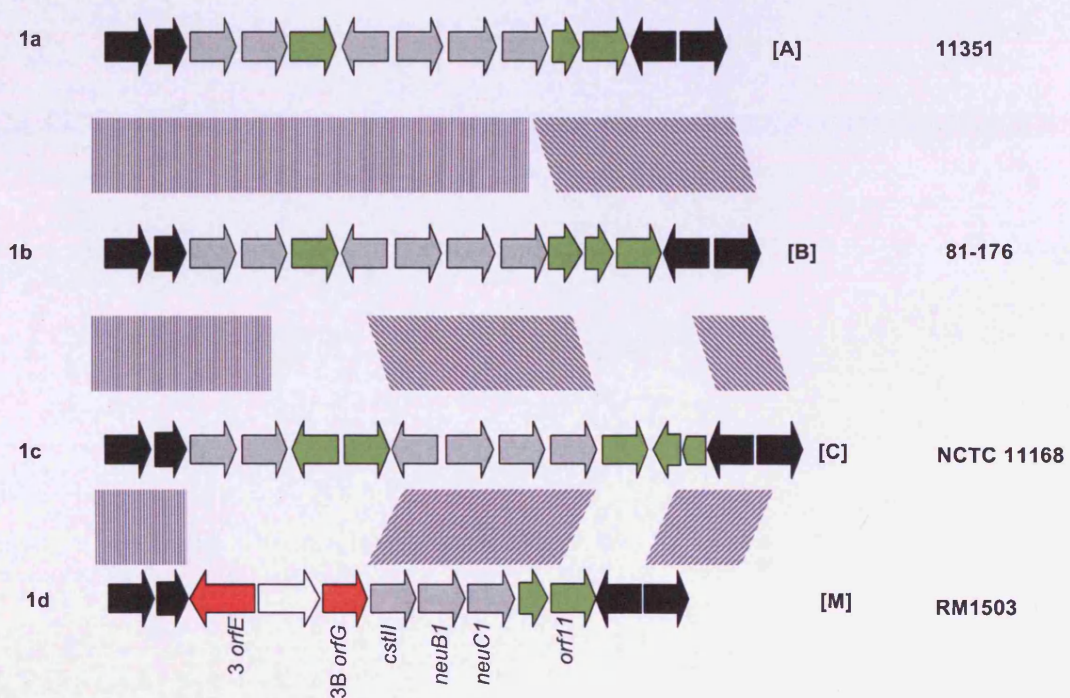


### 4.6.4 Correlations between LOS group and other typing methods

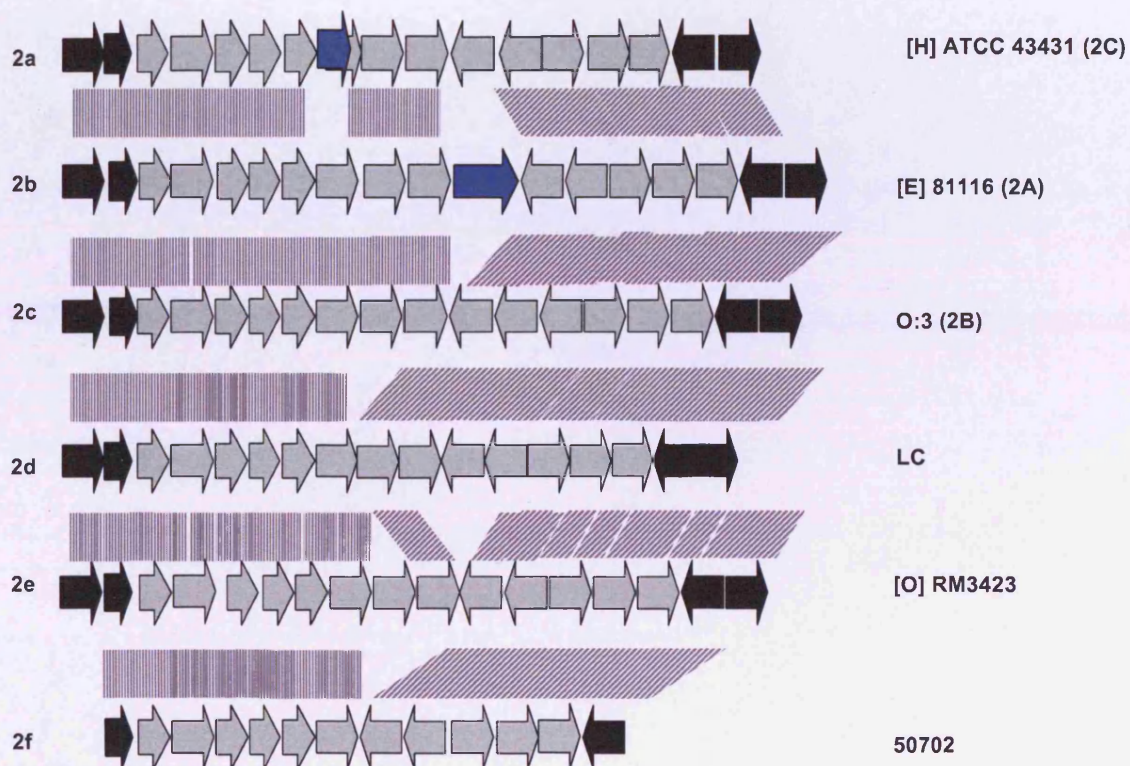
Re-examination of the MLST results with the new information about the novel strains re-enforces a link between MLST clonal complex and LOS group as discussed in Chapter 3, section 3.2.5. Eight of the nine strains classified as sub-group 1C belong to clonal complex ST21, 50% of the sub-group 1B strains are clonal complex ST48 strains. Most of the group 2 strains belong to clonal complex ST45 and most group 3 strains belong to ST257. Clonal complexes ST21 and ST45 are the two most common clonal complexes associated with human isolates in the UK (Dingle *et al.*, 2002). To enable definitive correlations to be drawn between these molecular methods a larger strain set would have to be examined but confirms the potential link between LOS group and ST seen by Millar (2003).

Comparisons between LOS group and serotype show no correlation, except that each serotype is only associated with one or two LOS group types. For instance, HS:2 is associated with group 1 or 2 LOS types and HS:31 is associated with LOS group 3. This link maybe due to the limited strain set used, both in number and origin. Other studies have also not found any links between Penner serotype and LOS group (Parker *et al.*, 2005). No correlation can be formed between phage type and LOS group as all the LOS groups appear to be represented in all the phage types tested.

**Figure 4.5 Schematic diagram showing the organisational classification for group 1.** The black arrows represent genes conserved throughout the group, such as *waaC* and *waaV*. Grey arrows represent genes that are shared by members of the same group. Genes which are divergent are highlighted by coloured arrows. Regions of similarity in the group are highlighted by the stripes. The red arrows represent ORFs from group 3 strains. Letters in square brackets indicate nomenclature adopted by Gilbert *et al.* (2002), Poly *et al.* (2004), Millar (2003) and Parker *et al.* (2005; unpublished). Example strains are indicated with each group. Adapted from Karlyshev *et al.*, 2005.

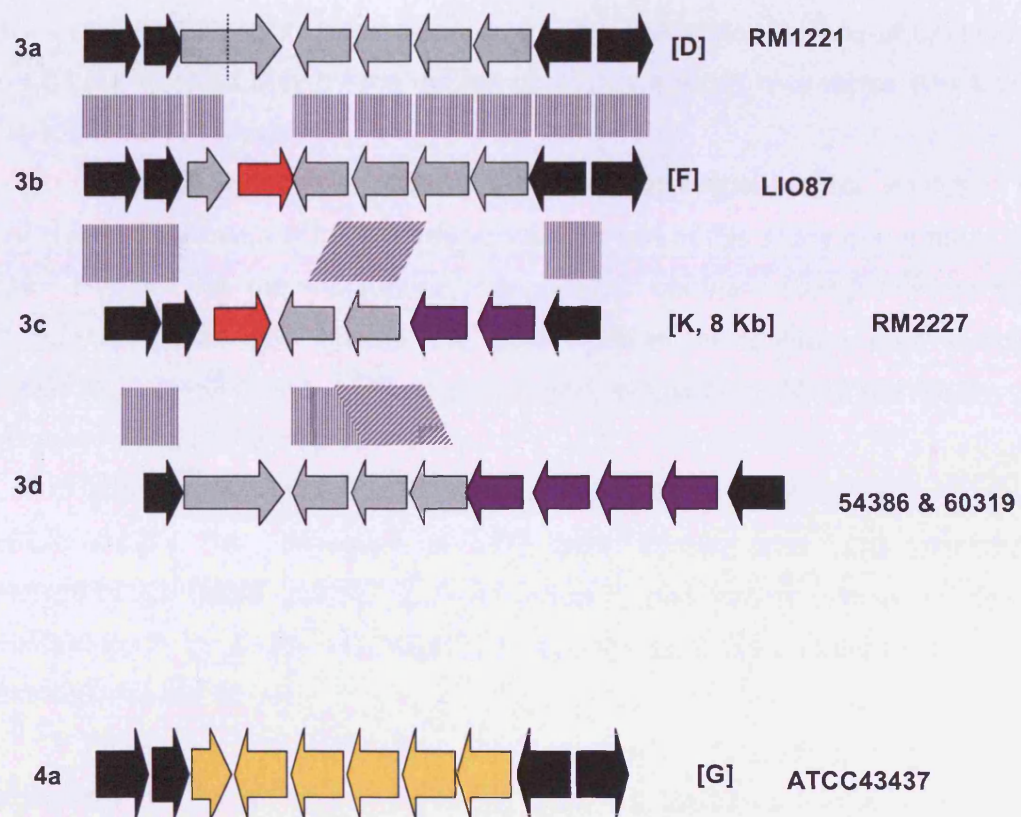


**Figure 4.6 Schematic diagram showing the organisational classification for group 2.** The black arrows represent genes conserved throughout the group, such as *waaC* and *waaV*. Grey arrows represent genes that are shared by members of the same group. Genes which are divergent are highlighted by coloured arrows. Regions of similarity in the group are highlighted by the stripes. Letters in parentheses indicate nomenclature adopted by Gilbert *et al.* (2002), Poly *et al.* (2004), Millar (2003) and Parker *et al.* (2005; unpublished). Example strains are indicated with each group. Adapted from Karlyshev *et al.* (2005). Various strains have had genes inserted and deleted as follow: O:3, *wlaUJ* deleted; LC, *wlaUG* and *wlaUI* deleted; RM3423, *wlaUI* and *wlaUJ* deleted; 50702, possible deletions between *wlaUE* and *wlaVB*; ATCC 43431, deletion of *wlaUI* and *wlaUG* and insertion of a butyryltransferase between *wlaUF* and *wlaUH*.



**Figure 4.7 Schematic diagram showing the organisational classification for groups 3 and 4.** The black arrows represent genes conserved throughout the group, such as *waaC* and *waaV*. Grey arrows represent genes that are shared by members of the same group. Genes which are divergent are highlighted by coloured arrows. Regions of similarity in the group are highlighted by the stripes. The purple arrows represent capsule genes. Letters in parentheses indicate nomenclature adopted by Gilbert *et al.* (2002), Poly *et al.* (2004), Millar (2003) and Parker *et al.* (2005; unpublished). Example strains are indicated with each group. Adapted from Karlyshev *et al.*, 2005.





### 4.6.5 Conclusion

LOS gene content is known to inform LOS structure, the assignment of LOS group by PCR screening and primer walking can also inform LOS gene content. Both these PCR based methods can only provide an overview of gene content in this region, only completely sequencing the region can provide complete information about this region. Another method of analysing LOS gene content may be the design of a locus specific microarray which will be discussed in Chapter 5.

Completion of the sequencing and subsequent data analysis is required to confirm any initial observations formed in this study about the LOS gene content for the incompletely sequenced strains. Completion of the sequence would also enable the establishment of recombination events occurring between the LOS and capsule loci and confirm the point of recombination in the cluster.

Subsequent determination of LOS core structure of all these strains would enable the correlation of LOS gene content and LOS structure, especially for those groups, such as group 3, and strains with novel gene content such as 54386 and 60319, where the LOS core structure has not been elucidated as yet.

Furthermore to enable the establishment of LOS group classification alongside molecular typing methods, such as MLST, a larger strain set encompassing strains of more diverse origin would need to be examined for both MLST and LOS type to establish a relationship.



## **Chapter 5: Design and validation of a locus specific microarray**

### **5.1 Introduction**

#### **5.1.1 Current typing methods**

Two serotyping methods form the basis of surveillance and epidemiological typing of *C.jejuni*: the Penner heat-stable (HS) scheme (Penner and Hennessy 1980) and the Lior heat-labile scheme (Lior *et al.*, 1982). Both of these schemes are labour-intensive, subject to variability and costly due to the large panel of antisera that are needed. The Campylobacter Reference Unit of the Health Protection Agency developed a modified Penner serotyping (Laboratory of Enteric Pathogens) scheme in order to reduce problems with non-specific agglutination and cross-reactivity (Frost *et al.*, 1998). These methods provide a system to classify *C.jejuni*; there are over 60 HS serotypes, more than 100 Lior serotypes and approximately 47 LEP serotypes. Other phenotypic techniques such as phage typing are also used to provide further discrimination, but some strains remain untypeable by these methods.

The genetic diversity of *C.jejuni* has been shown by methods such as ribotyping, Pulsed Field Gel Electrophoresis (PFGE) and PCR-based methods, such as Random Amplified Polymorphic DNA analysis (RAPD). These methods, although highly discriminatory, are subject to inter-lab variation and there difficulties in comparing results. Campynet was established to help the standardisation of three methods; *fla*-PCR RFLP (Restriction Fragment Length Polymorphism), PFGE and AFLP (amplified fragment length polymorphism;

[www.medvetnet.org/cms/templates/docphp?id=65](http://www.medvetnet.org/cms/templates/docphp?id=65)).

Multilocus sequence typing (MLST) has become the 'gold standard' for molecular typing and is easily comparable between laboratories. Dingle *et al.* (2001) have used this method to confirm the genetic diversity of *C. jejuni* and have shown the bacterium to be weakly clonal. Strains are grouped into clonal complexes when two or more sequence types (ST) share identical alleles at four or more loci. Members of each clonal complex are believed to have arisen from a common ancestor. In a further study, Dingle *et al.* (2002) identified 379 ST from 814 human and animal isolates, with 63% of the human isolates belonging to six clonal complexes. MLST has also been used to indicate the contribution of recombination and mutation in the diversification of clonal complexes. Changes in alleles may arise by point mutation or recombination and the population structure of *C. jejuni* is thought to have arisen by recombination (Dingle *et al.*, 2001b). Although MLST provides details of the genetic diversity of housekeeping genes, it does not enable further characterisation of genetic variation at other loci. Together with PFGE and RAPD, MLST only provides a very limited snapshot of the *Campylobacter* genome. As yet none of the typing methods described above have been able to link phenotypic characteristics with pathogenicity, virulence or ecological niche.

### **5.1.2 Microarrays**

The availability of the whole genome sequence of *C. jejuni* allowed the production of microarrays and the development of 'genomotyping' or genome indexing, which promised correlations between phenotype, pathogenicity and virulence to be drawn. Microarrays can be divided into three types; the Affymetrix GeneChip™ which involves the *in situ* synthesis of oligonucleotides on silica wafers at high density to produce one channel of information; membrane arrays which involve the printing of reporter elements on nylon membranes that are used in a similar way to Southern blot hybridisation and glass slide microarrays where double-stranded DNA or single-stranded oligonucleotides are spotted onto the glass support at high density using specially designed robotic equipment. Glass microarrays incorporate two

colour technology, enabling the user to label samples with different fluorescent dyes for competitive hybridisation.

Two types of hybridisations are performed with microarrays; those that examine the transcriptome and gene expression (Stintzi 2003; Elvers *et al.*, 2005; Stintzi *et al.*, 2005; Sampathkumar *et al.*, 2006) and those that look at comparative genomics or the differences in gene content at a whole genome level. The comparison of the genome sequenced variant of NCTC 11168 and the original clinical isolate of NCTC 11168 by Gaynor *et al.* (2004) revealed that variation in the transcriptome can also be a mechanism for variation where molecular genotyping methods such as PFGE and MLST did not reveal genetic differences.

### 5.1.3 Genomotyping of *C.jejuni*

Dorrell *et al.* (2001) performed the first whole genome comparison of *C.jejuni* against a clone array constructed from sequencing clones constructed from the sequencing library of the NCTC 11168 genome. Amplified clone inserts representative of the coding sequences were spotted onto glass slides. Only 34.5% of the clones selected represented a single gene, with 65.5% of the clone inserts containing one or more adjacent gene fragment. This gave rise to cross-hybridisation and distorted the hybridisation signal for the desired gene and the reporter element. This study showed on a genomic level that the central variable region of the capsule biosynthesis locus was responsible for the serotype. Four strains with the O:2 serotype showed complete conservation in capsule biosynthesis locus and strains with different serotypes showed variation in this region. This whole genome comparison also showed that approximately 1300 of the 1654 coding sequences were common to all strains and the remaining 354 (21%) genes were absent or highly divergent amongst strains. The microarray data also reflected the variability of the LOS outer core. Pearson *et al.* (2003) found variable regions located in seven distinct genomic regions or plasticity regions. Three of these plasticity regions contained the genes for the *C.jejuni* glycome (capsule, LOS biosynthesis and flagellin modification loci). Perhaps

due to a larger number of strains or to the microarray design (one reporter element per CDS), only 16% of the genes present in strain NCTC 11168 were absent or highly divergent in the test strains.

In a later study, Taboada *et al.* (2004) compared the whole genome analysis of 51 strains to the data from three previous studies (Dorrell *et al.*, 2001; (Leonard II *et al.*, 2003); Pearson *et al.*, 2003) and showed that a large proportion of the variable genes were absent or highly divergent only in single strains. Taboada *et al.* (2004) proposed that large regions of the *Campylobacter* genome were genetically stable and of the highly divergent cds identified, 95% of these genes had divergent neighbours and showed intraspecies variability.

In an effort to find markers to identify potential GBS-antecedent strains, Leonard II *et al.* (2004) performed whole genome comparisons with 12 GBS associated isolates and 12 enteritis associated isolates. Findings confirmed significant genetic heterogeneity but failed to discover GBS specific markers, with the many regions of variation being the same amongst both groups of isolates.

Champion *et al.* (2005) performed microarray analysis with Bayesian-based algorithms to discover genetic markers for the source of infection. In total 111 isolates from diverse environmental and animal sources, and human isolates with diverse clinical presentations were analysed against a *Campylobacter* ORF specific whole genome microarray. The analysis of these strains enabled them to be grouped in two distinct groups; a livestock- and a non-livestock-associated clade. Strains that were grouped in these clades may not have grouped together by traditional serotyping methods. Strains that produced different clinical presentations were not from a common ancestor and did not form distinct clonal populations, in agreement with the analysis of GBS and enteritis strains by Leonard II *et al.* (2004). Analysis also revealed that over 55% of the clinical isolates were related to strains from non-livestock sources suggesting that environmental sources contribute more to campylobacteriosis than previously suspected. Analysis of the capsule biosynthesis, N-linked glycosylation and LOS biosynthesis loci showed that they could not be assigned to a specific clade. Interestingly the presence of six genes within the O-linked flagellar glycosylation locus, putatively involved

in the carbohydrate modification of the flagella proved indicative of the strains belonging to the livestock clade.

Genomotyping can be used as a complementary genotyping tool and indeed Leonard *et al.* (2003) used DNA microarrays to further differentiate strains that had been clustered epidemiologically by RAPD and PFGE. Whole genome comparisons by microarray may also prove to be highly complementary to MLST but as yet no studies examining MLST and the whole genome have been published.

Another application of microarray technology was studied by Poly *et al.* (2004), which still was unable to fully answer the problem of microarray providing information about unknown or divergent sequences. A shotgun library from the non-sequenced strain ATCC 43431 was printed on glass slides and after competitive genomic hybridisation with the NCTC 11168 sequenced strain, DNA unique to ATCC 43431 was identified. In total 84 Kb of unique DNA was discovered using this approach including potential molecular markers. This process was repeated with the then non-sequenced strain 81-176 and 87 new genes were identified (Poly *et al.*, 2005).

The whole genome sequence of *Campylobacter jejuni* RM1221 showed another type of diversity within the species and contains four *Campylobacter jejuni* integrated elements (CJIEs); genomic islands with Mu-like phage, phage-related endonucleases and integrated plasmids. Parker *et al.* (2006) combined a PCR-based assay with comparative genomic hybridisation (CGH) with a microarray designed to contain NCTC 11168 and RM1221-specific ORFs to examine *C.jejuni* and *C. coli* strains. The strains were screened by PCR for the presence of CJIE-like clusters and 55% of *C.jejuni* strains were found to contain one or more of these clusters. A subset of 35 strains was then examined by CGH and 21.5% of the genes were found to be highly divergent or absent in one or more strains. Diversity was also seen to occur within the CJIEs and in agreement with other studies the majority of genetic diversity occurred at known hypervariable regions.

Microarray technology goes hand in hand with some inherent limitations. PCR amplicon based arrays due to the size may be subject to cross-hybridisation leading the false-positive signals. Although microarray design has evolved to limit this problem, it may still occur, for instance

between paralogous gene families, i.e. Cj0617 and Cj1318 and the *tlp* genes (Dorrell *et al.*, 2005). Microarrays are also unable to provide information about gene rearrangements, gene insertions, chromosomal location, small deletions and point mutations. Novel genes that are not included on the microarray are also not detected. Many microarrays also do not include intergenic regions and therefore variations in promoter sequences and small RNAs cannot be scrutinised.

Microarray technology promises to be a quick and easy method for genotyping strains and as yet while all of these studies have served to increase knowledge regarding conserved and variable genes in the genome, no concise method of typing the glycome has been developed. The ability to vary these structures is important in campylobacter's lifestyle, therefore the ability to type the genes relating to these structures may provide an indication of lifestyle, environmental niche or virulence.

The aim of this study was to develop a locus-specific microarray-based tool that could be used to determine gene content, new genetic material and gene context for the lipooligosaccharide region, providing a complementary tool to other molecular typing techniques.

## **5.2 Microarray design and rationale**

The use of microarray technology for molecular typing is especially limited by the inability to detect changes in gene content, whether that involves gene insertions, gene deletions or gene rearrangements. To solve this issue, it was hypothesised that if reporter elements were designed to intergenic regions or the boundaries of open reading frames any changes in gene content at these regions maybe detected. Application and advancement of this technology is pertinent for variable regions such as the LOS core biosynthesis cluster. The aim of this microarray was to develop a tool that would indicate the LOS core gene content for a clinical strain, but also provide details about changes in gene order and gene context. The microarray

needed to be highly specific and be able to distinguish changes in gene content indicated by changes in the nucleotide sequence.

To test the specificity and viability of the different microarray technologies the preliminary microarray experiments involved the testing of both arrayed PCR products derived from the gene boundaries and 70-mer oligonucleotides derived from gene boundaries. The gene boundary PCR amplicons were derived from the NCTC 11168 sub-group 1C LOS region using the primers designed by BμG@S for the *Campylobacter jejuni* whole genome microarray, as listed in Appendix 1. This version of the microarray was called GMv1 and was used to establish this approach.

Oligonucleotide design was performed by Operon (Germany) for both gene-specific and gene-boundary 70-mers. For the gene-boundary oligonucleotides the Operon design team were supplied with 100 bp of each intergenic or boundary region. This limited the available sequence and altered the parameters they used for oligonucleotide design. Where open reading frames were separated by intergenic regions, the oligonucleotides were designed to cover the intergenic region and the start or end of the gene. No gene-boundary oligonucleotides could be designed for AB-wlaNA, C-1136-1137, C-*waaF*, D-12953, D-12951-12950, Etgh022-tgh160 and RM1221-orfA-orfB within acceptable parameters for T<sub>m</sub>, secondary structure and repetitive sequences.

It was predicted that oligonucleotides would provide the best approach and as such gene-specific oligonucleotides were designed for use in later versions of the microarray. For this aspect of the oligonucleotide design Operon used their normal primer design criteria with supplied sequences and all the open reading frames were successfully represented by a gene-specific oligonucleotide (see Appendix 2). Seven of the genes used in MLST were also included in the design process for use as controls; *aspA*, *glnA*, *gltA*, *glyA*, *pgm*, *tkt* and *dnaA*. When open reading frames appeared to be gene fusions, produced two-domain proteins or contained premature stop codons, they were represented by two separate oligonucleotides. The locus specific microarray was printed alongside the Bacterial Microarray Group's (BμG@S) *Campylobacter jejuni* version 2 PCR product whole genome microarray, which performed a role as an internal control for the hybridisations as well as

enabling comparisons at a whole genome level. The sequences of the oligonucleotides designed by Operon were checked against the provided target sequences and also for cross-hybridisation against the other LOS sub-groups by a BlastN-algorithm (results provided in Appendix 2).

### ***5.3 Validation and results from GMv1 locus specific microarray***

#### **5.3.1 Microarray Validation and results**

The first version of the locus-specific microarray, GMv1, consisted of a subset of gene-boundary oligonucleotides designed to the group 1 strains and PCR amplicons derived from the sub-group 1C LOS core region.

The microarray was initially tested using genomic DNA from LOS sub-class control strains in various combinations. Hybridisations were carried out as detailed in Chapter 2, following the standard protocol used by the BµG@S group. Image acquisition was performed using an Affymetrix 428 scanner and Imagen 5.2 software. Further data handling was performed using Bluefuse, which provides software based quality control. This program also assigns a 'PON' value which allows the assessment of binary changes between the test and control samples. It also assigns confidence values to hybridisation for each reporter, which are rated by flags. These values in combination were used to assess whether the reporter element detected the gene content of the relevant LOS sub-class control strain and any changes in the gene content presented by different sub-classes.

Initial validation was only performed using boundary oligonucleotides from group 1 LOS type. All of the test samples were hybridised against NCTC 11168 and dye swaps were performed to ensure that the results produced were not biased by the dye used (data not shown). The complete data set produced after analysis by Bluefuse can be found in Appendix 2. Two CSSS strains were also included in this study; 34806 which was designated as having an unknown LOS group type by PCR screening and 59653 which was designated as sub-group 1a.



The initial observation found on scanning the slides was that the signal generated by the locus-specific array, including both oligonucleotides and PCR products was extremely low compared to the signal produced by the whole genome microarray. Second to this the gene boundary PCR products had a much lower signal than the oligonucleotides and during the analysis these results were discarded on the basis of poor confidence levels. Table 5.1 summarises the results obtained with the various control strains and CSSS strains, where the confidence flag was above C. Red boxes indicate signal in the 'test' or Cy5 channel, green boxes indicate signal in the 'control' or Cy3 channel and yellow boxes indicate signal from both channels. The letter N is indicative of an oligo that produced little or no signal in either channel and as such the confidence was flagged below C. Where asterisks have been used as highlighters the result is unexpected. The results produced by the sub-group 2A strain, O:3, should be discounted as poor signal was produced with the majority of oligonucleotides. The most likely explanation for this may be that the DNA hybridised to the microarray was of poor quality.

### **Sub-Group 1A and 1B Boundary Oligonucleotides**

These oligonucleotides performed with mixed results. Little or no signal was produced with these oligos when they were tested using NCTC 11168 and strains that did not belong to group 1 and therefore they can be used to discriminate LOS type at a group level. Unfortunately signal was produced using sub-group 1B oligos with sub-group 1A strains, but not with group 1C control, suggesting cross hybridisation within these genomes. *In silico* analysis of these oligos reveals that B-*neuC-cgtAII* binds within the neuC-neuA boundary of sub-group 1A strains with an 8 bp mismatch. No explanation is apparent for the hybridisation of the other sub-group 1B oligo to group 1A strains or the hybridisation of the sub-group 1C control to the AB-*orf11-waaV* oligonucleotide. Although as the whole genome sequences for some of these strains are not available, cross-hybridisation in other parts of the genome cannot be ruled out.

The CSSS strain 59653, which had previously been classified as a sub-group 1A LOS type cannot be classified as such from the preliminary microarray results. More positive results were seen for the CSSS strain 34806, classified as 1B/3A, as signal for AB-*cgtA*, AB-*neuC-neuA*, and AB-*orf11-waaV* was seen in the Cy5 channel with this strain indicating that these sequences are present within this genome.

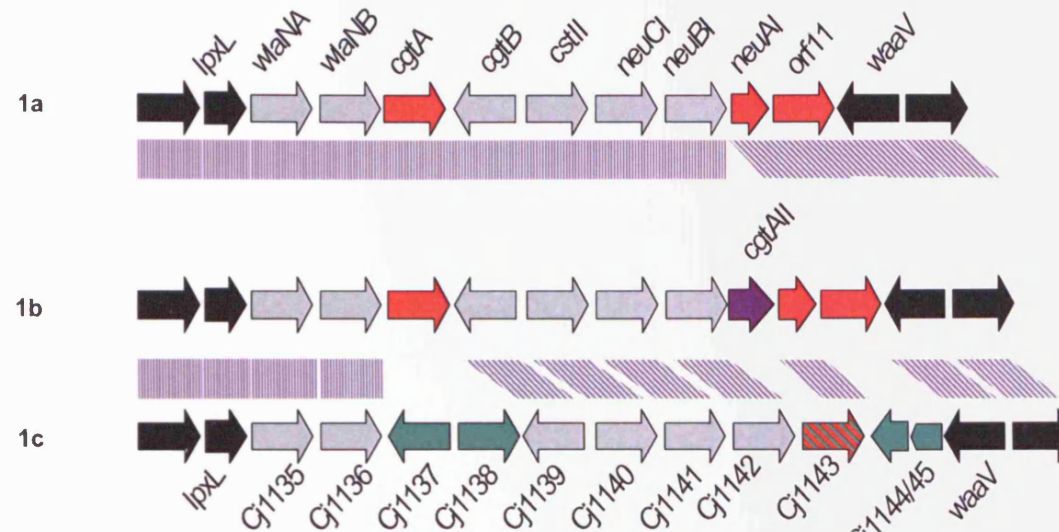
### **Sub-Group 1C Gene Boundary Oligonucleotides**


















































































Despite *in silico* analysis, there was expected to be some cross-hybridisation between the majority of the sub-group 1C oligonucleotides with other group 1 strains. The discrimination between these sub-classes was expected to be provided by the sub-group 1A and 1B specific oligonucleotides as well as oligos between C-1135 and C-*waaV*. Signal was produced from all of the sub-group 1C oligonucleotides in both channels with sub-group 1A strains. Although there appears to be shared sequence amongst some of the boundary oligos with conserved genes such as *galE* and *wlaA*, these occurrences do not explain the signal produced. Similarly signals from the 1C oligonucleotides were also seen in both channels from competitive hybridisation with sub-group 3A strains, RM1221 and O73. When tested with the sub-group 1B and sub-group 2A strains, the 1C oligonucleotides can be used to discriminate the LOS type. No difference could be determined between C-*waaF* and C-*gmhA*, where strains such as 81116 are known to contain gene insertions (Oldfield *et al.*, 2000).

**Table 5.1 Hybridisation results from the first generation locus specific microarray**, where N signifies signal was nil or under the confidence level, \* indicates a result unexpected by BLAST analysis, ■ indicates signal in test strain channel, ■ indicates signal in both the test and control channels and ■ indicates signal in the control channel (NCTC11168).














































































































All the oligonucleotides used in this experiment were designed to the gene-boundaries and oligonucleotide designation to each sub-group is indicated at the start of the oligonucleotide name and by the font colour.

The diagram below shows the different LOS gene content for group 1. The black arrows represent genes conserved throughout the groups, such as *waaC* and *waaV*. Grey arrows represent genes that are shared by members of the same group. Genes which are divergent are highlighted by red arrows. Regions of similarity in the group are highlighted by the stripes



BoundaryOligo	Samples								
	NCTC 11168 as 'Control' in Cy3 channel								
	11351(1A)	480(1A)	81-176(1B)	81116(2A)	O:3(2A)	RM1221(3A)	O73(3A)	59653(1A)	34806(1B/3A)
<b>AB-cgtA</b>		* 	*N	N	N	N	N	N	
<b>AB-cgtA-cgtB</b>				N	N	N	N	N	*N
<b>AB-neuC-neuA</b>	*N			N	N	N	N	N	
<b>AB-neuA-orf11</b>			*N	N	N	N	N	N	*N
<b>AB-orf11-waaV</b>	* 	* 	* 	N	N	N	N	N	
<b>B-neuC-cgtAll</b>	*N	* 		N	N	N	N	N	N
<b>B-cgtAll-neuA</b>	* 	* 		N	N	N	N	N	N
<b>C-galE</b>					*N	*N			
<b>C-1132</b>									
<b>C-waaC</b>					*N				
<b>C-waaC-lpxL</b>									
<b>C-lpxL-1135</b>					N				
<b>C-1135-1136</b>					N				
<b>C-1137</b>	* 	* 			N	N	N	* 	
<b>C-1138</b>	* 	* 		N	N			* 	N

Chapter 5: Design and validation of a locus specific microarray

Boundary Oligo	Samples								
	NCTC 11168 as 'Control' in Cy3 channel								
	11351(1A)	480(1A)	81-176(1B)	81116(2A)	O:3(2A)	RM1221(3A)	O73(3A)	59653(1A)	34806(1B/3A)
<b>C-1138-1139</b>	* 	* 			* 	* 	* 		N
<b>C-1140</b>	* 	* 			N	* 	* 		N
<b>C-1140-1141</b>	* 	* 			N	* 	* 		N
<b>C-1141-1142</b>	* 	* 			N	* 	* 		N
<b>C-1142-1143</b>	* 	* 		N	N	* 	* 	N	N
<b>C-1143-1144</b>	* 	* 			N	* 	* 		
<b>C-1144</b>	* 	* 	* 	* 	* 	* 	* 		
<b>C-1145</b>	* 	* 		* 	N	* 	* 		
<b>C-1145-waaV</b>	* 	* 		N		* 			
<b>C-waaV</b>					*N				
<b>C-waaF-gmhA</b>					N				
<b>C-gmhA-waaD</b>									
<b>C-waaD-waaE</b>									
<b>C-waaE-1152</b>					N			* 	

### 5.3.2 Discussion

The first problem that was noted from these preliminary hybridisations was the low signal produced compared to the whole genome array. This is important as it affects subsequent data analysis. The signal quality produced from the whole genome microarray argues against this low signal intensity being caused by poor quality DNA, although this is a limiting factor and due to the size of the oligonucleotides (70bp) compared to the size of PCR products (400-700 bp). The genomic DNA for these initial experiments was extracted using CTAB, sodium chloride and chloroform and this DNA proved to be a poor template when it was used in long range PCR (as discussed in Chapter 4).

The protocol used for labelling the DNA was a direct labelling approach. Direct incorporation of Cy dyes by some polymerases is sometimes inefficient due to steric hindrance by the large Cy groups. The different Cy dyes also incorporate to a different extent. Indirect labelling involves the incorporation of amino-allyl modified dNTPs into the DNA and the Cy dyes are bound to the target DNA by ester bonding. Both labelling methods have been used in the literature but the direct labelling protocol has been used successfully by B<sub>u</sub>G@S in a number of comparative genomic hybridisation experiments (Dorrell *et al.*, 2001; Rajakumar *et al.*, 2004; Champion *et al.*, 2005; Stabler RA *et al.*, 2005; Witney *et al.*, 2005).

Due to the length of the oligonucleotides compared to the PCR products, the labelling could be altered to produce shorter DNA strands. This also may be achieved by sonicating or digesting the DNA to fragment it. Alternative methods for improving the Cy dye incorporation and the signal intensity involve labelling long range PCR products (long range PCR was not reproducible at this time), labelling the DNA asymmetrically or labelling the DNA using gene-specific primers. Successful increases in signal intensity have previously been achieved using PCR amplification with degenerate primers (Sergeev *et al.*, 2004).

The hybridisation temperatures and salt concentrations used for these experiments had high stringency but potentially these conditions could have been altered to improve the hybridisation conditions.

As only group 1 oligonucleotides were assessed it was also difficult to rationalise whether the cross-hybridisation problems encountered were due to the high amount of sequence homology amongst strains of this group. Looking at the strains from other groups with these oligonucleotides also meant that only sequence divergence or absence could be inferred.

Some of the results from these proof of principle experiments were promising and some were disappointing. Problems with DNA quality and labelling could not be excluded as the cause of the low signal intensity.

There were solutions to all these potential pitfalls and as such a second generation locus – specific microarray was produced and factors such as changing hybridisation conditions and improving the DNA quality were explored.

#### **5.4 Results from the second generation locus-specific microarray, GMv2**

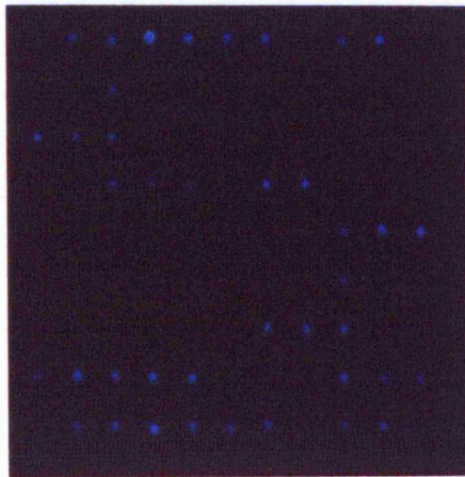
The GMv2 locus-specific microarray was again printed with the *Campylobacter jejuni* microarray. Control oligonucleotides, such as *C-galE*, were printed in eight different positions and the other oligos were printed twice. The locus-specific microarray comprised of both gene-specific and gene-boundary oligonucleotides.

##### **5.4.1. Increasing the signal intensity**

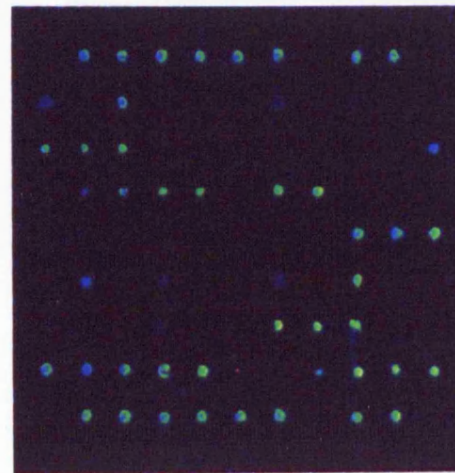
The first issue that needed resolving from the preliminary experiments involved improving the signal intensity. The first experiment performed involved using various hybridisation solutions; BpG@S, Pronto (Corning) and one containing of Denhardt's solution and formamide. These three hybridisation solutions were tested with the direct labelling of digested and

**Figure 5.1 Scanned images produced after testing the arrayed oligonucleotides with different hybridisation solutions.** The genomic DNA from NCTC 11168 was labelled with Cy5 and hybridised against the locus-specific microarray. Three different hybridisation solutions were tested; the B $\mu$ G@S hybridisation solution, the Corning Pronto™ kit and a hybridisation solution containing formamide and Denhardt's solution. The signal is increased with the Corning kit and the formamide containing hybridisation buffer compared to the standard B $\mu$ G@S solution.

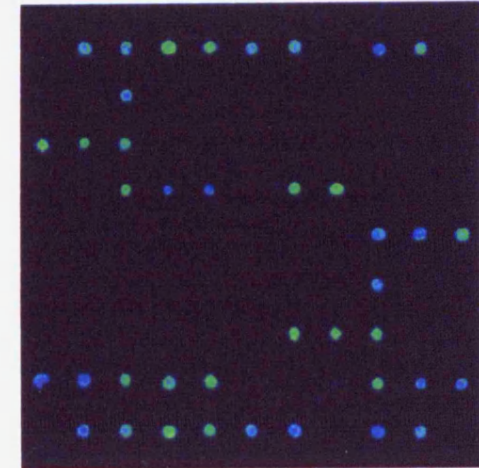




BμG@s Hyb conditions



Corning Pronto™



Formamide-based  
hybridisation  
solution

non-digested NCTC 11168 labelled with Cy5. The incorporation of Cy5 was also measured using a Nanodrop spectrophotometer (data not shown). Figure 5.1 shows the image produced after scanning the slide with labelled non-digested DNA. No difference in signal intensity or Cy dye incorporation was seen between the digested and undigested samples (data not shown). The addition of formamide in the commercial kit, Pronto, and the other hybridisation solution, both produced higher signal intensities than the normal BpG@S buffer. As the non-commercial formamide containing solution deviated less from the BpG@S standard hybridisation conditions, this protocol was used.

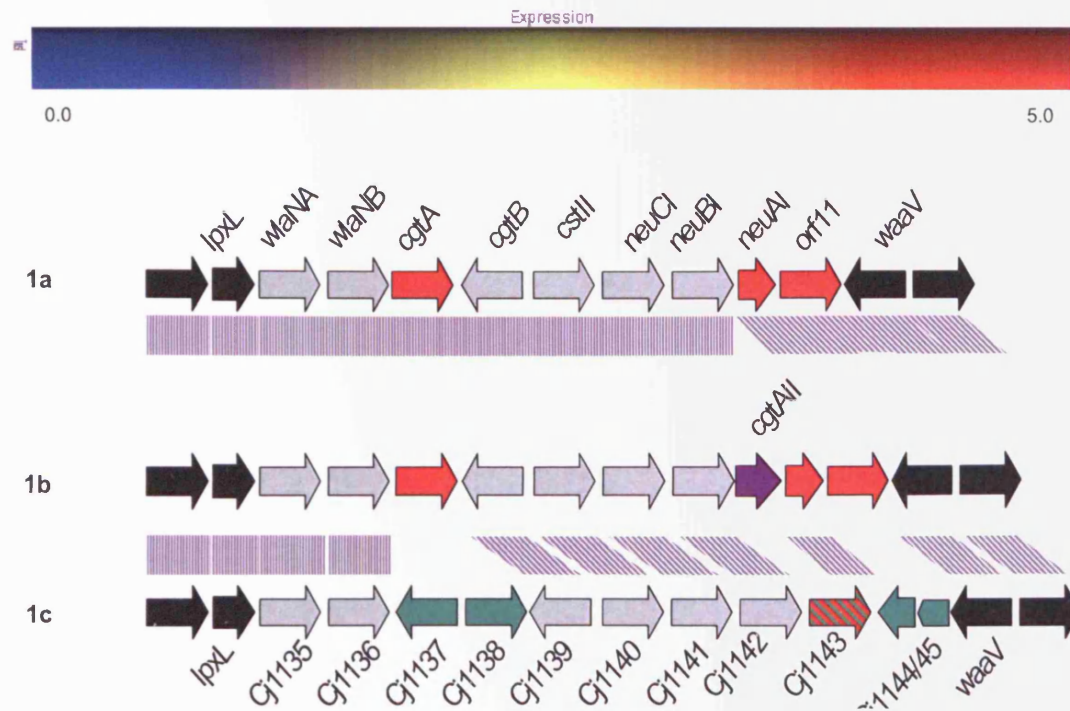
#### **5.4.2 Proof of principle and validation experiments**

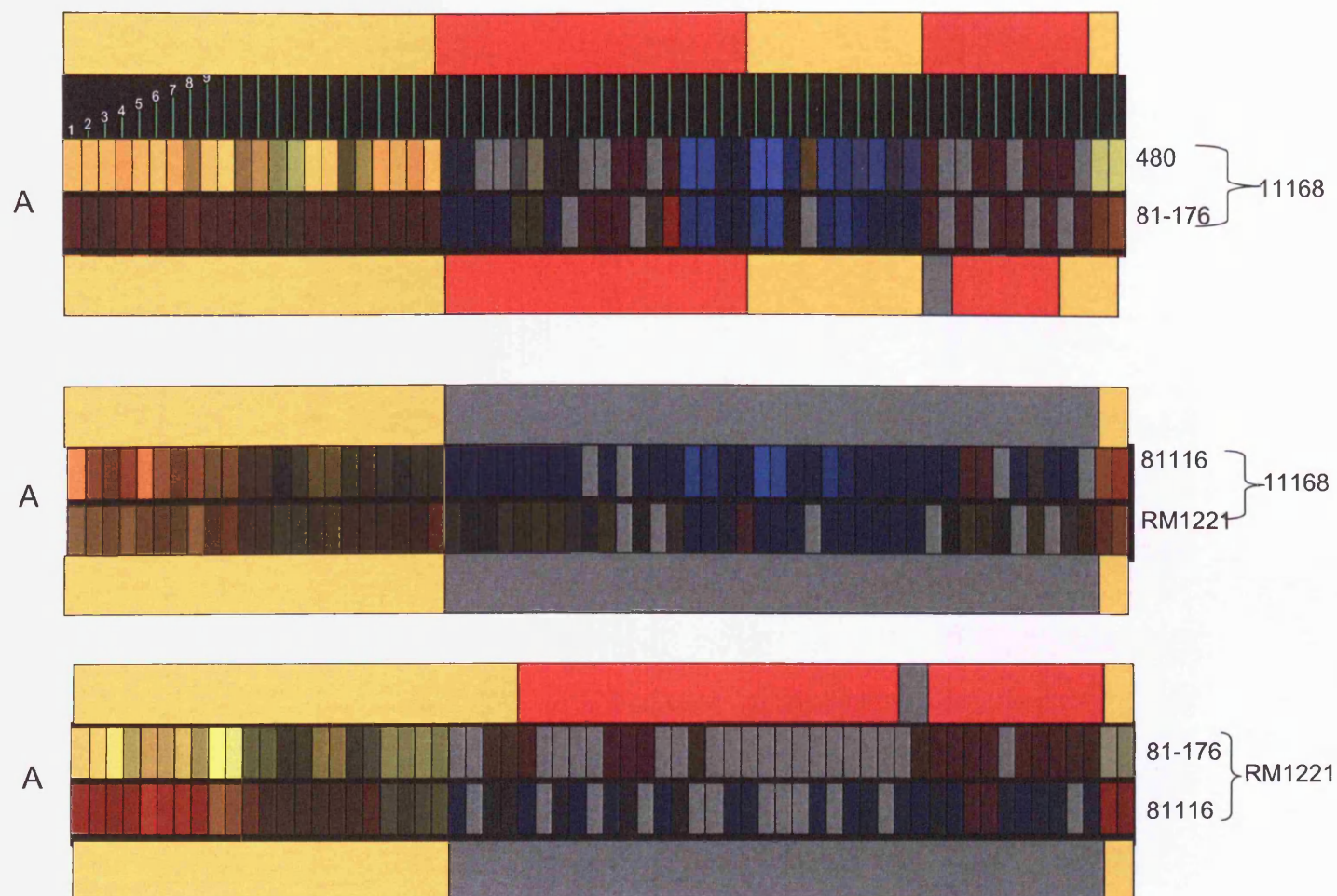
After testing different protocols to discover if changing the hybridisation conditions improved the overall signal intensity from the locus specific microarray, validation experiments were performed using the control strains for each group. These hybridisations were performed as two colour experiments with the genome strains NCTC 11168 or RM1221 as the control sample. The oligonucleotides for each group were used to create an ordered list or 'virtual genome'. The oligonucleotides and genes present in each sub-group virtual genome are listed in Appendix 2. Oligonucleotides representing *lpxL* and *waaV* are included in all virtual genomes. All the figures presented display log ratios and are filtered by reporter elements that are flagged as marginal and present.

##### **Group 1 oligonucleotides**

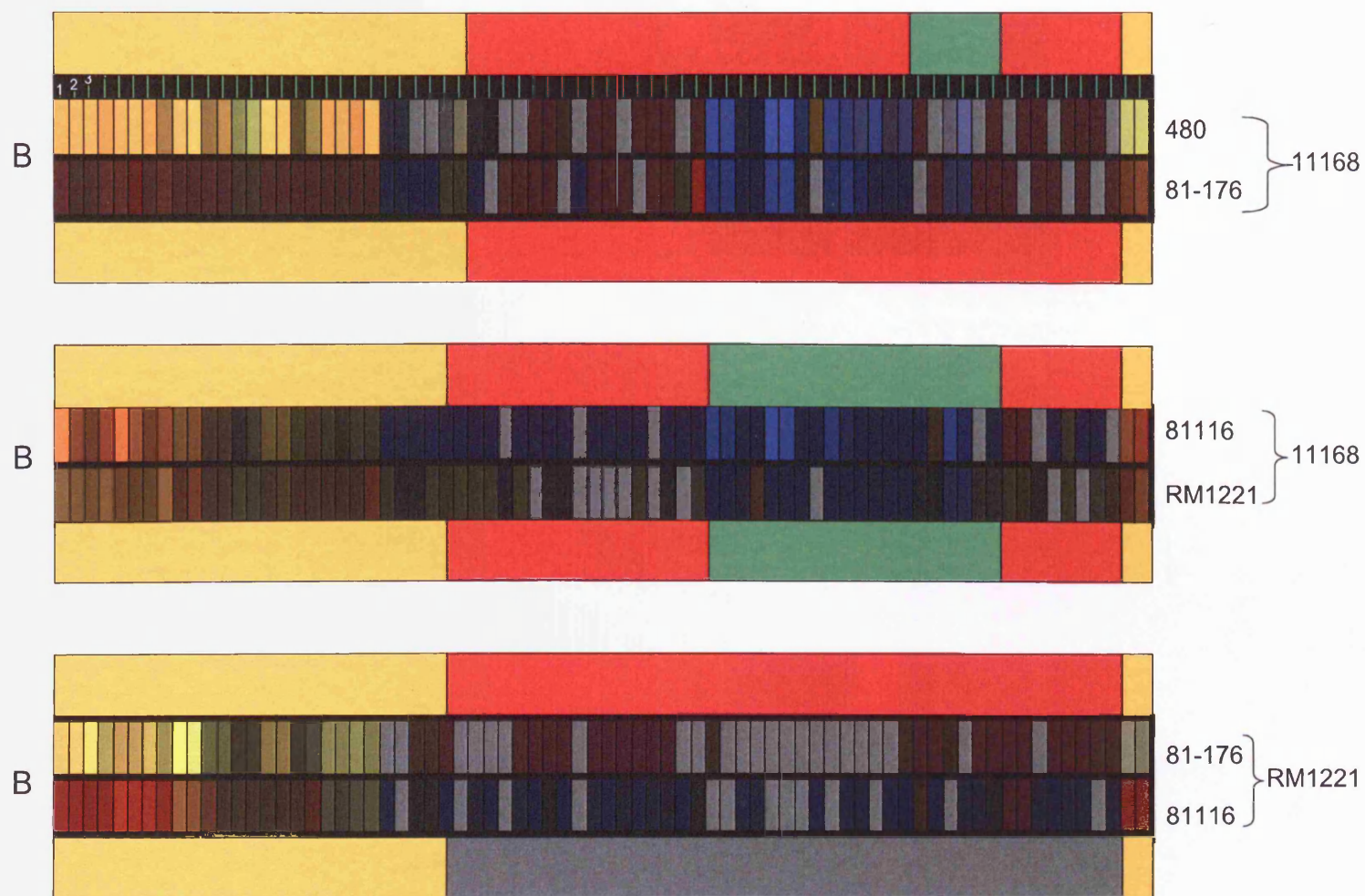
Generally positive signal was seen in the group-specific oligonucleotides with the relevant strains (see figure 5.2 a, b and c). Disappointingly the gene-boundary specific primers for sub-group 1A did not perform as expected and signal was not seen in both duplicate reporter

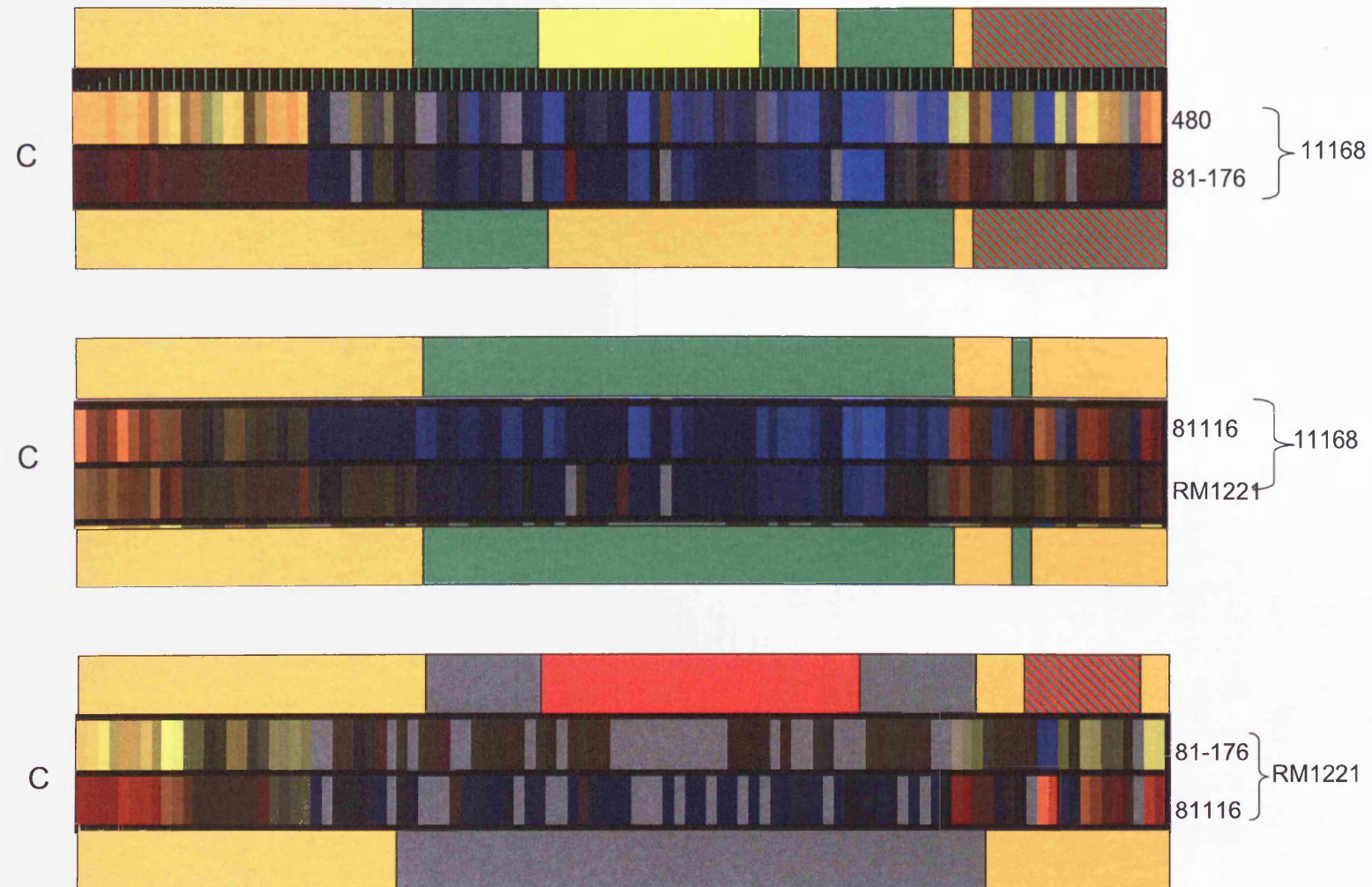
**Figure 5.2 a, b and c Hybridisation of the control strains to group 1 specific oligonucleotides.** Each Genespring image is separated into two strains, which are indicated on the right hand side of the image together with the strain hybridised in the control channel. The coloured boxes above or below the image indicate the expected results, where grey represents no signal, yellow indicates signal in both channels, red indicates signal in the test channel and green indicates signal in the control channel. The Genespring trust bar and a schematic representation of group 1 gene content are represented below. The Genespring trust bar provides an indication of signal intensity, where blue (0.0) indicates low or no signal and red (5.0) indicates high signal intensity.











oligonucleotides performed better in the competitive hybridisation of 81-176 against RM1221 than with 81-176 against NCTC 11168. The group 1c 70-mers confirmed previous deviations from the sequenced strain in the LOS region (Gilbert *et al.*, 2002; Karlyshev *et al.*, 2005). Differences between *waaF* and *gmhA* could only be examined using sub-group 1c oligonucleotides. The only indicator of a change in gene content between these genes was the lack of signal in the reporter elements representing *waaE*. There was signal present from the indicative reporter element, *waaF-gmhA* for all strains.

### **Group 2 oligonucleotides**

The group 2 70-mers were only tested with the control strain 81116 (2A) in competitive hybridisations with NCTC 11168 or RM1221 in the control channel (figure 5.3). These group-specific oligonucleotides performed well with positive signal in the test channel. No cross-hybridisation was seen between the different control strains and no signal was seen from the sub-group 2B specific 70-mers with the sub-group 2A control strain.

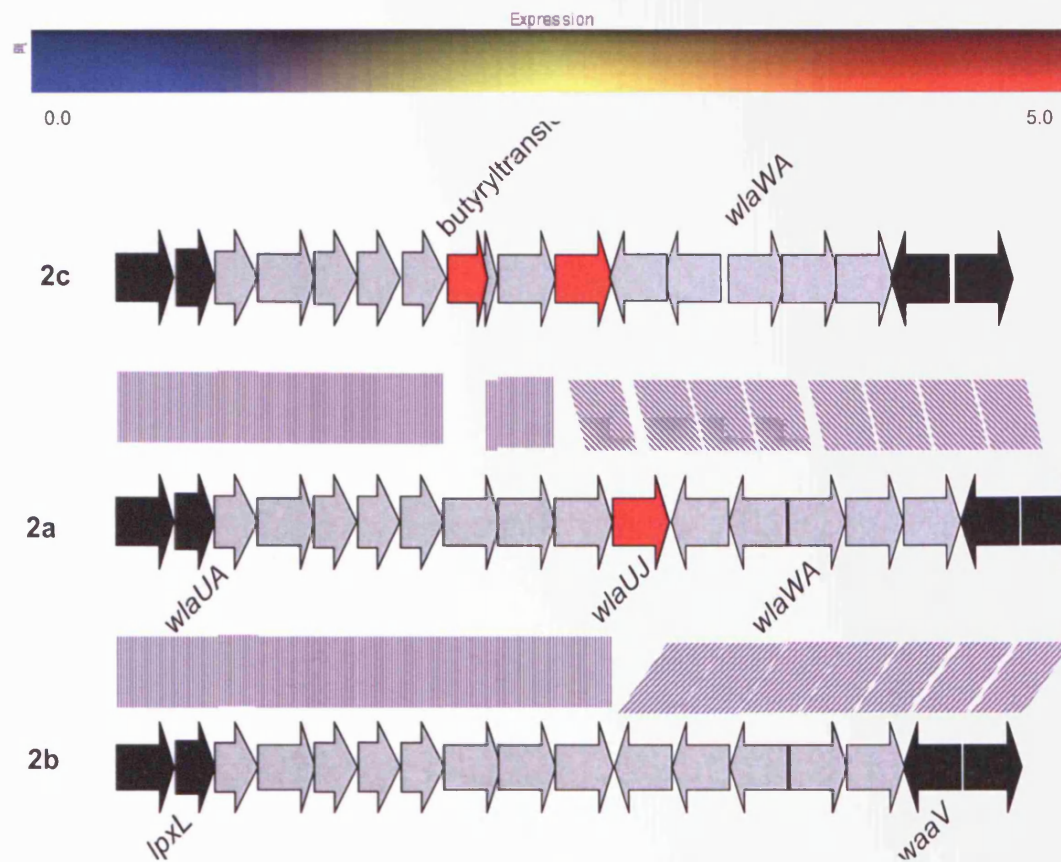
### **Group 3 and 4 oligonucleotides**

The only control strain available to assess the effectiveness of the group 3 and 4 oligonucleotides was the sub-group 3A strain, RM1221 (see figure 5.4). The boundary-specific and gene-specific oligonucleotides for group 3B successfully highlighted the absence of LIO87-orfG in RM1221 as the only difference between these strains.

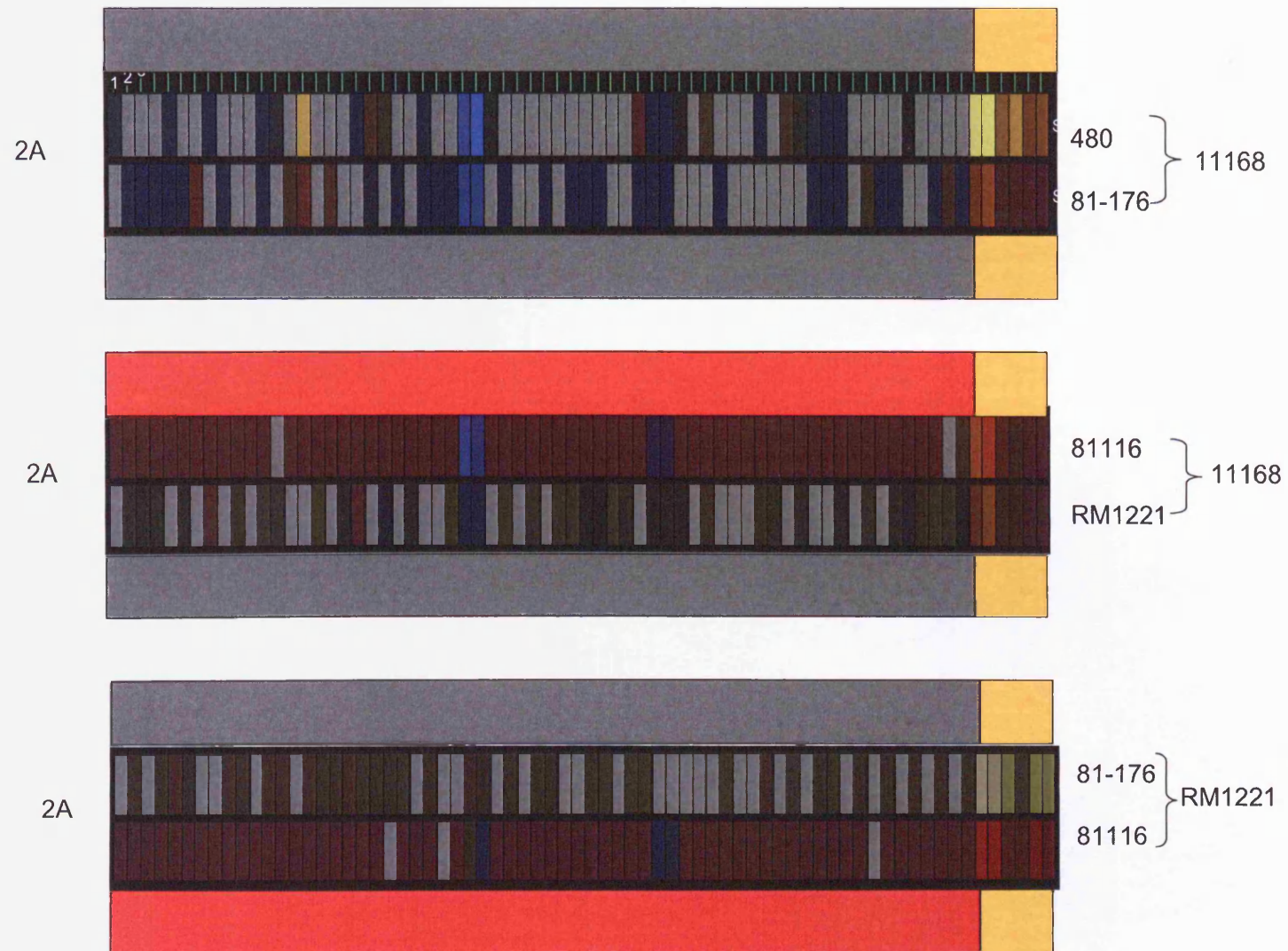
The group 4 specific oligonucleotides showed some cross-hybridisation in the 70-mers representing the boundary between *orf7-waaV* and *waaV*, although this is not unexpected as there is some conservation amongst these genes.

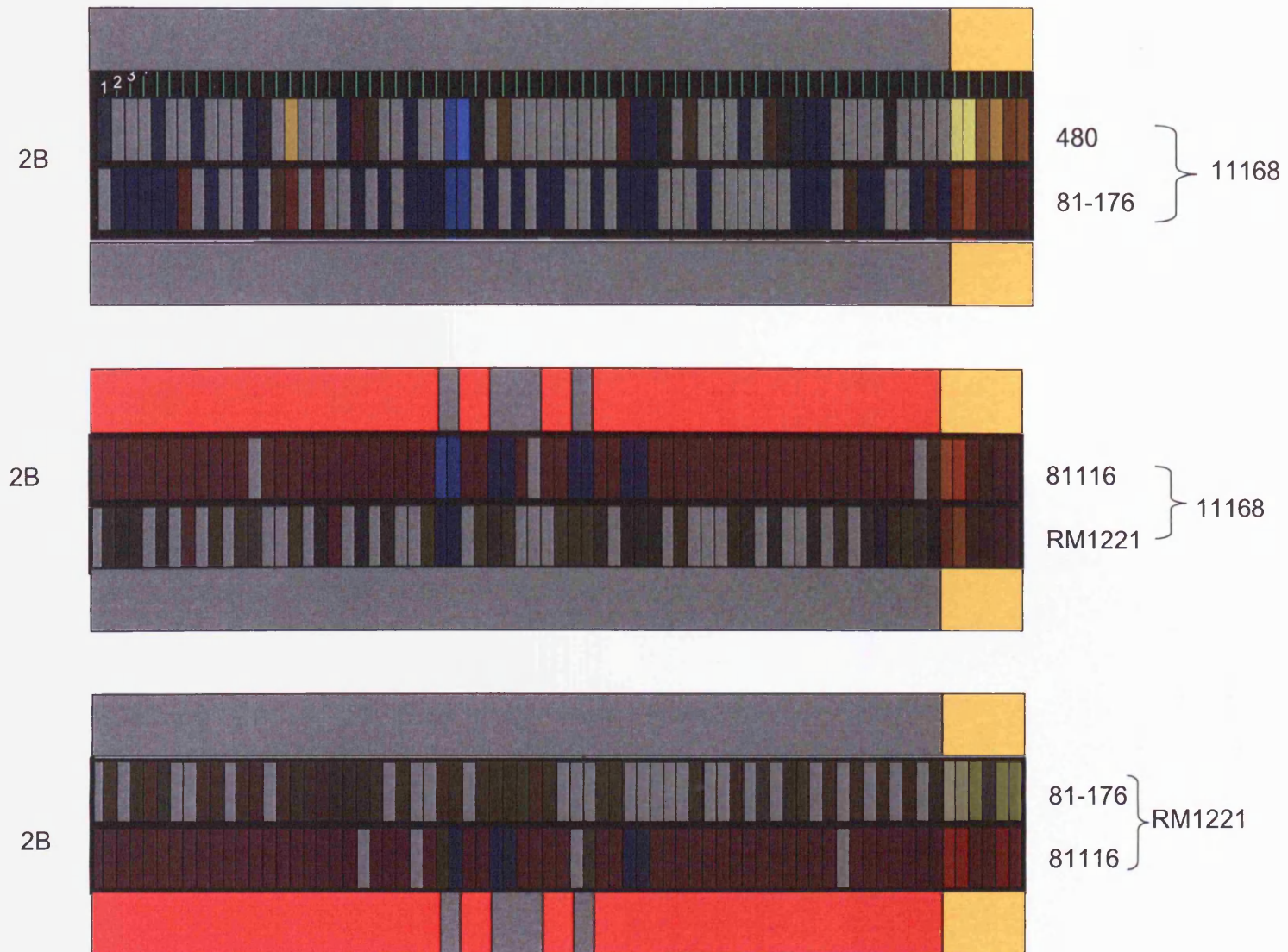


**Figure 5.3 Hybridisation of the control strains to group 2 specific oligonucleotides.** Each Genespring image is separated into two strains, which are indicated on the right hand side of the image together with the strain hybridised in the control channel. The coloured boxes above or below the image indicate the expected results, where grey represents no signal, yellow indicates signal in both channels, red indicates signal in the test channel and green indicates signal in the control channel. The Genespring trust bar and a schematic representation of group 2 gene content are represented below. The Genespring trust bar provides an indication of signal intensity, where blue (0.0) indicates low or no signal and red (5.0) indicates high signal intensity.

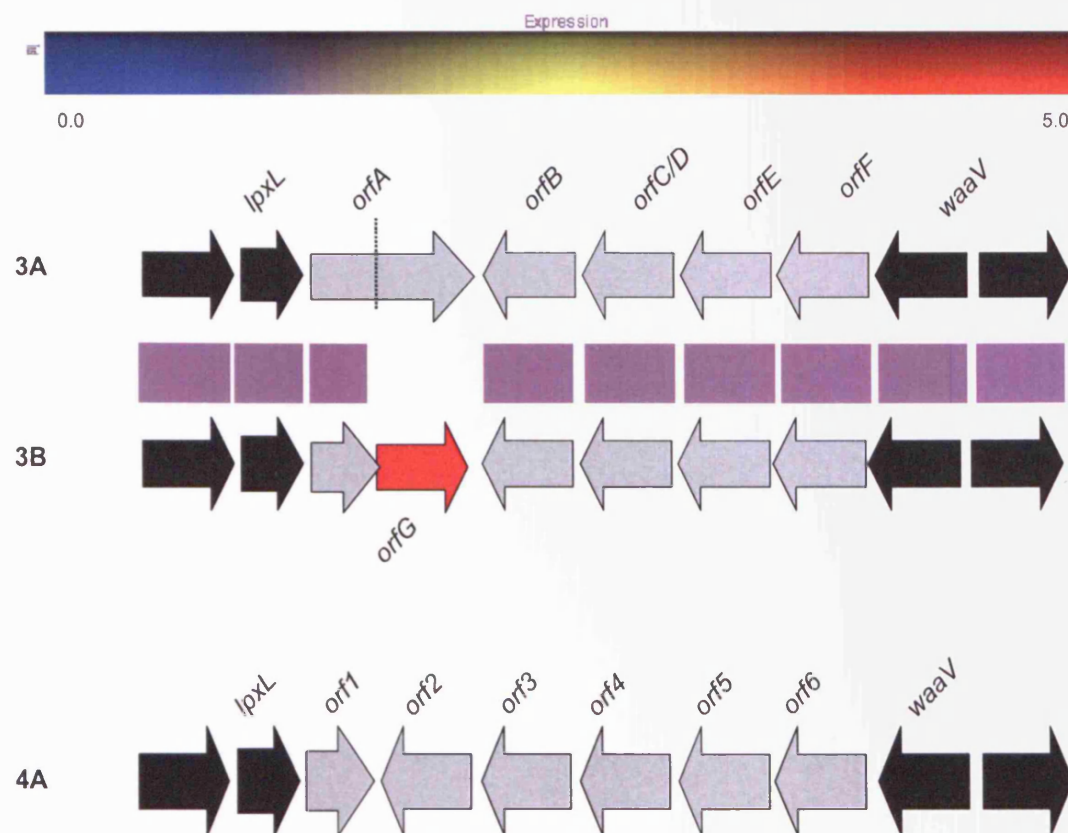




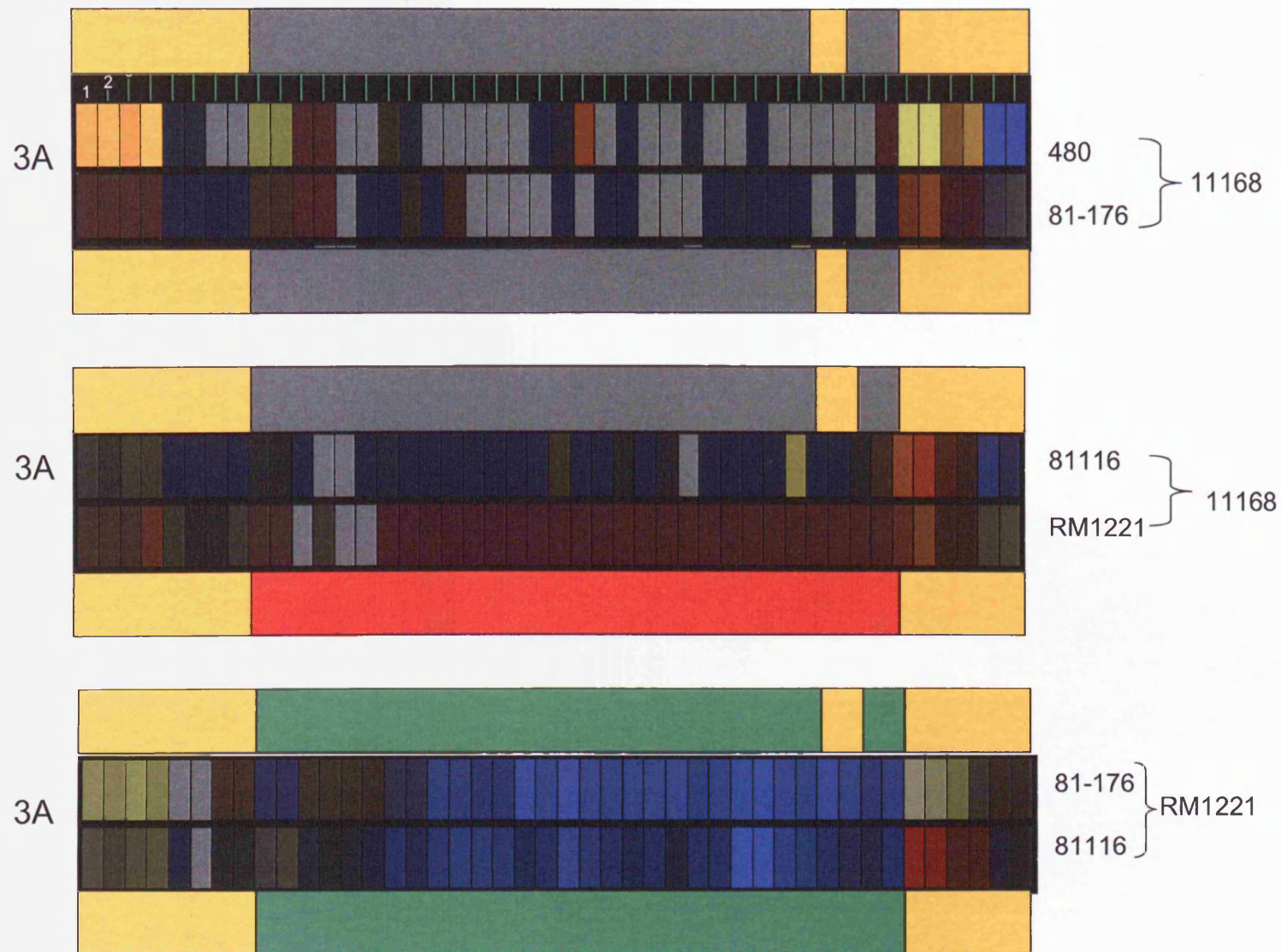


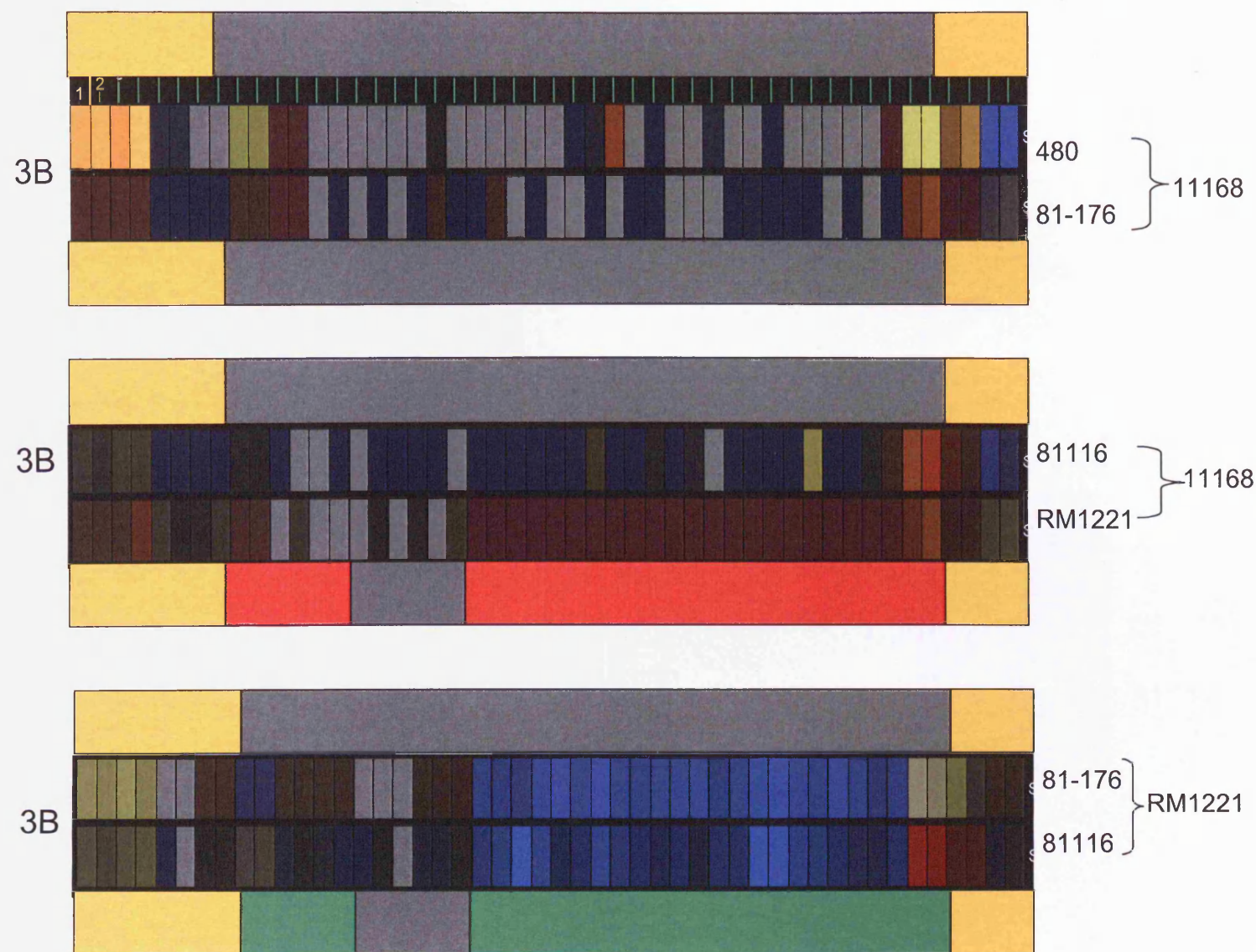


**Figure 5.4 Hybridisation of the control strains to group 3 and 4 specific oligonucleotides.** Each Genespring image for group 3 is separated into two strains, which are indicated on the right hand side of the image together with the strain hybridised in the control channel. The Genespring image generated for group 4 has not been separated as none of the control strains are expected to produce signal with these oligonucleotides. The coloured boxes above or below the image indicate the expected results, where grey represents no signal, yellow indicates signal in both channels, red indicates signal in the test channel and green indicates signal in the control channel. The Genespring trust bar and a schematic representation of group 3 and 4 gene content are represented below. The Genespring trust bar provides an indication of signal intensity, where blue (0.0) indicates low or no signal and red (5.0) indicates high signal intensity.

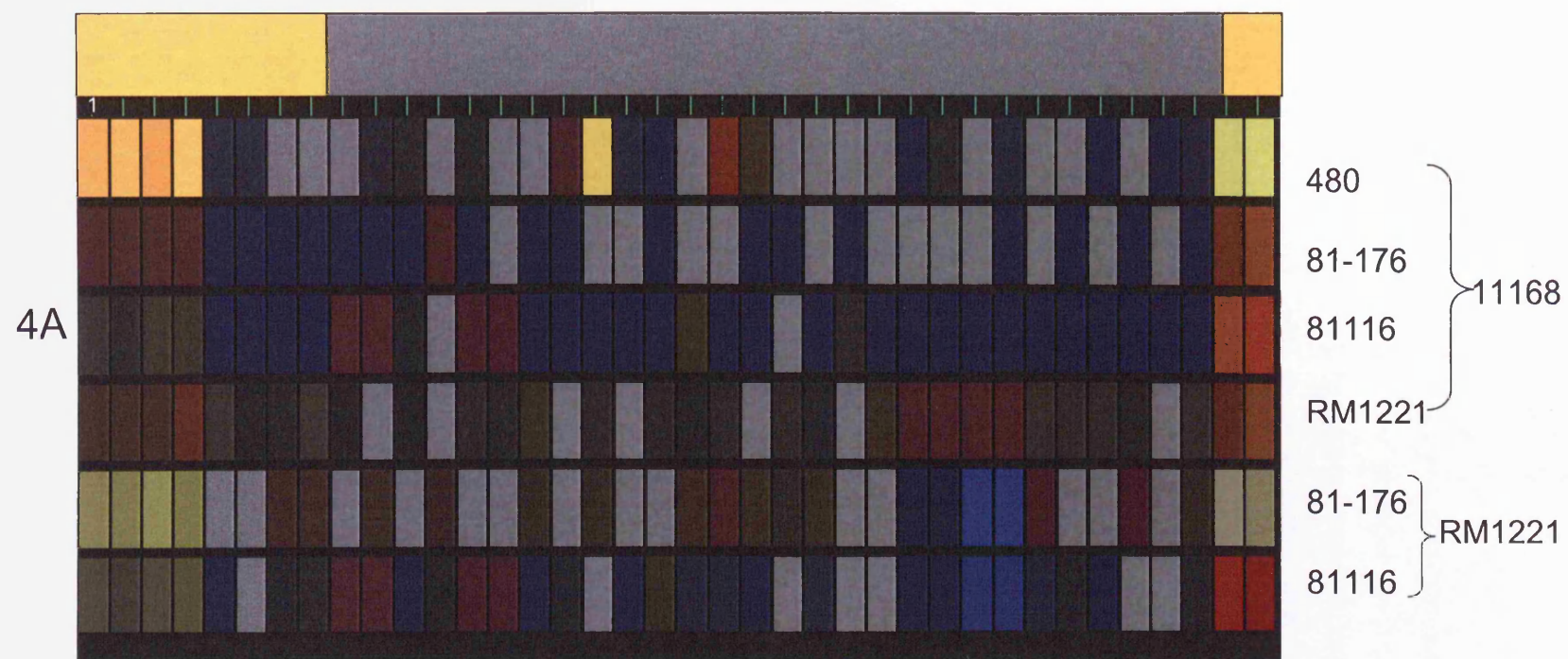












### 5.4.3 Microarray analysis of CSSS strains

To test the microarray, further hybridisations were carried out with the clinical strain collection. The chosen samples included unclassified strains and those that had been assigned to a group on the basis of the PCR screen. Strains 34086 (UC) and 51566 (UC) were excluded from the following analysis as the hybridisations performed produced data of poor quality despite repetition.

Hybridisations with these 20 strains were carried out with a universal control (SpotQC) that should produce signal with all the reporter elements as the use of a single common reference or pooled common reference can bias the hybridisation and subsequent data analysis (Gadgil *et al.*, 2005). The data analysis for these experiments was carried out using Imagene, Bluefuse and Genespring. The hybridisation results for NCTC 11168 with each set of 70-mers are displayed at the top of each image for comparison. Gene lists for each 'virtual genome' can be found in Appendix 2.

#### Sub-Group 1A

Figure 5.5 shows the hybridisation patterns produced with the clinical strains for the sub-group 1A 70-mers. Only one of the clinical strain, 59653, which had been classified as sub-group 1A by PCR, obviously stands out as a group 1A strain by this analysis. Those oligos that produce no signal with this strain belong to the 'control' oligos or sub-group 1C. The 70-mer, AB-*orf11-waaV*, that previously exhibited cross-hybridisation in the first round of experiments does not appear here.

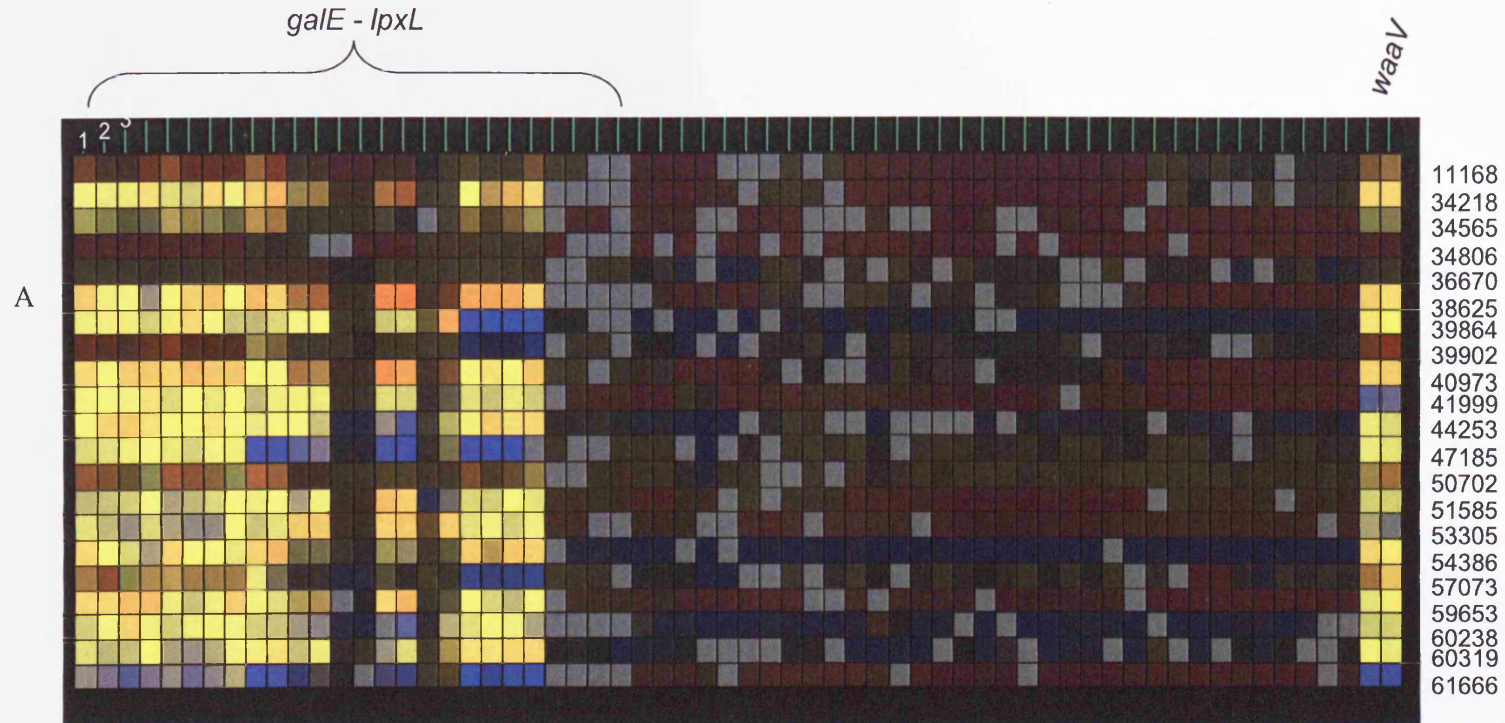
#### Sub-Group 1B

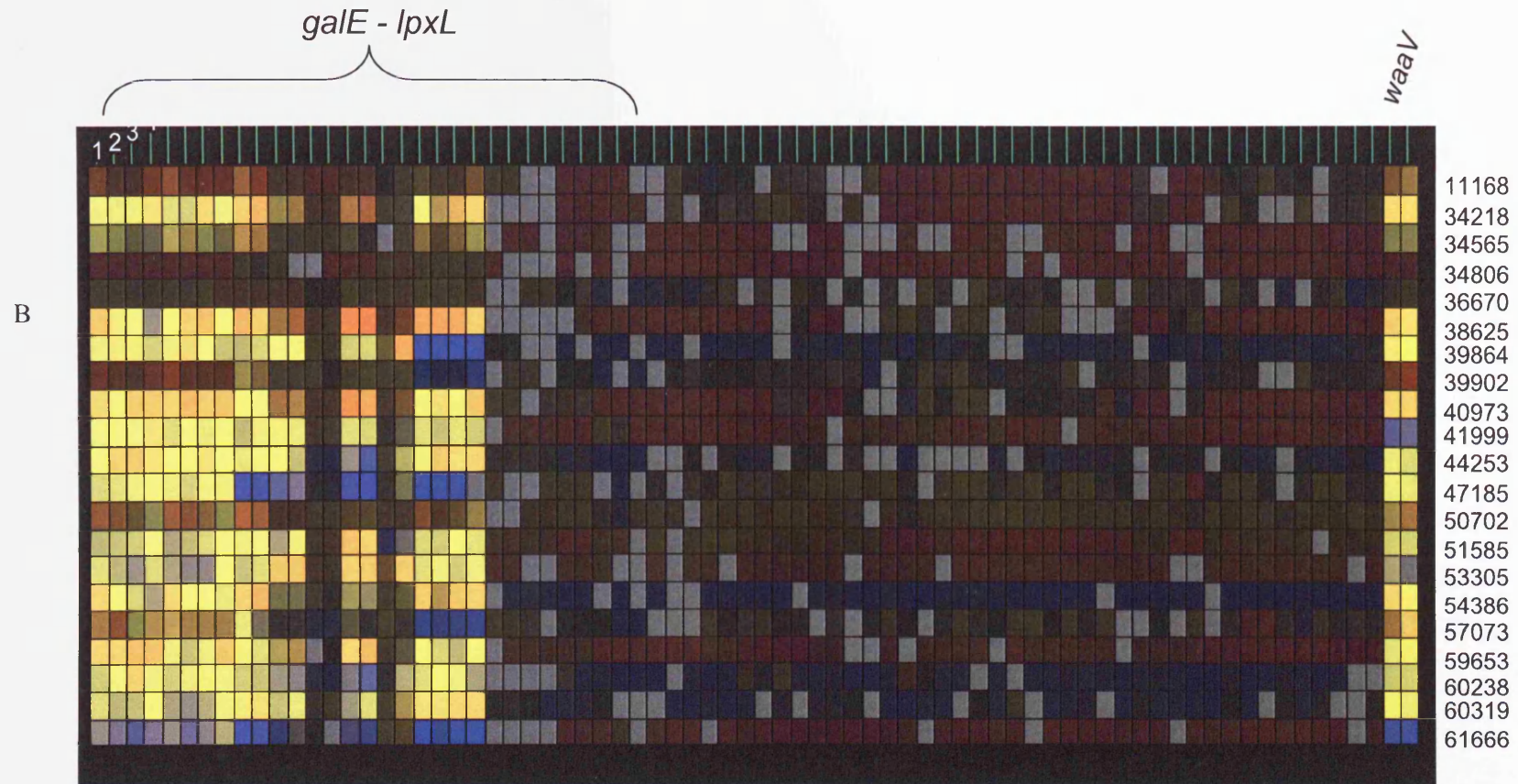
Strains 40973 and 41999 have both been previously classified as sub-group 1B following PCR. Microarray analysis does not show conclusively that these strains belong to this group conclusively (Figure 5.5 B). PCR analysis of strain 34806 classified this strain as having multiple class types.

**Figure 5.5 Hybridisation of the CSSS strains to Sub-Group 1a, 1b and 1c specific oligonucleotides.** The diagram was produced using the ordered list option in GeneSpring GX. The 'trust' bar is indicated below. The Genespring trust bar provides an indication of signal intensity, where blue (0.0) indicates low or no signal and red (5.0) indicates high signal intensity. Oligonucleotides between *galE* and *lpxL*, and *waaV* are expected to be present and therefore produce a yellow signal.



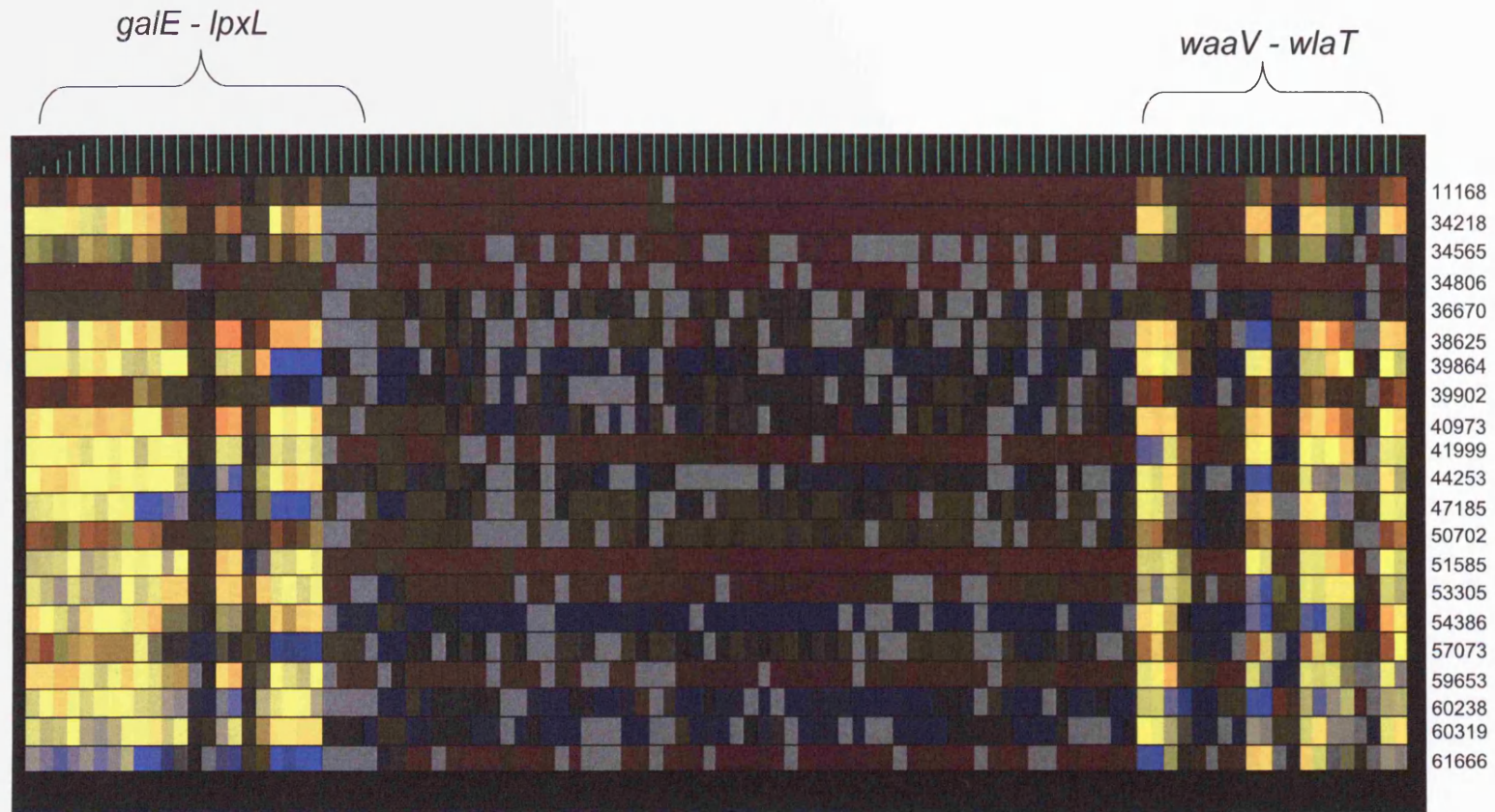








C



Examination of the microarray results clearly shows hybridisation occurring with all 70-mers except those corresponding to sub-group 1C genes.

#### Sub-Group 1C

Figure 5.5 C shows the typical results seen with *Campylobacter* whole genome microarrays. Strain 34218, which is unclassified by PCR, exhibits a similar hybridisation pattern to NCTC 11168. Another strain 51585, which was classified as a sub-group 1C/2A strain by PCR, also exhibits similar hybridisation patterns to NCTC 11168. Another strain 57073, similarly classified as sub-group 1C/2A by PCR shows poor hybridisation to the sub-group 1C oligos and the same can be said with the only sub-group 1C strain tested, 44253.

#### Sub-Groups 2A and 2B

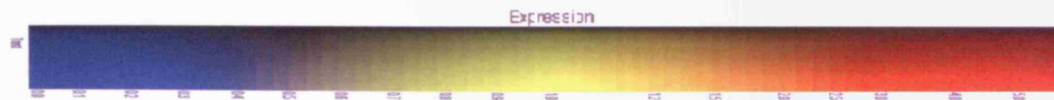
Hybridisation of strains 53305 and 57073 to group 2 specific primers shows these strains have considerable homology to these sequences. Strain 61666 appears to be a group 2 strain as well. Strains 47185, 39902 and 39864, appear to hybridise with the oligos designed to the extremities of the group 2 LOS cluster (Figure 5.6).

#### Sub-Groups 3A, 3B and 4A

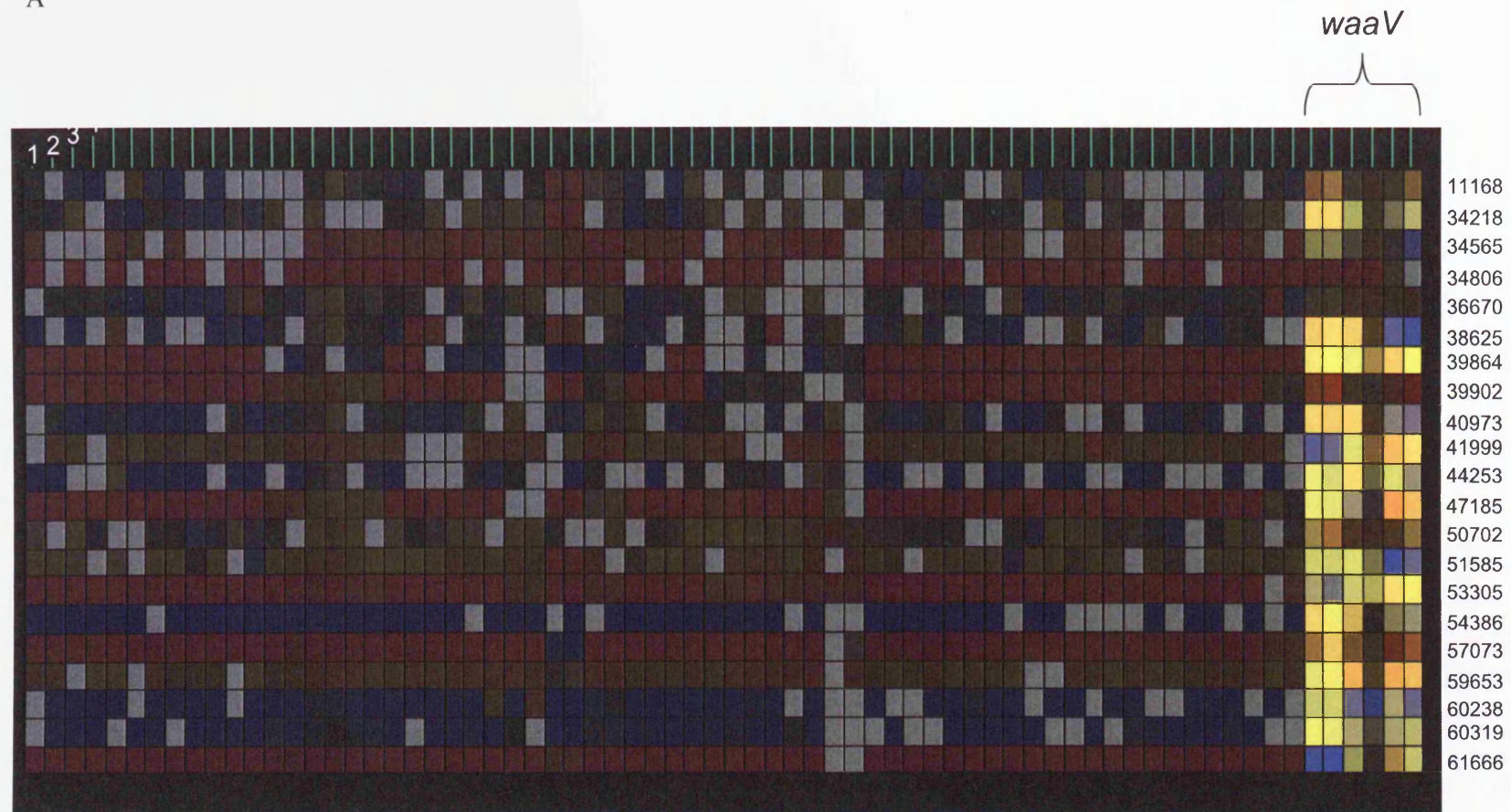
Microarray analysis of the strains 36670, 60319 and 60238 confirmed the PCR classification of these strains (Figure 5.7). The sub-group 3B specific primers, *LIO87-orfA-orfG*, *LIO87-orfG* and *LIO87-orfG-B*, performed well in hybridisation and could be used to discriminate between sub-groups. The strain 44253, which had been classified as sub-group 1C, hybridised well with the group 3A 70-mers.

Mixed results were produced with two strains that had been classified as belonging to multiple groups; 39864 (2B/3A) did not produce signal with either group 3 or 4 oligos and the testing these 70-mers with 34806 confirmed

**Figure 5.6 Hybridisation of the CSSS strains to Sub-Groups 2A and 2B specific oligonucleotides.** The diagram was produced using the ordered list option in GeneSpring GX. The 'trust' bar is indicated below. The Genespring trust bar provides an indication of signal intensity, where blue (0.0) indicates low or no signal and red (5.0) indicates high signal intensity.

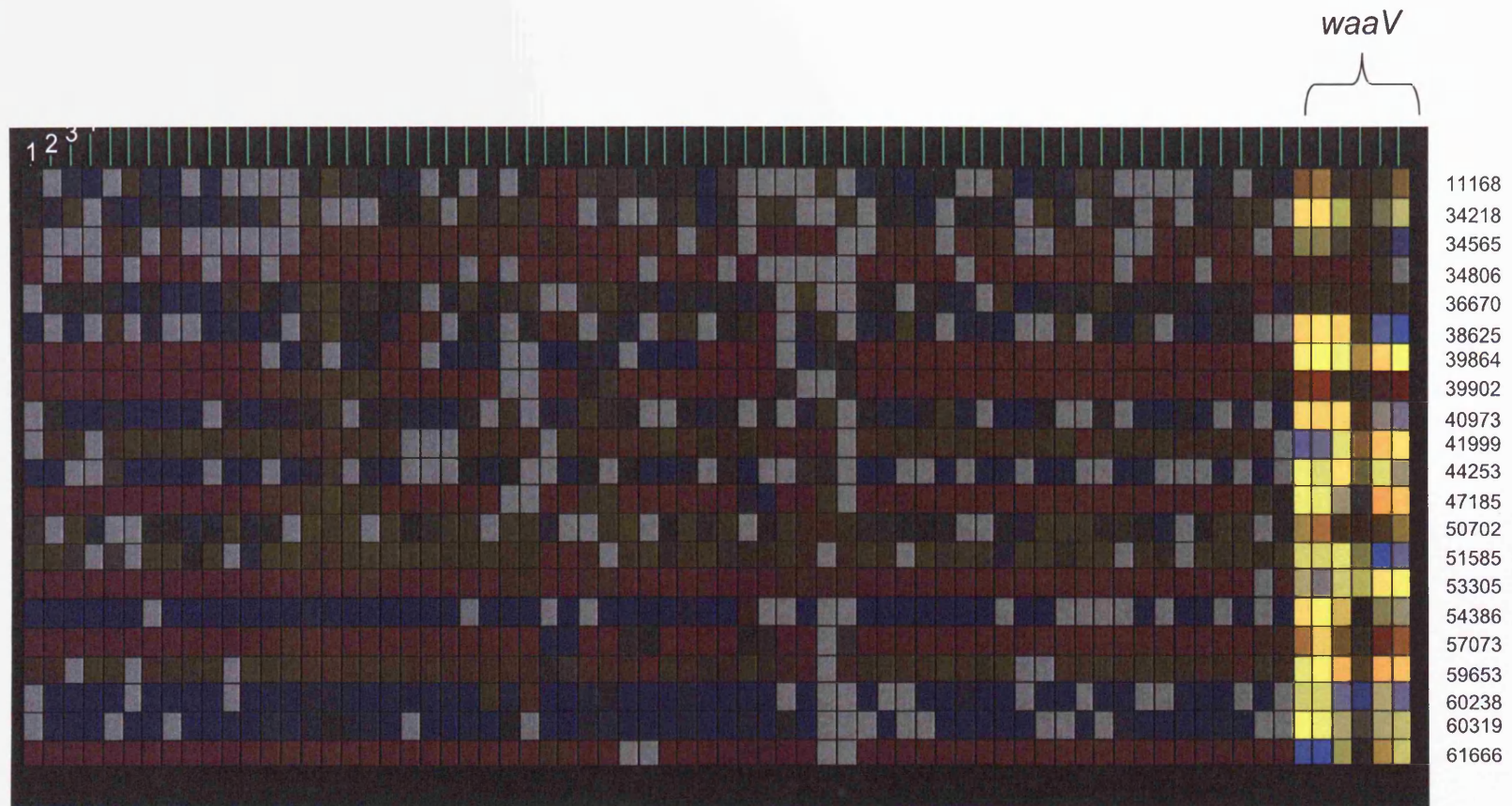


A





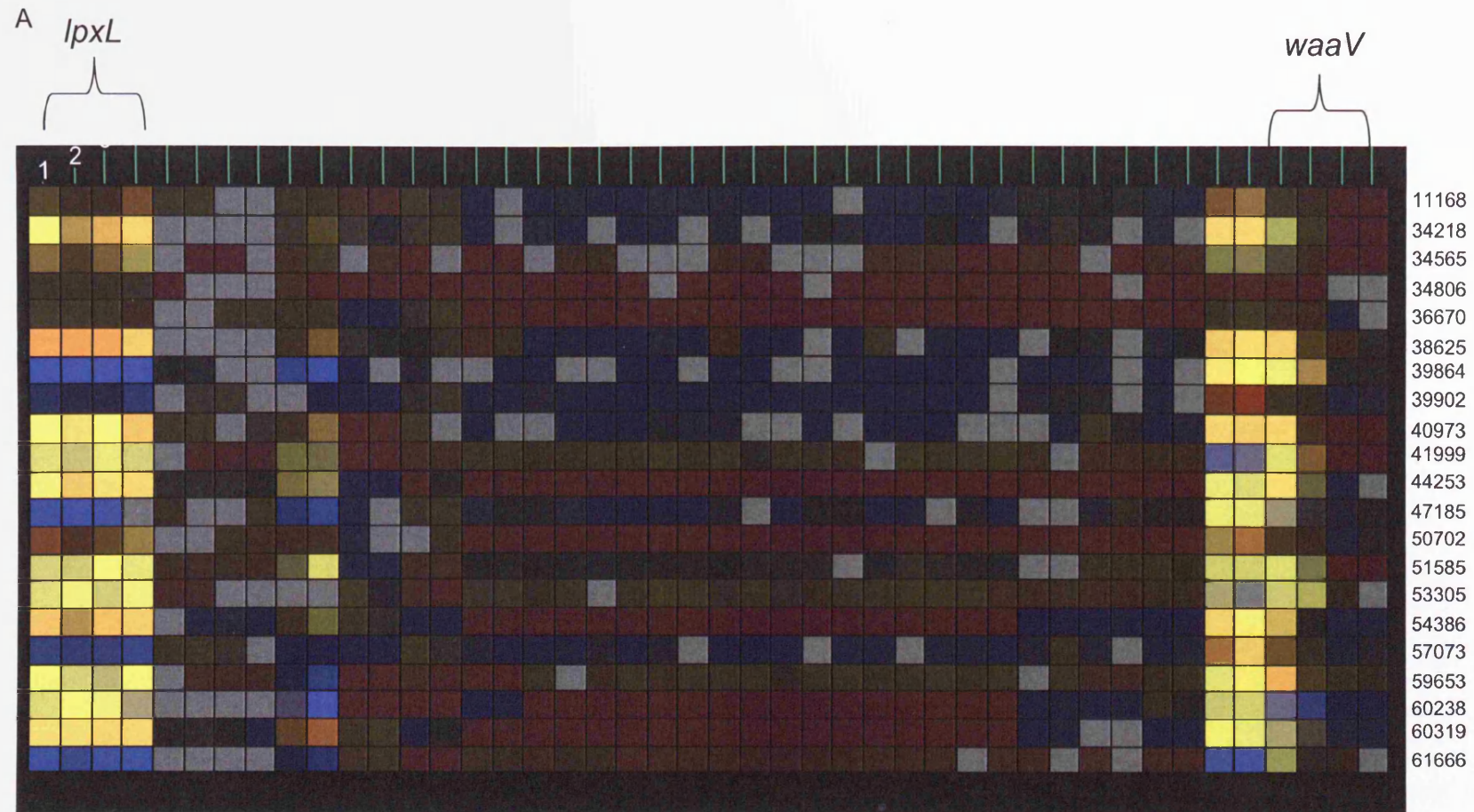
B

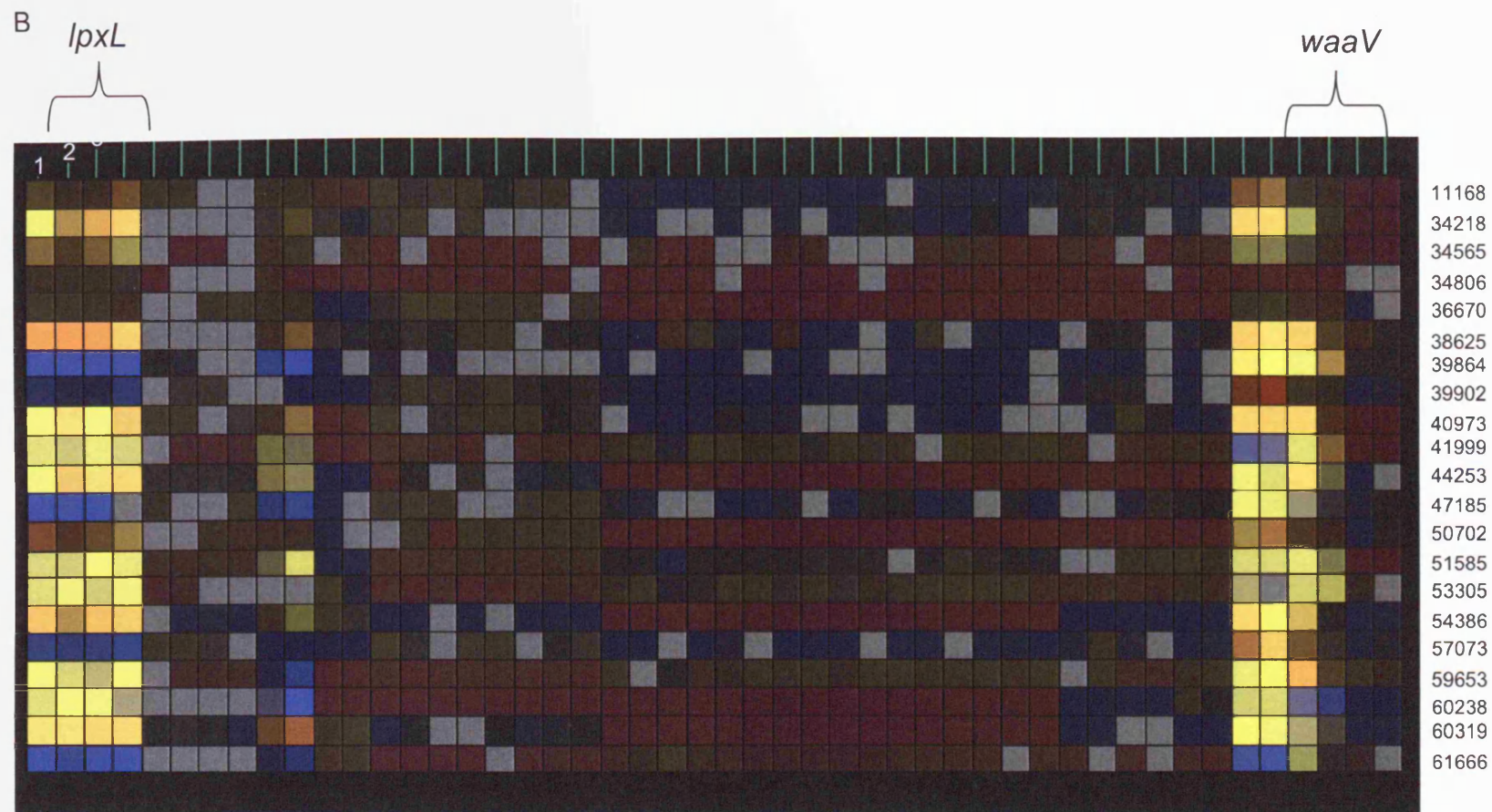


**Figure 5.7 Hybridisation of the CSSS strains to Sub-Groups 3A, 3B and 4A specific oligonucleotides.** The diagram was produced using the ordered list option in GeneSpring GX. The 'trust' bar is indicated below. The Genespring trust bar provides an indication of signal intensity, where blue (0.0) indicates low or no signal and red (5.0) indicates high signal intensity.

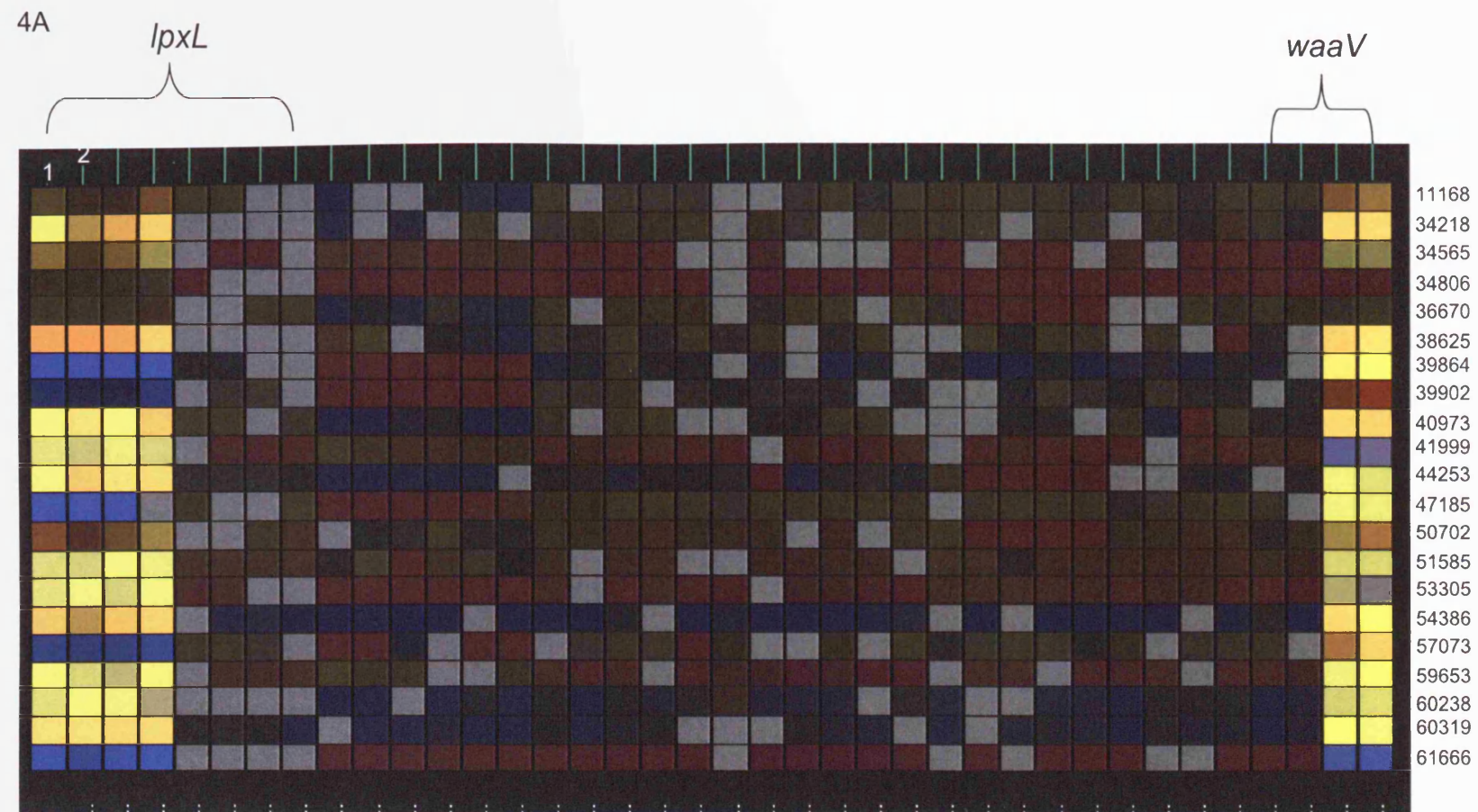












## 5.5 Discussion

### 5.5.1 Improving signal intensity

Analysis of genomic DNA by oligonucleotide microarrays presents some obstacles that may not be an issue with PCR-based microarrays. The highly complex nature of genomic DNA may result in several potential targets in the genome sharing identity with a probe and the reporter elements having high melting temperatures. Preliminary experiments with fragmented DNA and hybridisation solution supplemented with formamide showed that altering hybridisation conditions increased the signal intensity. Other methods of increasing the detection threshold of oligonucleotides include using strategies to amplify the genomic DNA. Most life sciences companies produce kits to enable the amplification of genomic DNA, but there are other methods available too. The use of  $\Phi$ 29 polymerase coupled with random hexamers, initiates replication at multiple sites and allows strand displacement resulting in the exponential amplification of the genomic DNA (Avarre *et al.*, 2006). This method has also been used in combination with Klenow-based random labelling, increasing the detection sensitivity compared to Klenow-based random labelling alone (Vora *et al.*, 2004). The introduction of spacer molecules between the oligonucleotides and the array surface can decrease the effect of steric hindrance on DNA hybridisation, increasing the hybridisation signal, but the influence of these molecules on specificity has not been investigated (Vora *et al.*, 2004). A more than 100-fold increase in signal sensitivity was seen using tyramide signal amplification (Vora *et al.*, 2004). This involves hybridisation with a biotinylated target and the addition of streptavidin-horse radish peroxidase catalyses the addition of fluorochrome-labelled tyramides to the slide surface. The addition of C<sub>0</sub>t-1 DNA reduces the hybridisation of repetitive elements and supplementation with tRNA which acts as a non-specific hybridisation blocker may both act to improve signal intensity and specificity (Van Ijperen and Saunders' 2004).

Changing the labelling method from a direct to an indirect method may affect hybridisation efficiency. The incorporation of the Cy labelled dNTPs is

inefficient due to steric hindrance, and from this it may also be inferred that the steric hindrance by these molecules may be a problem when binding with oligonucleotides. Direct labelling also causes early chain termination and incomplete elongation of the labelled single stranded product, leading to potential loss of signal and small products that are more likely to cross-hybridise (Tomioka *et al.*, 2005)

### 5.5.2 Validation of the second generation microarray

Analysis of the group 1 oligonucleotides shows that it is difficult to discriminate between sub-groups 1A and 1B; this is partly because some oligos did not perform very well. This is also to be expected as these groups only differ by the insertion of one gene between *neuC1* and *neuA1*. Several studies have been performed to examine oligonucleotide specificity and Kane *et al.* (2000) showed that for any 50-mer any non-target gene which displays more than 75% identity over the target length may show cross-hybridisation. *In silico* analysis by Schröder *et al.* (2001) showed that 70-mers are less specific than 50-mers as they are more likely to contain stretches of complementary sequences comprised of 15 or more bases. The importance of base mismatches also depends on their distance from the slide surface.

Later one-colour experiments showed that the sub-group 1C specific oligonucleotides performed well. The group 2 oligonucleotides performed well against 81116 which could easily be identified as belonging to this group. The sub-group 2A specific oligonucleotides as expected did not produce signal with 81116 and therefore the oligonucleotides may be used to discriminate between sub-groups.

### 5.5.3 Locus-specific Microarray analysis of clinical strains

Table 5.2 shows a summary of the microarray results for the CSSS strains. Those sub-groups indicated by +\* highlight limited binding for those oligonucleotides. Of the 12 strains tested, which were previously unclassified or classified into multiple groups, six strains could be assigned to group or sub-group after microarray analysis. The strains that can now be assigned to

a group require further discrimination by PCR or sequencing. A further two strains hybridised to more than one group or sub-group. The last four strains hybridised to some groups to a limited extent. Some strains appeared to hybridise to more than one set of oligonucleotides, indicating potential cross-hybridisation that was not highlighted by the *in silico* analysis. As genomic DNA was used, it may be that signal is produced because the oligonucleotides hybridise to another part of the genome.

The boundary oligonucleotides performed with mixed results. Where sequence in the test DNA was absent or divergent, the boundary oligonucleotide failed to give a signal. Further work to improve the stringency of the hybridisation and to improve specificity of the oligonucleotides may improve the ability to see a change in gene content using boundary oligonucleotides.

#### **5.5.4 Correlation of Locus-specific Microarray analysis with primer walking and sequencing analysis**

In comparison to the sequencing data, the microarray data correctly classified nine of fourteen strains into sub-groups. The microarray indicated a absence or divergent sequence in the LOS cluster of strains 54386 and 60319 from the known sequence but not due to the boundary sequences.

Strains 38625 and 50702, classified as group 2 strains by sequencing, were not classified as belonging to this group by microarray analysis, indicating that the sequence of strain RM3423 is divergent from the sequences used for the microarray design. The microarray also highlighted that the sequences for *wlaUD*, *wlaUF*, *wlaUG*, *wlaUI* and *wlaUJ* may be absent or divergent in strain 39864; *wlaUC*, *wlaUD*, *wlaUF*, *wlaUG*, *wlaUJ* and *wlaWC* may be absent or divergent in strain 47185. The microarray also indicated that genes *wlaUE-wlaVB* may be absent or divergent in strain 51585. Strains 53305, 57073 and 61666 for which no product could amplified by long range PCR could be classified as group 2 strains with sequence absence or divergence from the group 2 oligonucleotides designed for the genes noted above.

#### 5.5.5 Whole genome microarray analysis and the glycome

The whole genome microarray produces an enormous amount of data, much of which proves redundant for typing experiments. Data analysis becomes the time-limiting and labour intensive process in microarray experiments. The glycome is important in *Campylobacter* and therefore the microarray results for these loci were examined (see Figure 5.8). There are some inter-strain differences in the *lpx* genes, perhaps due to the presence of sequence variation as these genes are generally conserved. There is also variation in presence of *wlaJ*, the only variable gene in the protein glycosylation locus (Wood *et al.*, 1999), although sequence divergence may explain the changing signal for *wlaD* and *wlaC*. Although this type of analysis is useful, it provides clear indication of the inherent problems with microarrays as genes can only be highlighted as absent or divergent on the basis of the sequenced strain.

**Table 5.2 summarises the hybridisation of each CSSS strain to the group specific oligonucleotides**, where the strain is highlighted by +\* there appears to be limited hybridisation. UC indicates the strain was unclassified by PCR, NC indicates the strain was not cultured, NS indicates that sequencing was not p and NP indicates no product was produced with long range PCR amplification. CPS indicates the presence of capsule genes in the LOS gene cluster.



Strain	PCR result	Seq	Group 1			Group 2		Group 3		Gp 4 A
			A	B	C	A	B	A	B	
34218	UC	1C	+	+	+	-	-	-	-	-
34565	UC	2	+	+	-	+	+	-	-	-
34806	1B/3A	3A	+	+	+	+	+	+	+	+
36670	3A	NS	-	-	-	-	-	+	+	-
38625	UC	2	+	+	-	-	-	-	-	-
39864	2A/3A	2A	-	-	-	+	+	-	-	-
39902	NC	NS	-	-	-	+	-	-	-	-
40973	1B/2A	1B	+	+	-	-	-	-	-	-
41999	1B	NS	+	+	+	-	+	-	-	+
44253	1C	NS	-	-	-	-	-	+	+	-
47185	UC	2	-	-	-	+	+	-	-	-
50702	UC	2	-	-	-	-	+	+	+	-
51585	1C/2A	2	+	+	+	+	+	-	-	-
53305	UC	NP	+	+	-	+	+	-	-	-
54386	UC	3/CPS	-	-	-	-	-	+	+	-
57073	1C/2A	NP	-	-	-	+	-	-	-	-
59653	1A	NS	+	+	+	-	+	-	-	+
60238	3B	NS	-	-	-	-	-	+	+	-
60319	UC	3/CPS	-	-	-	-	-	+	+	-
61666	UC	NP	+	+	+	-	-	-	-	-

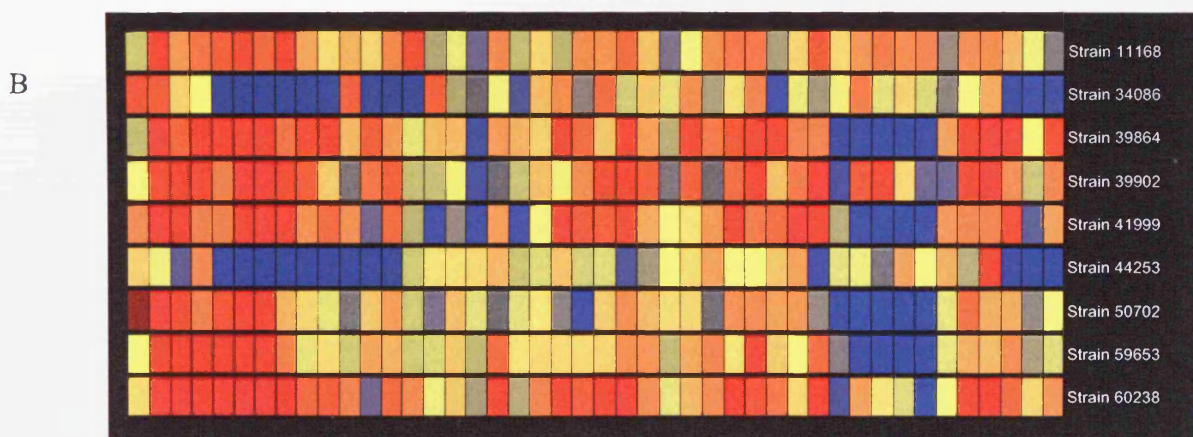
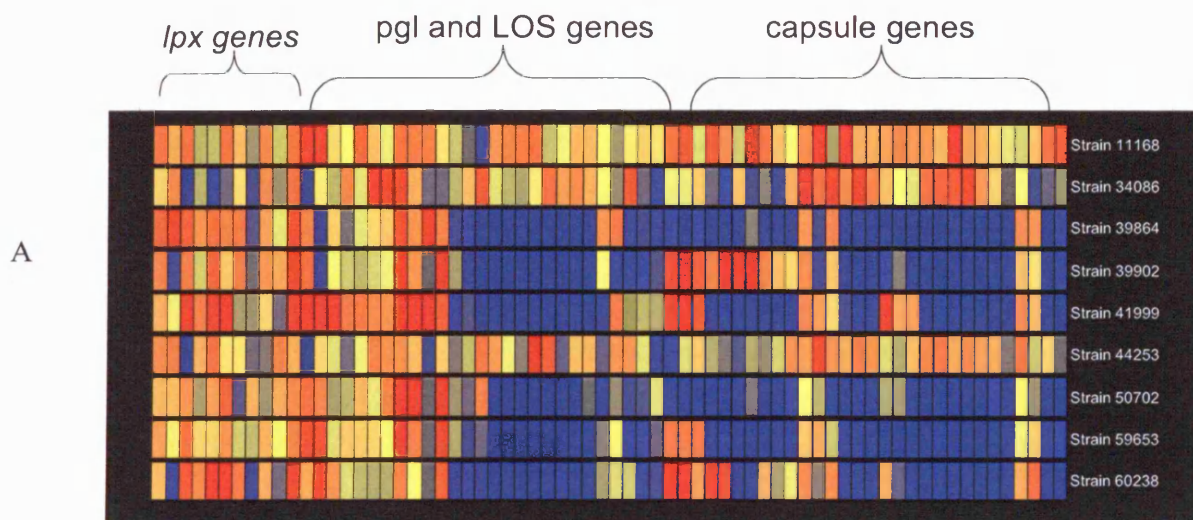
**Figure 5.8 indicating comparisons of A) the surface polysaccharide loci and B) surface appendages using the whole genome microarray and selected CSSS strains.**

A) includes *lpx* genes, *pgl* and LOS genes and capsule genes

B) includes flagellar structural and glycosylation genes.

The Genespring trust bar provides an indication of signal intensity, where blue (0.0) indicates low or no signal and red (5.0) indicates high signal intensity.





## 5.6 Conclusion

Various locus-specific microarrays have been produced to type genetic variants. These include a microarray for typing genetic variation in the LEE (locus of enterocyte effacement), which in combination with PCR successfully enabled the screening of many strains (Garrido *et al.*, 2006). Unlike these locus-specific microarrays, this LOS-specific microarray aimed to determine genetic change and gene context.

This locus-specific microarray successfully discriminated the LOS core biosynthesis cluster to a group level. It also provided supplementary information to the PCR screen of these strains, confirming that only the implication of group type can be inferred from the PCR analysis. The boundary sequences did not however provide a clear picture of changes in gene content.

The discriminatory power of this technique may be improved by increasing signal intensity as discussed previously and by using a more focused amplification technique. Although these experiments included the *Campylobacter* whole genome microarray, much of the data from the whole genome microarray produced proves redundant as it does not address variation in the LOS cluster. Therefore long range PCR could be used to amplify the LOS region to provide more specific results. The environmental microbiology community has used PCR based amplification of microbial communities in association with microarray technology and have noted PCR biases, which for their needs may not provide the true diversity of the community they are examining (Palmer *et al.*, 2006). PCR amplification prior to hybridisation has been used successfully with locus-specific microarrays (Seregeev *et al.*, 2004; Garrido *et al.*, 2006). Seregeev *et al.* (2004) examined various PCR amplification techniques and found that using degenerate primers based on conserved regions provided the best method to limit false-positive or false-negative results. Garrido *et al.* (2006) found that long range PCR stopped the signal biases of normal PCR and multiplex PCR amplification as all the genes had the same copy number.

Microarray technology is continually progressing and currently tiling arrays are being developed. On a whole genome scale this technology, which involves the design of overlapping oligonucleotides, can be used for sequencing by

hybridisation and single nucleotide polymorphism (SNP) detection (Jackson *et al.*, 2006; Pandya *et al.*, 2006). Other applications of tiling arrays include the analysis of alternative splicing in higher eukaryotes, identification of RNA-binding motifs and the use of ChIP (chromatin immunoprecipitation) on chip to identify the DNA sequences by transcription factors (Grainger *et al.*, 2005; Mockler and Ecker 2005; Grainger *et al.*, 2005b).

Jackson *et al.* (2006) used *E. coli* O157:H7 tiling arrays to identify more than 150 SNPs as well as deletions and changes in gene copy number. This sequence variation was confirmed by sequencing and PCR, but the tiling array proved to be useful in determining allelic diversity. Two studies have been performed to test tiling arrays in a more focused application as an alternative approach to MLST with varying success. van Leeuwen *et al* (2003) examined the use of tiling arrays for MLST of *Staphylococcus aureus* finding the results to be reproducible and concordant with conventional MLST. When applied to *Neisseria meningitidis*, Swiderek *et al.* (Swiderek *et al.*, 2005) found that the high density of polymorphic sites and the changing GC content of this bacterium led to many false positive results.

Despite these issues, tiling arrays present an ideal technology to enable the identification of changes in gene content and gene context, and further work should entail the development of a glycome targeted tiling array. The only other stumbling block will be the amount of data produced, even for a locus-specific microarray.

## Chapter 6: Determining the minimal LOS core gene content for viability

### 6.1 Introduction

#### 6.1.1 LOS mutants in *Campylobacter jejuni*

The genes required for LOS core biosynthesis in *C. jejuni* NCTC 11168 are clustered together with coding sequences involved in the general protein glycosylation, heptose biosynthesis and an acetyltransferase gene, *lpxL* (see figure 6.1).

The functions of the majority of LOS biosynthesis genes in both the sequenced strain and other strains have been determined including; *gne* [*galE*] (Fry *et al.*, 2000; Bernatchez *et al.*, 2005), *waaC* (Kanipes *et al.*, 2006) *wlaN* and *cgtB* (Gilbert *et al.*, 2000; Linton *et al.*, 2000b), *cgtA* and *neuA/cgtA* (Gilbert *et al.*, 2000; Gilbert *et al.*, 2002; Guerry *et al.*, 2002), *cstII* (Gilbert *et al.*, 2000), *neuB* (Karlyshev *et al.*, 2000; Millar 2003; Sundaram *et al.*, 2004), and *neuC* (Guerry *et al.*, 2000; Guerry *et al.*, 2002), *lpxL* (Millar 2003) *waaF* (Oldfield *et al.*, 2002; Kanipes *et al.*, 2004), and *gmhB* (*wlaT*; Karlyshev *et al.*, 2005a). Figure 6.2 summarises the known and predicted functions of the enzymes encoded by these genes.

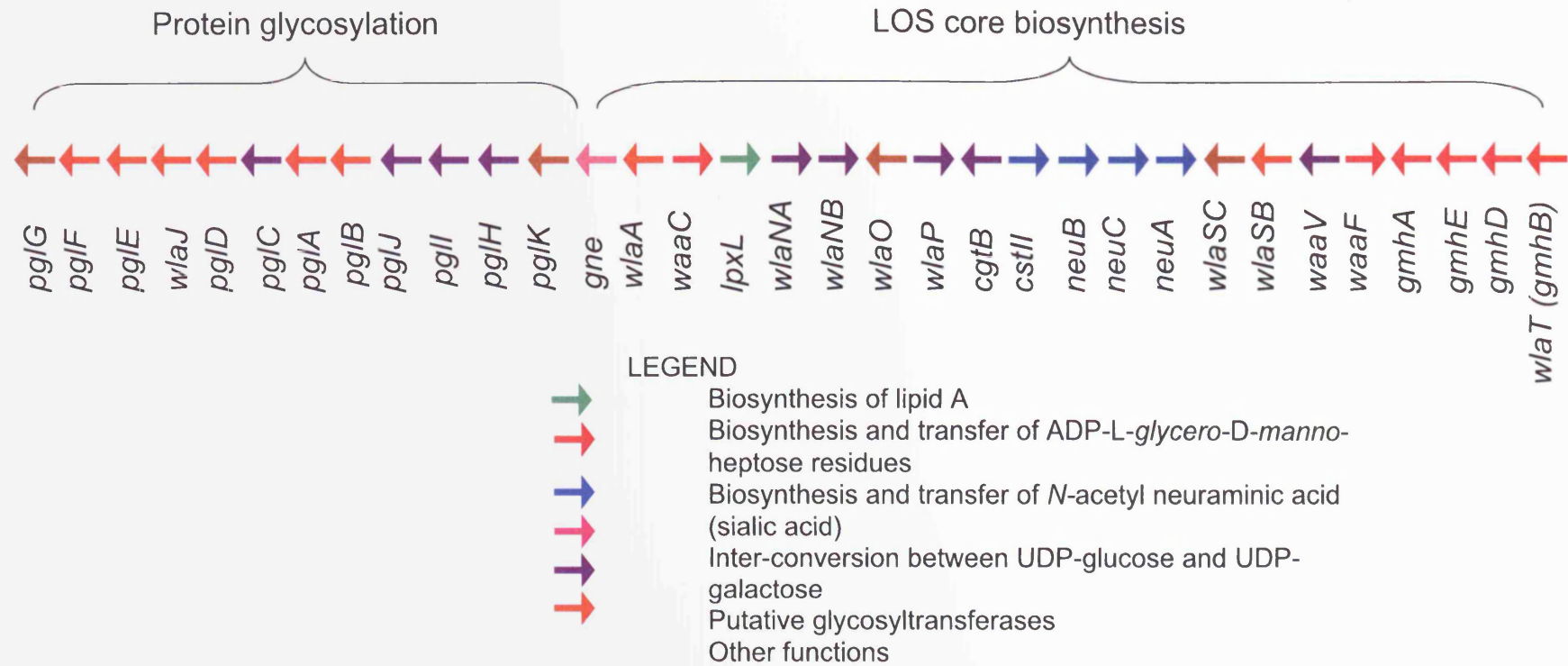
The enzyme Gne, or GalE, is a bifunctional UDP-GlcNAc/Glc 4-epimerase supplying UDP-GalNAc and UDP-Gal for LOS biosynthesis, and UDP-GalNAc to the capsule and general protein glycosylation pathways (Fry *et al.*, 2000); (Bernatchez *et al.*, 2005). Mutation of *gne* causes truncation of LOS and a reduction in adherence and invasion *in vitro*.

The gene product of *waaC*, a heptosyltransferase, catalyses the transfer of the first L-glycero-D-manno-heptose residue to the 3-deoxy-D-manno-octulosonic acid (Kdo)-lipid A moiety of LOS (Klena *et al.*, 1998). Initial studies showed the inability to mutate *waaC* in *C. jejuni* and the function of this gene was confirmed by complementation of a heptose deficient *Salmonella enterica* serovar Typhimurium with a plasmid encoded copy of

**Figure 6.1. Representational diagram of the general protein glycosylation and lipooligosaccharide core biosynthesis gene cluster in NCTC 11168.**

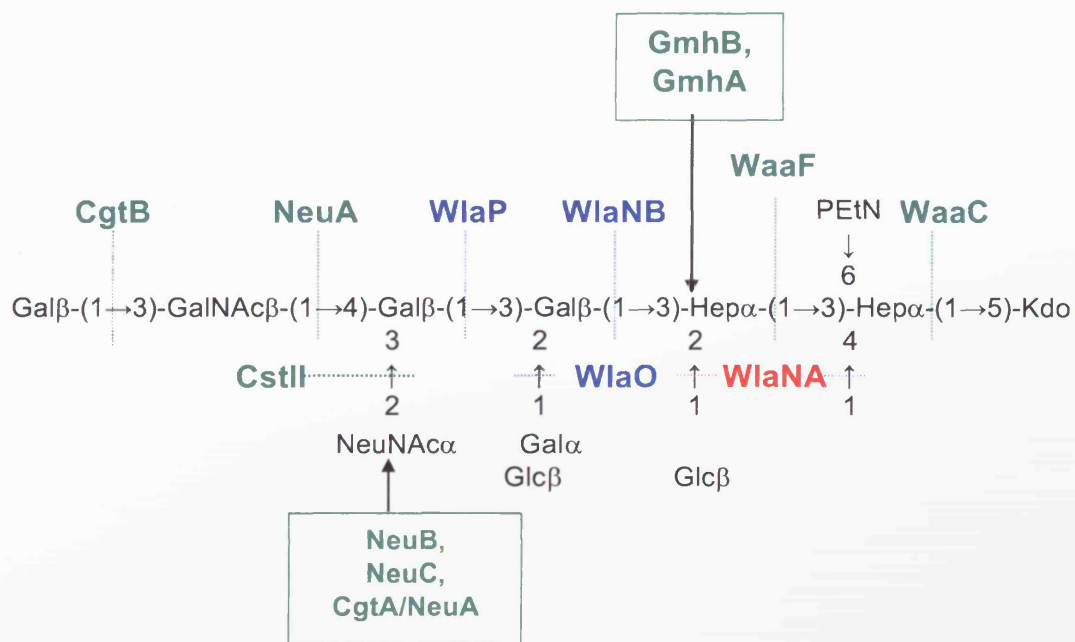
Brackets highlight genes involved in each system. Arrows show the direction of transcription and indicate putative or elucidated function. This diagram is not to scale.







**Figure 6.2: Structure and biosynthetic enzymes of the lipooligosaccharide core from *C. jejuni* NCTC 11168.** The core structure from *C. jejuni* NCTC 11168 is detailed. Enzymes for which there is experimental evidence of function are shown in green. Enzymes for which there is no experimental evidence but sequence similarity to other enzymes of known function are shown in blue. *WlaNA*, shown in red, is a two-domain transferase of which the first domain bears similarity to LgtF of *Neisseria*, and adds the glucose indicated with the blue line. The second domain is thought to add the glucose indicated with the red line, as this sugar is missing in several *C. jejuni* strains that have a premature translational stop in the C-terminus of *wlaNA*. Enzymes that are involved in synthesising a particular residue are boxed. Adapted from Millar, 2003



*waaC* from *C.coli*. Recently a *waaC* mutant was constructed in *C. jejuni* producing truncated LOS that was deficient in the sugar residues proximal to the Kdo-lipid A moiety (Kanipes *et al.*, 2006). Mutation in the *waaC* gene also led to loss of the 3-O-methyl group from the capsule polysaccharide. Complementation of *waaC* mutant restored both the LOS and capsule defect. Phenotypic testing of the *waaC* mutation showed hypersensitivity to antibiotics and detergents.

The enzyme LpxL is involved in the biosynthesis of the lipid A moiety and the *lpxL* mutant strain appeared to show a lack of fatty acid acylation as expected (Miller, 2003). Interestingly either a majority of truncated LOS core was produced with only one KDO and the second heptose apparent or small quantities of the wild type LOS core were observed. Preliminary analysis showed the mutant to be attenuated in adherence and invasion, but hypersensitivity to antibiotics and detergents could not be established. Due to the different structures observed, and the ability to produce a viable molecule, it is apparent that the LOS synthesis, assembly and transport pathway of *C. jejuni* differs from that of *E. coli* and *Salmonella*, as these organisms are unable to survive with mutations that affect the addition of both Kdo moieties as this leads to lethal under-acylation of lipid A (Millar, 2003).

The attempts of Phongsisay *et al.* (2007) to create a mutant in the *lpxL* gene of the *C. jejuni* strain HB93-13 by insertion and deletion mutagenesis strategy, similar to that used by Millar (2003), resulted in no double crossover events being seen. This inability to produce a *lpxL* mutant lead to their conclusion that unlike the previous findings of Millar (2003) the *lpxL* gene is essential for *C. jejuni* survival.

Open reading frame *Cj1139*, *wlaN* or *cgtB*, encodes a  $\beta$ -1,3-galactosyltransferase using a GalNAc residue as an acceptor for the terminal galactose (Linton *et al.*, 2000; Linton *et al.*, 2000b). This gene also contains an intragenic homopolymeric tract of eight or nine G residues leading to a premature end to translation and the core oligosaccharide switching between a GM<sub>1</sub>- and GM<sub>2</sub>- like epitope.

Complementation studies in a *N*-acetyl neuraminic acid deficient *E. coli* strain were also used to confirm the *N*-acetyl neuraminic acid synthase activity of NeuB1, which is involved in LOS sialylation (Linton *et al.*, 2000a).

Mutation of *neuB1* caused LOS to lose cholera toxin binding and therefore GM<sub>1</sub> ganglioside mimicry.

A two-domain protein is encoded by *neuA1*, with the N-terminal domain transferring N-acetylglucosamine to the LOS core and the C-terminal domain involved in the synthesis of sialic acid. In some *C. jejuni* strains, the two domains are encoded by separate genes, where *cgtA* has the function of the N-terminal domain (Gilbert *et al.*, 2000).

Oldfield *et al.* (2002) characterised the function of the second heptosyltransferase, *waaF* in *C. jejuni* NCTC 11828. The *waaF* mutant produces a truncated LOS, which misses the outer core. This deep-rough mutant confirmed the role of LOS in host-pathogen interaction and in adherence and colonisation (Oldfield *et al.*, 2002). Further studies in *C. jejuni* 81-176 also confirmed that a *waaF* mutant is attenuated in its ability to invade *in vitro* and that it was also hypersensitive to antibiotics and detergents (Kanipes *et al.*, 2004).

Although the actions of many of the gene products in the LOS gene cluster have been elucidated and functions of other genes have been predicted (Parkhill *et al.*, 2001, Fouts *et al.*, 2005), experimental confirmation of these functions would be ideal. By making a deep rough mutant with the minimal gene content for viability, the interactions of the products of these genes can also be examined. The construction of a large deletion mutant will remove the effects of polarity and minimise the build up of carbohydrate which can result from insertional mutagenesis of individual genes. The effect of different LOS structures can then be examined by in tissue culture systems, screened for differences in bacteriophage specificity and tested *in vivo*, to provide insight into the role of LOS in enteric disease

### 6.1.2 Predicting the minimal LOS gene content for viability

Studies using *E. coli* to create mutants lacking LOS have produced conditionally lethal mutants. Vorachek-Warren *et al.*, (2002) constructed a triple mutant in *E. coli* creating insertional mutations in the genes, *lpxM*, *lpxP* and *lpxL*, three late acetyltransferases. This triple mutant was unable to

produce penta-acylated lipid A molecules, and although the outer membrane localisation and content of lipid A appear almost normal, export of the tetra-acylated lipid A to the outer membrane is reduced. The mutant was unable to grow in nutrient broth at any temperature, unable to reach stationary phase growth and was hypersensitive to antibiotics. An *E. coli waaQ* mutant was recently used to look at the colonisation of the mouse large intestine but a truncated core oligosaccharide was observed due to polar effects potentially inactivating the whole *waa* operon (Møller *et al.*, 2003). Unlike a nonpolar *waaQ* mutant, not only was a truncated core observed but the polar mutant exhibited hypersensitivity to SDS and novobiocin. Further experiments looking at the polar effects of the kanamycin cassette used were unfortunately not performed.

*E. coli* and *S. enterica* require the addition of 2 Kdo residues to lipid A before acylation of the lipid A moiety, implying that the Kdo residues are required to prevent lethal under-acylation of lipid A. Another member of the family *Enterobacteriaceae*, *Yersinia pestis* is viable with LPS lacking Kdo (Tan and Darby 2005), indicating a tolerance to under-acylation or acylation of lipid A prior to Kdo addition.

The *lpxL* mutant produced in *C. jejuni* had a truncated core and the core only contained one Kdo residue (Millar 2003). Other bacteria such as *H.influenzae*, *V.cholerae* and *B.pertussis* are also viable with one Kdo residue (Gronow and Brade 2000). Two fractions were observed in the LOS preparations from this mutant where the major fraction was truncated as above and consisted of two heptose residues from one Kdo sugar and the addition of PEtN to one heptose. The minor fraction observed was able to bind cholera toxin, and although there was a slight increase in mobility, it appears that the inner core structure has not altered. This may be due to inefficient transfer of the Kdo residue or partial complementation of *lpxL* by an unknown enzyme, such as in *E.coli* (Vorachek-Warren *et al.*, 2002).

Two mutants have been created in *N. meningitidis*; one that only harbours lipid A (Tzeng *et al.*, 2002) and another *lpxA* mutant with no trace of LPS (Steeghs *et al.*, 1998). *LpxA* is an enzyme involved in the first step of the lipid A biosynthesis pathway, adding the O-linked 3-OH fatty acid to UDP-N-acetylglucosamine and previous attempts to mutate this gene in other bacteria

have resulted in a non-viable phenotype. In the *N.meningitidis* mutant no endotoxin was detected when an assay measuring endotoxic activity was performed. The *lpxA* mutant although viable, grew at a reduced rate and despite the outer membrane appearing normal, the mutant was reliant on the capsule polysaccharide for viability (Steeghs *et al.*, 2001).

The minimum LOS gene content required for viability in *C. jejuni* is predicted to reflect that the situation in *N. meningitidis* where only Lipid A is required for viability. Evidence for this lies in the ability to mutate the *lpxL* gene which is involved in Lipid A acylation, and produce viable cells.

### 6.1.3 Current mutagenesis strategies

Mutagenesis allows determination of gene function by phenotypic methods and other appropriate assays. Current methods involve insertional mutagenesis with an antibiotic resistance cassette. This can often cause polar effects providing misleading phenotypic information and this is especially important in large transcriptional units. Currently only three antibiotic resistance cassettes are in use in *C. jejuni* and confer kanamycin, chloramphenicol or tetracycline resistance. All of these genes have been cloned from *C. coli* plasmids, as antibiotic resistance genes used in *E.coli* are not functional in *C. jejuni* (van Vliet *et al.*, 1998).

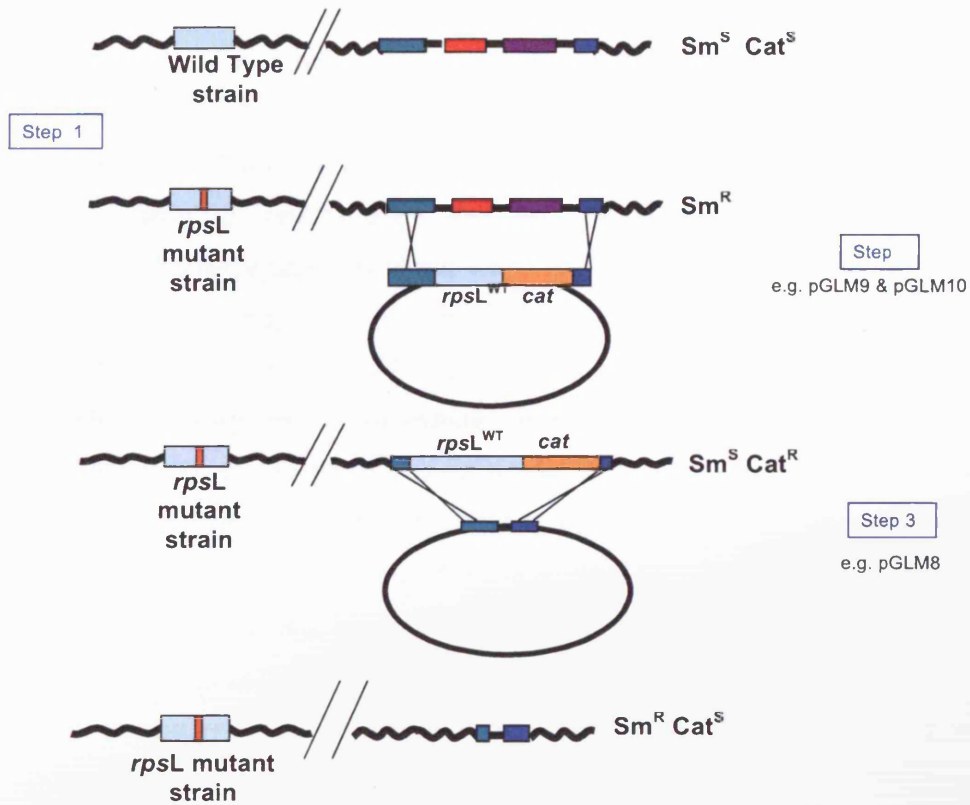
Hendrixson *et al.* (2001) published a method to construct defined chromosomal deletion mutants, see Figure 6.3. This method utilises the ability of *C. jejuni* to form spontaneous mutations in its *rpsL* gene (a ribosomal protein subunit which is present in multiple copies) conferring streptomycin resistance ( $\text{Sm}^R$ ). The wild-type *rpsL* gene, *rpsL*<sup>WT</sup>, which is dominant and confers streptomycin sensitivity, can then be used to create an antibiotic cassette with the chloramphenicol acetyltransferase gene, *cat*. This antibiotic cassette can be inserted into a plasmid containing the gene of interest. These suicide plasmids can be electroporated into the  $\text{Sm}^R$  *C. jejuni* strain and transformants can be selected for chloramphenicol resistance and streptomycin sensitivity as the *rpsL*<sup>R</sup> mutation is recessive. The cassette is then removed by creating plasmids containing flanking genes and

electroporating this into the intermediate mutant, allowing selection for streptomycin resistance and chloramphenicol sensitivity. Although this strategy is advantageous as it provides a method to produce non-polar mutations, it is a multi-step process and produces a deletion mutant in a *rpsL<sup>R</sup>* mutated background, which has unknown effects and may complicate subsequent phenotypic characterisation.

Despite these potential pitfalls, this mutagenesis strategy was employed to make large deletion mutants in the LOS region as it enables the user to delete multiple genes of interest, while only using one antibiotic resistance cassette.

**Figure 6.3 Diagrammatic Illustration of the defined deletion mutagenesis strategy employed by Hendrixson *et al.*, (2001).** Step 1 involves growing and verifying a streptomycin resistant strain. Step 2 entails the construction of a cassette with the wild type *rpsL* gene and the chloramphenicol resistance gene, *cat* and the gene or genes of interest. This is then electroporated in to the streptomycin resistant *C. jejuni*. Intermediates are selected for streptomycin sensitivity and chloramphenicol resistance. Step 3 involves the introduction of a plasmid containing the flanking genes to be introduced into this intermediate strain to allow crossing out of the cassette. The antibiotic resistance profile of the strains is detailed on the right hand side of the diagram.





## **6.2 Construction of a *wlaP* - *wlaSC* deletion mutant.**

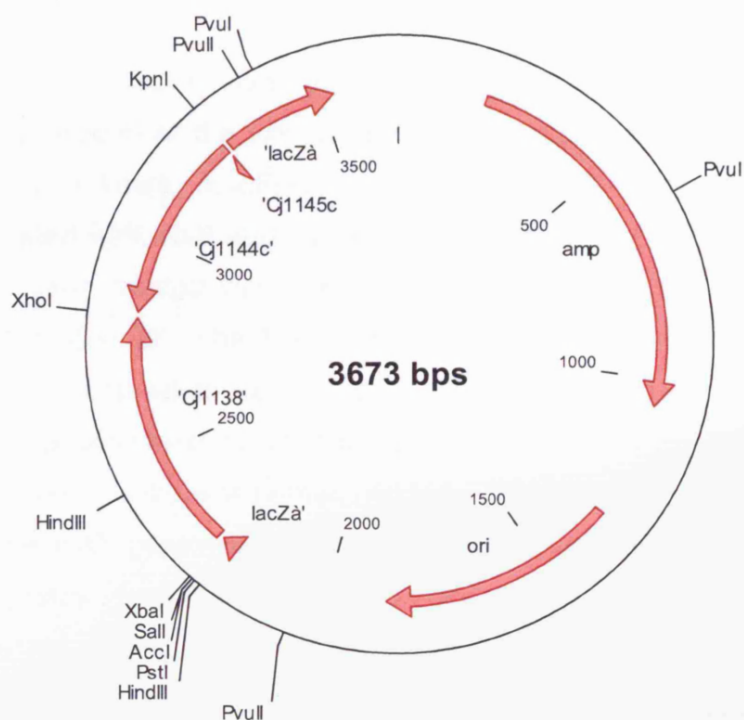
### **6.2.1 Cloning of *wlaP* and *wlaSC* into pUC19**

Initially a small deletion mutant was constructed to establish the methodology. This small deletion mutant involved the removal of the genes between *wlaP* and *wlaSC*, therefore removing the GM1 like epitope, which could easily be confirmed by Western blotting with cholera toxin. Details of the methodology used to clone *wlaP* and *wlaSC* into the vector, pUC19 can be found in Chapter 2. Details of the primers used and the plasmids constructed can be found in Appendix 1. The genes, *wlaP* and *wlaSC* were amplified using the Triplmaster mix system and primers; *wlaPF*(*Xba*I), *wlaPR*(*Xho*I), *wlaSCF*(*Xho*I) and *wlaSCR*(*Kpn*I). These amplified fragments were subjected to a double digest with the restriction enzymes detailed above and ligated into pUC19 in a three-way ligation. After transformation into *E. coli* DH5 $\alpha$ E, colonies containing the recombinant plasmid were identified via blue-white screening. The plasmid DNA was extracted and digested with *Xho*I to confirm the size of the plasmid (data not shown). The DNA inserted into the plasmid's multiple cloning polylinker was sequenced using the universal primers M13F and M13R to ensure sequence errors had not been introduced either during PCR or in plasmid replication. No errors were found during sequence analysis (data not shown). This plasmid construct was named pGLM8.

### **6.2.2 Obtaining a streptomycin resistant strain.**

A spontaneous streptomycin resistant strain of *C. jejuni* NCTC 11168 was obtained by innoculating the wild type strain onto agar plates containing streptomycin. Three streptomycin resistant mutant strains were isolated and the *rpsL* gene was sequenced to identify the mutated base. All three strains had the same point mutation, from C to T causing an amino acid substitution at position 91 from leucine to proline. Studies in *E.coli* have indicated that the

**Figure 6.4 Schematic diagram of pGLM8, where *Cj11138* is *wlaP* and *Cj11144* is *wlaSC*.**



amino acid at this position is susceptible to mutation, although the amino acid at this position may differ (Taylor *et al.*, 1998).

### 6.2.3 Construction of the CAT-*rpsL*<sup>WT</sup> cassette.

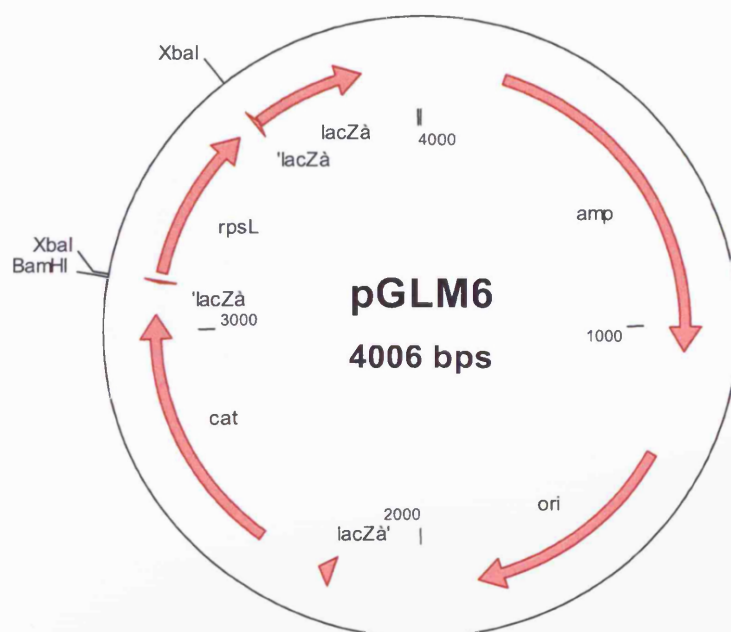
The first step involved in constructing the antibiotic cassette involved PCR amplification of the wild type *rpsL* gene using the TripleMaster™ mix system and primers; *rpsLF*(*Xba*I) and *rpsLR*(*Xba*I). The amplified product was digested with *Xba*I and ligated into *Xba*I digested pUC19. The plasmid had also been treated with shrimp alkaline phosphatase to prevent it from self-ligating due to cohesive ends. The recombinant plasmid was then identified by blue-white screening and the orientation of the gene in the plasmid was confirmed by digestion of the DNA with *Pvu*II and *Spe*I (data not shown). The plasmid was named pGLM4.

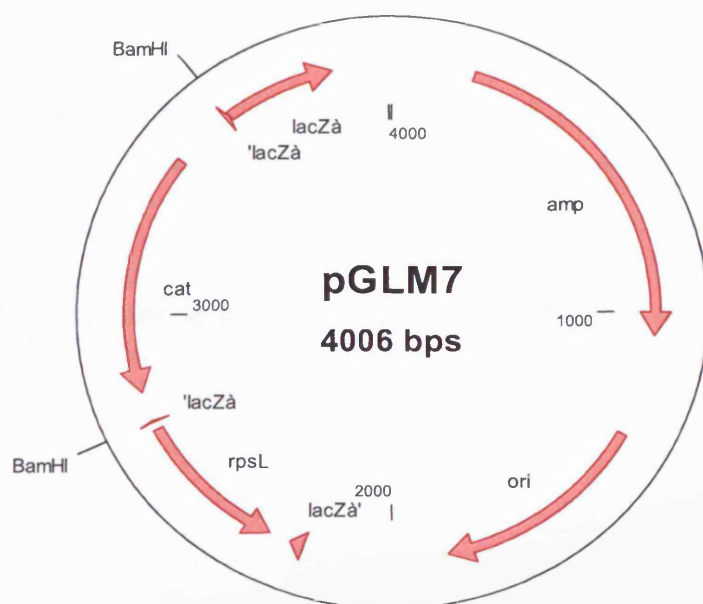
The CAT gene was digested from the plasmid, pAV35, with *Bam*HI. The digested fragment was then cloned into digested and alkaline phosphatase treated pGLM4. Recombinant plasmids were selected on the basis of chloramphenicol resistant colonies and the orientation of the gene was confirmed by double digestion of the plasmid DNA with *Cla*I and *Pvu*I (data not shown). The fidelity of the final constructs, pGLM6 and pGLM7, was then confirmed by sequencing using the universal sequencing primers, M13F and M13R. No errors were evident from the amplification or cloning process (data not shown). A map of the final construct can be found in figure 6.4.

### 6.2.4 Inserting the CAT-*rpsL*<sup>WT</sup> cassette into pGLM8

The three-way ligation to produce pGLM8, had resulted in the production of an *Xho*I site. The CAT-*rpsL*<sup>WT</sup> cassette was amplified using primers CAT*rpsL*WF(*Xho*I) and CAT*rpsL*WR(*Xho*I). After digestion of the amplified product, it was then cloned into digested and alkaline phosphatase treated pGLM8. Transformants containing the recombinant plasmid were selected on the basis of chloramphenicol resistance and the size of the

**Figure 6.5 Schematic diagrams detailing the final constructs of pGLM6 and pGLM7.**







plasmid was confirmed by digestion with *Xho*I. The orientation of the cassette was confirmed by PCR with M13F and M13R in combination with the inverse primers designed to anneal to the CAT cassette, CATINVf and CATINVR.

The plasmid in which the cassette was inserted in the same orientation as the amp<sup>R</sup> cassette was named pGLM9 and that where the cassette was in the opposite orientation to amp<sup>R</sup> was named pGLM10 (data not shown). Maps of the final constructs, pGLM9 and pGLM10, can be seen in figure 6.6.

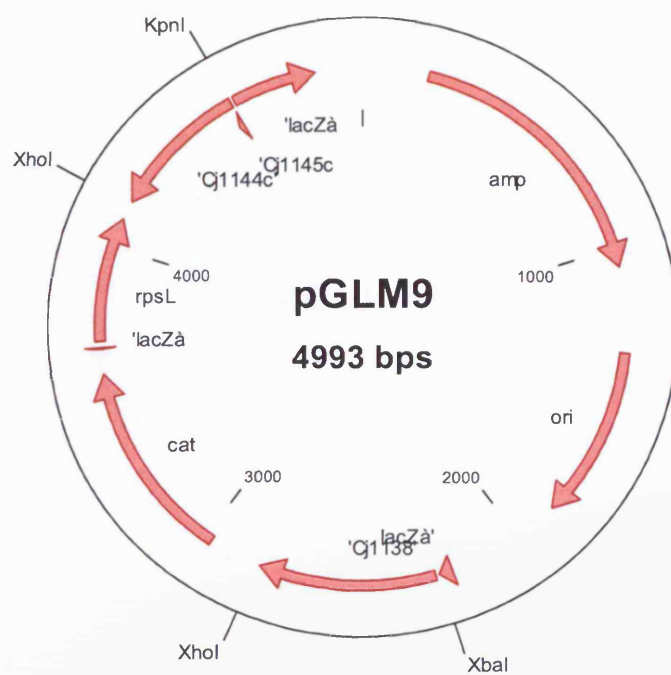
### 6.2.5 Allelic exchange between pGLM9 and pGLM10 with *C. jejuni* NCTC 1168<sup>Strep<sup>R</sup></sup>

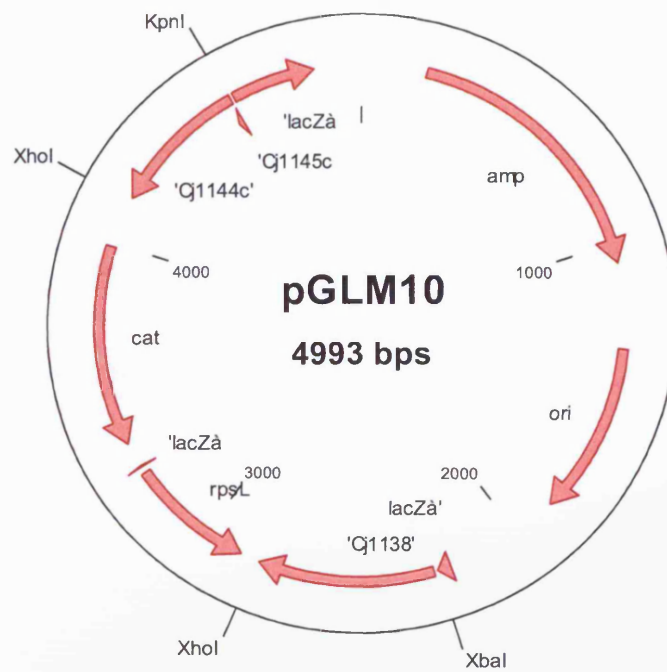
Competent *C. jejuni* cells were prepared, as detailed previously, and the plasmids (pGLM9 and pGLM10) were introduced into the cells by electroporation. The deletion mutants were selected by their ability to grow on plates containing chloramphenicol and the presence of the CAT-*rpsL*<sup>WT</sup> antibiotic cassette was also confirmed by the sensitivity of the mutants to streptomycin. No colonies were obtained after transformation with pGLM10. PCR screening was performed to confirm the 6 Kb deletion using the initial cloning primers; *wla*PF(*Xba*I) and *wla*SCR(*Kpn*I). This not only served to confirm the deletion, but also the presence of one antibiotic cassette (figure 6.7).

The deletion of the genes between *wlaP* and *wlaSC* was also confirmed using a primer walking approach. Using primers designed to the LOS region by BμG@S (Bacterial Microarray Group at St George's, see Chapter 4, section 2), various 'anchor' primers were used in combination with primers designed to each open reading frame. Figure 6.8 shows the results of this experiment in comparison to the cognate wild type strain, where no PCR products are produced in the mutant strain for the genes between *wlaP* and *wlaSC*. The primers used in this experiment are detailed in Appendix 1.

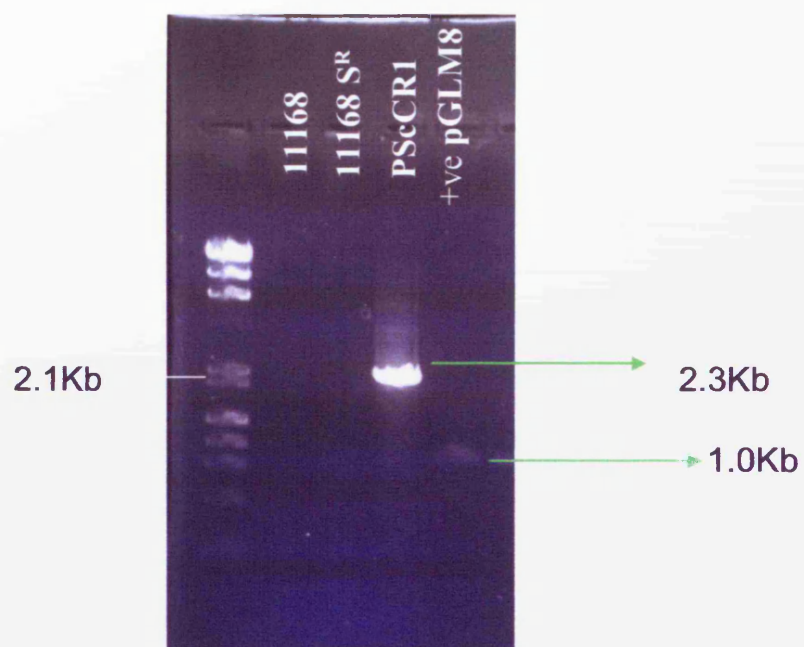
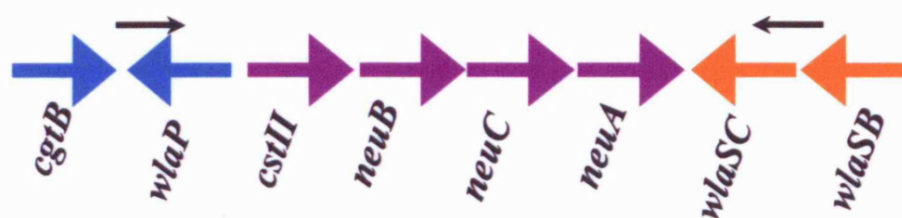
Further confirmation of the deletion size was also determined by Southern blot and sequencing. The genomic DNA of the wild type, streptomycin resistant strains and the deletion mutant were digested with the

**Figure 6.6 Schematic diagrams detailing the final constructs of pGLM9 and pGLM10.**

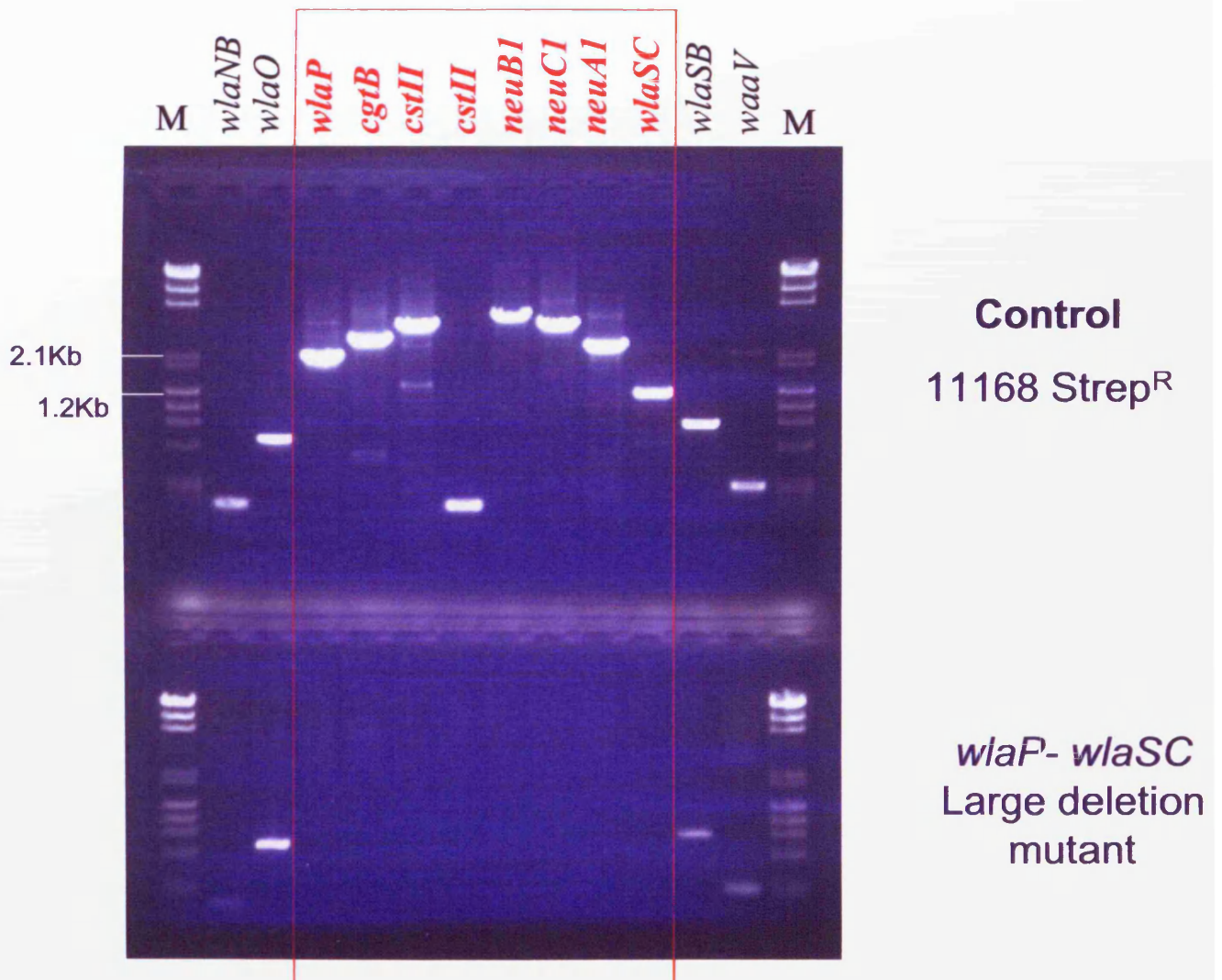
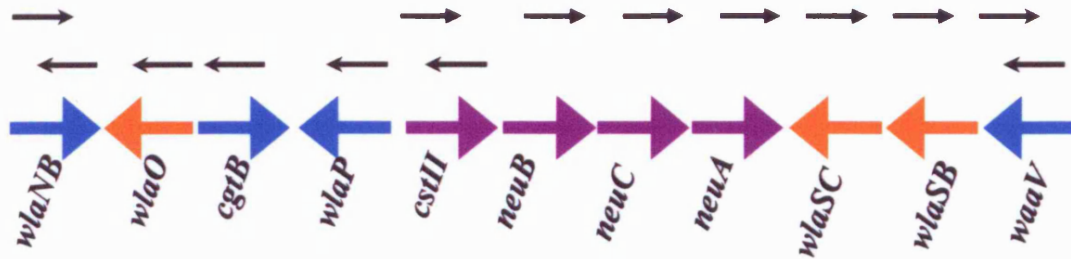




**Figure 6.7 Gel photograph showing the results of the PCR using *wlaPF(XbaI)* and *wlaSCR(KpnI)*. The PCR was performed with the wild-type, streptomycin resistant and deletion mutant strains, with the cloning plasmid as a positive control. The wild type and streptomycin resistant strains (lanes 1 and 2) produce a 6.9 Kb product, which is not amplified using the PCR system. The deletion mutant (lane 4) produces an amplicon of 2.3 Kb, showing a 4.6 Kb deletion and presence of the CAT-*rpsL*<sup>WT</sup> cassette. Lane 5 shows the 0.9 Kb product produced by the pGLM8 as the positive control. This deletion has also been confirmed by primer walking, see figure 6.8. The schematic diagram above the gel photograph shows the gene context and the black arrows indicate the genes in which the primers anneal.**

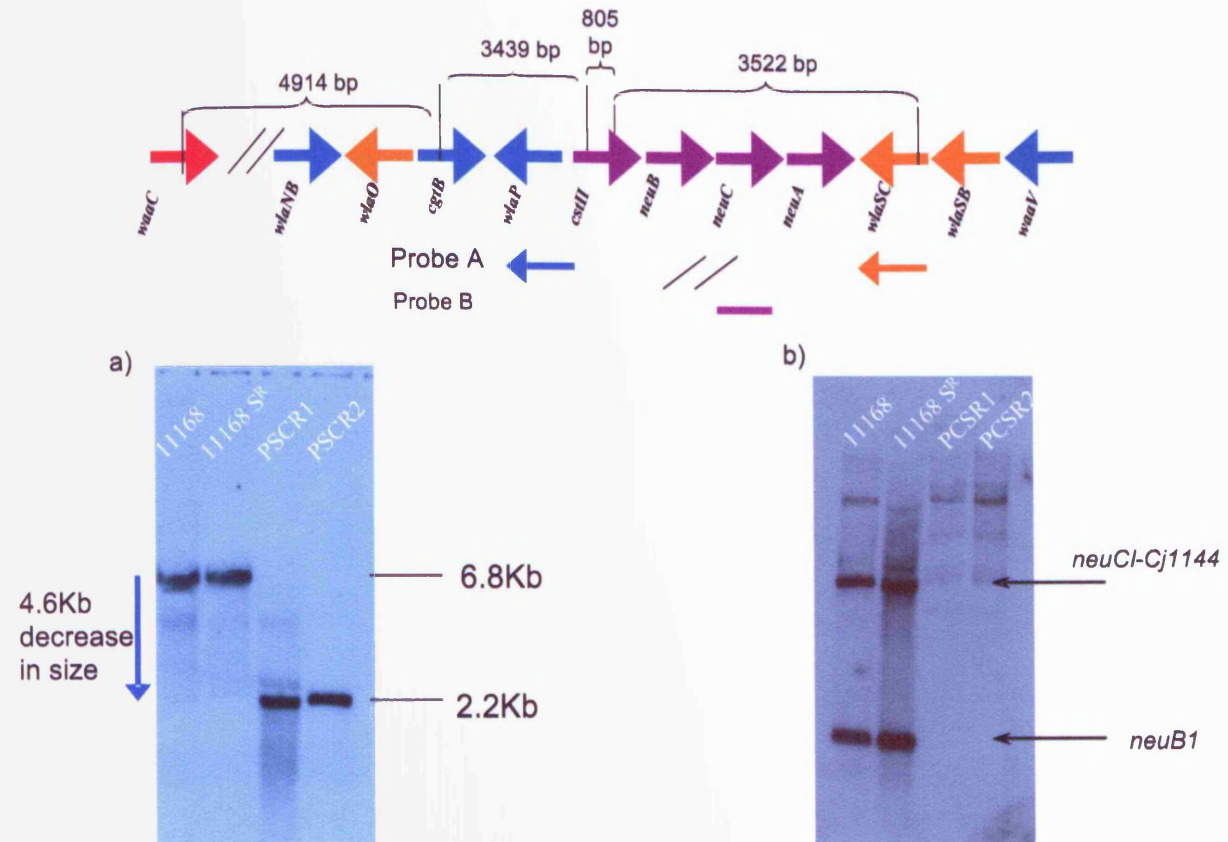


**Figure 6.8. Gel photograph showing primer walking results on *C. jejuni* NCTC 11168 and the *wlaP*-Sc deletion mutant. A forward primer was used as an anchor primer in combination with a reverse primer for each gene from *wlaNB* to *cstII*. A reverse anchor primer designed to *waaV* was used in combination with forward primers from *cstII* to *neuB* to *waaV*. The letter M, indicates the molecular weight markers detailed in chapter 2. The schematic diagram above the gel photograph indicates the gene content and the black arrows represent the primer position.**





**Figure 6.9a shows the results of the Southern blot probed with the *wlaP-wlaSC* fragment. Genomic DNA was digested with *Bgl*II and probed with *wlaP-wlaSC*. The deletion mutant shows a 4.6 Kb decrease in size as expected. Figure 6.9b shows the results of a Southern blot after probing with a *neuBI-neuCI* fragment. After digestion with *Bgl*II, the deletion mutants show the loss of a *neuCI-Cj1145* fragment (~3.5 Kb) and a *neuBI* fragment (805 bp).**



restriction enzyme, *Bgl*II, which loses two restriction sites in the LOS core biosynthesis region of the mutant. Two probes were used to confirm the deletion; the *wlaP-wlaSC* genes were amplified from the original plasmid and a *neuB1-neuC1* PCR product amplified from genomic DNA. In Figure 6.9b, this latter probe can be seen to hybridise to the wild type strains on two separate fragments and not hybridise at all to the deletion strain, indicating that these genes have been deleted. The second probe, *wlaP-wlaSC*, can be seen to hybridise to a much smaller fragment in the deletion strain, with a decrease in size of approximately 4.6 Kb (see Figure 6.9a). Blots were also probed with the CAT-*rpsL*<sup>WT</sup> cassette to confirm its presence within the genome, and with pUC19 to confirm that no plasmid DNA had integrated into the genome (data not shown).

### **6.3 Phenotypic analysis of the *wlaP-wlaSC* deletion mutant**

#### **6.3.1 Analysis of LOS from the *wlaP-wlaSC* deletion mutant by Tricine SDS-PAGE and western blotting**

The LOS from the deletion mutant was extracted and compared against the wild type strains by separation on tricine SDS-PAGE. The SDS-PAGE gels were then subjected to silver staining or western blotting. Silver staining showed that the core epitopes produced by the deletion mutant were truncated compared to that produced by the wild type strains (Figure 6.10a). Western blotting was performed using cholera toxin. The result of the western blot hybridised with the cholera toxin (see Figure 6.10b) show loss of cholera toxin binding in the deletion mutant, and therefore loss of the GM<sub>1</sub> – like epitope. The lectin binding potential of this deletion mutant was also explored using peanut agglutinin and wheatgerm agglutinin but despite several attempts using SDS-PAGE gels with different lengths and differing acrylamide concentrations no lectin binding could be seen (Linton *et al.*, 2002). The predicted LOS structure of this mutant is detailed in Figure 6.11.

**Figure 6.10a. Analysis of LOS by tricine-SDS PAGE.** The figure shows an 18% (v/v) SDS-PAGE gel with LOS extracts from wild type (lane 1), streptomycin resistant (lane 2) and *wlaP-wlaSC* deletion mutant (lane 3) stained with silver nitrate using the method of (Tsai and Frasch 1982). The size of the WT core is indicated, as is the truncated core of the *wlaP-wlaSC* deletion mutant. The double band present in lane 3 (Streptomycin resistant strain) shows phase variation causing switching from a GM<sub>1</sub>- to a GM<sub>2</sub>- like epitope. This is not seen in the corresponding lane of the western blot as cholera toxin only binds the GM<sub>1</sub>- like epitope. **Figure 6.10b Analysis of LOS by western blot with cholera toxin subunit B**, where cholera toxin subunit B binds to the parent strains but not to the *wlaP-wlaSC* large deletion mutant.

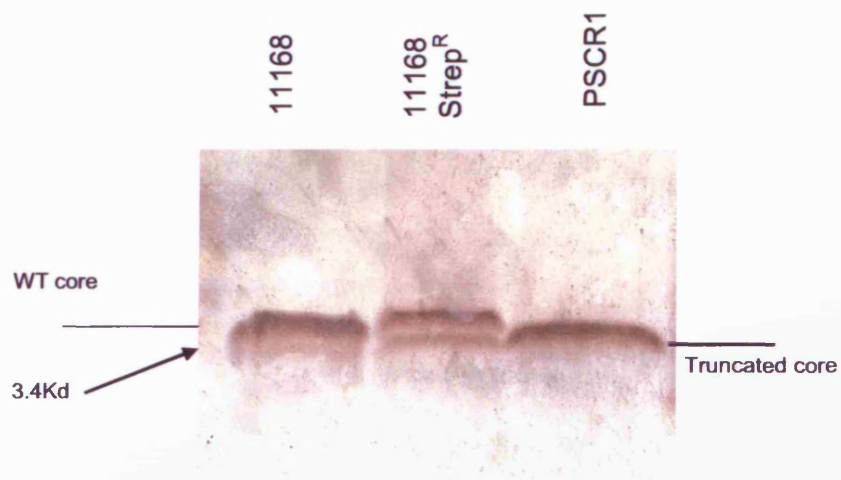


Figure 6.10a

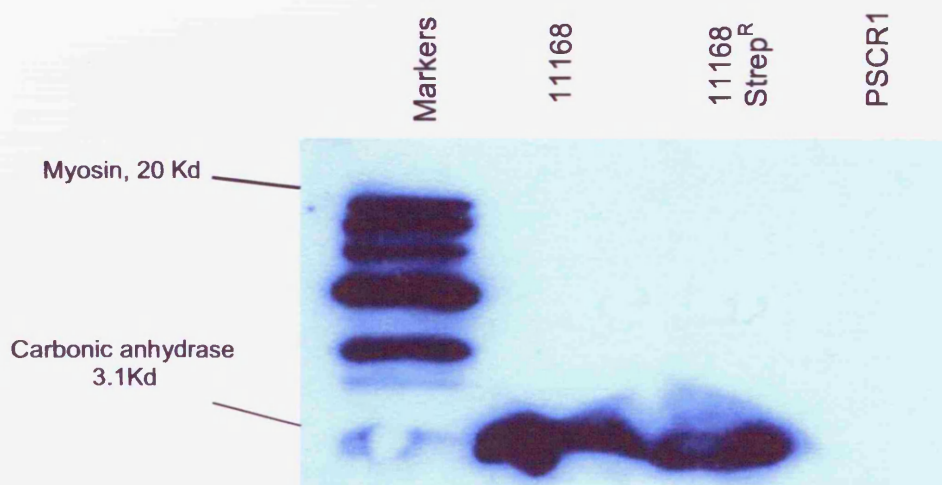
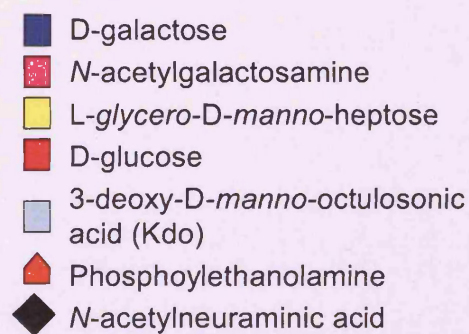
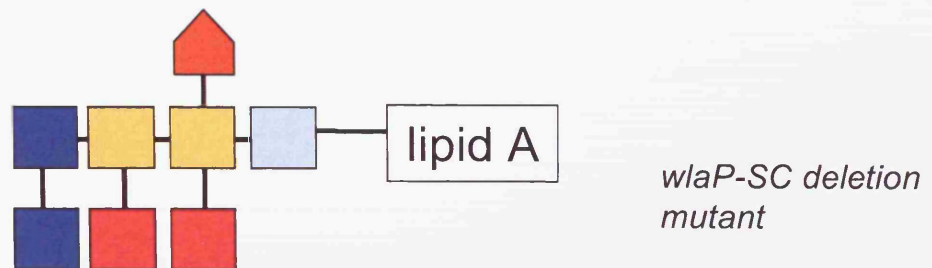
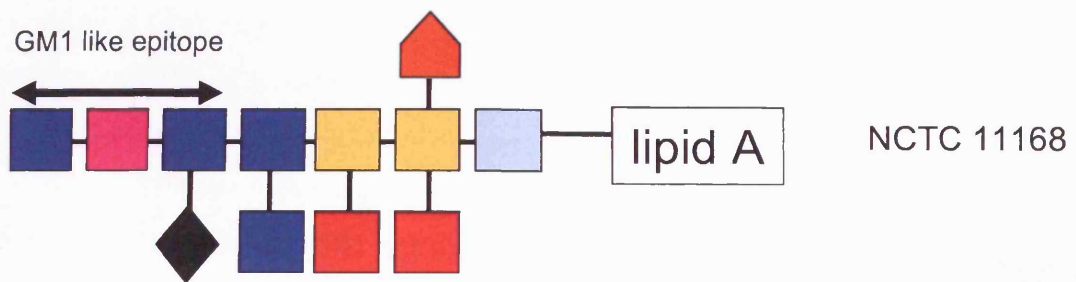


Figure 6.10b

**Figure 6.11.** A schematic diagram indicating the proposed structure of the LOS core for the *wlaP-wlaSC* deletion mutant and the wild-type NCTC 11168.





## **6.4 Construction of a *wlaA-wlaT* large deletion mutant**

### **6.4.1 Cloning of *wlaA*, *wlaT* and CAT-*rpsL*<sup>WT</sup> into pUC19.**

The rationale behind constructing this large deletion mutant was that the minimal LOS core gene content required for viability would be confirmed. The construction of a *wlaP*-SC deletion mutant established proof of principle for the deletion mutagenesis strategy.

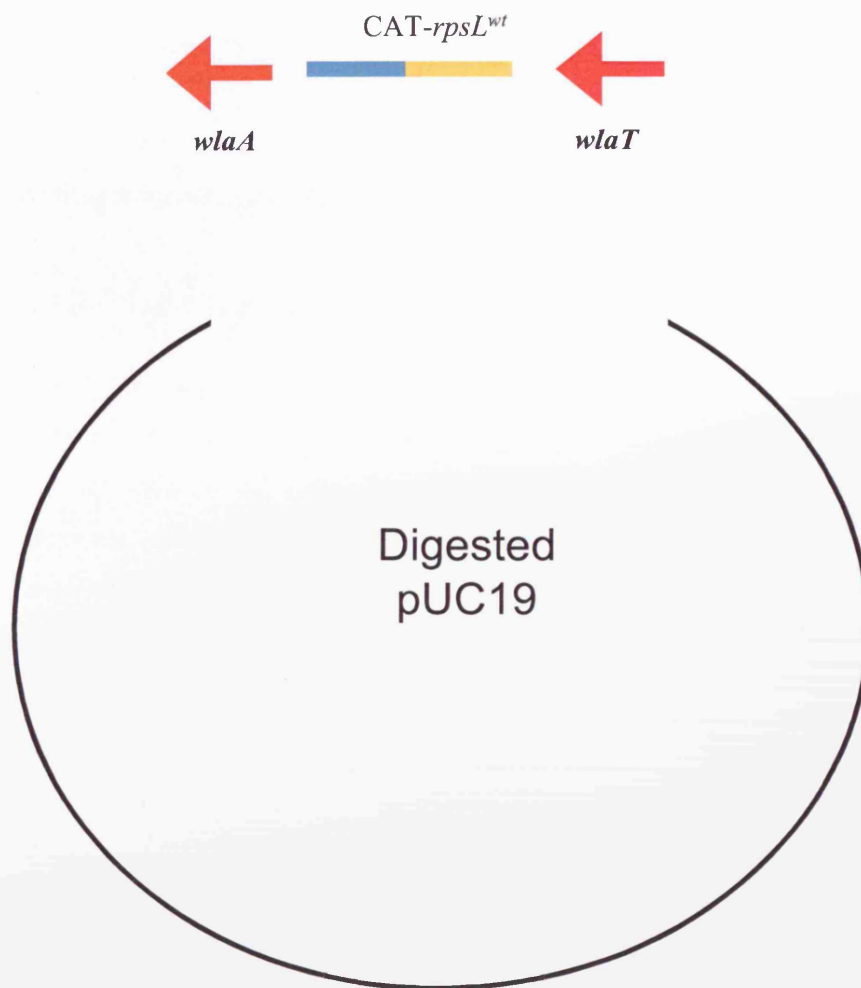
The genes, *wlaA* and *wlaT* (*gmhB*) were amplified using the Triplmaster mix system and primers; *wlaAF*(*Xba*I), *wlaAR*(*Xho*I), *1152F*(*Xho*I) and *1152R*(*Kpn*I). These amplified fragments were subjected to a double digest with the restriction enzymes detailed above and ligated together with digested CAT-*rpsL*<sup>WT</sup> into pUC19 in a four-way ligation (see figure 6.12). After transformation into *E. coli* DH5αE, colonies containing the recombinant plasmid were identified via blue-white screening and chloramphenicol resistance. The plasmid DNA was extracted and digested with *Xho*I to confirm the size of the plasmid and *Kpn*I to confirm the orientation of the antibiotic cassette (data not shown). The DNA inserted into the plasmid's multiple cloning polylinker was sequenced using the universal primers M13F and M13R to ensure sequence errors had not been introduced either during PCR or in plasmid replication. No errors were found during sequence analysis (data not shown). These plasmid constructs were named pGLM12 and pGLM13.

### **6.4.2 Allelic recombination between pGLM12 and pGLM13 with *C. jejuni* NCTC 11168<sup>StrepR</sup>**

The plasmids were introduced into competent *C. jejuni* cells by electroporation and deletion mutants were selected on the basis of chloramphenicol resistance. Colony PCR using *wlaAF* and *1152R* was used



**Figure 6.12 schematic diagram showing the four separate components of the four way ligation used to construct pGLM12 and pGLM13.**



to screen the transformants for the deletion (data not shown). After the extraction of genomic DNA, the transformants were analysed with primer walking, using primers from Cj1130f –Cj1153r. Examination of these PCRs revealed that the LOS region was not deleted and that recombination had occurred within the LOS region (data not shown). This experiment was repeated several times, with varying DNA concentration without success.

#### 6.4.3 Inserting a kanamycin resistance cassette into pGLM12

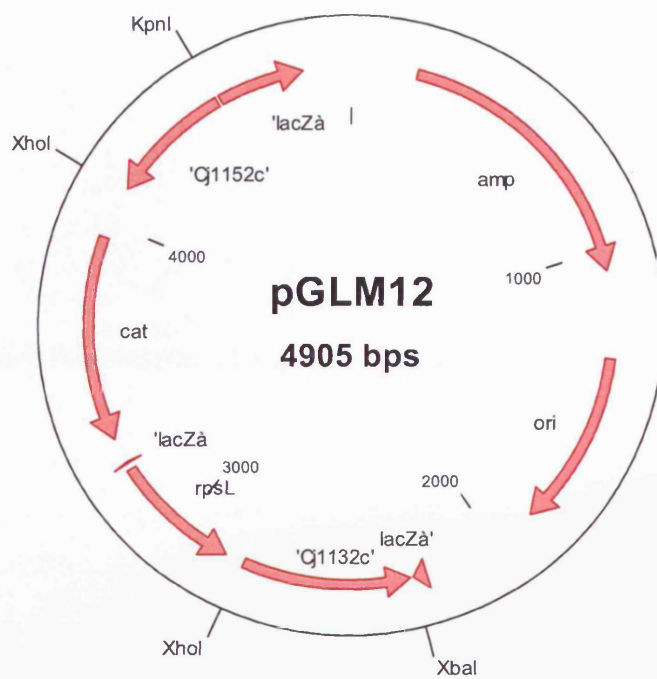
As directly electroporating pGLM12 or pGLM13 into competent *C. jejuni* NCTC 11168::Strep<sup>R</sup> did not yield any deletion mutants, the transformation was attempted using competent 11168::wlaP-wlaSC cells. To enable selection for the larger deletion mutant, an alternative selectable marker was required as the competent cells were chloramphenicol resistant. A kanamycin resistance cassette was therefore used to replace the CAT-*rpsL*<sup>WT</sup> cassette in pGLM12.

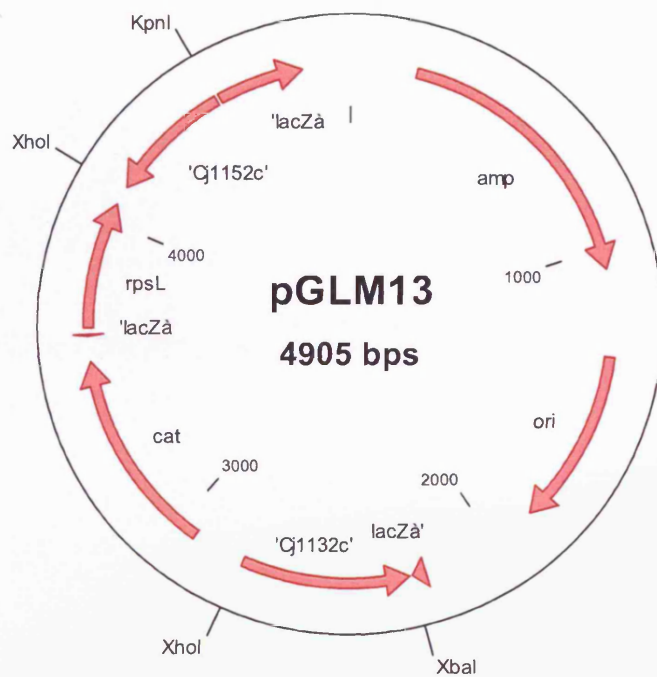
The switching of the antibiotic cassettes was performed by using a number of procedures; digestion of pGLM12 with *Xho*I, gel extraction of the plasmid and filling in/chewing back of the cohesive ends with *Pfu* polymerase to produce a blunt ended vector. This vector was then ligated with an *Eco*RV digested kanamycin resistance cassette. After electroporation into *E. coli* DH5αE, transformants were selected on the basis of kanamycin resistance. The extracted plasmid DNA was then digested with a *Xba*I and *Kpn*I double digest to check the integrity of the insert DNA, and *Kpn*I and *Pst*I single digests to check the orientation of the kanamycin resistance cassette (data not shown). The plasmids were named pGLM14 and pGLM15, and figure 6.13 shows the maps of these plasmids.

#### 6.4.4 Allelic exchange between *C. jejuni* NCTC 11168::wlaP-wlaSC, pGLM14 and pGLM15

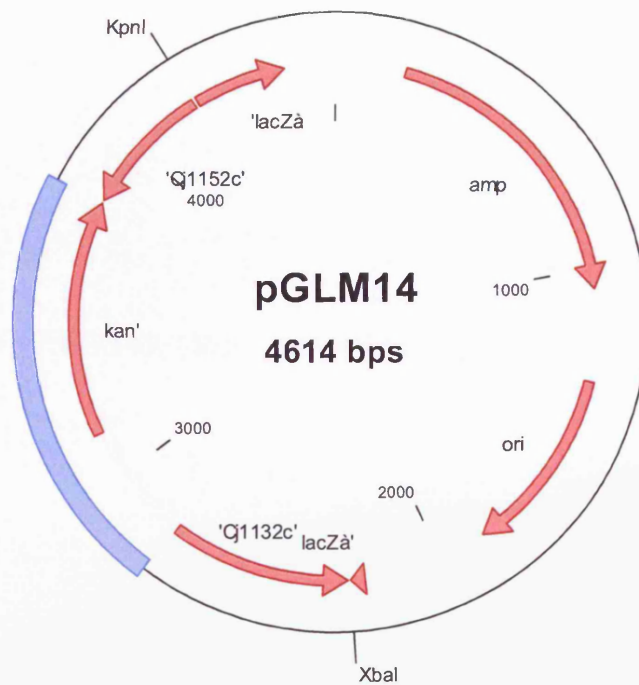
After preparation of competent cells, the plasmids, pGLM14 and pGLM15 were introduced into the cells by electroporation. One transformant

**Figure 6.13 Schematic diagrams representative of the final plasmid constructs for pGLM12 and pGLM13.**

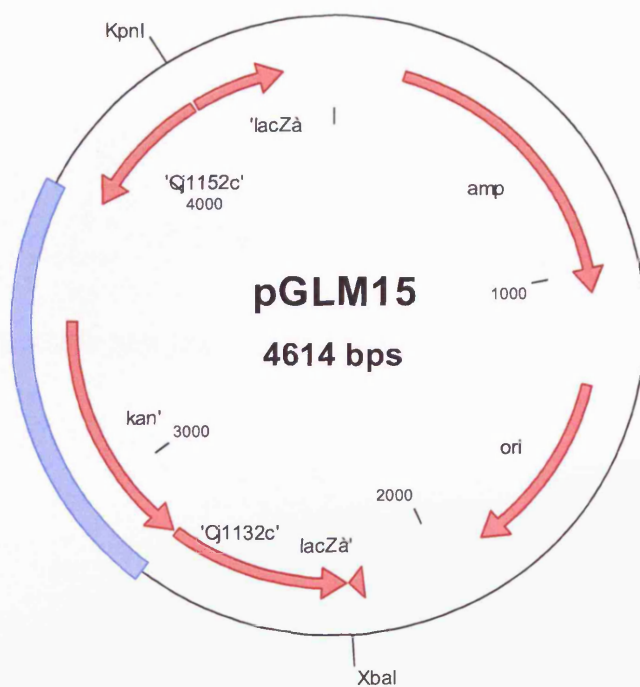




**Figure 6.14: Schematic diagrams of the final constructs for pGLM14 and pGLM15**







**Figure 6.15. Gel photograph showing the amplicons produced by *cj1130F* and *cj1153R* primers. The *wlaA-T* deletion mutant produced the expected 3.8 Kb amplicon and the wild type strain does not produce a product.**

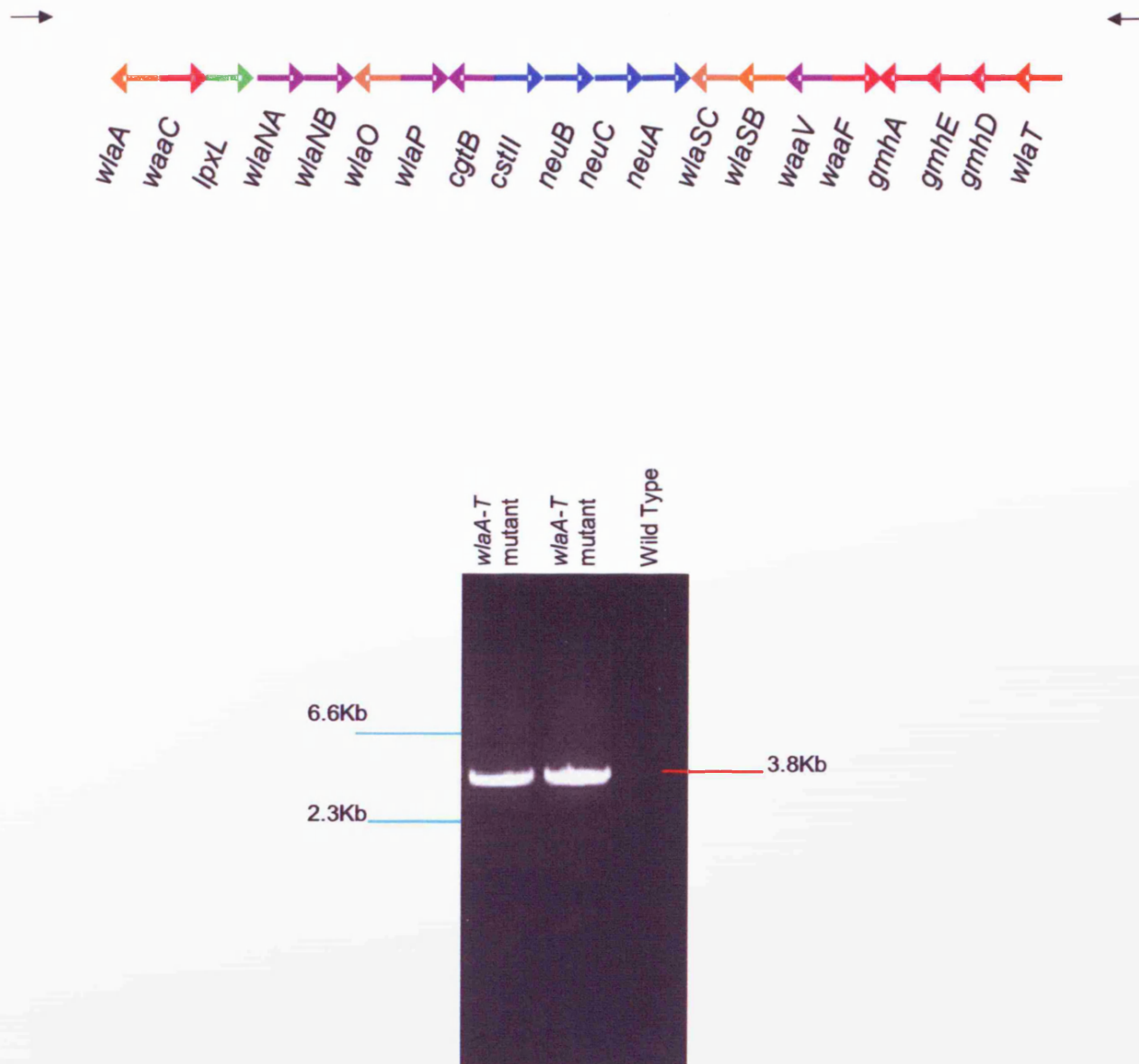
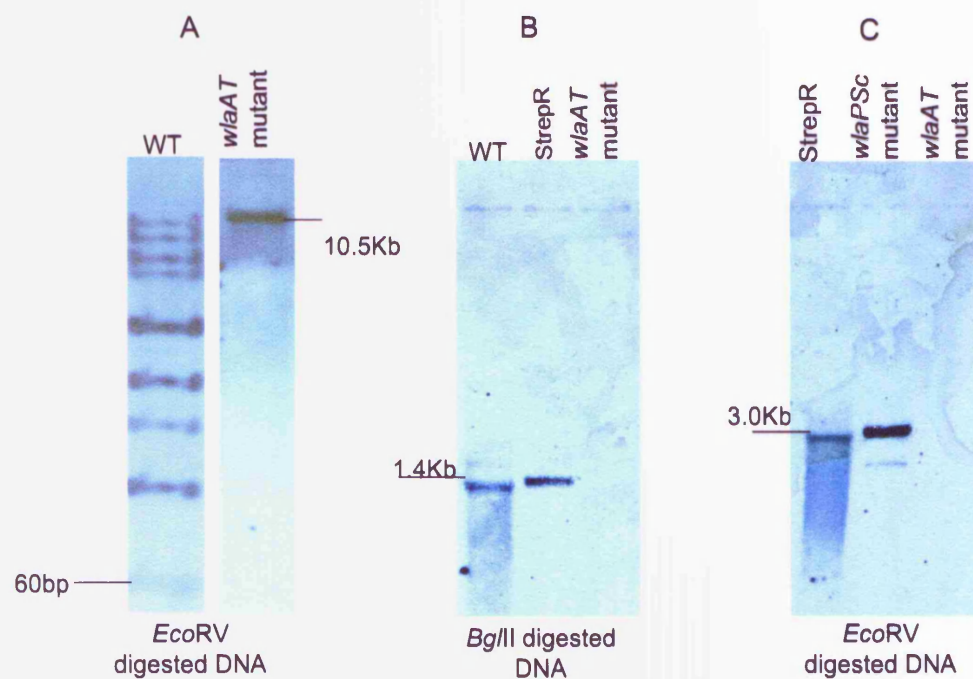
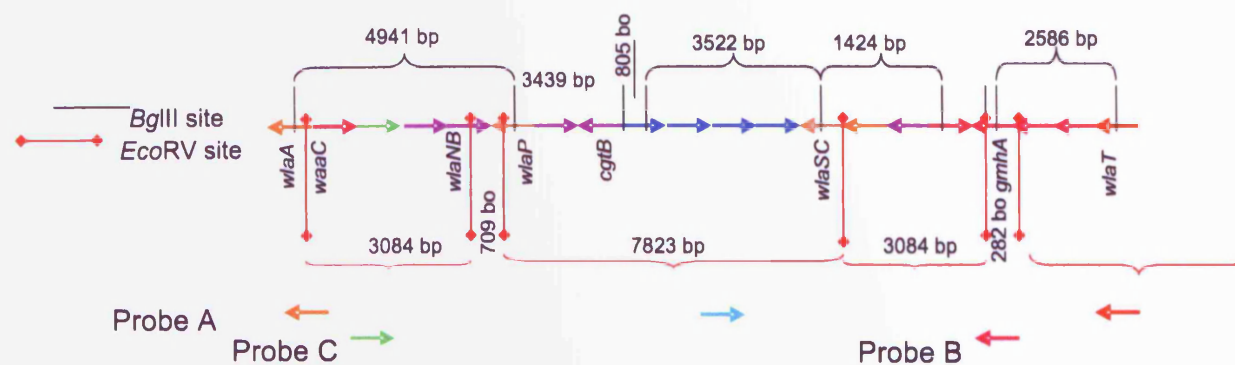


Figure 6.15

**Figure 6.16. showing Southern blots to analyse the *wlaA-wlaT* large deletion mutant.** Panel **A** shows a genomic DNA digested with *Bgl*II and probed with a *wlaAwlaTKAN* fragment amplified from pGLM14, which hybridises to 9 fragments in the wild type strain and 1 fragment in the *wlaA-wlaT* large deletion mutant. Panel **B** shows the results of hybridising a *gmhA* (*Cj1149*) probe to *Bgl*II digested genomic DNA where signal is only seen in the wild type and streptomycin resistant strains at ~1.4Kb. Panel **C**: genomic DNA was digested with *EcoRV* and probed with a *lpxL* (*Cj1134*) fragment resulting in signal from the wild type, streptomycin resistant and *wlaP-wlaSC* deletion strains at ~3Kb. The difference in the fragment size of the parent strains in panel B maybe due to under-digested genomic DNA as no other differences have been seen between the wild type and streptomycin resistant strains.

## Chapter 6: Determining the minimal LOS core gene content for viability



was selected on the basis of kanamycin resistance, and was noted to be very slow growing, taking over 72 hours longer than other mutants to grow.

After extraction of the genomic DNA, PCR was performed using the primers Cj1130cr and Cj1153cf. As can be seen in Figure 6.15, this PCR produces the expected 3.8kb amplicon and no amplicons from the *wlaP-wlaSC* deletion mutant or the wild type strain (amplicons would be 14.8Kb and 20.7Kb respectively). The amplicons produced were then sequenced using internal primers to confirm the boundaries of the deletion and the absence of sequencing errors (data not shown).

Further confirmation of the deletion was sought by Southern blot analysis. The genomic DNA was digested with *Bgl*II, which cut to produce nine fragments within the wild type LOS cluster. Figure 6.16a shows the results of the blotted *Bgl*II digested DNA when probed with the *wlaA-KAN-wlaT* fragment, which hybridises over 9 fragments in the wild type and streptomycin strains and once in the *wlaA-wlaT* deletion mutant. The blots were also probed with a *wlaA-wlaT* fragment and produced the correct results (data not shown). The genomic DNA was also probed with an amplicon corresponding to *Cj1134* and an amplicon corresponding to *Cj1149*, both these probes should have hybridised to one fragment in the wild type and small deletion strains and should not hybridise at all to the large deletion mutant (Figure 6.16b and c).

## **6.5 Phenotypic analysis of the *wlaA-wlaT* large deletion mutant**

### **6.5.1 Analysis of LOS from the *wlaA-wlaT* deletion mutant by Tricine SDS-PAGE and western blotting**

The LOS from the deletion mutant was extracted and compared against the wild type strains by separation on tricine SDS-PAGE. The SDS-PAGE gels were then subjected to silver staining or western blotting. Silver staining showed that no core was detected in the deletion mutant (figure 6.17). Western blotting was performed using cholera toxin and the lectin,

**Figure 6.17. Analysis of LOS by Tricine-SDS PAGE.** The figure shows an 18% (v/v) SDS-PAGE gel with LOS extracts from wild type (lane 1), streptomycin resistant (lane 2) and *wlaP-wlaSC* deletion mutant (lane 3) stained with silver nitrate using the method of (Tsai and Frasch 1982). The size of the WT core is indicated, as is the truncated core of the *wlaP-wlaSC* deletion mutant.

**Figure 6.18 Analysis of LOS by western blot with cholera toxin subunit B,** where cholera toxin subunit B binds to the parent strains but not to the *wlaP-wlaSC* large deletion mutant.

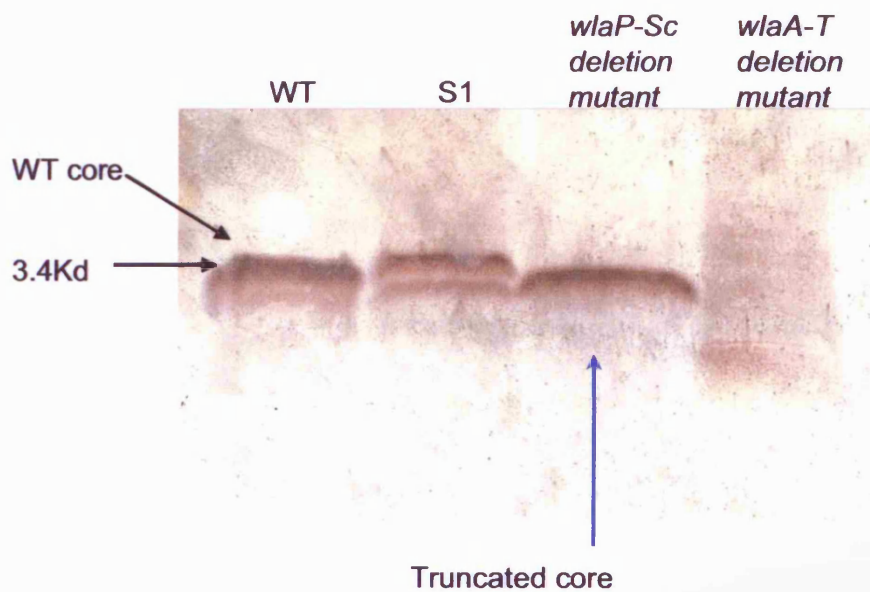


Figure 6.17

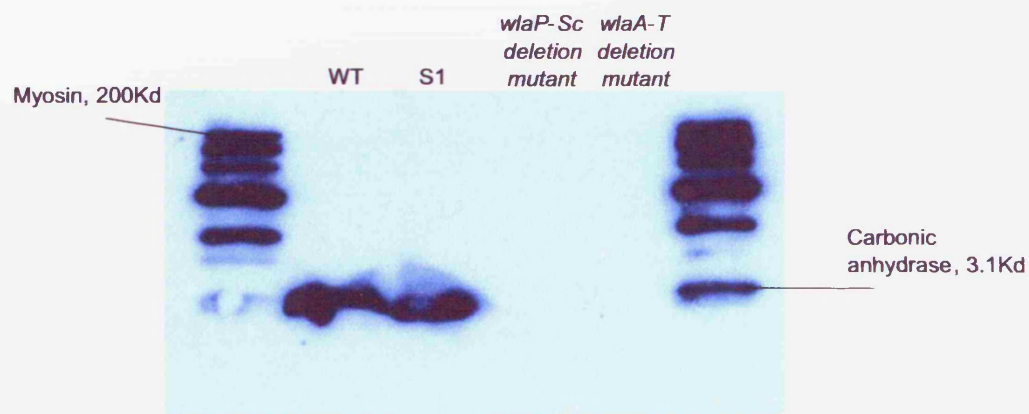
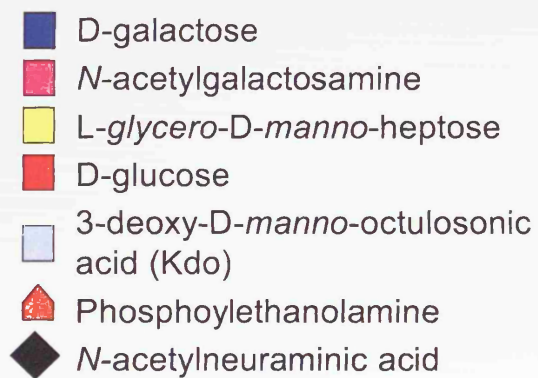
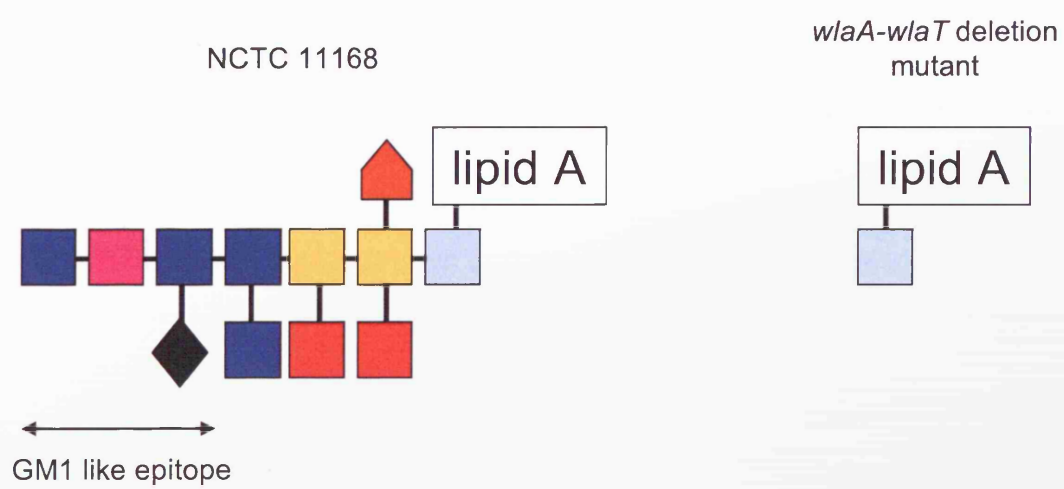


Figure 6.18



**Figure 6.19.** A schematic diagram indicating the proposed structure of the LOS core for the *wlaA-wlaT* deletion mutant and the wild-type NCTC 11168





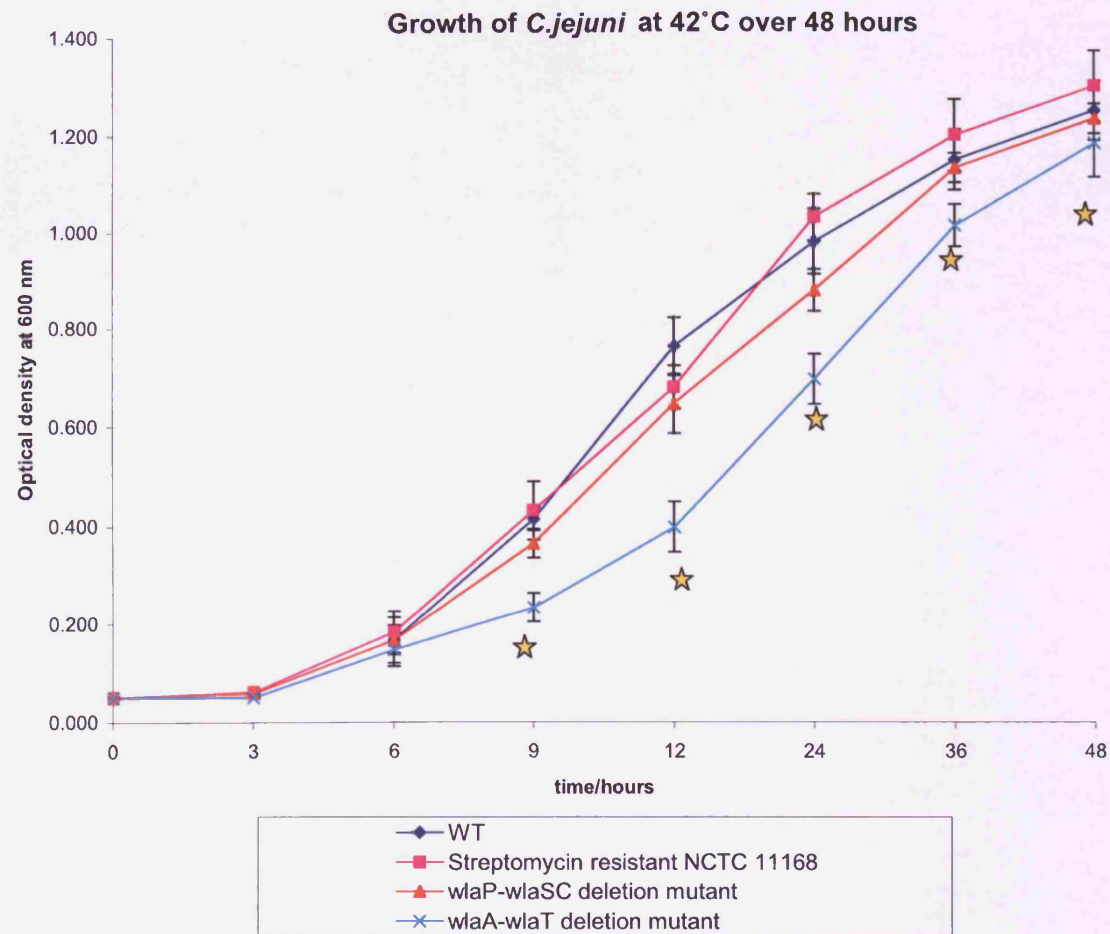
wheat germ agglutinin (RCA) as instead of a primary antibody. The result of the western blot hybridised with the cholera toxin (see figure 6.18) show loss of cholera toxin binding in the deletion mutant, and therefore loss of the GM<sub>1</sub> – like epitope. The binding of lectins (Linton *et al.*, 2002) to this mutant was examined by western blot, but problems were encountered as with the *wlaP-wlaSC* mutant.

## **6.6 Phenotypic analysis of the *C.jejuni* *wlaP-SC* and *wlaA-T* deletion mutants**

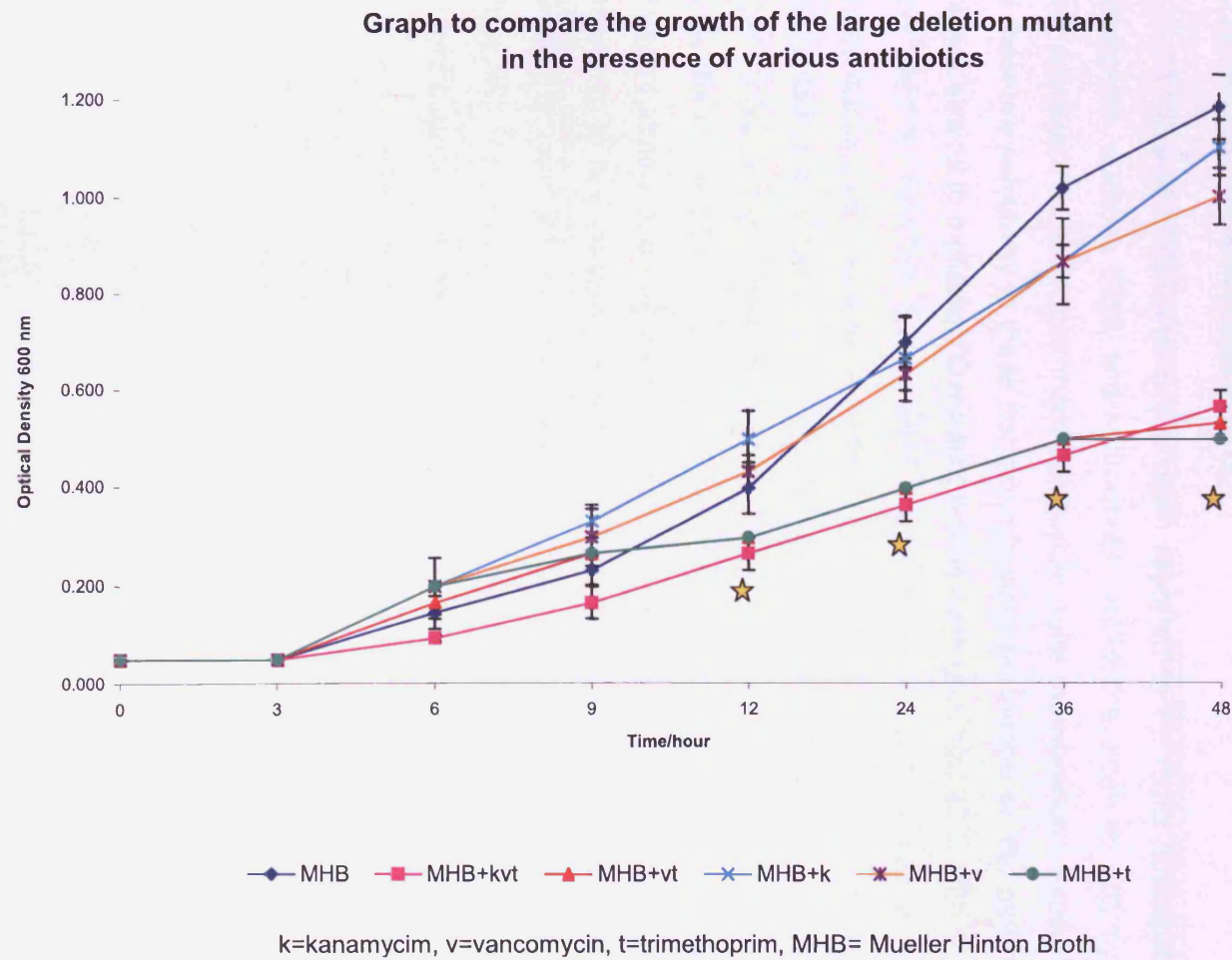
### **6.6.1 Growth patterns of the *C. jejuni* *wlaP-SC* and *wlaA-T* deletion mutants.**

The growth characteristics of deep rough mutants of both *C. jejuni* and other bacteria are altered in comparison to wild type strains. The *wlaA-wlaT* deletion mutant took longer to recover on MHA plate after electroporation than other mutants, and as such the growth assay was extended beyond the usual 24 hours incubation. To examine the growth of the deletion mutants in comparison to the wild type and streptomycin resistant strains, 5ml of MHB was inoculated with an overnight broth to an OD<sub>600 nm</sub> of 0.05 and was incubated with shaking at 42°C. The optical density was measured at 3 hourly intervals for 12 hours and then at 24, 36 and 48 hours. Figure 6.20 shows the average of six biological replicates for each time point and the bars indicate the standard error. The growth pattern of the wild type and streptomycin resistant is identical. The *wlaP-wlaSC* deletion mutant showed no difference in growth over time and there was no statistically significant difference between the growth of this mutant and the parent strains. The *wlaA-wlaT* large deletion strain exhibited a longer lag phase than the parent strains and did not appear to reach the same optical density. The Student's t-test was performed on the result, comparing the wild-type and *wlaA-wlaT* large deletion strains and there was a statistical difference at time points 3, 9, 12 and 24 hours. Details of the statistics used can be found in Chapter 2 and the detailed workings can be seen in Appendix 2, Table 6.1. The results of this

**Figure 6.20 comparing the growth of the parent and mutant strains at 42°C over 48 hours.** The experiment was repeated 6 times and the error bars indicate the standard error at each time point (n=6). See Appendix 2 for full details of OD measurements. Measurements where the difference between the wild type and *wlaAT* mutant is significant are indicated by ★



**Figure 6.21 comparing the growth of the *wlaA-T* mutant strain in various antibiotics at 42°C over 48 hours.** The experiment was repeated 3 times and the error bars indicate the standard error at each time point (n=3). See Appendix 1 for full details of OD measurements. Measurements where the difference between the untreated and trimethoprim treated media is significant are indicated by ★



growth assay did not reflect the negative effect on growth when grown with antibiotics; therefore the growth of this large deletion strain was examined in various antibiotic combinations. This growth assay was performed as before, in triplicate and is detailed in figure 6.21. The addition of all three antibiotics; kanamycin, vancomycin and trimethoprim caused the most significant reduction in growth rate. Singly the addition of trimethoprim alone had the most significant effect on growth.

#### **6.6.2 Sensitivity of the *C. jejuni* *wlaP-SC* and *wlaA-T* deletion mutants to SDS, novobiocin and polymyxin B.**

LPS and LOS mutants are often reported to be more susceptible to detergents, such as SDS, and hydrophobic antibiotics, such as polymyxin B and novobiocin, due to their more permeable outer membranes. Experiments to measure sensitivity to these compounds were performed on the parent and mutant strains in triplicate. Overnight broths were grown at 42°C with shaking and used to inoculate 5ml of MHB that had been supplemented to a final concentration as detailed in table 6.2. The supplemented MHB was inoculated to an OD<sub>600nm</sub> of 0.1 and the cultures were judged to have grown after 16 hours if they had reached an OD<sub>600nm</sub> of 0.2. Figure 6.22 a,b and c, show the mean OD<sub>600nm</sub> at each concentration for the wild type, streptomycin resistant strains and the *wlaP-SC* and *wlaA-wlaT* large deletion mutants. No difference in the sensitivity was seen between the parent and *wlaP-SC* large deletion mutant, where growth was inhibited at a minimum concentration of 25µg ml<sup>-1</sup> of novobiocin and 1µg ml<sup>-1</sup> of polymyxin B. The *wlaP-SC* deletion mutant appeared slightly more sensitive to SDS than the parent strains. The *wlaA-wlaT* large deletion mutant showed a much greater susceptibility to these compounds as the strain did not grow even in the lowest concentrations.



**Table 6.1 detailing the final working concentrations of each treatment used in the assaying the sensitivity of the parent and mutant strains of *C. jejuni*.**

## Chapter 6: Determining the minimal LOS core gene content for viability

Sample No.	Treatment/final working concentration $\mu\text{g ml}^{-1}$		
	SDS	Novobloclin	Polymyxin B
1	400	50	2
2	200	25	1
3	100	12.5	0.5
4	50	6.25	0.25
5	25	3.2	0.125
6	12.5	-	0.0625

### **6.6.3 Invasive potential of the *C. jejuni* *wlaP-SC* and *wlaA-T* deletion mutants.**

Adherence and invasion assays using standard protocols were carried out by Paul Everest (Glasgow) to compare the efficiency of adherence and invasion to CaCo-2 intestinal cells between the wild type *C. jejuni* NCTC 11168, Streptomycin resistant NCTC 11168 and the deletion mutants (MacCallum *et al.*, 2006). The streptomycin and WT strains are the same except for the point mutation in the *rpsL* gene, so this strain has been included to show that the phenotype of the streptomycin resistant strain is the same as the wild-type strain. The genome strain of NCTC 11168 is poorly invasive and therefore low levels of invasion are expected for the parent strains. As is shown in figure 6.23, the *wlaA-T* deletion mutant is unable to invade CaCo-2 cells indicating that the LOS core is important in invasion and adherence. It can not be ruled out that the lack of invasion with this mutant may be because the alteration in the outer membrane has caused it to be more sensitive to tissue culture media and further experiments would need to be performed to ensure that the loss of adherence and invasion is due to the deletion of the LOS core.

The *wlaP-Sc* mutant is 10 x more invasive than the parent strains and this is reproducible perhaps indicating that the loss of this part of the outer core and the ability to switch off the expression of the genes required to construct this epitope is advantageous in the invasion process as seen in previous studies (Linton *et al.*, 2000b; Guerry *et al.*, 2002).

### **6.6.4 Outer membrane protein composition of the *C. jejuni* *wlaP-SC* and *wlaA-T* deletion mutants.**

As there were observed differences in the membrane permeability as observed by the sensitivity assays and in the invasion potential of the *wlaA-wlaT* deletion mutant compared to the wild type, the outer membranes of these strains were compared. Outer membrane preparations were extracted as detailed in Chapter 2. The samples were run on an SDS-PAGE gel and

stained with Commassie blue or silver nitrate (see figure 6.24). No gross changes were noted in the deletion mutants compared to the parent strains in either procedure. The sensitivity of the *w/A-w/aT* strain to antibiotics suggests that the membrane may be leaky and susceptible to loss of proteins and other components in the preparation of the outer membrane that may otherwise stay intact.

### **6.7 Analysis of the LOS from 11168::*wlaP*-SC and 11168::*wlaA*-T by Capillary Electrophoresis - Mass Spectroscopy**

Capillary Electrophoresis Mass Spectroscopy (CE-MS) analysis of these mutants was carried out by Jianjuin Li from the Institute of Biological Sciences, National Research Council of Canada. The core structure of the *wlaA*-T large deletion mutant was predicted to contain one Kdo as shown in Figure 6.13. The structure of 11168::*wlaP*-SC is shown in figure 6.25 and is as predicted in figure 6.8, where the outer core has been deleted.

The late acetyltransferase gene, *lpxL*, was previously deleted (Millar, 2003) and CE-MS analysis of the LOS from the *lpxL* mutant showed that the lipid A molecule was altered but the extent of the alteration was not determined. In the *wlaA*-T mutant, the lipid A is formed of 2 C12 fatty acid chains rather than 2 C16 fatty acid chains (see figure 6.26). Unlike the predicted result for the core structure in figure 6.13, the outer core of the *wlaA*-T mutant contained 2 Kdo moieties.





### **6.8 Restoration of LOS core region in the *wlaA*-T large deletion mutant**

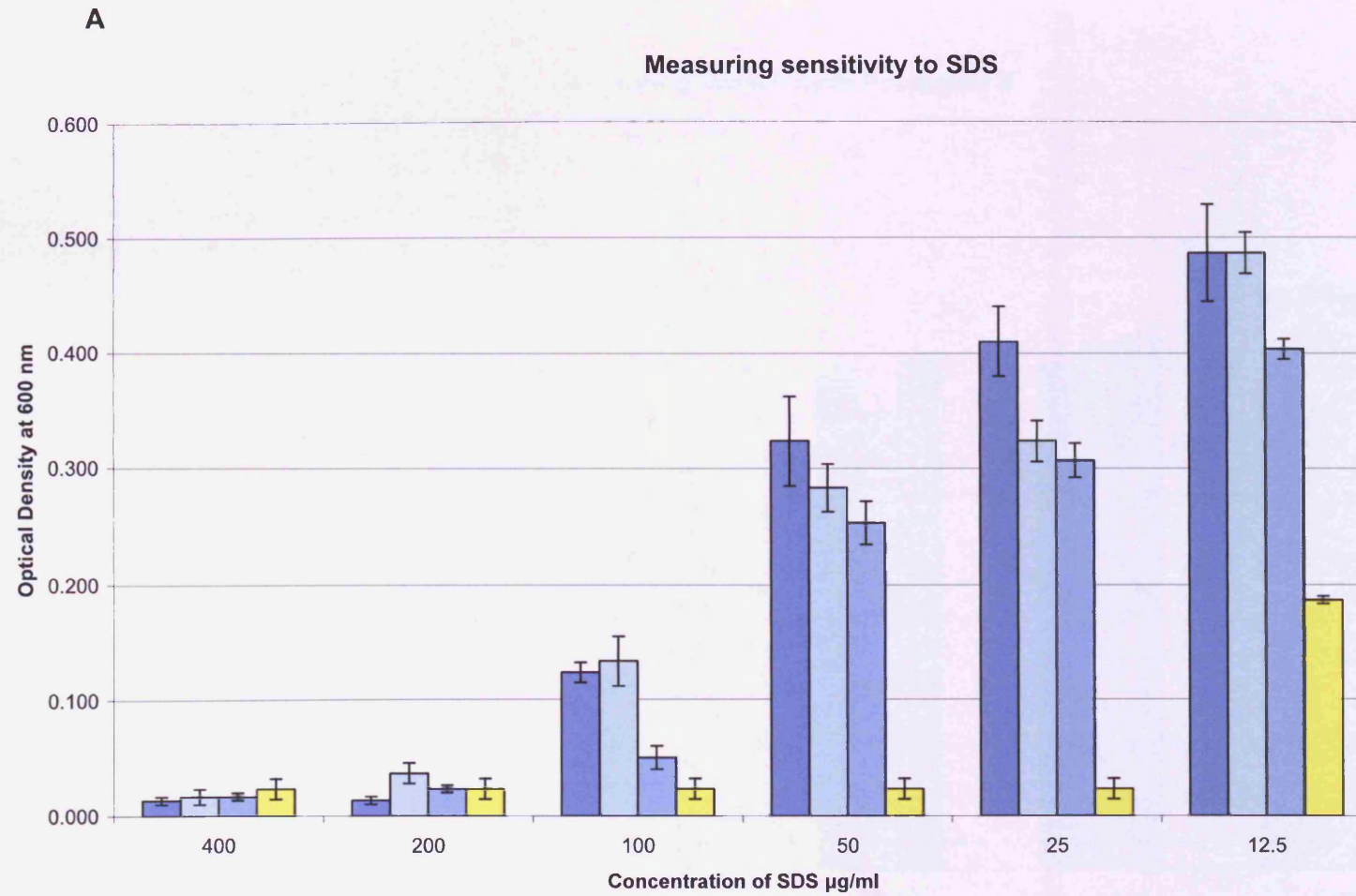
#### **6.8.1 Natural transformations with RM1221, NCTC11168::*wlaP*-SC and NCTC11168::*waaF***

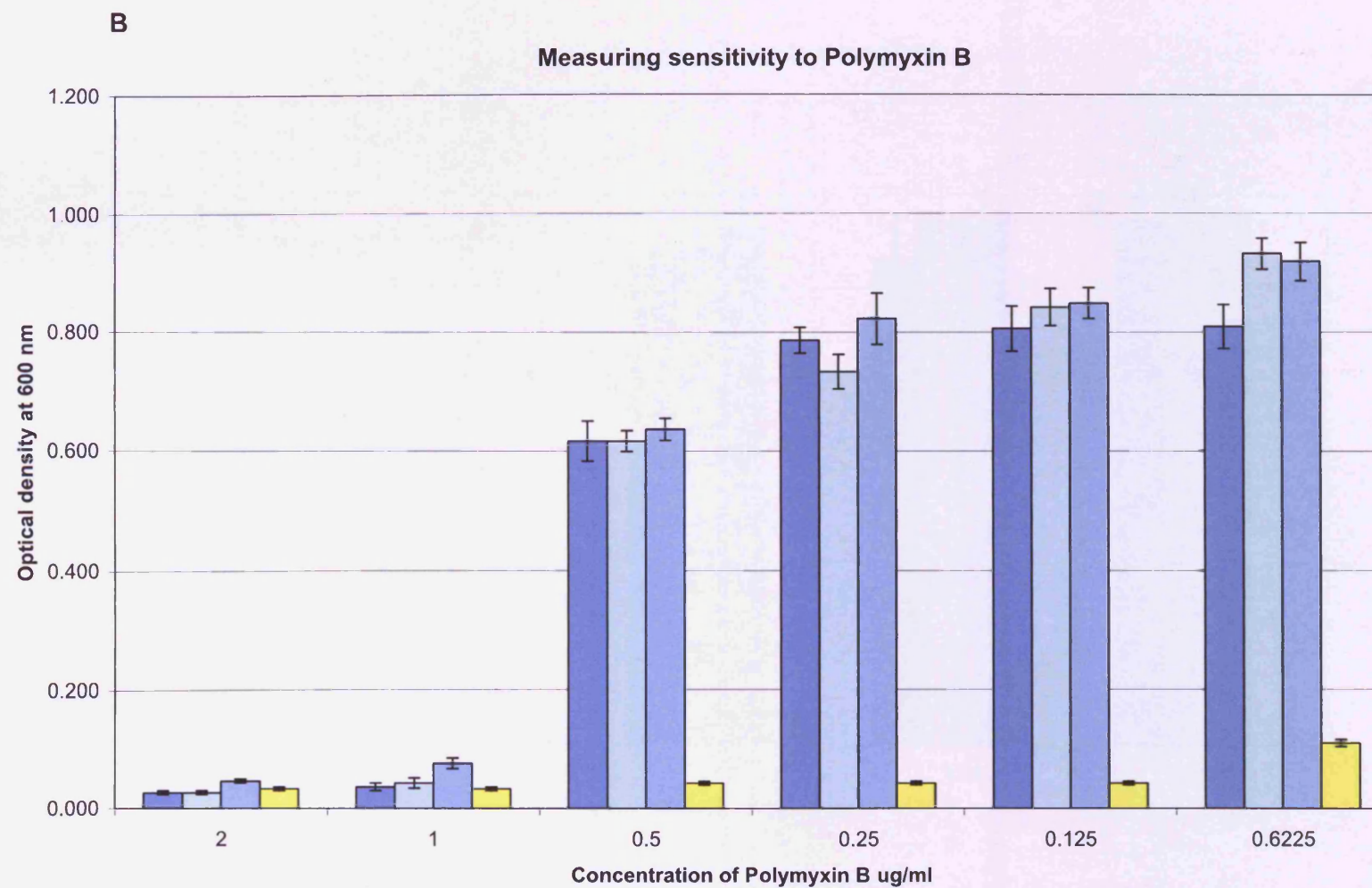
*C. jejuni* has been shown to acquire LOS genes by horizontal transfer and natural transformation (Gilbert, 2004; Phongsisay, 2006). Natural transformation was used as a method to see if the LOS biosynthesis genes

could be replaced in the *wlaA-wlaT* large deletion. Various concentrations of genomic DNA from the strains; RM1221, *wlaP*-SC large deletion mutant and the *waaF* mutant (Oldfield, 2001) were used in the natural transformation process. No transformants were recovered after the experiment with any of the above strains. Another experiment was conducted using 11168 $\Delta$ *cheV*, which has an average transformation efficiency of  $1.6 \times 10^4$  cfu/g (O.Bridle, pers. communication) and again no transformants were recovered implying the loss of natural competence in the *wlaA-wlaT* large deletion strain. The ability of this mutant to take-up DNA by electroporation has not been tested.

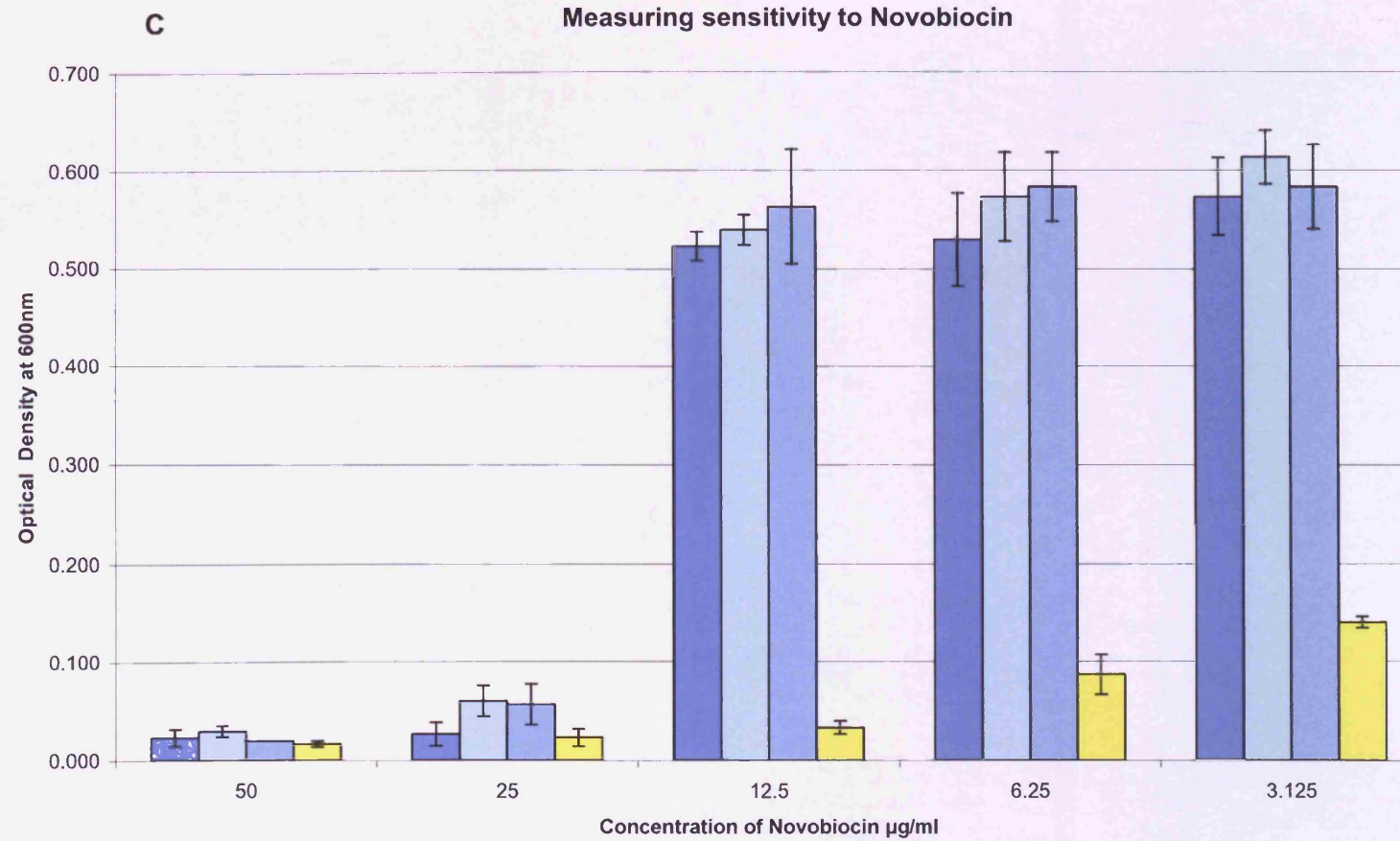
**Figure 6.22 comparing the sensitivity of the deletion mutants to A) SDS, B) Polymyxin B and C) Novobiocin.** The assays were performed in triplicate and the error bars indicate the standard error, where n=3. The complete measurements can be found in Appendix 1.

-  Wild type NCTC 11168
-  Streptomycin resistant NCTC 11168
-  *wlaP-wlaSC* deletion mutant
-  *wlaA-wlaT* deletion mutant



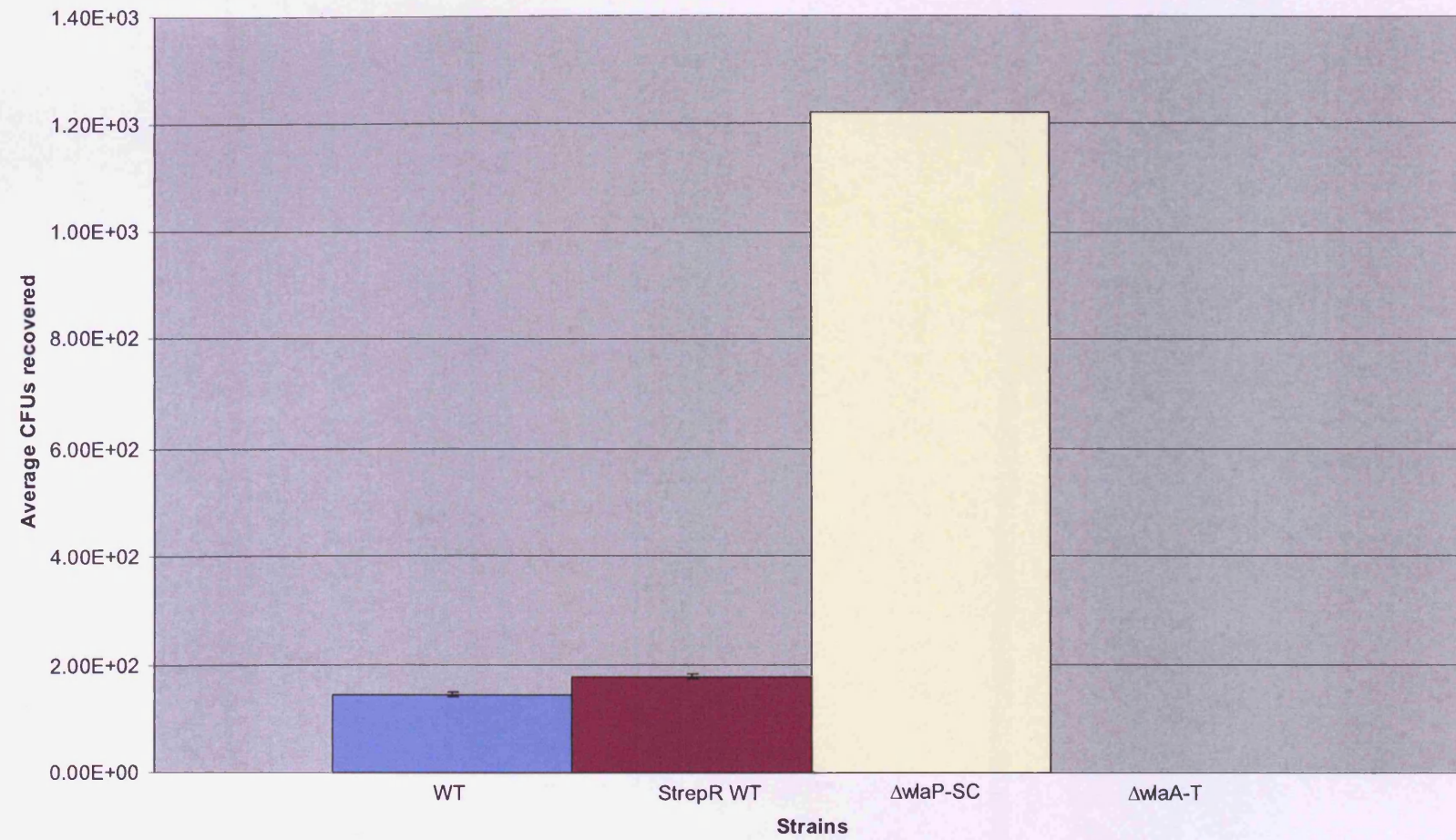




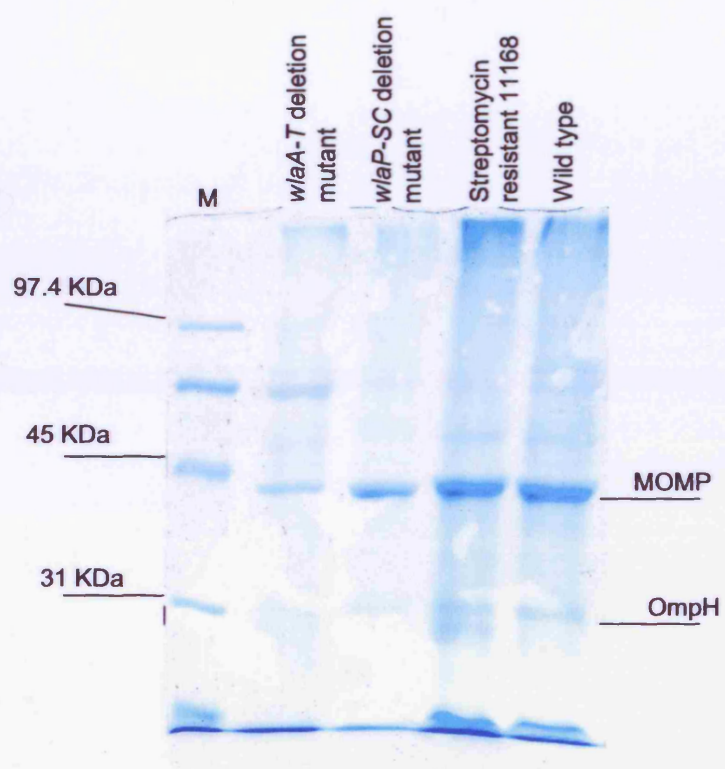


**Figure 6.23** Graph showing the results of *in vitro* invasion experiments with the parent and mutant strains, where average CFUs recovered represents the average number of colony forming units recovered from lysed CaCo-2 cells after bacteria were allowed to adhere and invade CaCo-2 monolayers (MacCallum *et al.*, 2006). The experiment was repeated in triplicate.

Graph to show the recovery of bacteria after an invasion assay with CaCo-2 cells



**Figure 6.24 Coomassie blue stained SDS-PAGE of outer membrane preparations from the parent and mutant strains.** Samples were run on a 15% (v/v) SDS-PAGE gel. The sample in each lane and the marker sizes are indicated on the diagram.

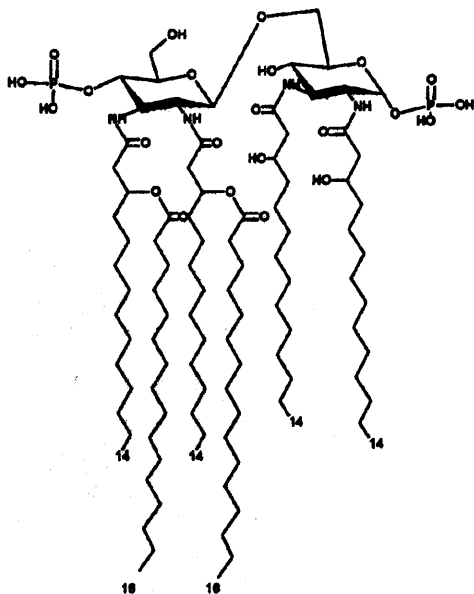


**Figure 6.25: CE-MS analysis of LOS from the *wlaP*-SC mutant.** As predicted the outer core 'GM1' like epitope has been deleted. CE-MS was performed by Jianjuin Li at the Institute of Biological Sciences, National Research Council of Canada as previously described (St. Michael *et al.*, 2002). The ESI-MS spectrum of WlaP-WlaSC LOS (Fig. 2a) showed triply charged ions at  $m/z$  1242.6, 1283.3 and 1310.1, together with their corresponding quadruply charged ions at  $m/z$  931.6, 962.3, and 982.4, respectively. The compositions for each glycoforms were derived from the mass spectrum as Neu5Ac<sub>1</sub>•Hex<sub>3</sub>•Hep<sub>2</sub>•KDO<sub>2</sub>•lipid A, Neu5Ac<sub>1</sub>•Hex<sub>3</sub>•Hep<sub>2</sub>•PEtn<sub>1</sub>•KDO<sub>2</sub>•lipidA and HexNAc<sub>1</sub>•Neu5Ac<sub>1</sub>•Hex<sub>3</sub>•Hep<sub>2</sub>•PEtn<sub>1</sub>•KDO<sub>2</sub>•lipid A, respectively. MS/MS spectrum of triply charged ion at  $m/z$  1242.6 is presented in Figure 2b. The spectrum shows a fragment ion at  $m/z$  290.2 indicating the existence of Neu5Ac. Other major fragment ions are derived from lipid A portion, such as doubly charged ion at  $m/z$  1001.0 and singly charged ions at  $m/z$  1922.4 and 1879.3, corresponding to the consecutive loss of PPEtn and two PEtn residues, respectively. The ion at  $m/z$  1212.4 corresponds to the composition Hex<sub>3</sub>•Hep<sub>2</sub>•PEtn<sub>1</sub>•KDO<sub>1</sub>. The ESI-MS analysis of O-deacylated LOS obtained from *C. jejuni* WlaP-WlaSC is shown in Figure 2c. The mass spectrum indicated a major triply charged ion at  $m/z$  1042.5 corresponding to the composition Neu5Ac<sub>1</sub>•Hex<sub>3</sub>•Hep<sub>2</sub>•KDO<sub>2</sub>•lipid A-OH. In addition, a minor ion corresponding to glycoform containing additional HexNAc could be identified at  $m/z$  1110.0. The product ion spectrum obtained from the triply charged ion at  $m/z$  1042.5 (Fig. 2d) showed a singly charged fragment ion at  $m/z$  1402.0, which revealed that the lipid A-OH was consisted of disaccharide (GlcN3N, GlcN3N), to which two *P* and four N-linked 14:0(3-OH) fatty acids were attached. The diagnostic ion of sialic acid was observed at  $m/z$  290.1. The fragment ion at  $m/z$  1212.4 indicating the core region with the composition Hex<sub>3</sub>•Hep<sub>2</sub>•PEtn<sub>1</sub>•KDO<sub>1</sub>.

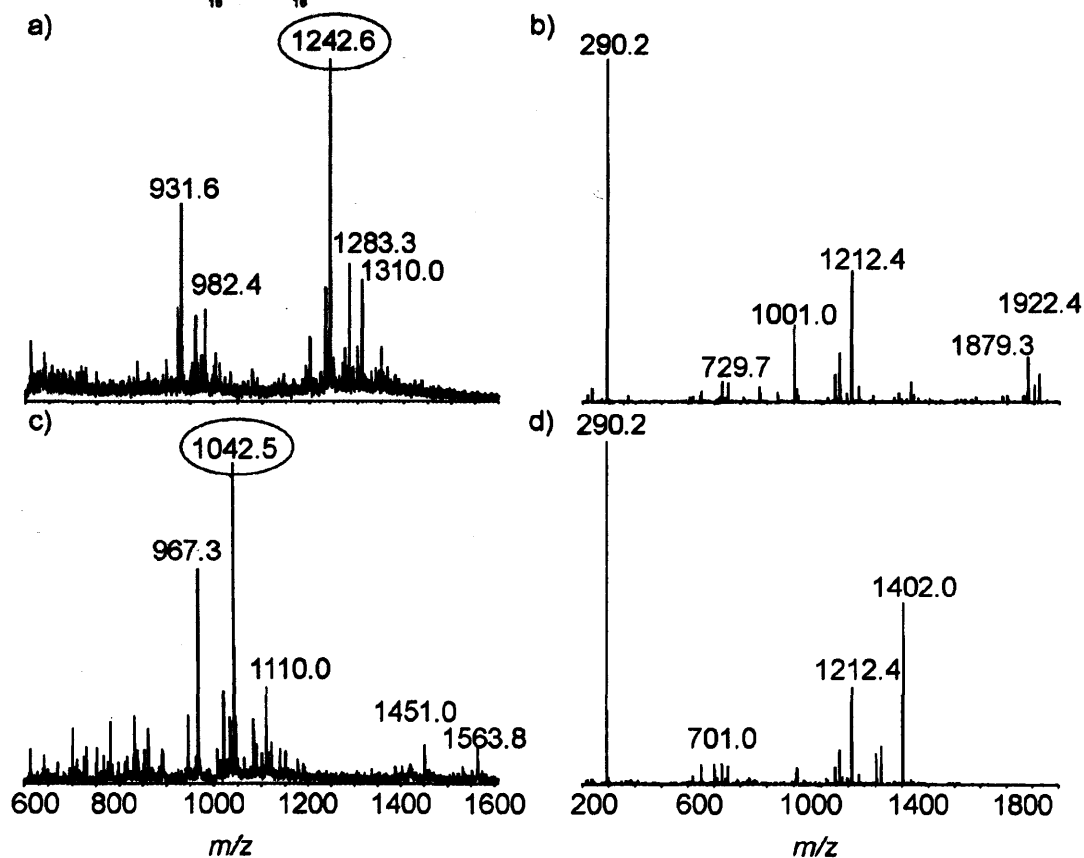
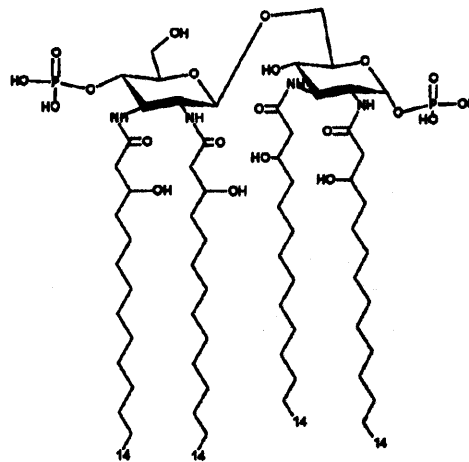
C. j. WlaP-wlaSC

Core: HexNAc<sub>0</sub>, NeuAc<sub>1</sub> Hex<sub>3</sub> Hep<sub>2</sub> KDO<sub>2</sub> PEA<sub>1</sub>

Lipid A



Lipid A-OH



a) Mass spectrum of intact LOS

b) Product ion spectrum of ion at  $m/z$  1242.2

c) Mass spectrum of O-deacylated LOS

d) Product ion spectrum of ion at  $m/z$  1042.5



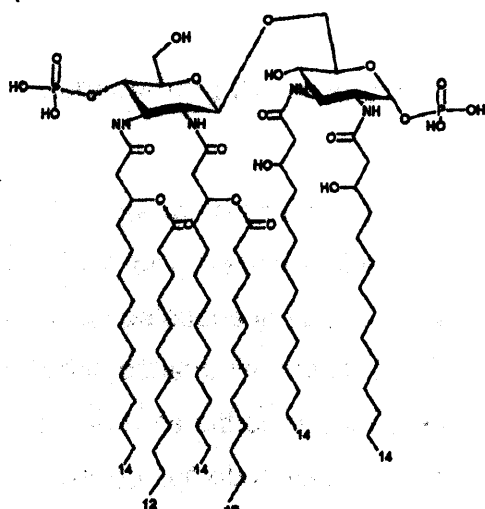
**Figure 6.26: CE-MS analysis of LOS from the *wlaA-T* mutant.** CE-MS was performed by Jianjuin Li at the Institute of Biological Sciences, National Research Council of Canada as previously described (St. Michael *et al.*, 2002). Figure 1a shows the mass spectrum obtained from *C. jejuni* strain WlaA-WlaT. The mass spectrum is dominated by molecular ion corresponding to triply ( $[M-3H]^{3-}$ ) and doubly ( $[M-2H]^{2-}$ ) charged ions. The most abundant ion at  $m/z$  1102.2 represents glycoform with the composition of  $KDO_2 \bullet$ lipid A. The ion at  $m/z$  1163.7 corresponds to the same glycoform but with an additional *PEtn* moiety. The extracted MS/MS spectrum of ion at  $m/z$  1102.2 is shown in Fig. 1b. The spectrum shown major ions at  $m/z$  882.1 ( $[M-2H]^{2-}$ ) and 1765.1 ( $[M-H]^-$ ), providing information on lipid A compositions. For this strain, lipid A portion is composed of a disaccharide  $\alpha$ -D-GlcN3N-(1 $\rightarrow$ 6)- $\alpha$ -D-GlcN3N carrying two phosphate groups at position O-1 and O-4' and substituted by six fatty acid chains (four *N*-linked 14:0(3-OH) linked in 2', 2, 3 and 3' positions and two O-linked 12:0 linked to the 14:0(3-OH) acyl chain in 2' and 3' positions). The fragment ions at  $m/z$  237.1 and 439.1 give evidence for the existence of KDO and  $KDO \bullet KDO$ . The ESI-MS spectrum of O-deacylated LOS (Fig. 1c) showed doubly charged ions at  $m/z$  921.1, 808.5 and 695.8. The mass difference of 226 Da indicated an incomplete lipid A-OH resulted from missing 14:0(3-OH) fatty acid. The product ion spectrum obtained from the ion at  $m/z$  921.1 is shown in Figure 1d. The observation of fragment ions at  $m/z$  219.12 and 439.1 suggests the presence of KDO and  $KDO \bullet KDO$ . The doubly charged fragment ion at  $m/z$  701.0 revealed that lipid A-OH, resulting from the cleavage between the KDO-lipid A bond under collision-induced dissociation, consisted of a  $\beta$ -(1 $\rightarrow$ 6)-linked disaccharide of *N*-acylated 14:0(3-OH) glucosamine residues, each residue being substituted with a phosphate group.



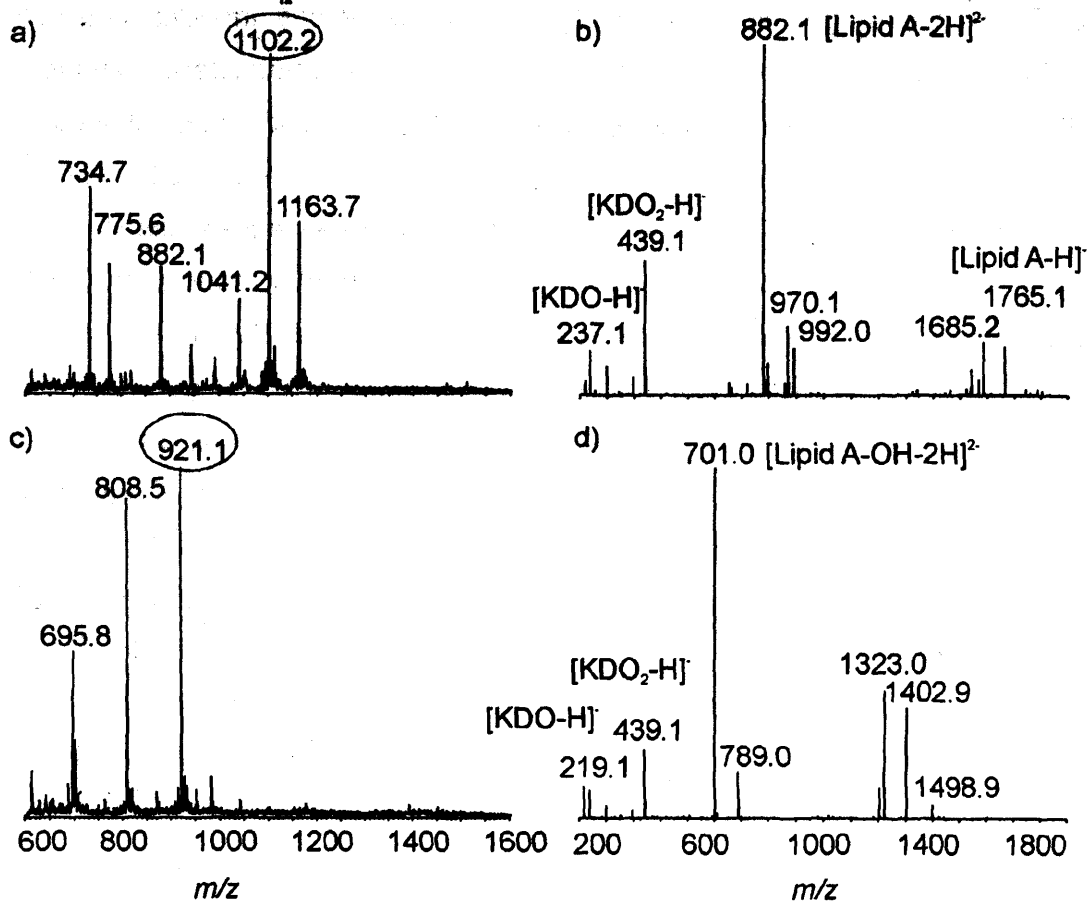
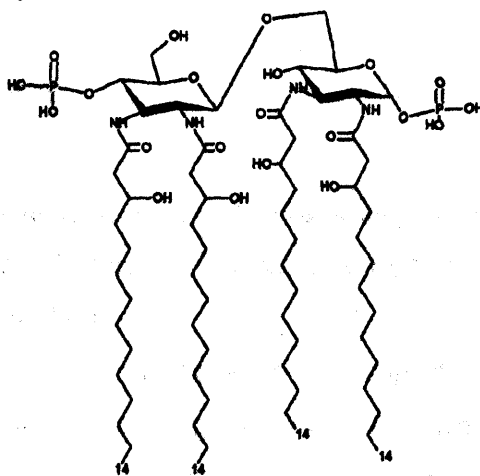
C. j. WaaA-wlaT

Core: KDO-KDO

Lipid A



Lipid A-OH



- a) Mass spectrum of intact LOS  
 b) Product ion spectrum of ion at  $m/z$  1102.2  
 c) Mass spectrum of O-deacylated LOS  
 d) Product ion spectrum of ion at  $m/z$  921.1

## 6.9 Discussion

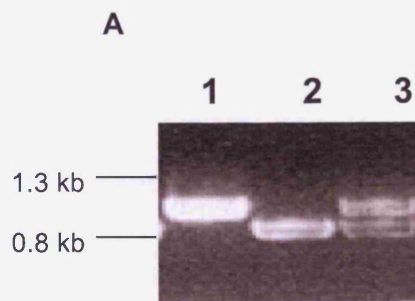
### 6.9.1 Establishing the mutagenesis strategy

The Hendrixson *CAT-rpsL*<sup>WT</sup> mutagenesis strategy had previously been established for mutating single genes (2001). In this work it has been used to make large deletions in sizeable transcriptional units.

The initial constructs made for this work also incorporated the imperfect inverted repeats either side of the protein glycosylation gene, *wlaJ*. This was due to an earlier observation in this lab (Millar, 2003) where the *wlaJ* gene was both present and absent in a single colony (see Figure 6.27a). This event was postulated to be due to the recombination of the imperfect inverted repeats located either side of the *wlaJ* gene as detailed in Figure 6.27b (Misawa *et al.*, 2001).

These sequences were incorporated into the *CAT-rpsL*<sup>WT</sup> cassette by PCR and cloned with the cassette in various orientations, reflecting their positions in the NCTC 11168 genome and alternative positions. This allowed one less step to be performed in the mutagenesis process (Figure 6.27c). Multiple rounds of replica plating and screening by colony PCR did not produce any chloramphenicol sensitive, streptomycin resistant colonies as the recombination event did not occur at a detectable frequency in this context. Problems were also encountered when a spontaneous mutation occurred in *rpsL*<sup>WT</sup> of the *CAT-rpsL*<sup>WT</sup> cassette giving rise to streptomycin resistance meaning mutants could only be selected on the basis of the loss of chloramphenicol resistance.

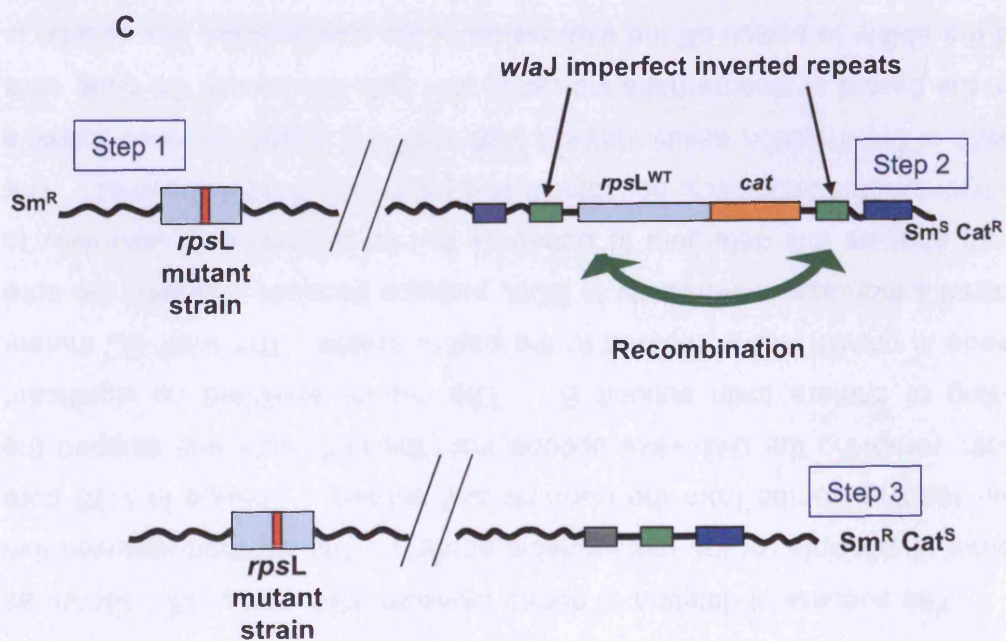
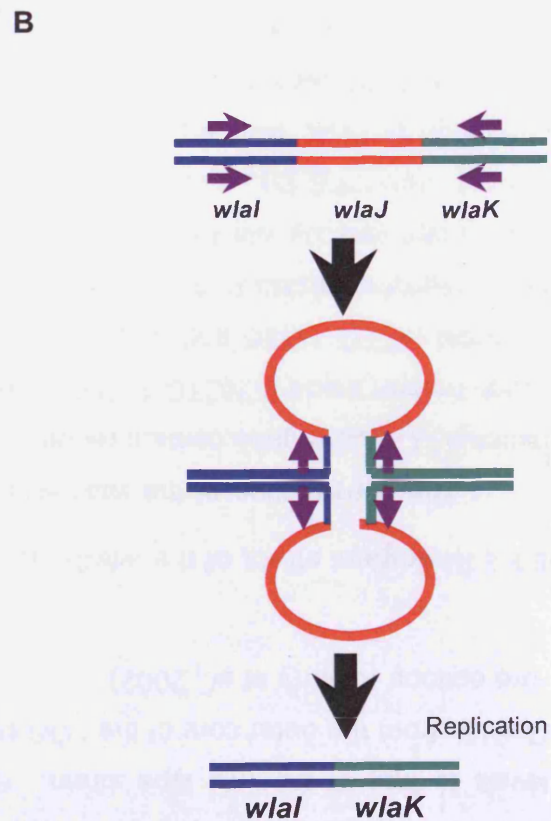
Figure 6.27 A) Gel photograph showing the presence and absence of *wlaJ* in a single colony. B) Schematic diagram postulating the mechanism of recombination which results in the loss of *wlaJ*. C) Schematic diagram showing the steps involved in the mutagenesis strategy with the imperfect inverted repeats. (Misawa *et al.*, 2001; Wood *et al.*, 1999)



Lane 1: *wlaJ* present

Lane 2: *wlaJ* absent

Lane 3: *wlaJ* absent & present



### 6.9.2 Biological effect of the *wlaP-wlaSC* deletion mutant

The successful deletion of genes between *wlaP* and *wlaSC* served as a proof of principle for the mutagenesis strategy. The deletion removed five open reading frames from the genome and caused a change in LOS core length, removing the GM<sub>1</sub>-like epitope from the LOS core and stopped the binding of cholera toxin subunit B. The mutant exhibited no significant change in growth rate compared to the parent strains. The *wlaP-SC* mutant showed a increase in sensitivity to SDS, perhaps because changing the core length enables this detergent to penetrate but no difference in sensitivity to the hydrophobic antibiotics, novobiocin and polymyxin B was observed. The results of the invasion assay indicate that *wlaP-SC* mutant is more invasive than the parent strains perhaps indicating that both the loss of the outer core and the ability to switch off the expression of the gangliosides like epitope is advantageous in the invasion process as seen by Guerry *et al.* (2002). Experiments with this 81-176 *cgtA* mutant showed no change in adherence but a significant increase in invasion with INT407 monolayers and an increase in serum resistance. Complementation of the mutation returned invasion levels to that of the wild type strain. Mutation of *cgtA* caused the loss of GalNAc from the outer core of the LOS structure which then resembles a GM<sub>3</sub>-like epitope (Guerry *et al.*, 2002).

### 6.9.3 Biological effect of the *wlaA-wlaT* deletion mutant

The construction of the *wlaA-wlaT* large deletion mutant confirms the minimal LOS core gene content required for viability. This is also the deepest rough mutant made in NCTC 11168. Attempts to directly mutate streptomycin resistant NCTC 11168 and delete the genes between *wlaA* and *wlaT* led to gross rearrangements in the LOS core biosynthesis locus. The deletion was then attempted by electroporating the plasmid into the *wlaP-SC* deletion mutant producing just one colony. The deletion of 16.9 Kb of DNA was confirmed by PCR, sequencing and Southern blotting.

No LOS core could be seen on tricine SDS-PAGE by silver staining and the loss of cholera toxin binding was also seen by western blotting. The predicted structure of the LOS of this mutant was expected to be the same as

that seen by Kanipes *et al.* (2006) where only the Kdo was present with no proximal sugars. The core structure of the *wlaA-T* mutant was shown to contain 2 Kdo moieties and no proximal sugars. The difference in results observed by the two studies could be as a result of the mutations being made in different strains of *C. jejuni*. In 81-176, the presence of one Kdo and the loss of the 3-O-methyl group from the capsule implies that this capsular change allows the strain to remain viable. The capsule biosynthesis region of 81-176 is smaller than that in NCTC 11168 but may express four different structures, whereas NCTC 11168 only has one potential capsule structure as detailed in Figure 6.28 (Karlýshev *et al.*, 2005). The affect of the LOS core deletion in the *wlaA-T* deletion strain on the capsule will have to be explored and may provide information about the interaction between these polysaccharides.

The *wlaA-T* large deletion mutant was more sensitive to detergents and hydrophobic antibiotics as with other deep rough mutants (Steeghs *et al.*, 1998). The large deletion strain also showed an altered growth pattern and sensitivity to the antibiotic trimethoprim, which inhibits outer membrane synthesis (Painter *et al.*, 1988; Tzeng *et al.*, 2002; Vorachek-Warren *et al.*, 2002; Møller *et al.*, 2003; Tan and Darby 2005). Like this mutant, there was a considerable increase in sensitivity to novobiocin and SDS with a *waaF* and *waaC* mutant suggesting the importance of this gene in the functioning of the outer membrane (Kanipes *et al.*, 2004; Kanipes *et al.*, 2006). It was also noted in the later paper that the *waaC* mutant was more sensitive to SDS and antibiotics than the *waaF* mutant, and this may be due to the depth of the mutation.

Analysis of the 81-176 *waaC* mutant by transmission electron microscopy also showed that the cell morphology had changed (Kanipes *et al.*, 2006) and the cell morphology of the *wlaA-T* deletion mutant should also be examined.

Comparing the outer membrane protein preparations of the parent and mutant strains shows that the alterations in the membrane of the *wlaA-T* mutant may have caused a change in the protein content of the membrane or the membrane instability may have caused the loss of membrane-associated proteins in the OMP extraction process.

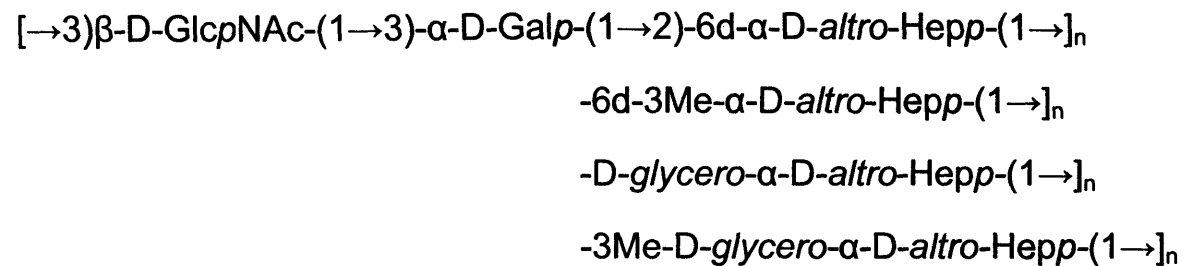
No *wlaA-T* large deletion mutant cells were recovered after lysing of the Caco-2 cells in the invasion assay, indicating that this mutant is unable to invade and that as suggested previously the LOS core oligosaccharide is important in invasion (McSweeney *et al.*, 1986; Oldfield *et al.*, 2002; Kanipes *et al.*, 2004; Kanipes *et al.*, 2006). The process of adhesion and invasion is not well understood yet now a role for LOS in the process can be confirmed, although further *in vivo* assays need to be performed to confirm the role of the LOS core in virulence. Despite the possibility of the tissue culture media having a detrimental effect on the invasive potential of the *wlaA-T* mutant, the results of assays with this mutant previous studies support a role for LOS in invasion. The deletion of *waaF* in NCTC 11168 and 81-176 caused a decrease in invasion *in vitro*. A decrease in invasive potential was also seen with a 81-176 *waaC* mutant indicating a role for the LOS core in invasion.

The inability of this large deletion strain to take up DNA through natural transformation may be because altering the outer membrane structure has also altered the competence of the bacterium. Experiments need to be conducted to see if this strain is able to take up DNA via electroporation.

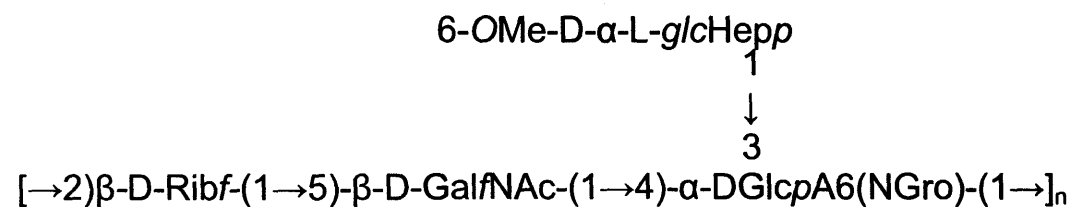
**Figure 6.28 Schematic representation of the *C. jejuni* 81-176 and NCTC 11168 capsule structure;** adapted from Karlyshev *et al.*(2005), where Altro, altrose; Gro, glycerol; Hep, heptose; Gal, galactose; GalNac, N-acetylgalactosamine; Glc, glucuronic acid; Me, methyl group.



**81-176 (HS:23/36)**



**NCTC 11168 (HS:2)**



#### **6.9.4 Conclusion**

The creation of the *wlaP*-SC deletion mutant established the mutagenesis strategy for the creation of large deletion mutants. The ability to create a *wlaA-wlaT* deletion mutant confirmed the minimum LOS gene content required for viability. This is also the deepest rough mutant created in *C. jejuni* NCTC 11168. The mutant can be used to examine the role of LOS in virulence and the interaction of LOS with other polysaccharide loci.

The inability of this strain to take up DNA by natural transformation implies that it is 'locked' and therefore maybe a vaccine candidate. This topic shall be discussed in Chapter 7.

## Chapter 7: General Discussion

*C. jejuni* remains an important gastrointestinal pathogen causing severe gastroenteritis, which exerts a major social and economic burden worldwide. Campylobacteriosis is also associated with rare neuropathological complications such as GBS and MFS as the most frequent antecedent infection. LOS has a role in the development of these neuropathies as a molecular mimic of host gangliosides.

The sequencing of the NCTC 11168 genome highlighted the importance of the glycome for this bacterium as a total of eight percent of its relatively small genome is involved in the production of polysaccharides. The important role of LOS as a constituent of the outer membrane is well established, as is an association with the development of GBS due to molecular mimicry, but the contribution of LOS to campylobacteriosis remains to be clearly established.

The current study aimed to address various aspects of the role of LOS in pathogenesis by establishing the amount of diversity in the LOS gene content of a clinical strain set and determining the gene content of any unclassified strains by several molecular methods. The importance of the core LOS molecule was demonstrated by deletion of the LOS core biosynthesis cluster, which established that *C. jejuni* was viable without the majority of the LOS core.

## **7.1 Diversity in LOS biosynthesis gene cluster**

### **7.1.1 Variation in LOS core biosynthesis gene content**

The work undertaken here forms the fifth study to examine the extent of gene content variation in the LOS biosynthesis gene cluster (Oldfield, 2000; Gilbert *et al.*, 2000; Gilbert *et al.*, 2002; Millar, 2003; Parker *et al.*, 2005). The gene content variation originally classified on a historical basis was reorganised into an organisationally based grouping system (Karlyshev *et al.*, 2005). The information gained in the previous gene content studies were used as a basis for the development of a PCR based screen which was used to classify 50 clinical strains on the basis of LOS cluster gene content. From the initial PCR screen, twelve strains remained unclassified and four strains produced amplicons indicating they may have belonged to more than one sub-group. These sixteen strains were subjected to further examination by both phenotypic and molecular methods. In this current work, a PCR based screen was successfully developed to enable the determination of LOS group type.

As with previous studies, the results from this PCR screen infer that the majority of strains can be categorised as group 1. Unlike the previous studies by Millar (2003) and Parker *et al.* (2005), the percentage of strains belonging to this group are much lower; 36% compared to ~ 60% and 64% respectively. This present study found that 24% of the strains examined could not be assigned a group or sub-group. The results of the study by Parker *et al.* (2005) are in agreement with this finding, with 22% of their strains being unclassified. Unlike the previous two studies, the current studies also found potential novel gene combinations within the LOS core gene locus. The previous studies drew correlations with environmental niche but the strains used in the current study were all clinical isolates related to gastroenteritis from the UK.

Long range PCR amplicons proved invaluable as a source of sequencing template, but also for comparing the size of the product against the size of known sub-group strains. Sequencing and annotation showed that strains that may have a particular group gene content may differ in their gene

expression due to missense mutations and insertion or deletions of bases, and hence their LOS structure. The inactivation of *orfA* (which has homology to *wlaNA* from NCTC 11168) by a single base insertion is seen in some group 3A strains, although the effect of this gene inactivation remains unclear as the LOS structure for this group is unknown (Gilbert *et al.*, 2005). This glucosyltransferase gene, *orfA* is also subject to non-phase variable frameshift mutations which lead to the expression of one or two-domain glucosyltransferases.

Sequencing analysis has also revealed the presence of homopolymeric tracts, which cause variation by slip strand mispairing in many of the strains. Nine ORFs in the LOS core biosynthesis cluster have been reported to contain phase variable poly G tracts including; *cgtA*, *wlaN*, *cstII*, *orf11*, *Cj1144/45*, *orfE*, *wlaUD* and *wlaUF* and there is also one example of a poly A tract in the *cstII* gene of the MF6 LOS gene cluster (Gilbert *et al.*, 2002; 2005). Phase variation is often associated with virulence genes in bacteria such as *H. influenzae*, *H. pylori* and *Neisseria* spp. and in *C. jejuni* homopolymeric tracts are commonly associated with the genes involved in carbohydrate biosynthesis.

Sequencing has also confirmed the potential for gene content variation in group 2, which previously consisted of three sub-groups, shows further differentiation by the insertion and deletion of genes. There have been various reports of gene deletions occurring in group 2 strains. The strain O:3 was reported to be missing the gene *wlaUJ* (Millar, 2003). There has only been one observation of insertion and deletion events occurring in this group as highlighted by the strains ATCC 43431 where genes *wlaUG* and *wlaUI* have been deleted, and a gene encoding a butyryltransferase has been inserted in between *wlaUF* and *wlaUH* (Parker *et al.*, 2005). Variation in this group appears to occur between genes *wlaUF* and *wlaVB* indicating that the deletion of these genes in strain 50702 may be a possibility. The current study has highlighted that group 2 LOS gene content varies to a much greater extent than previously discovered.

Evidence of recombination and lateral gene transfer has previously been seen amongst the group 1 LOS type members, and further sequencing of the incomplete strains may reveal them as intermediaries in this process.

Gilbert *et al.* (2002; 2005) observed that based on sequence identity group 1B strains evolved from group 1A strains with strains such as ATCC 43438 representing an evolutionary intermediate. The events that occurred to produce group 1B strains involved the duplication of *cgtA*, lateral transfer to alter the sequence from *cgtA*I to *cgtB*, and finally the insertion of a 'C' terminus cassette in this region which leads to a large number of amino acid substitutions. Group 1B has also been proposed as the evolutionary intermediary between group 1A and group 1C strains, as group 1C strains also contain the duplicated *cgtA*II gene as a gene fusion with *neuA*. Other recombination events have occurred to cause the deletion of *cgtA*I and the insertion of *wlaO* and *wlaP* in its place, as well as the insertion of *wlaSB* between *neuA* and *waaV* (Gilbert *et al.*, 2005).

This study also highlighted that possible recombination occurs between polysaccharide loci, as two strains contained capsule genes in their LOS cluster. It should also be highlighted that these clinical strains were isolated in different parts of the UK and 5 months apart, although no detail about the clinical presentation of the diseases caused by these strains are available. The occurrence of capsule genes within the LOS cluster further deepens the notion of interaction between the capsule and LOS polysaccharides.

The information available from previous studies also enabled the development of a locus specific microarray. Microarrays are inherently limited by their inability to inform about changes in gene content, whether that is gene rearrangements, gene insertions or deletions, chromosomal location and SNPs. It was therefore postulated that if oligonucleotides were designed to the boundaries of each ORF, changes in the gene content may be detected.

The ability to validate both the first and second generation microarrays was limited by detection issues and by the small set of strains used. The restricted design also influenced oligonucleotide specificity. PCR amplification prior to hybridisation has been used successfully with locus-specific microarrays (Seregeev *et al.*, 2004; Garrido *et al.*, 2006). Seregeev *et al.* (2004) examined various PCR amplification techniques and found that using degenerate primers based on conserved regions provided the best method to limit false positive or false negative results. Garrido *et al.* (2006)

found that long range PCR stopped the signal biases of normal PCR and multiplex PCR amplification as all the genes had the same copy number.

Greater specificity may be gained from this approach by using more specific template for the hybridisation, such as long range PCR amplicons and by developing a group-based tiling microarray to encompass the known sequences for this locus. On a whole genome scale tiling microarray technology involves the design of overlapping oligonucleotides and can be used for sequencing by hybridisation and single nucleotide polymorphism (SNP) detection (Pandya *et al.*, 2006; Jackson *et al.*, 2006). Jackson *et al.* (2006) used *E.coli* O157:H7 tiling arrays to identify more than 150 SNPs as well as deletions and changes in gene copy number. This sequence variation was confirmed by sequencing and PCR, but the tiling array proved to be useful in determining allelic diversity. Two studies have been performed to test tiling arrays in a more focused application as an alternative approach to MLST with varying success. van Leeuwen *et al.* (2003) examined the use of tiling arrays for MLST of *Staphylococcus aureus* finding the results to be reproducible and concordant with conventional MLST. Although in the application of this technique to *Neisseria meningitidis*, Swiderek *et al.* (2005) found that the high density of polymorphic sites and the changing GC content of this bacterium led to many false positive results.

Characterisation of the LOS gene cluster provides more and more evidence for diversity in this region and confirms that this region is a hotspot for genetic exchange and rearrangements. The extent of variability in the LOS cluster of this bacterium is surprising as other bacteria, such as *Neisseria* spp. are limited in the variation of their LOS genes. The LOS grouping system has been extended to include the further diversification observed in this current study and from sequences in the public databases,

### **7.1.2 Role of LOS in *C. jejuni* pathogenesis**

The ability of this bacterium to vary its LOS gene content and therefore its LOS structure leads to the conclusion that LOS and its variability has an important role in *C. jejuni* pathogenesis. The presence of sialic acid in the LOS core has been linked to the development of GBS and MFS (Moran *et al.*,

1996), as well as providing protection against the host's complement and immune system (Vogel and Frosch 1999). A recent study of GBS and MFS associated strains showed that unsialyated strains can also play a role in the aetiology of these neuropathies (Godschalk *et al.*, 2007). This previous study also showed that strains harbouring the same LOS group gene content may express different LOS structures. Five different LOS structures were found within thirteen sub-group 1A strains and in the strain set of 26 strains, 14 strains expressed a mixture of at least 2 different core structures. The information gained from this current study only serves to highlight that knowledge of LOS gene content can inform but not predict the final LOS structure produced. This is because SNPs and other mechanisms of variation influence and alter gene expression. The extent of the influence imposed by the capsule polysaccharide should also be examined as it may alter LOS structure. Strains with novel gene content should be examined by CE-MS to enable any further correlations to be drawn between LOS gene content and LOS structure.

Further studies need to be conducted to see if a link can be established between the different LOS structures, inflammation, invasion and clinical presentation. The strains used in this current study were UK gastroenteritis isolates but no conclusions can be drawn about LOS group and clinical presentation as no specific information is available for each isolate. Studies have observed strains expressing mixed LOS core structures and the mechanisms which cause the mixed core structures to be expressed need to be explored. A further examination of what core structure is expressed at each stage of infection should be performed. It also remains to be determined if a strain from other groups also express multiple LOS core structures. In *Y.pestis*, LOS modification has been shown to under the control of the PhoPQ regulatory system (Hitchen *et al.*, 2002) and in *S. enterica* serovar Typhimurium PmrA and PmrB determinants have been shown to be responsible for remodelling LPS which is necessary for resistance to ferric iron and polymyxin B in this bacterium (Nishino *et al.*, 2006).



### 7.1.3 LOS cluster gene content and typing methods.

*C.jejuni* has a weakly clonal population and is subject to genetic recombination. The LOS core gene cluster has a lower GC content than the genome average indicating frequent genetic exchange and recombination at this locus. MLST was used to investigate the genetic background of each strain. Together with the results of Miller (2003) correlations can be drawn between LOS group and clonal complex. The majority of strains for each clonal complex belong to one LOS group. In order to view the distribution of the different LOS biosynthesis gene clusters from an epidemiological context, MLST typing was performed on the majority of the isolates. The STs that have been identified thus far have been grouped into clonal-complexes of genetically related strains. A total of twenty percent of the isolates typed belonged to the ST21 complex, which was the largest clonal complex identified from 800 isolates (Dingle *et al.*, 2002). Perhaps due to the large nature of this complex, it is also the most genetically heterogeneous identified to date as it is composed of many sequence types. In agreement with the MLST results of Millar (2003), the majority of group 1C strains appear to cluster in this complex. This could indicate that either group 1C gene content is only acquired through recombination with strains from the ST21 complex, or LOS structures synthesised from the genes in the group1C loci provides a selective advantages for colonisation of the niches inhabited by the ST21 strains. The clonal complex ST21 is the most common complex associated with human gastroenteritis and is also associated with GBS and MFS antecedent strains (Dingle *et al.*, 2001; Dingle *et al.*, 2002). The second most commonly associated human isolate clonal complex is ST45 and 20% of the CSSS strains belong to this ST complex. There are five strains from the ST48 complex in this current study and they all contain genes relating to the group 1B LOS gene content type.

Complexes ST21, ST48 and ST206 are thought to form a supercomplex of closely related genotypes that undergo frequent horizontal exchange (Dingle *et al.*, 2002). The strains in this supercomplex exhibit a high level of antigenic diversity in relation to their Penner serotype and

variation at the *flaA* short variable region. The results of the present study infer that either this is not the case at the LOS gene cluster or that exchange at the LOS cluster is disadvantageous. There are no strains representing the ST206 complex, but the majority of strains from the other two complexes contain group 1 LOS gene content. To determine the relationship between LOS gene content and ST complex, an extended strain collection needs to be examined. Analysis of LOS group alongside MLST may provide further discriminatory power to this molecular technique.

Examination of LOS group and serotype showed a loose association by LOS group as each serotype was only associated with strains from 2 groups. This may be due to the small strain set or their limited origin and an extended strain set would need to be examined to establish this correlation. It may also indicate that although the capsule is the main sero-determinant in serotyping, LOS core structure does also influence this typing method and selection is on the association of capsule polysaccharide and LOS or on the compatibility of polysaccharide structures.

## **7.2 Minimal LOS gene content for viability**

### **7.2.1 Deletion of *wlaP-wlaSC***

The construction of the 11168:: *wlaP-SC* strain served two purposes. One aspect of this was to show that the mutagenesis strategy could be used to delete multiple open reading frames as the technique had only been applied to the mutation of single genes. Initially plasmids were constructed using the imperfect inverted repeats that are postulated to cause the looping out of *wlaJ*, in the hopes that they would cause the removal of the antibiotic cassette. The inverted repeats were incorporated into the antibiotic cassettes by PCR in various combinations, but despite this the recombination event did not take place at a detectable level. The frequency of the recombination event was attempted in the mutation of single open reading frames (the previously established *waaF* mutation) and also in the multiple ORF context. Experiments were begun to establish the frequency of the recombination

events by cloning the cassette in to the *C. jejuni* plasmid pMW10 (Wösten *et al.*, 1998). This plasmid is a promoterless *lacZ* shuttle vector and as such loss of chloramphenicol resistance and  $\beta$ -galactosidase activity in *C. jejuni* 480 would indicate that a recombination event had taken place. Although initial colony PCR established that recombination was occurring (data not shown) and further studies would have to be performed to form any definitive conclusions.

The construction of the *wlaP*-SC deletion mutant removed five open reading frames and caused a change in LOS core length. This strain was shown to be more invasive *in vitro* which confirmed that the ability to switch the expression of the GM1-like epitope on and off using homopolymeric tracts is advantageous in this bacterium, as seen previously by Guerry *et al.* (2002).

### 7.2.2 Deletion of *wlaA-wlaT*

The construction of the 11168::*wlaA-T* large deletion mutant confirmed the minimum LOS gene content required for the viability of *C. jejuni* cells and is the deepest rough mutant constructed to date, removing the LOS core from the bacterium.

Differences were seen in the CE-MS analysis of the LOS between this mutant and a previously constructed *waaC* mutant in 81-176, where one Kdo was observed in the core (Kanipes *et al.*, 2006). Two Kdo moieties were observed in the core structure of the *wlaA-T* large deletion mutant and this difference implies variation in results due to different strains being used. One explanation of this may be due to the differences in the capsule polysaccharide of these strains, suggesting that the capsule exerts a more protective effect on the 81-176 *waaC* mutant. The capsule polysaccharide of the *waaC* mutant lacked a 3-O-methyl group changing the immunoreactivity of the capsule leading to conclusions that mutating *waaC* affects the regulation of capsule biosynthesis genes.

Another potential explanation for the differences observed between these mutant strains may be due to differences in the enzyme which adds the

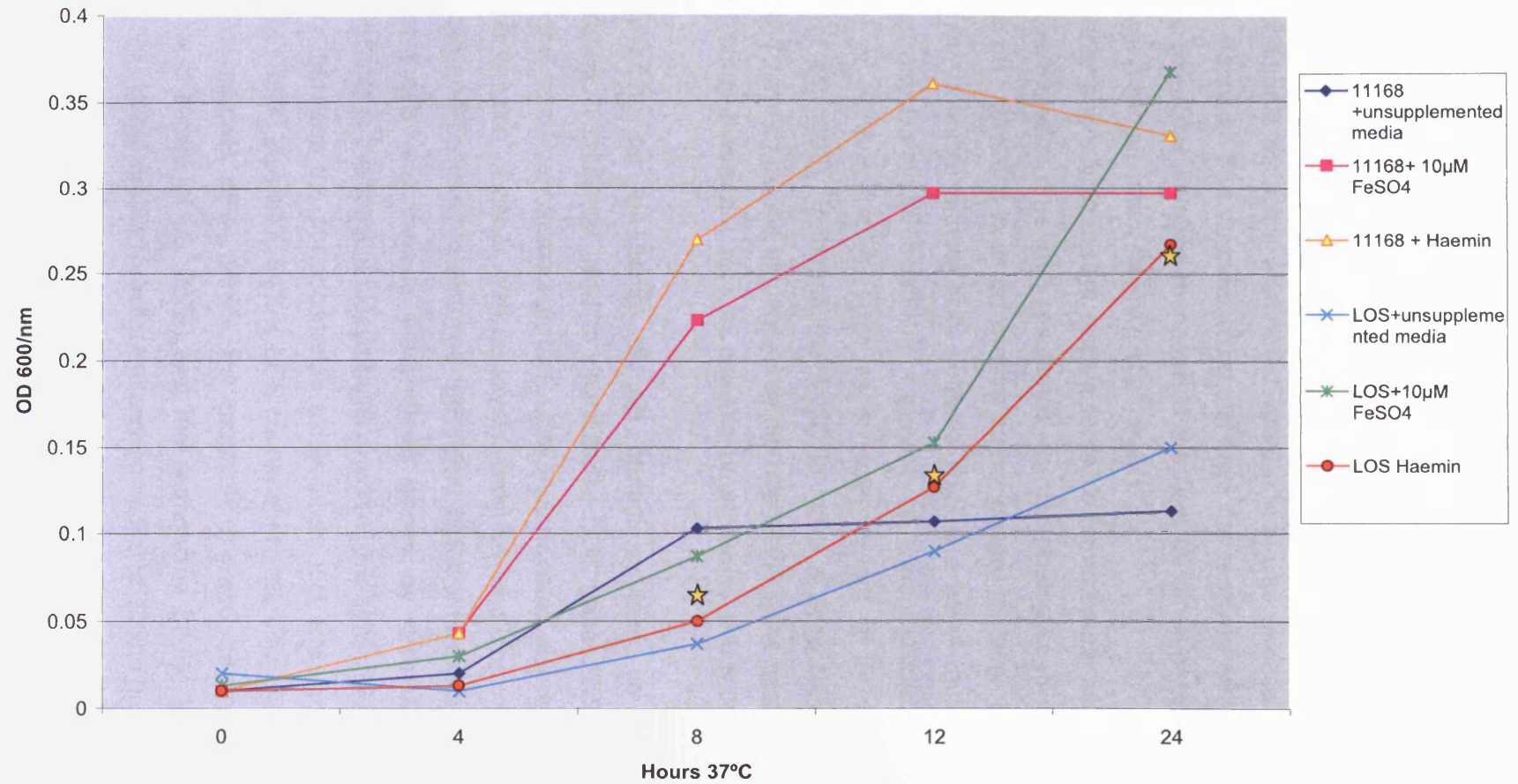
Kdo, KdtA. The KdtA homologues from 81-176 and NCTC 11168 share 96% amino acid identity, although they may have differing enzyme activities.

The extreme outer membrane sensitivity of the *wlaA-T* mutant to antibiotics and detergents confirmed the importance of LOS as a barrier to these toxic substances. Outer membrane function was also seen to be compromised after outer membrane preparations were examined by SDS-PAGE. Further confirmation of an impairment in membrane function was seen with growth assays in minimal media (MEM $\alpha$ ) to assess the bacterial uptake of iron (in the form of haemin; performed by C. Miller, University of Leicester.) Figure 7.1 shows the differences between growth of wild type NCTC 11168 and the *wlaA-T* deletion mutant in iron replete, iron limited and unsupplemented media. There are significant differences between the mutant and wild type growth rate in minimal media and in the ability to take up iron in any form, suggesting that deleting the LOS core has had a significant effect on the iron acquisition systems and potentially all systems that utilise outer membrane receptors. Further investigation at a transcriptome and proteome level would serve to confirm these observations.

The inability to recover any cells of the *wlaA-T* large deletion mutant after invasion assays with CaCo-2 cells also serves to confirm the role of the LOS core in the processes of adherence and invasion as noted previously (McSweeney *et al.*, 1986; Oldfield *et al.*, 2002). Initially invasion assays need to be performed to assess the affect of the tissue culture media on this large deletion mutant, although comparisons with *waaC* and *waaF* mutants argue against this being the cause of no bacterial cells being recovered (Oldfield *et al.*, 2002; Kanipes *et al.*, 2004; Kanipes *et al.*, 2006). The growth assay performed to assess iron uptake was performed in minimal media and no adverse effects were seen with this media. Comparatively the growth of the mutant in the minimal media was much slower but it still survives. Further investigations with this mutant *in vivo* or *ex vivo* may enable conformation of the nature of the interaction between LOS and the host. Other aspects of this interaction may also be investigated as *C. jejuni* LOS is a known potentiator of the maturation of dendritic cells (Hu *et al.*, 2006).

**Figure 7.1 Graph comparing the growth of NCTC 11168 and the *wlaA-T* large deletion mutant in iron replete and iron limited conditions.** Each measurement was repeated in triplicate by C. Miller. Measurements where the OD600 is significantly different between the 11168:haemin and LOS mutant: haemin samples are indicated by ☆

## 11168 and LOS mutant growth with porcine haemin



Further to these observations altering the outer membrane structure has also altered the competence of the bacterium and it is unable to take up DNA by natural transformation. Experiments need to be conducted to confirm this observation and to see if this strain is able to take up DNA via electroporation.

If the 'locked' nature of this mutant is confirmed it would make a suitable candidate for a live attenuated vaccine strain as genetic exchange has been stopped. The vaccine would still have the required antigenicity to produce a protective immune response in the host as other polysaccharide loci and lipid A remain intact (Meinke *et al.*, 2004).

Further phenotypic assays remain to be performed such as the effect of this large mutation on serum resistance, cell morphology and motility. It would also be interesting to determine if a further deeper mutation is possible and *C. jejuni* is viable without Kdo, such as *yrbH* and *waaA* mutants of *Y. pestis* (Tan and Darby, 2005) and *kdtA* mutants of *N.meningitidis* (Tzeng *et al.*, 2002) or without lipid A, such as an *lpxA* mutant of *N.meningitidis* (Steeghs *et al.*, 1998). Although the ability to make a deep rough mutant without *wlaT* (*gmhB*) maybe influenced by the role of the gene product in not only LOS biosynthesis but also in capsule biosynthesis (Karlyshev *et al.*, 2005a).

Depending on the ability of this mutant to take up DNA by electroporation, it could be used to characterise LOS genes from diverse strains in the same genetic background. The mutant could also be used to confirm the function of genes with putatively assigned functions. Further to this, it could also be used to examine the extent of the interaction between LOS biosynthesis and capsule biosynthesis genes by creating defined deletion mutants in the capsule biosynthesis region of this strain. Evidence of interactions between the polysaccharide biosynthesis loci already exists as the general protein glycosylation and LOS cluster share the bifunctional gene, Cj1129 (Bernatchez *et al.*, 2005), the mutation of *waaC* affects capsule biosynthesis (Kanipes *et al.*, 2006) and the sequencing of the LOS gene cluster of two clinical strains revealed the presence of ORFS with homology to capsule genes in the LOS core biosynthesis loci.

### **7.3 Final Conclusions**

This work has confirmed that LOS has role in *C. jejuni* pathogenesis and in host-cell interactions. Clearly the ability of this pathogen to vary its repertoire of LOS epitopes proves advantageous in campylobacteriosis and in various ecological niches. This work has served to highlight that the extent of variation in the LOS cluster is much greater than at first realised. It has also highlighted another potentially greater source for recombination as capsule biosynthesis genes were found in the LOS biosynthesis cluster of two strains. This work has also served to re-enforce an interaction between the capsule polysaccharide and LOS gene cluster and this relationship requires further examination. The ability to construct a deep rough mutant without the LOS core oligosaccharide also establishes the role of the LOS core in the *C. jejuni* outer membrane and its role in invasion.



## Appendix 1

**Table A1.1 PCR primer and Oligonucleotide sequences**

Restriction sites are underlined and the enzyme shown in parentheses.

### Chapter 3

Primer Name	Nucleotide Sequence 5'-3'	Target gene
<b>Class determination</b>		
Class ABF	CTTTAGCAGCAGCTATTGTTGGAG	<i>neuCI</i>
Class ABR	CCCAAGTGTGATACCCTGTTTAAG	<i>orf 11</i>
Class BIIF	AAATACCGCCGAAACGATTACC	<i>cgtAll</i>
Class BIIR	ACGAGCAGGGATTATTGCTAAG	<i>neuAI</i>
Class CF	TTGGGAGTTAACTGCATAGAAGAG	<i>wlaNB</i>
Class CR	CCTTTATATACCCAGCATGTAAATAGC	<i>wlaP</i>
Class CIIF	AGTATGTTACCTGCCATACAAAGAG	<i>neuAI</i>
Class CIIR	ATCGGAGTAGCAGCGTTTAGTG	<i>Cj1145c</i>
Class DER	CTCCCAAATTTCCAACCTCGTTTATC	<i>wlaUI</i>
Class DFor	GGTGCTCTCGGAAACTATATTCATC	<i>wlaUI</i>
Class DERII	CCAAATTCAATGTGCAACTATCAAG	<i>wlaUJ</i>
ATCC43431F	TAGATAGCGGAAGCACAGATGA	<i>butyltransferase</i>
Class FGF1	GCAGCTAAATATGGTGCGATTGC	<i>orfA</i>
Class FGR1	ACTCCTTGTAGCCGAAGATATGC	<i>orfB</i>
Class GF2	GCGGCGGAGAAAGAGTTTG	<i>orfG</i>
Class HF	GATAGCGGAAGCGCAGATGATAG	<i>orf3</i>
Class HR	AATCAGTAGTGATTATATGGGTGTAATGAG	<i>orf4</i>
<i>gneF</i>	TCCAAAGCCGCTAAATGTGC	<i>gne</i>
<i>gneR</i>	TCTTATTAGCGGTGGTGCAG	<i>gne</i>

**Chapter 4**

Primer Name	Nucleotide Sequence 5'-3'	Target gene
<b>Primer walking - Subclass 1A &amp; 1B</b>		
1134F	TGCCTGAGTGTATCTTGCAATTT	<i>lpxL</i>
1134R	GCTTGCTCCCATTGATAATTC	<i>lpxL</i>
1135R	GCCTTGCTGTATCTTCTTTGCT	<i>wlaNA</i>
wlaNBr	GATCCTACACTTAATGCTCGATTTG	<i>wlaNB</i>
AcgtAF	AAGATCCAAAATCTCCTTTGAATCC	<i>cgtA</i>
A1cgtAr	AAGCAAGCAATCTCCTGGTTC	<i>cgtA</i>
A2cgtAr	TGCTAAACAATCTCCAGGAGC	<i>cgtA</i>
43456-1R	CTGCACACATATAGACCCCTGA	<i>cstII(A)</i>
43456-1F	ATTATTGCTGGAAATGGACCAA	<i>cstII(A)</i>
BneuBr1	GCTAAAGGCTGCATCTACC	<i>neuB</i>
neuCR1	GTGGCGTAAGCTATCGTC	<i>neuC</i>
neuAR1	CATCATTACAAATCCCTGCTAAATC	<i>neuA</i>
classABR	CCCAAGTGTGATACCCTGTTTAAG	<i>orf11</i>
waaVr	CTAAACACTTAGCCCAAACCG	<i>waaV</i>
cgtbr	GGGAGTGATGAAAGTATAGATATAGC	<i>cgtB</i>
43438-1r	TCATATCCCAAATGAGCATCAG	<i>cstII(B)</i>
43438-1f	ATTTCTGGAAATGGACCAAGTTT	<i>cstII(B)</i>
cgtAllr	TGAAAGGCCTGAAATTCTTCTAAAC	<i>cgtAll</i>
<b>Primer walking – Subclass 1C</b>		
1131CF	TGTGAGCGAAACAAGTCCTTTA	<i>gne</i>
1131CR	TTTTTCCCAATCAAAGCAGAT	<i>gne</i>
1132CF	TTTTTGATTGTGAGAGTGTGCC	<i>wlaA</i>
1133R	GATATGCTTGGCATTGTTGGATT	<i>waaC</i>
1134R	GCTTGCTCCCATTGATAATTC	<i>lpxL</i>
1135F	ATTTGCACATTTACCACAGTGC	<i>wlaNA</i>
1135R	GCCTTGCTGTATCTTCTTTGCT	<i>wlaNA</i>
1136R	GCCTCTTTAGCATTTTCTCCAA	<i>wlaNB</i>
1137CF	TGGAAAAATGAGCAATCAAGTG	<i>wlaO</i>
1138R	CTGGTTTCCACCCAACATAACAT	<i>wlaP</i>
1139CF	TAAAATTTTGATATGGGTGGG	<i>cgtB</i>
1140F	AGCTGTGGCTCTGGGATATAAA	<i>cstII</i>
1140R	ATCACATGCCATTGAATGATTG	<i>cstII</i>
1141R	GTATGGCTTCAAATCCTTTGC	<i>neuBI</i>
1142R	AGCTCTCCGCCTTCTATATGTG	<i>neuCI</i>
1143R	TTAACAATTGATCCCCTCCATC	<i>neuAI</i>
1144R	GGGGTGTTATAGGACCTTGTT	<i>wlaSC</i>
1145CF	TCCAGATCCCCTATCGTCTCTA	<i>wlaSB</i>
1146CF	GAAGAAAATTTGGGCACTTTTG	<i>waaV</i>
1146CR	CTAAACACTTAGCCCAAACCG	<i>waaV</i>
1148R	TCATTTGTGATGAAAAGATCGC	<i>waaF</i>
1149CF	GTGGAAAAAGAATGGCAAGAAC	<i>gmhA</i>
1150CF	TTTAAGCGATTATGCCAAAGGT	<i>waaE</i>
1151CF	TGAAAATAGCAATCACAGGTGG	<i>waaD</i>
1152CF	TGCTTTTTATAGCGACAAACCA	<i>wlaT</i>

# Appendix 1

Primer Name	Nucleotide Sequence 5'-3'	Target gene
<b>Primer walking – Subclasses 2D &amp; 2E</b>		
1131CR	TGTGAGCGAAACAAGTCCTTTA	<i>gne</i>
1131CF	TTTTTCCCAATCAAAGCAGAT	<i>gne</i>
wlaXR	GCGAGAGTGTACCAGATGTTG	<i>wlaX</i>
1134R	GCTTGCTCCCATTGATAATTC	<i>lpxL</i>
81116-24r	TAGCCCCATTTTGTATTGCT	<i>wlaUA</i>
81116-25r	GGCTATTTCTTCTAAGCAGGCA	<i>wlaUB</i>
81116-26r	TTTTGAAGCTGAATAAGGGGAA	<i>wlaUC</i>
81116-26f	ATTTTAGTTACTGGCGGAGCTG	<i>wlaUC</i>
129600r	GCATGATCACCTCTTATTGAATC	<i>wlaUD</i>
11828-2f	CAAACACTAGCAATTTTGCTCCA	<i>wlaUE</i>
11828-3f	ATCTTTTGTAACGATAGCCCCC	<i>wlaUF</i>
11828-4f	ATCGATTGGAAAACAGTGTTGA	<i>wlaUG</i>
11828-4r	CCAATAAAAATGACAAGGGCAT	<i>wlaUG</i>
11828-5f	AAAATATCCCCTGTATCGCATT	<i>wlaUH</i>
11828-6f	ACCAAGTTTTCCAAGTATCCCA	<i>wlaUI</i>
11828-7f	TTGCTTATGAGGAGGGATAGGA	<i>wlaUJ</i>
11828-1f	ATTCGTTCCAAATTTCTGCTT	<i>wlaVB</i>
11828-1r	TCAAAAGATAATGCATTGCCA	<i>wlaVB</i>
129592r	GCTATACAAAGAGGGATTATAGG	<i>wlaVA</i>
11828-8f	AGTTTGGGCTCCAAATATCTCA	<i>wlaWA</i>
11828-9f	CCAATTTTTGTAATCTTGCGGT	<i>wlaWB</i>
11828-10f	TTTGATCGTCATCAATTATTCCA	<i>wlaWC</i>
1146CF	GCCACAACCTTCGATCATAATCC	<i>waaF</i>
1148r	GCATTTCTATCCCGAACAAACATC	<i>waaV</i>
11828-11f	CCATTGTATCATTTCCGCCTAT	<i>orf1</i>
11828-12f	TGTCAGCGTAGTCCAAATCATC	<i>orf2</i>
tgh020r	GCATGATCACCTCTTATTGAATC	<i>tgh020</i>
tgh0160r	TTATGAAAGGCTACAACTCTATCG	<i>tgh0160</i>
tgh004r	GTATACTTCCGGCTATACAAAGAGG	<i>tgh004</i>
81116-1f	GCCAAATTCAATGTGCAACTAT	<i>tgh011</i>

# Appendix 1

Primer Name	Nucleotide Sequence 5'-3'	Target gene
<b>Primer walking – Subclasses 3F, 3G &amp; 4H</b>		
1134F	TGCCTGAGTGTATCTTGCATTT	<i>lpxL</i>
1134R	GCTTGCTCCCATTTGATAATTC	<i>lpxL</i>
GorfAR	AAACATAATCAGGCCACCAAC	<i>orfA</i>
1135R	GCCTTGCTGTATCTTCTTTGCT	<i>wlaNA</i>
Lio87orfBr	CTTGTAGCCGAAGATATGCTTGC	<i>orfB</i>
Lio87orfC/Dr	CAGCTAGAAATGCGGGTTTGG	<i>orfC/D</i>
Lio87orfEr	GACCCTGATGATTATTTAGAACTTAATG	<i>orfE</i>
Lio87orfFr	GATTTAATCTTGGCTTGATCATGG	<i>orfF</i>
waaVr	GCATTTCTATCCCGAACAAACATC	<i>waaV</i>
Class GF2	GCGGCGGAGAAAGAGTTTG	<i>orfA</i>
Lio87orfGr	TCTTCCCACGCTAAGTGCAAC	<i>orfG</i>
81116-24r	TAGCCCCATTTTTGTTATTGCT	<i>orf1</i>
43437-orf4cf	TCATGCCTAATCCCTTGGGTG	<i>orf4</i>
43437-orf5cf	AGCCAAAGCAATAGAAAGTTGTATC	<i>orf5</i>
43437-orf7cf	TGGAATTTGTTATTGGCCTATTGTG	<i>orf6</i>

Appendix 1

Primer Name	Nucleotide Sequence 5'-3'
<b>Sequencing Primers</b>	
LOSXLf	AAGCGTCCTATTATCTTCACAACTGCACACTATGG
LOSXLr	ATGCCACAACCTTTCTATCATAATCCCGCTT
LRR1	CATTATGAATTTAATCCAAATGGAAGATATG
44406F1	CAAGCACAAAGTTGCGAAG
LRF2	GATTGGATTTTAAGCATTGATGCTGATGAAGTGCTTG
38625f2	GCAAACCTTAGAAGAAGTAC
39865/47185f2	GATAGCGGAAGCACAGATGATAGTC
50702/51585F2	CCATCATTTAAAGCATACAAAGTAGCAAG
LRF3	GAACATTTAGAATTTGCACAAAGAAGAC
34086r3	GATAATGATCTAAGCAAAGCTAGAATCAC
34218f3	CGCAAAGTAATTTTACAAGGTTCAAGAG
34218r3	GGTATACACCTAGAGGTTATAAAGGTATTG
34565/44406r3	CCTTCCAACCTTTTGCATAATGAGGTAG
34806/51566r3	CCTTGGGGTATTATAGGTCCTTGG
38625f3	AGTGGAAGTAGACTTTATCCTAGC
38625/47185r3	CAAGCCACATTATATCCTCATCAGTCAAACC
39864r3	CATAGGAGAAATACATTCTATATTGTTTGG
40973f3	CGCTGATAGTGAATATATAATAGTTATTGATGG
44406f3	CGATCAGCGTTTGCAATGCCTTAGGG
47185f3	CAAGAGCGTCCAGAAGGTTTAGC
50702/51585f3	GGAGTGGTAGAAATGGATCAAAATGG
50702r3	CCTTGCTCATCAAATTCTTGTAATTC
54386r3	GGCAATGTAAATCCTATAGGAATTCTG
60319r3	GGCTATCAATATGCTTGAAGTTC
34086f4	CAGAGTCCAGATATTCCAATAAATTCC
34086r4	GAAGGAGTTAAGGCAGCTAATTC
34218f4	GTATTATGGTTTGACCATGCTTG
34565f4	GTTTCAGCTAGAAACGTAGGTATAG
34806r4	GAAGTTAATGCTTGTGAAGAGTGTG
38625r4	CAAATGATACAGATATTCTATCTTTC
39864f4	GAGCGTTTGAGTTACGGC
40973r4	TATCGCTGCTTACAAGGACTTTGC
44406f4	GGTTTAAGTTCAGCTAGAAACGTAGG
47185f4	GCCTACAATAACAAGTGGATAGATAATG
47185r4	CAGCGTCTAGCCAAATTCCAC
50702/51585f4	TGCGGTTATAAATTTTGCAGC
51566/54386/61666f4	CCGCAAGCATATCTTCGGCTAC
54386/61666r4	GCGGCTGCAACTGATCTAGCG
54386/61666f5	CCCGCATTCTAGCTGAAGTAAAGCC
54386r5	GATGGCACTGATGATTGTAGGGAAAGTC
51585r5	GGGTTCCATTGCCACCACCAAC
51585f5	GATGTTGTGATTGGAGATAATGTTACG
51566r5	CCCCAATTACACACATACAATAAGC
50702f5	GGTCATGATAGAAGATATGCCATAG
47185f5	CCATGGGTTCCATTGCCACCAC

Appendix 1

Primer Name	Nucleotide Sequence 5'-3'
<b>Sequencing Primers</b>	
40973r5b	CCCCCCCCCCTTTG
40973r5a	CCCCCCCCCCCAAG
39864f5	CTCATCAGTCAAACCTCTTTGTATAGC
38625r5	GAGATGGCTATATGTAAAGATCATTGC
34565f5	GCAGCAAGCAGGATTATAACTG
34218r5	GGCACTAAATTGTAATAAATGGCATTG
34218f5	GCGTGAGTTTGTGTAAAAAGC
34086r5	CATGTTGGAGTATTGGTATTGTTAAGC
34086f5	GCTTCCTTGCCAAGCAAATAAACTCAC
60319r5	GCAGCTATTAAGGGAAATGAAGAG
54386r6	GCACTTGTGCAACTTATGGAGAAC
54386r6_2	CATCAGTGCAGCTGAAATTGTAAGTG
51585r6	GCACACCCTATCTCACCAAGACTTA
51566f6	TCCAAACCCGCATTTCTAGCTG
50702r6	GAGCAAGTGCTTATAATGATGAGTG
47185f6	GCGAAGATTGCATGCTTGTTTATACC
44406f6	GTTGCAAGTTTTATACCAAAGATGAATG
39864r6_2	ATCACCTGATATGGTAACTAAAG
39864f6	CCAGCGTCTAGCCAAATTCCACC
34806r6	GTTGGCAAGATGATTGAAATTTTAGG
34565f6	CCTGCTATCCAAAGAGGTGTTATAGG
34218r6	GGGTGTTCTAGCTTATATTGAATATG
34086f6	GCTTAACAATACCAATACTCCAAC
60319f6	CCCGCATTTCTAGCTGAACTTAAGCC
54386r7	GTAGGTTGTGGGCTTAGTGGAG
51585r7	GTTGCAAGCTTTATTCCACAAGGAG
51585f7	CCATCGTCTAAATAAACTTGCCTATAATGACC
50702r7	CCAACATCACCTTTAAATCCATGGG
50702f7	CATCAAATAAGCATCATCTTCC
47185f7	GCAGAAAGTATGCATGAAGATAGG
39864f7	GCATCATCTTCCAATATCAAACACTC
34806f7	GCTGCCTTAACTCCTTCATATCTAGCTC
34218r7	GCATAGAAGAATGTGTACCTAGAATGG
34086r7	CAATGCCTAAAATTTCAATCATCTTGC
60319f7_2	CAATACTCCAACATGGTAATCTAAG
54386f8a	CCCCCCCCCCCCCCCCCTTAAC
54386f8b	CCCCCCCCCCAGCTG
54386f8	CACACTCTTCACAAGCATTAAAGTTC
50702f8	GTATACTTCCGGCTATACAAAGCG
39864f8	CCGGCTATACAAAGAGGGATTATAGG
34806f8	GGAAATTTAAGGAGTGAAGTTGTG
34565f8	ACTCCACAATCATCCACTAC
34218r8_2	AGTTTAGAATTAGTGCGATTTGTAG
34218f8	CGATAAAGATCCTTTGTGTATTAAAGAG
54386r8	GAATTTCTTGTGAGTATGAAGAG

# Appendix 1

Primer Name	Nucleotide Sequence 5'-3'
<b>Sequencing Primers</b>	
54386r9	GGTGCAACCTATCAAGATTTCC
51585r9	GCTGGAGGTAGTGGAACCTAGG
34565f9_2	GATAATTCACACTATATCGGACACAG
34218r9	GGATAATATAATAAGGCATCTTCTG
60319r9	GCGCTCCAGATAATAAAATATTC
60319r9_2	GCGGGTTGGATATAGCCAAAGGAG
54386f10	GACGCTTTGCAAGTTATAATAAGG
60319r1-2	CAATGAAAGTAGTGATAATACC
54386fb	CCAATACTCCAACATGGTAATCTAAG
54386r1	GCAGTAGGAACCTCCACCGCATCCAG
54386fa	CAGGTAAATCTTGAATGATAAAGTC
51585r2-3	CTTATCAGTCAAACCTCTTTGTATAGCATC
51585f2-3	TCCAAACCCGCATTTCTAGCTG
51566f5-6	GTGTGCAAATATCTATATCATC
51566f1-2	CAAGGATGCAAAATGAATCTAAAGC
50702f3-4	GTTACTGGCGGAGCTGGTTTTATAGG
39864f4-5	TGCATATGATGTTTAAATAAG
39864af	GGGGCTATATCCTGATTATAACATAAG
38625f2-3	GGAGTGCAACAATGAAAGGC
34806f2-3	GGCAAGAAGATGAAGGGTTTAGG
34806f3-4	CTGCAAGCATATCTTCGGCTACAAGG
34218f3c	CACATTTGGCAAGAAGATGAGGG
34218f1-2	GTAGTGATAATACCCTAAATATAGC
34218f1-2	CAATGAAAGTAGTGATAATACC
60319r4-5	CCTGGACCTGGATATGGAGGTTC
60319f6-7	GCATCTTCTGCCATAATTAATC
54386f9-10	GAAGTGAATTATCAAATGGGAGCAAG
54386f6-7	CTCGATGTAATTTATATACG
54386f	CAATACAATCAAGCTCTAAATAATCATC
51585r7-8	TGATCATCCCTGTTTATGCCTATAAAG
51585f4-5	TGGAGATGGTAAGAATATAAGAGATTG
51585f2-3	CATCAGTGCAGCTGAAATTGTAAGTG
51566fa	GCAAGCTTAGAAGAACTGACACTTTATC
51566f1-2	GGATGCAAAATGAATCTAAAGCAAATAAG
50702f4-5	CGCTCAAATTCCAATATATGGAGATGGTAAG
50702f1-2	CATAGACGCTTTGCTAGTTATAATCAAGG
44406f2-3	GGCAAGAAGATGAAGGGTTTAGGCTTAG
39864r7-8	TAGCGAAGATTGCATGCTTGTTTATACC
39864r3-4	GCATGATATTGCAAGAGATGAATATCTG
39864f6-7	GGTTCCATTGCCACCACC
39864f5-6	GCATGATATTGCAAGAGATGAATATCTG
39864f3-4	GCTTTGTTTTGATGCCTTTGTACATAG
34806r4-5	AATAGTTCTGTTGGGACTAATGATATTG
34806f2-3	GAACGCCTTAAGTTAGTGCTTGATAGTG
34218f3-4	GCAGCAAGTAGGATGATAACAGCTTTGG

---

Appendix 1

---

Primer Name	Nucleotide Sequence 5'-3'
<b>Sequencing Primers</b>	
34086f4-5	CCCCAATTACACACATACAATAAGC
47185f3-4	GCTAAAATCACAACATTTAGGAAGAG



## Chapter 6

Primer Name	Nucleotide Sequence 5'-3'	Target gene
<b>Cloning Primers</b>		
Strep(Xba)F	<u>GCTCTAGAGCTCAAAGAAAGGAATTATTGTGCCTACC</u>	<i>rpsL</i>
Strep(Xba)R	<u>GCTCTAGAGCTCTAACGGATTTGTCTGTATGAATG</u>	<i>rpsL</i>
wlaPF(Xba)	<u>GCTCTAGAGCTTAGATTCTGATGATTATTGGGAATTAAAC</u>	<i>wlaP</i>
wlaPR(Xho)	<u>CCGCTCGAGCGGTACAAATCGCACTAATTCTAAACTG</u>	<i>wlaP</i>
wlaSCF(Xho)	<u>CCGCTCGAGCGGGTTGTCCATAAACTCCCATTG</u>	<i>wlaSC</i>
wlaSCR(Kpn)	<u>GGGGTACCCCGGTATGGGTAGATCTTGATATG</u>	<i>wlaSC</i>
CATrpsLF(Xho)	<u>CCGCTCGAGCGGTTTTCATACCAATTTTAAAGCTCTGCTCGGCGGTGTTCCCTTC</u>	
CATrpsLR(Xho)	<u>CCGCTCGAGCGGTTGTTGTTCTAACAATAATTTTAGCTCTAACGGATTTGTCTGTATGAATG</u>	
CATrpsLF1(Xho)	<u>CCGCTCGAGCGGTTTTAATAACAATCTTGTTGTTCTGCTCGGCGGTGTTCCCTTCCAAG</u>	<i>CATrpsL<sup>WT</sup></i>
CATrpsLR1(Xho)	<u>CCGCTCGAGCGGAAAAGTATGGTTAAAAATTGGAAGCTCTAACGGATTTGTCTGTATGAATG</u>	<i>CATrpsL<sup>WT</sup></i>
CATrpsLR2(Xho)	<u>CCGCTCGAGCGGTTTTAATAACAATCTTGTTGTTAGCTCTAACGGATTTGTCTGTATGAATG</u>	<i>CATrpsL<sup>WT</sup></i>
CATrpsLF2(Xho)	<u>CCGCTCGAGCGGAAAAGTATGGTTAAAAATTGGAAGCTCTAACGGATTTGTCTGTATGAATG</u>	<i>CATrpsL<sup>WT</sup></i>
wlaAF(Xba)	<u>GCTCTAGAGCCATTACCTCATCGCCATGCACATC</u>	<i>wlaA</i>
wlaAR(Xho)	<u>CCGCTCGAGCGGGTGCCAGATGTTGAGCTTATCC</u>	<i>wlaA</i>
Cj1152F(Xho)	<u>CCGCTCGAGCGGAAGTGCCAATATCAGCATTTAATCC</u>	<i>wlaT</i>
wlaTR(Kpn)	<u>GGGGTACCCCGGTTTAGATAGAGACGGTGTGATTAATATAG</u>	<i>wlaT</i>

# Appendix 1

Primer Name	Nucleotide Sequence 5'-3'	Target gene
<b>Primer Walking</b>		
1128cr	TAAGCATTTTTACTCGCCCCTA	<i>wlaC</i>
1128cf	CTTGATGCAAAAAGCAAATGAG	<i>wlaC</i>
1129cf	AATATTGCTTGATTGCTGCGA	<i>wlaB</i>
1130cf	TGCGGTGCCTAGAATTTATCTT	<i>wlaB</i>
1131CR	TTTTTCCCAATCAAAGCAGAT	<i>gne</i>
1132CF	TTTTTGATTGTGAGAGTGTGCC	<i>wlaA</i>
1133R	GATATGCTTGGCATTGTTGGATT	<i>waaC</i>
1134R	GCTTGCTCCCATTTGATAATTC	<i>lpxL</i>
1135F	ATTTGCACATTTACCACAGTGC	<i>wlaNA</i>
1135R	GCCTTGCTGTATCTTCTTTGCT	<i>wlaNA</i>
1136R	GCCTCTTTAGCATTTTCTCCAA	<i>wlaNB</i>
1137CF	TGGAAAAATGAGCAATCAAGTG	<i>wlaO</i>
1138R	CTGGTTTCCACCCAACAT	<i>wlaP</i>
1139CF	TAAAATTTTGATATGGGTGGG	<i>cgtB</i>
1140F	AGCTGTGGCTCTGGGATATAAA	<i>cstII</i>
1140R	ATCACATGCCATTGAATGATTG	<i>cstII</i>
1141R	GTATGGCTTCAAATCCTTTGC	<i>neuBI</i>
1142R	AGCTCTCCGCCTTCTATATGTG	<i>neuCI</i>
1143R	TTAACAATTGATCCCCTCCATC	<i>neuAI</i>
1144R	GGGGTGTTATAGGACCTTGGTT	<i>wlaSC</i>
1145CF	TCCAGATCCCCTATCGTCTCTA	<i>wlaSB</i>
1146CF	GAAGAAAATTTGGGCACTTTTG	<i>waaV</i>
1146CR	CTAAACACTTAGCCCAAACCG	<i>waaV</i>
1148R	TCATTTGTGATGAAAAGATCGC	<i>waaF</i>
1149CF	GTGGAAAAAGAATGGCAAGAAC	<i>gmhA</i>
1150CF	TTTAAGCGATTATGCCAAAGGT	<i>waaE</i>
1151CF	TGAAAATAGCAATCACAGGTGG	<i>waaD</i>
1152CF	TGCTTTTTATAGCGACAAACCA	<i>wlaT</i>
1153R	CAAGGTGCGATCATGAAACTTA	<i>Cj1153</i>
1154CF	ATCATTTTGGCAACTTTGCTTT	<i>Cj1154</i>

## Appendix 1

Primer Name	Nucleotide Sequence 5'-3'	Target gene
<b>Sequencing Primers</b>		
M13Flong	GTTGTAAACGACGGCCAGTG	pUC19 MCS
M13Rlong	GGAAACAGCTATGACCATGATTAC	pUC19 MCS
CATinvF	GGAATGTCCGCAAAGCCTAATCC	<i>cat<sup>R</sup></i>
CATinvR	GCGGTCCTGAACTCTTCATGTC	<i>cat<sup>R</sup></i>
seqKANr	AGCCATCATGCCCTTCAA	<i>Kan<sup>R</sup></i>
STMKANINVF	CTGGGGATGAAGCCTGATTG	<i>Kan<sup>R</sup></i>
Cj1152invseqF	GTAGTTTCTGCTTTGGCATGTCTTGG	<i>Cj1152</i>
wlaAseqinvR	CCACGAATATTGTGAAAGTGATG	<i>wlaA</i>

**Table showing raw data for the growth assays with parent and strains.**

# Appendix 1

Time/h	Sample/ OD <sub>600</sub>						
	WT1	WT2	WT3	WT4	WT5	WT6	Wtmean
0	0.000	0.050	0.050	0.050	0.050	0.050	0.050
3	0.060	0.050	0.060	0.060	0.060	0.060	0.058
6	0.100	0.200	0.300	0.100	0.100	0.200	0.167
9	0.400	0.400	0.500	0.400	0.400	0.400	0.417
12	0.600	0.800	0.800	0.800	0.900	0.700	0.767
24	0.800	1.000	1.100	0.900	1.000	1.100	0.983
32	1.200	1.200	1.300	1.300	1.300	1.300	1.267
48	1.300	1.300	1.400	1.350	1.400	1.400	1.358
	PSc1	PSc2	PSc3	PSc4	PSc5	PSc6	Pscmean
0	0.050	0.050	0.050	0.050	0.050	0.050	0.050
3	0.060	0.050	0.060	0.060	0.060	0.060	0.058
6	0.200	0.200	0.200	0.100	0.200	0.100	0.167
9	0.400	0.400	0.300	0.300	0.400	0.400	0.367
12	0.500	0.700	0.600	0.600	0.700	0.800	0.650
24	0.900	0.900	0.800	0.900	0.800	1.000	0.883
32	1.100	1.100	1.200	1.100	1.200	1.100	1.133
48	1.200	1.200	1.200	1.300	1.300	1.200	1.233
	Strep1	Strep2	Strep3	Strep4	Strep5	Strep6	Strepmean
0	0.050	0.050	0.050	0.050	0.050	0.050	0.050
3	0.060	0.060	0.060	0.060	0.060	0.060	0.060
6	0.200	0.300	0.200	0.100	0.100	0.200	0.183
9	0.400	0.600	0.400	0.300	0.400	0.500	0.433
12	0.700	0.700	0.700	0.600	0.600	0.800	0.683
24	1.000	0.900	1.100	1.100	1.100	1.000	1.033
32	1.100	1.000	1.300	1.300	1.300	1.200	1.200
48	1.200	1.100	1.400	1.400	1.400	1.300	1.300
	A52m1	A52m2	A52m3	A52m4	A52m5	A52m6	52mMean
0	0.050	0.050	0.050	0.050	0.050	0.050	0.050
3	0.050	0.050	0.050	0.050	0.050	0.050	0.050
6	0.100	0.080	0.100	0.200	0.200	0.200	0.147
9	0.200	0.300	0.300	0.200	0.200	0.200	0.233
12	0.500	0.400	0.300	0.400	0.300	0.500	0.400
24	0.700	0.700	0.800	0.600	0.600	0.800	0.700
32	1.000	1.000	1.100	0.900	1.000	1.100	1.017
48	1.300	1.300	1.200	1.000	1.100	1.200	1.183

**Table showing raw data for the growth assays with parent and strains with antibiotics**

# Appendix 1

Time/h	Sample/ OD <sub>600</sub>							
	52vt1	52vt2	52vt3	52vtMean	52v1	52v2	52V3	52vMean
0	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050
3	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050
6	0.100	0.200	0.200	0.167	0.100	0.300	0.200	0.200
9	0.300	0.300	0.200	0.267	0.200	0.400	0.300	0.300
12	0.300	0.300	0.300	0.300	0.500	0.400	0.400	0.433
24	0.400	0.400	0.400	0.400	0.700	0.600	0.600	0.633
32	0.500	0.500	0.500	0.500	0.800	0.900	0.900	0.867
48	0.500	0.600	0.500	0.533	0.900	1.000	1.100	1.000
	52t1	52T2	52T3	52Tmean	52k1	52k2	52K3	52Kmean
0	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050
3	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050
6	0.100	0.300	0.200	0.200	0.100	0.300	0.200	0.200
9	0.300	0.300	0.200	0.267	0.300	0.300	0.400	0.333
12	0.300	0.300	0.300	0.300	0.500	0.400	0.600	0.500
24	0.400	0.400	0.400	0.400	0.700	0.500	0.800	0.667
32	0.500	0.500	0.500	0.500	0.900	0.700	1.000	0.867
48	0.500	0.500	0.500	0.500	1.100	1.000	1.200	1.100
	52kvt1	52kvt2	52kvt3	52kvtmean				
0	0.050	0.050	0.050	0.050				
3	0.050	0.050	0.050	0.050				
6	0.100	0.080	0.100	0.093				
9	0.200	0.100	0.200	0.167				
12	0.300	0.200	0.300	0.267				
24	0.400	0.400	0.300	0.367				
32	0.500	0.500	0.400	0.467				
48	0.600	0.600	0.500	0.567				

Time/h	Sample/ OD <sub>600</sub>						
	52m1	52m2	52m3	52m4	52m5	52m6	52mMean
0	0.050	0.050	0.050	0.050	0.050	0.050	0.050
3	0.050	0.050	0.050	0.050	0.050	0.050	0.050
6	0.100	0.080	0.100	0.200	0.200	0.200	0.147
9	0.200	0.300	0.300	0.200	0.200	0.200	0.233
12	0.500	0.400	0.300	0.400	0.300	0.500	0.400
24	0.700	0.700	0.800	0.600	0.600	0.800	0.700
32	1.000	1.000	1.100	0.900	1.000	1.100	1.017
48	1.300	1.300	1.200	1.000	1.100	1.200	1.183
	WT1	WT2	WT3	WT4	WT5	WT6	Wtmean
0	0.000	0.050	0.050	0.050	0.050	0.050	0.050
3	0.060	0.055	0.060	0.060	0.060	0.060	0.059
6	0.100	0.200	0.200	0.100	0.100	0.200	0.150
9	0.400	0.400	0.500	0.400	0.400	0.400	0.417
12	0.600	0.800	0.800	0.800	0.900	0.700	0.767
24	0.800	1.000	1.100	0.900	1.000	1.100	0.983
32	1.200	1.200	1.300	1.300	1.300	1.300	1.267
48	1.300	1.300	1.400	1.350	1.400	1.400	1.358

**Table showing the *t*-test calculations and levels of significance comparing the wild type and mutant strains for growth assays.** The two tailed *t*-test was used and the critical values for *P* determined using Reed (*et al.*, 1999).

Using the formula;

$$t = \frac{Y_1 - Y_2}{\sqrt{(SE_1)^2 + (SE_2)^2}}$$

where SE= the sample standard error,  
y= the sample mean  
n= degrees of freedom=  
(sample size <sub>1</sub> -1)+(sample size <sub>2</sub> -1)



## Appendix 1

	WT	PSc	WT-PSc	$(SE1)^2+(SE2)^2$	$\sqrt{((SE1)^2+(SE2)^2)}$	$t$	n	Significant: yes/no (P value)
0	0.050	0.050	0.000	0.000000	0.0000	0.0000	10	no
3	0.058	0.058	0.000	0.000011	0.0032	0.0000	10	no
6	0.167	0.167	0.000	0.003089	0.0558	0.0000	10	no
9	0.417	0.367	0.050	0.001444	0.0379	1.3176	10	no
12	0.767	0.650	0.117	0.007222	0.0850	1.3769	10	no
24	0.983	0.883	0.100	0.006444	0.0802	1.2461	10	no
32	1.150	1.133	0.017	0.004556	0.0675	0.2517	10	no
48	1.250	1.233	0.017	0.004556	0.0675	0.2517	10	no

	WT	52m	WT-52m	$(SE1)^2+(SE2)^2$	$\sqrt{((SE1)^2+(SE2)^2)}$	$t$	n	Significant: yes/no (P value)
0	0.050	0.050	0.000	0.000000	0.0000	0.0000	10	no
3	0.058	0.050	0.008	0.00001	0.0032	2.6352	10	yes (0.05)
6	0.167	0.147	0.020	0.00220	0.0469	0.4264	10	no
9	0.417	0.233	0.183	0.00056	0.0237	7.7472	10	yes (0.001)
12	0.767	0.400	0.367	0.00356	0.0597	6.1453	10	yes (0.001)
24	0.983	0.700	0.283	0.00456	0.0675	4.1958	10	yes (0.01)
32	1.150	1.017	0.133	0.00367	0.0606	2.2009	10	no
48	1.250	1.183	0.067	0.00367	0.0607	1.097477	10	no

	WT	PSc	WT-PSc	$(SE1)^2+(SE2)^2$	$\sqrt{((SE1)^2+(SE2)^2)}$	$t$	n	Significant: yes/no (P value)
0	0.050	0.050	0.000	0.000000	0.0000	0.0000	10	no
3	0.058	0.058	0.000	0.000011	0.0032	0.0000	10	no
6	0.167	0.167	0.000	0.003089	0.0558	0.0000	10	no
9	0.417	0.367	0.050	0.001444	0.0379	1.3176	10	no
12	0.767	0.650	0.117	0.007222	0.0850	1.3769	10	no
24	0.983	0.883	0.100	0.006444	0.0802	1.2461	10	no
32	1.150	1.133	0.017	0.004556	0.0675	0.2517	10	no
48	1.250	1.233	0.017	0.004556	0.0675	0.2517	10	no

	WT	52m	WT-52m	$(SE1)^2+(SE2)^2$	$\sqrt{((SE1)^2+(SE2)^2)}$	$t$	n	Significant: yes/no (P value)
0	0.050	0.050	0.000	0.000000	0.0000	0.0000	10	no
3	0.058	0.050	0.008	0.00001	0.0032	2.6352	10	yes (0.05)
6	0.167	0.147	0.020	0.00220	0.0469	0.4264	10	no
9	0.417	0.233	0.183	0.00056	0.0237	7.7472	10	yes (0.001)
12	0.767	0.400	0.367	0.00356	0.0597	6.1453	10	yes (0.001)
24	0.983	0.700	0.283	0.00456	0.0675	4.1958	10	yes (0.01)
32	1.150	1.017	0.133	0.00367	0.0606	2.2009	10	no
48	1.250	1.183	0.067	0.00367	0.0607	1.097477	10	no

**Table showing the *t*-test calculations and levels of significance comparing the wild type and mutant strains in growth assays with antibiotics. The two tailed *t*-test was used and the critical values for *P* determined using Reed (*et al.*, 1999).**

Using the formula;

$$t = \frac{Y_1 - Y_2}{\sqrt{((SE_1)^2 + (SE_2)^2)}}$$

where SE= the sample standard error

y= the sample mean

n= degrees of freedom= (sample size <sub>1</sub> -1)+(sample size <sub>2</sub> -1)

## Appendix 1

	WT	52kvt	WT-52kvt	$(SE1)^2+(SE2)^2$	$\sqrt{((SE1)^2+(SE2)^2)}$	t	n	Significant: yes/no (P value)
0	0.050	0.050	0.000	0.000000	0.000000	0.000000	7	no
3	0.058	0.050	0.008	0.000006	0.002366	3.521476	7	yes (0.01)
6	0.167	0.093	0.073	0.002244	0.047375	1.547930	7	no
9	0.417	0.167	0.250	0.001667	0.040829	6.123112	7	yes (0.001)
12	0.767	0.267	0.500	0.004667	0.068313	7.319224	7	yes (0.001)
24	0.983	0.367	0.617	0.005667	0.075277	8.191914	7	yes (0.001)
32	1.150	0.467	0.683	0.004778	0.069122	9.885955	7	yes (0.001)
48	1.250	0.567	0.683	0.004778	0.069122	9.885955	7	yes (0.001)

	52m	52kvt	52m-52kvt	$(SE1)^2+(SE2)^2$	$\sqrt{((SE1)^2+(SE2)^2)}$	t	n	Significant: yes/no (P value)
0	0.050	0.050	0.000	0.000000	0.000000	0.0000	7	no
3	0.050	0.050	0.000	0.000000	0.000000	0.0000	7	no
6	0.147	0.093	0.053	0.00120	0.034641	1.5396	7	no
9	0.233	0.167	0.067	0.00200	0.044721	1.4907	7	no
12	0.400	0.267	0.133	0.00378	0.061482	2.1687	7	yes (0.01)
24	0.700	0.367	0.333	0.00378	0.061482	5.4217	7	yes (0.001)
32	1.017	0.467	0.550	0.00300	0.027928	19.6932	7	yes (0.001)
48	1.183	0.050	1.133	0.00567	0.075299	15.0510	7	yes (0.001)

	52m	52t	52m-52t	$(SE1)^2+(SE2)^2$	$\sqrt{((SE1)^2+(SE2)^2)}$	t	n	Significant: yes/no (P value)
0	0.050	0.050	0.000	0.000000	0.000000	0	7	no
3	0.050	0.050	0.000	0.000000	0.000000	0	7	no
6	0.147	0.200	-0.053	0.00449	0.067007	-0.7959	7	no
9	0.233	0.267	-0.033	0.00200	0.044721	-0.7454	7	no
12	0.400	0.300	0.100	0.00267	0.051672	1.9353	7	no
24	0.700	0.400	0.300	0.00267	0.051672	5.8058	7	yes (0.001)
32	1.017	0.500	0.517	0.00189	0.043474	11.8845	7	yes (0.001)
48	1.183	0.500	0.683	0.00456	0.067528	10.1193	7	yes (0.001)

Table showing raw data accumulated for the sensitivity assays

# Appendix 1

Novobiocin Concentration $\mu\text{g/ml}$ / $\text{OD}_{600}$					
Sample	50	25	12.5	6.25	3.125
WT1	0.010	0.010	0.500	0.440	0.520
WT2	0.020	0.020	0.520	0.550	0.550
WT3	0.040	0.050	0.550	0.600	0.650
Wtmean	0.023	0.027	0.523	0.530	0.573
sd	0.015	0.021	0.025	0.082	0.068
Std Er	0.009	0.012	0.015	0.047	0.039
S1	0.020	0.050	0.520	0.510	0.630
S2	0.040	0.090	0.530	0.550	0.560
S3	0.030	0.040	0.570	0.660	0.650
SMEAN	0.030	0.060	0.540	0.573	0.613
sd	0.010	0.026	0.026	0.078	0.047
Std Er	0.006	0.015	0.015	0.045	0.027
PSC1	0.020	0.060	0.570	0.520	0.610
PSC2	0.020	0.020	0.460	0.590	0.500
PSC3	0.020	0.090	0.660	0.640	0.640
PSCMEAN	0.020	0.057	0.563	0.583	0.583
sd	0.000	0.035	0.100	0.060	0.074
Std Er	0.000	0.020	0.058	0.035	0.043
A52k1	0.020	0.040	0.040	0.050	0.140
A52K2	0.010	0.010	0.020	0.090	0.150
A52K3	0.020	0.020	0.040	0.120	0.130
A52KMEAN	0.017	0.023	0.033	0.087	0.140
sd	0.006	0.015	0.012	0.035	0.010
Std Er	0.003	0.009	0.007	0.020	0.006

## Appendix 1

SDS concentration $\mu\text{g/ml}$ / $\text{OD}_{600}$						
Sample	400	200	100	50	25	12.5
WT1	0.020	0.020	0.120	0.400	0.470	0.570
WT2	0.010	0.010	0.110	0.290	0.390	0.450
WT3	0.010	0.010	0.140	0.280	0.370	0.440
Wtmean	0.013	0.013	0.123	0.323	0.410	0.487
sd	0.006	0.006	0.015	0.067	0.053	0.072
Std Er	0.003	0.003	0.009	0.038	0.031	0.042
S1	0.010	0.050	0.160	0.320	0.350	0.520
S2	0.010	0.020	0.090	0.250	0.330	0.460
S3	0.030	0.040	0.150	0.280	0.290	0.480
SMEAN	0.017	0.037	0.133	0.283	0.323	0.487
sd	0.012	0.015	0.038	0.035	0.031	0.031
Std Er	0.007	0.009	0.022	0.020	0.018	0.018
PSC1	0.020	0.020	0.060	0.290	0.330	0.390
PSC2	0.020	0.030	0.030	0.240	0.310	0.420
PSC3	0.010	0.020	0.060	0.230	0.280	0.400
PSCMEAN	0.017	0.023	0.050	0.253	0.307	0.403
sd	0.006	0.006	0.017	0.032	0.025	0.015
Std Er	0.003	0.003	0.010	0.019	0.015	0.009
A52k1	0.040	0.040	0.040	0.040	0.040	0.190
A52K2	0.010	0.010	0.010	0.010	0.010	0.180
A52K3	0.020	0.020	0.020	0.020	0.020	0.190
A52KMEAN	0.023	0.023	0.023	0.023	0.023	0.187
sd	0.015	0.015	0.015	0.015	0.015	0.006
Std Er	0.009	0.009	0.009	0.009	0.009	0.003

## Appendix 1

Polymyxin B concentration $\mu\text{g/ml}$ /OD <sub>600</sub>						
Sample	2	1	0.5	0.25	0.125	0.0625
WT1	0.030	0.030	0.640	0.770	0.750	0.750
WT2	0.020	0.030	0.550	0.760	0.790	0.800
WT3	0.030	0.050	0.660	0.830	0.880	0.880
Wtmean	0.027	0.037	0.617	0.787	0.807	0.810
sd	0.006	0.012	0.059	0.038	0.067	0.066
Std Er	0.003	0.007	0.034	0.022	0.038	0.038
S1	0.020	0.030	0.610	0.720	0.870	0.980
S2	0.030	0.040	0.590	0.690	0.780	0.890
S3	0.030	0.060	0.650	0.790	0.880	0.930
SMEAN	0.027	0.043	0.617	0.733	0.843	0.933
sd	0.006	0.015	0.031	0.051	0.055	0.045
Std Er	0.003	0.009	0.018	0.030	0.032	0.026
PSC1	0.050	0.060	0.650	0.740	0.800	0.860
PSC2	0.040	0.080	0.600	0.840	0.860	0.930
PSC3	0.050	0.090	0.660	0.890	0.890	0.970
PSCMEAN	0.047	0.077	0.637	0.823	0.850	0.920
sd	0.006	0.015	0.032	0.076	0.046	0.056
Std Er	0.003	0.009	0.019	0.044	0.026	0.032
A52k1	0.030	0.030	0.040	0.040	0.040	0.100
A52K2	0.030	0.030	0.050	0.050	0.050	0.120
A52K3	0.040	0.040	0.040	0.040	0.040	0.110
A52KMEAN	0.033	0.033	0.043	0.043	0.043	0.110
sd	0.006	0.006	0.006	0.006	0.006	0.010
Std Er	0.003	0.003	0.003	0.003	0.003	0.006

Table detailing Invasion assay raw data



## Appendix 1

---

	Expt 1			
	WT	strep WT	wlaP-SC	wlaA-T
1	0.00E+00	2.00E+01	1.00E+02	0.00E+00
2	0.00E+00	0.00E+00	1.00E+02	0.00E+00
3	0.00E+00	1.00E+01	1.60E+02	0.00E+00
4	1.00E+01	0.00E+00	7.00E+01	0.00E+00
5	1.00E+01	1.00E+01	1.60E+02	0.00E+00
6	1.00E+01	2.00E+01	1.50E+02	0.00E+00
7	2.00E+01	2.00E+01	1.00E+02	0.00E+00
8	3.00E+01	4.00E+01	1.30E+02	0.00E+00
9	3.00E+01	2.00E+01	1.30E+02	0.00E+00
ave	1.22E+01	1.56E+01	1.22E+02	0.00E+00
sd	12.0185	12.36033	31.53481	0.00E+00
avexdiIn10	1.22E+02	1.56E+02	1.22E+03	0.00E+00

	Expt 2			
	WT	strep WT	wlaP-SC	wlaA-T
1	3.00E+01	2.00E+01	1.10E+02	0.00E+00
2	2.00E+01	0.00E+00	6.00E+01	0.00E+00
3	1.00E+01	0.00E+00	1.20E+02	0.00E+00
4	1.00E+01	1.00E+01	1.40E+02	0.00E+00
5	0.00E+00	1.00E+01	1.30E+02	0.00E+00
6	1.00E+01	2.00E+01	1.50E+02	0.00E+00
7	3.00E+01	1.00E+01	1.30E+02	0.00E+00
8	3.00E+01	2.00E+01	1.40E+02	0.00E+00
9	1.00E+01	2.00E+01	1.20E+02	0.00E+00
ave	1.67E+01	1.22E+01	1.22E+02	0.00E+00
sd	11.18034	8.333333	26.35231	0.00E+00
avexdiIn10	1.67E+02	2.00E+02	1.22E+03	0.00E+00

Table showing raw data for growth assay in iron controlled conditions

# Appendix 1

NCTC 11168 Sample				
Time/h	1	2	3	Mean
0	0.01	0.01	0.01	0.01
4	0.02	0.02	0.02	0.02
8	0.12	0.09	0.10	0.10
12	0.12	0.10	0.10	0.11
24	0.12	0.12	0.10	0.11
positive sample				
Time/h	1	2	3	Mean
0	0.01	0.01	0.01	0.01
4	0.04	0.04	0.05	0.04
8	0.23	0.22	0.22	0.22
12	0.30	0.30	0.29	0.30
24	0.29	0.30	0.30	0.30
porcine haemin				
Time/h	1	2	3	Mean
0	0.01	0.01	0.01	0.01
4	0.05	0.04	0.04	0.04
8	0.28	0.26	0.27	0.27
12	0.35	0.36	0.37	0.36
24	0.32	0.33	0.34	0.33
LOS mutant				
Time/h	1	2	3	Mean
0	0.02	0.02	0.02	0.02
4	0.01	0.01	0.01	0.01
8	0.04	0.04	0.03	0.04
12	0.09	0.09	0.09	0.09
24	0.15	0.15	0.15	0.15
positive				
Time/h	1	2	3	Mean
0	0.01	0.02	0.01	0.01
4	0.03	0.03	0.03	0.03
8	0.08	0.09	0.09	0.09
12	0.15	0.15	0.16	0.15
24	0.36	0.36	0.37	0.37
haemin				
Time/h	1	2	3	Mean
0	0.01	0.01	0.01	0.01
4	0.02	0.01	0.01	0.01
8	0.06	0.04	0.05	0.05
12	0.12	0.13	0.13	0.13
24	0.27	0.28	0.25	0.27

## Bibliography

- Adak, G. K. (2002). Trends in indigenous foodborne disease and deaths, England and Wales: 1992 to 2000. Gut **51**: 832-841.
- Adak, G. K., S. M. Meakins, H. Yip, B. A. Lopman and S. J. O'Brien (2005). Disease risks from Foods, England and Wales, 1996-2000. EID **11**(3): 365-372.
- Ahmed, I. H., G. Manning, T. M. Wassenaar, S. Cawthraw and D. G. Newell (2002). Identification of genetic differences between two *Campylobacter jejuni* strains with different colonization potentials. Microbiology **148**(Pt 4): 1203-12.
- Allos, B. M. (1997). Association between *Campylobacter* infection and Guillain-Barre syndrome. J Infect Dis **176 Suppl 2**: S125-8.
- Allos, B. M. (2001). *Campylobacter jejuni* infections: update on emerging issues and trends. Clin. Infect. Dis. **32**: 1201-1206.
- Allos, B. M., F. T. Lippy, A. Carlsen, R. G. Washburn and M. J. Blaser (1998). *Campylobacter jejuni* strains from patients with Guillain-Barre syndrome. Emerg Infect Dis **4**(2): 263-8.
- Amor, K., D. E. Heinrichs, E. Firdich, K. Ziebell, R. P. Johnson and C. Whitfield (2000). Distribution of core oligosaccharide types in lipopolysaccharides from *Escherichia coli*. Infect Immun. **68**(3): 1116-24.
- Appelmek, B. J., R. Negrini, A. P. Moran and E. J. Kuipers (1997). Molecular mimicry between *Helicobacter pylori* and the host. Trends in Microbiology **5**(2): 70-73.
- Aspinall, G., C. Lynch, H. Pang, S. RT and A. P. Moran (1995). Chemical Structure of the core region of *Campylobacter jejuni* O:3 lipopolysaccharide and an associated polysaccharide. European Journal of Biochemistry **231**: 570-578.
- Aspinall, G., A. G. McDonald and H. Pang (1992a). Structures of the O chains from lipopolysaccharides of *Campylobacter jejuni* serotypes O:23 and O:36. Carbohydrate Research **231**: 13-30.
- Aspinall, G., A. G. McDonald, H. Pang, L. Kurjanczyk and J. L. Penner (1993a). Lipopolysaccharide of *Campylobacter coli* Serotype O:30. J. Biol Chem **268**(9): 6263-6268.
- Aspinall, G. O., A. G. McDonald, H. Pang, L. A. Kurjanczyk and J. L. Penner (1994). Lipopolysaccharides of *Campylobacter-Jejuni* Serotype-O-19 - Structures of Core Oligosaccharide Regions from the Serostrain and 2 Bacterial Isolates from Patients with the Guillain-Barre- Syndrome. Biochemistry **33**(1): 241-249.
- Aspinall, G. O., A. G. McDonald, T. S. Raju, H. Pang, L. A. Kurjanczyk, J. L. Penner and A. P. Moran (1993c). Chemical-Structure of the Core Region of *Campylobacter-Jejuni* Serotype O-2 Lipopolysaccharide. European Journal of Biochemistry **213**(3): 1029-1037.
- Aspinall, G. O., A. G. McDonald, T. S. Raju, H. Pang, S. D. Mills, L. A. Kurjanczyk and J. L. Penner (1992b). Serological Diversity and Chemical Structures of *Campylobacter-Jejuni* Low-Molecular-Weight Lipopolysaccharides. Journal of Bacteriology **174**(4): 1324-1332.
- Aspinall, G. O., A. G. McDonald, T. S. Raju, H. Pang, A. P. Moran and J. L. Penner (1993b). Chemical Structures of the Core Regions of

- Campylobacter-Jejuni Serotypes O-1, O-4, O-23, and O-36 Lipopolysaccharides. European Journal of Biochemistry **213**(3): 1017-1027.
- Avarre, J.-C., P. de Lajudie and B. G. (2006). Hybridisation of genomic DNA to microarrays: A challenge for the analysis of environmental samples. J. Microbiol. Methods doi:10.1016/j.mimet.2006.11.007.
- Bacon, D. J., R. A. Alm and D. H. Burr (2000). Involvement of a plasmid in virulence of *Campylobacter jejuni* 81-176. Infection & Immunity. **68**(8): 4384-4390.
- Bacon, D. J., C. M. Szymanski, D. H. Burr, R. P. Silver, R. A. Alm and P. Guerry (2001). A phase-variable capsule is involved in virulence of *Campylobacter jejuni* 81-176. Molecular Microbiology **40**(3): 769-777.
- Bannerjee, A., R. Wang, S. N. Uljon, P. A. Rice, E. C. Gotschlich and D. C. Stein (1998). Identification of the gene (*lgtG*) encoding the lipooligosaccharide  $\beta$  chain synthesizing glucosyl transferase from *Neisseria gonorrhoeae*. PNAS **95**: 10872-10877.
- Bernatchez, S., C. M. Szymanski and N. Ishiyama (2005). A Single bifunctional UDP-GlcNAc/Glc 4-Epimerase supports the synthesis of three cell surface glycoconjugates in *Campylobacter jejuni*. J Biol Chem **280**(6): 4792-4802.
- Bertone, P., M. Gerstein and M. Snyder (2005). Applications of DNA tiling arrays to experimental genome annotation and regulatory pathway discovery. Chromosome Research **13**: 259-274.
- Birnboim, H. C. and J. Doly (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. NAR **7**: 1513-1523.
- Black, R. E., M. M. Levine, M. L. Clements, T. P. Hughes and M. J. Blaser (1988). Experimental *Campylobacter jejuni* infection in humans. J. Infect. Dis. **157**(3): 472-479.
- Blattner, F. R., G. Plunkett and B. C.A. (1997). The complete genome sequence of *Escherichia coli* K-12. Science **277**: 1453-1462.
- Boer, P., J. A. Wagenaar, R. P. Achterberg, J. P. Putten, L. M. Schouls and B. Duim (2002). Generation of *Campylobacter jejuni* genetic diversity in vivo. Mol Microbiol **44**(2): 351-9.
- Bos, M. P. and J. Tommassen (2004). Biogenesis of the Gram-negative bacterial outer membrane Current Opin Micro **7**: 610-616.
- Buffone, G. J. and G. J. Darlington (1985). Isolation of DNA from biological specimens without extraction with phenol. Clin Chem **31**(1): 164-5.
- Champion, O. L., M. W. Gaunt, O. Gundogdu, A. Elmi, J. Hinds, N. Dorrell and B. W. Wren (2005). Comparative phylogenomics of the food-borne pathogen *Campylobacter jejuni* reveals genetic markers predictive of infection source. PNAS **102**(44): 16043-16048.
- Coward, C., A. J. Grant, C. Swift, J. Philp, R. Towler, M. Heydarian, J. A. Frost and D. J. Maskell (2006). Phase-variable surface structures are required for infection of *Campylobacter jejuni* by bacteriophages. Appl. Environ. Microbiol **72**(7): 4638-4647.
- CSSS (2003). Ethnicity and *Campylobacter* infection: a population-based questionnaire survey. J Infect **47**: 210-216.
- CSSS (2005). Is *Campylobacter jejuni* enteritis a weekend disease? J Infect **50**: 265-
- De Bolle, X., C. D. Bayliss, D. Field, T. van de Ven, N. J. Saunders, D. W.

- Hood and E. R. Moxon (2000). The length of a tetranucleotide repeat tract in *Haemophilus influenzae* determines the phase variation rate of a gene with homology to type III DNA methyltransferases. Mol Microbiol **35**(1): 211-22.
- Dingle, K. E., F. M. Colles, D. Falush and M. C. J. Maiden (2005). Sequence typing and comparison of population biology of *Campylobacter coli* and *Campylobacter jejuni*. J Clin Micro **43**(1): 340-347.
- Dingle, K. E., F. M. Colles, R. Ure, J. A. Wagenaar, B. Duim, F. J. Bolton, A. J. Fox, D. R. A. Wareing and M. C. J. Maiden (2002). Molecular characterisation of *Campylobacter jejuni* clones: A basis for epidemiologic investigation. Emerg. Infect. Dis. **8**(9): 949-955.
- Dingle, K. E., F. M. Colles, D. R. A. Wareing, R. Ure, A. J. Fox, F. J. Bolton, H. J. Bootsma, R. J. L. Willems, R. Urwin and M. C. J. Maiden (2001a). Multilocus sequence typing system for *Campylobacter jejuni*. J Clin Micro **39**(1): 14-23.
- Dingle, K. E., N. van den Braak, F. M. Colles, L. J. Price, D. L. Woodward, R. F.G., H. P. Endtz, A. van Belkum and M. C. J. Maiden (2001b). Sequence typing confirms that *Campylobacter jejuni* strains associated with Guillain Barré and Miller Fisher syndromes are of diverse genetic lineage, serotype and flagella type. J Clin Micro **39**(9): 3346-3349.
- Doig, P., N. Kinsella, P. Guerry and T. J. Trust (1996). Characterization of a posttranslational modification of campylobacter flagellin - identification of a sero-specific glycosyl moiety. Mol. Microbiol. **19**(2): 379-387.
- Doig, P., R. J. Yao, D. H. Burr, P. Guerry and T. J. Trust (1996). An environmentally regulated pilus-like appendage involved in *Campylobacter* pathogenesis. Molecular Microbiology **20**(4): 885-894.
- Donahue Jr, R. A. and F. R. Bloom (1998). Focus. **20**.
- Dorrell, N., O. L. Champion and B. W. Wren (2005). Advances in *Campylobacter jejuni* comparative genomics through whole genome DNA microarrays. Campylobacter: Molecular and Cellular Biology. J.M.Ketley and M. E. Konkel, Horizon Biosciences: 79 - 99.
- Dorrell, N., J. A. Mangan, K. G. Laing, J. Hinds, D. Linton, H. Al-Ghusein, B. G. Barrell, J. Parkhill, N. G. Stoker, A. V. Karlyshev, P. D. Butcher and B. W. Wren (2001). Whole genome comparison of *Campylobacter jejuni* human isolates using a low-cost microarray reveals extensive genetic diversity. Genome Research **11**(10): 1706-1715.
- Elvers, K. T., S. M. Turner, L. M. Wainwright, G. Marsden, J. Hinds, J. A. Cole, R. K. Poole, C. W. Penn and S. F. Park (2005). NssR, a member of the Crp-Fnr superfamily from *Campylobacter jejuni*, regulates a nitrosative stress-responsive regulon that includes both a single-domain and a truncated haemoglobin. Mol Micro **57**(3): 735-750.
- Everest, P. H., H. Goossens, J. P. Butzler, D. Lloyd, S. Knutton, J. M. Ketley and P. H. Williams (1992). Differentiated caco-2 cells as a model for enteric invasion by *Campylobacter jejuni* and *Campylobacter coli*. Journal of Medical Microbiology **37**(5): 319-325.
- Fitzgerald, C., K. Stanley, S. Andrew and K. Jones (2001). Use of pulsed-field gel electrophoresis and flagellin gene typing in identifying clonal groups of *Campylobacter jejuni* and *Campylobacter coli* in farm and clinical environments. Appl Environ Microbiol **67**(4): 1429-36.
- Fouts, D. E., E. F. Mongodin, R. E. Mandrell, W. G. Miller and D. A. Rasko

- (2005). Major structural differences and novel potential virulence mechanisms from the genomes of multiple *Campylobacter* species. *PLoS Biology* **3**(1): 72- 85.
- Friedman, C. R., J. Neimann, H. C. Wegner and R. V. Tauxe (2000). Epidemiology of *Campylobacter jejuni* infections in the United States and other industrialized nations. *Campylobacter*. I. Nachamkin and M. J. Blaser. Washington, D.C, American Society for Microbiology: 121-138.
- Frost, J. A. (2001). Current epidemiological issues in human campylobacteriosis. *Sym Ser Soc Appl Microbiol* **30**: 85S-95S.
- Frost, J. A., A. N. Oza, R. T. Thwaites and B. Rowe (1998). Serotyping scheme for *Campylobacter jejuni* and *Campylobacter coli* based on direct agglutination of heat-stable antigens. *J Clin Microbiol* **36**(2): 335-9.
- Fry, B. N., S. Feng, Y. Y. Chen, D. G. Newell, P. J. Coloe and V. Korolik (2000). The *galE* gene of *Campylobacter jejuni* is involved in lipopolysaccharide synthesis and virulence. *Infect Immun* **68**(5): 2594-601.
- Fry, B. N., V. Korolik, J. A. ten Brinke and M. T. T. Pennings (1998). The lipopolysaccharide biosynthesis locus of *Campylobacter jejuni* 81116. *Microbiology (Reading)* **144**: 2049-2061.
- Fry, B. N., N. J. Oldfield, V. Korolik, P. J. Coloe and J. M. Ketley (2000). Genetics of *Campylobacter* lipopolysaccharide biosynthesis. *Campylobacter*. I. Nachamkin and M. J. Blaser. Washington, D.C, American Society for Microbiology: 381-403.
- Gadgil, M., W. Lian, C. Gadgil, V. Kapur and W. S. Hu (2005). An analysis of the use of genomic DNA as a universal reference in two channel DNA microarrays. *BMC Genomics* **6**(66).
- Garrido, P., M. Blanco, M. Moreno-Paz, C. Briones, G. Dahbi, J. Blanco, J. Blanco and V. Parro (2006). STEC-EPEC oligonucleotide microarray: a new tool for typing genetic variants of the LEE pathogenicity island of human and animal shiga-toxin producing *Escherichia coli* (STEC) and enteropathogenic *E.coli* (EPEC) strains. *Clin Chem* **52**(2): doi:10.1373/clinchem.2005.059766.
- Gaynor, E. C., S. Cawthraw, G. Manning, J. K. MacKichan, S. Falkow and D. G. Newell (2004). The Genome-sequenced variant of *Campylobacter jejuni* NCTC 11168 and the original clonal clinica isolate differ markedly in colonization, gene expression and virulence-associated phenotypes. *J Bac* **186**(2): 503-517.
- Gilbert, M., J. R. Brisson, M. F. Karwaski, J. Michniewicz, A. M. Cunningham, Y. Wu, N. M. Young and W. W. Wakarchuk (2000). Biosynthesis of ganglioside mimics in *Campylobacter jejuni* OH4384. Identification of the glycosyltransferase genes, enzymatic synthesis of model compounds, and characterization of nanomole amounts by 600-mhz (1)h and (13)c NMR analysis. *J Biol Chem* **275**(6): 3896-906.
- Gilbert, M., P. C. R. Godschalk, M.-F. Karwaski, C. Wim Ang, A. van Belkum, J. Li, W. W. Wararchuk and H. P. Endtz (2004). Evidence for acquisition of the lipooligosaccharide biosynthesis loci in *Campylobacter jejuni* GB11, a strain from a patient with Guillain Barré syndrome, by horizontal exchange. *Infection & Immunity*. **72**(2): 1162-

- 1165.
- Gilbert, M., P. C. R. Godschalk, C. T. Parker, H. P. Endtz and W. W. Wakarchuk (2005). Genetic bases for the variation in the lipooligosaccharide outer core of *Campylobacter jejuni* and possible association of glycosyltransferase genes with post-infectious neuropathies. Campylobacter. Molecular and cellular biology. J. M. Ketley and M. E. Konkel. London, Horizon Biosciences: 219-248.
- Gilbert, M., M.-F. Karwaski, S. Bernatchez, N. Martin Young, E. N. Taboada, J. Michniewicz, A.-M. Cunningham and W. Wakarchuk (2002). The genetic basis for variation in the lipooligosaccharide of the mucosal pathogen, *Campylobacter jejuni*. J. Biol Chem **277**(1): 327-337.
- Gillespie, I. A., S. J. O'Brien, J. A. Frost, G. K. Adak, P. Horby, A. V. Swan, M. J. Painter, K. R. Neal and CSSS. (2002). A Case-case comparison of *Campylobacter coli* and *Campylobacter jejuni* infection: A tool for generating hypotheses. Emerg Infect Dis **8**(9): 937-42.
- Godschalk, P. C. R., M. L. Kuijf, J. Li, F. St. Michael, C. Wim Ang, B. C. Jacobs, M.-F. Karwaski, D. Brochu, A. Moterassed, H. P. Endtz, A. van Belkum and M. Gilbert (2007). Structural characterisation of *Campylobacter jejuni* lipooligosaccharide outer cores associated with Guillain -Barré Syndrome. Infect.Immun **75**(3): 1245-1254.
- Godschalk, P. C. R., A. van Belkum, N. van den Braak, D. van Netten, C. W. Ang, B. C. Jacobs, M. Gilbert, H. A. Verbrugh and H. P. Endtz (2001b). Comparative analysis of *Campylobacter jejuni* genes involved in the biosynthesis of lipooligosaccharide in strains derived from patients with Guillain-Barre or Miller Fisher syndrome. 11th International Workshop on Campylobacter, Helicobacter and related Organisms September 01-05, 2001 Freiburg, Germany; IJMM International Journal of Medical Microbiology **291**(Supplement 31): 127-128.
- Goodyear, C. S., G. M. O'Hanlon, J. J. Plomp, E. R. Wagner and I. Morrison (1999). Monoclonal antibodies raised against Guillain Barré syndrome associated *Campylobacter jejuni* lipopolysaccharides react with neuronal gangliosides and paralyze muscle-nerve preparations. J.Clin Invest. **104**(6): 697-708.
- Goon, S., J. F. Kelly, S. M. Logan, C. P. Ewing and P. Guerry (2003). Pseudaminic acid, the major modification on *Campylobacter* flagellin, is synthesised via the Cj1293 gene. Mol Micro **50**(2): 659-671.
- Grainger, D. C., S. J. W. Busby, D. Hurd, J. Holdstock and M. Harrison (2005b). How do bacteria turn their genes on and off? Genetic Engineering New **25**(17).
- Grainger, D. C., D. Hurd, M. Harrison, J. Holdstock and S. J. W. Busby (2005). Studies of the distribution of *Escherichia coli* cAMP-receptor protein and RNA polymerase along the *E.coli* chromosome. PNAS **102**(49): 17693-17698.
- Gronow, S. and H. Brade (2000). Lipopolysaccharide biosynthesis: which steps do bacteria need to survive? J Endotoxin research **7**(1): 4-23.
- Guerry, P. (1997). Nonlipopolysaccharide surface antigens of *Campylobacter* species. J. Infect. Dis. **176**(S2): S122-S124.
- Guerry, P., P. Doig, R. A. Alm, D. H. Burr, N. Kinsella and T. J. Trust (1996). Identification and characterization of genes required for posttranslational modification of *Campylobacter coli* VC167 flagellin.



- Mol. Microbiol. **19**(2): 369-378.
- Guerry, P., C. P. Ewing, T. E. Hickey, M. M. Prendergast and A. P. Moran (2000). Sialylation of lipooligosaccharide cores affects immunogenicity and serum resistance of *Campylobacter jejuni*. Infection and Immunity **68**(12): 6656-6662.
- Guerry, P., C. P. Ewing, M. Schirm, M. Lorenzo, J. Kelly, D. L. Pattarini, G. Majam, P. Thibault and S. M. Logan (2006). Changes in flagellin glycosylation affect *Campylobacter* autoagglutination and virulence. Mol Micro **60**(2): 299-311.
- Guerry, P., C. M. Szymanski, M. M. Prendergast, T. E. Hickey, C. P. Ewing, D. L. Pattarini and A. P. Moran (2002). Phase variation of *Campylobacter jejuni* 81-176 lipooligosaccharide affects ganglioside mimicry and invasiveness in vitro. Infect Immun **70**(2): 787-93.
- Hadden, R. D. M. and N. A. Gregson (2001). Guillain Barré syndrome and *Campylobacter jejuni* infection. J. Appl. Microbiol **90**: 146S-154S.
- Heinrichs, D. E., J. A. Yethon, P. A. Amor and C. Whitfield (1998b). The assembly system for the outer core portion of R1- and R4-type lipopolysaccharides of *Escherichia coli*. The R1 core-specific beta-glucosyltransferase provides a novel attachment site for O-polysaccharides. J Biol Chem **273**(45): 29497-505.
- Heinrichs, D. E., J. A. Yethon and C. Whitfield (1998a). Molecular basis for structural diversity in the core regions of the lipopolysaccharides of *Escherichia coli* and *Salmonella enterica*. Molecular Microbiology **30**(2): 221-232.
- Hendrixson, D. R., B. J. Akerley and V. J. DiRita (2001). Transposon mutagenesis of *Campylobacter jejuni* identifies a bipartite energy taxis system required for motility. Mol Microbiol **40**(1): 214-24.
- Hendrixson, D. R. and V. J. DiRita (2004). Identification of *Campylobacter jejuni* genes involved in commensal colonisation of the chick gastrointestinal tract. Molecular Microbiology **52**(2): 471-484.
- Hitchen, P. G., J. L. Prior, P. C. F. Oyston, M. Panico and B. W. Wren (2002). Structural characterization of lipo-oligosaccharide (LOS) from *Yersinia pestis*: regulation of LOS structure by the PhoPQ system. Molecular Microbiology **44**(6): 1637-1650.
- Hofreuter, D., S. Odenbreit and R. Haas (2001). Natural transformation competence in *Helicobacter pylori* is mediated by the basic components of a type IV secretion system. Mol Microbiol **41**(2): 379-91.
- Hofreuter, D., J. Tsai, R. O. Watson, V. Novik, B. Altman, M. Benitez, C. Clark, T. Jarvie, L. Du and J. E. Galán (2006). Unique Features of a highly pathogenic *Campylobacter jejuni* strain. Infect.Immun **74**(8): 4694-4707.
- Hughes, R. and J. Rees (1997). Clinical and epidemiological features of Guillain Barré syndrome. J Infect Dis **176**: S92-8.
- Jackson, S. A., M. K. Mammel, I. R. Patel, T. Mays, T. J. Albert, J. E. LeClerc and T. A. Cebula (2006). Interrogating genomic diversity of *E.coli* O157:H7 using DNA tiling arrays. Forensic Science International doi:10.1016/j.forsciint.2006.06.079.
- Jin, S., A. Joe, J. Lynett, E. K. Hani, P. Sherman and V. L. Chan (2001). JlpA, a novel surface-exposed lipoprotein specific to *Campylobacter jejuni*, mediates adherence to host epithelial cells. Mol Microbiol **39**(5): 1225-

- 36.
- Kane, M. D., T. A. Jatcoe, C. R. Stumpf, J. Lu, J. D. Thomas and S. J. Madore (2000). Assessment of the sensitivity and specificity of oligonucleotide (50mer) microarrays. Nucleic Acids Res **28**: 4552-4557.
- Kanipes, M. I., L. C. Holder, A. T. Corcoran, A. P. Moran and P. Guerry (2004). A deep-rough mutant of *Campylobacter jejuni* 81-176 is noninvasive for intestinal epithelial cells. Infection & Immunity. **72**(4): 2452-2455.
- Kanipes, M. I., E. Papp-Szabo, P. Guerry and M. A. Monteiro (2006). Mutation of *waaC*, encoding heptosyltransferase I in *Campylobacter jejuni* 81-176, affects the structure of both lipooligosaccharide and capsular carbohydrate. J Bac **188**(9): 3273-3279.
- Kaniuk, N. A., M. A. Monteiro, C. T. Parker and C. Whitfield (2002). Molecular diversity of the genetic loci responsible for lipopolysaccharide core oligosaccharide assembly within the genus *Salmonella*. Mol. Microbiol **45**: 1305-1381.
- Karlyshev, A. V., O. L. Champion, C. Churcher, J. R. Brisson, H. C. Jarrell, M. Gilbert, D. Brochu, F. St Michael, J. Li, W. Wakarchuk, I. Goodhead, M. Sanders, K. Stevens, B. White, J. Parkhill, B. W. Wren and C. M. Szymanski (2005a). Analysis of *Campylobacter jejuni* capsular loci reveals multiple mechanisms for the generation of structural diversity and the ability to form complex heptoses. Molecular Microbiology **55**(1): 90-103.
- Karlyshev, A. V., O. L. Champion, G. W. P. Joshua and B. W. Wren (2005). The polysaccharide capsule of *Campylobacter jejuni*. Campylobacter. Molecular and Cellular Biology. J. M. Ketley and M. E. Konkel, Horizon Biosciences: 249-273.
- Karlyshev, A. V., P. H. Everest, S. Cawthraw, D. G. Newell and B. W. Wren (2004). The *Campylobacter jejuni* general glycosylation system is important for attachment to human epithelial cells and in the colonisation of chicks. Microbiology (Reading) **150**: 1957-1964.
- Karlyshev, A. V., J. M. Ketley and B. W. Wren (2005b). The *Campylobacter jejuni* glycome. FEMS Microbiol Reviews **29**: 377-390.
- Karlyshev, A. V., D. Linton, N. A. Gregson, A. J. Lastovica and B. W. Wren (2000). Genetic and Biochemical evidence of a *Campylobacter jejuni* capsular polysaccharide that accounts for Penner serotype specificity. Molecular Microbiology **35**(3): 529-514.
- Karlyshev, A. V., D. Linton, N. A. Gregson, A. J. Lastovica and B. W. Wren (2000). Genetic and biochemical evidence of a *Campylobacter jejuni* capsular polysaccharide that accounts for Penner serotype specificity. Mol Microbiol **35**(3): 529-41.
- Karlyshev, A. V., D. Linton, N. A. Gregson and B. W. Wren (2002). A novel paralogous gene family involved in phase-variable flagella-mediated motility in *Campylobacter jejuni*. Microbiology **148**(Pt 2): 473-80.
- Karlyshev, A. V., M. V. McCrossan and B. W. Wren (2001). Demonstration of polysaccharide capsule in *Campylobacter jejuni* using electron microscopy. Infect Immun **69**(9): 5921-4.
- Karlyshev, A. V. and B. W. Wren (2001a). Detection and initial characterisation of novel capsular polysaccharide among diverse *Campylobacter jejuni* strains using alcian blue dye. J.Clin Microbiol

- 39(1): 279-284.**
- Ketley, J. M. (1997). Pathogenesis of enteric infection by *Campylobacter*. Microbiology (Reading) **143**: 5-21.
- Klena, J. D., S. A. Gray and M. E. Konkel (1998). Cloning, sequencing and characterisation of the lipopolysaccharide biosynthetic enzyme heptosyltransferase I gene (*waaC*) from *Campylobacter jejuni* and *Campylobacter coli*. Gene **222**: 177-185.
- Knirel, Y. A., B. Lindner, E. V. Vinogradov, N. A. Kocharova, S. N. Senchenkova, R. Z. Shaikhutdinova, S. V. Dentovskaya, N. K. Fursova, I. V. Bakhteeva, G. M. Titareva, S. V. Balakhonov, O. Holst, T. A. Gremyakova, G. B. Pier and A. P. Anisimov (2005). Temperature dependent variations and intraspecies diversity of the structure of the lipopolysaccharide of *Yersinia pestis*. Biochemistry **44**: 1731-1743.
- Koga, M., C. W. Ang, N. Yuki, B. C. Jacobs, P. Herbrink, F. G. van der Meche, K. Hirata and P. A. van Doorn (2001). Comparative study of preceding *Campylobacter jejuni* infection in Guillain-Barre syndrome in Japan and The Netherlands. J Neurol Neurosurg Psychiatry **70(5)**: 693-5.
- Konkel, M. E., S. G. Garvis, S. L. Tipton, D. E. Anderson and W. Cieplak (1997). Identification and molecular cloning of a gene encoding a fibronectin-binding protein (CadF) from *Campylobacter jejuni*. Mol. Microbiol. **24(5)**: 953-963.
- Konkel, M. E., B. J. Kim, V. Rivera-Amill and S. G. Garvis (1999). Bacterial secreted proteins are required for the internalization of *Campylobacter jejuni* into cultured mammalian cells. Mol Microbiol **32(4)**: 691-701.
- Konkel, M. E., J. D. Klena, V. Rivera-Amill, M. R. Monteville, D. Biswas, B. Raphael and J. Mickelson (2004). Secretion of virulence proteins from *Campylobacter jejuni* is dependent on a functional flagellar export apparatus. JBAC **186(11)**: 3296-3303.
- Leonard II, E. E., T. Takata, M. J. Blaser, S. Falkow, L. S. Tompkins and E. C. Gaynor (2003). Use of an open-reading frame-specific *Campylobacter jejuni* DNA microarray as a new genotyping tool for studying epidemiologically related isolates JID **187**: 691-694.
- Leonard II, E. E., L. S. Tompkins, S. Falkow and I. Nachamkin (2004). Comparison of *Campylobacter jejuni* isolates implicated in Guillain-Barré Syndrome. Infection & Immunity. **72(2)**: 1199-1203.
- Linton, D., E. Allan, A. V. Karlyshev, A. D. Cronshaw and B. Wren (2002). Identification of N-acetylgalactosamine-containing glycoproteins PEB3 and CgpA in *Campylobacter jejuni*. Mol Microbiol **43(2)**: 497-508.
- Linton, D., M. Gilbert, P. G. Hitchen, A. Dell, H. R. Morris, W. W. Wararchuk, N. A. Gregson and B. W. Wren (2000b). Phase variation of a beta-1.3 galactosyltransferase involved in generation of the GM<sub>1</sub> - like lipooligosaccharide of *Campylobacter jejuni*. Molecular Microbiology **37(3)**: 501-514.
- Linton, D., A. V. Karlyshev, P. G. Hitchen, H. R. Morris, A. Dell, N. A. Gregson and B. W. Wren (2000a). Multiple N-acetyl neuraminic acid synthetase (*neuB*) genes in *Campylobacter jejuni*: Identification and characterisation of the gene involved in sialylation of lipooligosaccharide. Molecular Microbiology **35(5)**: 1120-1134.
- Lior, H., D. L. Woodward, J. A. Edgar, L. J. Laroche and P. Gill (1982).

- Serotyping of *Campylobacter jejuni* by slide agglutination based on heat-labile antigenic factors. *J Clin Micro* **15**(5): 761-8.
- MacCallum, A. J., D. Harris, G. Haddock and P. H. Everest (2006). *Campylobacter jejuni*-infected human epithelial cell lines vary in their ability to secrete interleukin-8 compared to *in vitro*-infected primary human intestinal tissue. *Microbiology* **152**(12): 3661-5.
- Mackinnon, F. G., A. D. Cox, J. S. Plested, C. M. Tang, K. Makepeace, P. A. Coull, J. C. Wright, R. Chalmers, D. W. Hood, J. C. Richards and E. R. Moxon (2002). Identification of a gene (*lpt3*) required for the addition of phosphoethanolamine to the lipopolysaccharide inner core of *Neisseria meningitidis* and its role in mediating susceptibility to bacterial killing and opsonophagocytosis. *Mol. Microbiol* **43**: 931-43.
- Maiden, M. C., J. A. Bygraves, E. Feil, G. Morelli, J. E. Russell, R. Urwin, Q. Zhang, J. Zhou, K. Zurth, D. A. Caugant, I. M. Feaver and M. Achtman (1998). Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *PNAS* **95**: 3140-5.
- Månsson, M., D. W. Hood, E. R. Moxon and E. K. H. Schweda (2003). Structural diversity in lipopolysaccharide expression in nontypeable *Haemophilus influenzae*. *Eur.J.Biochem* **270**: 610-624.
- Marchant, J., B. Wren and J. Ketley (2002). Exploiting genome sequence: predictions for mechanisms of *Campylobacter* chemotaxis. *Trends Microbiol* **10**(4): 155-9.
- McNally, D. J., H. C. Jarrell, J. Li, N. H. Kheiu, E. Vinogradov, C. M. Szymanski and J. R. Brisson (2005). The HS:1 serostrain of *Campylobacter jejuni* has a complex teichoic acid-like capsular polysaccharide with nonstoichiometric fructofuranose branches and O-methyl phosphoramidate groups. *FEBS Jour* **272**: 4407-4422.
- McSweeney, E. and R. I. Walker (1986). Identification and characterization of two *Campylobacter jejuni* adhesins for cellular and mucous substrates. *Infect. Immun.* **53**(1): 141-148.
- Meinke, A., T. Henics and E. Nagy (2004). Bacterial genomes pave the way to novel vaccines *Curr Opin Microbiol* **7**(3): 314-20.
- Michael, F. S., C. M. Szymanski, J. Li, K. H. Chan, N. H. Khieu, S. Larocque, W. W. Wakarchuk, J.-R. Brisson and M. A. Monteiro (2002). The structures of the lipooligosaccharide and capsule polysaccharide of *Campylobacter jejuni* genome sequenced strain NCTC 11168. *European Journal of Biochemistry* **269**(21): 5119-5136.
- Millar, L. (2003). Diversity and function of the Lipooligosaccharide Biosynthesis genes from *Campylobacter jejuni*. *Dept of Genetics, University of Leicester. PhD Thesis*: 292.
- Miller, W. G. and R. E. Mandrell (2005). Methods for epidemiological analysis of *Campylobacter jejuni*. *Campylobacter: Molecular and cellular biology*. J. M. Ketley and M. E. Konkel. London, Horizon Biosciences.
- Misawa, N., K. Kawashima, F. Kondo, B. M. Allos and M. J. Blaser (2001). DNA diversity of the *wla* gene cluster among serotype HS:19 and non-HS:19 *Campylobacter jejuni* strains. *J Endotoxin Res* **7**(5): 349-58.
- Mockler, T. C. and J. R. Ecker (2005). Applications of DNA tiling arrays for whole-genome analysis. *Genomics* **85**: 1-15.
- Moen, B., A. Oust, Ø. Langsrud, N. Dorrell, G. L. Marsden, J. Hinds, A.

- Kohler, B. W. Wren and K. Rudi (2005). Explorative multifactor approach for investigating global survival mechanisms of *Campylobacter jejuni* under environmental conditions. Applied and Environmental Microbiology **71**(4): 2086-2094.
- Møller, A. K., M. P. Leatham, T. Conway, P. J. M. Nuijten, L. A. M. de Haan, K. A. Krogfelt and P. S. Cohen (2003). An *Escherichia coli* MG1655 lipopolysaccharide deep-rough core mutant grows and survives in mouse cecal mucus but fails to colonize the mouse large intestine. Infection & Immunity. **71**(4): 2142-2152.
- Moran, A. P. (1995). Biological and serological characterization of *Campylobacter jejuni* lipopolysaccharides with deviating core and lipid A structures. Fems Immunology and Medical Microbiology **11**(2): 121-130.
- Moran, A. P. (1997). Structure and conserved characteristics of *Campylobacter jejuni* lipopolysaccharides. Journal of Infectious Diseases **176**(S2): S115-S121.
- Moran, A. P., H. Annuk and M. M. Prendergast (2005). Antibodies induced by ganglioside-mimicking *Campylobacter jejuni* lipooligosaccharide recognise epitopes at nodes of Ranvier. J Neuroimmun. **165**: 179-185.
- Moran, A. P., B. J. Appelmelk and G. O. Aspinall (1996b). Molecular mimicry of host structures by lipopolysaccharides of *Campylobacter* and *Helicobacter* spp: Implications in pathogenesis. Journal of Endotoxin Research **3**(6): 521-531.
- Nachamkin, I. (2001). *Campylobacter* Enteritis and the Guillain-Barre Syndrome. Curr Infect Dis Rep **3**(2): 116-122.
- Nachamkin, I. (2002a). Chronic effects of *Campylobacter* infection. Microbes and Infection **4**: 399-403.
- Nachamkin, I., B. M. Allos and T. Ho (1998). *Campylobacter* species and Guillain-Barre syndrome. Clin Microbiol Rev **11**(3): 555-67.
- Nachamkin, I., J. Liu, M. Li, H. Ung, A. P. Moran, M. M. Prendergast and K. Sheikh (2002b). *Campylobacter jejuni* from patients with Guillain Barré Syndrome preferentially expresses a GD<sub>1a</sub> -like epitope. Infection & Immunity. **70**(9): 5299-5303.
- Nishino, K., F. F. Hsu, J. Turk, M. J. Cromie, M. M. S. M. Wösten and E. A. Groisman (2006). Identification of the lipopolysaccharide modifications controlled by the *Salmonella* PmrA.PmrB system mediating resistance to Fe(III) and Al(III). Mol Micro **61**(3): 645-654.
- Oldfield, N. J. (2000). The genetic basis of lipopolysaccharide biosynthesis in *Campylobacter jejuni*. Department of Genetics. Leicester, University of Leicester.
- Oldfield, N. J., A. P. Moran, L. A. Millar, M. M. Prendergast and J. M. Ketley (2002). Characterization of the *Campylobacter jejuni* heptosyltransferase II gene, waaF, provides genetic evidence that extracellular polysaccharide is lipid A core independent. Journal of Bacteriology **184**(8): 2100-2107.
- Oldfield, N. J., A. P. Moran, L. A. Millar, M. M. Prendergast and J. M. Ketley (2002). Characterization of the *Campylobacter jejuni* heptosyltransferase II gene, waaF, provides genetic evidence that extracellular polysaccharide is lipid A core independent. J Bacteriol **184**(8): 2100-7.

- On, S. (2005). Taxonomy, Phylogeny and Methods for Identification of *Campylobacter* species. Campylobacter: Molecular and Cellular Biology. J. M. Ketley and M. E. Konkel, Horizon Biosciences: 13-42.
- Painter, G. R., R. Grunwald and B. Roth (1988). Interaction of the Antifolate Antibiotic Trimethoprim with Phosphatidylcholine Membranes: A  $^{13}\text{C}$  and  $^{31}\text{P}$  Nuclear Magnetic Resonance Study. Molecular Pharmacology **33**: 551-558.
- Palmer, C., E. M. Bik, M. B. Eisen, P. B. Eckburg, T. R. Sana, P. K. Wolber, D. A. Relman and P. O. Brown (2006). Rapid quantitative profiling of complex microbial populations. Nucleic Acids Res **34**(1): e5.
- Pandya, G., M. Holmes, Y. Bai, B. Verratti, P. Jarrahi, V. K., K. Moffat Saeed, S. Sunkara, L. Lindler, R. Fleishmann and S. Peterson (2006). Whole genome resequencing and SNP genotyping of category A biodefense agent *Francisella tularensis*. ASM 2006.
- Parker, C. T., S. T. Horn, M. Gilbert, W. G. Miller, D. L. Woodward and R. E. Mandrell (2005). Comparison of *Campylobacter jejuni* lipooligosaccharide biosynthesis loci from a variety of sources. J Clin Micro **43**(6): 2771-2781.
- Parker, C. T., B. Quiñones, W. G. Miller, S. T. Horn and R. E. Mandrell (2006). Comparative genomic analysis of *Campylobacter jejuni* strains reveals diversity due to genomic elements similar to those present in *C.jejuni* RM1221. J Clin Micro **44**(11): 4125-4135.
- Parkhill, J., G. Dougan, K. D. James, N. R. Thomson, D. Pickard and J. Waln (2001). Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi CT18. Nature **413**: 848-852.
- Parkhill, J., B. W. Wren, K. Mungall, J. M. Ketley, C. Churcher, D. Basham, T. Chillingworth, R. M. Davies, T. Feltwell, S. Holroyd, K. Jagels, A. V. Karlyshev, S. Moule, M. J. Pallen, C. W. Penn, M. A. Quail, M. A. Rajandream, K. M. Rutherford, A. H. van Vliet, S. Whitehead and B. G. Barrell (2000). The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hypervariable sequences. Nature **403**(6770): 665-8.
- Pearson, B., C. Pin, J. Wright, K. I'Anson, T. Humphrey and J. Wells (2003). Comparative genome analysis of *Campylobacter jejuni* using whole genome microarrays. FEBS Letts **554**: 224-230.
- Pei, Z., C. Burucoa, B. Grignon, s. Baqar, X. Huang, D. J. Kopecko, A. L. Bourgeois, J. Fauchere and M. J. Blaser (1998). Mutation in the *peb1A* locus of *Campylobacter jejuni* reduces interactions with epithelial cells and intestinal colonisation of mice. Infection & Immunity. **66**(3): 938-943.
- Penner, J. L. and G. Aspinall (1997). Diversity of lipopolysaccharide structures in *Campylobacter jejuni*. JID **176**(Suppl 2): S135-8`.
- Penner, J. L. and J. N. Hennessy (1980). Passive hemagglutination technique for serotyping *Campylobacter fetus* subsp. *jejuni* on the basis of soluble heat-stable antigens. J. Clin. Microbiol. **12**(6): 732-737.
- Perna, N. T., G. Plunkett, V. Burland, B. Mau, J. D. Glasner, D. J. Rose, G. F. Meyhew, P. S. Evans, J. Gregor, H. A. Kirkpatrick, G. Pósfal, J. Hackett, S. Klink, A. Boutin, Y. Shao, L. Miller, E. J. Grotbeck, N. Wayne Davis, A. Lim, E. T. Dimslanda, K. D. Potamousis, J. Apodaca, T. S. Anantharaman, J. Lin, G. Yen, D. C. Schwartz, R. A. Welch and

- F. R. Blattner (2001). Genomic sequence of enterohaemorrhagic *Escherichia coli* O157:H7. Nature **409**: 529-33.
- Phongsisay, V., V. N. Perera and B. N. Fry (2006). Exchange of lipooligosaccharide synthesis genes creates potential Guillain-Barré syndrome-inducible strains of *Campylobacter jejuni*. Infection & Immunity **74**(2): 1368-1372.
- Phongsisay, V., V. N. Perera and B. N. Fry (2007). Expression of the *htrB* gene is essential for responsiveness of *Salmonella typhimurium* and *Campylobacter jejuni* to harsh environments. Microbiology **153**: 254-262.
- Pickett, C. L. and R. B. Lee (2005). Cytolethal distending toxin. Campylobacter: Molecular and Cellular Biology. J. M. Ketley and M. E. Konkel, Horizon Bioscience: 385-396.
- Pier, G. B. (2000). Peptides, *Pseudomonas aeruginosa*, polysaccharides and lipopolysaccharides - player in the predicament of cystic fibrosis patients. Trends in Microbiology **8**(6): 247-250.
- Poly, F., D. Threadgill and A. Stintzi (2004). Identification of *Campylobacter jejuni* ATCC 43431 - specific genes by whole microbial genome comparisons. JBAC **186**: 4781-4795.
- Poly, F., D. Threadgill and A. Stintzi (2005). Genomic diversity in *Campylobacter jejuni*: Identification of *C.jejuni* 81-176 specific genes. J.Clin Microbiol **43**(5): 2330-2338.
- Prendergast, M. M., D. R. Tribble, s. Baqar, D. A. Scott, J. A. Ferris, R. I. Walker and A. P. Moran (2004). In vivo phase variation and serologic response to lipooligosaccharide of *Campylobacter jejuni* in experimental human infection. Infection & Immunity **72**(2): 916-922.
- Raetz, C. R. H. and C. Whitfield (2002). Lipopolysaccharide endotoxins. Annu. Rev. Biochem. **71**: 635-700.
- Rajakumar, K., J. Shafi, R. J. Smith, R. A. Stabler, P. W. Andrew, D. Modha, G. Bryant, P. Monk, J. Hinds, P. D. Butcher and M. R. Barer (2004). Use of genome level-informed PCR as a new investigational approach for analysis of outbreak-associated *Mycobacterium tuberculosis* isolates. J Clin Micro **42**(5): 1890-6.
- Regué, M., N. Climent, N. Abitiu, N. Coderch, M. S., L. Izquierdo, M. Altarriba and J. M. Tomas (2001). Genetic characterisation of the *Klebsiella pneumoniae waa* gene cluster, involved in core lipopolysaccharide biosynthesis. J Bac **183**(12): 3564-73.
- Ribot, E. M., C. Fitzgerald, K. Kubota, B. Swaminathan and T. J. Barrett (2001). Rapid pulsed-field gel electrophoresis protocol for subtyping of *Campylobacter jejuni*. J Clin Micro **39**: 1889-94.
- Ridley, K. A., J. D. Rock, Y. Li and J. M. Ketley (2006). Heme utilization in *Campylobacter jejuni*. J Bac **188**(22): 7862-75.
- Sampathkumar, B., S. Napper, C. D. Carrillo, P. Willson, E. N. Taboada, J. H. E. Nash, A. A. Potter, L. A. Babuik and B. J. Allan (2006). Transcriptional and translational expression patterns associated with immobilized growth of *Campylobacter jejuni*. Microbiology (Reading) **152**: 567-577.
- Saunders, N. J., A. C. Jeffries, J. F. Peden, D. W. Hood, H. Tettelin, R. Rappuoli and E. R. Moxon (2000). Repeat-associated phase variable genes in the complete genome sequence of *Neisseria meningitidis*

- strain MC58. Mol Microbiol **37**(1): 207-15.
- Schmidt, H. and M. Hensel (2004). Pathogenicity islands in bacterial pathogenesis. Clin Micro Rev **17**(1): 14-56.
- Schnaitman, C. A. (1971). Solubilization of the cytoplasmic membrane of *Escherichia coli* by Triton X-100. J Bacteriol **108**(1): 545-52.
- Schnaitman, C. A. and J. D. Klena (1993). Genetics of lipopolysaccharide biosynthesis in enteric bacteria. Microbiol. Rev. **57**: 655-682.
- Schröder, S., J. Weber and H. Paul (2001). 50 nucleotide long probes on microarrays enable high signal intensity and high specificity. Genomic Discovery, MWG AG Biotech Application note 14.
- Sergeev, N., D. Volokhov, V. Chizhikov and A. Rasooly (2004). Simultaneous Analysis of Multiple Staphylococcal Enterotoxin genes by an Oligonucleotide Microarray Assay. J Clin Micro **42**(5): 2134-2143.
- Skirrow, M. B. (1977). *Campylobacter* enteritis: a "new" disease. British Medical Journal **2**: 9-11.
- Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J Mol Biol **98**(3): 503-17.
- St. Michael, F., C. M. Szymanski, J. Li, K. H. Chan, N. H. Khieu, S. Laroque, W. W. Wararchuk, J.-R. Brisson and M. A. Monteiro (2002). The structures of the lipooligosaccharide and capsule polysaccharide of *Campylobacter jejuni* sequenced strain NCTC 11168. Eur.J.Biochem **269**: 5119-36.
- Stabler RA, Marsden GL, Witney AA, Li Y, Bentley SD, Tang CM and Hinds J (2005). Identification of pathogen-specific genes through microarray analysis of pathogenic and commensal *Neisseria* species. . Microbiology **151**(9): 2907-22.
- Steeghs, L., H. de Cock, E. Evers, B. Zomer, J. Tommassen and P. van der Ley (2001). Outer membrane composition of a lipopolysaccharide-deficient *Neisseria meningitidis* mutant. EMBO J **20**(24): 6937-6945.
- Steeghs, L., R. den Hartog, A. den Boer, B. Zomer, P. Roholl and P. van der Ley (1998). Meningitis bacterium is viable without endotoxin. Nature **392**: 449-450.
- Stintzi, A. (2003). Gene expression profile of *Campylobacter jejuni* in response to growth temperature variation. JBAC **185**(6): 2009-2016.
- Stintzi, A., D. Marlow, K. Palyada, H. Naikare, R. Panciera, L. Whitworth and C. Clarke (2005). Use of genome-wide expression profiling and mutagenesis to study the intestinal lifestyle of *Campylobacter jejuni*. Infection & Immunity. **73**(3): 1797-1810.
- Summers, D. K. and H. L. Withers (1990). Electrotransfer: direct transfer of bacterial plasmid DNA by electroporation. NAR **18**(8): 2192.
- Sundaram, A. P., L. Pitts, K. Muhammad, M. Betenbaugh, R. W. Woodard and W. F. Vann (2004). Characterisation of *N*-acetylneuraminic acid synthase isoenzyme 1 from *Campylobacter jejuni*. Biochem.J. **383**: 83-89.
- Sunshine, M. G., B. W. Gibson, J. J. Engstrom, W. A. Nichols, B. D. Jones and M. A. Apicella (1997). Mutation of the *htrB* gene in a virulent *Salmonella typhimurium* strain by intergeneric transduction: strain construction and phenotypic characterization. J. Bacteriol. **179**: 5521-5533.
- Swiderek, H., H. Claus, M. Frosch and U. Vogel (2005). Evaluation of



- custom-made DNA microarrays for multilocus sequence typing of *Neisseria meningitidis*. Int Jour Med Micro **295**: 39-45.
- Szymanski, C. M., D. H. Burr and P. Guerry (2002). Campylobacter protein glycosylation affects host cell interactions. Infect Immun **70**(4): 2242-4.
- Szymanski, C. M., M. King, M. Haardt and G. D. Armstrong (1995). *Campylobacter jejuni* motility and invasion of Caco-2 cells. Infection & Immunity. **63**(11): 4295-4300.
- Szymanski, C. M., S. M. Logan, D. Linton and B. W. Wren (2003b). *Campylobacter*- a tale of two protein glycosylation systems. Trends in Microbiology **11**(5): 233-238.
- Szymanski, C. M., F. St Michael, H. C. Jarrell, J. J. Li, M. Gilbert, S. Larocque, E. Vinogradov and J. R. Brisson (2003a). Detection of conserved N-linked glycans and phase-variable lipooligosaccharides and capsules from *Campylobacter* cells by mass spectrometry and high resolution magic angle spinning NMR spectroscopy. Journal of Biological Chemistry **278**(27): 24509-24520.
- Szymanski, C. M. and B. W. Wren (2005). Protein glycosylation in bacterial mucosal pathogens. Nature Rev Micro **3**: 225 - 237.
- Szymanski, C. M., R. Yao, C. P. Ewing, T. J. Trust and P. Guerry (1999). Evidence for a system of general protein glycosylation in *Campylobacter jejuni*. Molecular Microbiology **32**(5): 1022-1030.
- Szymanski, C. M., R. Yao, C. P. Ewing, T. J. Trust and P. Guerry (1999). Evidence for a system of general protein glycosylation in *Campylobacter jejuni*. Mol Microbiol **32**(5): 1022-30.
- Taboada, E. N., R. R. Acedillo, C. D. Carrillo, W. A. Findlay, D. T. Medeiros, O. L. Mykytczuk, M. J. Roberts, C. Alexander Valencia, J. M. Farber and J. H. E. Nash (2004). Large scale comparative genomics meta-analysis of *Campylobacter jejuni* isolates reveals low level of genomic plasticity. J Clin Micro **42**(10): 4566-4576.
- Tan, L. and C. Darby (2005). *Yersinia pestis* is viable with endotoxin composed of only lipid A. J Bac **187**(18): 6599-6600.
- Tomioka, K., M. Peredelchuk, D. Volokhov, A. Selvapandian, K. Stabler, J. Mellquist-Riemenschneider, V. Chizhikov, G. Kaplan, H. Nakhasi and R. Duncan (2005). A multiplex polymerase chain reaction microarray assay to detect bioterror pathogens in blood. JMD **7**(4): 486-494.
- Tong, H. H., Y. Chen, M. James, J. Van Deusen, D. B. Welling and T. F. DeMaria (2001). Expression of cytokine and chemokine genes by human middle ear epithelial cells induced by formalin-killed *Haemophilus influenzae* or its lipooligosaccharide *htrB* and *rfaD* mutants. Infection & Immunity. **69**(6): 3678-3684.
- Tsai, C.-M. and C. E. Frasch (1982). A sensitive silver stain for detecting lipopolysaccharides in polyacrylamide gels. Anal Biochem **119**: 115-119.
- Tzeng, Y., A. Datta, V. Kolli, R. Carslon and D. Stephens (2002). Endotoxin of *Neisseria meningitidis* composed of only intact lipid A: inactivation of the meningococcal 3-deoxy-D-manno-octulosonic acid transferase. J Bac **184**(9): 2379-2388.
- van Belkum, A. (2002). Molecular typing of micro-organisms: At the centre of diagnostics, genomics and pathogenesis of infectious diseases? Journal of Medical Microbiology **51**(1): 7-10.

- van Belkum, A., N. van den Braak, P. Godschalk, W. Ang, B. Jacobs, M. Gilbert, W. Wakarchuk, H. Verburgh and H. Endtz. (2001). A *Campylobacter jejuni* gene associated with immune-mediated neuropathy. Nature Med **7**(7): 752-3
- Van der Ley, P., L. Steeghs, F. J. Hamstra, J. Ten Hove, B. Zomer and L. Van Alphen (2001). Modification of lipid A biosynthesis in *Neisseria meningitidis* *lpxL* mutants: influence on lipopolysaccharide structure, toxicity, and adjuvant activity. Infection & Immunity **69**(10): 5981-5990.
- Van Ijperen, C. and N. A. Saunders (2004). Microarrays for bacterial typing: realistic hope or holy grail? Methods in Molecular Biology, Humana Press Inc. **266**: 213-227.
- van Leeuwen, W. B., C. Jay, S. Snijders, N. Durin, B. Lacroix, H. A. Verbrugh, M. C. Enright, A. Troesch and A. van Belkum (2003). Multilocus sequence typing of *Staphylococcus aureus* with DNA array technology. J Clin Micro **41**(7): 3323-3326.
- van Vliet, A. H. M. and J. M. Ketley (2001). Pathogenesis of enteric *Campylobacter* infection. J. Appl. Microbiol **90**: 45S-56S.
- van Vliet, A. H. M., J. M. Ketley, S. F. Park and C. W. Penn (2002). The role of iron in *Campylobacter* gene regulation, metabolism and oxidative stress defense. FEMS Microbiol Rev **26**: 173-186.
- van Vliet, A. H. M., A. C. Wood, J. Henderson, K. Wooldridge and J. M. Ketley (1998). Genetic manipulation of enteric *Campylobacter* species. Methods in microbiol. **27**: 408-419.
- Vimr, E., C. Lichtensteiger and S. Steenbergen (2000). Sialic acid metabolism's dual function in *Haemophilus influenzae*. Molecular Microbiology **36**(5): 1113-1123.
- Vogel, U. and M. Frosch (1999). Mechanisms of neisserial serum resistance. Molecular Microbiology **32**(6): 1133-1139.
- Vora, G. J., C. E. Meador, D. A. Stenger and J. D. Andreadis (2004). Nucleic acid amplification strategies for DNA-based pathogen detection. Appl. Environ. Microbiol **70**(4): 3047-3054.
- Vorachek-Warren, M. K., S. Ramirez, R. J. Cotter and C. R. H. Raetz (2002). A triple mutant of *Escherichia coli* lacking secondary acyl chains on lipid A. J Biol Chem **277**(16): 14194-14205.
- Wacker, M., D. Linton, P. G. Hitchen, M. Nita-Lazar, S. M. Haslam, S. J. North, M. Panico, H. R. Morris, A. Dell, B. W. Wren and M. Aebi (2002). N-linked glycosylation in *Campylobacter jejuni* and its functional transfer into *E.coli*. Science **298**: 1790-1793.
- Wang, G., Z. Ge, D. A. Rasko and D. E. Taylor (2000). Lewis antigens in *Helicobacter pylori*: biosynthesis and phase variation. Molecular Microbiology **36**(6): 1187-1196.
- Wassenaar, T. M., N. M. Bleumink-Pluym and B. A. van der Zeijst (1991). Inactivation of *Campylobacter jejuni* flagellin genes by homologous recombination demonstrates that *flaA* but no *flaB* is required for invasion. EMBO J **10**(8): 2055-61.
- Whitehouse, C. A., P. B. Balbo, E. C. Pesci, D. L. Cottle, P. M. Mirabito and C. L. Pickett (1998). *Campylobacter jejuni* cytolethal distending toxin causes a G<sub>2</sub> phase cell cycle block. Infection & Immunity **66**(5): 1934-1940.
- Whitfield, C., N. Kaniuk and E. Fridrich (2003). Molecular insights into the

- assembly and diversity of the outer core oligosaccharide in lipopolysaccharides from *Escherichia coli* and *Salmonella*. J Endo Res **9**(4): 244-249.
- Witney, A. A., G. L. Marsden, M. T. G. Holden, R. A. Stabler, S. E. Husain, J. K. Vass, P. D. Butcher, J. Hinds and J. A. Lindsay (2005). Design, validation and application of a seven-strain *Staphylococcus aureus* PCR product microarray for comparative genomics. Applied and Environmental Microbiology **71**(11): 7504-7514.
- Wood, A. C., N. J. Oldfield, C. A. O'Dwyer and J. M. Ketley (1999). Cloning, mutation and distribution of a putative lipopolysaccharide biosynthesis locus in *Campylobacter jejuni*. Microbiology **145** ( Pt 2): 379-88.
- Wooldridge, K. G., J. A. Morrissey and P. H. Williams (1992). Transport of ferric-aerobactin into the periplasm and cytoplasm of *Escherichia coli* K12: role of envelope-associated proteins and effect of endogenous siderophores J Gen Microbiol. **138**(3): 597-603.
- Wooldridge, K. G., P. H. Williams and J. M. Ketley (1994). Iron-responsive genetic regulation in *Campylobacter jejuni*: Cloning and characterisation of a *fur* homolog. JBAC **176**(18): 5852-5856.
- Wosten, M. M., M. Boeve, W. Gastra and B. A. van der Zeijst (1998). Cloning and characterization of the gene encoding the primary sigma-factor of *Campylobacter jejuni*. FEMS Microbiol Lett **162**(1): 97-103.
- Wösten, M. M. S. M., M. Boeve, M. G. A. Koot, v. Nuenen and V. D. Zeijst (1998). Identification of *Campylobacter jejuni* promoter sequences J Bac **180**(3): 594-599.
- Wren, B. W., J. Henderson and J. M. Ketley (1994). A PCR-based strategy for the rapid construction of defined bacterial deletion mutants. Biotechniques **16**: 994-996.
- Wren, B. W., D. Linton, N. Dorrell and A. V. Karlyshev (2001). Post genome analysis of *Campylobacter jejuni*. Journal of Applied Microbiology **90**: 36S-44S.
- Yao, R. and P. Guerry (1996). Molecular cloning and site-specific mutagenesis of a gene involved in arylsulfatase production in *Campylobacter jejuni*. J Bac **178**(11): 3335-3338.
- Yao, R. J., D. H. Burr, P. Doig, T. J. Trust, H. Y. Niu and P. Guerry (1994). Isolation of motile and nonmotile insertional mutants of *Campylobacter jejuni* - the role of motility in adherence and invasion of eukaryotic cells. Mol. Microbiol. **14**(5): 883-893.
- Young, V. B. and L. S. Mansfield (2005). *Campylobacter* infection - Clinical context. Campylobacter. Molecular and Cellular Biology. J. M. Ketley and M. E. Konkel, Horizon Bioscience: 1-12.
- Yuki, N. (1997). Molecular mimicry between gangliosides and lipopolysaccharides of *Campylobacter jejuni* isolated from patients with Guillain Barré syndrome and Miller Fisher syndrome. J.Infect.Dis **176**(Suppl2): S150-3.
- Yuki, N. (2001). Infectious origins of, and molecular mimicry in, Guillain Barré and Fisher syndrome. Lancet Infect Dis **1**: 29-37.
- Yuki, N., N. Miyatani, S. Sato, M. Hirabayashi, M. Yamazaki, N. Yoshimura, Y. Hayashi and T. Miyatake (1992). Acute relapsing sensory neuropathy associated with IgM antibody against B-series gangliosides containing a GalNAc beta 1-4(Gal2-3 alpha NeuAc8 alpha NeuAc) beta 1

## Bibliography

---

- configuration. Neurology **42**: 686-9.
- Yuki, N., T. Taki, F. Inagaki, T. Kasama, M. Takahashi, K. Saito, S. Handa and T. Miyatake (1993). A bacterium lipopolysaccharide that elicits Guillain Barré syndrome has a GM1 ganglioside-like structure. J.Exp.Med. **178**: 1771-1775.