Genomic Diversity in Naturally Transformable *Streptococcus pneumoniae*

Dr Donald James Inverarity

BSc MBChB MRCP MSc DTMH DLSHTM FRCPath

A Thesis Submitted in Fulfilment of the Requirements for the Degree of PhD

Division of Infection and Immunity
Faculty of Biomedical and Life Sciences
University of Glasgow

February 2009

© Donald James Inverarity

Abstract

Infections due to *Streptococcus pneumoniae* (the pneumococcus) remain a substantial source of morbidity and mortality in both developing and developed countries despite a century of research and the development of effective therapeutic interventions (such as antibiotic therapy and vaccination). The ability of the pneumococcus to evade multiple classes of antibiotic through several genetically determined resistance mechanisms and its evasion of capsular polysaccharide based vaccines through serotype replacement and capsular switching, all reflect the extensive diversity and plasticity of the genome of this naturally transformable organism which can readily alter its genome in response to its environment and the pressures placed upon it in order to survive.

The purpose of this thesis is to investigate this diversity from a genome sequence perspective and to relate these observations to pneumococcal molecular epidemiology in a region of high biodiversity, the pathogenesis of certain disease manifestations and assess for a possible bacterial genetic basis for the pneumococcal phenotypes of, “carriage” and, “invasion.”

In order to do this, microarray comparative genomic hybridization (CGH) has been utilized to compare DNA from a variety of pneumococcal isolates chosen from 10 diverse serotypes and Multilocus Sequence Types and from clinically relevant serotypes and sequence types (particularly serotypes 3, 4 and 14 and sequence types ST9, ST246 and ST180)) against a reference, sequenced pneumococcal genome from an extensively investigated serotype 4 isolate – TIGR4.

Microarray comparison of the transcriptional profiles of several isolates has also been undertaken to compare gene expression from isolates of serotype 1 (ST227 and ST306) and serotype 3 (ST180) related to particular disease states and exposure of a multi-resistant pneumococcus to an antimicrobial (clarithromycin) commonly used to treat pneumococcal pneumonia.

Table of Contents

Abstract 2
List of Tables....... 7
List of Figures 9
Acknowledgements 12
Author's Declaration 14
Abbreviations 15

1 Introduction 18
1.1 Taxonomy of Streptococcus pneumoniae 18
1.2 Phenotypic identification and characteristics of Streptococcus pneumoniae 19
1.3 Genetic typing of Streptococcus pneumoniae 21
1.4 Pneumococcal Carriage 28
1.5 Manifestations of Pneumococcal Infection 31
1.6 Studies of the Genomic Diversity of Streptococcus pneumoniae 35
2 Materials and Methods 63
2.1 General Procedures 63
2.2 Microarray Protocols 67
2.3 Gene Sequencing and Multi Locus Sequence Typing 72
2.3.1 Gene Sequencing 72
2.3.2 Multi Locus Sequence Typing 73
2.4 Analysis of Genome Sequences Using the Artemis Comparison Tool 73
2.5 Quantitative Real Time Polymerase Chain Reaction 74
2.5.1 cDNA Synthesis 74
2.5.2 Quantitative Real Time Polymerase Chain Reaction Protocol 74
3 Microarray and Genome Sequencing Approaches to the Study of Pneumococcal Genomic Diversity 76
3.1 What is a Microarray? 76
3.2 Applications of Microarrays 77
3.3 Microarray Design 79
3.4 Microarray Analysis Methods 80
3.5 Microarray Validation Experiments 82
3.5.1 Validation of DNA CGH 82
3.6 Comparisons of Microarray DNA CGH results with Sequenced Pneumococcal Genomes 89
3.6.1 Choice of Diverse Sequenced Strains 89
3.6.2 Comparison of Microarray DNA CGH results with Genome Sequence Data. 89
3.6.3 Genomic Diversity Identified by Genome Sequence Data Alone 94
3.6.4 Using Pneumococcal Genome Sequences to Identify Gene Insertions 97
3.7 Discussion 100
3.7.1 Advantages and Disadvantages of a Genome Sequencing Approach to mapping Genomic Diversity compared to microarray based CGH investigations 100
3.7.2 The Distributed Genome Hypothesis and a Pneumococcal Supragenome 102
3.7.3 Microarray Limitations 102
4 Genomic Diversity Observed in Phenotypically Diverse Pneumococcal Isolates 104
4.1 Choice of Diverse Pneumococcal Isolates 104
4.2 Microarray Results 106
4.3 Discussion 106
4.3.1 Regions of Diversity in the TIGR4 and R6 Genomes 106
4.3.2 Diversity at Particular Genetic Loci 111
4.4 Conclusions 115
5 Genomic Diversity in a Multilocus Sequence Type Associated with Invasive Pneumococcal Disease (IPD) 116

5.1 Serotype 14 and ST9: associations with IPD 116
5.2 Choice of Isolates 116
5.3 Microarray Results 117
5.4 Discussion 117
5.4.1 Overall diversity in ST9 117
5.4.2 Regions of Diversity in the TIGR4 and R6 genomes 120
5.4.3 Diversity at Particular Genetic Loci 120
6 Genomic Diversity within a Multilocus Sequence Type Accounting for Invasive Pneumococcal Disease and Carriage 123
6.1 A Review of the Associations of Serotype 3 and ST180 with Pneumococcal Carriage or Invasive Disease 123
6.2 Serotype 3 ST180 Isolates Analysed by Microarray CGH 124
6.3 DNA CGH Comparison of Carriage and Invasive Isolates of ST180 125
6.4 RNA Expression Differences in Carriage and Invasive Isolates of ST180 127
6.4.1 Microarray results 127
6.4.2 Quantitative Real Time PCR results 129
6.5 Discussion 130
6.5.1 Overall genomic diversity in Serotype 3 ST180 isolates 130
6.5.2 A Carriage Genotype versus an Invasive Genotype for ST180 131
7 Genomic Diversity in Isolates of the Same Serotype and Multilocus Sequence Type Related to Clinical Manifestation and Outcome 140
7.1 Background 140
7.2 Epidemiological details for Serotype 4 ST246 Test Isolates 141
7.3 Clinical Manifestations and Outcomes for Serotype 4 ST246 Test Isolates 142
7.4 Bacteriophage induction 142
7.5 Microarray DNA CGH Results 142
7.6 Discussion 146
8 Genomic Diversity in Nosocomial Outbreaks of Pneumococcal Disease 150
8.1 Pneumococcal Outbreaks – Definition and Features 150
8.2 Pneumococcal Typing Methods and Limitations in Outbreak Investigations 152
8.2.1 Serotyping 152
8.2.2 Molecular typing 153
8.3 Background to Chosen Outbreaks 153
8.3.1 Serogroup 1 ST227 153
8.3.2 Serogroup 4 ST206 154
8.4 Microarray CGH analysis of Outbreak Isolates 154
8.4.1 CGH Results from Serogroup 1 Outbreak 155
8.4.2 Discussion of CGH Results for Serogroup 1 Outbreak 157
8.4.3 CGH Results from Serogroup 4 Outbreak 159
8.4.4 Discussion of CGH Results from the Serogroup 4 Outbreak 161
8.5 Discussion 162
8.5.1 Possible role for Microarrays in Public Health Outbreak Investigations 162
8.5.2 Genomic Diversity of Chosen Outbreak Related Strains 162
9 Genomic Diversity in a Paediatric Carriage Population in the Bolivian Amazon 163
9.1 Background 163
9.1.1 Reasons for this Study 163
9.1.2 Association of Pneumococcal Carriage and Acute Otitis Media 163
9.1.3 Issues regarding Pneumococcal conjugate vaccination in Latin America for the prevention of otitis media and carriage 164
9.1.4 Pneumococcal surveillance in Latin America 165
9.1.5 Location of Study 166
9.2 Materials and Methods 167
9.2.1 Specimen Collection, Storage and Transportation 168
9.2.2 Epidemiological Data Collection and Analysis 169

9.2.3 Ethical Considerations 169
9.3 Results 170
9.3.1 Epidemiological Data 170
9.3.2 Antibiotic Resistance 170
9.3.3 Serotyping of Bolivian Isolates 171
9.3.4 MLST of Bolivian Isolates 172
9.4 Discussion 174
9.4.1 Epidemiological Data 175
9.4.2 Implications for Otitis Media 175
9.4.3 Antibiotic Resistance 176
9.4.4 New MLST Profiles 179
9.4.5 New Insights into Existing Clonal Complexes 180
9.4.6 Common Serotypes and STs which comprise them 182
9.4.7 Serotype Distributions in Latin America and Bolivia 183
9.4.8 Implications for Conjugate Vaccine Implementation 187
9.4.9 Implications for Protein Vaccines 188
9.5 Conclusions 188
10 Genomic Diversity and Gene Expression in Specific Invasive Pneumococcal Disease Manifestations 189
10.1 Background 189
10.2 Serotype 1 Bacteraemic Pneumococcal Pneumonia with Parapneumonic Complications 190
10.2.1 Background 190
10.2.2 Choice of Isolates for Microarray Studies 193
10.2.3 DNA CGH Experiments 194
10.2.4 RNA Expression Experiments 194
10.2.5 Discussion Regarding a Genetic Basis for Pneumococcal Parapneumonic Complications 197
10.3 Pneumococcal Meningitis 202
10.3.1 Serotype 3 Association with Meningitis 202
10.3.2 Serotype 3 ST180 DNA CGH Hybridizations 202
10.3.3 Serotype 3 ST180 RNA Expression Experiments 205
10.3.4 Discussion regarding Serotype 3 Associated Pneumococcal Meningitis ..207
10.4 Pneumococcal Cerebral Abscess 212
10.4.1 Background 212
10.4.2 Serotype and MLST distribution of Cerebral Abscess Associated Pneumococci in Scotland 213
10.4.3 Serotype 3 ST180 DNA CGH Hybridizations 214
10.4.4 RNA Expression of Serotype 3 ST180 Cerebral Abscess Associated Isolates 214
10.4.5 Discussion regarding pneumococcal cerebral abscesses 215
10.5 Overall Discussion Regarding Gene Associations with Invasive Pneumococcal Disease Manifestations 216
11 Diversity of Pneumococcal Gene Expression in Response to an Antibiotic 220
11.1 Physicochemical Properties of Clarithromycin 220
11.2 Mechanism of Action of Clarithromycin 221
11.3 Resistance Mechanisms associated with Clarithromycin 221
11.4 Putative Effects of Clarithromycin at Subtherapeutic Concentrations 221
11.4.1 Anti-inflammatory and Immunomodulatory Effects 221
11.4.2 Anti-neoplastic Effects 222
11.4.3 Transcriptional modulation 222
11.4.4 Disruption of Quorum Sensing 223
11.4.5 Effects on Virulence Factors 223
11.5 Role of Clarithromycin in the Management of IPD 223

11.6 Experimental Design to Assess Pneumococcal Gene Expression in the Presence of Subtherapeutic Clarithromycin Concentrations 224
11.6.1 Growth of Strain South Africa 2507 for RNA Extraction 224
11.6.2 RNA Extraction from Strain South Africa 2507 224
11.6.3 Microarray Comparative Genomic Hybridization Analysis 225
11.6.4 Quantitative RT-PCR Validation of Microarray Data 225
11.7 Strain South Africa 2507 226
11.7.1 Antibiotic sensitivities 226
11.7.2 Calculation of Clarithromycin Minimum Inhibitory Concentration 226
11.7.3 Growth Curves for Strain South Africa 2507 in Brain Heart Infusion 227
11.7.4 Typing of Strain South Africa 2507 229
11.8 Microarray results 229
11.9 Quantitative Real Time PCR results 230
11.10 Discussion 231
12 Concluding Thoughts 241
13 References 243
14 List of Publications 303

List of Tables

Table 1-1 Common pneumococcal serotypes associated with IPD in various countries.....35
Table 1-2 Regions of diversity in the TIGR4 genome, size and putative virulence functions as determined by CGH in previous studies. 43
Table 2-1 Primer sequences used for MLST...73
Table 3-1 Normalized log ratios determined by Genespring GX 7.3.1 for genes used to determine an appropriate cut-off value. 82
Table 3-2 Sanger sequenced isolates which have been compared with microarray DNA CGH results. 89
Table 3-3 Comparison of size of regions of diversity identified by microarray CGH with the size of the same regions of diversity identified from sequence data.............................91
Table 4-1 Details of isolates of diverse serotype and MLST used in microarray DNA CGH experiments. 105
Table 4-2 Highly variable genes with multiple homologs compared with CGH results. ..112
Table 6-1 Background to serotype 3, ST180 isolates used in the microarray CGH experiments. 124
Table 6-2 Genes identified by Genespring GX 7.3.1. whose expression detected by microarray are either significantly upregulated or downregulated in invasive pneumococcal disease. 127
Table 6-3 Predictive strengths of the top 40 genes whose expression is more associated with invasive pneumococcal disease than carriage in serotype 3 ST180 isolates.............128
Table 6-4 Genes from the two regions SP0473-SP0478 and SP2159-SP2166 which code for components of phosphotransferase systems and which are variably present in the genomes of serotype 3, ST180 isolates. 131
Table 6-5 Genes identified by Genespring GX 7.3.1. which are involved in lactose metabolism. 134
Table 7-1 Basic epidemiological information about the source patients from which serotype 4, ST246 isolates were received. 141
Table 7-2 Brief clinical histories of the cases from which serotype 4, ST246 isolates were received. 142
Table 8-1 Pneumococcal serogroups which can cause disease outbreaks. 151
Table 8-2 Presentation of pneumococcal disease or carriage associated with outbreaks. .152
Table 9-1 Odds ratios for pneumococcal carriage risk factors estimated from univariate and multivariate models. 170
Table 9-2 Antibiotic sensitivity for the 54 optochin sensitive Bolivian isolates. 171
Table 9-3 MLST profiles for the newly discovered sequence types currently unique to carriage isolates from Bolivia. 173
Table 9-4 Serotype and geographical associations of sequence types identified in Bolivia, which have also been identified in other regions 174
Table 9-5 Serotypes accounting for IPD in various Latin American countries over 3 decades. 185
Table 9-6 Serotypes associated with pneumococcal carriage in various Latin American countries. 186
Table 10-1 Details of isolates used in microarray experiments in Chapter 10. 194
Table 10-2 Genes which are significantly up or down-regulated when isolates are grown to midlog under standardised conditions...195
Table 10-3 Genes identified as being associated with the occurrence of parapneumonic complications when grown to midlog in Brain Heart Infusion...196
Table 10-4 Comparison of Normalized Expression Ratios of genes differentially expressed when expression levels in CSF are compared to blood...206
Table 10-5 Comparison of the fold change differences in expression for genes identified as being downregulated in CSF in this study with results from similar work by Orihuela *et al* (Orihuela *et al*., 2004b). 206

Table 10-6 Comparison of results for genes identified as significantly up or down regulated in CSF with their fold change levels as detected in mouse blood by Orihuela *et al* (Orihuela *et al*., 2004b). 207
Table 10-7 Genes identified by Genespring GX 7.3.1. as upregulated or downregulated in cerebral abscess associated serotype 3 ST180 isolates compared to non cerebral abscess related serotype 3 ST180 isolates...214
Table 10-8 Genes identified by Genespring GX 7.3.1. as predictive of a cerebral abscess associated phenotype when compared to non cerebral abscess related serotype 3 ST180 isolates...215
Table 11-1 Comparison of the effects of sublethal erythromycin and clarithromycin on pneumococcal gene expression..237

List of Figures

Figure 1-1Schematic diagram of the pneumococcal genome. 39
Figure 1-2 Proportions of cases of hospitalised pneumococcal lobar pneumonia due to serotypes 1 and 3 documented in Glasgow over the 20th Century. 45
Figure 1-3 Annual number of cases of pneumonia (all types) reported in Glasgow (1920-1972). 45
Figure 1-4 Proportions of cases of hospitalised pneumococcal lobar pneumonia in Edinburgh due to serotypes 1 and 3 over the 20th Century. 46
Figure 1-5 Proportion of blood cultures growing pneumococci at Boston City Hospital due to serotype 1 and serotype 3 (1935-1974). 46
Figure 1-6 Number of episodes of invasive serotype 1 associated disease seen in Scotland. 47
Figure 1-7 MLST sequence types associated with serotype 1 capsule in Scotland causing invasive pneumococcal disease (2002-2006). 48
Figure 1-8 MLST sequence types associated with serotype 3 capsule in Scotland causing invasive pneumococcal disease (2002-2006). 51
Figure 1-9 Relationships of MLST sequence types constructed using eBURST version 3 of all isolates in the MLST database which express serotype 3 capsule. 52
Figure 1-10 Relationships of MLST sequence types constructed using eBURST version 3 of all isolates in the MLST database which express serotype 19A capsule. 55
Figure 1-11 Pie charts demonstrating the diversity of MLST sequence types (clonal complex or CC) which express serotype 19A capsule before and after pneumococcal conjugate vaccine introduction. 56
Figure 2-1 Results of microarray CGH dye swap experiments for isolate 06-1805 (ST227). 68
Figure 3-1 Comparison of normalization methods for CGH of Sample 03-4183...............81
Figure 3-2 Comparison of microarray CGH result with PCR results from the ST246 pneumococcal isolates used in Table 3-1. 83
Figure 3-3 An investigation into the accuracy of hybridization and diversity of genes coding for the pneumococcal pilus islet *rlrA* investigated by PCR. 85
Figure 3-4 Comparison of SP0464 in the TIGR4 and INV104B genomes using the Artemis Comparison Tool (ACT). 86
Figure 3-5 Comparison of TIGR4 gene expression from RNA extracted after growth in different batches of BHI grown on different days. 87
Figure 3-6 Comparison of TIGR4 gene expression from batches of RNA extracted from culture using the same batch of BHI and grown on the same day. 88
Figure 3-7 ACT comparison of the TIGR4 genome with the ATCC700669 genome.........90
Figure 3-8 Comparison of microarray CGH results for R6 genes with genome sequence data for the same genes. 93
Figure 3-9 ACT comparison of the spr0320-spr0323 genes in OXC141 (A), ATCC700669 (B) and INV200 (C). 95
Figure 3-10: Comparison of gene insertions not present in the TIGR4 genome but present in the ATCC700669 genome using ACT. 98
Figure 3-11 Comparison of PTS system genes in the TIGR4 genome with the ATCC700669 genome using ACT. 99
Figure 3-12 Comparison of the TIGR4 genome with the OXC141 genome using ACT demonstrating the insertion site of the phiOXC prophage. 100
Figure 4-1 Whole pneumococcal genome view of the DNA CGH results for the 10 chosen isolates of diverse sequence type generated by Genespring GX 7.3.1..............................106
Figure 4-2 Demonstration of a new region of diversity (SP0726 – SP0731) in the TIGR4 genome using Genespring GX 7.3.1. 108
Figure 4-3 Demonstration of a new region of diversity (SP2180 – SP2183) in the TIGR4 genome using Genespring GX 7.3.1. 109

Figure 4-4 Demonstration of a, “hot spot,” for recombination events (SP1309 to SP1337) in the TIGR4 genome using Genespring GX 7.3.1. 110
Figure 5-1 Comparison of diversity within genomes of serotype 14 associated ST9 isolates and non serotype 14 ST9 isolates using Genespring GX 7.3.1. 117
Figure 5-2 Location of a new region of diversity SP1933 – SP1936 identified in isolate 04-1548. 120
Figure 6-1 Comparison of diversity within genomes of serotype 3 associated ST180 invasive and carriage associated isolates using Genespring GX 7.3.1. 125
Figure 6-2 Serotype 3, ST180 microarray CGH results demonstrating the presence of the genes SP2159 to SP2166 (which code for phosphotransferase system genes) in the Dutch isolate 03-4156 and the Bolivian isolate 07-2838. 126
Figure 6-3 RT-PCR results comparing expression of SP0110 in the invasive isolate 00-3946 and the carriage isolate OXC141. 129
Figure 6-4 Rapid ID 32 Strep API results for three of the carriage associated ST180 isolates (07-2838, 03-4283 and 03-4156) for which microarray CGH demonstrated differences in the complement of PTS associated genes. 135
Figure 6-5 Results of Serotype 3 ST180 CGH studies superimposed onto sugar metabolism pathways which relate to phosphotransferase system genes. 136
Figure 6-6 Results of Serotype 3 ST180 gene expression studies superimposed onto sugar metabolism pathways which relate to phosphotransferase system genes. 137
Figure 6-7 Screen view of Genespring GX 7.3.1 demonstrating the maintenance of baseline expression of SP1722 in isolate 03-4156. 138
Figure 6-8 Screen view of Genespring GX 7.3.1 demonstrating that the expression of SP1884 which is involved in trehalose metabolism. 139
Figure 7-1 DNA CGH results for serotype 4, ST246 isolates. 143
Figure 7-2 PCR validation results for genes where are least one of the 6 ST246 isolates did not hybridize on the microarray. 145
Figure 7-3 Genomic diversity of the transcriptional regulator gene SP1130....................145
Figure 7-4 Demonstration of the relationship of ST246 to other closely related sequence types using e-BURST version 3 and the MLST database. 149
Figure 8-1 Comparison of CGH results for serotype 1 outbreak associated isolates.........155
Figure 8-2 Genelists created from CGH analysis of dye swap experiments for serotype 1, ST227 isolates 96-5891 and 96-5892. 156
Figure 8-3 Relationships of MLST sequence types constructed using eBURST version 3 of all isolates in the MLST database which express serotype 1 capsule. 158
Figure 8-4 Comparison of CGH results for serotype 4 outbreak isolates. 159
Figure 8-5 Genelists created from CGH analysis of dye swap experiments for serogroup 4, ST206 isolates. 160
Figure 8-6 e-BURST version 3.0 representation of serogroup 4 related MLST clonal complexes. 161
Figure 9-1 Map of Bolivia. 167
Figure 9-2 Serotype distribution of 53 Bolivian pneumococcal carriage isolates. 172
Figure 9-3 New Sequence Types associated with serotype 23F which provide new insight into their relationship within a clonal complex. 180
Figure 9-4 e-BURST version 3 diagram of serotype 38 pneumococci. 183
Figure 9-5 Distribution of pneumococcal serotypes causing invasive disease in Bolivia (n=45) from 2000-2003. 184
Figure 10-1 Comparison of the Mean Expression Levels for genes SP0054, SP0045 and SP2153 by RT-PCR. 197
Figure 10-2 CGH comparisons of isolates 99-4038 (cultured from blood) and 99-4039 (cultured from CSF). 203
Figure 10-3 Comparison of microarray DNA CGH results for isolates 99-4038 and 99-4039 with PCR results for the same genes using DNA from the same isolates................204

Figure 10-4 Comparison of the significant differences in expression of SP0376 and SP1872 in isolates 99-4038 (blood origin) and 99-4039 (CSF origin)..210
Figure 10-5 RT-PCR results relating to gene expression of SP0376 (*RitR*) in blood and CSF and the upregulation of SP1869 and SP1872 in CSF. 211
Figure 10-6 Distribution of pneumococcal serotypes associated with cerebral abscesses in Scotland 1993-2007 (n=9). 213
Figure 10-7 Distribution of pneumococcal MLSTs associated with cerebral abscesses in Scotland 1993-2007 (n=9). 213
Figure 11-1 Structure of clarithromycin. 220
Figure 11-2 Growth of isolate South Africa 2507 to determine MIC to clarithromycin. ..227
Figure 11-3 Growth curve of isolate South Africa 2507 when grown in Brain Heart Infusion broth. 228
Figure 11-4 Comparison of the growth of South Africa 2507 in Brain Heart Infusion broth without clarithromycin added and the influence of adding clarithromycin. 228
Figure 11-5 Mean Normalized Expression Levels identified by RT-PCR for genes SP0740, SP0800 and SP1631. 230
Figure 11-6 Mean Normalized Expression Levels identified by RT-PCR for genes *ErmB* and pneumolysin. 231
Figure 11-7 Comparison of strain South Africa 2507 gene expression demonstrated by microarray at midlog and 15 minutes later growing in the presence and absence of subtherapeutic clarithromycin (5mg/L)...233

Acknowledgements

This research would not have been completed without the help and support of many individuals from various organisations.

At the University of Glasgow, thanks are due to Professor Tim Mitchell as my supervisor and for hosting the work in his laboratory, facilitating the progress of the experimental work along with his research assistants and my cohort of PhD student colleagues. Thanks are also due to Dr Nuno Silva for an introduction to microarray technology, Dr Johanna Jefferies for an introduction to MLST and the staff of the Henry Wellcome Molecular Biology Support Unit for assistance using their equipment for quantification of nucleic acids and scanning microarrays.

Sincere thanks is given to the staff of the Bacterial Microarray Group at St George's Hospital, University of London as without their training in the use of their microarrays and the bioinformatics interpretation of microarray results, this work would not have progressed. Particular thanks are due to Dr Jason Hinds, Dr Kate Gould and Dr Lucy Brooks.

The assistance of the staff of the Scottish Meningococcal and Pneumococcal Reference Laboratory was above and beyond that expected. Thanks are due to Dr Giles Edwards for permission to use the laboratory facilities, strain collection, equipment and supporting the idea of performing a pneumococcal carriage study in Bolivia. The work in Bolivia would have ground to a halt without backup support by Dr Rosein Ure, Mrs Jennifer Reid, Mrs Louise Thom, Mr Andrew Small, Mr Donald Kier, Mr Kevin Scott and Mrs Barbara Denham. Thanks for the training in serotyping and MLST and supporting the research with such enthusiasm and humour. Particular thanks to Dr Mathew Diggle for making time to explain so many aspects of molecular biology, training in the use of MLST and gene sequencing, quantitative RT-PCR and catching a vision to risk life and limb in the Bolivian Amazon as assistant clinical microbiologist.

Sincere thanks to Dr Diego and Mrs Jo Santana-Hernandez and the staff of Fundacion Totai for hosting and facilitating our research in Bolivia, for their enthusiasm, patience, perseverance and faith which enabled insurmountable obstacles to be surmounted. Thanks too to Dr Peter and Mrs Adelina Altstadt for allowing us unlimited access to their microbiology laboratory in Trinidad, Beni (and putting up with our dreadful Spanish) –

muchas gracias. Thanks to Medical, Wire and Equipment, United Kingdom for donating swabs for the work in Bolivia for free. Thanks are also due to the Reference Laboratory for Meningococci, Madrid without whose help we would have been unable to transport our pneumococcal isolates out of Bolivia.

Within the National Health Service, thanks to Dr Bishan Thakker of the Department of Clinical Microbiology, Glasgow Royal Infirmary for providing details relating to the serotype 1 outbreak featured in Chapter 8 and Dr Alisdair MacConnachie, Consultant Infectious Disease Physician, Gartnavel Hospital, Glasgow for details of the case of pneumococcal endocarditis featured in Chapter 7. Thanks too are due to the medical and laboratory staff of the Departments of Medical Microbiology at Glasgow Royal Infirmary and the Western Infirmary of Glasgow which enabled the API work featured in Chapter 6 to be performed and for keeping me abreast of changes in the world of clinical microbiology.

Thanks to Mr Nicholas Croucher at the Wellcome Trust Sanger Institute, Hinxton for training in the use of the Artemis Comparison Tool.

Thanks are also due to the Wellcome Trust for funding the research and the Meningitis Association Scotland for fund raising for the work in Bolivia.

Thanks too for all the friends and family who have been so patient and encouraging over the three years.

This work is dedicated to my grandfather, Mr George Inverarity (23/01/1900-08/02/1935), who died as a younger man than I am as I write this, from pneumococcal meningitis.

Author's declaration

This thesis embodies the original work of the author unless otherwise stated.

Donald James Inverarity

Abbreviations

ACT	Artemis Comparison Tool
ANOVA	Analysis of Variance
AOM	Acute Otitis Media
BµG@S	Bacterial Microarray Group at St George's Hopsital, University of London
BHI	Brain Heart Infusion
BLAST	Basic Local Alignment Search Tool
Bp	Base pair
CAP	Community Acquired Pneumonia
CFU	Colony Forming Unit
cDNA	complementary Deoxyribonucleic acid
CGH	Comparative Genomic Hybridization
CLSI	Clinical and Laboratory Standards Institute
CSF	Cerebrospinal Fluid
dATP	2'-deoxyadenine 5'-triphosphate
dCTP	2'-deoxycytidine 5'-triphosphate
dGTP	2'-deoxyguanosine 5'-triphosphate
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleoside triphosphate
DTT	Dithiothritol

dTTP 2’-deoxythymidine 5’-triphosphate

eBURST electronic Based Upon Related Sequence Types

EDTA Ethylenediaminetetraacetic acid

EMBL European Molecular Biology Laboratory

GAL Genepix Array List

GLIMMER Gene Locator and Interpolated Markov ModelER

IAA Isoamyl Alcohol

IPD Invasive Pneumococcal Disease

IS Insertion Sequence

Kb Kilobase pair

Mb Megabase pair

MIAME Minimum Information About Microarray Experiment

MIC Minimum Inhibitory Concentration

ml millilitre

MLST Multilocus Sequence Typing

MRC Medical Research Council

NCBI National Center for Biotechnology Information

nm nanometre

ORF Open Reading Frame

PAHO Pan American Health Organisation

PBP	Penicillin Binding Protein
PCR	Polymerase Chain Reaction
PCV-13	13 Valent Pneumococcal Vaccine
PCV-7	7 Valent Pneumococcal Vaccine
PMEN	Pneumococcal Molecular Epidemiology Network
PTS	Phosphotransferase Systems
RNA	Ribonucleic acid
Rpm	Revolutions per minute
SDS	Sodium dodecyl sulphate
SMPRL	Scottish Meningococcal and Pneumococcal Reference Laboratory
SNPs	Single Nucleotide Polymorphisms
ST	Sequence Type
TE Buffer	Tris EDTA buffer
TIFF	Tagged Image File Format
TIGR	The Institute for Genomic Research
TrisHCl	Tris (hydroxymethyl) aminomethane Hydrochloride
WHO	World Health Organisation
µl	microlitre
µg	microgram

1 Introduction

Infections due to *Streptococcus pneumoniae* (the pneumococcus) remain a substantial source of morbidity and mortality in both developing and developed countries despite a century of research and the development of therapeutic interventions such as multiple classes of antibiotics and vaccination. The World Health Organisation estimates that in developing countries 814,000 children under the age of five die annually from invasive pneumococcal diseases (Scott, 2007) with an estimated 1.6 million deaths affecting all ages globally (WHO, 2007).

The ability of some isolates of *S. pneumoniae* to evade many antibiotic classes through several distinct genetically determined resistance mechanisms and pneumococcal evasion of capsular polysaccharide based vaccines through serotype replacement and capsular switching, along with the discovery of 91 distinct polysaccharide capsular types and over 4000 Multi Locus Sequence Types (MLST) all reflect the extensive diversity and plasticity of the genome of this naturally transformable organism which can readily alter its genome to survive in response to its environment and the pressures placed upon it.

The purpose of this thesis is to investigate this genomic diversity from a genomic sequence perspective using contemporary microarray and sequencing technology and to relate these findings to the epidemiology of invasive pneumococcal disease (IPD) and carriage with the intent of further understanding a genetic basis for pneumococcal disease pathogenesis and the response of pneumococci to specified currently available therapeutic options.

1.1 Taxonomy of Streptococcus pneumoniae

The taxonomy of *Streptococcus pneumoniae* (the pneumococcus) has undergone a dynamic process since separate initial descriptions in saliva by Pasteur (Pasteur, 1881) and as *Micrococcus lanceolatus* by Sternberg (Sternberg, 1881). For instance, in 1903 Schottmuller described *Streptococcus mucosus* (Schottmuller, 1903) but within a couple of years (Park *et al*., 1905, Collins, 1905) it was reclassified as a pneumococcus (a term used since 1886 (Fraenkel, 1886)) on the basis of its biochemistry and specific agglutination and agglutinin absorption tests despite its markedly different mucoid morphology on solid agar. From 1926 the genus Diplococcus was used until 1974 when it changed to *Streptococcus pneumoniae* (Musher, 2005). More recently, it has become clear

that the species boundaries between *S. pneumoniae* and particularly *Streptococcus mitis* and *Streptococcus oralis* are not easily defined (Whatmore *et al*., 2000).

Such a catalogue of reclassifications and disputed identification reflects the remarkable phenotypic diversity of the pneumococcus, which in turn reflects the substantial genomic and transcriptional diversity of the organism, resulting from a constellation of several interacting mechanisms – the investigation of which is the focus of this thesis.

1.2 Phenotypic identification and characteristics of Streptococcus pneumoniae

Streptococcus pneumoniae is a Gram positive coccus which can be identified by its production of α-haemolysis on blood agar, bile solubility, inhibition by ethyl hydrocupereine (optochin) and catalase negativity (Musher, 2005). It is noteworthy that the Gram stain itself was developed by Christian Gram as a means to distinguish the pneumococcus from other pathogens in specimens taken from patients with lobar pneumonia (Gram, 1884). One or more of these tests may be inconclusive and further phenotypic tests such as agglutination with anti-pneumococcal polysaccharide capsule antibodies or genotypic tests can be required to distinguish the isolate from closely related oral streptococci (Whatmore *et al*., 2000, Hanage *et al*., 2005) although even some of these tests may also be inconclusive in rare instances (Dowson, 2004).

1.2.1.1 A History of Serotyping *Streptococcus pneumoniae*

In 1909 Neufeld and Haendel raised antibodies to pneumococci in horses and donkeys, allowing serological identification (Neufeld and Handel, 1909). This was repeated with a second strain in 1910 (Neufeld and Handel, 1910). In 1913, A.R Dochez and L.J. Gillespie at the Rockefeller Institute in New York (Dochez and Gillespie, 1913) and Frederick Lister at the South African Institute for Medical Research in Johannesburg (Lister, 1913) published serologically based typing schemes for pneumococci. The Rockefeller Institute scheme described Types I, II, III and IV of which Types I, II and III were identified serologically and Type IV was a group for pneumococci found not to react with their three sera. There was overlap with Lister's typing scheme as Type I corresponded to Lister's Group C, Type II corresponded with Lister's Group B and Type III corresponded to Lister's Group E (Urquhart, 1921). Consequently the Lister scheme fell into disuse.

To complicate matters, the Rockefeller Institute Type II was found to be a heterogeneous group (Urquhart, 1921) and probably consisted of several different contemporary serotypes. It is likely that since Type 1 was homogeneous and since clinical descriptions of Type 1 associated disease patterns match those described for the contemporary serotype 1, that the Rockefeller Type 1 pneumococcus is equivalent to the modern serotype 1 pneumococcus.

Avery *et al* in 1917 at the Rockefeller Institute, proposed that *Pneumococcus mucosus*, described as, “larger, rounder, and less lanceolate than other types of pneumococcus,” and possessing, “a large distinct capsule,” which grew on blood agar with colonies which were, “moist, mucoid, and confluent” should become known as their Type III pneumococcus (Avery *et al*., 1917). This phenotypic description matches that seen for contemporary serotype 3 isolates although mucoid colonies can be seen in serotypes 1, 3, 6, 8, 18, 20, 23 and 25 (Gransden *et al*., 1985). Avery *et al* addressed this in 1917 saying,

> “For these reasons the identification of Type III pneumococcus by morphologic and cultural characteristics is not always absolute, and the diagnosis should be established by immunologic methods, when Type III serum is available (Avery *et al*., 1917).”

However, it was not until 1934 when type III could be distinguished serologically from the phenotypically similar Type VIII (Finland and Sutliff, 1934) so only after 1934 is it fair to conclude that descriptions of the Type III pneumococcus are equivalent to the contemporary serotype 3 pneumococcus.

Standardisation of the Rockefeller typing scheme in Great Britain is documented by Glynn and Digby (Glynn and Digby, 1923), who note that from 1920, diagnostic and treatment sera for use in Britain by the Medical Research Council was supplied by the Rockefeller Institute. Standardisation of antipneumococcal serum had been practiced at the Rockefeller Institute since 1917 (Cole and Moore, 1917).

The Quellung technique for serotyping the pneumococcus became the preferred method for serotyping in 1932 (Neufeld and Etinger-Tulczynska, 1932) although first described in 1902 (Neufeld, 1902). Between the end of World War II and the early 1980s there was little international agreement regarding pneumococcal nomenclature based on serotyping and the reasons for this are reviewed by Henrichsen (Henrichsen, 1999). The most recent serotype to be discovered was 6C in 2007 (Park *et al*., 2007) bringing the current total of

recognised serotypes to 91 and reinforcing observations by Waltman *et al* (Waltman *et al*., 1991) that there is often substantial diversity within serogroups.

1.3 Genetic typing of Streptococcus pneumoniae

Typing methodologies based on the pneumococcal genetic structure have been developed. These genetic typing schemes are reviewed below although it is worth noting that many early methods were only useful for examining genetic relatedness in temporally and geographically restricted studies (Robinson *et al*., 1998).

The need for standardised molecular typing schemes has been due in part to the international spread of clones of multiply antibiotic resistant pneumococci (Hermans *et al*., 1997b) and after the establishment of the Pneumococcal Molecular Epidemiology Network (PMEN) in 1997, guidelines for molecular typing surveillance were published (McGee *et al*., 2001a). It is becoming apparent though that, even with current technologies,

> "the clearest picture of invasiveness and genetic relatedness can be viewed when typing methods are used collectively (Obert *et al*., 2007)."

1.3.1.1 Pulsed Field Gel Electrophoresis (PFGE)

PFGE was first described for the pneumococcus by Lefevre *et al* (Lefevre *et al*., 1993) and has been used in several studies to establish the genetic relatedness of pneumococcal isolates (Hermans *et al*., 1995, Obert *et al*., 2007, Porat *et al*., 2006, McGee *et al*., 2001b, Tsolia *et al*., 2002, Rossi *et al*., 1998, Watanabe *et al*., 2003, Hall *et al*., 1996). In general it consists of generating fragments of chromosomal DNA using the restriction enzyme *Sma*I and then separating them using gel electrophoresis. As experimental conditions are not always identical between studies, the comparison of results over time and between different laboratories is fraught with difficulties. It is labour intensive and does not lend itself well to computerised interpretation (Hermans *et al*., 1995). Maiden *et al* also propose that PFGE is better suited to investigating outbreaks as it is discriminatory enough to distinguish between strains circulating within a geographical area but is not discriminatory enough for long term epidemiological surveillance (Maiden *et al*., 1998).

1.3.1.2 Restriction Fragment End Labelling (RFEL)

RFEL was described by Van Steenbergen *et al* in 1995 (Van Steenbergen *et al*., 1995). Although similar to PFGE, smaller DNA fragments are produced by RFEL and

pneumococcus specific reference bands can be produced which aids computer assisted analysis of results (Hermans *et al*., 1995). It has been used to identify clones of antibiotic resistant pneumococci (Hermans *et al*., 1997b) as well as clustering within penicillin susceptible strains (Overweg *et al*., 2000a).

1.3.1.3 Restriction Fragment Length Polymorphisms (RFLP)

RFLP was an uncommon method used mainly in the 1990s as a means of assessing clonal relationships between similar isolates of the same serotype (Swiatlo *et al*., 1996, Robinson *et al*., 1998). Pneumococcal genomic DNA was digested with the restriction enzyme *Hind*III, separated by gel electrophoresis and pressure or vacuum transferred to a nylon membrane where hybridization of fragments with probes of a recognised pneumococcal gene (such as *psp*A) or an insertion sequence (such as IS*1167*) occurred. The resulting banding patterns are used to denote different subtypes. Often RFLP was used in combination with another method of investigating genetic relationships such as MLEE (Robinson *et al*., 1998) and penicillin binding protein gene fingerprinting (Hermans *et al*., 1997b).

1.3.1.4 Penicillin Binding Protein Gene Fingerprinting

The discovery that alleles for penicillin binding proteins (pbps) in the pneumococcus were heterogenous permitted their use in an early molecular typing scheme for the pneumococcus. The genes for pbps 1A, 2B and 2X were amplified from genomic DNA by Polymerase Chain Reaction (PCR). The purified genes for pbps were then digested by restriction enzymes using methodologies similar to RFLP (Munoz *et al*., 1991, Hermans *et al*., 1997a, McGee *et al*., 2001b, Swiatlo *et al*., 1996, Kell *et al*., 1993, Coffey *et al*., 1991). Like many early typing schemes, although significant genomic diversity could be demonstrated, gel results were difficult to interpret, hard to standardise and therefore hard to reproduce by different laboratories.

1.3.1.5 BOX fingerprinting

BOX fingerprinting for pneumococci is a method which has been superseded by contemporary techniques. Essentially genomic DNA was digested by a restriction enzyme and separated by gel electrophoresis and hybridized against an oligonucleotide probe (a section of or the entire BOX repeat – a repetitive mosaic sequence comprised of boxA, boxB or boxC subunits found in the pneumococcal genome (Koeuth *et al*., 1995, Van Belkum *et al*., 1996)). This allowed clustering of strains sharing identical fingerprints or banding patterns. BOX fingerprint clusters often did not correspond to clustering based on

serotyping (Hermans *et al*., 1995, Rodriquez-Barradas *et al*., 1997, McGee *et al*., 2001b, Mollerach *et al*., 2004) or RFLP results (Robinson *et al*., 1999). BOX fingerprinting has been used in combination with RFLP using the IS*1167* sequence to determine IS*1167-boxA* genotypes but it was concluded that,

> "markers such as IS*1167* and *boxA* are not ideal for dealing with the large genotypic diversity seen in cross-sectional samples of pneumococci (Robinson *et al*., 2001)."

It did have an application in investigating pneumococcal outbreaks (Ertugrul *et al*., 1997).

1.3.1.6 Multi Locus Enzyme Electrophoresis (MLEE)

MLEE was developed in the mid 1980s (Selander *et al*., 1986) and used extensively to investigate genetic diversity of the pneumococcus in the 1990s (Hall *et al*., 1996). This method detects variation within housekeeping enzymes and their differing electrophoretic mobilities (Feil *et al*., 2000a). It was used to demonstrate diversity of penicillin resistance genes (Kell *et al*., 1993), capsule (Sibold *et al*., 1992) and Immunoglobulin A1 protease (Lomholt, 1995) and often used in studies in combination with other methods such as RFLP (Robinson *et al*., 1998), PFGE (Hall *et al*., 1996) and penicillin binding protein gene fingerprinting (Coffey *et al*., 1991). However, MLEE can produce ambiguous results as it,

> "relies on the indirect assignment of alleles based on the electrophoretic mobility of enzymes, and indistinguishable mobility variants may be encoded by very different nucleotide sequences (Maiden *et al*., 1998)."

It did, however, pave the way for Multi Locus Sequence Typing.

1.3.1.7 Multi Locus Sequence Typing (MLST)

The development of Multi Locus Sequence Typing (MLST) (Enright and Spratt, 1999b, Enright and Spratt, 1998) generated a genotyping method which sequences seven conserved housekeeping genes *(aroE, gdh, gki, recP, spi, xpt, ddl)* assigning each unique allele with a number, allowing relatedness between isolates to be identified with a numerical code and overall sequence type which is easily reproducible and comparable between laboratories thereby providing unambiguous results which are electronically portable and suitable for global surveillance studies via the internet (Obert *et al*., 2007). It has been successfully automated (Jefferies *et al*., 2003).

The depositing of MLST sequence data in an internet based repository[1] greatly enhances a global understanding of bacterial epidemiology and the development of software such as eBURST (electronic Based Upon Related Sequence Types) allows the construction of identifiable genetic lineages and clusters (Feil *et al.*, 2004). The contributions of recombination and mutation to clonal divergence can also be assessed using this data (Feil *et al.*, 2000b).

MLST has a further advantage as it can be performed after PCR amplification of DNA extracted directly from clinical material or non-viable organisms and, unlike serotyping, can be performed on uncapsulated organisms (Hanage *et al.*, 2005) which can both improve diagnostic yields and epidemiological surveillance (Maiden *et al.*, 1998). The advantages, disadvantages and clinical applications of MLST are reviewed by Sullivan *et al* (Sullivan *et al.*, 2005).

1.3.1.8 Multi Locus Variable Number of Tandem Repeat Analysis (MLVA)

MLVA was proposed for use as a typing scheme for epidemiological studies by Koeck *et al* (Koeck *et al.*, 2005). They used a set of 16 polymorphic tandem repeat sequences identified from published fully sequenced pneumococcal genomes from which 49 genotypes of pneumococci were identified. Although it was claimed to be cheaper and requiring less expertise than MLST, it has not competed well against the popularity of MLST.

1.3.1.9 Multi Invasive Locus Sequence Typing (MILST)

Multi Virulence Sequence Typing is a further development of the MLST scheme which adds virulence genes to the housekeeping genes used in the profiling of bacteria. It has been used for typing *Listeria monocytogenes* with greater discriminatory power than MLST alone (Zhang *et al.*, 2004). This typing scheme has been adapted for pneumococci by Obert *et al*, to form a scheme which they termed MILST (Obert *et al.*, 2007) and which they propose not as a replacement for MLST but as an expansion of the scheme.

1.3.1.10 Microarray Comparative Genomic Hybridization (CGH)

One of the most utilised applications of microarray technology is to allow comparison of bacterial genomes and identify conserved regions and regions of diversity. The first

[1] http://www.mlst.net {accessed 10th December 2008}

applications of this technology to examine the pneumococcal genome demonstrated that some genes such as hyaluronidase, neuraminidase A or neuraminidase B, autolysin or pneumolysin were highly conserved while sequence variation could be demonstrated for cell surface proteins such as pneumococcal surface protein C, other choline binding proteins and the trimethoprim resistance gene dihydrofolate reductase (*dhfr*) (Hakenbeck *et al*., 2001). The absence of genes such as some capsular genes when comparing the R6 and TIGR4 genomes resulted in an understanding that there were significant regions of diversity in the pneumococcal genome and attempts at identifying a set of “core” genes. This has also been possible for CGH studies of *Enterococcus faecalis* (Aakra *et al*., 2007) and *Staphylococcus aureus* (Lindsay *et al*., 2006). Cassone *et al* review examples of the application of CGH to the study of intra- and inter- species genomic diversity for *Mycobacteria*, *Salmonella* and *Yersinia* species (Cassone *et al*., 2007).

Hollingshead and Briles observe that,

> “there is also valuable information to be gleaned from knowing the sequence of more than one genome within the same species. Additional genome sequence data for well-chosen strains will yield important clues to solving certain puzzles about the biochemical diversity, the virulence and pathogenesis range and/or the evolution of bacterial species (Hollingshead and Briles, 2001).”

The CGH studies which follow in subsequent chapters are designed to enhance such understanding of the diversity of the pneumococcal genome by comparing clinically relevant pneumococcal isolates with the established sequenced pneumococcal genomes TIGR4 and R6 and also assess diversity in more recently sequenced pneumococcal genomes.

1.3.1.11 Transcriptome Microarrays

Rather than comparing gene sequences from genomic DNA, bacterial transcriptome microarrays assess the presence or absence of expression of genes by taking total messenger RNA expressed by the organism, under carefully controlled conditions, and synthesizing cDNA from the RNA by reverse transcription. The cDNA is then hybridized against the DNA sequences of the microarray. As such, they are,

> “a powerful tool to dissect regulatory networks (Rimini *et al*., 2000).”

The first transcriptome microarrays were for partial or full genomes (*Escherichia coli* (Richmond *et al*., 1999, Chuang *et al*., 1993) and *Saccharomyces cerevisiae* (De Risi *et al*., 1997, Wodicka *et al*., 1997)).

De Saizieu *et al*, 1998 used the first pneumococcal transcript oligonucleotide microarray, which contained only 100 genes, to investigate gene expression during competence (De Saizieu *et al*., 1998). With regard to pneumococcal disease pathogenesis, microarrays have been employed in the investigation of the genetic basis of autolysis (Sublett *et al*., 2004), the genetic basis of capsule loss (Ogburn and Dowson, 2004), competence (Mascher *et al*., 2003, De Saizieu *et al*., 1998, Peterson *et al*., 2004, Peterson *et al*., 2000, Rimini *et al*., 2000), mechanisms of iron uptake (Ulijasz *et al*., 2004b), the vancomycin stress response (Sublett *et al*., 2004) and the transcriptional adaptation of pneumococci to various environmental changes (Pandya *et al*., 2005, Novak *et al*., 1999, Hendriksen *et al*., 2004). One potential limitation of this approach though is that microarrays can only detect genes which are on the array so whole pathways may not be represented if they involve novel genes.

1.3.1.12 Genome Sequencing

The first published whole genome sequences for bacteria began to appear in the mid-1990s beginning with that of *Haemophilus influenzae* in 1995 (Fleischmann *et al*., 1995). As early as 1991, partial sequences of a pneumococcal genome (located on *Sma*I and *Apa*I fragments) were used to construct a map of the genome (Gasc *et al*., 1991).

Initial sequencing of regions of a pneumococcal genome (R6) with the aim of identifying potential drug targets (amino-acyl tRNA synthetase and ribosomal protein genes) are described by Baltz *et al* (Baltz *et al*., 1998). The first comparative analysis of a draft pneumococcal genome was published in 2001 where the draft genome of G54 (a serotype 19F strain) was compared with the genome of *Streptococcus pyogenes* (Dopazo *et al*., 2001). Currently, according to the National Center for Biotechnology Information (NCBI) Genome Project[2], there are six completely sequenced pneumococcal genomes with a further 14 in progress and a further 16 at an assembly stage. The list of genomes, "in progress," includes the ATCC700669 serotype 23F isolate (used in Chapter 3) at the Wellcome Trust Sanger Institute but does not include the OXC141, INV104B or INV200 isolates (used in Chapter 3) which have also been sequenced by conventional Sanger sequencing (using dideoxynucleotide sequence termination and capillary electrophoresis with fluorescent marker excitation by laser (Ryan *et al*., 2007, Metzker, 2005)) and are available from the Wellcome Trust Sanger Institute website[3]. Neither does it include the

[2] http://www.ncbi.nlm.nih.gov/sites/entrez?db=genomeprj&cmd=Retrieve&dopt=Overview&list_uids=12328 {accessed 17th November 2008}

[3] http://www.sanger.ac.uk/Projects/S_pneumoniae/ {accessed 17th November 2008}

sequencing of isolates 03-2672, 03-3038, 06-1370 (used in Chapter 10), nor 02-1198, 03-4156, 03-4183, 07-2838, 99-4038 and 99-4039 (used in Chapters 6 and 10) which are being sequenced at the Wellcome Trust Sanger Institute using high-throughput 454 Life Sciences Corporation (Roche, USA) pyrosequencing methodologies (Margulies *et al.*, 2005) recently described in the sequencing of *Salmonella typhi* (Holt *et al.*, 2008) and pneumococci (Hiller *et al.*, 2007).

Knowing the genome sequence of a pathogen doesn't necessarily mean that knowledge regarding gene function is known although it can be inferred (Tang and Moxon, 2001). From these inferences, potential antimicrobial drug targets (Brazas and Hancock, 2005) or vaccine candidate molecules (Wizemann *et al.*, 2001) can be identified although sequence similarity should be combined with other evidence such as experimental data to assign function (Kaushik and Sehgal, 2008). Comparative genome analysis is useful for assigning function to unknown genes as gene and protein homologs in different bacteria are likely to be functionally linked and have similar degrees of conservation within the same pathways (Martin *et al.*, 2003b).

Bacteria which show substantial diversity (even within the same serotype) have a gene repertoire which can be divided into a "core" genome comprising essential genes such as housekeeping genes, a "dispensable" genome and "strain specific genes" which may confer selective advantages such as adaptation to particular environments (Kaushik and Sehgal, 2008, Fraser-Liggett, 2005). When an ST180 serotype 3 strain (0100993) was compared with TIGR4, a variety of chromosomal rearrangements and variations due to mobile genetic elements were discovered along with 17 novel genes not present in TIGR4 (Oggioni *et al*, 2001). The first full comparison of pneumococcal genomes was published by Bruckner *et al* in 2004, who compared the TIGR4 and R6 genomes. They noted that the two genomes differed in size (2 versus 2.16Mb) and in approximately 10% of their gene content (Bruckner *et al.*, 2004). R6 was noted to contain 69kb of genetic material not seen in TIGR4 and TIGR4 has 157kb of genetic material not seen in R6 (Bruckner *et al.*, 2004). This is similar to the situation in non-typeable *Haemophilus influenzae* isolates where around 10% of the gene content of 10 isolates was identified as novel compared to the sequenced *H. influenzae* Rd KW20 genome with many of these unique genes non-uniformly distributed in the genome (Shen *et al.*, 2005). Shen *et al*, were later to describe 58 novel genes from 8 pneumococcal isolates (serotypes 3, 6A, 9F, 11, 14, 18C, 19F and 23F) which were not present in either the TIGR4 or R6 genomes (Shen *et al.*, 2006a). Consequently, attempts have been made to try to identify a core pneumococcal genome (Hiller *et al.*, 2007, Obert *et al.*, 2006). It is becoming clear though that the sequence of a

single pneumococcal genome (e.g. TIGR4) and even of the multiple pneumococcal genomes becoming available (Hiller *et al*., 2007) may not adequately represent the full extent of genomic diversity of this bacteria.

1.4 Pneumococcal Carriage

Pneumococci predominantly colonise the human nasopharynx and in the vast majority of instances do not progress to cause invasive disease. In the first two years of life, 95% of children can be colonized with pneumococci and 73% can carry at least two serotypes simultaneously (Obaro and Adegbola, 2002). Nasopharyngeal colonization can begin as early as the day of birth. The duration of carriage for a particular serotype is commonly 2.5 to 4.5 months and the duration of carriage decreases with each successive pneumococcal serotype. This duration of carriage is inversely correlated with age (Gray *et al*., 1980) as pneumococcal carriage declines as children grow older (Bogaert *et al*., 2004).

Historical observations regarding the carrier state for Type I and Type III pneumococci are often different in the pre-antibiotic era when compared with modern times. Modern accounts tend to suggest that it is uncommon to find serotype 1 being carried by individuals (Brueggemann and Spratt, 2003, Brueggemann *et al*., 2003) and it has been claimed that,

> “serotype 1 is almost never carried (Brueggemann *et al*., 2004),”

It is erroneous to conclude that such a state cannot occur. Several historical accounts from the pre-antibiotic era make it clear that a healthy carrier state of Type I did and can occur in individuals in close contact with cases and for prolonged periods in those who had recovered from Type I associated lobar pneumonia without antibiotic therapy.

Dochez and Avery in 1915 noted carriage of Type I pneumococci in 3 of 66 pneumococcal carriers (4.5%) who were, “normal individuals.” They were unable to assess Type III carriage rates due to a lack of Type III antisera (Dochez and Avery, 1915). Stillman in 1917 noted that,

> “although pneumococcus is present in the mouths of about fifty percent of normal individuals, it is extremely rare to find pneumococcus of Type I in the normal mouth except in individuals who have been in intimate association with patients suffering from lobar pneumonia (Stillman, 1917).”

This was a conclusion based on his earlier observation in 1916 that only 4 of 172 (2.3%) pneumococcal orophayrngeal carriers carried Type 1 and 44 of the 172 (26%) carried Type III. There was persistent carriage of Type III for over 3 months in some instances (Stillman, 1916) and Type I carriage of up to 90 days after recovery from pneumococal pneumonia (Dochez and Avery, 1915). Meyer in 1920 demonstrated higher Type III carriage rates than for Type I which was not detected (Meyer, 1920). Strom in 1932 also documents a long period of Type I carriage for 73 days after recovery from pneumonia (Strom, 1932).

Stillman in 1917 cultured 107 healthy household contacts of 28 patients admitted with Type I lobar pneumonia and found that 15% were Type I carriers (Stillman, 1917). Almost a third of cases gave rise to a carrier state in at least one of their immediate associates. For these healthy contacts, the average period of carriage was 25 days for Type I - undoubtedly providing a human reservoir for further infection. Avery *et al*, also in 1917, notes a Type I carriage rate of 13% in the contacts of cases of lobar pneumonia and only 0.33% in controls who had no contact with cases of lobar pneumonia and noted that the Type I carriers harboured it for 3 to 4 weeks (Avery *et al.*, 1917). Similar differences between contacts and controls with regard to Type I carriage are documented in Glasgow in 1932 (Christie, 1932).

In a carriage study in 1919, using specimen collection methodology remarkably similar to the current World Health Organisation (WHO) standard method (O'Brien and Nohynek, 2003), Sailer *et al* in Georgia, USA performed nasopharyngeal swabbing on 700 soldiers and detected a pneumococcal carriage rate of 16%. Of the carriers, 5.4% were Type I and 4.5% were Type III. They observed that their carriage rates were influenced by recent outbreaks of measles and influenza (Sailer *et al.*, 1919). There is also an association with higher pneumococcal carriage rates during times of epidemic upper respiratory tract infection (Gordon, 1921). Strom in 1932 documents the decline in serotype 1 carriage rates after a serotype 1 outbreak in a Norwegian orphanage in 1931 finding serotype 1 carriage rates of 33% at the height of the outbreak dropping to 16.9% 3 months later but being undetectable at all times in neighbouring orphanages unaffected by the outbreak (Strom, 1932). The serotype 1 pneumococcal pneumonia outbreak at the State Hospital in Worcester, Massachusetts also noted a carriage rate of almost 10% (Smillie *et al.*, 1938).

Other risk factors for pneumococcal carriage are well recognised and include overcrowding (Hodges *et al.*, 1946, DeMaria *et al.*, 1980, Hoge *et al.*, 1994), concurrent outbreaks of invasive pneumococcal disease and pneumonia (DeMaria *et al.*, 1980, Dagan

et al., 2000, Hoge *et al.*, 1994) and adults in close contact with children (Hendley *et al.*, 1975), ethnicity, family size, smoking and recent antibiotic use (Bogaert *et al.*, 2004). Children are thought to be the most important reservoir for pneumococci (Leiberman *et al.*, 1999, Kyaw *et al.*, 2002),

Pneumococcal carriage rates vary substantially between studies and populations as demonstrated in the review by Bogaert *et al* (Bogaert *et al.*, 2004). Hodges and MacLeod suggest that serotype 1 and serotype 3 carriage rates varied so much in early carriage studies, because of differences in the methodology for specimen processing. They compared several methods and mouse inoculation (favoured by the early Rockefeller Institute studies in New York) yielded higher recovery rates of serotype 1 and serotype 3 pneumococci than direct plating of throat swabs onto blood agar (Hodges *et al.*, 1946).

The demonstration that different cultural methods affect the observed carriage rates could readily account for why serotype 1 carriage rates were apparently higher at the start of the 20th century when mouse inoculation was common but which dropped in later years when most carriage study methodology involved variants of plating nasopharyngeal swabs onto blood agar (Hodges *et al.*, 1946). Differences in the frequency of sampling and in the populations studied could also account for difference in carriage rates (Davies and Lockley, 1987). It is reasonable to conclude though that serotype 1 carriage does still occur in communities, that it does increase when numbers of cases of pneumococcal pneumonia due to serotype 1 increase in the community but that it may not be as readily detected by current standard methods as other serotypes such as serotype 3.

The distribution of serotypes identified in the nasopharynx is usually wider than those causing IPD in the same population but varies substantially between human populations. In Europe and the USA the commonly seen serotypes found in carriage studies are 6B, 14, 19F and 23F (Bogaert *et al.*, 2004, Kyaw *et al.*, 2002). This has a direct impact on the coverage of serotypes included in conjugate vaccines which can also vary substantially (Bogaert *et al.*, 2004). The prevalence of nasopharyngeal carriage is considerably higher in developing countries and the probability of exposure to a non-vaccine serotype considerably greater (Scott, 2007).

The mechanism of colonisation is reviewed by Bogaert *et al* (Bogaert *et al.*, 2004). Extracellular neuraminidase decreases viscosity of mucous on epithelial cell surfaces and exposes N-acetyl-glycosamine receptors which interact with pneumococcal surface-associated proteins such as *PsaA*. Cytokines from the host epithelial cells result in

upregulation of platelet activating factor receptors. Pneumococcal cell wall phosphocholine attaches to these receptors. The choline binding protein *CbpA* binds to exposed sialic acid residues and lacto-N-neotreatose on the epithelial surface binding to the polymeric immunoglobulin receptor (*pIgR)* which facilitates transcytosis. Phase variation is also involved with transparent phase variants showing greater adherence than opaque variants. It is significant that pneumococcal conjugate vaccination but not pneumococcal polysaccharide vaccination has been shown to reduce carriage of vaccine included serotypes as well as IPD (O'Brien and Dagan, 2003).

1.5 Manifestations of Pneumococcal Infection

1.5.1.1.1 Acute Otitis Media

Acute otitis media (AOM) is the commonest pneumococcal disease manifestation (Hausdorff *et al*., 2002) and is felt to be a major driving force in the establishment of antimicrobial resistant pneumococci (Hausdorff *et al*., 2002). This possibly relates to the frequent development of biofilms at this site promoting persistent infection and facilitating horizontal gene transfer (McEllistrem *et al*., 2007).

In a multinational study, including several countries from the Americas and Europe, it was determined that serotypes 19F and 23F each comprised 13-25% of middle ear fluid isolates, 14 and 6B comprised 6-18% and 6A, 19A and 9V each comprised 5-10% despite differences in location. In children aged 6 months to 59 months, 60-70% of the causative serotypes are covered by the 7-valent pneumococcal conjugate vaccine (PCV-7) but in children less than 6 months old or over 5 years old this coverage drops to 40 to 50%. Serotypes 1, 3 and 5 are important causes of AOM in children under 6 months (Hausdorff *et al*., 2002). In Southern Israel, serotypes 1, 3, 5, 18C, 19A and 19F were associated with the development of AOM (Shouval *et al*., 2006). Serotypes 6A, 6B, 15A and non-typeable pneumococci do not tend to cause AOM (Shouval *et al*., 2006).

PCV-7 is effective at preventing acute otitis media (Eskola *et al*., 2001) although not recurrent otitis media (Veenhoven *et al*., 2003).

1.5.1.1.2 Pneumonia

The pneumococcus is the commonest bacterial pathogen to cause community acquired pneumonia (up to 35% of cases) in adults requiring hospitalisation (Moine *et al*., 1995). Predisposing factors for the development of pneumococcal pneumonia include advanced age, smoking, chronic obstructive pulmonary disease, cancer, HIV, diabetes, chronic heart

failure, alcohol abuse, liver disease, neurological disease, recent hospitalisation and previous pneumonia (Musher *et al*., 2000)

Co-infection or recent infection with influenza has been recognised as an association with pneumococcal pneumonia since the 1918 influenza pandemic (McLelland, 1918) but has only recently been clearly demonstrated to result in 12%-30% of excess cases of pneumococcal pneumonia (Grabowska *et al*., 2006).

The introduction of penicillin reduced mortality from pneumococcal pneumonia from 60% to 9% (Flippin *et al*., 1951).

1.5.1.1.3 Pneumococcal Bloodstream Infections

The importance of obtaining blood cultures during pneumococcal infections was recognised even in the early 20th century (McLelland, 1918). In 1938 it was recognised that a blood culture positivity rate of 20% could be found for patients hospitalised with pneumococcal pneumonia (Musher *et al*., 2000). Blood culture positivity rates are influenced by the severity of infection and level of bacteraemia, the culture method used (with closed automated systems being more sensitive) and whether there was prior administration of antibiotics or pneumococcal vaccination. The presence of pneumococci in the bloodstream can result from invasion at the site of colonisation (primary) or from invasion at the site of an established infection (secondary) (Musher, 2004). Likewise, the detection of pneumococci in the bloodstream may be transient and inconsequential or form part of more significant and life-threatening manifestations of the systemic inflammatory response syndrome, severe sepsis or septic shock (Munford, 2005).

1.5.1.1.4 Meningitis

Meningitis, the infection of cerebrospinal fluid and the meningeal coverings of the brain, is one of the most serious manifestations of invasive pneumococcal disease. Pneumococci account for about 37% of all adult cases of acute bacterial meningitis (Durand *et al*., 1993). Even when treated appropriately with penicillin or a third generation cephalosporin (in penicillin susceptible cases) or vancomycin or chloramphenicol (for penicillin non-susceptible cases), the case fatality rates can vary from 11 - 60% (Stanek and Mufson, 1999). Of those who survive, up to 54% develop neurological sequelae which may persist for a lifetime (Stanek and Mufson, 1999, Bohr *et al*., 1984). Recently, it has been determined that early co-administration of dexamethasone with antibiotics may reduce the severity of these neurological sequelae in adults (Van de Beek *et al*., 2004) as well as children (McIntyre *et al*., 1997).

In animal models, the infecting serotype appears to determine recognisable differences in cerebral histopathology with serotype 1 causing more cortical haemorrhage, serotype 3 cortical necrosis and abscess formation and serotype 9V, subcortical abscess formation (Ostergaard *et al*., 2004).

The pathogenesis of pneumococcal meningitis was reviewed extensively by Weber (Weber, 2004) and Koedal *et al* (Koedel *et al*., 2002) and involves the co-ordinated activities of several virulence factors including neuramindase A, IgA protease, hyaluronidase, capsule, pneumococcal surface protein A and C, pneumolysin, choline binding protein A, phase variation and autolysin.

1.5.1.1.5 Conjunctivitis

Pneumococcal conjunctivitis tends to manifest as outbreaks (Leighton *et al*., 2003, Martin *et al*., 2003a, Medeiros *et al*., 1998) although sporadic cases are recognised (Porat *et al*., 2006). Serotype 3 and non-typeable pneumococci are associated with the development of acute conjunctivitis (Berron *et al*., 2005, Shouval *et al*., 2006). It is thought that non-typeable pneumococci have a selective predisposition to cause acute conjunctivitis (Porat *et al*., 2006) in both outbreaks (Medeiros *et al*., 1998, Leighton *et al*., 2003, Martin *et al*., 2003a) and sporadic cases (Shouval *et al*., 2006). It is therefore unfortunate that current vaccination strategies have no effect on non-typeable pneumococci (Porat *et al*., 2006, Martin *et al*., 2003a).

1.5.1.1.6 Endocarditis

Between 0.8-3.4% of patients with pneumococcal bacteraemia develop endocarditis (Lindberg and Fangel, 1999). There is an association with alcoholism (Bruyn *et al*., 1990) and a predilection for the aortic valve (Lindberg and Fangel, 1999, Bruyn *et al*., 1990). Affected valves can rapidly ulcerate and fail (Stewart and Flint, 1919). Embolic phenomena are common (Finland and Barnes, 1970), as are paravalvular abscesses (Bruyn *et al*., 1990). Even with appropriate antibiotic therapy, there is a fatal outcome in as high as 40-50% of cases (Finland and Barnes, 1970, Bruyn *et al*., 1990)

1.5.1.1.7 Cerebral Abscess

Pneumococcal cerebral abscesses are uncommon but are associated with significant morbidity (40% of survivors are left with neurological deficits) and high mortality (a case fatality rate of 35% in one series) despite appropriate antibiotic therapy and surgery (Grigoriadis and Gold, 1997). Serotype 3 pneumococci have historically been associated with cerebral abscesses (Fincher, 1946, Anonymous, 1970, Colman and Hallas, 1983).

1.5.1.1.8 Pleural Empyema

The epidemiology of pneumococcal pleural empyema was well documented by Glynn and Digby in 1923 for the Medical Research Council (Glynn and Digby, 1923). During the 1920s serotype 1 related pneumococcal empyema was a common manifestation with a high mortality of up to 53% of serotype 1 related disease and 77% of all cases of pneumococcal empyema in one British series (Whittle, 1929). Serotype 3 related pneumococcal empyema was rare (6% of all cases of pneumococcal empyema) (Whittle, 1929). These observations suggest that the current increase in serotype 1 related paediatric pneumococcal empyema being experienced in the United Kingdom (Fletcher *et al*., 2006, Eastham *et al*., 2004) may not necessarily be a new phenomenon but may be the initial signs of the serotype 1 pneumococcal population in this country reverting back to a previously high incidence.

1.5.1.2 Genomic Diversity in Invasive Pneumococcal Disease

It is unwise to make generalisations regarding the global epidemiology of IPD precisely because of the extent of genomic diversity and the fact that different geographical regions (Hausdorff *et al*., 2001) and different human host populations within a geographical region (Hausdorff *et al*., 2001, Flannery *et al*., 2004, Greenwood, 1999) can demonstrate vastly different incidence of IPD, with different pneumococcal serotypes and sequence types contributing. In addition, IPD is the result of the interaction between a dynamic population of bacteria and a dynamic human host population with changes over time being evident, especially in response to the use of antibiotics and vaccination policies (Kristinsson, 2008).

However, the informed use of antibiotics and vaccination programme planning depends on knowledge of regional pneumococcal epidemiology. For this reason, several global and regional networks currently exist which record pneumococcal epidemiology and antibiotic resistance patterns such as the WHO *Streptococcus pneumoniae* Global Disease Burden project, the Pneumococcal Global Serotype Project, Global Alliance for Vaccines and Immunizations Pneumococcal Vaccines Accelerated Development and Introduction Plan (GAVI PneumoADIP), Pneumococcal Molecular Epidemiology Network (PMEN), South Asian Pneumococcal Alliance (SAPNA), Asian Network for Surveillance of Resistant Pathogens (ANSORP), The Network for Surveillance of Pneumococcal Disease in the East Africa Region (NETSPEAR) or the Pan American Health Organisation (PAHO). Likewise there are also national surveillance programmes throughout the world such as the Centres for Disease Control and Prevention (CDC) Active Bacterial Core Surveillance/Emerging Infections Program Network in the USA and the Health Protection Agency (HPA) and Health Protection Scotland (HPS) in the United Kingdom.

Different studies have associated different serotypes with IPD and the results of some of these studies are displayed below in Table 1-1.

Serotype associated with IPD	Country	Reference
1, 4, 14, 18C	England	(Brueggemann *et al.*, 2003)
1, 5, 7	England, USA, Papua New Guinea, Kenya, Canada and Iceland	(Brueggemann *et al.*, 2004)
1, 4, 7F, 9V, 12F, 18C	Sweden	(Sandgren *et al.*, 2004)
1, 5, 12F	Israel	(Shouval *et al.*, 2006)

Table 1-1 Common pneumococcal serotypes associated with IPD in various countries.

Most recently, the Pneumococcal Global Serotype Project identified the most common serotypes causing IPD in children under 5 years of age. This shows that 7 to 11 serotypes account for over 80% of disease in every region and that Asia and Africa share the top 8 serotypes. Serotype 14 is most commonly isolated in all regions but in older children in Asia and Africa, serotype 1 was most common (O'Brien, 2008). This is compatible with earlier observations (Kyaw *et al.*, 2002, Hausdorff *et al.*, 2000a, Hausdorff *et al.*, 2000b).

In Scotland, at the start of the 21st century, the serogroups most commonly associated with IPD were 14, 9, 6, 19, 23, 8 and 4 (Kyaw *et al.*, 2003) although enhanced surveillance has shown that this has changed between 2003 and 2006 (Lamb *et al.*, 2008). Again using Scotland as an example, it appears that the number of MLST clones associated with IPD which circulate in a particular geographical region are limited and generally stable (Clarke *et al.*, 2004b, Clarke *et al.*, 2005, Clarke *et al.*, 2004a, Jefferies *et al.*, 2008)). This appears also to be true of antibiotic resistant clones as the number of such clones to be found in any individual country tends to be few (McGee *et al.*, 2001a, Smith *et al.*, 2006).

1.6 Studies of the Genomic Diversity of *Streptococcus pneumoniae*

Different species of bacteria demonstrate different degrees of genomic diversity. For instance, *Helicobacter pylori* (Alm *et al.*, 1999, Bjorkholm *et al.*, 2001) and *Campylobacter jejuni* show extensive genomic diversity (Dorrell *et al.*, 2001) while *Mycobacterium tuberculosis* does not and appears to show a high level of genomic conservation (Behr *et al.*, 1999). *S. pneumoniae* is an organism which exhibits a high degree of genomic diversity and natural populations appear to maintain a fluid state of genetic exchange in order to adapt to their environment. This section will document some

of the historical milestones which have aided understanding of this diversity of pneumococcal genomes, the mechanisms generating this diversity, the role of biofilms and regions of the pneumococcal genome which have been identified as being highly variable.

1.6.1.1 Historical background

In 1991 using restriction fragment digests of the pneumococcal strain R6 (an avirulent, serotype 2 unencapsulated strain derived from the strain D39 (Lanie *et al.*, 2007)), Gasc *et al* estimated this pneumococcal genome to be between 2,240 and 2,270kbp (Gasc *et al.*, 1991). It wasn't until 2001 that the fully sequenced genome for R6 was published (Hoskins *et al.*, 2001). Also published in 2001, was the sequenced genome of an isolate termed TIGR4 (Tettelin *et al.*, 2001), as it was a serotype 4 isolate being investigated by The Institute of Genomic Research (TIGR) which originated from the blood of a 30 year old man in Kongsvinger, Norway during the 1990s (Mann *et al.*, 2006, Orihuela *et al.*, 2004b, Bruckner *et al.*, 2004, Tettelin *et al.*, 2001). Although investigation of natural variation of individual alleles was possible by PCR and sequencing (Whatmore and Dowson, 1999), comparison of entire genomes was expedited by the manufacture of DNA microarrays based on the R6 and TIGR4 genomes allowing direct comparisons between these genomes and those of other clinical isolates by using comparative genomic hybridization (Bruckner *et al.*, 2004, Hakenbeck *et al.*, 2001).

Comparing the R6 and TIGR4 genomes was a milestone in realising the potential extent of diversity of pneumococcal genomes. The two genomes differ in size (R6 being 2Mb and TIGR4 being 2.16Mb and more similar to the other draft genomic sequence at that time for serotype 19F strain G54 which was 2.1Mb (Dopazo *et al.*, 2001)). R6 and TIGR4 differ in around 10% of their genes (Bruckner *et al.*, 2004). The R6 genome contains 69kb in six regions which are absent from TIGR4 and TIGR4 contains 157kb in twelve clusters which are absent from R6 (Bruckner *et al.*, 2004). Analysis of further genomes has identified many more genes which are not present in either R6 or TIGR4 genomes (Shen *et al.*, 2006a). As further isolates have been compared with R6 and TIGR4 genomes, it has become apparent that there are recognised regions of diversity within the pneumococcal genome (Bruckner *et al.*, 2004, Embry *et al.*, 2007, Obert *et al.*, 2006, Shen *et al.*, 2006a, Silva *et al.*, 2006, Hakenbeck *et al.*, 2001) and attempts have been made to try to elucidate whether there is a "core" genome required by the pneumococcus for either invasive disease or asymptomatic carriage (Obert *et al.*, 2006). Often spurious conclusions are drawn since no apparent acknowledgement that host factors may be influencing invasive disease presentation is appreciated (Hiller *et al.*, 2007, Obert *et al.*, 2006). Such comparisons are,

however, compatible with a distributed genome hypothesis which states that pathogenic bacteria possess a "supragenome" or gene pool which is much larger than that of any single isolate and that a large set of "non-core" genes is accessed to generate genomic diversity (Hiller *et al*., 2007). These studies have also demonstrated that, in addition to allelic variation, the pneumococcal genome demonstrates intra-species genic variation which refers to the absence or presence of certain genes (Hiller *et al*., 2007).

1.6.1.2 Mechanisms of Natural Diversity of the Pneumococcus

1.6.1.2.1 Horizontal gene transfer and competence

> "Horizontal gene transfer, or the acquisition of exogenous genetic material and its subsequent stable incorporation into a recipient genome, has been, and continues to be, a central force that drives bacterial evolution (Joyce *et al*., 2002)."

Joyce *et al* also note that,

> "Gene transfer events have been revealed through analyses of genome sequences, which differ in guanine and cytosine (G+C) content and codon usage at chromosomal locations that have recently acquired foreign DNA (Joyce *et al*., 2002)."

This acquisition of exogenous DNA by the pneumococcus depends on competence (Claverys, 2000), a state whereby DNA can be acquired from the extracellular environment during a short period in the pneumococcal growth cycle which is carefully controlled by the *com* operon of three genes (*comC*, *comD* and *comE*) which encode competence stimulating peptide (CSP), histidine kinase and a response regulator. CSP is exported into the extracellular environment by *comA* and *comB* and when CSP reaches a critical concentration, *comD* is activated, phosphorylating *comE* and resulting in the upregulation of several genes involved in permitting competence (Whatmore *et al*., 1999). *ComC* and *comD* themselves have multiple alleles (Whatmore *et al*., 1999, Pozzi *et al*., 1996).

The earliest descriptions of horizontal gene transfer in the pneumococcus related to the acquisition of penicillin resistance. Dowson *et al* in 1989 hypothesised that altered pbp2X genes arose from interspecies recombinational events with other species of streptococci (Dowson *et al*., 1989). In 1991, Laible *et al*, demonstrated that pbp2X had highly divergent regions likely resulting from interspecies recombinational events (Laible, 1991). Also in 1991, isolates of serogroup 19 and serogroup 23 were studied which had identical genes for pbp1A, pbp2A and pbp2X and it was concluded that these penicillin binding protein genes had been exchanged by horizontal gene transfer although the possibility of

horizontal exchange of capsular genes was not fully dismissed (Coffey *et al*., 1991). (Capsular switching as an example of horizontal gene transfer will be discussed below and the allelic diversity of pneumococcal surface proteins and other virulence factors will also be reviewed.) Further work published in 1993 found evidence of pbp2B gene transfer from *Streptococcus mitis* to the pneumococcus (Dowson *et al*., 1993) and that there was horizontal transfer of pbp2B from the pneumococcus to *Streptococcus oralis* (Coffey *et al*., 1993) showing that the horizontal transfer of genes in pneumococci was a bilateral process. Further evidence of the extent of horizontal gene transfer was later seen when isolates of *S. mitis* were discovered which harboured pneumolysin and autolysin – virulence factors which were thought, until that time, to be specific to the pneumococcus (Whatmore *et al*., 2000, Neeleman *et al*., 2004a, Neeleman *et al*., 2004b). Bruckner *et al* conclude that the high numbers of insertion elements evident in pneumococcal genomes and the high number of PTS systems suggests high flexibility in these genomes (Bruckner *et al*., 2004).

This accumulated evidence demonstrates that several pneumococcal genes exhibit many polymorphic alleles, many of which are mosaic genes which have been generated by intragenic recombination (Maiden, 1998, Bruckner *et al*., 2004). Feil *et al* have estimated that in the pneumococcal genome, recombination generates new alleles at a frequency ten time higher than mutation and that a single nucleotide is fifty times more likely to change through recombination than mutation (Feil *et al*., 2000a).

1.6.1.2.2 Capsular switching

Capsular switching is a form of horizontal gene transfer whereby the capsular genes coding for one serotype are exchanged for those of a different serotype (Coffey *et al*., 1998a, Coffey *et al*., 1999, Ramirez and Tomasz, 1999, Coffey *et al*., 1991). It is thought that this occurs relatively frequently in mixed pneumococcal populations (Coffey *et al*., 1998a) which are often found in the nasopharynx. The capsular gene cassette, flanked by genes *dexB* and *aliA*, is found at the same chromosomal location for all pneumococcal serotypes except serotype 37 (Claverys, 2000). This cassette contains up to 19 genes, several of which are serotype specific (Claverys, 2000). The serotype 37 capsule is coded by a single gene, *tts*, located 820kb distant to the capsular gene locus (Llull *et al*., 1999, Llull *et al*., 2000). Recently it has been observed that not only can the capsular locus be transferred spontaneously but also the adjacent pbps (Figure 1-1) can also be involved in a spontaneous recombination event both *in vitro* (Trzcinski *et al*., 2004, Coffey *et al*., 1999) and *in vivo* (Brueggemann *et al*., 2007, Coffey *et al*., 1999).

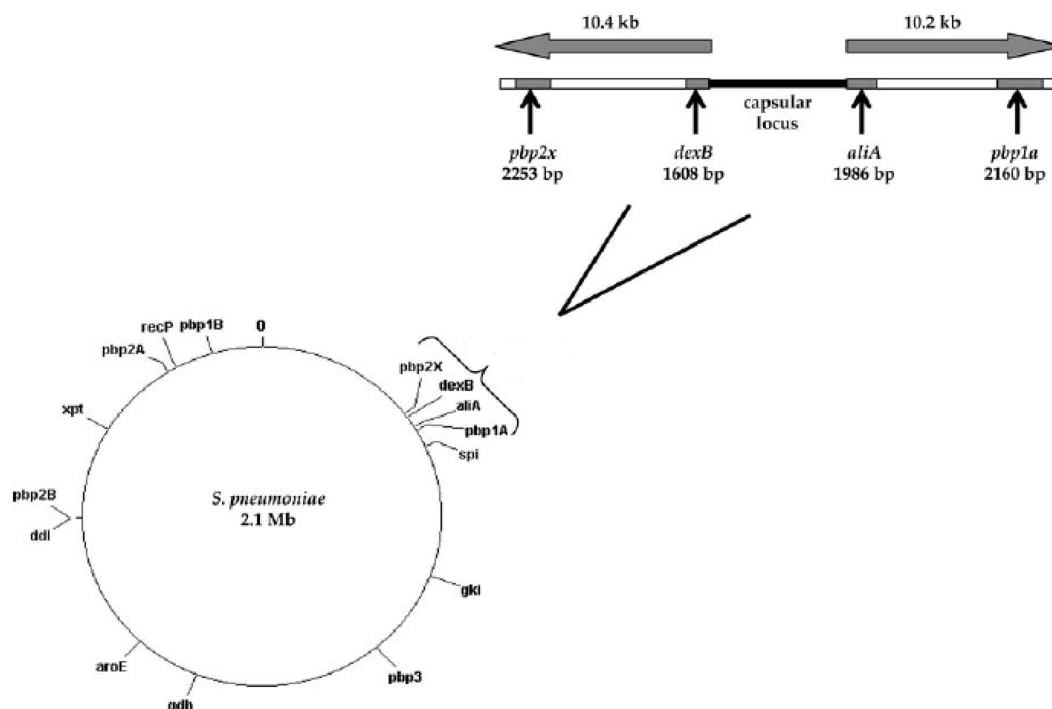


Figure 1-1Schematic diagram of the pneumococcal genome.

Adapted from (Brueggemann *et al.*, 2007) this identifies the location of the capsular locus and its flanking genes *(aliA and dexB)* in relation to pbps *(pbp1A, pbp1B, pbp2A, pbp2B, pbp2X, pbp3)* and housekeeping genes used for determination of MLST sequence types *(aroE, gdh, gki, recP, spi, xpt, ddl).*

The net effect of this is that a pneumococcal isolate with a particular sequence type can exist as several different serotypes depending on which capsular genes it contains and expresses (Jefferies *et al.*, 2004, Coffey *et al.*, 1998b). The relationship between serotype and genome is complex and switching capsules can increase or decrease virulence (Kelly *et al.*, 1994). Genetic factors other than just capsule influence virulence (Obert *et al.*, 2006).

This is of significant concern as there is accumulating evidence that after the introduction of PCV-7 within a population, serotype switching allows pneumococcocal populations to gradually escape the effects of the limited valency vaccines by disposing of the capsular genes affected by vaccination and replacing them with serotypes which the vaccine does not cover (Brueggemann *et al.*, 2007).

1.6.1.2.3 Serotype replacement

Serotype replacement is also driven by the introduction of pneumococcal conjugate vaccines. By targeting only a limited number of serotypes, it is observed that non-vaccine

related serotypes can replace serotypes covered by vaccines in the nasopharynx, permitting increased carriage and transmission of non-vaccine serotypes in the community and subsequently, invasive disease from non-vaccine serotypes (Spratt and Greenwood, 2000, Brueggemann *et al*., 2007). Unlike capsular switching (Ramirez and Tomasz, 1999), the genotype of these replacement serotypes maintain a sequence type in a relationship which is the same as that which was seen prior to vaccine introduction. Evidence for this is particularly dramatic for an increased incidence of serotype 1 related, complicated pneumonia in Utah, USA (Byington *et al*., 2005c), serotype 3 related otitis media (McEllistrem *et al*., 2007, McEllistrem *et al*., 2005) and 19A invasive disease in Alaska, USA (Singleton *et al*., 2007) and New York, USA (Pichichero and Casey, 2007).

One beneficial effect of serotype replacement however, is a decreased incidence of penicillin resistance among pneumococci after conjugate vaccine introduction due to their predominant serotypes being included in such vaccine formulations (Byington *et al*., 2005c, Spratt and Greenwood, 2000). Of greater concern though are manifestations of serotype replacement resulting in the emergence of multiply antibiotic resistant pneumococci of non-vaccine serotypes (Pichichero and Casey, 2007).

1.6.1.2.4 Phase variation

The pneumococcus exhibits phase variation whereby one isolate of the same serotype and MLST can exist in two distinct phenotypes (Bruckner *et al*., 2004, Weiser *et al*., 1994). These two forms, opaque and transparent, are most distinct when grown on translucent solid media (Weiser *et al*., 1994). Ring *et al* found that the transparent phenotype of phase variation increased the ability of pneumococci to cross the blood brain barrier by as much as six fold compared to the opaque phenotype (Ring *et al*., 1998). The transparent phenotype has more cell wall phosphorylcholine, less capsular polysaccharide and has different surface proteins than its opaque counterpart indicating diversity of associated cell surface components which may influence virulence (Ring *et al*., 1998, Weiser and Kapoor, 1999). Higher rates of transformation occur in transparent variants compared to opaque so less capsular material (transparent) appears to enhance transformation (Weiser and Kapoor, 1999). Opaque variants survive better than transparent in the bloodstream and appear more resistant to phagocytosis (Obaro and Adegbola, 2002). Weiser *et al* conclude,

> "An isolate should be considered a mixed population of phenotypes which differ in amounts of capsular polysaccharide, teichoic acid and choline binding proteins (Weiser and Kapoor, 1999)."

1.6.1.2.5 Bacteriophages

It is estimated that as many as 75% of clinical pneumococcal isolates contain temperate bacteriophages (Ramirez *et al*., 1999) of which four have been sequenced (Lopez and Garcia, 2004). Bacteriophages have also been found commonly in nasopharyngeal carriage isolates (Sa-Leao *et al*., 2002). Lytic phages appear to contribute to natural transformation of the pneumococcus by expanding the reservoir of exogenous DNA available for incorporation into the pneumococcal genome (Lopez *et al*., 2000, Ramirez *et al*., 1999) and may also alter virulence by creating chimeric enzymes through recombination after excising DNA from genes (Lopez *et al*., 2000).

1.6.1.2.6 Plasmids

It is thought that plasmids are not a common means by which diversity is introduced into the pneumococcal genome (Bruckner *et al*., 2004). The role of conjugation in pneumococcal diversity is also unclear (Bruckner *et al*., 2004). However, the acquisition and loss of plasmids by pneumococci has been demonstrated as comparison of the genome of R6 with its progenitor strain D39, shows loss of its pDP1 plasmid (Lanie *et al*., 2007).

1.6.1.2.7 The role of biofilms in propagating pneumococcal genomic diversity

It is established that pneumococci form biofilms and that these contribute to certain disease manifestations such as otitis media (McEllistrem *et al*., 2007, Hall-Stoodley *et al*., 2006). DNA release and transformation are a part of the biofilm-related life cycle and readily occurs in many bacteria with released DNA stabilising the biofilm structure (Molin and Tolker-Nielsen, 2003). Phase variation is also a feature of pneumococcal biofilms (McEllistrem *et al*., 2007). It is likely that such an environment conducive to horizontal gene transfer has a role in generating genomic diversity in pneumococcal populations in order to aid survival in changing environmental conditions (Boles *et al*., 2004).

1.6.1.2.8 Regions of diversity in the pneumococcal genome

Several groups have used microarray CGH to demonstrate regions of diversity in the pneumococcal genome using the TIGR4 genome as reference. Initial reports of these regions were by Hakenbeck *et al* who detected 10 clusters where hybridization signals indicated no hybridization in 20 diverse pneumococcal isolates from different serotypes and different geographical regions when hybridized against a serotype 4 strain (KNR.7/87) using an Affymetrix oligonucleotide array (Hakenbeck *et al*., 2001). These regions varied in size from 9kb to 37kb each (Hakenbeck *et al*., 2001).

In studies identifying regions of diversity in the TIGR4 genome (Silva *et al*., 2006, Bruckner *et al*., 2004, Embry *et al*., 2007), there has been no standardised definition of what should characterise such a region. This has resulted in some groups considering there to be 13 regions of diversity of 3.7kb to 40.3kb in size (Embry *et al*., 2007, Obert *et al*., 2006) while others have considered regions of diversity to be as small as 1.7kb and so document 25 such regions (Silva *et al*., 2006). Several regions of diversity relate to genes which have an identified virulence function (Embry *et al*., 2007).

Bruckner *et al* have also identified 6 regions of diversity within the R6 genome (Bruckner *et al*., 2004) and Shen *et al*, identified 58 novel sequences in clinical isolates not present in TIGR4 or R6, some of which were virulence associated (Shen *et al*., 2006a).

Region of Diversity	Location in TIGR4 Genome	Size (kb)	Encoded Virulence Determinants	References
RD 1	SP0067-0074	9.0	Zinc Metalloproteinase (ZmpC)	(Silva *et al.*, 2006, Bruckner *et al.*, 2004)
RD 2	SP0109-0115	5.8		(Silva *et al.*, 2006, Bruckner *et al.*, 2004)
RD 3	SP0163-0168	5.6		(Silva *et al.*, 2006, Bruckner *et al.*, 2004, Obert *et al.*, 2006)
RD 4	SP0346-0360	14.2	Capsular Polysaccharide synthesis operon	(Silva *et al.*, 2006, Bruckner *et al.*, 2004)
RD 5	SP0378-0380	3.3		(Silva *et al.*, 2006)
RD 6	SP0394-0397	5.4		(Silva *et al.*, 2006, Bruckner *et al.*, 2004)
RD 7	SP0460-0468	12.6		(Silva *et al.*, 2006, Bruckner *et al.*, 2004)
RD 8	SP0473-0478	7.1		(Silva *et al.*, 2006)
RD 9	SP0531-0544	5.6		(Silva *et al.*, 2006, Bruckner *et al.*, 2004)
RD 10	SP0643-0648	11.0	RlrA pathogenicity islet	(Silva *et al.*, 2006)
RD 11	SP0644-0666	8.0		(Silva *et al.*, 2006)
RD 12	SP0692-0700	4.4		(Silva *et al.*, 2006, Obert *et al.*, 2006)
RD 13	SP0888-0891	1.7		(Silva *et al.*, 2006, Bruckner *et al.*, 2004)
RD 14	SP0949-0954	7.9		(Silva *et al.*, 2006)
RD 15	SP1050-1065	11.9	Pneumococcal Pathogenicity Island 1 (PPI1)	(Silva *et al.*, 2006, Bruckner *et al.*, 2004, Obert *et al.*, 2006)
RD 16	SP1129-1147	9.2		(Silva *et al.*, 2006, Bruckner *et al.*, 2004, Obert *et al.*, 2006)
RD 17	SP1315-1352	33.7		(Silva *et al.*, 2006, Bruckner *et al.*, 2004, Obert *et al.*, 2006)
RD 18	SP1433-1444	12.1		(Silva *et al.*, 2006, Bruckner *et al.*, 2004)
RD 19	SP1612-1622	10.3		(Silva *et al.*, 2006, Bruckner *et al.*, 2004, Obert *et al.*, 2006)
RD 20	SP1756-1773	34.8	PsrP-sec Y2A2 pathogenicity island	(Silva *et al.*, 2006, Bruckner *et al.*, 2004, Obert *et al.*, 2006)
RD 21	SP1793-1799	5.3		(Silva *et al.*, 2006, Bruckner *et al.*, 2004)
RD 22	SP1828-1830	3.2		(Silva *et al.*, 2006)
RD 23	SP1911-1918	3.2		(Silva *et al.*, 2006)
RD 24	SP1948-1955	9.4		(Silva *et al.*, 2006, Bruckner *et al.*, 2004)
RD 25	SP2159-2166	5.3		(Silva *et al.*, 2006, Bruckner *et al.*, 2004, Obert *et al.*, 2006)

Table 1-2 Regions of diversity in the TIGR4 genome, size and putative virulence functions as determined by CGH in previous studies.

Region of Diversity	Location in R6 Genome	Size (kb)	Function	Reference
RD 1	spr0102 to spr0119	14.4	Arginine biosynthesis	(Bruckner *et al.*, 2004)
RD 2	spr0311 to spr0323	9.4	Capsule biosynthesis (equivalent to SP0347-0360)	(Bruckner *et al.*, 2004)
RD 3	spr0955 to spr0971	17.1	Macrolide efflux *mefE* related (equivalent to SP1054-1064)	(Bruckner *et al.*, 2004)
RD 4	spr1184 to spr1198	14.3	ABC transporter (equivalent to SP1309-1337)	(Bruckner *et al.*, 2004)
RD 5	spr1403 to spr1404	9.6	Cell wall anchor protein	(Bruckner *et al.*, 2004)
RD 6	spr1618 to spr1621	4.1	ABC transporter (equivalent to SP1796-1799)	(Bruckner *et al.*, 2004)

Table 1-3 Regions of Diversity in the R6 genome, size and putative virulence functions as determined by CGH.

1.6.1.3 Genomic Diversity and Temporal Changes Illustrated by Emerging and Significant Serotype and MultiLocus Sequence Type Combinations

It is clear that pneumococcal populations are dynamic and although there is much concern regarding how vaccination policies are altering these population structures, it is evident that noticeable changes in the population structures were occurring prior to the introduction of conjugate vaccines. This section will focus on a selection of serotype and MLST combinations which illustrate this dynamic diversity and which feature in later chapters.

1.6.1.3.1 Serotype 1

Figure 1-2 demonstrates the burden of disease caused by serotype 1 (Type I) pneumococci respectively in Glasgow from the 1920s until the 1980s. There is a marked decline in serotype 1 disease over the 20th century, even recognising that many of the original papers in the pre-antibiotic era include both patients who received anti-pneumococcal serum and those who did not in their datasets and that case definitions are not standardised. Figure 1-3 illustrates declining numbers of reported cases of pneumonia in Glasgow during the 1950s and 1960s (which may be a product of the Clean Air Act passed in 1956, the use of antibiotics after MRC trials demonstrated their utility in 1951, the associated decline in cases of pulmonary tuberculosis or a general recovery from the social and economic effects of World War II) which could be influencing the reported numbers of cases of serotype 1 associated lobar pneumonia from the 1950s but it is noteworthy that the same decline in serotype 1 reporting was seen in Edinburgh (Figure 1-4) and in Boston, USA (Figure 1-5) suggesting that a genuine decline in serotype 1 associated pneumococcal disease was occurring in the United Kingdom and the USA during the 1950s and 1960s.

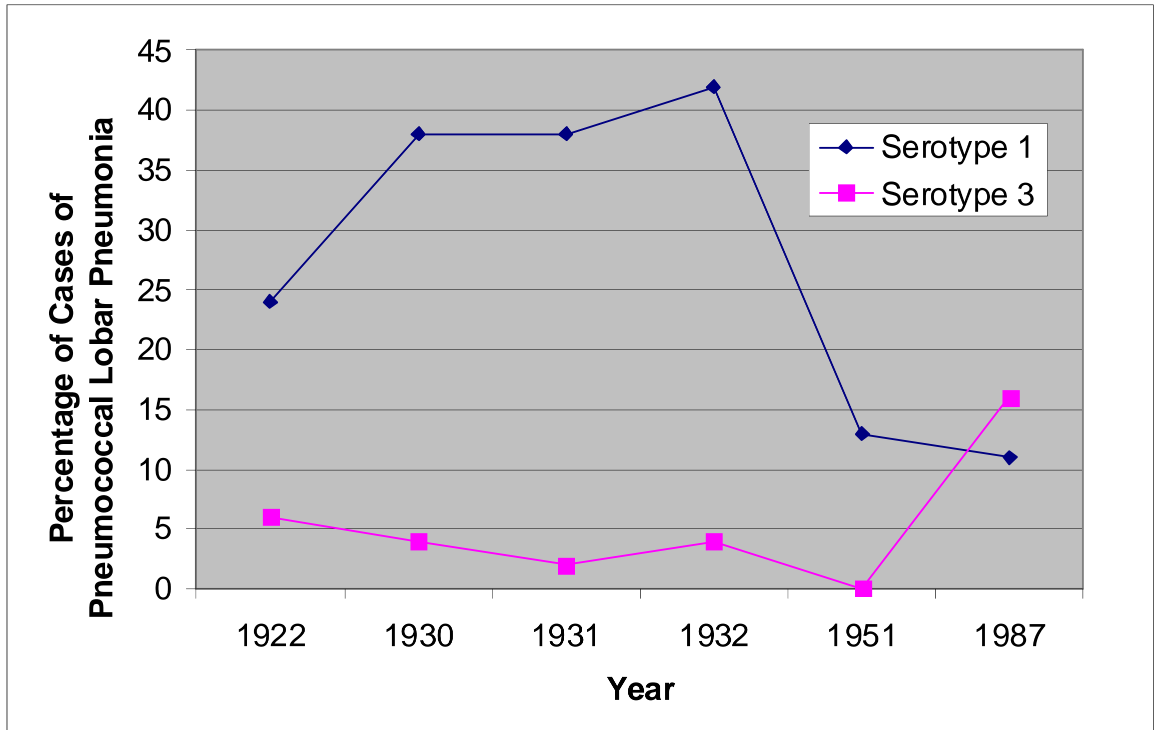


Figure 1-2 Proportions of cases of hospitalised pneumococcal lobar pneumonia due to serotypes 1 and 3 documented in Glasgow over the 20th Century.

Data taken from (Grant, 1922, Christie, 1932, Grist *et al.*, 1952, Smart, 1987, Cowan *et al.*, 1932)

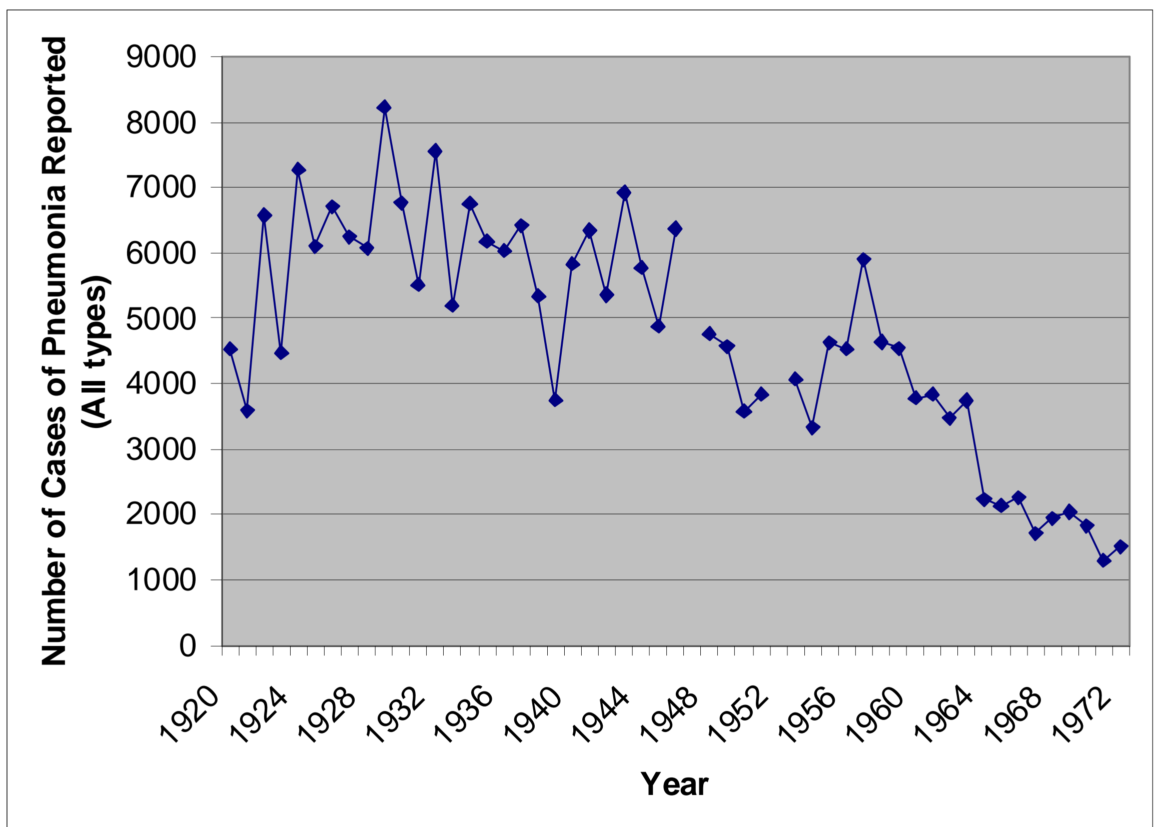


Figure 1-3 Annual number of cases of pneumonia (all types) reported in Glasgow (1920-1972).

Data taken from Reports of the Medical Officer of Health, City of Glasgow 1920 to 1971. It is noteworthy that 1922, 1930, 1952, 1957 and 1970 were documented as years when there were recognised influenza epidemics in Glasgow.

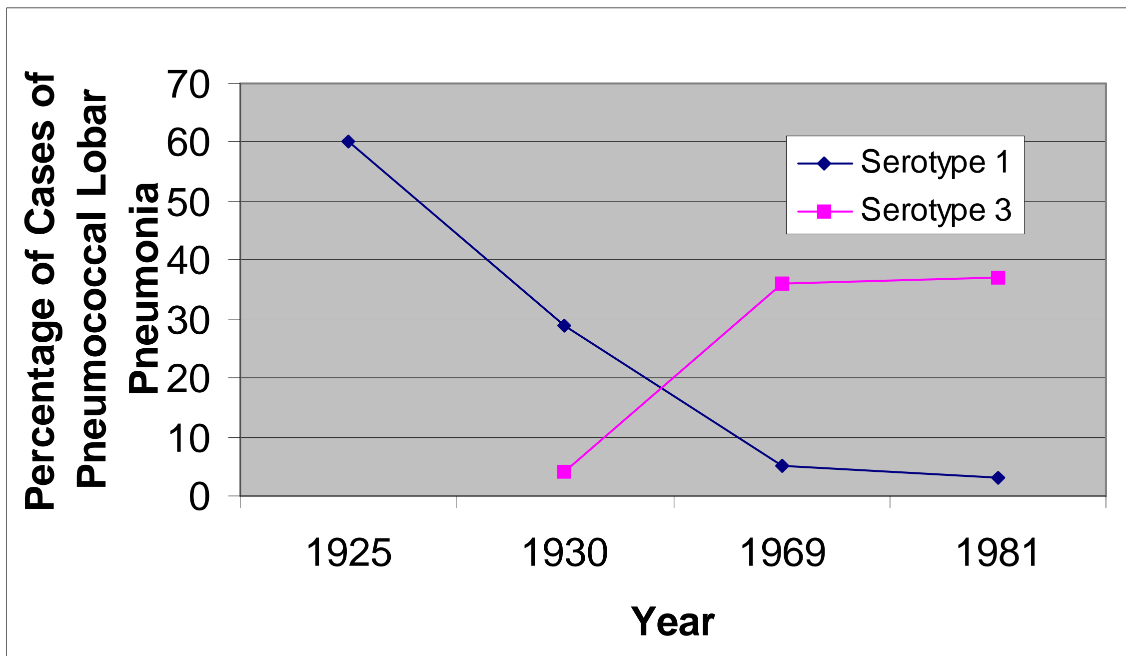


Figure 1-4 Proportions of cases of hospitalised pneumococcal lobar pneumonia in Edinburgh due to serotypes 1 and 3 over the 20th Century.

Records from Edinburgh reproduce the same dramatic fall in the proportion of cases of serotype 1 associated disease that was documented in Glasgow during the 1950s and 1960s (Figure 1-2) and Boston, USA (Figure 1-4) and concurrent rise in the proportion due to serotype 3.

Data taken from (Morgan *et al.*, 1983, Calder *et al.*, 1970, Davidson, 1925, Alston and Stewart, 1930)

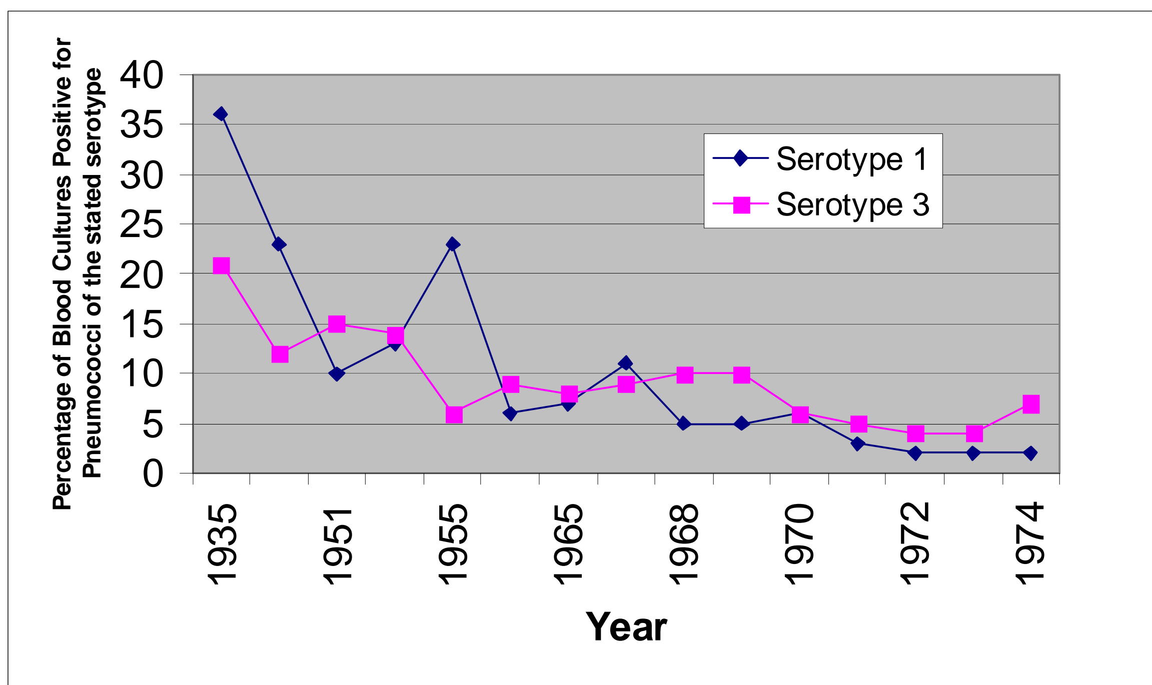


Figure 1-5 Proportion of blood cultures growing pneumococci at Boston City Hospital due to serotype 1 and serotype 3 (1935-1974).

This also shows a dramatic fall in cases of serotype 1 associated disease during the 1950s and 1960s similar to that seen in Scotland but without an associated rise in proportion of cases due to serotype 3 during the early 1970s. Data from (Finland and Barnes, 1977b).

It is clear from contemporary reports of IPD in both Scotland (Kirkham *et al*., 2006, Lamb *et al*., 2008, Diggle and Edwards, 2006) and England and Wales (George *et al*., 2006) as well as other European countries (Hausdorff, 2008, Hanquet *et al*., 2008) that cases due to serotype 1 are now increasing in incidence. Figure 1-6 documents this increase in Scotland since 1993.

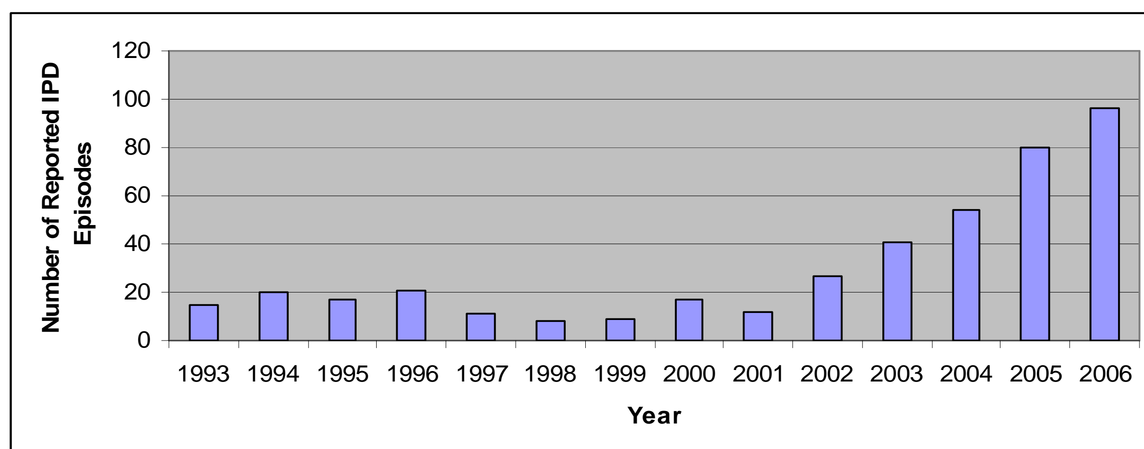


Figure 1-6 Number of episodes of invasive serotype 1 associated disease seen in Scotland. Data relates to episodes reported to the Scottish Meningococcal and Pneumococcal Reference Laboratory 1993-2006.

Serotype 1 accounts for > 6% of IPD in many geographical regions, but it is not included in the PCV-7 (Hausdorff *et al*., 2000b) although it is included in a new 13 valent conjugate vaccine (Kieninger *et al*., 2008). Unlike Europe and the USA, it is one of the commonest serotypes causing IPD in children in selected studies from Africa and Asia (Hausdorff *et al*., 2000b) and Latin America (Coral *et al*., 2001). It is associated with very high odds ratios for invasiveness in children (Brueggemann *et al*., 2004). This serotype also has an association with causing outbreaks of IPD (Jefferies *et al*., 2007) and is more commonly cultured from blood than CSF (Hausdorff *et al*., 2000a). Serotype 1 has also been determined to have a low case fatality rate in Sweden (Sandgren *et al*., 2005, Berg *et al*., 2006) and a lower relative risk of death in Denmark (Martens *et al*., 2004).

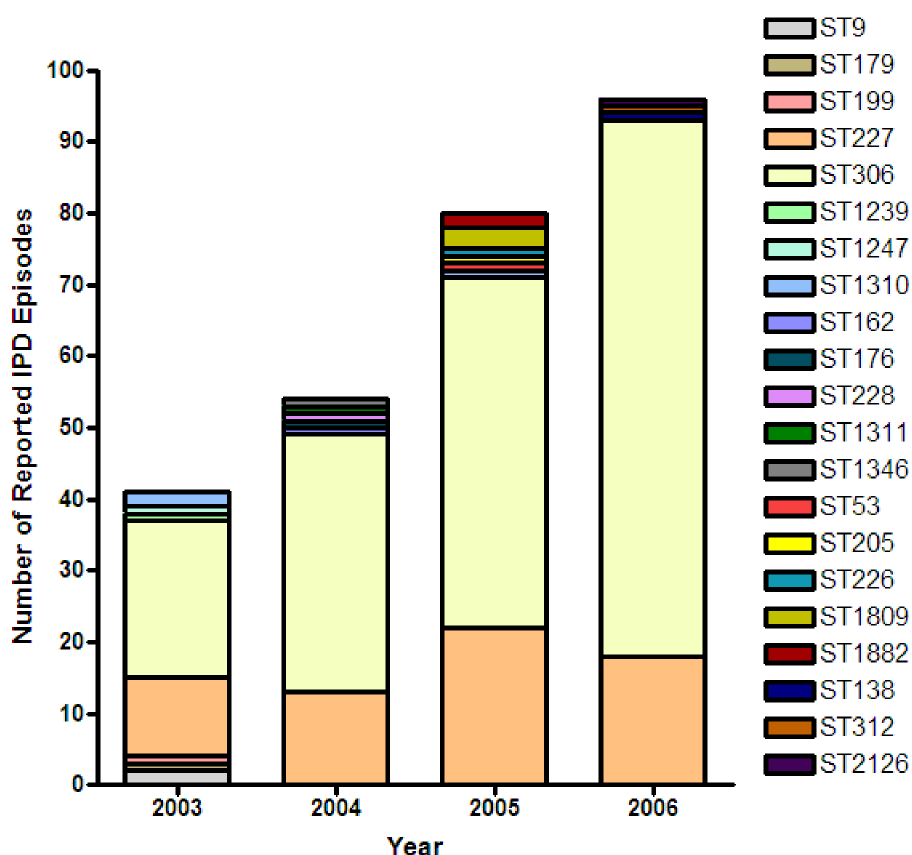


Figure 1-7 MLST sequence types associated with serotype 1 capsule in Scotland causing invasive pneumococcal disease (2002-2006).

1.6.1.3.1.1 ST227

Within the serotype 1 related IPD cases in Scotland since 2002, there are several sequence types represented but predominantly ST227 and ST306 (Figure 1-7). The proportion due to ST227 has remained stable since 2002. It was the predominant, stable serotype 1 clone in the USA between 1993 and 2002 (Brueggemann and Spratt, 2003, Gonzalez *et al.*, 2004). In the immediate period after PCV-7 introduction in Utah (pre 2003), ST227 accounted for all cases of serotype 1 related paediatric empyema (Byington *et al.*, 2005c) but since 2003 this has not been the case (Byington *et al.*, 2008). ST 227 is associated with a low case fatality rate in Sweden (Sjostrom *et al.*, 2006).

1.6.1.3.1.2 ST306

It is of some concern that the proportion of serotype 1 related IPD due to ST306 is increasing each year in Scotland (Lamb *et al.*, 2008). This phenomenon was also documented 10 years ago in Sweden (Hedlund *et al.*, 2003, Henriques Normark *et al.*, 2001). ST306 accounts for the majority of the current increase in serotype 1 IPD in Scotland (Lamb *et al.*, 2008, Jefferies *et al.*, 2008). It is possible that ST306 is a new clone which has entered the Scottish population but it is also possible that it is a clone that was

prevalent during the early 20th century, was lost from the Scottish pneumococcal population for whatever reason and is now returning, potentially because of a lack of herd immunity to it as people who may have been exposed to it and developed some immunity during the 1920s and 1930s come to the end of their lifespan. It is certainly clear that a proportion of elderly, non-vaccinated, patients in a Glasgow Geriatric Hospital had pre-existing immunity to serotype 1 pneumococci (Thakker *et al.*, 1998) presumably the result of exposure to it earlier in their lifetimes.

ST306 has also recently emerged in Utah, USA as a common clone associated with paediatric empyema (Byington *et al.*, 2008) although it is not clear whether this is a phenomenon associated with serotype replacement resulting from pressure produced by PCV-7 or whether it is solely a property of ST306 which allows it to expand rapidly in a population as it is known that ST306 harbours mutations in virulence factors which may alter its behaviour (Kirkham *et al.*, 2006). ST306 is also the dominant serotype 1 clone found in some South Pacific islands and has been responsible for invasive disease outbreaks in 1999 and 2007 (Le Hello *et al.*, 2008.).

The association of ST306 with greater survival has also been documented in a mouse model of pneumonia and bacteraemia where ST306 was associated with the production of less tumor necrosis factor when compared to other sequence types associated with invasive disease (Sandgren *et al.*, 2005). ST 306 is also associated with a low case fatality rate in Sweden (Sjostrom *et al.*, 2006) and, unusually, has been detected as a carriage associated isolate in Portugal (Nunes *et al.*, 2007).

1.6.1.3.2 Serotype 3

Serotype 3 pneumococci have always been noted to have a distinctive phenotypic appearance with substantial mucoid capsule (Schottmuller, 1903) (or “slime layer” (Wood and Smith, 1949)). More recently it has been determined that duplications in the *cap3A* gene in the type 3 capsule locus are associated with high frequency phase variation (Waite *et al.*, 2001). Properties of this polysaccharide capsule allow it to interact with Dendritic cell-specific ICAM 3 grabbing non-integrin (DC-SIGN) while most other pneumococcal serotypes do not (Koppel *et al.*, 2005). Likewise, only serotype 3 pneumococci appear to have a gene called *gadA* which encodes a protein similar to human glutamate decarboxylase (Garcia and Lopez, 1995).

Serotype 3 is a common cause of AOM (Hausdorff *et al.*, 2000a, McEllistrem *et al.*, 2007, Shouval *et al.*, 2006) where biofilm formation may be important in its pathogenesis. It also

causes acute conjunctivitis and it is postulated that serotype 3 possesses virulence factors which predispose it to mucosal sites (Shouval *et al*., 2006).

Serotype 3 is rarely carried by Israeli children (Shouval *et al*., 2006) although it is a serotype 3 associated MLST (ST180) which is most strongly associated with pneumococcal carriage by children in Oxfordshire (Brueggemann *et al*., 2003).

In Sweden, greater severity of IPD and higher case fatality rates from IPD are due to serotype 3 disease when compared to all other serotypes (Sjostrom *et al*., 2006). Serotype 3 associated IPD also has the highest relative risk of death in Denmark (Martens *et al*., 2004).

Globally, serotype 3 consistently comprises only 1-2% of IPD in young children (Hausdorff, 2007). In a meta-analysis of serogroup specific odds ratios for invasiveness in children , serotype 3 was one of the least likely to cause IPD (Brueggemann *et al*., 2004). Paradoxically, it is often one of the commonest serotypes to cause IPD in adults and the elderly where it is associated with severe clinical manifestations and poor outcomes (Inverarity *et al*., 2008, Gransden *et al*., 1985). Figure 1-2 demonstrates the burden of disease caused by serotype 3 (Type III) pneumococci in Glasgow and Figure 1-4 from Edinburgh from the 1920s until the 1980s. The serotype 3 population does not appear to be increasing at present in Scotland (Lamb *et al*., 2008). There are many sequence types among serotype 3 isolates in Scotland but this serotype is predominantly comprised of ST180 as seen in Figure 1-8 and Figure 1-9. It is noteworthy that serotype 3 isolates of different genotypes may have different virulence in mice (Kelly *et al*., 1994, Ren *et al*., 2003). Genes associated with virulence have been determined in serotype 3 pneumococci by means of signature-tagged mutagenesis (Lau *et al*., 2001).

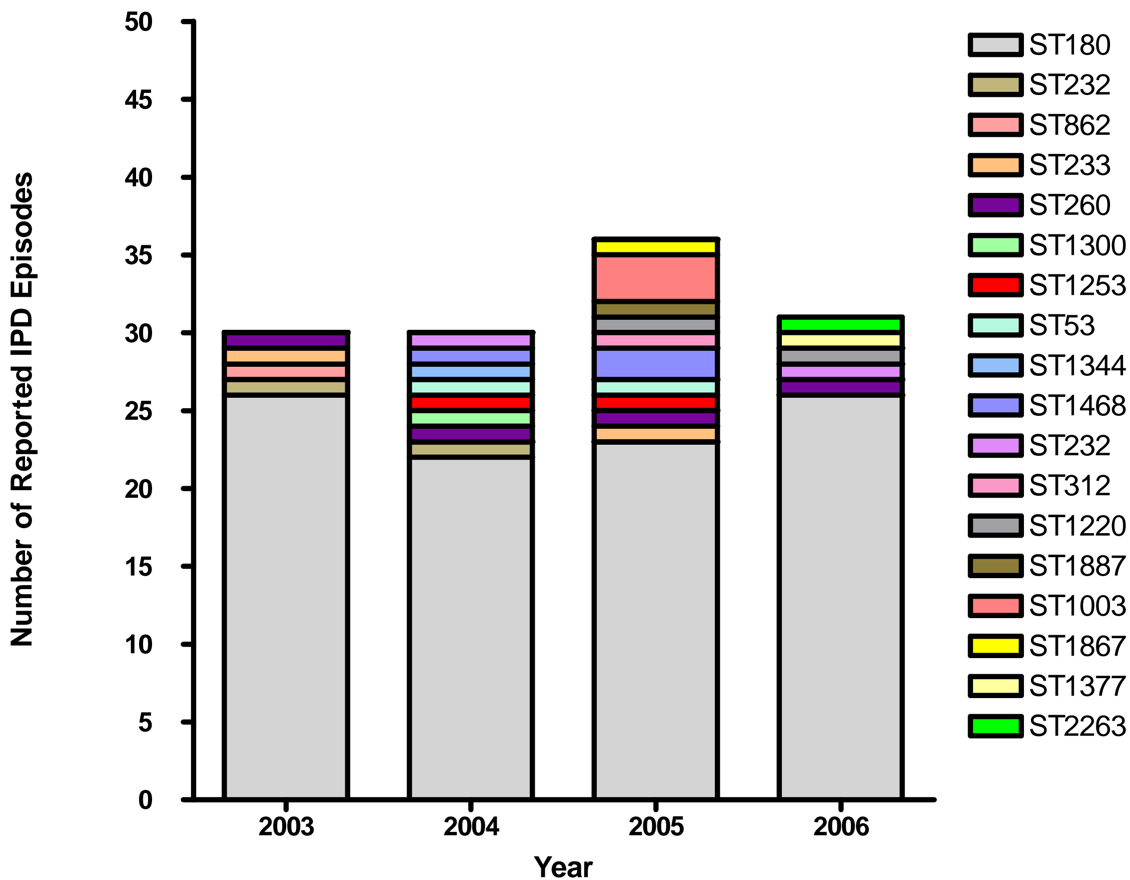


Figure 1-8 MLST sequence types associated with serotype 3 capsule in Scotland causing invasive pneumococcal disease (2002-2006).

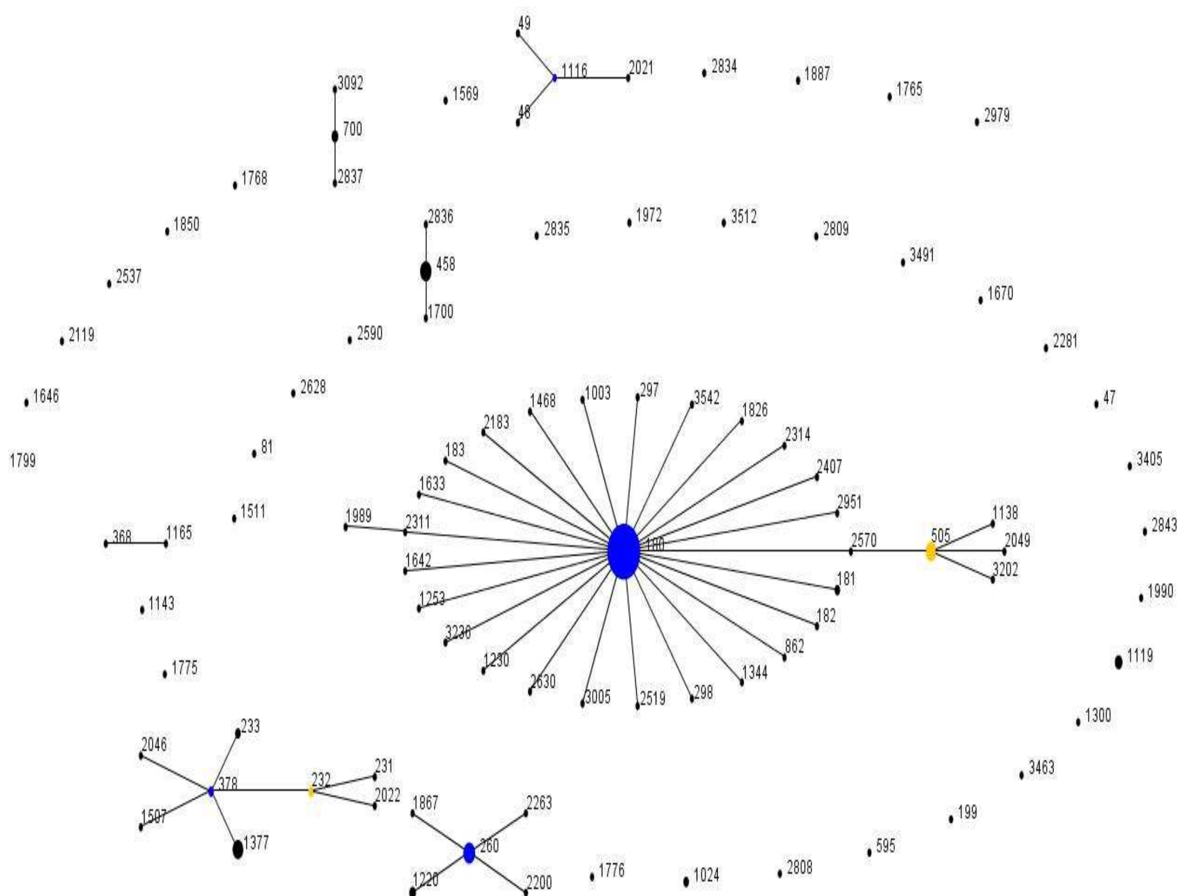


Figure 1-9 Relationships of MLST sequence types constructed using eBURST version 3[4] of all isolates in the MLST database which express serotype 3 capsule.

Blue dots indicate STs which are founders of clones from which single locus variants are demonstrated in black. Yellow dots indicate STs which are subgroup founders. ST180 is the dominant founder clone of serotype 3.

In countries where use of PCV-7 is more established, serotype 3 is increasing in incidence, which is likely to be a result of serotype replacement (Byington *et al.*, 2008, Hicks *et al.*, 2007).

1.6.1.3.2.1 ST180

Serotype 3 ST180 has emerged since 2001 as a cause of paediatric pneumococcal empyema in Utah, USA (Byington *et al.*, 2008) and in Spain (Obando *et al.*, 2008) and has also emerged as a non vaccine serotype in Alaska, USA causing IPD in adults and children (Miernyk *et al.*, 2008) as well as paediatric empyema (Singleton *et al.*, 2008). In addition, ST180 is a common sequence type associated with AOM (McEllistrem *et al.*, 2005). ST180 is associated with the highest case fatality rate from IPD in Sweden (Sjostrom *et al.*,

[4] http://spneumoniae.mlst.net/eburst {accessed 20th December 2008}

2006). It has also been determined in children from Oxfordshire that ST180 rarely caused IPD but was highly associated with nasopharyngeal carriage (Brueggemann *et al*., 2003).

The neuraminidase gene *nanB* appears to be present in all ST180 isolates while *nanC* is absent from all ST180 isolates (Pettigrew *et al*., 2006).

1.6.1.3.3 Serotype 4

Serotype 4 pneumococci are an important cause of IPD and can cause severe invasive disease in animal models (Sandgren *et al*., 2005) although in humans they have been associated with milder disease manifestations (Sjostrom *et al*., 2006).

1.6.1.3.4 Serotype 14

At the start of the 21st century, serotype 14 was the commonest serotype to cause IPD in children in the USA, Europe (including Scotland (Kyaw *et al*., 2003, Clarke *et al*., 2004c)) and Latin America and was still among the top five serotypes associated with paediatric IPD in Africa and Asia (Hausdorff *et al*., 2000b). It is more commonly cultured from blood than CSF (Hausdorff *et al*., 2000a) and is associated with an intermediate level of mortality (Henriques *et al*., 2000). Serotype 14 pneumococci also interact with the dendritic pathogen receptor DC-SIGN (Koppel *et al*., 2005).

Serotype 14 has recently been identified by PCR as a cause of paediatric empyema in the United Kingdom from culture negative pleural fluid samples (Sheppard *et al*., 2008). Overuse of macrolide antibiotics has been cited as contributing to the spread of multiply resistant clones of serotype 14 pneumococci (including ST9) in Europe (Dias and Canica, 2004).

1.6.1.3.4.1 ST9

In children from Oxfordshire, ST9 was mostly associated with causing IPD (Brueggemann *et al*., 2003). ST9 is almost 5 times more common in blood than it is in non-invasive sites (Amezaga *et al*., 2002) and has been associated with meningitis (Urwin *et al*., 1996). It is also associated with increased expression of a variant form of glyceraldehyde-3-phosphate dehydrogenase (GAPDH) which may relate to its virulence (Cash *et al*., 1999). The ST9 clone is associated with the M phenotype for macrolide resistance which is associated with the *mefA* gene but with a sequence more commonly associated with *Streptococcus pyogenes* (Amezaga *et al*., 2002). The PMEN clone, England ^{14}-9, is a serotype 14, ST9 clone which can be found in Scotland but here often has lower antimicrobial minimum inhibitory concentrations compared to the original clone suggesting some diversity within the clone (Smith *et al*., 2006).

1.6.1.3.5 Serotype 19A

Serotype 19A is not prevented by PCV-7. In countries which introduced this vaccine into their paediatric schedule around the year 2000, such as the USA, there was later a noticeable increase in the incidence of serotype 19A related IPD (Kristinsson, 2008, Weatherholtz *et al*., 2008, Singleton *et al*., 2008, Carvalho *et al*., 2008, Pai *et al*., 2005, Hicks *et al*., 2007). To complicate matters there is evidence from countries which did not introduce the vaccine, such as Scotland and Germany, that serotype 19A was responsible for significant levels of IPD (Ruckinger *et al*., 2008, Lamb *et al*., 2008, Clarke *et al*., 2004c) and that the incidence of serotype 19A IPD had also increased (Kristinsson, 2008, Hanquet *et al*., 2008, Byington *et al*., 2006). So it was not initially clear whether this had been an example of serotype replacement in response to the vaccine or the acquisition of a more virulent clone of serotype 19A. It is also apparent that serotype 19A can be found causing asymptomatic carriage (see Chapter 9).

More perplexing was the discovery that the circulating serotype 19A population in the USA since the introduction of PCV-7 was genetically diverse as shown in Figure 1-10 and Figure 1-11 (Pai *et al*., 2005). For instance, some isolates are resistant to multiple antibiotics while others are not (Brueggemann *et al*., 2007, Kristinsson, 2008, Pichichero and Casey, 2007, Singleton *et al*., 2007). More strikingly, Brueggemann *et al* demonstrate that although some serotype 19A isolates are genotypically identical to strains circulating in the USA since the 1990s (and likely to have been selected by serotype replacement), there are others which clearly have evidence of having been produced as the result of capsular switching (Figure 1-11) whereby the capsule is that of 19A while the genotype is identical to that of a serotype 4 clone (Brueggemann *et al*., 2007). They postulate that at least two vaccine escape capsular switch events have occurred. The main event resulted in a penicillin nonsusceptible serotype 19A with genotype ST695 which arose from recombination between a recipient serotype 4 ST695 and a donor serotype 19A ST199. This likely occurred around 2003 and resulted in clonal expansion through the Northeast USA because of its antibiotic resistance (Hanage *et al*., 2008, Pai *et al*., 2005). A second recombination event occurred between a donor serotype 19A ST2365 and a recipient serotype 4 ST247 (Brueggemann *et al*., 2007). There is also evidence from Utah that a serotype 19A of sequence type ST667 (which is more usually seen in serotypes 14 and 19) is causing IPD (Byington *et al*., 2008) and from Atlanta and Massachusetts that a serotype 19A of sequence type ST320 (acquired from the serotype 19F multiply resistant PMEN clone Taiwan^{19F}-14) accounts for a multiply antibiotic resistant clone which is resistant to penicillins, macrolides, clindamycin, tetracyclines, co-trimoxazole, cefuroxime and meropenem (Chancey *et al*., 2008, Carvalho *et al*., 2008, Hanage *et al*., 2008). In Alaska

though, the serotype 19A clone which has expanded is associated with ST 172 and has been associated with a fall in antibiotic resistance (Singleton *et al*., 2007).

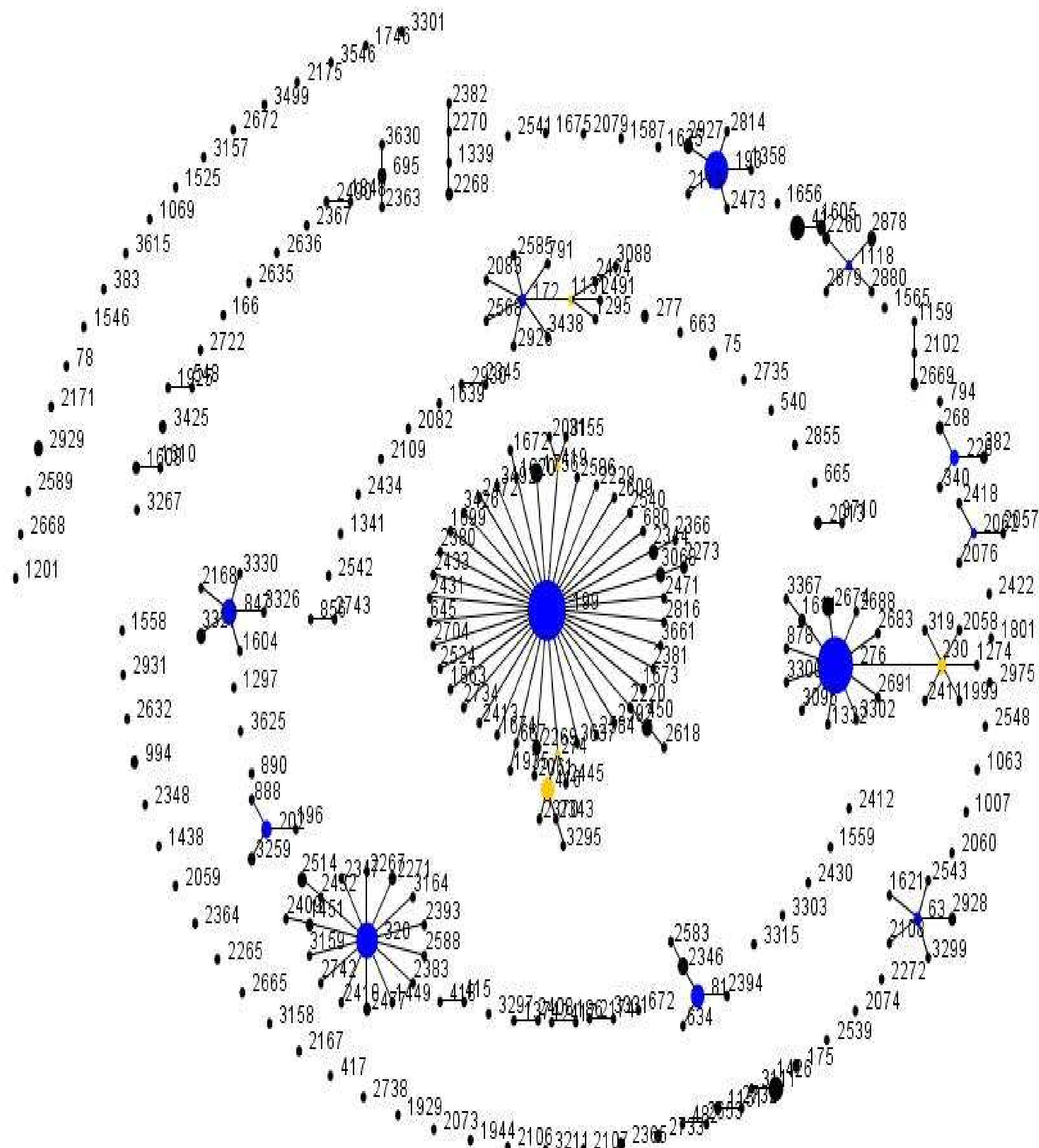


Figure 1-10 Relationships of MLST sequence types constructed using eBURST version 3[5] of all isolates in the MLST database which express serotype 19A capsule.

Blue dots indicate STs which are founders of clones from which single locus variants are demonstrated in black. Yellow dots indicate STs which are subgroup founders. ST199 is the dominant founder clone of serotype 19A. There is substantially greater diversity within STs which express serotype 19A capsule than for serotype 1 or 3.

[5] http://spneumoniae.mlst.net/eburst {accessed 20th December 2008}

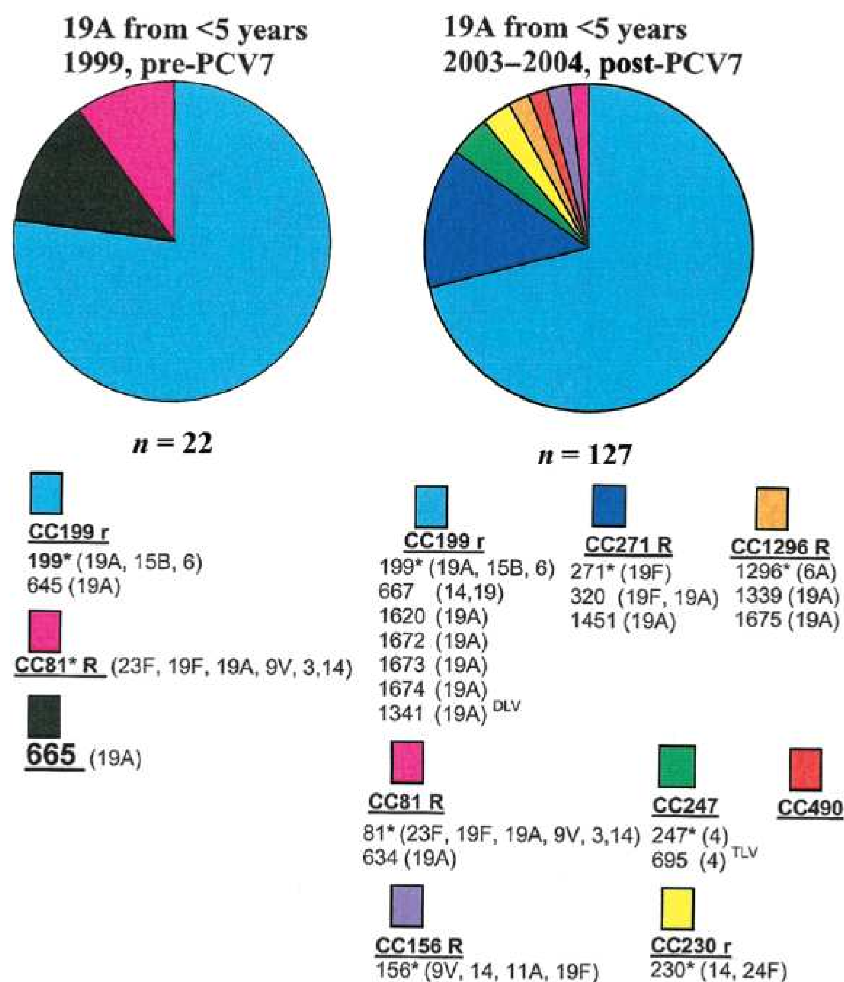


Figure 1-11 Pie charts demonstrating the diversity of MLST sequence types (clonal complex or CC) which express serotype 19A capsule before and after pneumococcal conjugate vaccine introduction.

Figure taken from Pai *et al*, 2005. This increase in diversity post conjugate vaccine introduction has also been seen in other expanding non vaccine serotypes such as 3 and 7F (Byington, 2007). R indicates resistance to multiple antimicrobials and r indicates intermediate resistance to penicillin using Clinical and Laboratory Standards Institute (CLSI) criteria.

Serotype 19A ST199 isolates are associated with pneumococcal haemolytic uraemic syndrome which is also increasing in incidence in the United Kingdom (Waters *et al.*, 2007).

1.6.1.4 Genomic Diversity and Virulence Associated Genes

1.6.1.4.1 Antibiotic Resistance Genes

1.6.1.4.1.1 Penicillin binding proteins (PBP)

The discovery of several allelic variants for pbp2X, pbp2B and pbp1A of the six penicillin binding proteins was pivotal in understanding the phenomenon of horizontal gene transfer

between pneumococcal isolates and between the pneumococcus and other streptococcal species (Hakenbeck *et al*., 1999). It has also been noted that when capsular switching occurs, that penicillin resistance genes may also be transferred at the same time (Brueggemann *et al*., 2007). Diversity in the pbp2B gene can also influence diversity in the *ddl* housekeeping gene during interspecies recombinational exchanges (Enright and Spratt, 1999a).

1.6.1.4.1.2 Macrolides

Resistance to macrolide antibiotics such as clarithromycin occurs mainly by two mechanisms. The first is methylation by a 23s-rRNA dimethyltransferase which prevents macrolides from binding to the ribosome (Retsema and Fu, 2001), resulting in resistance to macrolides, lincosamides and streptogramins (the MLS_B phenotype) (Jain and Danziger, 2004). Several erythromycin ribosome methylase genes (*erm* genes) produce the MLS_B phenotype (Jain and Danziger, 2004). The *ermB* gene predominates in pneumococci (Jain and Danziger, 2004) and is carried by transposons, allowing transmission of this gene to occur horizontally and clonally (Okitsu *et al*., 2005). The MLS_B phenotype results in high levels of resistance to macrolides.

The second mechanism is due to antibiotic efflux which can occur via two classes of pump – the ATP-binding-cassette (ABC) transporter superfamily and the major facilitator superfamily (MFS). The macrolide efflux pump *(mefA)* predominates in pneumococci and results in low to moderate resistance (the M phenotype with Minimum Inhibitory Concentrations of 1-64 mg/L) (Jain and Danziger, 2004, Hyde *et al*., 2001). The presence of the *mefA* gene in R6 and its absence in TIGR4 features in one region of diversity which is apparent when these sequenced genomes are compared (Bruckner *et al*., 2004).

A third unique mechanism which developed rapidly in a patient with serotype 3 pneumococcal pneumonia and which contributed to a fatal outcome, resulted from an insertion of an 18bp tandem repeat in the L22 ribosomal protein and was described by Musher *et al* (Musher *et al*., 2002).

Macrolide resistance predominantly affects serogroups 6, 9, 14, 19 and 23 (Jacobs, 2002). In serotype 14, macrolide resistance is particularly associated with ST9 in Scotland (Clarke *et al*., 2004b).

1.6.1.4.1.3 Tetracyclines

Six diverse alleles for the *tet(M)* gene which confers tetracycline resistance in the pneumococcus have been identified (Doherty *et al*., 2000).

1.6.1.4.1.4 Trimethoprim

Two main groups of mutations in the dihydrofolate reductase gene have been identified in pneumococcal isolates (Adrian, 1997). One of these mutations (Ile100-Leu) confers a 50 fold increase in the 50% inhibitory dose (ID_{50}) of trimethoprim (Adrian, 1997).

1.6.1.4.2 Capsular genes

Diversity of capsular genes and their expression is discussed earlier in section 1.6.1.2.2.

1.6.1.4.3 Surface Proteins

1.6.1.4.3.1 Neuraminidase

The *nanA* gene is thought to be universally present in all pneumococcal strains (Pettigrew *et al*., 2006) while *nanB* was present in 96% of a series of isolates and *nanC* was present in only 51% of the same series (Pettigrew *et al*., 2006). *NanA* itself has substantial sequence diversity with three large regions of diversity – mosaic blocks A, C and D – and regions of insertions. Indeed, when comparing the R6 and TIGR4 genomes, it was noticed that the *nanA* gene in TIGR4 was smaller than that in R6 (Bruckner *et al*., 2004). Point mutations and frameshift mutations have also been recognised in the *nanA* sequence (Calum Johnston, University of Glasgow, personal communication).

1.6.1.4.3.2 Choline Binding Proteins

1.6.1.4.3.2.1 Pneumococcal Surface Protein A (*PspA*)

PspA and *PspC* are choline binding proteins associated with virulence. There is particular interest in *PspA* as it is immunogenic and has potential use for a protein based vaccine (Mollerach *et al*., 2004, Heeg *et al*., 2007). The *PspA* genes are mosaic genes (Hollingshead *et al*., 2000) and *PspA* sequences can be classified into three main families by the variability in their alpha helices (Roche *et al*., 2003) although over 95% of strains belong to family 1 or family 2 (Roche *et al*., 2003, Coral *et al*., 2001, Mollerach *et al*., 2004). Families 1 and 2 of *PspA* are over 50% divergent by sequence analysis (Ren *et al*., 2003). Each family can be further subdivided into six clades (Hollingshead *et al*., 2000, Heeg *et al*., 2007).

1.6.1.4.3.2.2 Pneumococcal Surface Protein C (*PspC*)

PspC is encoded by a heterogeneous group of mosaic genes (Brooks-Walter *et al*., 1999). It has a virulence role in adherence, invasion and evasion of complement (Kerr *et al*., 2006) but as it is highly polymorphic (Iannelli *et al*., 2002), there is substantial strain to strain variation and this influences virulence (Kerr *et al*., 2006). Such sequence variation is noticeable when performing microarray DNA CGH experiments (Hakenbeck *et al*., 2001).

1.6.1.4.3.2.3 Other Choline Binding Proteins (cbps)

In their comparison of the R6 and TIGR4 genomes, Bruckner *et al* identified differences between the two for several other cbps (Bruckner *et al*., 2004). One of these, cpbG, has been noted to exist in a truncated form due to a TTTA repeat causing a frameshift and premature stop codon (Mann *et al*., 2006) although different clinical presentations can be associated with either form (Mann *et al*., 2006).

1.6.1.4.4 Pneumolysin

Genetic diversity in the structure of pneumolysin has recently been shown to be associated with different biological behaviour of the toxin (Jefferies *et al*., 2007, Kirkham *et al*., 2006) indicating that genetic differences in the bacterial virulence genes could account for different disease manifestations. Sequence diversity in the pneumolysin gene has been recognised since 2005 (Hanage *et al*., 2005).

1.6.1.4.5 Zinc Metalloproteinases

The pneumococcus exhibits large proteases on its surface which are predominantly zinc metalloproteinases and early sequenced genomes showed there to be up to four per isolate (Chiavolini *et al*., 2003). These are particularly important for invasive infections and virulence in serotype 4 pneumococci, none appear important for virulence in serotype 3 pneumococci and only two are required for virulence in serotype 19F (Chiavolini *et al*., 2003). Two of these have been characterised and show substantial genomic diversity – IgA1 protease and ZmpC.

1.6.1.4.5.1 Immunoglobulin A1 protease

This class of proteases allow the pneumococcus to evade host mucosal immunoglobulin which explains its substantial diversity as it is required to cleave structurally diverse substrates (Poulsen *et al*., 1998). Pneumococcal strains demonstrate variation in the number and sequence of repeat regions to facilitate this (Poulsen *et al*., 1998). There are also substantial sequence similarities between the pneumococcus and other species of

alpha haemolytic streptococci (Poulsen *et al*., 1998). This high level of diversity can be seen in results of microarray DNA CGH experiments where failure to hybridize indicates substantial divergence from the probe on the microarray (Hakenbeck *et al*., 2001).

1.6.1.4.5.2 Zinc metalloproteinase C (ZmpC)

ZmpC specifically cleaves human matrix metalloproteinase 9 (MMP-9) (Oggioni *et al*., 2003). It is absent from the R6 genome (Hoskins *et al*., 2001, Chiavolini *et al*., 2003). It has been determined to be present in 26% of isolates and has an association with pneumococcal isolates from patients with pneumonia (Chiavolini *et al*., 2003) and was not found in isolates from nasal or conjunctival swab isolates (Oggioni *et al*., 2003). It has therefore been suggested that it has a role in virulence and pathogenicity in the lung (Oggioni *et al*., 2003).

1.6.1.4.6 Two-component signal transduction systems (TCSs)

Two component systems allow pneumococci to respond to changes in their environment (Paterson *et al*., 2006). The two components are two proteins – a membrane associated sensor histidine kinase which on receipt of a specific stimulus phosphorylates an aspartate residue in a cytoplasmic cognate response regulator which invariably results in alteration of levels of gene transcription (Paterson *et al*., 2006). There are thirteen TCSs and one orphan response regulator described in the pneumococcus (Paterson *et al*., 2006, Standish *et al*., 2007). Allelic variation in these TCSs is recognised (Reichmann and Hakenbeck, 2000). All thirteen TCSs have been sequenced and several are homologous with genes of other Gram positive bacteria (Lange *et al*., 1999). For at least one of the TCSs, regulation of gene expression is strain dependent (Hendriksen *et al*., 2007, Blue and Mitchell, 2003).

1.6.1.4.7 Pneumococcal pilus

A pneumococcal pilus is encoded by the *rlrA* islet (LeMieux *et al*., 2006), which includes genes for three pilus subunits (RrgA, RrgB and RrgC) and can be found in some but not all pneumococcal isolates (Paterson and Mitchell, 2006) where it has a role in adherence and may be involved in virulence (Barocchi *et al*., 2006, Gianfaldoni *et al*., 2007). Its presence is associated with serotypes 4, 6B, 9V and 14 while it has been documented as absent from serotypes 1, 7F, 8 and 12B and its presence appears to be a clonal property (Aguiar *et al*., 2008, Moschioni *et al*., 2008, Basset *et al*., 2007) where it is particularly associated with ST156 and ST162 (Sjostrom *et al*., 2007). Sequencing of the *rlrA* islets has identified three clade types with homology of 88-92% (Moschioni *et al*., 2008). Diversity of sequence in the gene SP0466 within the *rlrA* islet resulting in false negative microarray CGH results

has been recently demonstrated (Sjostrom *et al*., 2007). A second pilus has also recently been described (Bagnoli *et al*., 2008).

1.6.1.5 Why Sequence the Pneumococcal Genome?

There are several practical applications resulting from investigating the diversity of pneumococcal genomes and their constituent genes, other than to merely catalogue species diversity. This can identify genes which may distinguish oral streptococci and so improve taxonomy (Hollingshead and Briles, 2001, Lan and Reeves, 2001). Additionally, highly variable genes may distinguish poor vaccine candidate genes from more conserved and useful ones (Hollingshead and Briles, 2001, Serruto *et al*., 2004) such as *lytA* (Whatmore and Dowson, 1999). The availability of genome sequence data facilitates the development of new technologies (Hinds *et al*., 2002b) which can be used to develop new diagnostic tests (Hinds *et al*., 2008) or improve antimicrobial drug discovery through the development of genomic expression profiling (Hollingshead and Briles, 2001) or the identification of novel virulence associated genes (Hollingshead and Briles, 2001, Fournier *et al*., 2007).

1.6.1.6 Aims of the Work Described in this Thesis

The aims of this work were to investigate:

1. The genomic diversity of defined populations of clinically relevant *Streptococcus pneumoniae* by means of a comparative genomic hybridization approach using a validated contemporary PCR product microarray, PCR and genomic sequencing.

2. Factors which may be causing genomic diversity such as heterogeneity of genes, gene deletions or insertion of genetic material.

3. The influence that genomic diversity may have on phenotype.

This was approached by:

- Comparison of microarray CGH results with genome sequence data for 4 fully sequenced pneumococcal isolates.

- Illustrating, using ten pneumococcal isolates of different serotypes and unrelated sequence types, where regions of the pneumococcal genome were variable.

- Illustrating differences in the extent of inter-serotype and intra-serotype genomic diversity by comparing five isolates of ST9 from different serotypes with five isolates of ST9 all of which were serotype 14.

- Illustrating the extent of diversity within two clones of identical serotype and sequence type (serotype 3, ST180 and serotype 4, ST246.)

- Investigating the degree of diversity which could be demonstrated in isolates retrieved from nosocomial outbreaks where isolates of the same serotype and sequence type (serotype 4, ST206 and serotype 1 ST227) were known to be linked epidemiologically.

- Documenting sequence type diversity involved in pneumococcal carriage in a paediatric population from a geographical region of high biodiversity.

The influence of genomic diversity on phenotype through the demonstration of microarray transcriptional profiles was investigated by:

- Comparison of IPD related serotype 3, ST180 isolates with serotype 3, ST180 isolates isolated from nasopharyngeal carriage to investigate whether there could be a genetic basis for the difference in "invasive" and "carriage" phenotype.

- Identification of transcriptional profiles from different clinical disease conditions:

    1. serotype 1 (ST227 and ST306) isolates associated with parapneumonic complications of pneumonia.

    2. serotype 3 (ST180) isolates from cases of cerebral abscess and meningitis.

- Demonstration of the transcriptional effects on a multiply antibiotic resistant pneumococcal isolate by a sub-therapeutic dose of clarithromycin.

2 Materials and Methods

2.1 General Procedures

2.1.1.1 Bacterial Strains and Growth Conditions

Details of the strains used in microarray experiments are displayed in Appendix 1. These details include serotypes, Multilocus Sequence Types, antibiotic sensitivities, details of the body fluid from which isolates were identified and the age of the source patients. Isolates were obtained from the strain collections of the Scottish Meningococcal and Pneumococcal Reference Laboratory (SMPRL) and the University of Glasgow Pneumococcal Research Group. All strains were taken from frozen stocks and grown overnight at 37°C on 5% horse blood agar (Oxoid, United Kingdom) in preparation for the manufacture of fresh glycerol stocks.

2.1.1.2 Glycerol Stock Manufacture and Culture Checks

Single colonies of test isolates grown overnight on 5% horse blood agar were inoculated with a sterile plastic loop into 10ml of Brain Heart Infusion broth (Oxoid, United Kingdom) and grown at 37°C in a water bath to an optical density of 0.6 at 600nm determined on a WPA biowave C08000 Cell Density Meter (WPA, United Kingdom). 1.2 ml of sterile glycerol (Riedel-de-Haen®, Germany) was added to the remaining 9ml of culture and mixed. 1ml aliquots were then pipetted into cryotubes (Sarstedt, Germany) and stored at -80°C. The purity of glycerol stocks was assessed by growth of 10µl plated onto 5% horse blood agar and incubated overnight at 37°C.

2.1.1.3 Antibiotic Susceptibility Testing

Antibiotic susceptibility were determined using disc susceptibility testing applying the Clinical and Laboratory Standards Institute (CLSI) methodology. Susceptibility to oxacillin, ampicillin, ciprofloxacin, clarithromycin and clindamycin were determined and these are documented in Appendix 1.

Isolates which were collected as part of the paediatric pneumococcal carriage study in Bolivia (Chapter 9) and isolate South Africa 2507 (Chapter 12) had antibiotic Minimum Inhibitory Concentrations (MICs) determined using E-tests® (AB Biodisk, Sweden). MICs were determined for penicillin, erythromycin, vancomycin, chloramphenicol, tetracycline and co-trimoxazole and documented in Appendix 2.

2.1.1.4 Serotyping

All serotyping of strains was performed at SMPRL using a coagglutination method (Smart, 1986) utilising sera from Statens Serum Institut, Denmark.

2.1.1.5 Nucleic Acid Extraction

2.1.1.5.1 DNA Extraction for Microarray Analysis and Polymerase Chain Reaction

From a pure frozen glycerol stock of the test pneumococcal isolate, 100µl was inoculated into 20mL Brain Heart Infusion and incubated at 37°C overnight. Purity of this broth culture was assessed by streaking 10µl aseptically onto 5% horse blood agar (Oxoid, United Kingdom) and incubating this overnight at 37°C.

The remaining culture was centrifuged at 4000 revolutions per minute (rpm) for 15 minutes (serotype 3 pneumococcal cultures required 30 minutes) at 4°C in a Sigma Laboratory centrifuge 4K15 (Philip Harris, United Kingdom). After decanting the supernatant, the pellet was resuspended in 1ml lysis buffer (consisting of 10µl 1M Tris pH 8.0 (Ambion/Applied Biosystems, United Kingdom); 200µl 0.5M EDTA (Ambion/Applied Biosystems, United Kingdom); 50µl 10% Sodium Dodecyl Sulphate (SDS) (Ambion/Applied Biosystems, United Kingdom); 740µl double distilled water). The resulting solution was transferred to a sterile 1.5ml microcentrifuge tube (Greiner Bio One, Germany) and incubated at 37°C for 1 hour in a digital dry bath (Accublock™ Digital Dry Bath, Labnet International, Inc., United Kingdom). Proteinase K (Invitrogen, United Kingdom) was added to a final concentration of 100µg/ml (5µg per ml from a 20mg/ml frozen stock). The resulting solution was incubated in the digital dry bath at 50°C for 3 hours.

RNaseA (Sigma-Aldrich, United Kingdom) was added to a final concentration of 20µg/ml (i.e. 2µl per ml of a 10mg/ml stock which had been boiled to remove DNase activity) and incubated at 37°C for 30 minutes. An equal volume of phenol: chloroform: isoamylalcohol (25:24:1) (Sigma-Aldrich, United Kingdom) was added and mixed by inverting the tubes sharply several times then centrifuged at 13000 rpm for 3 minutes at room temperature (Eppendorf centrifuge 5417C, USA). The upper phase was removed without disturbing the lower phase and added to a fresh microcentrifuge tube. 0.2 volumes of 10M ammonium acetate (Fisher Scientific, United Kingdom) were added along with 600µl absolute ethanol (Fisher Scientific, United Kingdom). The tubes were then inverted gently then spun at 30 minutes 13000 rpm to pellet the DNA.

The supernatant was carefully decanted and the remaining DNA pellet was air dried for 15-20 minutes upside down on a paper towel then re-suspended in TE buffer (200 µl). Once re-suspended, the DNA was stored at -20°C. The DNA was quantified on a Nanodrop ND-1000 spectrophotometer (Agilent Technologies, United Kingdom).

2.1.1.5.2 RNA Extraction for Microarray Analysis

From a pure frozen glycerol stock of the test pneumococcal isolate, 100µl was inoculated into 15mL Brain Heart Infusion and incubated at 37°C until an optical density at 600nm of 0.6 was reached. Purity of this broth culture was assessed by streaking 10µl aseptically onto 5% horse blood agar and incubating this overnight at 37°C.

10ml of broth culture was centrifuged at 5000 rpm at room temperature for 5 minutes in 15 ml centrifuge tubes after which the supernatant was discarded and the pellet frozen immediately in liquid nitrogen. Frozen pellets were then stored at -80°C until RNA extractions for all the hybridizations in a particular experiment could be performed in parallel.

Due to the possible action of RNases, it was considered essential to include a step early in the cell harvest and RNA extraction protocol which inactivates RNases (Conway and Schoolnik, 2003), thereby limiting bias in the transcript representation due to differential turnover rates of RNA. On comparing a liquid nitrogen freezing step with the addition of RNAprotect Bacteria Reagent (Qiagen, United Kingdom) freezing in liquid nitrogen was not only more cost effective but yielded higher concentrations of RNA which were of better quality.

Fresh lysozyme TE buffer was made. For 1ml of lysozyme TE buffer, 10µl 1M Tris HCl pH8.0 (Ambion/Applied Biosystems, United Kingdom), 2µl of 0.5M EDTA pH8.0 (Ambion/Applied Biosystems, United Kingdom) and 15mg of lysozyme (Sigma-Aldrich, United Kingdom) were added to 1ml of nuclease free water (Ambion/Applied Biosystems, United Kingdom).

To begin the extraction, 200µl of lysozyme TE Buffer (15 mg/ml) was added to the pellet, vortexed for 10 seconds using a rotamixer (Hook and Tucker Instruments, United Kingdom) and incubated at room temperature for 15 minutes with vortexing for 10 seconds every 2 minutes. 700µl of RLT Buffer from a Qiagen RNeasy Mini Kit (Qiagen RNeasy MINI KIT, Qiagen, United Kingdom) was added and vortexed for 10 seconds.

The resulting lysate was transferred to a sterile 1.5ml microcentrifuge tube (Greiner Bio One, Germany) containing 25-50mg of 100µm glass beads (Sigma-Aldrich, United Kingdom). Using a Hybaid Ribolyser (Hybaid, United Kingdom) set at speed 4, three runs of 20 seconds were used to facilitate the disruption of cells. The solution was then centrifuged at 13000 rpm (Eppendorf centrifuge 5417C, USA) for 10 seconds. The supernatant (approximately 900µl) was recovered to a new 1.5ml microcentrifuge tube (Greiner Bio One, Germany). 500µl of Ethanol 100% (Fisher Scientific, United Kingdom) was added and mixed without vortexing.

700µl of this solution was added to the RNeasy Mini column (Qiagen- RNeasy MINI KIT, Qiagen, United Kingdom) and centrifuged at 13000 rpm for 30 seconds at room temperature (Eppendorf centrifuge 5417C, USA). The flow-through was discarded and the remaining 700µl added to the same RNeasy Mini column and centrifuged at 13000 rpm for 30 seconds. 350µl of RW1 Buffer was added into the RNeasy Mini column (Qiagen- RNeasy MINI KIT, Qiagen, United Kingdom) and centrifuged at 13000 rpm for 5 minutes at room temperature.

10µl DNase I stock solution was added to 70µl Buffer RDD (Qiagen RNase-Free DNase Set, Qiagen, United Kingdom) and this 80µl was pipetted directly onto the RNeasy Mini column silica-gel membrane, and placed on the benchtop at room temperature for 15 minutes.

350µl of Buffer RW1 was pipetted into the RNeasy column then centrifuged for 30 seconds at 13000 rpm at room temperature and the flow through was discarded. A further 700µl of RW1 Buffer was added into the RNeasy Mini column and centrifuged at 13000 rpm for a further 30 seconds. The flow-through was again discarded and 500µl of RPE Buffer added into the RNeasy Mini column and centrifuged at 13000 rpm for 30 seconds. After discarding the flow through, an additional 500µl of RPE Buffer was added to the RNeasy Mini column and centrifuged at 13000 rpm for 2 minutes to dry the silica-gel membrane.

To elute, the RNeasy Mini column was transferred to a new 1.5 ml collection tube. 50µl of nuclease free water (Ambion/Applied Biosystems, United Kingdom), was pipetted directly onto the silica-gel membrane (30µl was added for serotype 3 pneumococcal RNA as the RNA yield was generally much lower than other serotypes). The tube was left to stand for 3 minutes and then centrifuged for 1 minute at 13000 rpm. A 5µl aliquot was removed for quality assessment on an Agilent 2100 bioanalyser (Agilent Technologies, United

Kingdom) where an RNA integrity number above 9 was considered high enough quality for further use and quantification on a Nanodrop ND-1000 spectrophotometer (Agilent Technologies, United Kingdom) while the remaining stock of RNA was stored at -80°C until required.

2.1.1.5.3 RNA Extraction for Quantitative Real Time Polymerase Chain Reaction (qRT-PCR)

In addition to the steps above, a second treatment to remove DNA was performed by adding 1µl Ambion TURBO DNA-*free*™ (Ambion/Applied Biosystems, United Kingdom) after elution into nuclease free water (Ambion/Applied Biosystems, United Kingdom) prior to assessing the RNA quality.

2.1.1.6 Polymerase Chain Reaction (PCR)

PCR of genes to confirm their presence or absence in test genomes was performed on a Techegene thermal cycler (Bibby Scientific, United Kingdom). This was set to 2 minutes of denaturation at 94°C followed by 35 cycles of 94°C for 30 seconds, X°C for 30 seconds (where X was 2°C lower than the primer melting temperature) and 40 seconds at 72°C. The final extension was set for 2 minutes at 72°C and then the reaction was held at 4°C. The reaction volume was 25µl consisting of 0.15µl Go Taq® DNA polymerase (Promega, USA), 5µl of 5xBuffer (Promega, USA), 2µl of Magnesium Chloride (Promega, USA), 0.5µl of 10mM dNTPs (Invitrogen, United Kingdom), 0.5µl forward primer, 0.5µl reverse primer, 1µl genomic DNA and 15.35µl PCR grade water.

2.1.1.7 Gel Electrophoresis

Prior to use in microarray experiments, the quality of genomic DNA was also assessed by gel electrophoresis using a 0.7% agarose gel (Agarose MP, Roche Diagnostics, Germany).

PCR products were run on a 2% agarose gel to assess the product size using a 100 bp DNA ladder (Promega, USA). SYBR® safe (Invitrogen, United Kingdom) was used to demonstrate bands of DNA.

2.2 Microarray Protocols

2.2.1.1 DNA Comparative Genomic Hybridization (CGH)

DNA CGH experiments were all performed using a common reference design with TIGR4 as the reference pneumococcal genome. Each experiment included fluorochome labelling

of DNA with a dye swap step so that for each test isolate of pneumococcal DNA (biological replicate), two microarrays were hybridized (technical replicates) – one with the test DNA labelled with fluorochrome Cy3 (GE Healthcare, United Kingdom) and one labelled with Cy5 (GE Healthcare, United Kingdom) (Churchill, 2002) as shown in Figure 2-1.

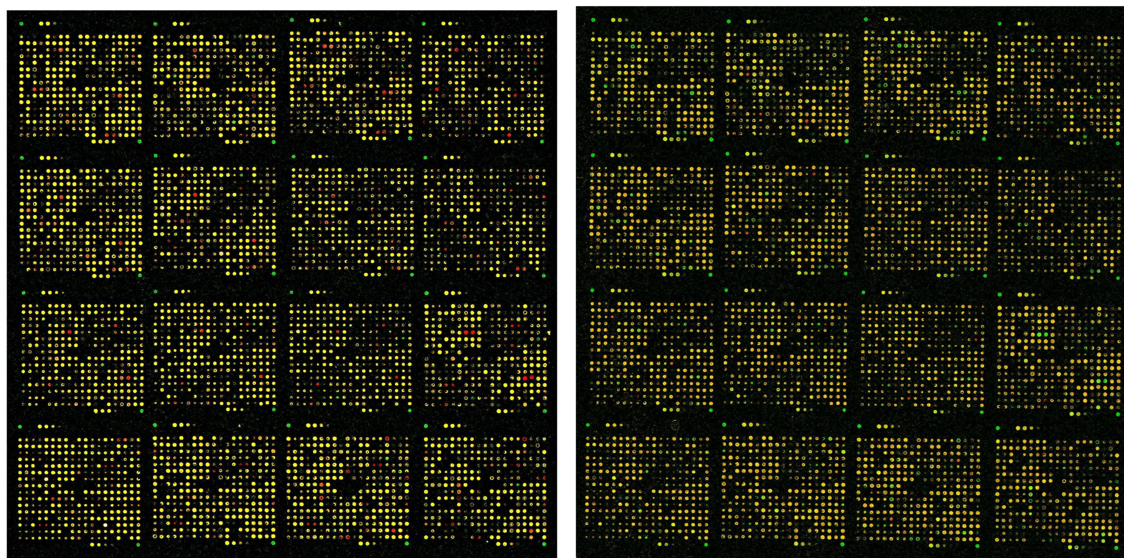


Figure 2-1 Results of microarray CGH dye swap experiments for isolate 06-1805 (ST227).

In both microarrays illustrated above, DNA from the test clinical isolate 06-1805 (used in Chapter 10) is competitively hybridized with DNA from the laboratory reference strain, TIGR4. On the left, the TIGR 4 genes are labelled with the fluorophore Cy3 and 06-1805 genes are labelled with the fluorophore Cy5. This labelling is reversed on the microarray on the right. When genes are present in TIGR4 but not the test clinical isolate they appear as red dots (left) or green dots (right). When genes are present in 06-1805 but not the TIGR4 genome they appear as green dots (left) and red dots (right) – these genes are identifiable in the laboratory reference strain R6 but are absent from TIGR4. Where genes are present in both 06-1805 and TIGR4 there is competition to hybridize on the array and the dots appear yellow.

For each microarray, one Cy3 labelled DNA sample and one Cy5 labelled DNA sample were prepared by heating (for 5 minutes at 95°C) 5µg of DNA and 1µl of random primers (Invitrogen, United Kingdom) made up to 41.5µl with nuclease free water (Ambion/ Applied Biosciences, United Kingdom). This was then snap cooled on ice and briefly centrifuged. To each was added 5µl 10xREact 2 buffer (Invitrogen, United Kingdom), 1µl dNTPs (5mM dATP, 5mM dGTP, 5mM dTTP and 2mM dCTP), 1.5µl Cy3 or Cy5 dCTP (GE Healthcare, United Kingdom) and 1µl Large fragment DNA Polymerase I (Klenow) (3-9U/µl) (Invitrogen, United Kingdom) and the solution was incubated at 37°C in the dark for 90 minutes in a Techegene thermal cycler (Bibby Scientific, United Kingdom).

Microarrays were prepared by soaking in a prehybridization solution preheated to 65°C (composed of 8.75ml 20xSSC (Ambion/Applied Biosciences, United Kingdom), 250µl

20% SDS (Ambion/Applied Biosciences, United Kingdom), 5ml (100mg/ml) Bovine Serum Albumin (Sigma-Aldrich, United Kingdom) and made up to 50ml of sterile double distilled water) for 20 minutes in a Coplin jar (Fisher Scientific, United Kingdom) and placed in a Techne Hybridizer HB-1D (Bibby Scientific, United Kingdom). Prehybridized arrays were rinsed in 400ml double distilled water for 1 minute and then in 400ml propan-2-ol (VWR International, USA) for 1 minute. Each array was placed in a 50ml centrifuge tube and centrifuged at 1500 rpm for 5 minutes then stored in a dust free box until ready for hybridization.

Cy3 and Cy5 labelled DNA samples were combined into a Qiagen Minelute purification column (Qiagen, United Kingdom). 500µl of Buffer PB was added and centrifuged at 13000 rpm for 1 minute (Eppendorf centrifuge 5417C, USA). The flow through was discarded and 500µl of Buffer PE was added and centrifuged at 13000 rpm for 1 minute. Again the flowthrough was discarded then 250µl of Buffer PE was added to the same column and centrifuged at 13000 rpm for a minute. Flow through was discarded and then the column was spun again for 1 minute at 13000 rpm then placed in a fresh 1.5ml collection tube. 15.9µl of nuclease free water (Ambion/Applied Biosciences, United Kingdom) was added to the membrane and centrifuged for 1 minute at 13000 rpm to elute DNA for hybridization.

14.9µl of the Cy3/Cy5 labelled DNA sample was added to 4.6µl of filtered 20xSSC and 3.5µl of 2% SDS and the resulting solution heated for 2 minutes at 95°C in a Techegene thermal cycler (Bibby Scientific, United Kingdom). Lifter slips (Erie Scientific Company, USA) were placed on the pre-hybridized microarrays and the DNA solutions pipetted onto the bottom left corner to allow the solution to be drawn across the microarray by capillary action. Prepared microarrays were placed in a hybridization cassette, sealed and then submerged in a water bath at 65°C in a Techne Hybridiser HB-1D (Techne, USA) in the dark for 20 hours.

After hybridization, arrays were transferred to a slide rack and washed with agitation in a pre-heated solution of 20ml 20xSSC, 1ml 20% SDS made up to 400ml with sterile double distilled water for 2 minutes and then transferred to a further solution of 1.2ml 20xSSC made up to 400ml with sterile double distilled water and washed with agitation for 4 minutes. Microarrays were then placed in 50ml centrifuge tubes and dried by centrifugation at 1500 rpm for 5 minutes. The hybridized microarrays were then scanned using ScanArray Express ™ (Packard Biosciences Biochip Technologies, Perkin Elmer).

2.2.1.2 RNA Comparative Hybridization

RNA expression experiments were also performed using a common reference design using TIGR4 RNA grown to midlog phase (Conway and Schoolnik, 2003). Unlike the DNA CGH experiments, 3 biological replicates of each test isolate RNA (i.e. three independent RNA extractions from different broth cultures) were hybridized against TIGR4 RNA (from the same batch of TIGR4 RNA) with only one technical replicate per biological replicate as required for statistical validity (Foster and Huber, 2002). The choice of using an RNA reference control grown to midlog phase was made as this is an established means of determining a baseline for gene expression. It is appreciated that this is not the only possible method and that alternatives such as using genomic DNA or a mixture of reference RNA from several sampling conditions have also been used (Conway and Schoolnik, 2003).

For each microarray, one Cy3 labelled cDNA sample (2-10µg) and one Cy5 labelled cDNA sample (2-10µg) were prepared by heating for 10 minutes at 70°C with 1µl of random primers (Invitrogen, United Kingdom) and made up to 11µl with nuclease free water (Ambion/ Applied Biosciences, United Kingdom). This was then snap cooled on ice and briefly centrifuged. To each was added 5µl 5xFirst strand buffer (Invitrogen, United Kingdom), 2.5µl DTT (100mM), 2.3µl dNTPs (5mM dATP, 5mM dGTP, 5mM dTTP and 2mM dCTP), 1.7µl Cy3 or Cy5 dCTP (GE Healthcare, United Kingdom) and 2.5µl SuperScript II (200U/µl) (Invitrogen, United Kingdom) and the solution was incubated at 25°C in the dark for 10 minutes then 42°C for 90 minutes in a Techegene thermal cycler (Bibby Scientific, United Kingdom).

Microarrays were prepared in an identical way to that described above in Section 2.2.1.1. Cy3 and Cy5 labelled DNA samples were combined into a Qiagen Minelute purification column (Qiagen, United Kingdom). 250µl of Buffer PB was initially added and centrifuged at 13000 rpm for 1 minute (Eppendorf centrifuge 5417C, USA). The rest of the procedure for preparing the hybridization solution, hybridizing the microarray and scanning is identical to that described in Section 2.2.1.1.

2.2.1.3 Microarray Normalization for DNA Comparative Genomic Hybridization

Tagged Image File Format (TIFF) images of the scanned microarrays created by ScanArray Express ™ (Packard Biosciences Biochip Technologies, Perkin Elmer) were

entered into Bluefuse for Microarrays 3.5 © with the Cy3 labelled image in Channel 1 and the Cy5 labelled image in Channel 2. The array gridmap files provided by the Bacterial Microarray Group at St George's Hospital, University of London (BμG@S, United Kingdom) were utilised. The post processing protocol was devised by Dr Jason Hinds of BμG@S and comprised of initial exclusion of unreliable results due to poor quality hybridizations with a confidence estimate of less than 0.1. Controls spots on the array were identified using an array gridmap GenePix Array List (GAL) file (SPv1_1_0_CGH_Gridmap.bcf) and data pertaining to control spot hybridizations was automatically removed from the analysis. To correct for spatial, intensity and dye related effects, normalization was performed using the option, "Global Lowess excluding all with text." Confidence flags were set at their default settings. Replicates of each dye swap were combined by fusion.

CGH was performed in Bluefuse for Microarrays 3.5© using a protocol devised by Dr Jason Hinds through the identification of a normal distribution of experimental variability and by identifying variability which was two standard deviations from the mean of this normal distribution for all the results for the TIGR4 and R6 genes represented on the microarray. Automated classification of regions of variability was performed by setting a ratio threshold for amplification as 1.0 and ratio threshold for deletion at -1.0 with the minimum number of clones included in the region in order for it to be classified as an amplification or a deletion set at 1. Dye swap processing was enabled.

Data analysis was completed in Genespring GX 7.3.1 (Agilent Technologies, USA) again using protocols devised by Dr Jason Hinds. Output_fused_CGH files were imported into Genespring GX 7.3.1 and further normalization was performed after data transformation to account for dye swaps. This normalization was performed using the, "Per spot and divided by control channel," protocol with a cross gene error model using the error model for one-colour data. The error model was based on deviation from 1. The generation of gene lists using Genespring GX 7.3.1 (Agilent Technologies, USA) was accomplished by importing the Bluefuse for Microarrays 3.5© generated output_fused.xls files to create an experiment whereby the microarray dye swaps for each strain could be analysed using the, "Filter on data file," option. To generate each gene list, a search was performed using the, "Type," column employing the search criteria, "Column values must be not equal to NO CHANGE," and "Value must appear in at least 1 of the selected columns." The resulting gene list could be saved or exported into Microsoft Office Excel 2003, Microsoft®, United Kingdom for comparison with further strains.

2.2.1.4 Microarray Normalization for RNA Expression

For RNA expression experiments, normalization was performed (again using protocols written by Dr Jason Hinds) by importing the Output_fused.xls files into Genespring GX 7.3.1 (Agilent Technologies, USA) for the 3 biological replicates of each isolate. When TIGR4 cDNA was labelled with Cy3 an initial "dye swap" normalization step was used for RNA expression experiments. An additional, "Per gene," normalization step was applied to specific samples (serotype 1 isolates) where the isolates being compared involved more than one MLST of *S. pneumoniae* to take into account clustering by MLST rather than by the clinical condition being investigated.

Statistical analysis of RNA expression data generated by Genespring GX 7.3.1 (Agilent Technologies, USA) was performed using the statistical analysis (ANOVA) tool. This performs a 1-way parametric test without assuming variances are equal. The false discovery rate was set at 0.05 resulting in a false discovery rate of about 5% of genes. Multiple testing correction was performed using a Benjamini and Hochberg False Discovery Rate. No *post hoc* tests were used.

The class prediction function on Genespring GX 7.3.1 (Agilent Technologies, USA) was used to generate lists of genes which were predictive of invasiveness, brain abscess or complicated pneumonia respectively, with a predictive strength calculation based on Fishers Exact Test. These lists were then imported into Microsoft Office Excel 2003, Microsoft®, United Kingdom for comparison.

2.3 Gene Sequencing and Multi Locus Sequence Typing

Gene sequencing and MLST of isolates (Enright and Spratt, 1998) was performed at the SMPRL using their protocols as outlined below.

2.3.1 Gene Sequencing

PCR products from reactions performed using primers noted in Appendix 3 (MWG Biotech AG, Germany) which were manufactured to generate the PCR products utilised in the manufacture of the microarray were used for sequencing. A semi automated PCR clean up (Clarke and Diggle, 2002) was performed using a RoboAmp liquid handling robot (MWG Biotech AG, Germany). Likewise, a sequence reaction using the RoboAmp thermocycler (MWG Biotech AG, Germany) with a semi automated sequence cleanup was

performed as per the published method (Clarke and Diggle, 2002). Sequencing was performed using a MegaBACE™ 1000 96-capillary sequencer (GE Healthcare, United Kingdom) and a published protocol (Jefferies *et al*., 2003).

2.3.2 Multi Locus Sequence Typing

MLST was performed using a previously published semi-automated protocol (Jefferies *et al*., 2003). The primers for the seven housekeeping genes used for MLST are listed below.

Housekeeping Gene	Primer Sequence (5'→3')
aroE forward	GCC TTT GAG GCG ACA GC
aroE reverse	TGC AGT TCA (G/A)AA ACA T(A/T)T TCT AA
gdh forward	ATG GAG AAA CCA GC(G/A/T/C) AG(C/T) TT
gdh reverse	GCT TGA GGT CCC AT(G/A) CT(G/A/T/C) CC
gki forward	GGC ATT GGA ATG GGA TCA CC
gki reverse	TCT CCC GCA GCT GAC AC
recP forward	GCC AAC TCA GGT CAT CCA GG
recP reverse	TGC AAC CGT AGC ATT GTA AC
spi forward	TTA TTC CTC CTG ATT CTG TC
spi reverse	GTG ATT GGC CAG AAG CGG AA
xpt forward	TTA TTA GAA GAG CGC ATC CT
xpt reverse	AGA TCT GCC TCC TTA AAT AC
ddl forward	TGC (C/T)CA AGT TCC TTA TGT GG
ddl reverse	CAC TGG GT(G/A) AAA CC(A/T) GGC AT

Table 2-1 Primer sequences used for MLST.

2.4 Analysis of Genome Sequences Using the Artemis Comparison Tool

The Artemis Comparison Tool (ACT) was developed as a means of visualising comparisons of complete genome sequences and their associated annotations (Carver *et al*., 2005). The genomes of the two microarray reference strains (TIGR4 and R6) were compared with the annotated sequences for fully sequenced pneumococcal genomes provided as EMBL files by Mr Nicholas Croucher, Wellcome Trust Sanger Institute using

the BLASTN algorithm from the web site DoubleACT which is hosted by the Health Protection Agency[6]. The resulting files were imported into ACT to allow comparison of either the TIGR4 or R6 genomes with other sequenced strains. These sequence comparisons could then be used to validate the microarray comparative genomic hybridizations using the same four test strains when hybridized competitively against TIGR4.

2.5 Quantitative Real Time Polymerase Chain Reaction

2.5.1 cDNA Synthesis

2µg of high quality RNA (RNA integrity number > 9 on Agilent 2100 Bioanalyser (Agilent Technologies, United Kingdom)) was mixed with 2µl Random Hexamers (Invitrogen™, United Kingdom) and 1 µl RNaseOUT™ (Invitrogen™, United Kingdom) and nuclease free water was added (Ambion/Applied Biosystems, United Kingdom) to a total volume of 17.5 µl. This was denatured at 70°C for 10 minutes in a Techegene thermal cycler (Bibby Scientific, United Kingdom), then put on ice. To each sample were added 6µl of 5x First Strand Buffer (Invitrogen™, United Kingdom), 3µl of 0.1M DTT (Invitrogen™, United Kingdom), 1.5µl of 10mM dNTP (Invitrogen™, United Kingdom) and 2µl of SuperScript® III (Invitrogen™, United Kingdom). The reaction was incubated at 42°C for 16 hours and then inactivated by heating to 70°C for 15 minutes. 1µl of *E. coli* RNase H (Invitrogen™, United Kingdom) was added and incubated at 37°C for 20 minutes. The concentration of the resulting cDNA was quantified using a Nanodrop ND-1000 spectrophotometer (Agilent Technologies, United Kingdom).

2.5.2 Quantitative Real Time Polymerase Chain Reaction Protocol

Quantitative Real Time PCR was performed at SMPRL on a LightCycler® 480 (Roche, United Kingdom). Pneumococcal DNA gyrase subunit A (SP1219) was used as the reference gene and its expression compared to appropriate test genes for each experiment. TIGR4 cDNA was used as a positive control and negative controls were nuclease free water and the RNA sample from which cDNA was manufactured (non-reverse transcription negative control). Standard curves were constructed using cDNA at dilutions of 1:5, 1:50, 1:500, 1:5000 and 1:50,000. Primers were used at a concentration of 1µM and

---

[6] http://www.hpa-bioinfotools.org.uk/pise/double_act.html# {accessed 17th November 2008}

the total reaction volume was 20µl (5µl cDNA, 10µl LightCycler® 480 SYBR Green I Master (Roche, United Kingdom), 3µl nuclease free water, 1µl forward primer and 1µl reverse primer). The protocol used consisted of a pre-incubation step for 5 minutes at 95°C and then 40 cycles of an amplification step consisting of 10 seconds at 95°C, 20 seconds at 50°C and 30 seconds at 72°C. The melting curve cycle consisted of 5 seconds at 95°C, 1 minute at 65°C followed by reaction termination at 72°C then cooling to 40°C for 10 seconds. The results of the real time PCR experiments were analysed using the software Q-gene (Muller *et al*., 2002) for assessment of relative expression levels for genes of interest. It is important to recognise though that qRT-PCR is a more quantitative method of comparing gene expression than by using microarrays which are semi-quantitative and so there can be notable differences in the level of gene expression detected by the two methods with microarrays often underestimating gene expression by two to ten fold (Conway and Schoolnik, 2003). The primers used for qRT-PCR experiments are displayed in Appendix 4.

3 Microarray and Genome Sequencing Approaches to the Study of Pneumococcal Genomic Diversity

3.1 What is a Microarray?

A microarray consists of a series of, up to several thousand, nucleic acid targets attached to a solid substrate. Hybridization of fluorescently labelled DNA or cDNA probes from test genomes using methodology developed in the 1990s from Southern blotting and reverse Northern blotting (Guo *et al.*, 1994) can allow assessment of the presence or sequence divergence of identifiable genes of interest or assessment of relative expression levels of genes (Bryant *et al.*, 2004, Ye *et al.*, 2001). There are two major classes of DNA microarrays – PCR product or "spotted" microarrays and oligonucleotide based microarrays (Ye *et al.*, 2001, Hinds *et al.*, 2002a, Bryant *et al.*, 2004, Frieberg and Brunner, 2002). In general, "spotted" microarrays are better for detecting the presence or absence of entire genes and are more economical (Finkelstein *et al.*, 2002) whereas oligonucleotide microrrays, which have a greater density of smaller probes, can detect sequence polymorphisms within a gene more readily (Bryant *et al.*, 2004, De Saizieu *et al.*, 1998, Cassone *et al.*, 2007). For "spotted" DNA microarrays, sequences corresponding to single open reading frames are preferable (Rimini *et al.*, 2000).

Several advancements in the last decade have driven the development of microarray technology. There has been a great expansion of techniques that have birthed methods for fluorescent tagging of DNA and fluorescence scanning. Techniques developed from on-chip photolithography and inkjet and microjet deposition allow the accurate "spotting" or application of DNA to minute areas of solid substrate such as glass microscope slides (which are inert at high temperature, allow covalent binding of DNA to the glass surface, tolerate high ionic washes, have low background fluorescence and permit competitive hybridization) (Hinds *et al.*, 2002a, Guo *et al.*, 1994, Dharmadi and Gonzalez, 2004), membranes (which are larger and only permit single channel hybridizations) (Hinds *et al.*, 2002a) or microchips. This allows DNA hybridization to be performed against thousands of genes simultaneously (Hinds *et al.*, 2002a, Kumar *et al.*, 2005). Fears of bioterrorist attacks and biological warfare have motivated a drive to develop technology that can rapidly identify pathogens (Wilson *et al.*, 2002, Pannucci *et al.*, 2004, Hashsham *et al.*, 2004) and has stimulated the miniaturisation of existing technologies for genomic

identification for application and utilisation outwith a laboratory. The complete sequencing of bacterial genomes has also provided a backbone template of genes against which strains of the same or similar organisms can be compared allowing studies of genomic diversity and genotyping, refinement of taxonomy and a greatly increased understanding of pathogenesis since gene expression studies need no longer be performed in isolation but in the dynamic context of enzymatic pathways and the adaptation of the organism to a range of changes in environmental exposures (Bryant *et al.*, 2004).

A growing number of human parasites as well as bacterial, fungal and viral pathogens have been studied using microarray technology in addition to *Streptococcus pneumoniae* (De Saizieu *et al.*, 1998, Peterson *et al.*, 2000, McCluskey *et al.*, 2002, Martin-Galiano *et al.*, 2004, McDaniel *et al.*, 2004, Shen *et al.*, 2006a, Orihuela *et al.*, 2004b). In an attempt to standardize data produced from different microarray platforms with different experimental designs, using different analysis methods, there exist guidelines as to the Minimum Information About a Microarray Experiment (MIAME) which should be accessible for any microarray generated dataset (Dharmadi and Gonzalez, 2004, Brazma *et al.*, 2001). In order to be compliant with the MIAME guidelines, all microarray data pertaining to these experiments are stored in BμG@SBase, the microarray data repository for the Bacterial Microarray Group at St George's Hospital (BμG@S), University of London.

3.2 Applications of Microarrays

3.2.1.1.1 Virulence studies

In the pneumococcus, microarrays allowed a comprehensive analysis of the timing and extent of activation of known components of the competence cycle along with the identification of additional loci not previously recognised as being involved in pneumococcal competence (Peterson *et al.*, 2000, Peterson *et al.*, 2004, Dagkessamanskaia *et al.*, 2004, De Saizieu *et al.*, 1998, De Saizieu *et al.*, 2000, Rimini *et al.*, 2000, Mascher *et al.*, 2006). The identification of genes involved in pneumococcal competence and transformation was also the subject of a recent study using genomic array footprinting (Burghout *et al.*, 2007). Orihuela *et al* have demonstrated different patterns of pneumococcal gene expression which occur when extracted from different anatomical locations in mice (Orihuela *et al.*, 2004b).

By comparing gene expression at different times in the growth of the pneumococcus, Ko *et al* (Ko *et al.*, 2006), showed that the late log or early stationary phase is the most virulent phase of pneumococcal growth. A growth phase dependent switch in virulence gene

expression has also been described for *Helicobacter pylori* (Thompson *et al*., 2003). An advantage of the microarray approach to studying gene expression is that they eliminate bias generated by prior selection of genes to study which are believed to be involved in pathogenesis (Cassone *et al*., 2007).

Microarrays based on the human genome have also been used to investigate the host response to pneumococcal virulence factors such as pneumolysin (McDaniel *et al*., 2004).

In bacteria other than pneumococci, combining microarray CGH with Bayesian phylogenies has allowed the identification of non pathogenic, low pathogenicity and highly pathogenic clades of *Yersinia enterocolitica* and a hypervirulent clade of *Clostridium difficile* by means of comparative phylogenomics (Howard *et al*., 2006, Stabler *et al*., 2006). Chizhikov *et al* have also used a oligonucleotide microarray to detect virulence factors in strains of *Salmonella*, *Shigella* and *Escherichia coli* (Chizhikov *et al*., 2001).

3.2.1.1.2 Drug Discovery and Development

As DNA CGH and RNA expression experiments can help identify putative gene functions, their roles in metabolic pathways or effects of regulatory systems, they are ideal for speeding up the process of identifying potential antimicrobial drug targets (Yin *et al*., 2004, Galperin and Koonin, 1999) or identifying distinct patterns of gene expression in the presence of antimicrobials. Bijlsma *et al*, (Bijlsma *et al*., 2007) and Burghout *et al* (Burghout *et al*., 2007) have recently used Genomic Array Footprinting (GAF) which combines random transposon mutagenesis and microarray technology to create a high throughput method of identifying essential genes in the pneumococcus which could be applied to aid identification of proteins to target for vaccine and antimicrobial drug development.

3.2.1.1.3 Diagnostics

It is possible to customise microarrays to address specific diagnostic issues (Kumar *et al*., 2005). One such application would be to rapidly identify the presence of antimicrobial resistance genes (Cassone *et al*., 2006, Zhu *et al*., 2007, Grimm *et al*., 2004, Call *et al*., 2003). With regard to the pneumococcus, there have been moves to develop a serotyping microarray although some of these prove unable to discriminate between serotypes, although they can discriminate between serogroups (Wang *et al*., 2007). Other serotyping microarrays, although more accurate (Hinds *et al*., 2008) cannot yet compete against the economy and rapidity of serotyping by co-agglutination. Use of microarray technology to perform MLST has also proven to be unsuccessful because of unacceptably high

misidentifications at polymorphic loci (Swiderek *et al*., 2005, Vernet *et al*., 2004). This has led to the conclusion that microarrays should not be considered a substitute for classical typing techniques (Cassone *et al*., 2007). Microarrays are also being developed which can discriminate between pneumococcal and other respiratory bacterial and viral infections (Lin *et al*., 2007).

Poor sensitivity and discriminatory power when identifying polymorphic loci, particularly when considering homologous sequences between bacterial species has limited the use of microarrays as a platform for the rapid, reliable identification of unknown organisms (Wilson *et al*., 2002), the simultaneous detection of multiple pathogens or discrimination of infection with closely related pathogens (Call, 2005, Palacios *et al*., 2007) although attempts are being made to resolve such issues (Call, 2005, Palacios *et al*., 2007, Liu *et al*., 2004). It has been estimated that hybridization will be successful if genetic dissimilarity between probe and target is less than 10-15% (Palacios *et al*., 2007). This is also influenced by the location of sequence dissimilarity as hybridization is less successful if polymorphisms are located centrally or throughout the probe sequence (Palacios *et al*., 2007). Successful hybridization is also influenced less by mismatched GC rich fragments than by mismatched AT rich fragments (Palacios *et al*., 2007). Microarrays are not well suited to enable a determination of pathogen viability or quantification of numbers of bacteria present such as would be required to estimate an infectious dose for a significant pathogen (Palacios *et al*., 2007).

Cassone *et al* have suggested that to overcome inherent sensitivity limitations with microarrays, amplification of selected sequences could be performed but notes that introducing this bias may not be acceptable in a clinical setting (Cassone *et al*., 2007). There would likely need to be a PCR-independent, whole genome amplification to avoid such bias. For such reasons as these, DNA microarrays may not be destined to become the, “standard laboratory tool,” which some have previously suggested (Ye *et al*., 2001).

3.3 Microarray Design

The technology required for microarray manufacture and the methodology behind their manufacture are reviewed by several authors (Kumar *et al*., 2005, Dharmadi and Gonzalez, 2004, Ye *et al*., 2001, Bryant *et al*., 2004, Hinds *et al*., 2002b, Hinds *et al*., 2002a, Finkelstein *et al*., 2002).

The SPv1.1 microarrays utilised in this work are manufactured by the Bacterial Microarray Group at St. George's Hospital, University of London (BμG@S) as a PCR product microarray which harbours the genes of the entire TIGR4 genome (Tettelin *et al*., 2001) plus 117 genes from the R6 genome (Hoskins *et al*., 2001) all of which are represented in duplicate and attached to an aminosilane coated glass slide (Hinds *et al*., 2002a). The microarray manufacturing procedures for this pneumococcal microarray are described by Hinds *et al* (Hinds *et al*., 2002a, Hinds *et al*., 2002b) and are identical to those published for their *Staphylococcus aureus* (Witney *et al*., 2005) and *Campylobacter jejuni* (Dorrell *et al*., 2001) arrays albeit SPv1.1 utilises pneumococcal primers and PCR products.

Optimised conditions relating to DNA or cDNA concentration used, temperature, buffer and salt concentrations are required for hybridization. Complementary hybridization between probe and target sequence results in stronger signal than between probe and mismatches, insertions or deletions (Kumar *et al*., 2005).

3.4 Microarray Analysis Methods

The fluorophores Cy3 and Cy5 have good photostability with a wide range of separation in their excitation (550nm for Cy3 and 649nm for Cy5) and emission spectra (570nm for Cy3 and 670nm for Cy5). The emitted light wavelength allows visualisation of Cy3 as red and Cy5 as green (Kumar *et al*., 2005).

DNA CGH and RNA expression experiments generate large amounts of complex data requiring tailored bioinformatics and biostatistical software, the use of which can have great impact on the interpretation of results (Kumar *et al*., 2005).

In order to address systematic variables introduced into microarray experiments through their inherent experimental design such as slide batch heterogeneity, differences in spot morphology or different degrees of background signal, a process of data normalization is required to allow direct comparisons between microarrays (Kumar *et al*., 2005, Finkelstein *et al*., 2002). For the purposes of this work, regressional normalization has been used by which a best fit slope is determined as a diagonal line on a scatter plot (Kumar *et al*., 2005, Dharmadi and Gonzalez, 2004). Normalization can be determined from this line by a variety of statistical methods in the software packages Bluefuse for Microarrays 3.5 © and Genespring GX 7.3.1 as illustrated below in Figure 3-1.

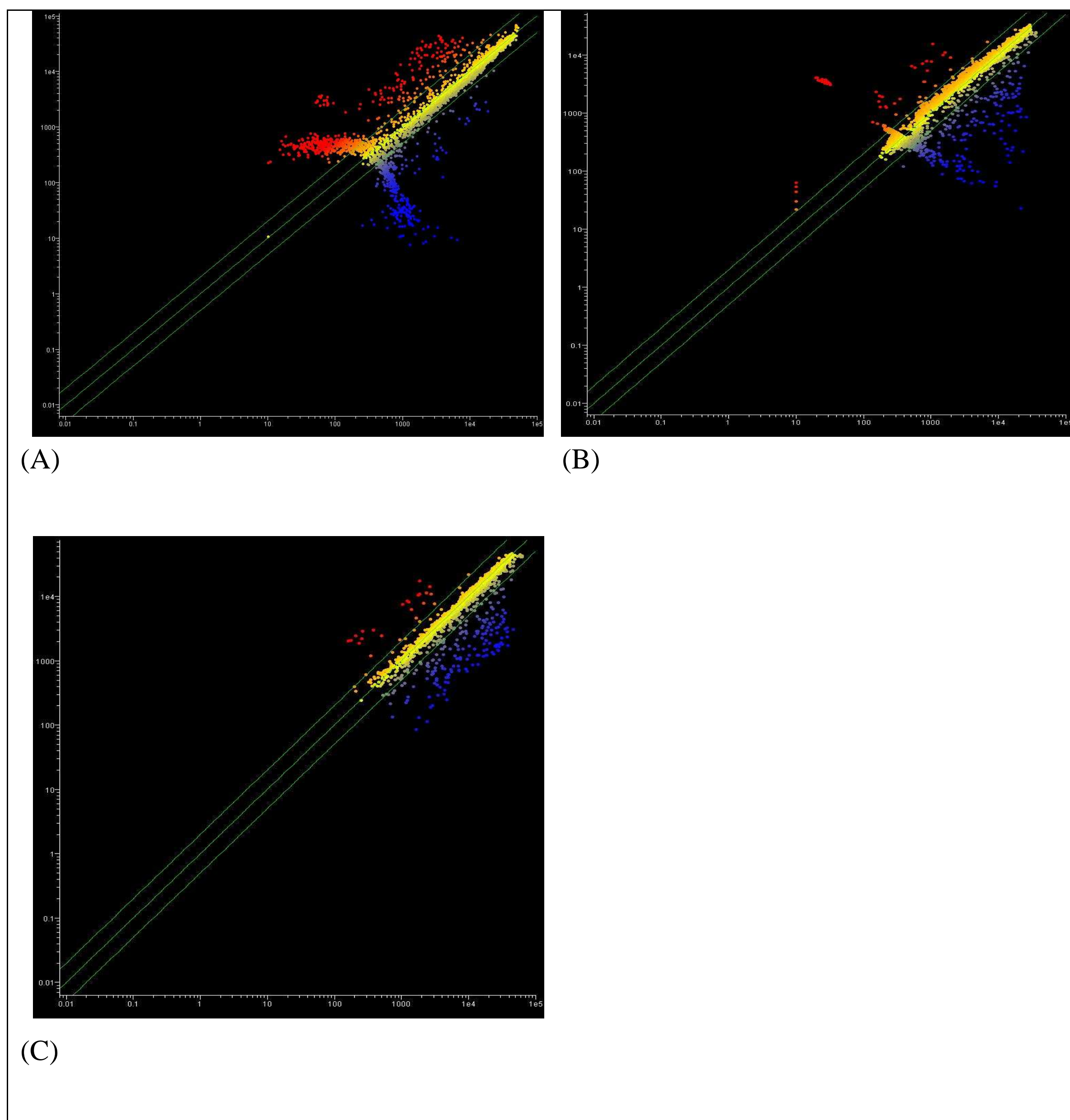


Figure 3-1 Comparison of normalization methods for CGH of Sample 03-4183.

Scatter plots A and B correspond to two separate dye swap experiments for sample 03-4183. In (A) the TIGR4 reference was labelled with Cy3 and in (B) the TIGR4 reference was labelled with Cy5. No normalization was performed using Bluefuse for Microarrays 3.5 © and Locally Weighted Scatterplot Smoothing (LOWESS) normalization was performed for each in Genespring GX 7.3.1. For both there are significant "tail artefacts" resulting from low signal hybridizations and inclusion of signal from control spots. (C) was generated using the normalization procedure described in Chapter 2 whereby the merging of both dye swaps, removal of low signal artefacts and control spot artefacts and LOWESS normalization were performed in Bluefuse for Microarrays 3.5 © with a further "Per spot-Divide by Control Channel" normalization step performed on this data in Genespring GX 7.3.1 resulting in substantially fewer false positive results.

The most appropriate normalization methodology for this work was determined by Dr Jason Hinds (BµG@S) and is described in Chapter 2. Normalization can be performed within a slide to account for efficiency of dye incorporation and the presence of staining artefacts, between the two slides of a dye swap DNA CGH experiment or across slide replicates of RNA expression experiments (Ye *et al*., 2001).

3.5 Microarray Validation Experiments

3.5.1 Validation of DNA CGH

A fully validated microarray should include evaluation of all of the probes on the microarray with all potential genomes to which they may be applied (Call, 2005). This validation has been performed for SPv1.1 by BμG@S (Hinds *et al.*, 2002a).

3.5.1.1 Determination of normalized log ratio cut off

The normalized log ratio cutoff is a means of determining the presence or absence of genes from DNA CGH data and is generated by comparing the hybridization signal strength between the clinical isolate and TIGR4. In order to determine a cut off for these experiments a method by Obert *et al* was adapted (Obert *et al.*, 2006). This identified genes SP0278, SP0410, SP0458, SP0568, SP0764, SP1018, SP0104, SP1381, SP1799, SP1975 and SP2142 on the basis that they have no paralogs, the genes are >50kb apart and no one physiological process was oversampled. The normalized log ratios for these genes as determined by Genespring GX 7.3.1 are displayed below in Table 3-1. PCR was performed for these genes on the six serotype 4, ST246 isolates which feature in Chapter 7.

Gene		Function	04-2239	03-5339	04-1342	06-1803	05-1109	06-1898
SP0104	*hdl*	Hydrolase, haloacid dehalogenase-like family	1.072	1.04	1.338	0.964	0.986	0.95
SP0278	*pepS*	aminopeptidase	0.972	0.924	0.97	0.96	0.931	0.966
SP0410	*ext*	Exfoliative toxin, putative	0.815	1.029	1.037	No Hyb	1.898	1.036
SP0458	*dinP*	DNA-damage inducible protein P	1.125	0.951	0.825	0.988	1.1	1.037
SP0568	*valS*	valyl-tRNA synthetase	0.97	0.943	1.029	0.938	1.027	1.122
SP0764	*pyrDa*	Dihydroorotate dehydrogenase A	1.004	1.034	1.132	0.907	1.064	1.197
SP1018	*tdk*	thymidine kinase	1.237	1.027	0.959	1.056	1.126	0.976
SP1381	*abcT*	ABC transporter, ATP binding protein	0.873	0.887	0.859	0.775	0.817	0.795
SP1799	*str*	sugar-binding transcriptional regulator	0.374	0.49	0.425	0.246	0.233	0.278
SP1975	*spoJ*	spoIIIJ family protein	0.883	0.989	0.948	1.027	1.07	1.018
SP2142	*rok*	ROK family protein	0.896	0.991	0.97	1.002	0.982	1.012

Table 3-1 Normalized log ratios determined by Genespring GX 7.3.1 for genes used to determine an appropriate cut-off value.

	04-2239		03-5339		04-1342		06-1803		05-1109		06-1898	
	Genespring	PCR	Genespring	PCR	Genespring	PCR	Genespring	PCR	Genespring	PCR	Genespring	PCR
SP0104												
SP0278												
SP0410												
SP0458												
SP0568												
SP0764												
SP1018												
SP1381												
SP1799												
SP1975												
SP2142												

Figure 3-2 Comparison of microarray CGH result with PCR results from the ST246 pneumococcal isolates used in Table 3-1.

Yellow indicates that hybridization occurred on the microarray or that a PCR product was obtained. Blue indicates that hybridization did not occur on the microarray or that a PCR product was not obtained.

On the basis of these investigations (excluding results for SP1799), a cutoff log ratio (fold change) of 2 was decided upon in order to have greater sensitivity by having low numbers of false negative hybridizations. Similar previous studies using this (Silva *et al*., 2006) and other (Obert *et al*., 2006) pneumococcal "spotted" arrays have used a cutoff log ratio of 1.5. However, using this "tighter" cutoff, it was evident that it was not uncommon to find poor agreement between microarray and PCR results (Silva *et al*., 2006) for genes whose normalized log ratio was near the cutoff value. However, even with a less stringent cutoff, it is clear that, as illustrated by gene SP1799 above, that there will still be instances of discrepant results when determining the presence or absence of a gene by microarray CGH or PCR. Probable reasons relate to the choice of microarray used as Joyce *et al* elucidate, saying,

> "typically spotted DNA arrays are not sensitive enough to detect the variability in genes with a single or a limited nucleotide polymorphism (Joyce *et al*., 2002)."

False positive hybridizations, which could account for the result for SP0278 in isolate 03-5339 can be the result of cross hybridization events which are more likely for genes with several alleles which could be homologous with other genes (Dorrell *et al*., 2001).

Occasionally false positive hybridizations may occur if a gene is absent but its adjacent genes are present (Dorrell *et al.*, 2001).

Microarray CGH may over-estimate genomic diversity if absence of hybridization is assumed to signify absence of a gene from a test isolate rather than diversity within its sequence. In addition a lack of hybridization can be observed where there is no sequence diversity between the PCR product and the microarray probe (see Chapter 7). Other groups have also documented this occurrence of false negative hybridization with rates as high as 41% (Peplies *et al.*, 2003). This may possibly be the result of folding of the DNA in the hybridization solution into a secondary structure incompatible with hybridization (Palacios *et al.*, 2007, Peplies *et al.*, 2003) or overlap of the microarray probe with a hypervariable region of the gene (Obert *et al.*, 2006). Whatever the mechanism is, these false negative hybridizations call into question the accuracy of identifying regions of diversity in the pneumococcal genome using microarray CGH.

3.5.1.2 Assessment of microarray hybridization accuracy in a variable region

Because of the unpredictable nature of false positive and false negative hybridizations for individual probes on microarrays, it is virtually impossible to determine a meaningful sensitivity and specificity for the overall test. Issues of false hybridization should be considered on an individual gene basis, particularly for known hypervariable genes. Obert *et al* consider this for the pneumococcal surface protein *pspA* when they realised that their microarray failed to hybridize for this gene in the majority of tested strains despite its successful amplification by PCR from the same strains (Obert *et al.*, 2006). Hybridization results described in subsequent chapters are compatible with this observation.

When the hypervariable genes which code for components of a pneumococcal pilus (SP0461-SP0468) were considered in serotype 1 isolates of ST227 and ST306, the SPv1.1 microarray failed to detect any of them although the majority of them were present and detectable using PCR (Figure 3-3).

Isolate tested	03-2672	04-2225	03-3038	06-1805	05-2739	05-1519	06-1370
MLST	ST306	ST227	ST306	ST227	ST227	ST306	ST306
Pneumonic complication	None	None	Empyema	Effusion	Empyema	Effusion	Effusion
Source of Bacteria	Blood	Blood	Pleural Pus	Blood	Blood	Blood	Blood
SP0461							
SP0462							
SP0463							
SP0464							
SP0465							
SP0466							
SP0467							
SP0468							

Figure 3-3 An investigation into the accuracy of hybridization and diversity of genes coding for the pneumococcal pilus islet *rlrA* investigated by PCR.

Yellow indicates the presence of the gene in the isolate by PCR. Blue indicates absence of the gene by PCR. All these genes failed to hybridize on the SPv1.1 microarray.

When the PCR products generated using the SP0464 primers were sequenced from the test isolates, they had 86-99% similarity to SP0464 present in TIGR4. SP0464 did not hybridize by CGH in the sequenced serotype 1 strain INV104B but again a product was detectable by PCR with a sequence which showed 93-96% similarity to the TIGR4 SP0464 as determined by BLAST software[7]. Using the Artemis Comparison Tool (ACT) it can be seen that a gene with 94% similarity to SP0464 is identifiable in the INV104B genome (Figure 3-4) where it is identified as *pmrA* which is additional evidence to suggest that the disparity between hybridization results and PCR for these genes is genuine and due to false negative hybridization rather than false positive PCR.

[7] http://blast.ncbi.nlm.nih.gov/Blast.cgi {accessed 8th January 2009}

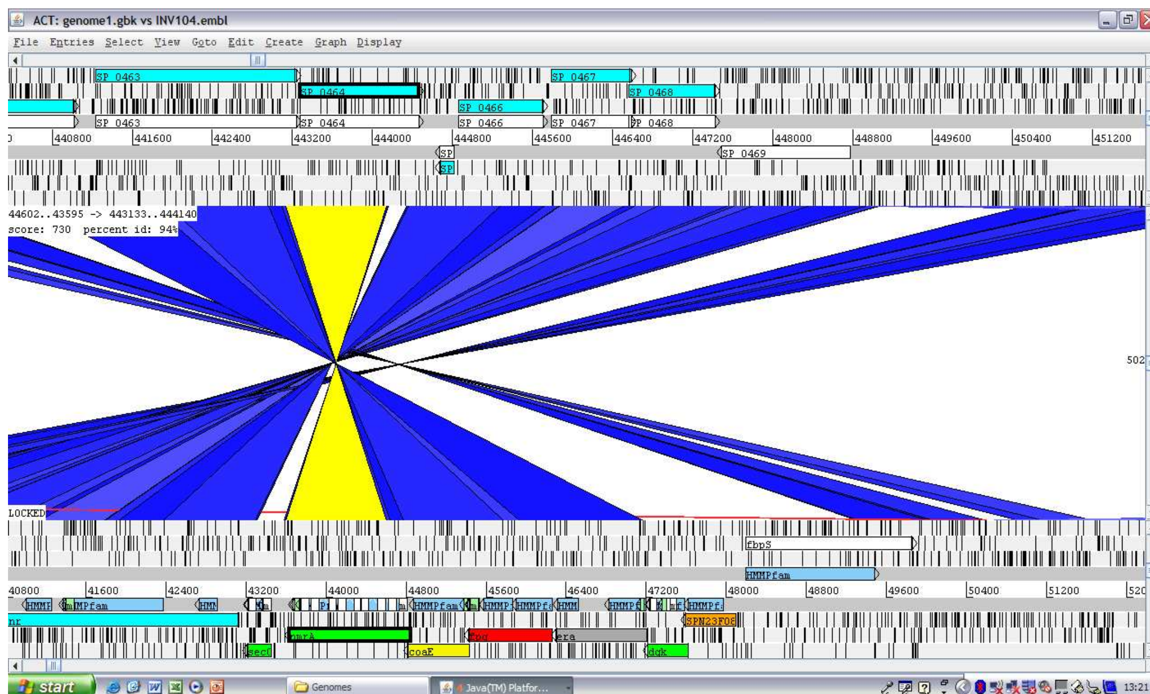


Figure 3-4 Comparison of SP0464 in the TIGR4 and INV104B genomes using the Artemis Comparison Tool (ACT).

SP0464 in TIGR4 (top of figure) is highlighted demonstrating 94% similarity (yellow triangles) to the *pmrA* gene in INV104B (bottom of figure).

This highlights that caution should be exercised in concluding that variable genes (such as the components of this pneumococcal pilus) are absent from particular serotypes when evidence is presented solely from hybridization based technology such as CGH or Southern blotting (Barocchi *et al*., 2006) or from PCR data pertaining to the whole pilus (Moschioni *et al*., 2008) rather than individual genes as noted above and by Bassett *et al* (Basset *et al*., 2007) as different methodologies can produce different results which can be misleading.

3.5.1.3 Validation of RNA Common Reference Hybridization Experimental Design

The common reference design for RNA expression experiments requires that there is little variation in the nucleic acid used as the control which in these experiments is always TIGR4 RNA. Rather than use TIGR4 DNA for our comparator in RNA expression experiments, RNA was preferable albeit total RNA rather than purely mRNA (De Saizieu *et al*., 1998). However, being more variable, initial hybridizations using TIGR4 RNA hybridized against TIGR4 RNA were performed in order to gauge how much variability inadvertently could be introduced into the experiment and determine means by which control TIGR4 RNA variability could be kept to a minimum.

3.5.1.3.1 Variation observed in TIGR4 RNA extracted from cultures grown in different media batches

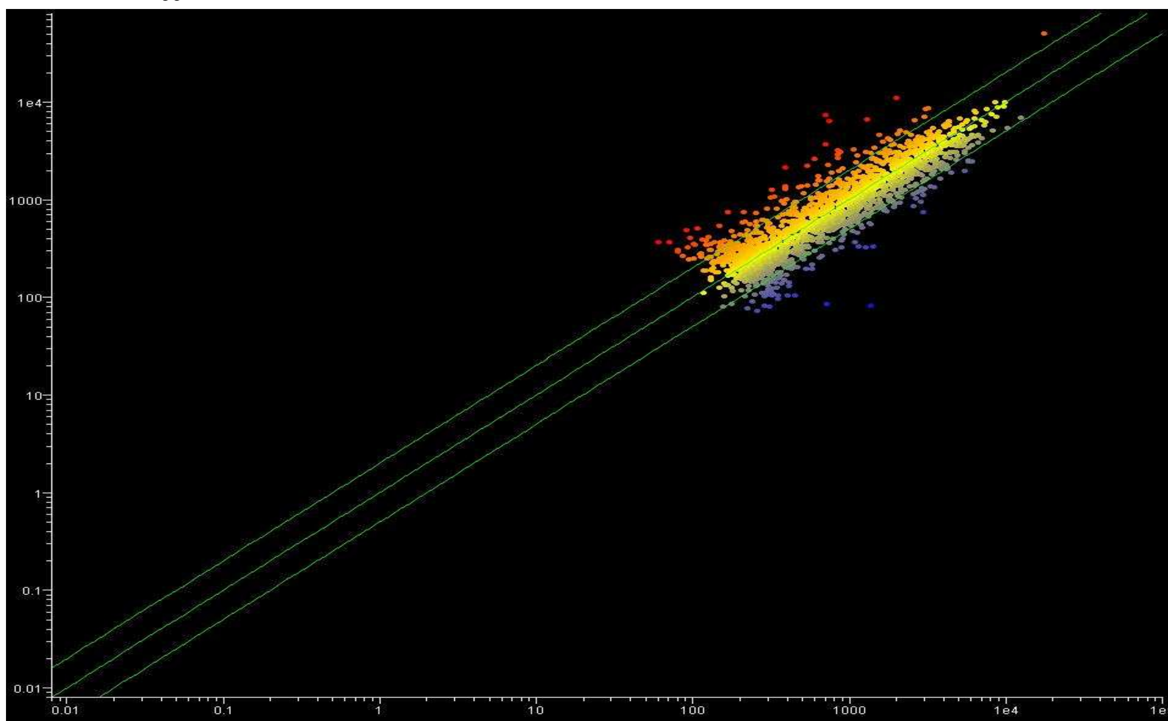


Figure 3-5 Comparison of TIGR4 gene expression from RNA extracted after growth in different batches of BHI grown on different days.

Both RNA extractions were performed on cultures grown to midlog optical density 0.6 at 600nm demonstrating that batch to batch variation substantially influences observable differences in gene expression with scatter evident extending past the 2 fold cut off lines around the line of equal fluorescence passing through the origin and which would have been considered as a significant difference in expression.

Although an increase in expression extending past the chosen fold change cutoff may be due to the cumulative effect of many different promoters for one gene (Peterson *et al*., 2004), this should not be the case when both the control and test isolates being hybridized are from the same genome and so it is legitimate to attribute the variation in Figure 3-5 above to having been introduced by the experimental design. Similar findings were considered by Gmuender *et al* who noted greater variability in hybridization of cDNA manufactured from *Haemophilus influenzae* RNA extracted from different cultures than if RNA was extracted from the same culture. They noted about 61 transcripts (about 3% of genes on their Affymetrix oligonucleotide based array) to lie above or below a 2 fold change cut off when RNA was taken from different batches of culture and attributed this increased variation to different media batches, slightly different inoculum sizes or slightly different optical densities when cells were harvested (Gmuender *et al*., 2001).

3.5.1.3.2 Variation observed in TIGR4 RNA extracted from cultures grown using the same batch of media

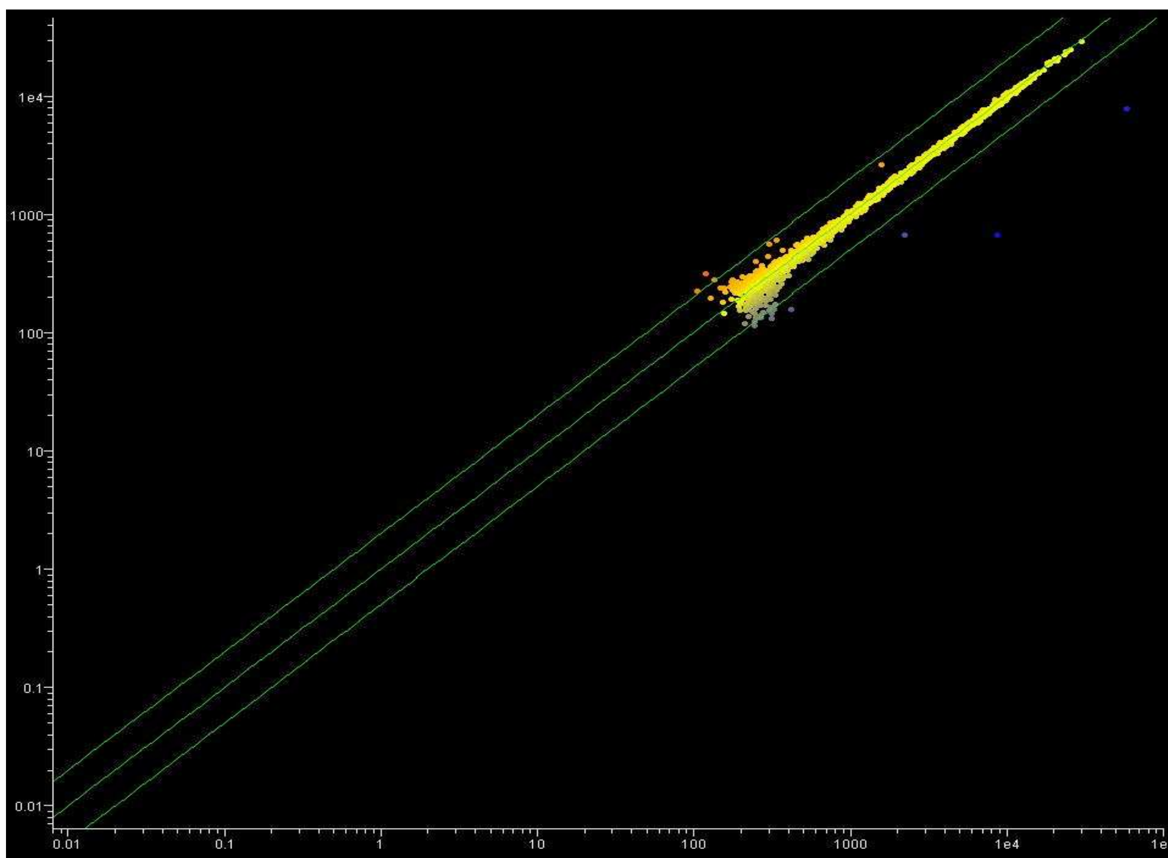


Figure 3-6 Comparison of TIGR4 gene expression from batches of RNA extracted from culture using the same batch of BHI and grown on the same day.

Both RNA extractions were performed on cultures grown to midlog optical density 0.6 at 600nm demonstrating substantially less variation in gene expression than Figure 3-5 with expressed genes generally all lying well within the lines of 2 fold cut off around the line of equal fluorescence passing through the origin.

Based on these results, whenever gene expression experiments have been performed in this work, the control channel TIGR4 RNA is always taken, for each hybridization, from the same overall culture grown in the same batch of media, from which aliquots have been taken for RNA extraction.

3.6 Comparisons of Microarray DNA CGH results with Sequenced Pneumococcal Genomes

3.6.1 Choice of Diverse Sequenced Strains

Isolates considered here have undergone Sanger sequencing. Listed below in Table 3-2 are the respective serotypes, MLST and size of genome. They demonstrate diversity in each of these categories and are of significance with regard to work described in subsequent chapters of this thesis. Genes of the TIGR4 and R6 genomes form the reference genes on the microarray which have been used in these studies for comparative genomic hybridizations. ATCC700669 is the reference strain of the Spain ^{23F}-1 Pneumococcal Molecular Epidemiology Network (PMEN) clone (Munoz *et al.*, 1991) which is the same PMEN clone as isolate South Africa 2507 used in Chapter 11. OXC141 is a carriage serotype 3 isolate of ST180 which forms part of the comparison of invasive and carriage serotype 3, ST180 isolates described in Chapter 6. INV104B is of the same serotype and sequence type as some of the isolates associated with parapneumonic manifestations considered in Chapter 10 and INV200 features in Chapter 5 in a comparison of CGH data from serotype 14, ST9 isolates.

Isolate	Serotype	MLST	Genome Length (Bp)
TIGR4	4	205	2160800
R6	Non Typeable	128	2038400
ATCC700669	23F	81	2220800
OXC141	3	180	2036800
INV104B	1	227	2140800
INV200	14	9	2098400

Table 3-2 Sanger sequenced isolates which have been compared with microarray DNA CGH results.

3.6.2 Comparison of Microarray DNA CGH results with Genome Sequence Data

3.6.2.1 Diversity of TIGR4 genes in test genomes

Appendix 5 compares the TIGR4 genes identified as not hybridizing in at least one of the four test sequenced genomes with their presence or absence from the sequencing data. The Artemis Comparison Tool can be used to visualise these differences between the TIGR4 reference genome and the test sequenced genomes. This is illustrated in Figure 3-7.

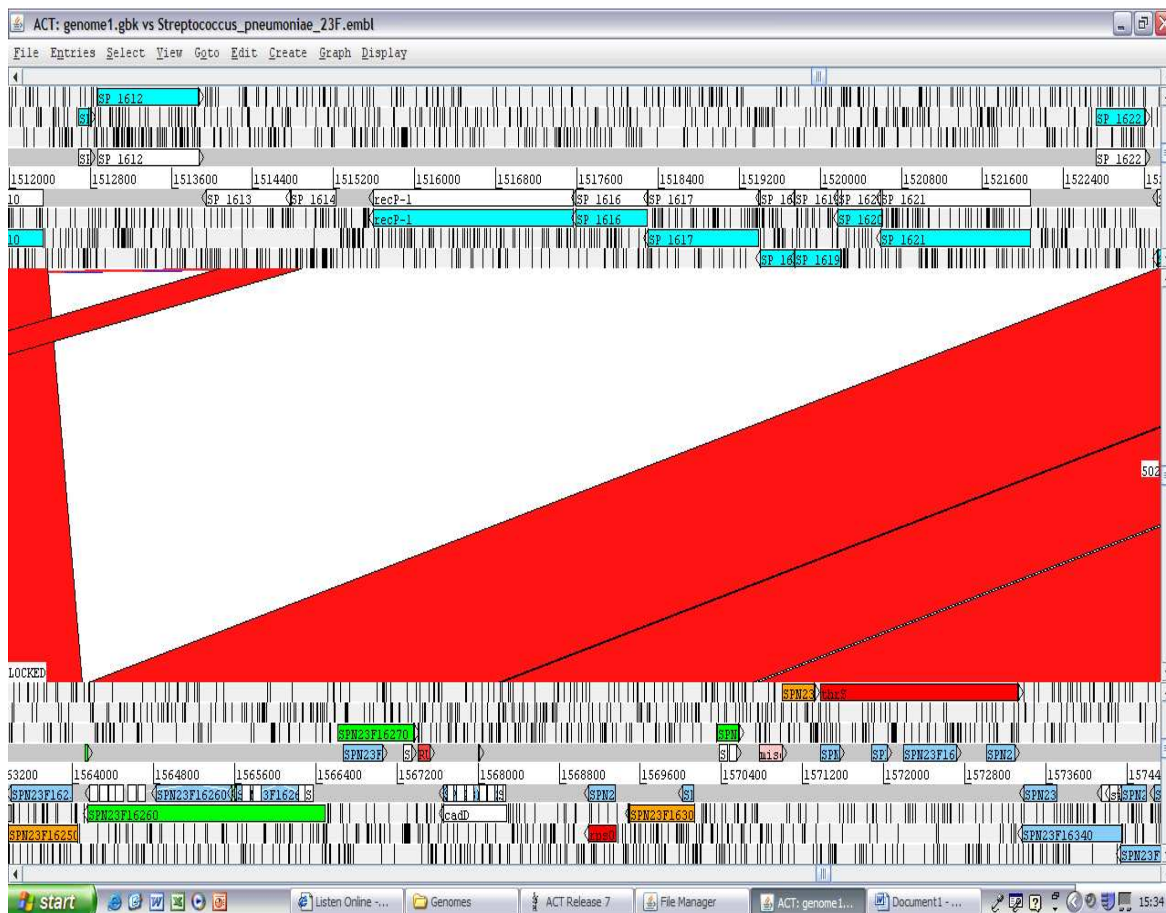


Figure 3-7 ACT comparison of the TIGR4 genome with the ATCC700669 genome.

The red lines indicate the presence of homologous genes in both sequences. This screen shot indicates that there are no homologs of the TIGR4 genes (top of figure) SP1615-1622 in the ATCC700669 genome (bottom of figure), consistent with the microarray CGH results which indicated no hybridization of DNA from ATCC700669 to the probes for these genes.

3.6.2.2 Regions of Diversity

The regions of diversity described in Chapter 1, Table 1-2 were identified from microarray CGH studies. On identifying these regions in the sequenced genomes, it is clear that the full extent of these regions has not been realised using the microarray CGH method which often underestimates the size of the region of diversity.

Region of Diversity	Size of Region identified by CGH (Silva *et al.*, 2006)	Size of Region identified by Sequence Comparison (unpublished data from Nicholas Croucher, Wellcome Trust Sanger Institute)
RD 1	SP0067-0074	SP0066-0075
RD 2	SP0109-0115	SP0107-0118
RD 3	SP0163-0168	SP0161-0173
RD 4	SP0346-0360	SP0342-0366
RD 5	SP0378-0380	SP0378-0381
RD 6	SP0394-0397	SP0390-0400
RD 7	SP0460-0468	SP0459-0471
RD 8	SP0473-0478	SP0471-0478
RD 9	SP0531-0544	SP0529-0544
RD 10	SP0643-0648	No diversity seen in ATCC700669, OXC141, INV104B or INV200
RD 11	SP0644-0666	No diversity seen in ATCC700669, OXC141, INV104B or INV200
RD 12	SP0692-0700	SP0690-0701
RD 13	SP0888-0891	SP0866-0892
RD 14	SP0949-0954	SP0948-0954
RD 15	SP1050-1065	SP1045-1067
RD 16	SP1129-1147	SP1128-1151
RD 17	SP1315-1352	SP1306-1354
RD 18	SP1433-1444	No diversity seen in ATCC700669, OXC141, INV104B or INV200
RD 19	SP1612-1622	SP1610-1623
RD 20	SP1756-1773	SP1754-1774
RD 21	SP1793-1799	SP1790-1800
RD 22	SP1828-1830	SP1827-1833
RD 23	SP1911-1918	No diversity seen in ATCC700669, OXC141, INV104B or INV200
RD 24	SP1948-1955	SP1946-1956
RD 25	SP2159-2166	SP2158-2167

Table 3-3 Comparison of size of regions of diversity identified by microarray CGH with the size of the same regions of diversity identified from sequence data.

In addition, comparison of the sequences for ATCC700669, OXC141, INV104B and INV200 identifies a further 47 newly identified regions of sequence diversity. These regions are: SP0020-0022, SP0027-0032, SP0054-0055, SP0079-0083, SP0137-0139, SP0141-0144, SP0267-0271, SP0338-0340, SP0376-0381, SP0390-0400, SP0489-0493, SP0502-0510, SP0517-0519, SP0529-0544, SP0566-0576, SP0737-0741, SP0783-0784, SP0794-0795, SP0825-0828, SP0873-0875, SP0881-0885, SP0866-0893, SP0906-0908, SP0938-0943, SP0948-0958, SP1008-1009, SP1018-1020, SP1035-1040, SP1153-1155, SP1173-1176, SP1212-1213, SP1291-1293, SP1356-1360, SP1466-1467, SP1546-1547, SP1563-1564, SP1706-1707, SP1739-1742, SP1782-1784, SP1827-1833, SP1848-1851, SP1885-1887, SP1903-1906, SP1984-1985, SP1988-1989, SP2092-2094, SP2135-2141 (Nicholas Croucher, Wellcome Trust Sanger Institute, personal communication).

3.6.2.3 Identification of R6 genes in test genomes

The four sequenced genomes were compared to the R6 genome, and no new regions of diversity were identified other than those described by Bruckner *et al* (Bruckner *et al*., 2004). The genome sequence data has an advantage over microarray CGH data in that it can definitively identify the absence of R6 genes. Although the R6 genes identified as present by microarray CGH were, in the majority of cases, correctly identified, the microarray approach generated a significant number of false negative hybridizations for R6 genes and underestimated the number of R6 genes present in each genome. There were two false positive results in the microarray CGH results as spr0118 and spr0966 were incorrectly identified as present in isolate ATCC700699. The results of the microarray CGH dye swap experiments and the Sanger sequencing results are compared in Figure 3-8 below. As CGH experiments were performed using TIGR4 and not R6 as the reference DNA, it is not possible to draw any conclusions about the non hybridization to the microarray probes as, for the R6 genes represented on the microarray, binding of test isolate DNA to the microarray probes is not competitive against reference R6 DNA in these instances.

Gene	Gene Function	CGH Data				Sequence Data			
		INV104B	INV200	OXC141	ATCC700669	INV104B	INV200	OXC141	ATCC700669
spr0067	Conserved Hypothetical Protein								
spr0104	Hypothetical Protein								
spr0105	Transporter, truncation								
spr0106	Transporter, truncation								
spr0107	Hypothetical Protein								
spr0108	Conserved Hypothetical Protein (similar to comB protein)								
spr0111	Hypothetical Protein								
spr0112	Hypothetical Protein								
spr0113	Hypothetical Protein								
spr0114	Hypothetical Protein								
spr0115	Hypothetical Protein								
spr0116	Hypothetical Protein								
spr0117	Hypothetical Protein								
spr0118	Hypothetical Protein								
spr0119	Hypothetical Protein								
spr0225	Hypothetical Protein								
spr0320	Type 2 capsule locus cps2L								
spr0321	Type 2 capsule locus cps2M								
spr0322	dTDP-glucose-4,6-dehydratase cpsN								
spr0323	dTDP-L-rhamnose synthase cpsO								
spr0416	Hypothetical Protein								
spr0491	Hypothetical Protein								
spr0493	Conserved Hypothetical Protein								
spr0703	Hypothetical Protein								
spr0800	Hypothetical Protein								
spr0955	Hypothetical Protein								
spr0956	Hypothetical Protein								
spr0957	Tn 5252, relaxase, truncation								
spr0958	Tn 5252, relaxase, truncation								
spr0959	Hypothetical Protein								
spr0960	Similar to positive transcriptional regulator MutR								
spr0966	Conserved Hypothetical Protein (Probable acid-CoA ligase)								
spr0972	Conserved Hypothetical Protein								
spr1042	Immunoglobulin A1 protease								
spr1093	Conserved Hypothetical Protein								
spr1114	Conserved Hypothetical Protein (Probable transcriptional regulator)								
spr1179	Conserved Hypothetical Protein								
spr1403	Hypothetical Protein								
spr1404	Conserved Hypothetical Protein								
spr1478	Hypothetical Protein								
spr1549	Hypothetical Protein								
spr1550	Conserved Hypothetical Protein (similar to MutR protein)								

Figure 3-8 Comparison of microarray CGH results for R6 genes with genome sequence data for the same genes.

Red indicates the presence of an R6 gene, identified by either method. Blue indicates the absence of the gene from sequence data.

3.6.3 Genomic Diversity Identified by Genome Sequence Data Alone

To illustrate how genomic diversity may be identified from pneumococcal genome sequence data alone using ACT and impact on the understanding of virulence, two examples of diversity affecting known virulence factors in these isolates are considered – serotype 2 capsule genes and IgA1 protease.

3.6.3.1.1 Serotype 2 capsule genes

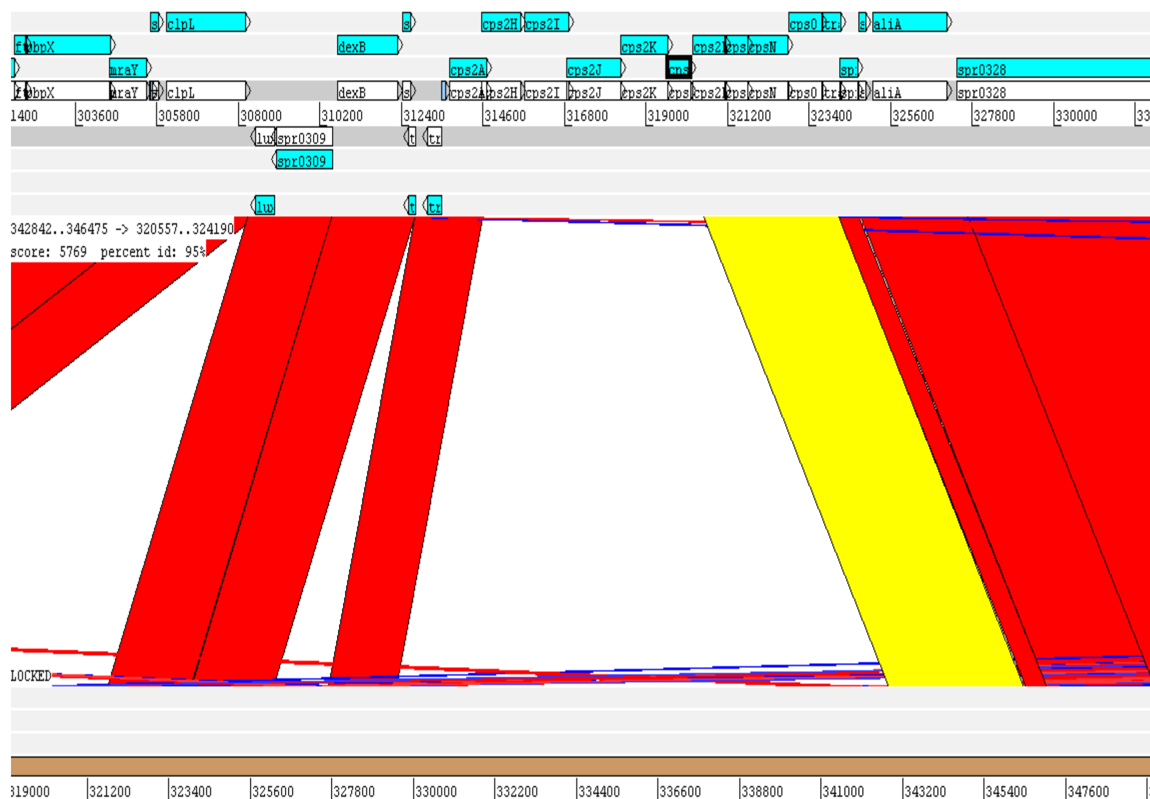
As R6 is an avirulent, non-capsulated laboratory strain derived from a serotype 2 strain, called D39, it is not surprising that remnants of the serotype 2 capsule locus remain in the R6 genome (Lanie *et al*., 2007). It is of interest that four of the serotype 2 capsule genes (cps2L, cps2M, cpsN and cpsO) are identifiable in the genomes of INV104B and ATCC700669 (neither of which are serotype 2) with 94-95% homology to the R6 versions of the genes suggesting that they may have been introduced by horizontal gene transfer or are fulfilling a function other than solely capsule polysaccharide synthesis. All four of these genes are involved in dTDP-rhamnose biosynthesis. This is illustrated below by ACT in Figure 3-9.

Figure 3-9 ACT comparison of the spr0320-spr0323 genes in OXC141 (A), ATCC700669 (B) and INV200 (C).

The R6 genome is displayed at the top of each figure and the test genome is displayed beneath. In (A) the genes spr0320-spr0323 are absent in OXC141 but present in both ATCC700669 (B) and INV104B (C) with a high degree of homology to their counterparts in R6 (identified by the yellow bands).

A. ACT comparison of R6 and OXC141 genomes at region spr0320-spr0323.

B. ACT comparison of R6 and ATCC700669 genomes at region spr0320-spr0323.



C. ACT comparison of R6 and INV200 genomes at region spr0320-spr0323.

3.6.3.2 IgA1 protease

A further interesting finding was the presence of the R6 gene spr1042 in all four of the test sequences. This is significant as it codes for IgA1 protease which is noted in Chapter 1 as a divergent gene in the pneumococcal genome allowing evasion of host mucosal immunoglobulin. Its homolog in TIGR4 is SP0071 which has been identified as a virulence factor required in a mouse pneumonia model (Hava and Camilli, 2002). Interestingly, the TIGR4 allele for this gene is absent from all these four sequenced strains using Sanger sequencing and does not hybridize from any of the test strains to the probe for SP0071 on the microarray, which is not surprising given its absence from their genomes.

Nevertheless all four strains have a gene for IgA1 protease but CGH comparisons made solely against the genome of TIGR4 as a reference genome could easily lead to the erroneous conclusion that these strains lack IgA1 protease and cast doubt over whether it genuinely is required for virulence. This highlights the fact that making assumptions about the presence or absence of genes based on comparisons with only one allele of a gene from a single reference genome produces a skewed understanding of “core” gene content in bacteria with such a high degree of genomic diversity as the pneumococcus. The absence of the reference genome allele of a gene does not necessarily equate with the absence of the gene *per se* as other alleles may be represented.

3.6.4 Using Pneumococcal Genome Sequences to Identify Gene Insertions

One significant limitation of a microarray CGH approach to understanding genomic diversity is that a microarray cannot identify genes in a genome if there are no probes for these genes represented on the microarray. Consequently, the microarray used for this work cannot identify additional pneumococcal genes which are not present in the TIGR4 genome or the R6 genome. It is clearly the case that the pool of genes which the pneumococcus can host is greater than those present in the TIGR4 and R6 genomes and, through transmissible genetic elements and horizontal gene transfer, the pool is not even limited by current bacterial species boundaries (Hakenbeck *et al*., 2001). Knowledge of the full test genome sequence and the ability to compare this with other pneumococcal genome sequences does allow the identification of regions of genetic material which have been inserted into the genome. Such inserted sequences can then be analysed to predict open reading frames *in silico* with software such as GLIMMER (Gene Locator and Interpolated Markov ModelER) (Delcher *et al*., 1999). Identification of putative functions for these genes can be performed using software such as BLAST (Basic Local Alignment Search Tool) (Altschul *et al*., 1990) by comparing known sequences in pneumococci and other bacteria *in silico* (Parkhill, 2002).

Of the four test sequenced genomes utilised in this chapter, only two (ATCC700669 and OXC141) are considered here with regard to the additional genetic material which they harbour as they have been sufficiently annotated to identify the putative functions of these inserted genes in most cases. In ATCC700669, 202 additional genes were identified in the genome when compared to TIGR4 using ACT (Appendix 6) and for OXC141, 123 additional genes were identified using ACT (Appendix 7).

In both these genomes, large regions of inserted genetic material appear to relate to the integration of genes of bacteriophage origin.

3.6.4.1 Genetic Material in the ATCC700669 Genome in addition to that seen in the TIGR4 genome

The full list of additional genes identified in the ATCC700669 genome is listed in Appendix 6. There is evidence of genetic insertion through bacteriophage integration and through transposon activity. A region of 49 genes inserted between the homologs of the

TIGR4 genes SP1563 and SP1564 appears to have originated from a bacteriophage and is illustrated below using ACT (Figure 3-10).

Figure 3-10: Comparison of gene insertions not present in the TIGR4 genome but present in the ATCC700669 genome using ACT.

This illustrates the absence in TIGR4 of some of the 49 genes present in ATCC700669 (bottom of figure) which appear inserted between the homologs of the TIGR4 genes (top of figure) SP1563 and SP1564.

Also of interest is a region inserted between the homologs of SP1806 and SP1807 which codes for components of a pneumococcal phosphotransferase (PTS) system (The diversity of pneumococcal genes coding for PTS systems is considered more in Chapter 6). This can also be easily visualised in ACT (Figure 3-11).

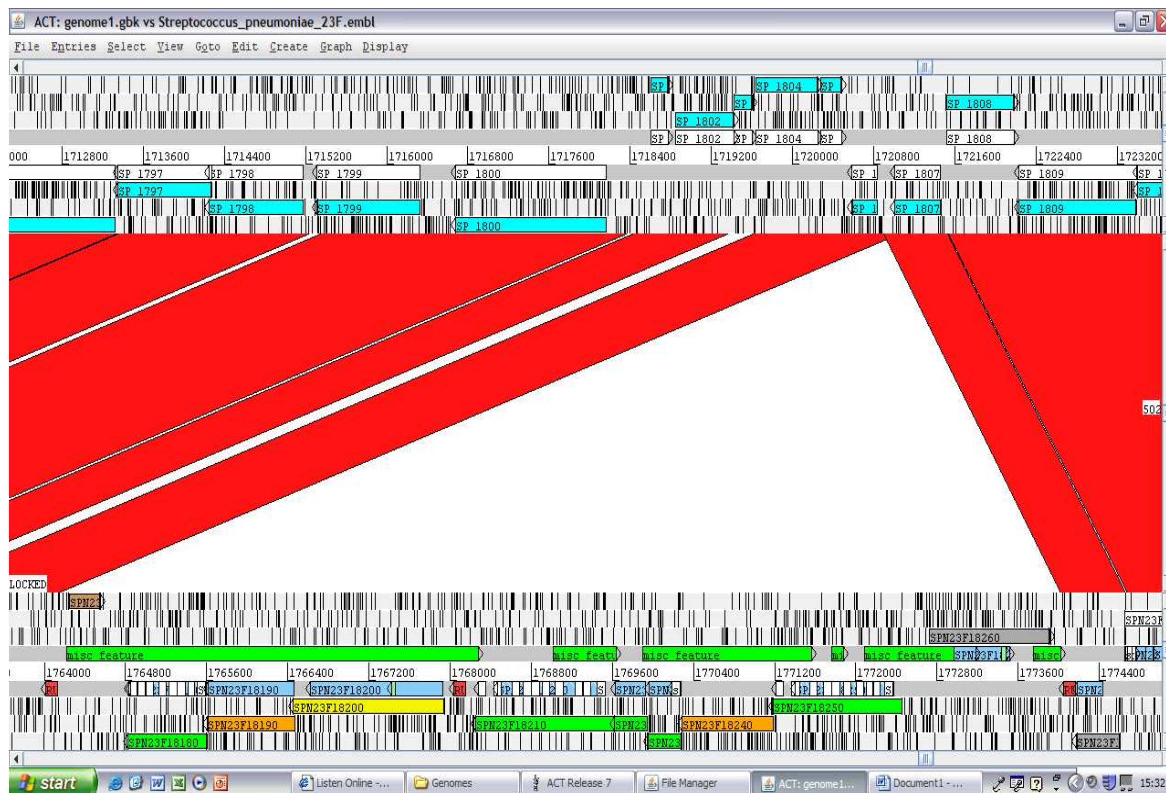


Figure 3-11 Comparison of PTS system genes in the TIGR4 genome with the ATCC700669 genome using ACT.

This illustrates the absence of some of the PTS system genes present in ATCC700669 (bottom of figure) which appear inserted between the homologs of the TIGR4 genes (top of figure) SP1806 and SP1807.

Of greater clinical significance is the finding of genes in the ATCC700669 genome involved in tetracycline resistance (SPN23F13050) and chloramphenicol resistance (SPN23F12590) as ATCC700669 is resistant to multiple classes of antimicrobial (Munoz *et al*., 1991).

3.6.4.2 Genetic Material in the OXC141 Genome in addition to that seen in the TIGR4 genome

As the OXC141 genome is smaller than that of TIGR4, it was expected that there would be less additional genetic material than that identified in the much larger ATCC700669 genome. On comparing the TIGR4 and OXC141 genomes, it is evident that between the OXC141 homologs of the TIGR4 genes SP0019 and SP0020, a sizeable amount of DNA has been inserted. This region has recently been found to code for a pneumococcal prophage called phiOXC (Dr Patricia Romero, University of Glasgow, personal communication). Genes identified by ACT as present in OXC141 but not in TIGR4 are displayed in Appendix 7.

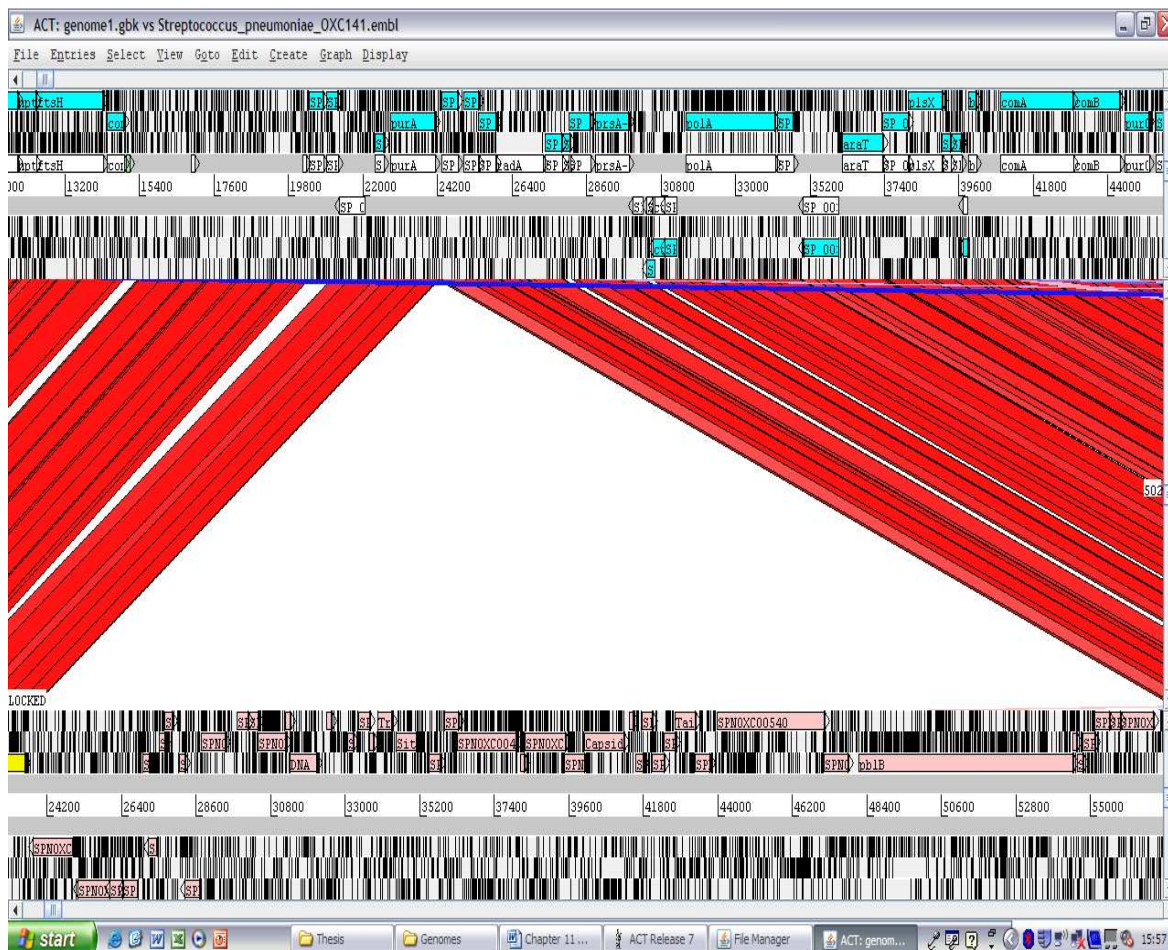


Figure 3-12 Comparison using ACT of the TIGR4 genome with the OXC141 genome using ACT demonstrating the insertion site of the phiOXC prophage.

This comprises 43 additional genes in the OXC141 genome (bottom of figure) which do not appear in the TIGR4 genome (top of figure). These genes are inserted between the OXC141 homologs of the TIGR4 genes SP0019 and SP0020.

3.7 Discussion

3.7.1 Advantages and Disadvantages of a Genome Sequencing Approach to mapping Genomic Diversity compared to microarray based CGH investigations

Full genome sequencing clearly has advantages over microarray based CGH in terms of the data it generates and its sensitivity and accuracy. It is not without its own hurdles though, some of which will now be considered. Although costs are falling, it has traditionally been expensive and time consuming to sequence a genome, requiring significant capital investment for sequencing hardware and bioinformatics support (Ryan *et al.*, 2007). The difficulties of correctly assembling fragments of the genome sequence (and the consequent need for specialist bioinformatics support) should not be underestimated,

particularly given the fact that pneumococcal genomes often exhibit runs of repeat sequences such as BOX and RUP.

It has been calculated that aiming for a US$1000 genome sequencing technology requires a 10^4 fold reduction in cost per base and a 10^3 fold increase in the number of bases read per second (Ryan *et al*., 2007). When starting this study of pneumococcal genomic diversity in 2005, such costs and the absence of access to a suitable technological platform on which to perform genomic sequencing meant that this was not a practical option for considering the genomes of several diverse isolates, particularly as a dye swap CGH experiment of a single pneumococcal isolate could be performed for £200 and still provide unique, albeit less accurate, insights into the genomic diversity of pneumococcal populations. Over the course of these investigations, access to sequencing technology has improved and the costs involved have decreased which has allowed a selection of the isolates used in the course of this work to be sequenced by 454 sequencing at the Wellcome Trust Sanger Institute, although these are still in the process of being assembled.

It is expected that in the future, whole genome sequencing will become the gold standard for investigations of the genomic diversity of pneumococci but, until then, microarray based CGH of pneumococcal genomes is an acceptable (and less expensive) option. This allows the investigation of a wider range of isolates while the microarray platform can also be used to gain new insights into the diversity of expression of pneumococcal genes as well as investigating the diversity of their presence. For these reasons, the majority of the genomic comparisons performed in this work have been performed using a microarray based approach.

Although full genome sequencing demonstrates advantages over microarray CGH, it is worth re-iterating the observation of Parkhill who warns that,

> "All but the most simple annotation is an interpretation of the sequence, and is thus subject to error and misinterpretation (Parkhill, 2002)."

It is also worth re-iterating that although genetic differences can be identified *in silico* between genomes, functional studies are still required to demonstrate the effects of these genes to truly establish a role for them in the pathogenesis of pneumococcal disease or carriage (Moscoso *et al*., 2005). Neverthless, these tools, which are in a rapid phase of development and improvement are, particularly when used in combination, establishing a much more informed understanding of the scale and dynamicity of genomic diversity in the pneumococcus and how that relates to its diverse disease presentations.

3.7.2 The Distributed Genome Hypothesis and a Pneumococcal Supragenome

These comparisons of pneumococcal isolates, by full genome sequencing or microarray CGH, suggest that any individual isolate of pneumococcus contains only a fraction of the possible combination of pneumococcal genes. There is a greater pool of genes represented by the wider global pneumococcal population (the pneumococcal supragenome) from which it can draw upon, under suitable conditions, to adapt to its environment, such as occurs during the phenomenon of capsular switching described in Chapter 1. Sequencing of such non-essential regions of the genome can greatly enhance understanding of the functions of such genes and identify potential roles in virulence or the generation of antigenic diversity (Mavroidi *et al*., 2004, Bentley *et al*., 2006, Bagnoli *et al*., 2008). For instance, a small change in sequence such as a single polymorphism in the rhamnosyl transferase gene, *wciP*, can alter capsule expression from that of serotype 6A to serotype 6B (Mavroidi *et al*., 2004).

Gaining an understanding of both the combination of genes present in clinical pneumococcal isolates and their allelic forms, may therefore improve understanding of pneumococcal disease pathogenesis and provide potential targets for therapeutic intervention or disease prevention. Observed redundancy in a bacterial genome is not unique to pneumococci but is also documented in many other bacteria such as *Escherichia coli* and *Mycoplasma genitalium* while in bacteria such as *Bacillus anthracis* there is very little diversity (Fraser-Liggett, 2005). Such observations and evidence of genomic diversity resulting from horizontal gene transfer have resulted in a “distributed genome” hypothesis for such bacteria where virulence traits are acquired by horizontal gene transfer to benefit the bacterial population as a whole rather than an individual organism (Ehrlich *et al*., 2008).

3.7.3 Microarray Limitations

Before considering the results of the microarray experiments in subsequent chapters, it is worth reiterating some of the limitations of this technique and their potential influence on the interpretation of results.

Microarrays can only detect complementary sequences of the genes which are represented on them during their manufacture.

Microarrays tell us nothing about the location of genes in the genome although this may influence hybridization if a probe for a deleted gene is flanked by probes for two genes which remain present and may result in false positive hybridization. False positive hybridizations probably occur infrequently and may also result from cross hybridization by homologous genes.

The absence of hybridization on a microarray is not synonymous with absence of a gene from the test genome but requires further investigation by PCR and sequencing. The presence of a PCR product of appropriate size may be due to a divergent gene, possibly containing single nucleotide polymorphisms, or the gene has the correct sequence but has formed a secondary structure in solution which prevents its hybridization. Determining whether a gene is absent is particularly difficult if the gene is known to be hypervariable and lack of hybridization may be due to the presence of an entirely novel sequence not represented on the microarray.

Microarrays cannot detect multiple copies of genes but the presence of multiple copies in an expression experiment may cause an apparent “increase” in expression.

4 Genomic Diversity Observed in Phenotypically Diverse Pneumococcal Isolates

4.1 Choice of Diverse Pneumococcal Isolates

Identified regions of diversity in the TIGR4 genome and those of other pneumococcal strains are outlined in Chapter 1 and 3. These have included strains from the USA (Shen *et al*., 2006a, Silva *et al*., 2006, Hakenbeck *et al*., 2001), United Kingdom (Silva *et al*., 2006, Joyce *et al*., 2002), South Africa (Silva *et al*., 2006, Bruckner *et al*., 2004, Hakenbeck *et al*., 2001), Norway (Silva *et al*., 2006), Germany (Bruckner *et al*., 2004, Hakenbeck *et al*., 2001), Papua New Guinea (Bruckner *et al*., 2004, Hakenbeck *et al*., 2001), Hungary (Bruckner *et al*., 2004, Hakenbeck *et al*., 2001, Joyce *et al*., 2002), Switzerland (Bruckner *et al*., 2004, Hakenbeck *et al*., 2001), France (Bruckner *et al*., 2004, Hakenbeck *et al*., 2001), Finland (Bruckner *et al*., 2004, Hakenbeck *et al*., 2001, Joyce *et al*., 2002), Taiwan (Joyce *et al*., 2002) and Spain (Bruckner *et al*., 2004, Hakenbeck *et al*., 2001, Joyce *et al*., 2002) and have often included Pneumococcal Molecular Epidemiology Network (PMEN) clones (Bruckner *et al*., 2004, Silva *et al*., 2006, Hakenbeck *et al*., 2001) which are of significance both because they are of public health concern due to antimicrobial resistance but also because they are examples of successful clones of pneumococci which demonstrate a selection advantage that has enabled their international spread (McGee *et al*., 2001a).

In 2001, Hakenbeck *et al* suggested,

> "Using a large number and a wide spectrum of genetically different strains, one could expect a gradual variability among the pneumococcal population, since despite the recognition of clonal spread, population analysis suggested a freely recombining structure characteristic of transformable organisms (Hakenbeck *et al*., 2001)."

This conclusion was based on data published by Hall *et al* who had demonstrated that the pneumococcus has an epidemic population structure (Hall *et al*., 1996).

This study has chosen 10 isolates for DNA CGH analysis and comparison as preliminary work before focusing on diversity within particular clonal complexes and clinical conditions. The isolates were chosen to represent isolates from different countries (United Kingdom, USA, South Africa and Bolivia) which included PMEN clones to represent isolates of different serotypes and of multilocus sequence types which had epidemic

potential but which were unrelated to each other. The absence of relationship between the sequence types was determined using eBURST software (Feil *et al*., 2004) as the algorithm it uses, based on the MLST allelic profiles, is more accurate at determining relationships within clonal complexes than software which generates phylogenetic trees. This is because frequent horizontal gene transfer in the pneumococcus has been noted to affect the construction of phylogenetic trees using housekeeping genes (Feil and Spratt, 2001).

Isolate	Serotype	MLST	Background Details
BAA-659	6A	ST376	American Type Culture Collection (ATCC) accession number for the PMEN clone N. Carolina ^{6A}-23. The clinical background for the isolate is not known.
BAA-340	14	ST67	ATCC accession number for the PMEN clone Tennessee^{14}-18 (Gherardi *et al*., 2000). The clinical background for the isolate is not known but the clone has also been detected causing IPD in Scotland (Smith *et al*., 2006).
05-1271	20	ST568	Blood culture isolate from a 12 year old girl in 2005 from Paisley, United Kingdom.
ATCC700904	19A	ST41	ATCC700904 is the accession number for the PMEN clone South Africa ^{19A}-13. Used in CGH experiments by Silva *et al* as PMEN13 (Silva *et al*., 2006). The clinical background for the isolate is unknown.
BAA-660	35B	ST377	ATCC accession number for the PMEN clone Utah ^{35B}-24. The clinical background for the isolate is not known but this clone has been detected in 2000 as a cause of IPD in Scotland (Smith *et al*., 2006).
05-2565	12F	ST218	Blood stream isolate from a 79 year old man in Glasgow, United Kingdom cultured in 2005.
05-1821	9V	ST156	Blood stream isolate from a 63 year old woman from Ayr, United Kingdom cultured in 2005.
OXC141	3	ST180	Carriage isolate from an individual from Oxford. It has been sequenced at the Wellcome Trust Sanger Institute.
ATCC51916	23F	ST37	ATCC accession number for the PMEN clone Tennessee ^{23F}-4. Identified in 1991 from a patient with meningitis who did not respond to treatment with a cephalosporin (McDougal *et al*., 1995, Sloas *et al*., 1992). This clone has also been detected as an infrequent cause of IPD in Scotland (Smith *et al*., 2006)
07-2839	9A	ST239	Isolated from the nasopharynx of a 9 year old girl in 2007 with chronic otitis media from Trinidad, Bolivia.

Table 4-1 Details of isolates of diverse serotype and MLST used in microarray DNA CGH experiments.

4.2 Microarray Results

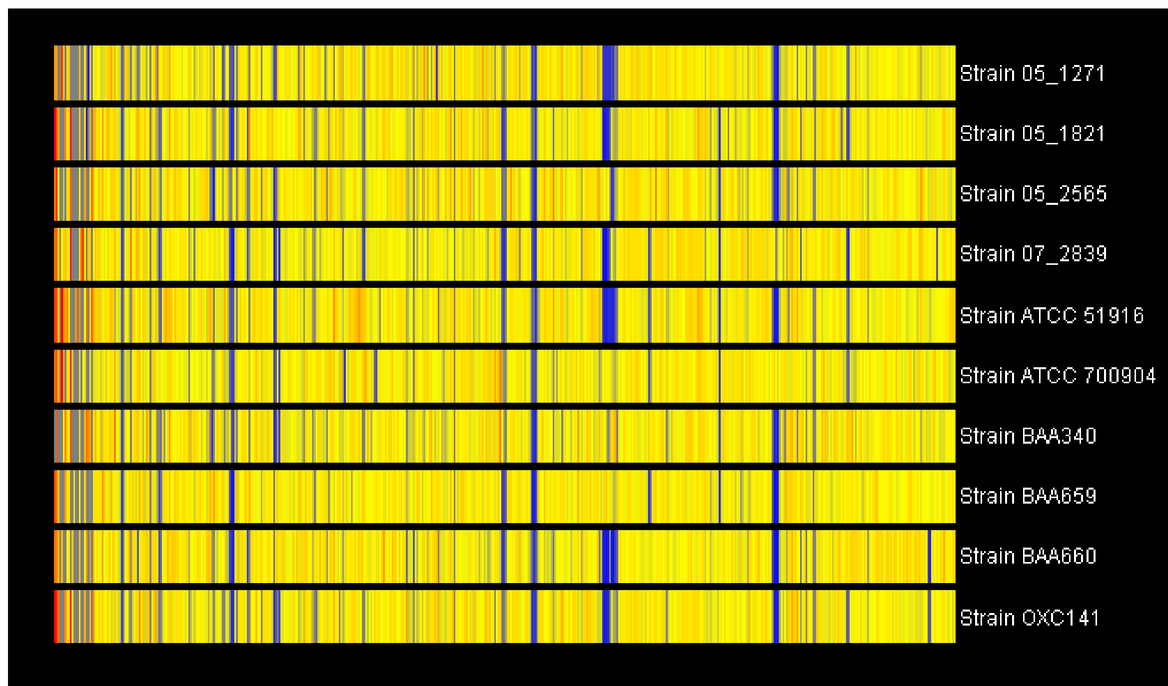


Figure 4-1 Whole pneumococcal genome view of the DNA CGH results for the 10 chosen isolates of diverse sequence type generated by Genespring GX 7.3.1.

Each horizontal bar represents a genome and each vertical coloured bar within these represents a gene. Shades of yellow indicate hybridization by both TIGR4 and the test strain DNA. Shades of blue indicate hybridization by TIGR4 DNA alone. Shades of red indicate hybridization by the test isolate DNA alone and grey indicates no hybridization by either TIGR4 DNA or the test strain DNA. It is clear that these hybridization patterns show few similarities between the 10 isolates.

4.3 Discussion

4.3.1 Regions of Diversity in the TIGR4 and R6 Genomes

Differences between the test isolate genomes and the TIGR4 genome are illustrated in Appendix 8. Of the R6 genes represented on the microarray which do not feature in the TIGR4 genome it was possible to identify from the hybridization patterns for the 10 test isolates all six regions of diversity described by Bruckner *et al* (Bruckner *et al*., 2004). This is illustrated in Appendix 9. For instance, the entire region spr0102 to spr0119 was present in strains ATCC700904 and 07-2839 but appeared almost entirely lacking hybridization for strain 05-1271. The other test isolates demonstrated many different combinations of hybridization or absent hybridization for these genes.

For the region spr0311 to spr0323, only genes from the smaller cluster spr0320 to spr0323 were found to hybridize. Again this could be the entire section (ATCC51916 and 05-1271) or only smaller sections (07-2839, ATCC700904 and BAA-659) or there could be a total lack of hybridization for the entire region (BAA-340, BAA-660, 05-2565, 05-1821 and OXC141).

Similar variable hybridization was seen at the regions of diversity spr0955 to spr0971, spr1403 to spr1404 and spr1618 to spr1621. Only isolate 07-2839 showed hybridization of any genes within the region of diversity spr1184 to spr1198.

In isolate ATCC700904 there appears to be a small region of diversity at spr1549 to spr1550 with hybridization occurring at these probes for this isolate alone. Both these genes code for hypothetical proteins although spr1550 has sequence similarity to a *MutR* protein. Spr1549 is 393bp in size and spr1550 is 852bp in size. These were noted as highly variable genes coding for a putative regulatory protein by Bruckner *et al* (Bruckner *et al*., 2004).

For these regions of diversity in the R6 genome, it would be of interest to perform CGH of DNA from the test strains against R6 DNA as competitive hybridization should then be seen for those genes highlighted in red above in Appendix 9.

Comparative hybridization of TIGR4 DNA against test isolate DNA has elucidated 4 regions of diversity which do not feature in the list of regions of diversity based on the TIGR4 genome identified by Silva *et al* (Silva *et al*., 2006) which are shown in Chapter 1 Table 1-2. The first is a 6.3kb region from SP0303 to SP0311 which mainly relates to components of a cellobiose phosphotransferase system. This region features as cluster 2 identified by Bruckner *et al* (Bruckner *et al*., 2004). A second region of 5.7kb from SP0726 to SP0730 was identified in isolate ATCC700904 which showed no hybridization, despite hybridizing well for all other strains (Figure 4-2) This was unexpected as the isolate ATCC700904 features in the work of Silva *et al* where it was named PMEN13 and this region was not identified despite using the same isolate on the same version of microarray with similar hybridization methodologies (Silva *et al*., 2006). This can be explained as the data analysis methods chosen (such as the normalization methods) are different between the two studies. Several genes within this region (SP0726, SP0728 and SP0729) have been associated with virulence in a serotype 4 mouse bacteraemic pneumonia model (Hava and Camilli, 2002).

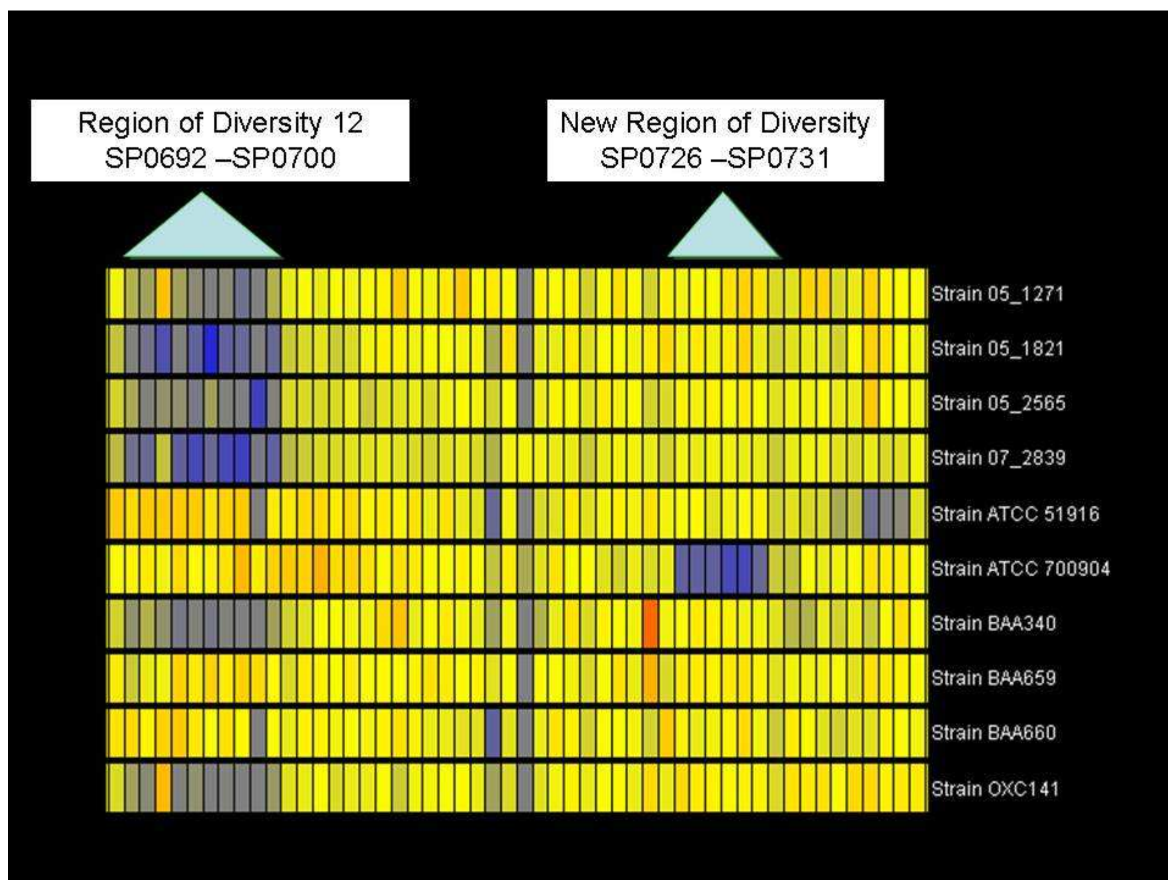


Figure 4-2 Demonstration of a new region of diversity (SP0726 – SP0731) in the TIGR4 genome using Genespring GX 7.3.1.

Each coloured bar represents a consecutive gene in the TIGR4 genome. Shades of yellow and orange indicate that hybridization occurred with DNA from both TIGR4 and the test isolate while shades of blue indicate no hybridization occurred with DNA from the test strain but unopposed hybridization by TIGR4 DNA had occurred.

A third shorter region of 5.1kb relates to a putative Type II restriction endonuclease which was identified as cluster 9 by Bruckner *et al* (Bruckner *et al*., 2004). It was also identified as present in serotype 6A isolates from cases of invasive pneumococcal disease but not serotypes 6B or 14 by Obert *et al* (Obert *et al*., 2006) and was not identified by Silva *et al* which again is likely to be due to the different data analysis methods utilised as Silva *et al* had also studied isolate BAA-659 which was named PMEN23 in their analysis (Silva *et al*., 2006).

The fourth region of diversity identified by this study which was not identified by Silva *et al* is a 2.9kb region found between SP2180 and SP2183 which was identified as not hybridizing in isolate 07-2839 from Bolivia. This is illustrated in Figure 4-3 below.

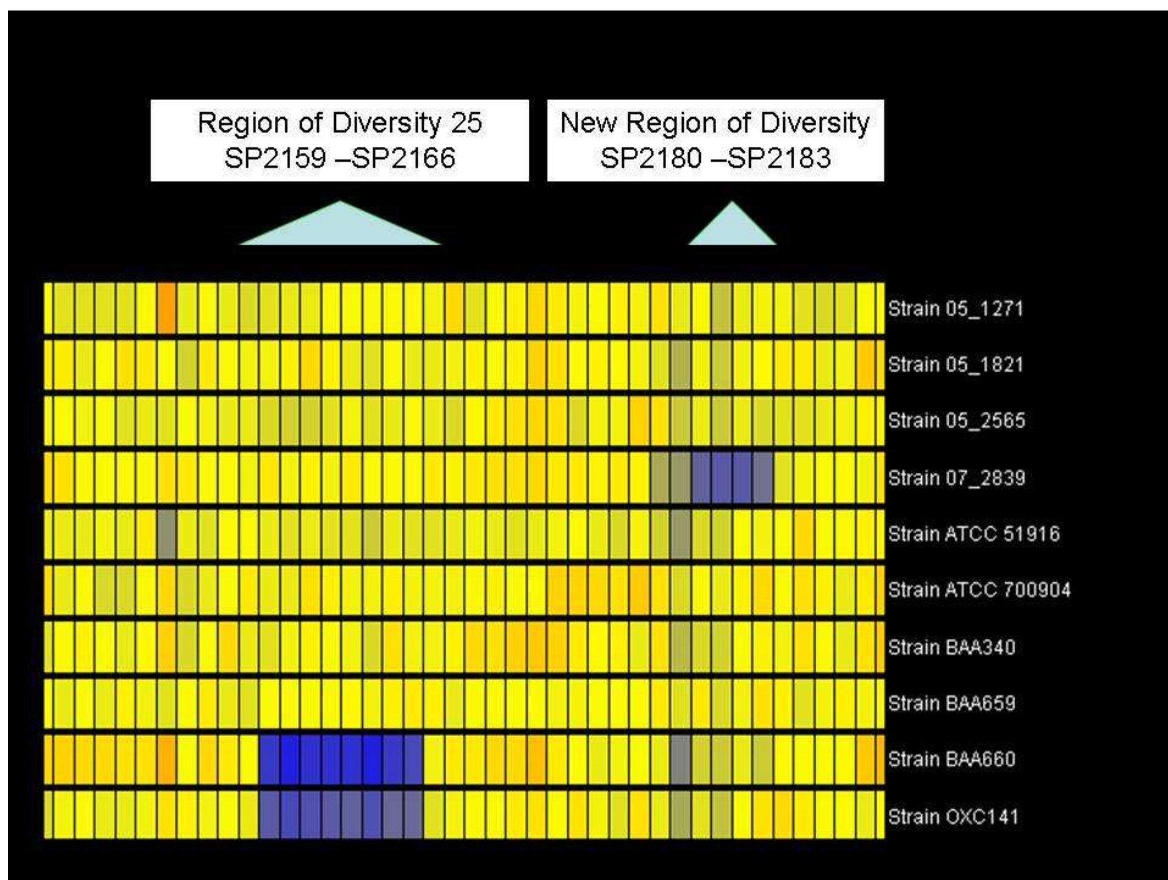


Figure 4-3 Demonstration of a new region of diversity (SP2180 – SP2183) in the TIGR4 genome using Genespring GX 7.3.1.

Each coloured bar represents a consecutive gene in the TIGR4 genome. Shades of yellow and orange indicate that hybridization occurred with DNA from both TIGR4 and the test isolate while shades of blue indicate no hybridization occurred with DNA from the test strain but unopposed hybridization by TIGR4 DNA had occurred.

The gene SP2179 falling immediately before this region was identified by Bruckner *et al* as a highly variable gene related to insertion sequence IS1380. The genes SP2180 and SP2181 have been previously identified as pseudogenes, particularly as SP2180 is interrupted by the IS1380 Spn1 element. However, gene expression studies to be described later in Chapter 10 using serotype 3 and serotype 1 pneumococci show SP2180, SP2182 and SP2183 to be expressed in both TIGR4 and the test isolates while SP2181 was not expressed in any of these expression experiments. SP2181 was expressed though by the strain South Africa 2507 but only in response to the addition of subtherapeutic clarithromycin (Chapter 11). Evidence of their expression under certain conditions does suggest that these genes may be functional.

Bruckner *et al* identified some regions which were not present in some of their battery of international test strains but were evident in TIGR4 and R6 (Bruckner *et al.*, 2004). SP0505 to SP0508 is such a region. However, this region appears to be present in all the 10

strains tested in this chapter. Bruckner *et al* also identify the region SP1309 - SP1337 as a, "hot spot," for recombination (Bruckner *et al*., 2004). These results also concur with this observation as this is a region of substantial diversity between strains as illustrated below in Figure 4-4. This region is within a slightly larger region of diversity extending from SP1309 - SP1347 which is part of the region of diversity 17 described by Silva *et al* and clusters 9 and 10 described by Bruckner *et al* (Bruckner *et al*., 2004).

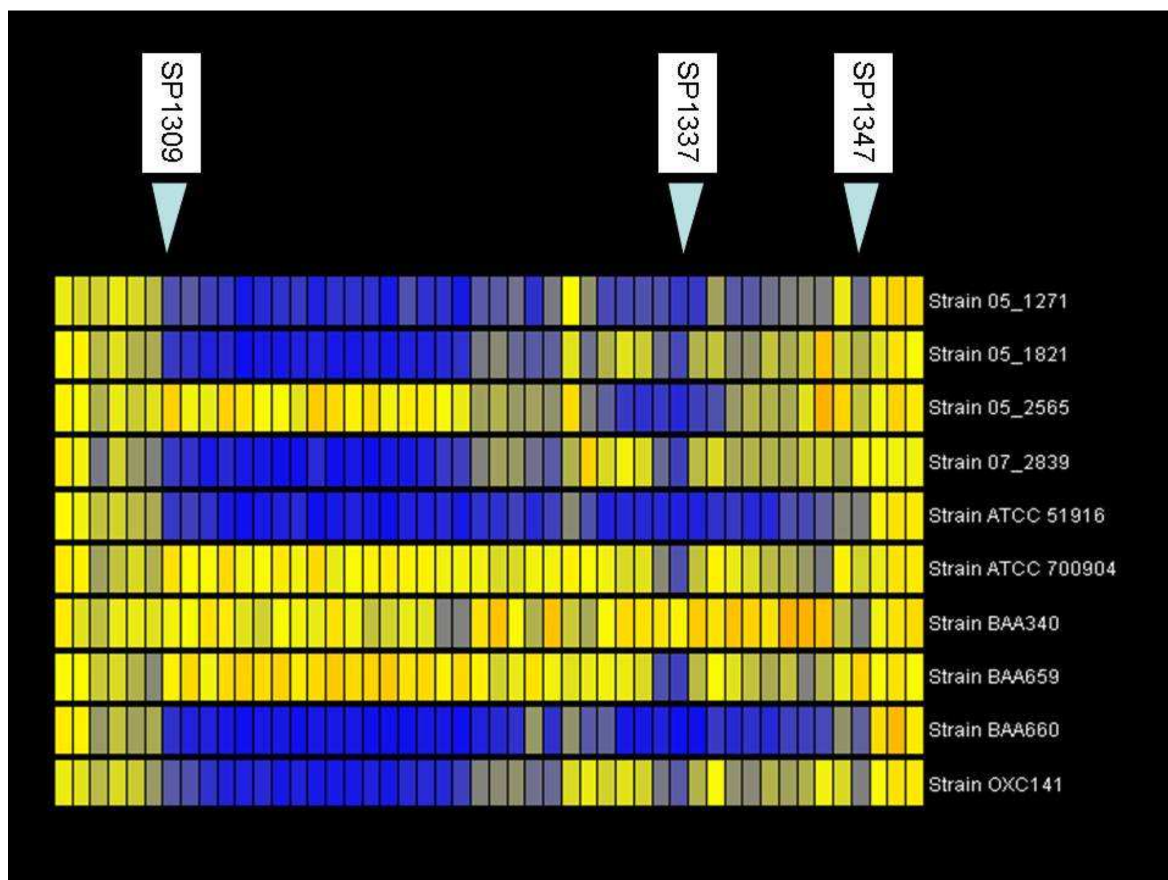


Figure 4-4 Demonstration of a, "hot spot," for recombination events (SP1309 - SP1337) in the TIGR4 genome using Genespring GX 7.3.1.

Each coloured bar represents a consecutive gene in the TIGR4 genome. Shades of yellow and orange indicate that hybridization occurred with DNA from both TIGR4 and the test isolate while shades of blue indicate no hybridization occurred with DNA from the test strain but unopposed hybridization by TIGR4 DNA had occurred.

4.3.2 Diversity at Particular Genetic Loci

4.3.2.1 Genes with variable homologs

Bruckner *et al*, identified several genes in the TIGR4 and R6 genomes which demonstrated highly variable homologs (Bruckner *et al*., 2004). These are illustrated in Table 4-2 along with results from the CGH of the above 10 isolates. These results suggest that either there is less variability within the genes of these test isolates or that cross hybridization has occurred between variable homologs and probes on this microarray or that the "spotted" microarray constructed by BμG@S identifies fewer variable genes than the oligonucleotide microarray used by Bruckner *et al* (Bruckner *et al*., 2004).

Gene	Putative Function	Hybridization Result in this Study
SP0071	Zinc metalloproteinase C	N
SP0147	Hypothetical protein	P
SP0257	Hypothetical protein	V
SP0298	Hypothetical protein	V
SP0328	IS1380	V
SP0332	Hypothetical protein	P
SP0378		P
SP0379		V
SP0380		P
SP0432	IS1167	P
SP0495	IS1380	V
SP0642	IS66	P
SP0643	IS66	P
SP0644	IS66	V
SP0666	Hypothetical protein	P
SP0714	IS1380	V
SP0813	IS30	P
SP0814	IS30	P
SP0826	Hypothetical protein	V
SP0836	IS1167	P
SP0949	IS1515	V
SP1015	IS1167	P
SP1188	Hypothetical protein	V
SP1189	Hypothetical protein	V
SP1262	IS1167	P
SP1292	SAP domain protein	P
SP1352	IS1380	P
SP1418	IS1380	P
SP1439	IS1380	V
SP1444	IS1380	P
SP1503	IS1380	V
SP1582	IS1167	P
SP1595	IS1380	V
SP1639	IS1167	P
SP1692	IS1167	P
SP1772	Cell Wall Anchor Protein	V
SP1927	IS1381	P
SP2093	Hypothetical protein	P
SP2179	IS1380	V
Spr0041	IS1178	Not represented on SPv1.1 microarray
Spr0042	IS1178	Not represented on SPv1.1 microarray
Spr0703	Putative sortase	V
Spr0986	IS1167 truncated	Not represented on SPv1.1 microarray
Spr0987	IS1167 truncated	Not represented on SPv1.1 microarray
Spr0988	IS1167 truncated	Not represented on SPv1.1 microarray
Spr1093	Hypothetical protein	N
Spr1403	Cell Wall Anchor Protein	V
Spr1484	Hypothetical protein	V
Spr1536	Neuraminidase A	Not represented on SPv1.1 microarray
Spr1549	Putative regulatory protein	V
Spr1550	Putative regulatory protein	V
Spr1675	IS1167	Not represented on SPv1.1 microarray
Spr1676	IS1167	Not represented on SPv1.1 microarray
Spr1716	IS1167	Not represented on SPv1.1 microarray

Table 4-2 Highly variable genes with multiple homologs compared with CGH results. P = gene hybridized with all test strains; N = gene hybridized with none of the test strains; V = gene variably hybridizing in test strains.

4.3.2.2 Diversity of zinc metalloproteinases SP1154, SP0664 and SP0071

In 1998, Poulsen *et al* identified a high degree of diversity of immunoglobulin A1 (IgA1) protease genes (SP1154) within the pneumococcus and other closely related streptococcal species (Poulsen *et al.*, 1998). This was due to repeat structures which varied in number and sequence which was thought to be related to immunogenicity and recognition of antigenic diversity (Poulsen *et al.*, 1998). IgA1 protease is a zinc metalloproteinase (zmp) and it has subsequently been documented that the pneumococcus has three zmps and that the other two, *ZmpB* (SP0664) and *ZmpC* (SP0071), show much sequence diversity also (Oggioni *et al.*, 2003, Chiavolini *et al.*, 2003). Whether these genes are absent, present or duplicated in the pneumococcal genome is also variable. SP0071 is absent from the R6 genome while SP1154 occurs twice in the genome of the sequenced strain G54 (Oggioni *et al.*, 2003). R6 and other pneumococcal genomes do still have a functional IgA1 protease despite SP0071 being absent as discussed in Chapter 3. Chiavolini *et al* identified significant sequence diversity of SP1154 between isolates of serotypes 2, 4, 6, 19F and 23F and of SP0664 for serotypes 4 and 23F (Chiavolini *et al.*, 2003). The results above demonstrate diversity as to whether hybridization occurred or not for SP1154 and SP0071. Where hybridization did not occur could be the result of the gene being absent or it could be present but with such a degree of sequence diversity that it does not hybridize to the probe sequence on the microarray which is based on the gene sequence in TIGR4. IgA1 protease is believed to be present in all pneumococcal strains and serotypes (De Paolis *et al.*, 2007). Oggioni *et al* also suggest that the presence of SP0071 may relate to either serotype or the body fluid of origin of the isolate. From a variety of clinical isolates they find SP0071 in serotypes 4, 6, 8, 9 and 19F. Over 50% (6/11) of sputum isolates had a *zmpC* homolog and only 1 CSF isolate. The CGH data from this study demonstrated hybridization of SP0071 only from DNA from a serotype 23F isolate with no hybridization from serotypes 3, 6A, 9A, 9V, 12F, 14, 20 or 35B which suggests that the relationship between presence of SP0071 with serotype is not as straight forward as proposed by Oggioni *et al* (Oggioni *et al.*, 2003). The issue of a possible relationship between body fluid of origin and presence of SP0071 could not be considered with this series of isolates as the clinical background for the majority of PMEN clones used is not known. This will be considered subsequently using CGH data from more adequately catalogued clinical isolates.

4.3.2.3 Diversity of SP1051 and Pneumococcal Pathogenicity Island 1

SP1051 is a virulence associated gene present within the pneumococcal pathogenicity island 1 which by PCR has been found to be absent from isolates of serotype 1, 6B, 17 and 35F (Brown *et al*., 2004). These serotypes did not feature in our chosen isolates for this study of diverse isolates and the CGH results for these isolates showed SP1051 to be present in all 10 test strains, consistent with the results of Brown *et al* (Brown *et al*., 2004).

4.3.2.4 Diversity of Pneumococcal Pilus Genes

The region of diversity at SP0460 to SP0470 is related to a pneumococcal pilus (SP0461-SP0468) and to diversity of choline binding proteins. The whole region was present in 3 of the isolates examined (serotypes 6A, 35B and 9A) and the region SP0460 to SP0468 failed to hybridize in its entirety for the other 7 isolates tested. This is consistent with the observations of Bruckner *et al* at this region (Bruckner *et al*., 2004) and Sjostrom *et al*, who did not identify any consistent relationship between the presence or absence of this region by CGH and serotypes 9V, 19F or 14 (Sjostrom *et al*., 2006).

4.3.2.5 Diversity of Choline binding proteins

It has been observed that choline binding protein A (SP2190) which is necessary for cell invasion is not expressed in 25% of clinical strains (Orihuela *et al*., 2003). A lack of hybridization was not identified for any of the tested isolates for SP2190 so, if this gene were not to be expressed in these isolates, it would not be due to its absence in the genome of the isolate.

4.3.2.6 Diversity of Neuraminidase Genes

Recently it has been described that the presence of neuraminidase C (*nanC*, SP1326) in clinical isolates is more variable than is the case for neuraminidase A or B (Pettigrew *et al*., 2006). In the isolates examined here 40% showed *nanC* to be present which is close to the 51% of isolates where it was identified by Pettigrew *et al* (Pettigrew *et al*., 2006). Again though, this phenomenon will be better investigated using a series of clinical isolates where the background to the isolates is more fully known.

4.4 Conclusions

These DNA CGH results relating to diverse isolates demonstrate substantial genomic diversity for pneumococcal isolates of diverse multilocus sequence types and serotypes and are consistent with the results of similar previous studies. In fact this work has shown such diversity to be greater than previously considered in CGH studies with two new regions of diversity identified. If a wider range of diverse isolates were assessed, it is probable that more new regions of diversity would be identified. It does appear though that identification of these regions is dependant on the data analysis methods which are chosen. These results also highlight a need for greater consensus as to what does and what does not comprise a region of diversity in the pneumococcal genome. Associations between CGH results and individual serotypes, sequence types and the clinical background of strains will be better achieved by more focussed studies of banks of particular sequence types and serotypes where the clinical background to strains has been adequately documented.

5 Genomic Diversity in a Multilocus Sequence Type Associated with Invasive Pneumococcal Disease (IPD)

5.1 Serotype 14 and ST9: associations with IPD

The significance of the ST9 clonal complex is reviewed in Chapter 1, along with its association with serotype 14 pneumococci. These studies of DNA CGH from 10 ST9 isolates were planned to assess whether genomic diversity could be identified within a clonal complex and also to compare the degree of diversity within isolates of the same serotype within that clonal complex compared to isolates of different serotypes within the clonal complex.

5.2 Choice of Isolates

Details of the 10 chosen ST9 isolates are outlined in Appendix 1. Nine of these were from the strain collection at SMPRL and were chosen to try to maximise observations on diversity by taking isolates from different years, with different antibiotic sensitivity profiles, source cultures from different body fluids, sourced from different geographical regions in Scotland and from different ages of patients. Even so, only one isolate originated from a cerebrospinal fluid specimen (01-5710). Five isolates were chosen of serotype 14 to assess whether overall genomic diversity is different in phenotypically similar isolates within a clonal complex. These five included four isolates from Scotland (00-1724, 01-5710, 02-2445 and 04-1870) and the Sanger sequenced strain INV200 which represents the PMEN clone England ^{14}-9 (Hall *et al*., 1996) – a clone which also infrequently causes IPD in Scotland (Smith *et al*., 2006). All 5 isolates are resistant to erythromycin consistent with previous observations that 95% of ST9 isolates in Scotland are erythromycin resistant (Clarke *et al*., 2004b) usually resulting from expression of the *mef*A gene and macrolide efflux – the M phenotype (Amezaga *et al*., 2002). The five isolates which were not of serotype 14 were of serotypes 8 (03-2105), 18 (02-1309), 19A (03-1051), 19F (04-1548) and 23F (04-1168).

5.3 Microarray Results

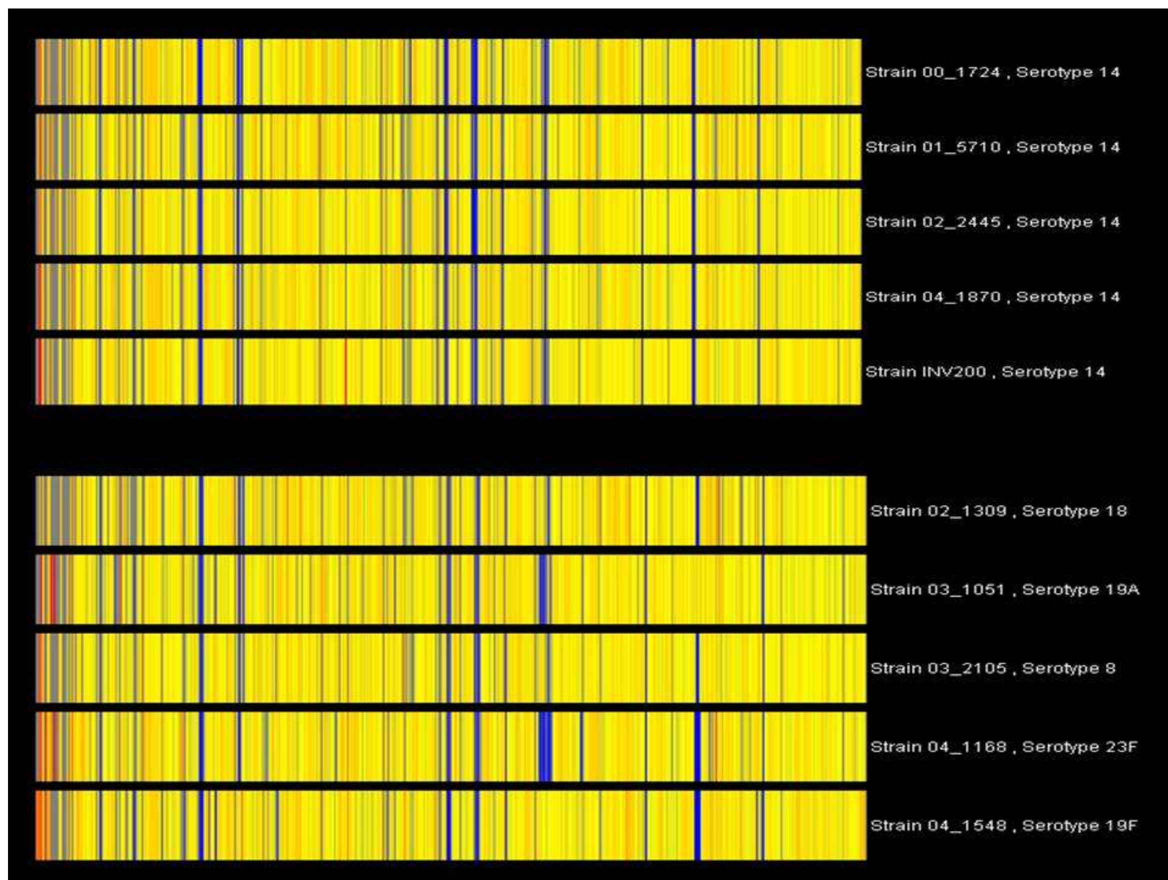


Figure 5-1 Comparison of diversity within genomes of serotype 14 associated ST9 isolates and non serotype 14 ST9 isolates using Genespring GX 7.3.1.

Shades of yellow and orange indicate that hybridization occurred with DNA from both TIGR4 and the test isolate while shades of blue indicate no hybridization occurred with DNA from the test strain but unopposed hybridization by TIGR4 DNA had occurred. Grey indicates that neither hybridization of TIGR4 or test isolate DNA occurred or the hybridization fluorescence was of such low intensity that it was excluded from analysis. Red indicates unopposed hybridization of test isolate DNA to probes representing genes from the R6 genome. The five serotype 14 isolates demonstrate less diversity than the five ST9 isolates of different serotypes. The list of genes from the TGR4 genome which are present or non-hybridizing in the ST9 strains are displayed in Appendix 10 and the list of genes from the R6 genome which are also present or non-hybridizing in the ST9 strains are in Appendix 11.

5.4 Discussion

5.4.1 Overall diversity in ST9

It has been recognised that serotype 14 strains of different genetic backgrounds have different propensities to cause pneumonia in mice and these differences are attributed to virulence factors other than the capsule (Mizrachi-Nebenzahl *et al*., 2004). These CGH studies of ST9 pneumococci (and a subgroup of serotype 14, ST9 pneumococci) allow

some investigation as to what some of these genetic differences between phenotypically identical (serotype 14) pneumococci may be and whether they relate to known virulence determinants.

Obert *et al* have performed CGH studies on serotype 14 pneumococcal isolates (12 were from cases of IPD and 8 were from asymptomatic nasopharyngeal carriage) of unknown sequence type. They have correlated the presence of certain genes with either an invasive or carriage phenotype for serotype 14 (Obert *et al.*, 2006). As all the serotype 14 isolates used in this study of ST9 isolates have an invasive phenotype, it is possible to assess whether the CGH results obtained fit with the phenotype expected from the results of Obert *et al*. Unfortunately, the majority of the genes identified by Obert *et al* as present only in serotype 14 isolates with the carriage phenotype can be found in the genomes of the ST9 serotype 14 isolates with invasive phenotypes. These misclassified genes are SP0029, SP0093, SP0096-SP0098, SP0117, SP0141, SP0173-SP0182, SP0186, SP0205, SP0256, SP0259, SP0288, SP0328, SP0336, SP0343, SP0398-SP0399, SP0495, SP0531-SP0532, SP0584, SP0617, SP0711, SP0714, SP0738-SP0740, SP0826, SP0875-SP0876, SP0877, SP0907, SP0949, SP1019, SP1065, SP1154, SP1172, SP1773, SP1175, SP1185, SP1198, SP1251, SP1307, SP1308, SP1337, SP1348-SP1349, SP1351-1352, SP1366, SP1418, SP1439, SP1441, SP1503, SP1595, SP1642, SP1672-SP1673, SP1677, SP1693, SP1719, SP1722, SP1902, SP1947, SP2003, SP2005, SP2158-SP2166, SP2179. It is likely that the results of Obert *et al* have been seriously confounded by not assessing a diverse enough sample of serotype 14 isolates from different clonal complexes or by assessing too small a sample of isolates.

Some of the genes identified by Obert *et al* in serotype 14 pneumococci as present only in isolates of invasive phenotype also appeared in the invasive serotype 14 ST9 isolates used in this chapter (shown in Appendix 10) but not ST9 isolates which were not serotype 14. These genes were SP0389-SP0397, SP0536, SP0540-SP0542, SP0607-SP0608, SP0627-SP0628, SP0691-SP0700, SP0982, SP1126, SP1315-SP1321, SP1324, SP1326-SP1331, SP1556, SP1612-SP1613, SP1679, SP1740, SP1741, SP1755-SP1759, SP1810, SP1911, SP2140 and SP2232. It is possible then that these genuinely are only associated with an invasive phenotype.

Some of the genes associated with carriage in serotype 14 are not seen in these invasive ST9 CGH results. These genes are SP1062-SP1064, SP1147, SP1334, SP1335, SP1338, SP1340-SP1345, SP1350 and SP1622. However as CGH was not performed on carriage ST9 isolates the validity of this association cannot be assessed from this data.

5.4.1.1 Macrolide Resistance

Pneumococcal genomic diversity within the ST9 complex has been alluded to before with regard to the antibiotic resistance genes which can be found in isolates from Scotland belonging to this clone (Amezaga *et al*., 2002). This is particularly the case regarding macrolide efflux. Bruckner *et al* associate the region of diversity SP0163 to SP0171 with macrolide efflux (particularly SP0168) and so it is not surprising that this has been a variable region detected by CGH in these ST9 isolates. SP0168 appeared not to hybridize in any of the isolates regardless of whether there was a macrolide resistant or sensitive phenotype. SP1110, another gene associated with a macrolide efflux pump function, appears to be present in all the isolates. The lack of hybridization of SP0168 could be due to the absence of the gene or presence of a divergent sequence which could potentially be altering its function as well as its ability to hybridize to the microarray probe.

Unfortunately the macrolide resistance gene *mef*A is not represented among the probes on the SPv1.1 microarray and so it has not been possible to assess for the presence of this gene by microarray CGH. A *mef*A sequence has been associated with a transposon Tn1207.1 in pneumococci (Santagati *et al*., 2000) and a macrolide efflux genetic assembly (MEGA) has also been described in pneumococci on a chromosomal insertion element which has more than four different insertion sites in the pneumococcal genome (Gay and Stephens, 2001). Macrolide resistance in pneumococci is also conferred by the *ermB* gene which is carried on transposons Tn1545 (Clewell *et al*., 1995, Courvalin and Carlier, 1987), Tn917 (McDougal *et al*., 1998) or Tn3872 (McDougal *et al*., 1998). These represent a potentially sizeable source of genetic diversity in the ST9 complex, with a direct effect on phenotype, which cannot be detected by CGH using the SPv1.1 microarray.

5.4.1.2 Glyceraldehyde-3-phosphate dehydrogenase (GAPDH)

Proteomic analysis has determined that there is an increase in expression of GAPDH in M phenotype pneumococci (Cash *et al*., 1999). Cash *et al*, identified three different forms of GAPDH. The TIGR4 genome contains two genes which code for GAPDH – SP1119 and SP2012. Both genes were identified by CGH as present in all five of the serotype 14, ST9 isolates tested. Microarray based expression studies of these isolates may help to elucidate the potential role of these two genes in the increased generation of GAPDH in M phenotype pneumococci.

5.4.2 Regions of Diversity in the TIGR4 and R6 genomes

The CGH results of the ST9 isolates have demonstrated a region of diversity within the isolate 04-1548 where hybridization of a 1.7kb region encompassing the TIGR4 genes SP1933 to SP1936 does not occur. This region has not been previously identified in the CGH studies of Bruckner *et al* (Bruckner *et al*., 2004) (who did not perform CGH on any serotype 14 isolates) or Silva *et al* (Silva *et al*., 2006) or Obert *et al* (Obert *et al*., 2006) (both of whom included several serotype 14 isolates in their experiments). SP1933, SP1934 and SP1935 code for hypothetical proteins and SP1936 codes for a putative Type II restriction modification system regulatory protein. This is illustrated below.

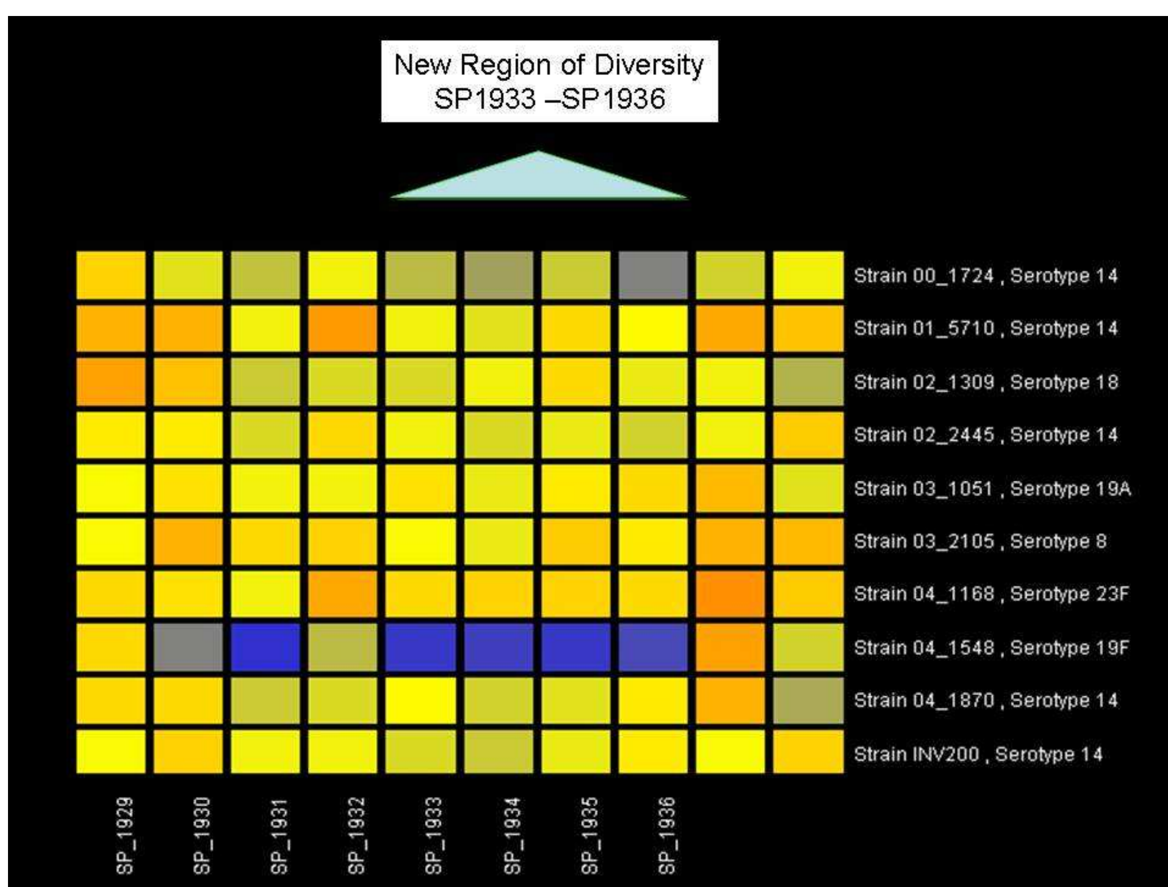


Figure 5-2 Location of a new region of diversity SP1933 – SP1936 identified in isolate 04-1548.

5.4.3 Diversity at Particular Genetic Loci

For illustrative purposes, the diversity observed at various regions known to be variable will be considered.

5.4.3.1 Genes with variable homologs

The presence or absence of hybridization for forty five genes identified by Bruckner *et al* (Table 4-2) as being highly variable in the TIGR4 and R6 genome could be assessed using the SPv1.1 microarray. When all 10 isolates of serotypes 8, 18, 14, 19A, 19F and 23F were considered, twenty seven genes (60%) were identified as having variable hybridization but if only serotype 14 isolates were considered this fell to only 5 genes (11%) having variable hybridization which demonstrates that there is substantially less diversity within isolates of the same serotype within a MLST clonal complex than between isolates of different serotypes of the same sequence type.

5.4.3.2 Diversity of zinc metalloproteinases SP0664, SP0071 and SP1154

The diversity of zinc metalloproteinases was discussed in Chapter 4. The CGH of ST9 isolates has allowed investigation of how much diversity occurs within a clonal complex. SP0071 (IgA1 protease) could not be identified by hybridization in the genomes of any of the tested ST9 isolates (although hybridization by TIGR4 DNA occurred in all the CGH experiments). The R6 gene coding for IgA1 protease, spr1042, was present only in the ST9, serotype 8 isolate 03-2105 which suggests that the other strains may have unique gene sequences which fulfil this function. As these isolates are from predominantly blood cultures with only one isolate from cerebrospinal fluid, it is possible to conclude that SP0071 may be infrequently found in blood or CSF isolates which is compatible with the observation of Oggioni *et al* (Oggioni *et al*., 2003). SP0664 was identified as present in all the ST9 isolates while SP1154 was variably present or not hybridizing in the test ST9 isolates.

5.4.3.3 Diversity of SP1051 and Pneumococcal Pathogenicity Island 1

As was the case with the CGH results of isolates of diverse serotype and sequence type (Chapter 4), these results of ST9 pneumococci from serotypes 8, 14, 18, 19A, 19F and 23F demonstrated that SP1051 was present in all the tested isolates which is again consistent with the findings of Brown *et al* (Brown *et al*., 2004).

5.4.3.4 Diversity of Pneumococcal Pilus genes

The *rrgC* gene (SP0464) has been used as a marker for the presence or absence of a pneumococcal pilus (Basset *et al*., 2007) where it has been identified as present in less than

10% of serotype 14 isolates (Basset *et al.*, 2007). In the CGH studies of Sjostrom *et al* which focussed particularly on the rlrA pilus islet (SP0461-SP0468) and which included three serotype 14 strains (two which were ST156 and one which was ST709), *rrgC* (SP0464) was present in both ST156 isolates but noted as absent in the ST709, serotype 14 isolate (Sjostrom *et al.*, 2006). In this series of ST9 isolates, we identified *rrgC* (SP0464) only in isolate 04-1548 (which is serotype 19F) and it failed to hybridize in any of the serotype 14 isolates consistent with the findings of Basset *et al* (Basset *et al.*, 2007).

5.4.3.5Diversity of Choline binding proteins

As was the case with the isolates in Chapter 4, a lack of hybridization was not identified for any of the tested ST9 isolates for choline binding protein A (SP2190) so, if this gene were not to be expressed in these isolates as was found in 25% of clinical pneumococcal isolates by Orihuela *et al* (Orihuela *et al.*, 2003), it would not be due to its absence from the genome of the isolate.

5.4.3.6Diversity of Neuraminidase Genes

In the isolates examined here 80% showed *nanC* (SP1326) to be present. It was present in all the serotype 14 isolates but did not hybridize in isolates which were serotype 19A and 23F. Pettigrew *et al* have made an association between the presence of *nanC* and CSF isolates (Pettigrew *et al.*, 2006). These CGH results are compatable with this association as *nanC* was identified in the CSF isolate which was tested (01-5710).

6 Genomic Diversity within a Multilocus Sequence Type Accounting for Invasive Pneumococcal Disease and Carriage

6.1 A Review of the Associations of Serotype 3 and ST180 with Pneumococcal Carriage or Invasive Disease

The dichotomy that serotype 3 pneumococci can cause disease with a high associated mortality in some individuals while being harmlessly carried in the nasopharynx of others has been recognised since the early 20th century (Blake, 1931). An association between serotype 3 pneumococci causing disease more commonly in the elderly than in children is also an established observation (Blake, 1931, Cecil *et al*., 1927) which remains true in several countries (Inostroza *et al*., 2001, Rahav *et al*., 1997, Shapiro and Austrian, 1994, Kyaw *et al*., 2003, Bescos *et al*., 2003, Martin and Brett, 1996). In children from Oxfordshire, a serotype 3 associated clone (ST180) had an odds ratio of invasiveness of only 0.1 and so was more associated with asymptomatic carriage (Brueggemann *et al*., 2003) although it is inappropriate to extrapolate this finding into an adult population as the relative risk of invasive disease due to serotype 3 increases over middle age (Scott *et al*., 1996). According to the MLST database[8] although predominantly serotype 3 associated, ST180 has also been associated with a serotype 19F capsule in Germany and non-typeable isolates in South Korea.

The ability to identify genes present in pathogenic invasive bacteria which are not present in asymptomatically carried commensals or define the genetic relationships between virulent and non-virulent clones of bacteria (in order to identify potential targets for therapeutic intervention) has long been a goal of bacterial comparative genomics (Hollingshead and Briles, 2001, Joyce *et al*., 2002). Molecular typing such as IS*1167-boxA* genotype fingerprinting demonstrated that carriage isolates belonged to patterns unique to carriage suggesting that clones associated with carriage may be distinguished from those associated with invasion (Robinson *et al*., 2001). It has also been suggested that more virulent clones of invasive pneumococci tend to have evolved from less virulent clones associated with carriage (Robinson *et al*., 2002).

---

[8] http://spneumoniae.mlst.net/ {accessed 6th December 2008}

In this chapter, the DNA of serotype 3, ST180 carriage and invasive disease associated isolates was used in microarray CGH experiments to investigate whether any genes are consistently present in carriage associated isolates and not invasive isolates or *vice versa*. An assessment is then made as to whether the patterns of gene expression under predetermined conditions for these isolates can be used to determine associations between carriage and invasive associated isolates.

6.2 Serotype 3 ST180 Isolates Analysed by Microarray CGH

Details of the isolates used in these CGH experiments are tabulated below:

Isolate	Phenotype	Clinical Details
OXC141	Carriage	Isolate from a carriage study in Oxford, United Kingdom.
03-4183	Carriage	Isolate from a carriage study from The Netherlands.
03-4185	Carriage	Isolate from a carriage study from The Netherlands.
03-4155	Carriage	Isolate from a carriage study from The Netherlands.
03-4156	Carriage	Isolate from a carriage study from The Netherlands.
07-2838	Carriage	Nasopharyngeal swab isolate obtained from a 9 year old girl in Trinidad, Bolivia during June 2007.
99-4038	Invasive	Blood culture isolate from a 31 year old man from Dundee, United Kingdom with meningitis.
99-4039	Invasive	Cerebrospinal fluid isolate from same 31 year old man from Dundee, United Kingdom with meningitis taken on same day as 99-4038.
00-3946	Invasive	Cerebral pus isolate from a 60 year old man from Edinburgh, United Kingdom.
06-1705	Invasive	Cerebral pus isolate from a 15 year old girl from Glasgow, United Kingdom.
02-1198	Invasive	Umbilical cord blood isolate from a neonate who died within hours of delivery in Glasgow, United Kingdom. This isolate is named P49 in the paper by Silva *et al* (Silva *et al.*, 2006). N.b. Maternal and neonatal death within hours of delivery as a consequence of fulminant infection due to pneumococcal endometritis is a recognised albeit exceedingly uncommon event (Gransden *et al.*, 1985).

Table 6-1 Background to serotype 3, ST180 isolates used in the microarray CGH experiments.

6.3 DNA CGH Comparison of Carriage and Invasive Isolates of ST180

Gene lists generated from these CGH experiments are displayed in Appendix 12 (TIGR4 genes) and Appendix 13 (R6 genes). Figure 6-1 is generated by Genespring GX 7.3.1. and indicates that two distinct patterns of hybridization occur with these isolates. Figure 6-2 focuses on the region SP2159-SP2166 coding for a phosphotransferase system - the presence or non-hybridizing of these genes divides the ST180 isolates into two distinct hybridization patterns.

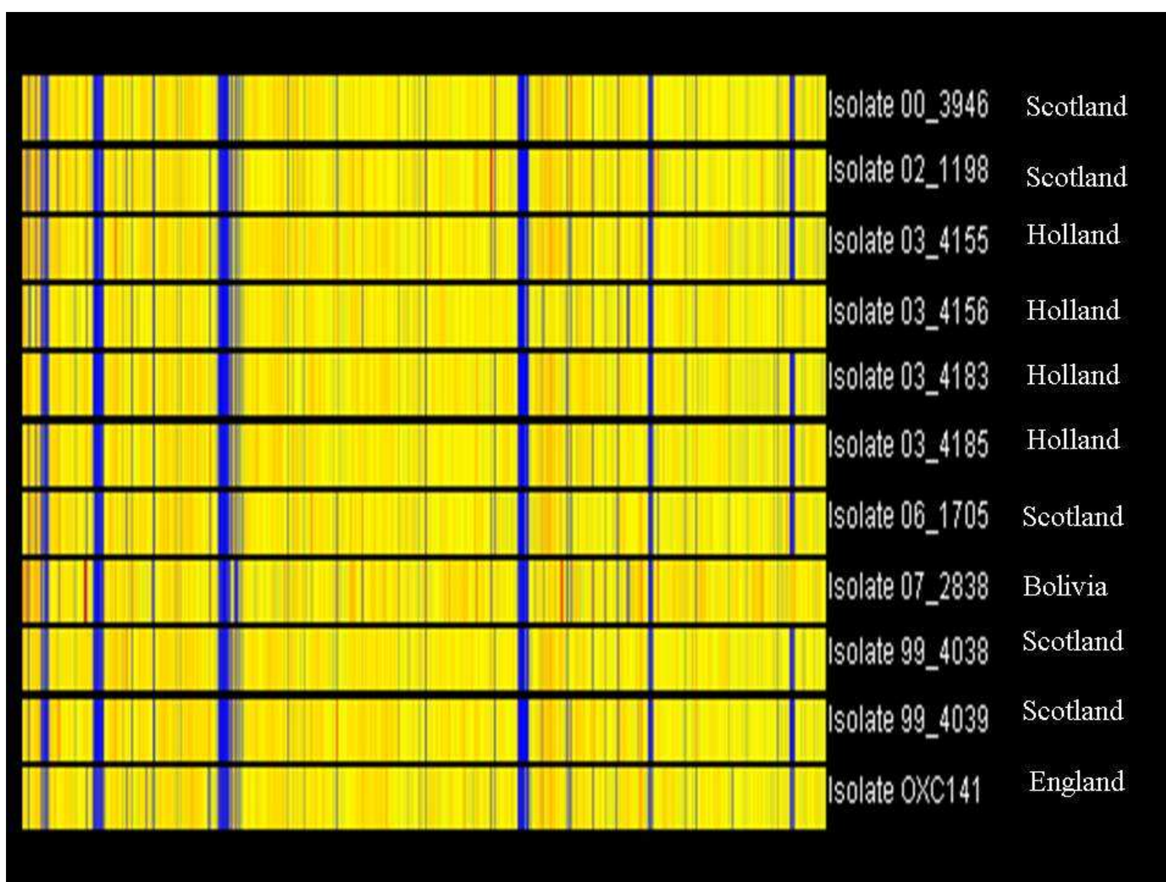


Figure 6-1 Comparison of diversity within genomes of serotype 3 associated ST180 invasive and carriage associated isolates using Genespring GX 7.3.1.

Each coloured bar represents a gene and these are aligned consecutively. Yellow indicates hybridization of DNA from both TIGR4 and the test strain. Blue indicates hybridization of DNA from TIGR4 but not the test strain and red indicates unopposed hybridization of test strain DNA to probes from the R6 genome. The names and countries of origin of the isolates are marked.

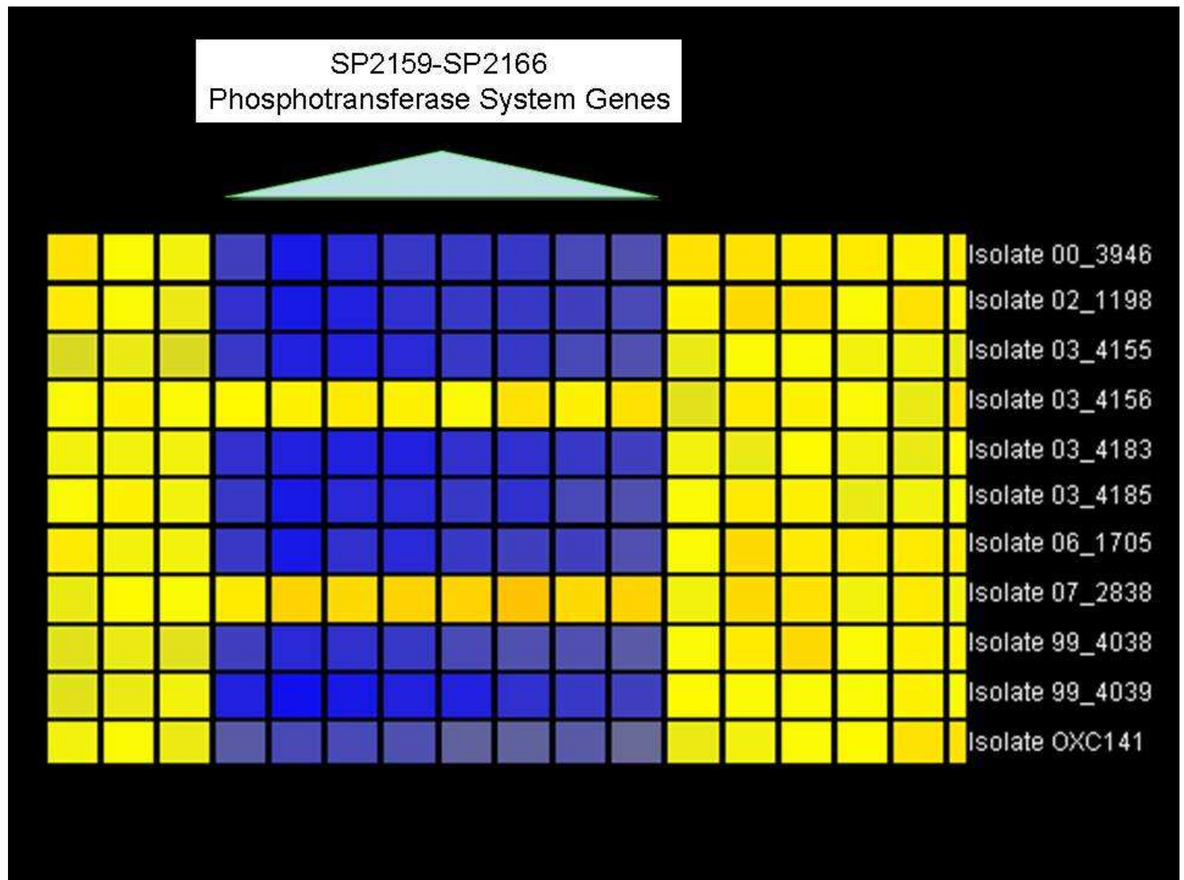


Figure 6-2 Serotype 3, ST180 microarray CGH results demonstrating the presence of the genes SP2159 to SP2166 (which code for phosphotransferase system genes) in the Dutch isolate 03-4156 and the Bolivian isolate 07-2838 which both have a carriage phenotype.

Each bar represents a pneumococcal gene. Yellow indicates hybridization of DNA from both TIGR4 and the test isolates. Blue indicates hybridization of DNA from TIGR4 but not the test isolate.

6.4 RNA Expression Differences in Carriage and Invasive Isolates of ST180

6.4.1 Microarray results

Genes Upregulated in IPD	Normalized Expression Level in Carriage Isolates	Normalized Expression Level in Invasive Isolates	Function
SP0576	0.9	22.9	Transcription antiterminator Lict
SP0640	1.0	4.5	Hypothetical protein
SP0877	1.0	4.0	Phosphotransferase system, fructose specific IIABC components
SP0875	1.0	3.3	Lactose phosphotransferase system repressor
SP0876	1.1	3.1	1-phosphofructokinase, putative
SP0358	1.0	2.7	Capsular polysaccharide biosynthesis protein Cps4J
SP0357	1.1	2.4	UDP-N-acetylglucosamine-2-epimerase
SP0248	1.0	2.3	Phosphotransferase system sugar-specific EIIA component
Genes Downregulated in IPD	**Normalized Expression Level in Carriage Isolates**	**Normalized Expression Level in Invasive Isolates**	**Function**
SP0110	0.5	0.1	ABC transporter membrane-spanning permease - amino acid transport
SP0111	0.7	0.2	Amino acid ABC transporter, ATP-binding protein, putative
SP0046	0.9	0.4	Amidophosphoribosyl transferase
SP1503	1.2	0.5	IS1380-Spn1, transposase
SP1418	1.3	0.5	IS1380-Spn1, transposase
SP1337	1.3	0.6	IS1380-Spn1, transposase
SP0050	1.6	0.6	Phosphoribosylaminoimidazolecarboxamide formyltransferase

Table 6-2 Genes identified by Genespring GX 7.3.1. whose expression detected by microarray are either significantly upregulated or downregulated in invasive pneumococcal disease.

(P<0.05 with Benjamini and Hochberg correction applied for multiple testing).

Gene	Predictive Strength for Invasive Disease	Function
SP0877	27.0	Fructose specific-phosphotransferase system IIBC component
SP0875	23.8	Lactose phosphotransferase system repressor
SP1474	20.9	Glycyl-tRNA synthetase beta chain
SP1592	16.6	Conserved domain protein
SP0830	16.6	Hypothetical protein
SP0066	16.1	Aldose-1-epimerase (mutarotase)
SP1097	14.9	Similar to GTP pyrophosphokinase
SP0876	14.7	Fructose-1-phosphate kinase
SP0833	14.7	Hypothetical protein
SP0831	14.7	Purine nucleoside phosphorylase (inosine phosphorylase)
SP1100	14.6	Phosphate acetyltransferase
SP0629	14.1	Conserved hypothetical protein
SP0828	12.8	Ribose-5-phosphate epimerase
SP0829	12.8	Phosphodeoxyribomutase
SP1701	12.7	Phospho-2-dehydro-3-deoxyheptonate aldolase
SP2014	12.7	IS630-Spn1, transposase Orf2
SP2096	12.7	Peptidase, M20/M25/M40 family
SP1563	12.7	Pyridine nucleotide-disulphide oxidoreductase family protein
SP0782	12.6	Conserved hypothetical protein
SP0712	12.0	Lactate oxidase, truncation
SP1417	12.0	Choline binding protein
SP2066	11.6	Threonine synthase
SP2192	11.6	Histidine kinase
SP0631	11.4	50S Ribosomal protein L1
SP0605	11.4	Fructose-bisphosphate aldolase
SP1523	11.4	SWF/SNF family ATP-dependent RNA helicase
SP0110	11.2	ABC transporter membrane-spanning permease - amino acid transport
SP1988	11.1	Immunity protein, putative
SP0834	11.1	Haemolysin-related protein
SP1193	11.1	Galactose-6-phosphate isomerase, LacA subunit
SP2002	11.1	Conserved hypothetical protein
SP1192	11.0	Galactose-6-phosphate isomerase LacB subunit
SP0715	10.9	Lactate oxidase
SP1099	10.9	Ribosomal large subunit pseudouridine synthase, RluD subfamily
SP0015	10.9	IS630-Spn1, transposase Orf1
SP0445	10.9	Acetolactate synthase large subunit
SP1098	10.9	Conserved hypothetical protein
SP1591	10.9	Proline dipeptidase
SP1337	10.8	IS1380-Spn1, transposase
SP0438	10.8	Glutamyl tRNA-Gln amidotransferase, subunit C

Table 6-3 Predictive strengths of the top 40 genes whose expression is more associated with IPD than carriage in serotype 3 ST180 isolates when grown to midlog at 37℃ in BHI.

Results generated using Genespring GX 7.3.1. gene predictor function.

6.4.2 Quantitative Real Time PCR results

The gene SP0110 which codes for a hypothetical protein and is identified as essential for virulence in a mouse pneumonia model (Hava and Camilli, 2002) was identified by Genespring GX 7.3.1. as significantly downregulated in invasive isolates (Table 6-2) and as being predictive of a serotype 3, ST180 isolate being invasive rather than carried asymptomatically (Table 6-3). However, in Appendix 12 SP0110 does not hybridize in all invasive or in all carriage isolates indicating either sequence divergence or absence of the gene. This seemingly contradictory phenomenon was investigated by qRT-PCR. No PCR product was generated for isolate 06-1705 confirming the CGH result. When the expression of SP0110 by 00-3946 (invasive phenotype) was compared to that by OXC141 (carriage phenotype), it was seen (Figure 6-3) that it was less expressed in the invasive isolate but this difference in expression did not reach statistical significance (p=0.547).

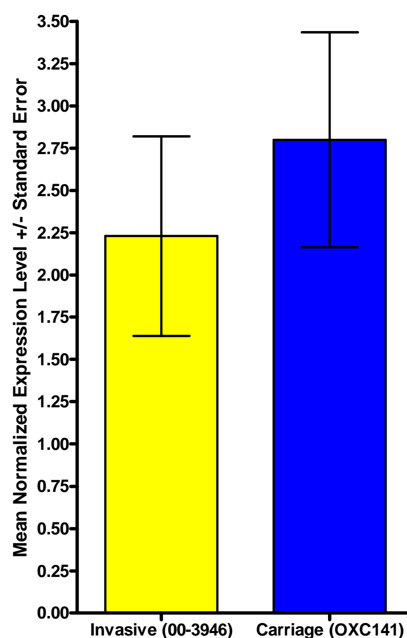


Figure 6-3 qRT-PCR results comparing expression of SP0110 in the invasive isolate 00-3946 and the carriage isolate OXC141.

The mean normalized expression was calculated using Q-gene (Muller *et al.*, 2002) from 3 replicates and the standard error of the means shown as error bars. Statistical comparison was made using an unpaired t-test (Muller *et al.*, 2002).

From this it can be concluded that the process of normalization which was used to pool expression data into the categories “invasive” and “carriage” has introduced artefacts in the results. Consequently, the genes identified by Genespring GX 7.3.1. as predictive of

invasive disease should not be considered to be predictive without further confirmation using further qRT-PCR experiments to compare expression levels for the other genes identified in Table 6-3.

6.5 Discussion

6.5.1 Overall genomic diversity in Serotype 3 ST180 isolates

The CGH experimental data demonstrates that within serotype 3, ST180 there is much similarity in the genomic content of the isolates despite coming from different patients, with different disease causing or carriage manifestations and from different countries. PFGE performed on serotype 3 isolates from cases of meningitis from Poland also demonstrated very little diversity within serotype 3 (Skoczynska and Hryniewicz, 2003).

It has been proposed by Obert *et al* that there is a core pneumococcal genome and that this includes genes such as SP0043-SP0056, SP0241-0242, SP0314, SP0663-0667, SP0730, SP0965, SP1002, SP1128, SP1204, SP1466, SP1469, SP1869-72, SP1923, SP1937, SP2141-SP2146 and SP2239. This was determined by CGH studies of serotype 6A, 6B and 14 (Obert *et al.*, 2006). The CGH results from these studies of serotype 3 isolates are compatible with the hypothesis of Obert *et al* as all of these genes hybridize from serotype 3 isolates with no variability seen.

Whereas substantial similarity is seen in the CGH results from isolates from Western Europe, there is greater diversity in isolate 07-2838 from Bolivia suggesting again that geographical location has had an influence on the genomic diversity of pneumococci (see Chapter 4 and 9). In 07-2838 there are distinctive hybridization patterns, as SP1335 (a hypothetical protein of 138 bp) and SP1336 (a Type II DNA modification methyltransferase of 1224 bp) are present whereas no hybridization is seen for them in any of the European isolates. Likewise SP1339-SP1341 (coding for hypothetical proteins and an ABC transporter ATP binding protein) do not hybridize in 07-2838 but are present in all tested European isolates.

Some of the genes identified by CGH as not hybridizing have been found to be essential for virulence in serotype 4 pneumococci in a mouse pneumonia model by signature tagged mutagenesis (Hava and Camilli, 2002). Similar identification of genes essential for virulence has been documented for serotype 3 pneumococci by signature tagged mutagenesis (Lau *et al.*, 2001) but the gene designations used in the study are not

comparable with either the TIGR4 or R6 genome nomenclature and so comparison of the results by Lau *et al* and the CGH data observed here is not easily done.

It appears that the serotype 3 ST180 genome can be subdivided into two further subcategories based on the hybridization or absence of hybridization of the regions SP0473-SP0478 (present in the two carriage associated isolates 03-4156 and 07-2838 but not hybridizing in any of the others) and SP2159 to SP2166 (again present in the two carriage associated isolates 03-4156 and 07-2838 but not hybridizing in any of the others). The functions of these genes are shown in Table 6-4 Serotype 3 pneumococci are also proficient at acquiring exogenous DNA. Shen *et al*, noted that a serotype 3 strain (BS71 ST180) had the greatest number of novel sequences which were not found in either the TIGR4 or R6 genomes (Shen *et al*., 2006a).

Gene	Function	Size (bp)
SP0473	ROK family protein, xylose repressor protein	1224
SP0474	PTS system, cellobiose-specific IIC component	1323
SP0475	Hypothetical protein	1941
SP0476	PTS, lactose specific IIA component	345
SP0477	6-phospho-beta-galactosidase	1413
SP0478	PTS, lactose specific IIBC components	1679
SP2159	Fucolectin related protein	3117
SP2160	Hypothetical protein	2295
SP2161	PTS system sugar specific EII component	801
SP2162	PTS system sugar specific EII component	774
SP2163	PTS system IIB component	471
SP2164	PTS system IIA component	432
SP2165	Fucose FucU protein	444
SP2166	L-fuculose phosphate aldolase	639

Table 6-4 Genes from the two regions SP0473-SP0478 and SP2159-SP2166 which code for components of phosphotransferase systems and which are variably present in the genomes of serotype 3, ST180 isolates.

6.5.2 A Carriage Genotype versus an Invasive Genotype for ST180

The ecological niche of invasive pneumococci, at some point prior to invasion, is to be carried asymptomatically no matter how brief is that time period of carriage (Robinson *et al*., 2001). Consequently, comparing the gene complement of pneumococci from invasive disease with those of carriage is difficult as, although it is relatively easy to identify those with an "invasive" phenotype from growth at normally sterile sites, there will always be uncertainty as to whether those cultured from nasopharyngeal "carriage" are truly commensal or are in a transient pre-invasive state or are a mixed population of the two states.

It has therefore proven difficult to definitively find robust associations between genotypes which are more associated with “invasion” than they are with “carriage” as the gene complement of “carriage” isolates may be similar or identical to those of the “invasive” phenotype when assessed by CGH if they happen to be present being carried but in a pre-invasive state. It is no surprise then to see that serotypes and genotypes which are found in the carrier state can also be identified in instances of invasive disease and *vice versa* (Muller-Graf *et al.*, 1999). It may be true that some serotypes which cause invasive disease are rarely found in carriage studies (for example, serotypes 1 and 5) (Scott *et al.*, 1996). This is perhaps due to a very short carrier state prior to invasion or as an artefact resulting from the methodology of the carriage study (Hodges *et al.*, 1946, Davies and Lockley, 1987, Smith *et al.*, 1993). There are many clear instances when such serotypes have been identified in carriers and so it is not the case that they are never carried (Chapter 1). Claiming strict relationships between genotypes and “carriage” or “invasive” phenotypes is unwise as associations with “carriage” found in one study often may be found in disease causing “invasive” isolates in another (Robinson *et al.*, 2001) as has been shown by comparing CGH of “invasive” serotype 14 isolates with genes and regions of diversity thought only to be related to a “carriage” phenotype (Obert *et al.*, 2006) (Chapter 5). Indeed none of the associations of regions of diversity associated with “carriage” or “invasive” phenotypes proposed by Obert *et al* (Obert *et al.*, 2006) hold true for these serotype 3 isolates as the regions are either present or non-hybridizing or variably present or non-hybridizing in both “invasive” and “carriage” phenotypes.

Rather than considering the phenotypes “carriage” and “invasion” as a direct consequence of the gene complement of bacteria, which invariably fails to make a robust association by any single method (Obert *et al.*, 2007), it is more helpful to think in terms of probability of “carriage” or “invasion.” This can be done through calculating odds ratios of a genotype being associated with a particular phenotype (Brueggemann *et al.*, 2003). Such an approach allows for the influences of gene expression differences and post-translational effects and, perhaps more importantly, the influences of host immunity and susceptibility to infection (Inostroza *et al.*, 2001).

There is evidence to suggest that for serotype 3 pneumococci, the sequence of the *cap3A* gene may rapidly change with the introduction of 11-239 bp duplications which result in a phenotype which lacks capsule while spontaneous reversion to the wild type gene returns the capsule expression (Waite *et al.*, 2001). Point mutations and deletions also occur in the serotype 3 capsular gene *cps3D* resulting in different phase variation phenotypes within a serotype 3 population (McEllistrem *et al.*, 2007). Transition between such genotypes

occurs in a biofilm environment (Waite *et al*., 2001, McEllistrem *et al*., 2007). The PCR product "spotted" microarray would not differentiate between the wild type *cap3A* or cps3D genes and their mutant genes as the serotype 3 capsular genes are not represented on this array which consists of genes from TIGR4 and R6 genomes which are not serotype 3. Consequently this microarray cannot differentiate differences in the very genes which may be involved in changing an invasive phenotype to a biofilm associated carriage phenotype (Waite *et al*., 2001) particularly as the variants of *cps3D* also alter transcription and post-transcriptional events in the serotype 3 biofilm (McEllistrem *et al*., 2007). However, this alternating gene phenomenon is unlikely to be the entire explanation for the recognised association of ST180 with carriage in children and invasive disease in adults which is likely also to include an age related change in host immunity.

The region SP2159-SP2166 is a recognised region of diversity (Silva *et al*., 2006, Bruckner *et al*., 2004, Obert *et al*., 2006) and its presence in the pneumococcal genome has been associated with invasive disease in serotype 6A isolates and associated with carriage in serotype 14 isolates (Obert *et al*., 2006). In these CGH investigations of serotype 3 ST180 isolates hybridization of the region SP2159-SP2166 only occurs in isolates 03-4256 and 07-2838 which are both carriage isolates. However a much larger sample size of carriage and invasive isolates of serotype 3 ST180 isolates would be required to see whether the presence of this region is genuinely only present in carriage isolates.

Obert *et al* also suggest that a region of diversity from SP1755-SP1772 is required for bloodstream entry in mice (Obert *et al*., 2006). However in none of the serotype 3 invasive or carriage isolates did the region SP1758-SP1772 hybridize suggesting that it may not be essential for invasion in this serotype.

Our observations are similar to those of Lindsay *et al*, who, using a very similar design of microarray based on *Staphylococcus aureus* genomes, demonstrated that there was no association between gene complement and invasive disease and that,

> "gene combinations necessary for invasive disease may also be necessary for nasal colonisation and that community-acquired invasive disease is strongly dependent on host factors (Lindsay *et al*., 2006)."

6.5.2.1 Genomic Differences in Phosphotransferase (PTS) Systems

It is significant that the regions SP0473-SP0478 and SP2159-SP2166 relate to genes associated with phosphotransferase systems (PTS) which are within known regions of

diversity (Bruckner *et al*., 2004, Silva *et al*., 2006, Obert *et al*., 2006). These genes are involved in the metabolism of lactose which involves the metabolism of fuculose (Table 6-5). Therefore, it was anticipated that this may result in differential metabolism of sugars between the two subcategories of ST180 isolate (Figures 6-1 and 6-2) which is discussed and demonstrated below.

ST180 Genes involved in Lactose metabolism	Gene Function
SP0474	PTS system, cellobiose-specific IIC component
SP0064	PTS system, IIA component
SP0305	Cellobiose phosphotransferase system IIB component
SP0308	Cellobiose phosphotransferase system IIA component
SP0310	Cellobiose phosphotransferase system IIC component
SP0321	PTS system IIA component
SP0323	PTS system IIB component
SP0324	PTS system IIC component
SP0325	PTS system IID component
SP0476	PTS system, lactose-specific IIA component
SP0478	PTS system, lactose-specific IIBC component
SP0577	PTS system, beta-glucosides-specific IIABC components
SP0645	PTS system IIA component, putative
SP0647	PTS system IIC component, putative
SP1684	PTS system IIBC components
SP2036	PTS system IIA component
SP2037	PTS system IIB component
SP2038	Ascorbate-specific PTS system enzyme IIC
SP2161	PTS system IID component
SP2162	PTS system IIC component
SP2163	PTS system IIB component
SP2164	PTS system IIA component
SP1185	PTS system, lactose-specific IIBC components

Table 6-5 Genes identified by Genespring GX 7.3.1. which are involved in lactose metabolism.

6.5.2.1.1 Phenotypic Differences in Sugar Metabolism

In order to assess whether there were differences in sugar metabolism in ST180 isolates with different complements of PTS system genes, three representative isolates were chosen. 07-2838 and 03-4156 both have the genes SP0473-SP0478 and SP2159-SP2166 while 03-4183 does not hybridize at these regions. Rapid ID 32 Strep API (BioMerieux®, France) strips were set up for the three isolates under identical conditions, cultured together for the same duration and developed and read concurrently. Figure 6-4 demonstrates the result.

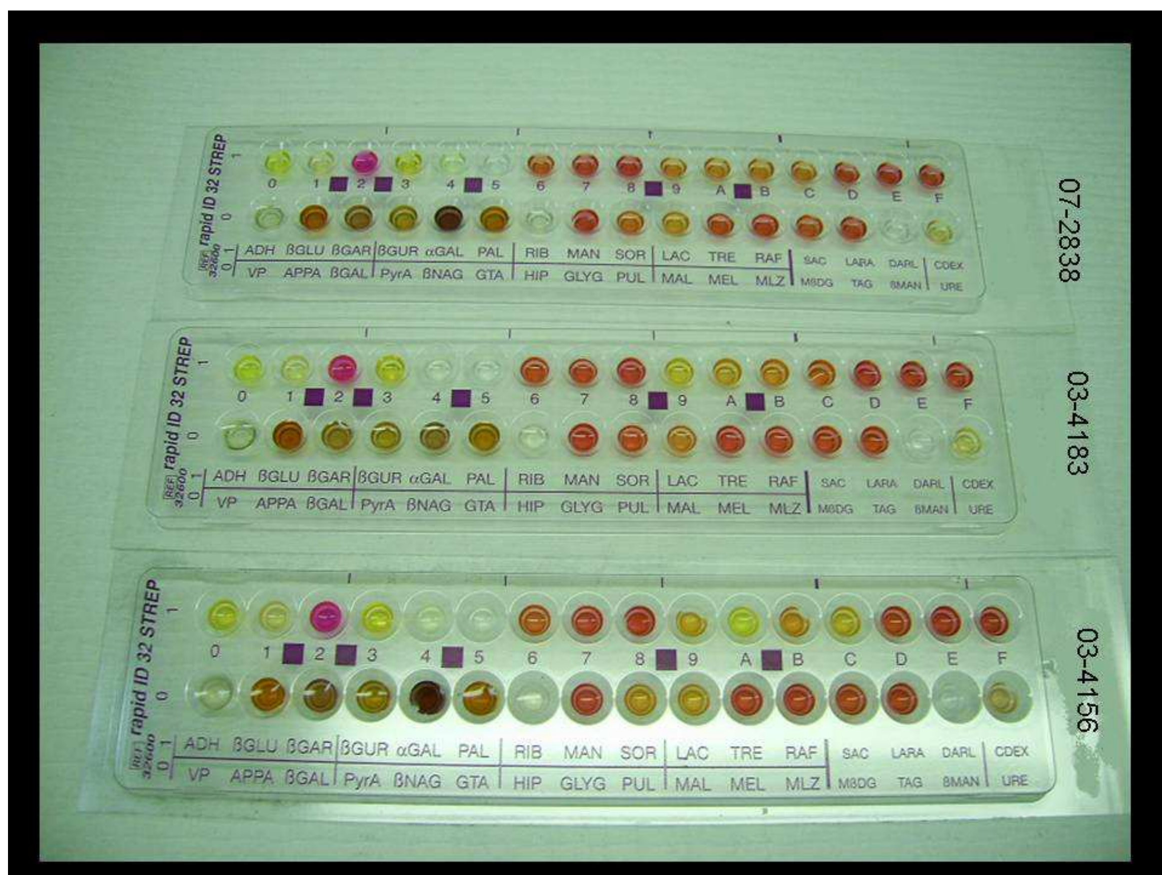


Figure 6-4 Rapid ID 32 Strep API results for three of the carriage associated ST180 isolates (07-2838, 03-4283 and 03-4156) for which microarray CGH demonstrated differences in the complement of PTS associated genes.

Minor differences in results were seen for metabolism of sucrose (SAC), D-trehalose (TRE) and D-lactose (LAC).

Although the assessment and comparison of colour change using the API system is crude and subjective it does suggest a difference in the metabolism of lactose for isolate 03-4183 compared to 03-4156 and 07-2838 which was predicted by the differences in gene content.

Figure 6-5 below demonstrates sugar metabolism pathways with the CGH results for all the ST180 isolates superimposed and the gene expression profiles for these genes are displayed in Figure 6-6.

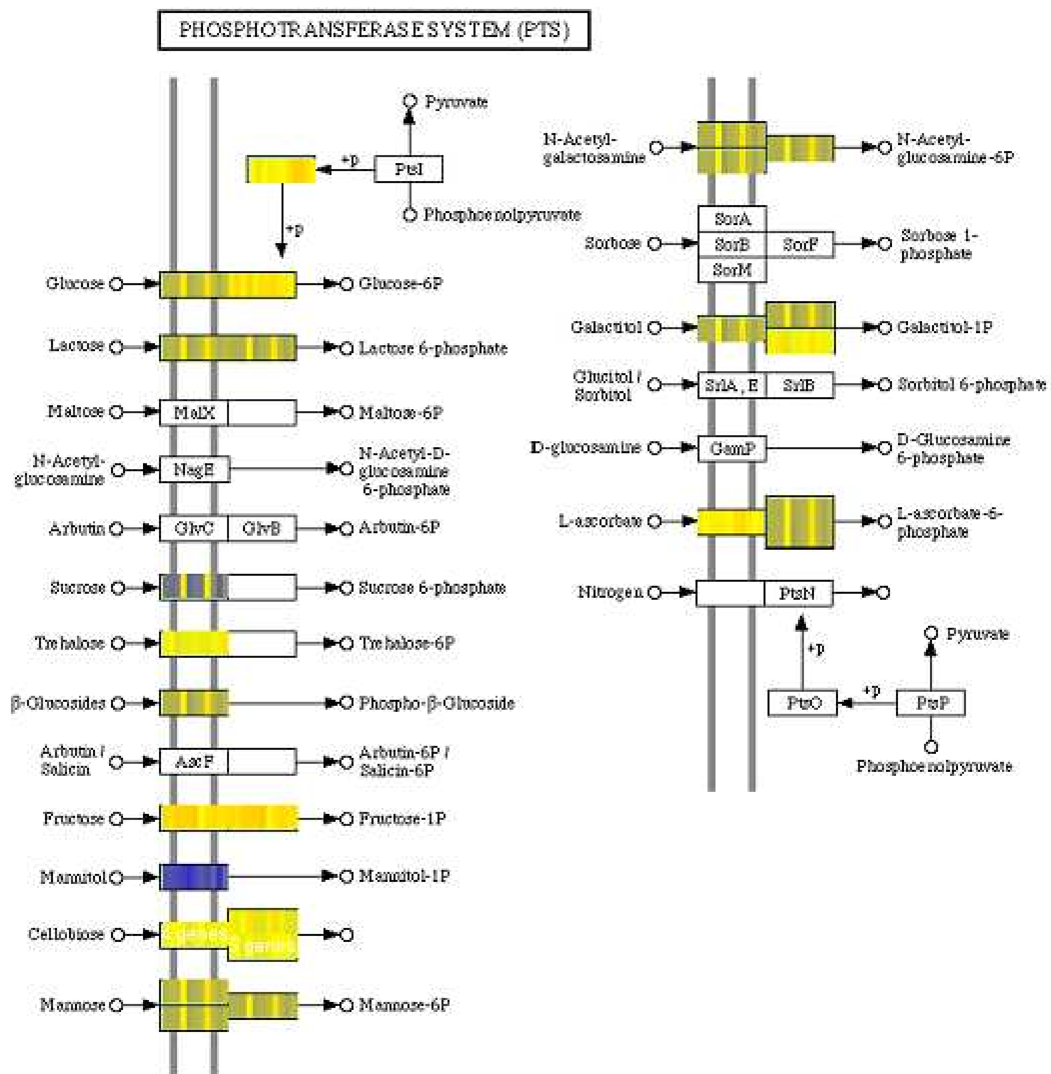


Figure 6-5 Results of Serotype 3 ST180 CGH studies superimposed onto sugar metabolism pathways which relate to phosphotransferase system genes.

In this figure generated by Genespring GX7.3.1, each coloured bar represents one of the ST180 isolates. Shades of yellow indicate the presence of genes in the pathway and shades of blue indicate lack of hybridization for genes in the pathway.

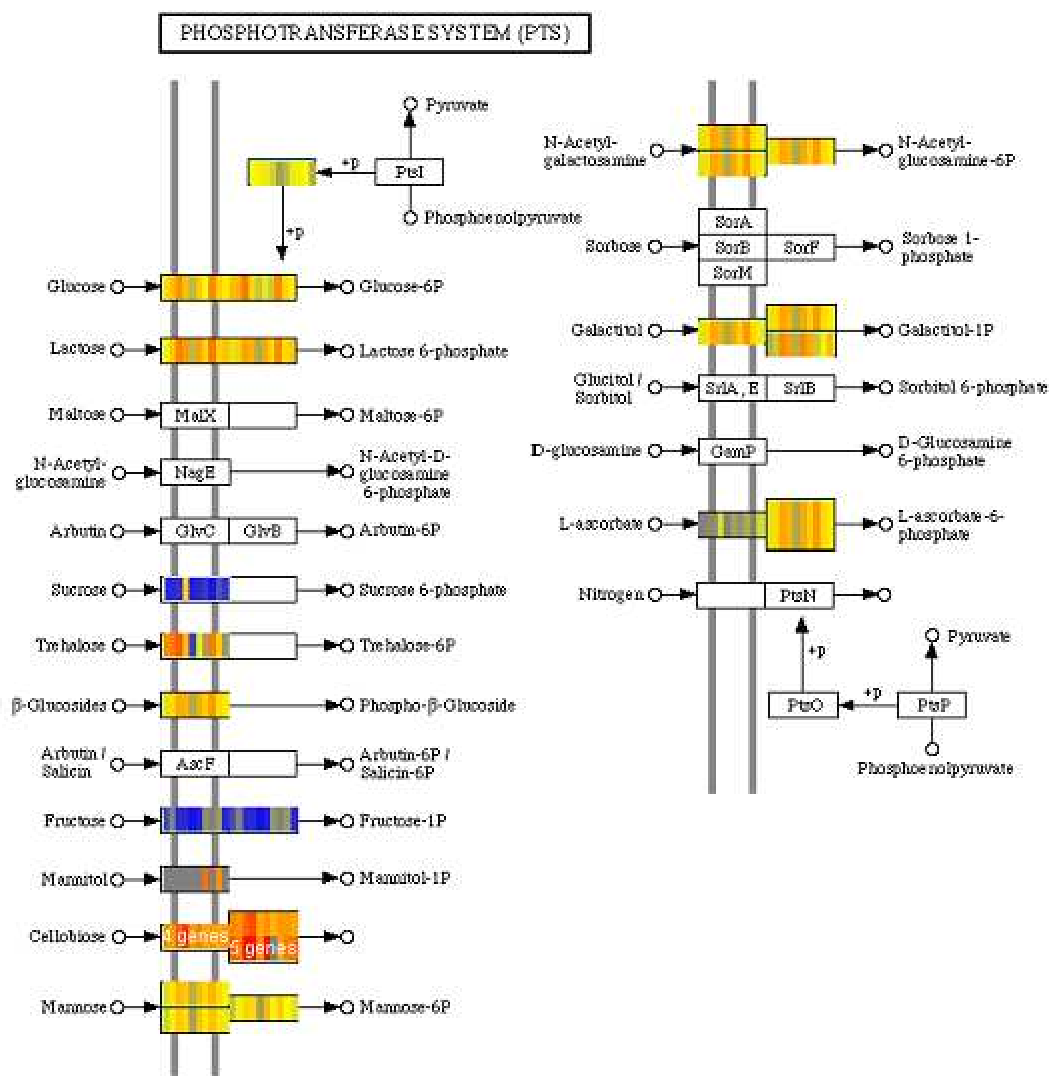


Figure 6-6 Results of Serotype 3 ST180 gene expression studies superimposed onto sugar metabolism pathways which relate to phosphotransferase system genes.

In this figure generated by Genespring GX7.3.1 each coloured bar represents one of the ST180 isolates. Shades of yellow indicate the expression of genes in the pathway at a baseline level, shades of red and orange indicate upregulation of the genes and shades of blue indicate downregulation for genes in the pathway.

Figure 6-5 demonstrates that the ST180s lack genes for mannitol metabolism which correlates with the negative result for mannitol (MAN) metabolism in all the API strips. (The finding that two of the isolates were expressing such genes in Figure 6-6 is spurious and suggests that there has been cross hybridization with this probe by cDNA manufactured from RNA generated by another gene).

It is also possible to see differences in sucrose metabolism which correlate with the API results. Below in Figure 6-7 are compared the expression levels of SP1722 (which is specific to sucrose metabolism) generated by Genespring GX 7.3.1 for the ST180 isolates. It demonstrates that the expression of this gene is significantly downregulated in all the

isolates except 03-4156 which maintains a baseline level of expression. This is compatable with the API (SAC) result in Figure 6-4 where sucrose (D-saccharose) metabolism was positive (yellow) in isolate 03-4156 but negative (red-orange) in the other isolates.

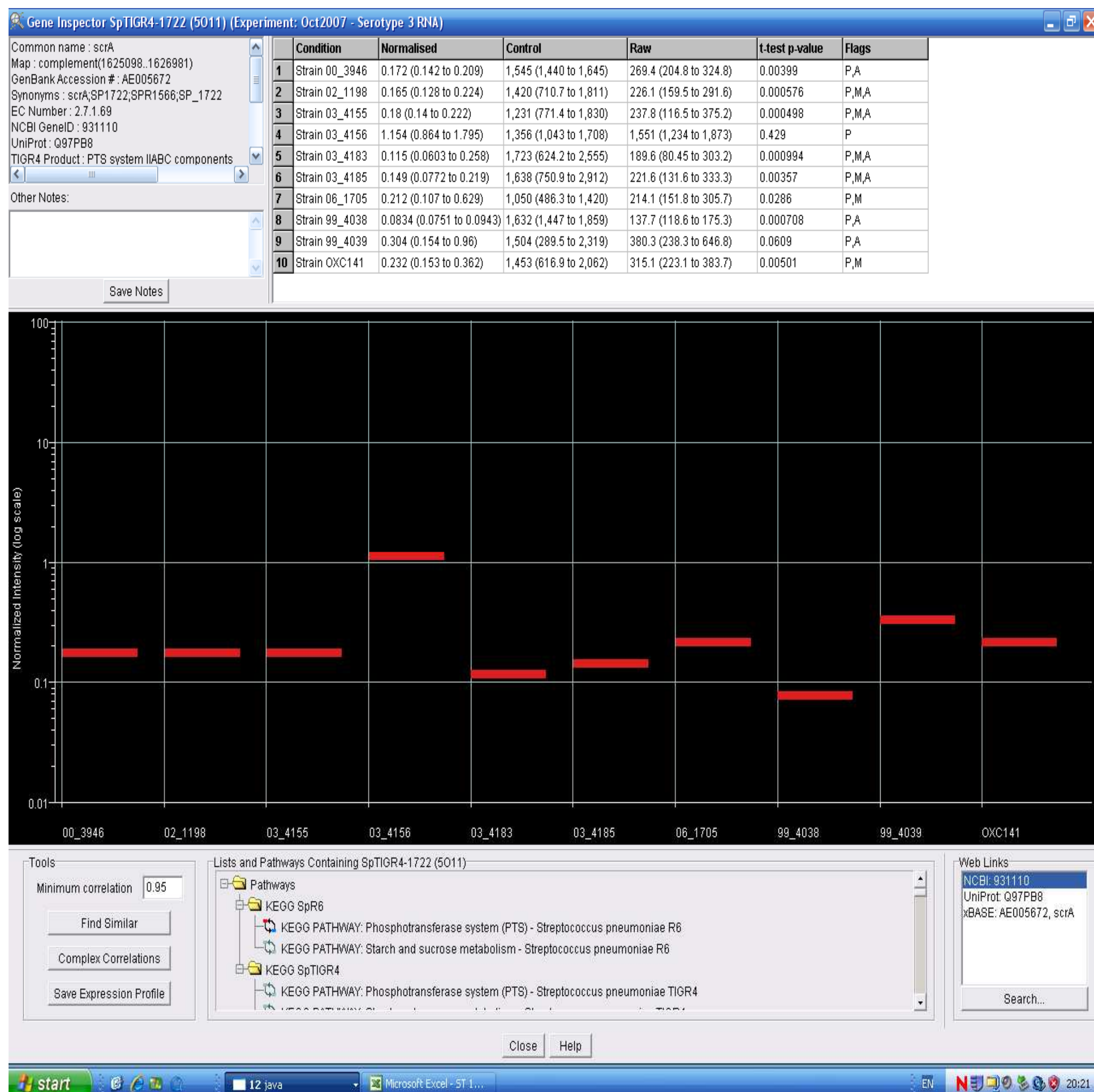
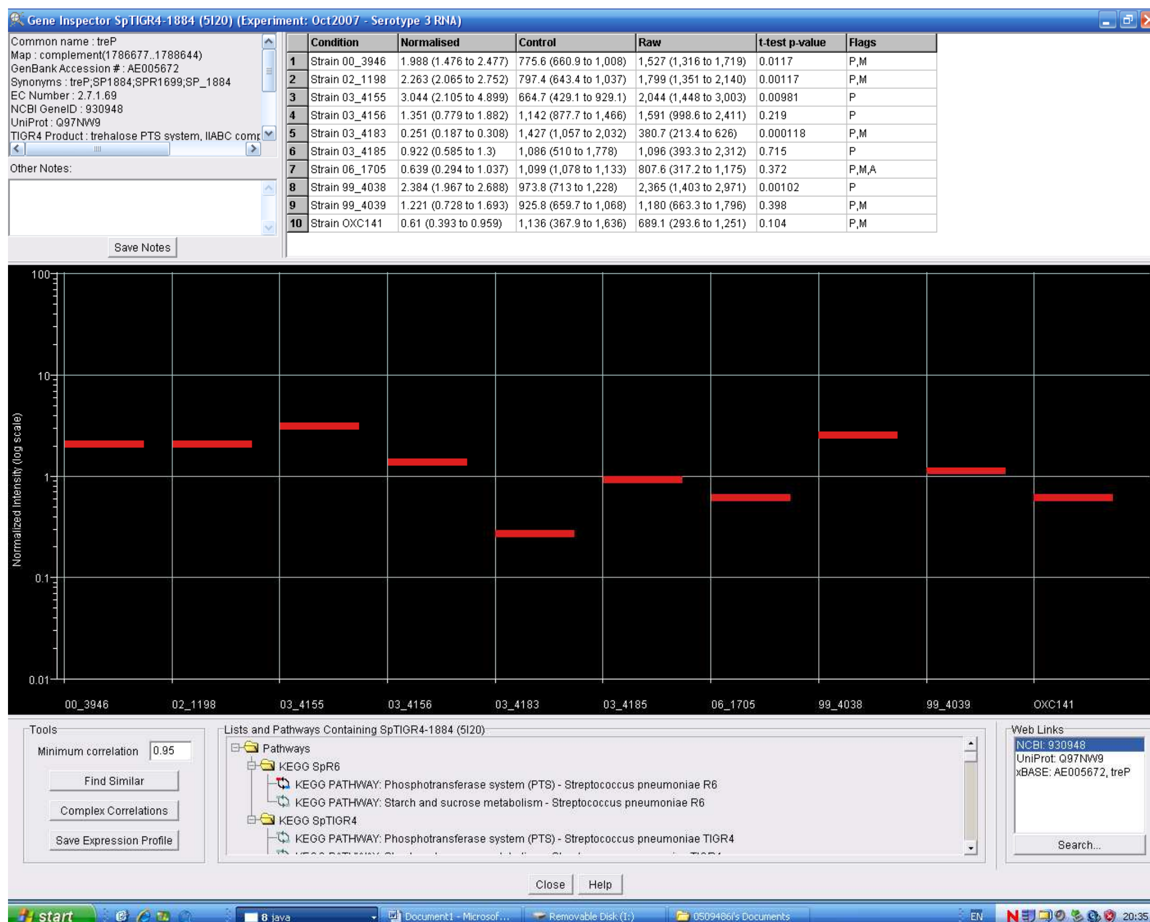


Figure 6-7 Screen view of Genespring GX 7.3.1 demonstrating the maintenance of baseline expression of SP1722 in isolate 03-4156.

Expression of this gene is downregulated in all other tested ST180 isolates.

A similar difference in the expression of genes for trehalose correlated with the API results in Figure 6-4. SP1884 is involved in trehalose metabolism and is downregulated in 03-4183 compared to 03-4156 as seen in Figure 6-8. This is compatible with the results in Figure 6-4 which show that trehalose metabolism is positive in 03-4156 but negative in 03-4183.



	Condition	Normalised	Control	Raw	t-test p-value	Flags
1	Strain 00_3946	1.988 (1.476 to 2.477)	775.6 (660.9 to 1,008)	1,527 (1,316 to 1,719)	0.0117	P,M
2	Strain 02_1198	2.263 (2.065 to 2.752)	797.4 (643.4 to 1,037)	1,799 (1,351 to 2,140)	0.00117	P,M
3	Strain 03_4155	3.044 (2.105 to 4.899)	664.7 (429.1 to 929.1)	2,044 (1,448 to 3,003)	0.00981	P
4	Strain 03_4156	1.351 (0.779 to 1.882)	1,142 (877.7 to 1,466)	1,591 (998.6 to 2,411)	0.219	P
5	Strain 03_4183	0.251 (0.187 to 0.308)	1,427 (1,057 to 2,032)	380.7 (213.4 to 626)	0.000118	P,M
6	Strain 03_4185	0.922 (0.585 to 1.3)	1,086 (510 to 1,778)	1,096 (393.3 to 2,312)	0.715	P
7	Strain 06_1705	0.639 (0.294 to 1.037)	1,099 (1,078 to 1,133)	807.6 (317.2 to 1,175)	0.372	P,M,A
8	Strain 99_4038	2.384 (1.967 to 2.688)	973.8 (713 to 1,228)	2,365 (1,403 to 2,971)	0.00102	P
9	Strain 99_4039	1.221 (0.728 to 1.693)	925.8 (659.7 to 1,068)	1,180 (663.3 to 1,796)	0.398	P,M
10	Strain OXC141	0.61 (0.393 to 0.959)	1,136 (367.9 to 1,636)	689.1 (293.6 to 1,251)	0.104	P,M

Figure 6-8 Screen view of Genespring GX 7.3.1 demonstrating that the expression of SP1884 which is involved in trehalose metabolism.

In isolate 03-4183 expression of SP1884 is markedly less than that of isolate 03-4156 which correlates with a difference in API result for trehalose (TRE) metabolism by API testing (Figure 6-4). It is also noteworthy that the above results suggest increased expression of SP1884 and trehalose metabolism in blood (99-4038) compared to cerebrospinal fluid (99-4039) from the same case of meningitis (Chapter 10).

Correlations between microarray CGH results for sugar metabolism genes which show diversity and sugar metabolism results observed by API testing have also been consistently documented by Aakra *et al* when comparing *Enterococcus faecalis* strains (Aakra *et al*., 2007). In pneumococci, Oggioni *et al* have noted that genomic variation in the PTS systems does influence sugar fermentation (Oggioni *et al*., 2008).

It may be that these subtle differences in sugar metabolism *in vitro* have a relationship to the transitioning serotype 3 capsule genes and their expression and are in some way involved in the phase variation noted above although unravelling whether there is such a relationship is beyond the scope of this thesis.

7 Genomic Diversity in Isolates of the Same Serotype and Multilocus Sequence Type Related to Clinical Manifestation and Outcome

7.1 Background

Serotype 4 pneumococci are an important cause of invasive pneumococcal disease (IPD) in humans and can cause severe invasive disease in animal models (Sandgren *et al*., 2005). Factors in addition to the pneumococcal capsule (a major virulence factor which determines the serotype) contribute to disease outcome (Mizrachi-Nebenzahl *et al*., 2004) and clonal properties other than serotype influence ability to cause invasive disease (Sandgren *et al*., 2005, Kerr *et al*., 2006).

Some virulence associated genes, such as *nanC*, may not be uniformly distributed within a single sequence type (Pettigrew *et al*., 2006) and within a sequence type, a virulence associated gene can demonstrate sequence diversity which may alter its function (Kirkham *et al*., 2006). The pneumococcus has many genes which are highly variable with multiple known alleles such as the pneumococcal surface protein PspA (Roche *et al*., 2003) but even within the housekeeping genes which are the basis of the MLST scheme and which are considered to be much more conserved, divergent sequences have been identified (Diggle and Clarke, 2005). With this high degree of variation in the pneumococcal genome, it was anticipated that sequence differences, such as point mutations or insertions resulting in frame shifts, may be identifiable which may be contributing to different clinical presentations.

It is known that bacteriophages are sources of DNA which, when integrated into bacterial genomes, result in greater genomic diversity. Exogenous DNA from bacteriophages can code for virulence associated proteins such as the lytic activity of the Pal enzyme belonging to the pneumococcal phage Dp-1(Lopez *et al*., 2000). Fully functional lysogenic phages, defective phages and remnant prophages are widespread amongst pneumococcal isolates of different serotypes and different geographical origins (Gindreau *et al*., 2000, Ramirez *et al*., 1999). It has been stated that the role of conjugative transposons and bacteriophage needs to be addressed in a clinical setting (Lopez, 2006).

It was decided to investigate a series of IPD cases caused by the same serotype and sequence type (serotype 4, ST246). Six isolates from patients of similar age, gender, racial background and geographical location were chosen for analysis. Their clinical manifestations were reviewed and the isolates used to perform microarray CGH, look for evidence of bacteriophage carriage and thereby assess whether genomic diversity within the ST246 clone was identifiable which might have contributed to the varied disease presentations. Isolate 06-1803 was primarily used because of the unusual clinical presentation of the source patient and five isolates were chosen of the same serotype and sequence type which originated from patients of similar characteristics using a database held at SMPRL.

7.2 Epidemiological details for Serotype 4 ST246 Test Isolates

Isolate	Year of Isolation	Age of Patient	Sex of Patient	Location	Racial Origin	Isolate Source	Serotype	MLST
03-5339	2003	37	Male	North East Glasgow	Caucasian	Blood culture	4	ST246
04-1342	2004	53	Male	North East Glasgow	Caucasian	Blood culture	4	ST246
04-2239	2004	42	Male	North East Glasgow	Caucasian	Blood culture	4	ST246
05-1109	2005	56	Male	North East Glasgow	Caucasian	Blood culture	4	ST246
06-1803	2006	49	Male	West Glasgow	Caucasian	Blood culture	4	ST246
06-1898	2006	64	Male	North East Glasgow	Caucasian	Blood culture	4	ST246

Table 7-1 Basic epidemiological information about the source patients from which serotype 4, ST246 isolates were received.

7.3 Clinical Manifestations and Outcomes for Serotype 4 ST246 Test Isolates

Isolate	Clinical History
03-5339	Diagnosed with bacteraemic pneumonia which did not require intensive care. Antibiotic therapy was not documented.
04-1342	Diagnosed with an acute confusional state and bacteraemic lobar pneumonia. Experienced severe sepsis during first 24 hours of admission. Gradually improved on intravenous amoxycillin.
04-2239	Presented with pneumonia and septic shock which required immediate intensive care management. Initially treated with intravenous ceftriaxone, clarithromycin, inotropes and recombinant activated protein C. Required intensive care management for over a month but survived.
05-1109	Diagnosed with bacteraemic pneumonia which did not require intensive care. Patient had an uneventful recovery on intravenous amoxycillin.
06-1803	Presented with concurrent pneumococcal meningitis, pneumonia and aortic valve endocarditis. Required emergency aortic valve replacement on day 2 of admission. Failed to settle on intravenous ceftriaxone and gentamicin (despite being a penicillin susceptible isolate *in vitro*). Still failed to settle after addition of intravenous vancomycin and developed an aortic root abscess. Finally recovered after 3 months of treatment with daptomycin.
06-1898	Presented with bacteraemic pneumonia and renal impairment. Patient responded to intravenous ceftriaxone and recovered uneventfully.

Table 7-2 Brief clinical histories of the cases from which serotype 4, ST246 isolates were received.

7.4 Bacteriophage induction

Bacteriophage induction for the six clinical isolates was performed by Dr Patricia Romero using the following methodology. Isolates were grown at 37°C until an optical density at 600nm of between 0.1-0.25 was reached. Mitomycin C was added to a final concentration of 100 ng/ml to induce the release of lysogenic bacteriophages. 200 µl of each culture (before and after the addition of mitomycin C) were added to wells of a 96-well plate in triplicate. Growth was monitored by optical density at 600nm in a plate reader (Fluostar OPTIMA, BMG LABTECH, Germany). Bacteriophage induction identified a bacteriophage only in isolate 05-1109.

7.5 Microarray DNA CGH Results

DNA microarray CGH was performed using the protocols outlined in Chapter 2. One large region of diversity from SP1129-SP1146 was evident and is magnified in Figure 7-1. In total, 46 gene loci where diversity was evident were identified by CGH (Figure 7-2).

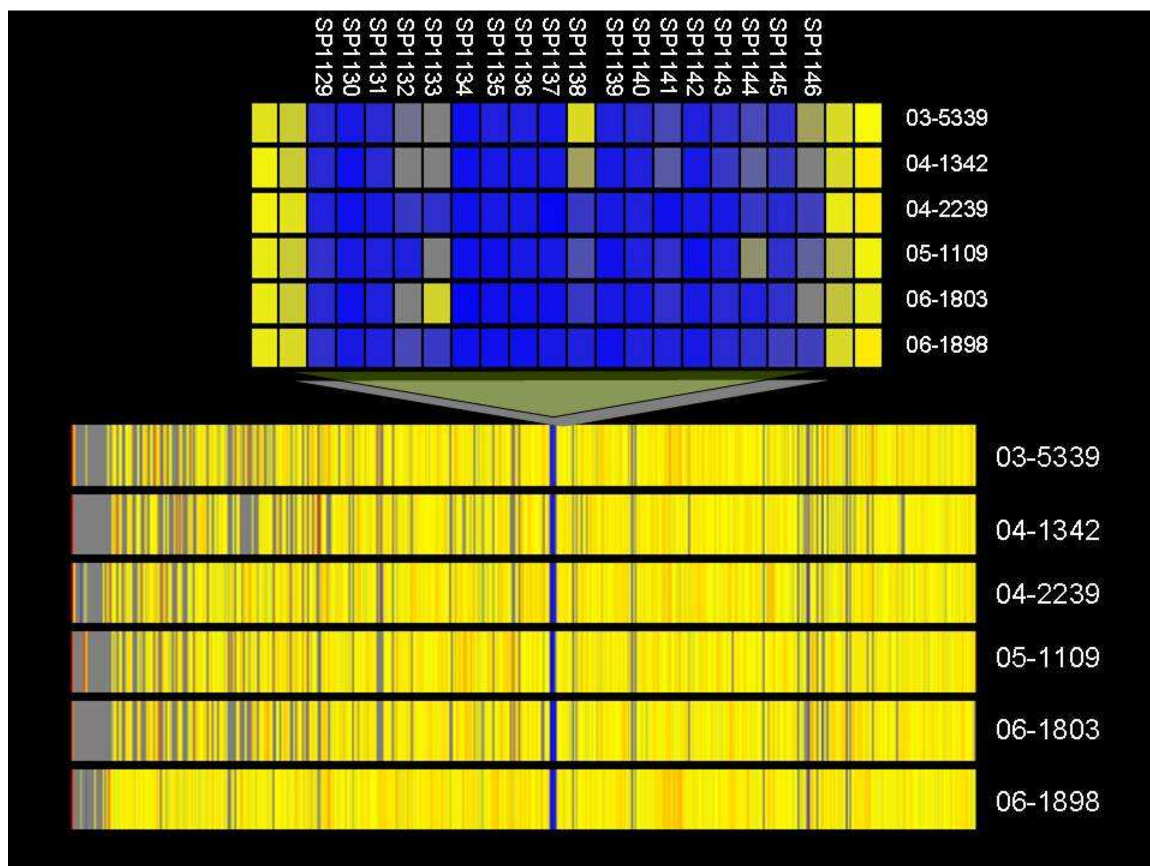


Figure 7-1 DNA CGH results for serotype 4, ST246 isolates.

This figure is produced by Genespring GX 7.3.1. The genes from the TIGR4 and R6 genomes featured on the microarray are illustrated as horizontal bars consisting of vertical lines representing consecutive genes. Yellow lines indicate where competitive hybridization has occurred. Red lines indicate where R6 genes are present and hybridizing in the test isolate but are absent from TIGR4. Blue lines indicate genes which are present and hybridizing in TIGR4 but not hybridizing with the test isolates. Grey lines indicate that the fluorescence intensity of the hybridization has not been high enough to reach a threshold set on Bluefuse for Microarrays 3.5 © to filter out low quality hybridizations. The region of diversity between SP1129 –SP1146 is magnified.

PCR (utilising the primers used to design the probes (Appendix 3) attached to the microarray (MWG Biotech AG, Germany)) demonstrated absence of the gene at 12 of these loci (Figure 7-2) and suggested that sequence divergence may account for detection of product at the other loci.

Gene loci	Gene Function	Associated with Virulence in Murine Pneumonia Model	Clinical Isolates					
			04-2239	03-5339	04-1342	06-1803	05-1109	06-1898
SP0031								
SP0068								
SP0111								
SP0165								
SP0309								
SP0514								
SP0532								
SP0570								
SP0573								
SP0949	IS1515, transposase, authentic frameshift	No						
SP1129	integrase/recombinase, phage integrase family	No						
SP1130	transcriptional regulator	No						
SP1131	transcriptional regulator, putative	No						
SP1132								
SP1134								
SP1135								
SP1136	conserved domain protein	No						
SP1137								
SP1138								
SP1139								
SP1140	hypothetical protein	No						
SP1141	hypothetical protein	No						
SP1142								
SP1143								
SP1144								
SP1145	hypothetical protein	No						
SP1158								
SP1181								
SP1189	hypothetical protein	No						
SP1254								
SP1342								
SP1343								
SP1350								
SP1353								
SP1696								
SP1718								
SP1770	glycosyl transferase, family 8	Yes						
SP1793								
SP1794	hypothetical protein	No						
SP1796	ABC transporter, substrate-binding protein	No						
SP1797								
SP1799								
SP1839								
SP1895								
SP1896								
SP1897								

Figure 7-2 PCR validation results for genes where are least one of the 6 ST246 isolates did not hybridize on the microarray.

Yellow indicates the presence of a PCR product of the same size as that in the TIGR4 genome (using the primers utilised in the manufacture of the microarray displayed in Appendix 3) while blue indicates the absence of a product. Where genes have been identified as absent (blue) their putative function is noted and whether they have been identified as virulence factors in the development of pneumonia in a mouse model (Hava and Camilli, 2002).

Five of the absent genes were absent in all 6 clinical isolates while 7 were strain dependent and could be absent or present in the ST246 clinical isolates.

Isolate 06-1803 demonstrated the greatest number of putative gene deletions, several of which were identified as absent solely in this isolate.

We chose three genes (SP1130, SP1136 and SP1342) where DNA in test isolates did not hybridize on the microarray but had demonstrable products using PCR and we sequenced the PCR products. This identified single nucleotide polymorphisms (SNPs) in several isolates when the sequenced genes were compared with their TIGR4 sequence using National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) software. The genetic diversity of SP1130 is illustrated in Figure 7-3.

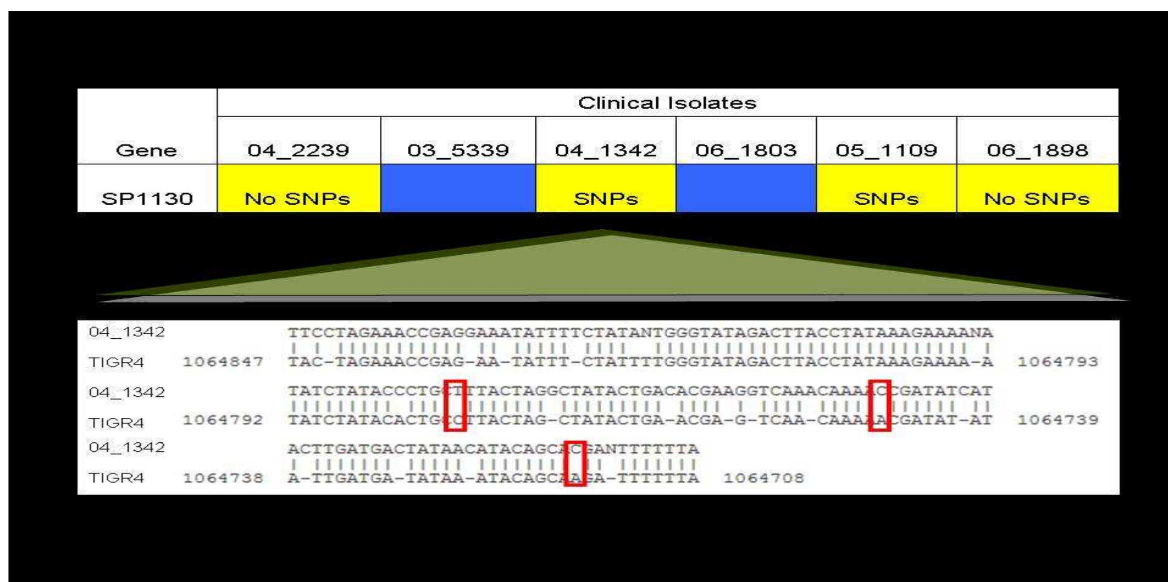


Figure 7-3 Genomic diversity of the transcriptional regulator gene SP1130.

The presence or absence of gene SP1130 in the 6 clinical ST246 serotype 4 isolates was determined using PCR and compared with results using microarray CGH where no hybridization occurred for this gene in any of the test isolates. Blue indicates where the gene was absent by PCR and yellow indicates that a product was detected which was the same size as the product from TIGR4. Single nucleotide polymorphisms (SNPs) were identified in the sequenced PCR product when compared to the TIGR4 sequence for isolates 04-1342 and 05-1109. The locations of some of these SNPs are identified in red boxes for part of the PCR product from isolate 04-1342.

7.6 Discussion

This investigation has demonstrated genetic diversity within clinical isolates of a single clonal complex (ST246) associated with one serotype (serotype 4) from patients with matched epidemiological characteristics using a PCR product microarray as a screening investigation. These DNA CGH results were validated using PCR. This work also assessed whether there may be additional integrated genetic material by bacteriophage induction.

A bacteriophage was identified only in isolate 05-1109 which had integrated 32-34kb of genetic material. This burden of genetic material represents a source of genomic diversity which would otherwise have remained undetected by microarray CGH analysis. This again highlights one of the limitations of microarray DNA CGH as it can only identify the presence of genes which feature in the reference genomes which were used to construct the probes on the microarray. Consequently, it may not detect additional genes acquired by horizontal gene transfer from other bacterial species or from bacteriophages, the acquisition of which may affect gene expression and subsequently the bacterial phenotype and disease presentation.

It has previously been recognised that the TIGR4 genome demonstrates evidence of a bacteriophage remnant in a 10.5kb cluster of 19 contiguous open reading frames (from SP1129 to SP1147) that is absent from the genome of R6 (Obregon *et al*., 2003). This corresponds exactly with the only significant region of diversity between these clinical isolates and TIGR4. This region of diversity has been previously identified (Silva *et al*., 2006) but the presence of genes in this region has previously been associated with a noninvasive phenotype when described before in serotype 6B isolates (Obert *et al*., 2006). This investigation of serotype 4 isolates from patients with bacteraemic pneumonia and other IPD manifestations suggests that if there really is an association with a noninvasive phenotype this association may be serotype specific.

The identification of some of the genes in the SP1129-SP1147 region in some isolates suggests that these isolates or their ancestors had carried a bacteriophage. Bacteriophages may therefore have a significant role in the generation of genomic diversity in the pneumococcus. For these isolates, it appears that genomic diversity (with a possible consequence being diversity of clinical manifestations) has arisen more as a result of acquiring additional genetic content rather than deletion of genetic material when compared to another serotype 4 strain (TIGR4) as deletion of individual genes within these isolates does not appear to be common when compared to the TIGR4 genome. It is

possible that the unique combination of gene deletions in isolate 06-1803 may have contributed to the unusual clinical presentation with simultaneous pneumonia, meningitis and endocarditis.

While considering the pathogenesis of pneumococcal endocarditis, Bruckner *et al* have identified SP1772, a gene coding for a cell wall anchor protein, as homologous to a platelet binding glycoprotein in *Streptococcus gordonii* (Bruckner *et al*., 2004) and Takamatsu *et al* have considered it important in the causation of endocarditis (Takamatsu *et al*., 2004). Our results do concur with this hypothesis as SP1772 did hybridize in all the test clinical strains, suggesting that all had the genetic potential to result in endocarditis if this hypothesis is correct. Only isolate 06-1803 however came from a patient with endocarditis, which suggests that there may be other pneumococcal genetic involvement required for the development of endocarditis or, perhaps more likely, a requirement for a particular host susceptibility such as underlying valvular pathology. It is noteworthy that in the CGH analysis performed by Bruckner *et al*, SP1772 appears absent from several of the serotypes analysed particularly serotype 3 (Bruckner *et al*., 2004). Our results from other DNA CGH experiments have also demonstrated an absence of hybridization for SP1772 in all serotype 3 isolates tested but also in serotype 23F (04-1168 and ATCC51916), serotype 6A (BAA659), serotype 14 (BAA340), serotype 20 (05-1271), serotype 35B (BAA660), serotype 12F (05-2565) and serotype 9V (05-1821). But before it can be concluded that the absence of SP1772 may preclude the development of endocarditis, it should be acknowledged that several case series of pneumococcal endocarditis from the mid-20th century (Austrian, 1957, Straus and Hamburger, 1966, Finland and Barnes, 1970) demonstrate that serogroups 3, 6, 14 and 20 can indeed cause endocarditis. As genetic analysis of these historical isolates is not possible, there is no conclusive proof of a pneumococcus without SP1772 being associated with endocarditis but given the universal lack of hybridization of SP1772 in all serotype 3 isolates analysed so far by CGH, a claim that possession of SP1772 is necessary to cause endocarditis is hard to support. Likewise, the glycosyl transferase SP1770, which was identified by signature tagged mutagenesis as being essential for virulence in a murine pneumonia model (Hava and Camilli, 2002) does not appear to be necessary in the pathogenesis of bacteraemic pneumococcal pneumonia in humans.

These results also highlight the impact that sequence diversity can have on how microarray DNA CGH results are interpreted. The results show that even small mutations in gene sequence may affect hybridization on our microarray which was evident when there was failure to hybridize genes which contained occasional SNPs in the test DNA compared to

its probe. Shen *et al* have also demonstrated that several pneumococcal genes contain numerous point mutations and small indels (Shen *et al*., 2006a). Detection of SNPs using microarray technology has been possible in the human genome project (Wang *et al*., 1998) but in order to do this with higher specificity for pneumococcal genomes, a new customised microarray would be required using an oligonucleotide probe approach rather than a PCR product probe (Kumar *et al*., 2005, Palacios *et al*., 2007, Dorrell *et al*., 2005). This approach would be less successful at detecting larger genes due to the shorter nature of the probes employed (Palacios *et al*., 2007). This has led to the conclusion that,

> "given the uncertainties about the extent of naturally occurring genetic polymorphisms in pathogen gene pools, the most practical hybridization-based detectors will probably rely on longer probes that are less sensitive to unexpected genetic polymorphism and that also provide greater analytic sensitivity compared with shorter oligonucleotide probes (Call, 2005)."

Silva *et al* note that isolates of the same serotype and sequence type can behave very differently *in vivo* in animal hosts. These observations on human cases due to the same serotype and sequence type with very different clinical courses agree with this observation.

Figure 7-4 below demonstrates that ST246 is a single locus variant of ST899 and a double locus variant of ST2365 and ST695. These serotype 4 related sequence types have been associated with capsular switching and acquisition of a serotype 19A capsule as a conjugate vaccine escape mutation (Brueggemann *et al*., 2007). Although it has not yet been seen, it would seem likely that such a closely related sequence type (ST246), which is of obvious virulence in humans, has the potential to switch capsules with serotype 19A to escape vaccine pressure but maintain the virulence associated with the ST246 clonal complex (Sandgren *et al*., 2004).

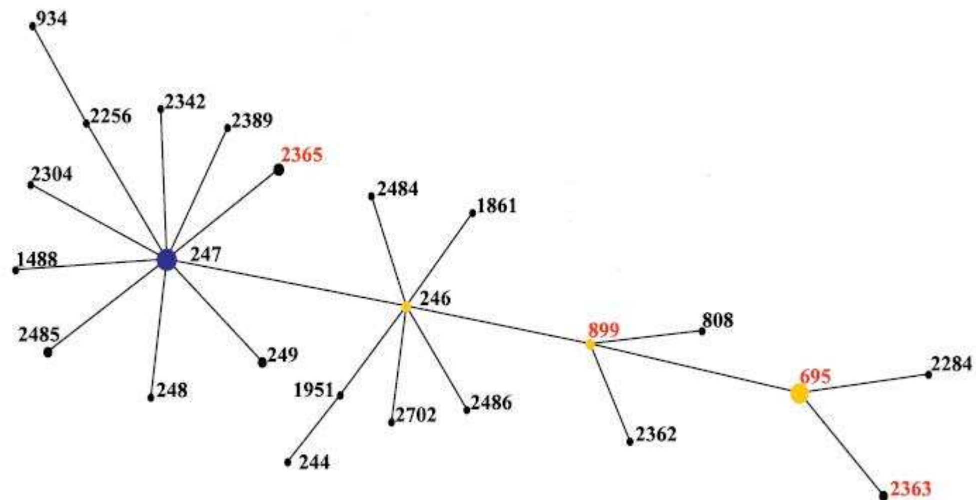


Figure 7-4 Demonstration of the relationship of ST246 to other closely related sequence types using e-BURST version 3 and the MLST database.

Adapted from *Brueggemann et al* (Brueggemann *et al.*, 2007). ST247 is found to be a founder clonal complex and ST246, ST899 and ST695 are co-founders. Sequence types highlighted in red have been found to have undergone capsular switching in recent years and have acquired and express the serotype 19A capsule.

8 Genomic Diversity in Nosocomial Outbreaks of Pneumococcal Disease

8.1 Pneumococcal Outbreaks – Definition and Features

According to Health Protection Agency draft guidelines, a cluster or outbreak of pneumococcal disease is defined as,

> "two or more cases of serious pneumococcal infection (confirmed or probable) reported from a closed setting within a four-week period."

Nosocomial outbreaks of IPD, although not common, are well described. Perhaps the first description of such a hospital acquired outbreak of pneumococcal pneumonia was at an asylum in 1903 (Sinigar, 1903). These outbreaks are associated with significant preventable morbidity and mortality particularly in elderly (Hansmann *et al.*, 2006, Thakker *et al.*, 1998, Bain *et al.*, 1990, Millar *et al.*, 1994, Cartmill and Panigrahi, 1992, Dawson *et al.*, 1992, Fiore *et al.*, 1998, Mandigers *et al.*, 1994, Bresnitz *et al.*, 2001, Kludt *et al.*, 1997, Gleich *et al.*, 2000, Bescos *et al.*, 2003, Tan *et al.*, 2003, Weiss *et al.*, 2001, Quick *et al.*, 1993, Barnes *et al.*, 1995, Gould *et al.*, 1987, Nuorti *et al.*, 1998, Gillespie *et al.*, 1997) and paediatric populations (Gupta *et al.*, 2007, Medeiros *et al.*, 1998, Leighton *et al.*, 2003, Craig *et al.*, 1999, Cherian *et al.*, 1994, Schroder and Cooper, 1930, O'Brien *et al.*, 2000, Strom, 1932, Melamed *et al.*, 2002, Radetsky *et al.*, 1981, Gilman and Anderson, 1938, Dagan *et al.*, 2000). Remarkably there is, as yet, no consensus as how best to prevent or manage them although in 2007 a Health Protection Agency working group issued draft interim guidelines for consultation in the United Kingdom.

Particular serogroups have a predilection for causing outbreaks (Table 8-1) possibly due to a genetic factor which has resulted in increased transmissibility. This is unlikely to be due solely to capsular properties as, "non-typeable," uncapsulated pneumococci can also cause outbreaks – particularly of conjunctivitis. Virtually all possible presentations of pneumococcal disease have been associated with outbreaks as shown in Table 8-2.

Pneumococcal Serogroup Associated with Outbreaks	References
Serogroup 1	(Thakker *et al.*, 1998, Gupta *et al.*, 2007, Musher *et al.*, 1997, Gilman and Anderson, 1938, Smeall, 1931, Park and Chickering, 1919, Gratten *et al.*, 1993, Mercat *et al.*, 1991, DeMaria *et al.*, 1980, Mackenzie *et al.*, 1940, Leimkugel *et al.*, 2005, Yaro *et al.*, 2006, O'Brien *et al.*, 2000, Strom, 1932, Dagan *et al.*, 2000, Proulx *et al.*, 2002)
Serogroup 2	(Smillie *et al.*, 1938)
Serogroup 3	(Bescos *et al.*, 2003)
Serogroup 4	(Hansmann *et al.*, 2006, Gleich *et al.*, 2000, Bain *et al.*, 1990, Clarke *et al.*, 2004a, Crum *et al.*, 2003)
Serogroup 5	(Schroder and Cooper, 1930, Melamed *et al.*, 2002)
Serogroup 6	(Cartmill and Panigrahi, 1992, Dawson *et al.*, 1992, Radetsky *et al.*, 1981)
Serogroup 8	(Birtles *et al.*, 2005, Berk *et al.*, 1985)
Serogroup 9	(Anonymous, 1992, Millar *et al.*, 1994, Mandigers *et al.*, 1994, Gillespie *et al.*, 1997, Crum *et al.*, 2003)
Serogroup 12	(Hoge *et al.*, 1994, Jorgensen *et al.*, 2005, Cherian *et al.*, 1994)
Serogroup 14	(Craig *et al.*, 1999, Fiore *et al.*, 1998, Bresnitz *et al.*, 2001, Tan *et al.*, 2003, Medeiros *et al.*, 1998)
Serogroup 19	(Clarke *et al.*, 2004a, Quick *et al.*, 1993)
Serogroup 23	(Carter *et al.*, 2005, Barnes *et al.*, 1995, Weiss *et al.*, 2001, Fry *et al.*, 2005, Gould *et al.*, 1987, Nuorti *et al.*, 1998)
Non-Typeable	(Hennink *et al.*, 2006, Martin *et al.*, 2003a, Feingold, 2003, Leighton *et al.*, 2003, Crum *et al.*, 2004, Carvalho *et al.*, 2003, Buck *et al.*, 2006, Ertugrul *et al.*, 1997)

Table 8-1 Pneumococcal serogroups which can cause disease outbreaks.

Pneumococcal Manifestation Observed in Outbreaks	References
Asymptomatic Carriage	(Jorgensen *et al.*, 2005, Gillespie *et al.*, 1997, Mackenzie *et al.*, 1940, Hoge *et al.*, 1994, Musher *et al.*, 1997, Bescos *et al.*, 2003, Carter *et al.*, 2005, Nuorti *et al.*, 1998, Radetsky *et al.*, 1981, Quick *et al.*, 1993, Barnes *et al.*, 1995)
Pneumonia	(Jorgensen *et al.*, 2005, Gillespie *et al.*, 1997, Mackenzie *et al.*, 1940, Lamb and Brannin, 1919, Anonymous, 1992, Smillie *et al.*, 1938, Schroder and Cooper, 1930, Hoge *et al.*, 1994, DeMaria *et al.*, 1980, Mercat *et al.*, 1991, Millar *et al.*, 1994, O'Brien *et al.*, 2000, Cartmill and Panigrahi, 1992, Dawson *et al.*, 1992, Proulx *et al.*, 2002, Musher *et al.*, 1997, Fiore *et al.*, 1998, Mandigers *et al.*, 1994, Gratten *et al.*, 1993, Strom, 1932, Hirsch and McKinney, 1919, Tan *et al.*, 2003, Gilman and Anderson, 1938, Park and Chickering, 1919, Bescos *et al.*, 2003, Gleich *et al.*, 2000, Kludt *et al.*, 1997, Bain *et al.*, 1990, Hansmann *et al.*, 2006, Dagan *et al.*, 2000, Subramanian *et al.*, 2003, Carter *et al.*, 2005, De Galan *et al.*, 1999, Nuorti *et al.*, 1998, Bresnitz *et al.*, 2001, Thakker *et al.*, 1998, Gupta *et al.*, 2007, Weiss *et al.*, 2001, Quick *et al.*, 1993, Fry *et al.*, 2005, Crum *et al.*, 2003, Berk *et al.*, 1985, Gould *et al.*, 1987)
Empyema	(Jorgensen *et al.*, 2005, Smillie *et al.*, 1938, DeMaria *et al.*, 1980, O'Brien *et al.*, 2000, Gupta *et al.*, 2007)
Meningitis	(Jorgensen *et al.*, 2005, Craig *et al.*, 1999, Yaro *et al.*, 2006, Leimkugel *et al.*, 2005, Birtles *et al.*, 2005, Hoge *et al.*, 1994, DeMaria *et al.*, 1980, Melamed *et al.*, 2002, Radetsky *et al.*, 1981)
Blood Stream Infection	(Jorgensen *et al.*, 2005, Cherian *et al.*, 1994, Tan *et al.*, 2003, Bescos *et al.*, 2003, Gleich *et al.*, 2000, Kludt *et al.*, 1997, Hansmann *et al.*, 2006, Dagan *et al.*, 2000, Carter *et al.*, 2005, Bresnitz *et al.*, 2001)
Otitis Media	(Gilman and Anderson, 1938, Dagan *et al.*, 2000)
Conjunctivitis	(Feingold, 2003, Leighton *et al.*, 2003, Martin *et al.*, 2003a, Hennink *et al.*, 2006, Cherian *et al.*, 1994, Crum *et al.*, 2004, Carvalho *et al.*, 2003, Buck *et al.*, 2006, Medeiros *et al.*, 1998, Ertugrul *et al.*, 1997)
Septic Arthritis	(Jorgensen *et al.*, 2005, DeMaria *et al.*, 1980)
Cellulitis	(Jorgensen *et al.*, 2005)
Peritonitis	(Malloch, 1922)

Table 8-2 Presentation of pneumococcal disease or carriage associated with outbreaks.

8.2 Pneumococcal Typing Methods and Limitations in Outbreak Investigations

8.2.1 Serotyping

Serotyping, although usually performed in a reference laboratory, is an important test which can help to rapidly identify links between epidemiologically connected cases of pneumococcal disease and clarify the composition of outbreak clusters. It does however have limitations when the cases are due to uncapsulated pneumococci which are “non-typeable” by serotyping. In such cases, molecular methods can be more significant in identifying or refuting links between cases.

8.2.2 Molecular typing

Molecular typing methodologies are reviewed in Chapter 1. Several of these have been used to investigate relationships between isolates collected in outbreak situations including PFGE (Nuorti *et al*., 1998, Carter *et al*., 2005, Subramanian *et al*., 2003, Martin *et al*., 2003a, Craig *et al*., 1999, Bescos *et al*., 2003, Barnes *et al*., 1995), PBP Gene fingerprinting (Gillespie *et al*., 1997, Barnes *et al*., 1995), Ribotyping (Cherian *et al*., 1994, Dagan *et al*., 2000), RFEL (De Galan *et al*., 1999), BOX-PCR (Ertugrul *et al*., 1997), Random Amplified Polymorphic DNA analysis (RAPD) (Melamed *et al*., 2002) fluorescence-based Amplified Fragment Length Polymorphism analysis (fbAFLP)(Hennink *et al*., 2006) and MLST (Birtles *et al*., 2005, Leimkugel *et al*., 2005, Clarke *et al*., 2004a, Yaro *et al*., 2006, Martin *et al*., 2003a) although several publications describing pneumococcal outbreaks predate the development of these methods. This chapter features the first description of microarray technology being utilised in the molecular typing of pneumococcal outbreak associated isolates and thereby the first to consider the clonal nature of such isolates at the level of the whole pneumococcal genome.

8.3 Background to Chosen Outbreaks

8.3.1 Serogroup 1 ST227

The significance of serogroup 1 pneumococci to the contemporary global epidemiology of this organism is reviewed in Chapter 1.

It was possible to identify isolates from two patients involved in a historical serogroup 1 related outbreak from a care of the elderly ward in Glasgow in 1996 (Thakker *et al*., 1998). These isolates were stored frozen at SMPRL. This allowed us to revisit the microbiology of the outbreak investigation with the benefit of MLST and microarray technology. Neither isolate was from the index case who had presented with an aggressive pneumonia and died within 24 hours of the onset of symptoms. Both isolates were from direct contacts of the index patient though. These secondary cases had developed bacteraemic pneumonia and the stored isolates had been obtained from blood cultures. One of these secondary cases also died within 24 hours of the onset of symptoms and the other died within 3 days of the onset of symptoms.

MLST of the two serogroup 1 isolates showed them to both be ST227. This itself is of significance as this outbreak predates the start of the expansion of the ST306 clone in

Scotland and is compatible with the hypothesis that ST227 was the dominant serogroup 1 clone in Scotland in the 1990s.

8.3.2 Serogroup 4 ST206

This outbreak occurred in 2002 in a care of the elderly ward in a hospital in the West of Scotland and affected four patients aged between 79 and 98. All the cases developed bacteraemic pneumonia and all succumbed to their infections. Isolates were available stored frozen at SMPRL from blood cultures for all four cases. Serotyping found all the isolates to be serogroup 4. These isolates had previously had MLST and all were ST206, a finding which has previously been used to illustrate the utility of MLST in outbreak investigations (Clarke *et al.*, 2004a).

Serogroup 4 pneumococci appear to have a particular association with outbreaks in elderly care facilities (Gleich *et al.*, 2000, Hansmann *et al.*, 2006, Clarke *et al.*, 2004a, Bain *et al.*, 1990, Kludt *et al.*, 1997).

8.4 Microarray CGH analysis of Outbreak Isolates

Unlike the microarray investigations so far which have been directed at identifying diversity in the pneumococcal genome, this chapter is focussed on assessing whether outbreak clinical isolates can be shown to be clonal. Hackenbeck *et al*, demonstrated that there is less diversity in isolates of the same PMEN clone than from different pneumococcal serogroups (Hakenbeck *et al.*, 2001). Isolates which are epidemiologically related in an outbreak setting are likely to be clonal, usually from a point source or index case. However, as the pneumococcus readily undergoes genetic recombination, the circulating population of pneumococci which do not form part of the outbreak may no longer demonstrate as clonal a population structure as the outbreak strains, allowing a distinction between the two populations to be made. CGH performed on a PCR product microarray should therefore be able to distinguish outbreak strains from those which are not related.

Microarrays have previously been used during outbreak investigations of rotavirus (Chizhikov, 2002), viral haemorrhagic fever (Palacios *et al.*, 2007), *Streptococcus pyogenes* (Smoot *et al.*, 2002) and *Campylobacter jejuni* (Leonard *et al.*, 2003) but this has been their first use in the investigation of outbreak related invasive pneumococcal disease. The investigation using outbreak related strains of *Streptococcus pyogenes* and

Campylobacter jejuni also utilized a PCR product "spotted" microarray which was able to successfully discriminate strains belonging to separate outbreaks whereas the rotavirus and viral haemorrhagic fever studies used oligonucleotide microarrays.

For these studies, genomic DNA was prepared and DNA CGH hybridizations and analysis was performed using the protocols in Chapter 2 with TIGR4 DNA as the reference genome. Dye swap experiments were performed for each isolate.

8.4.1 CGH Results from Serogroup 1 Outbreak

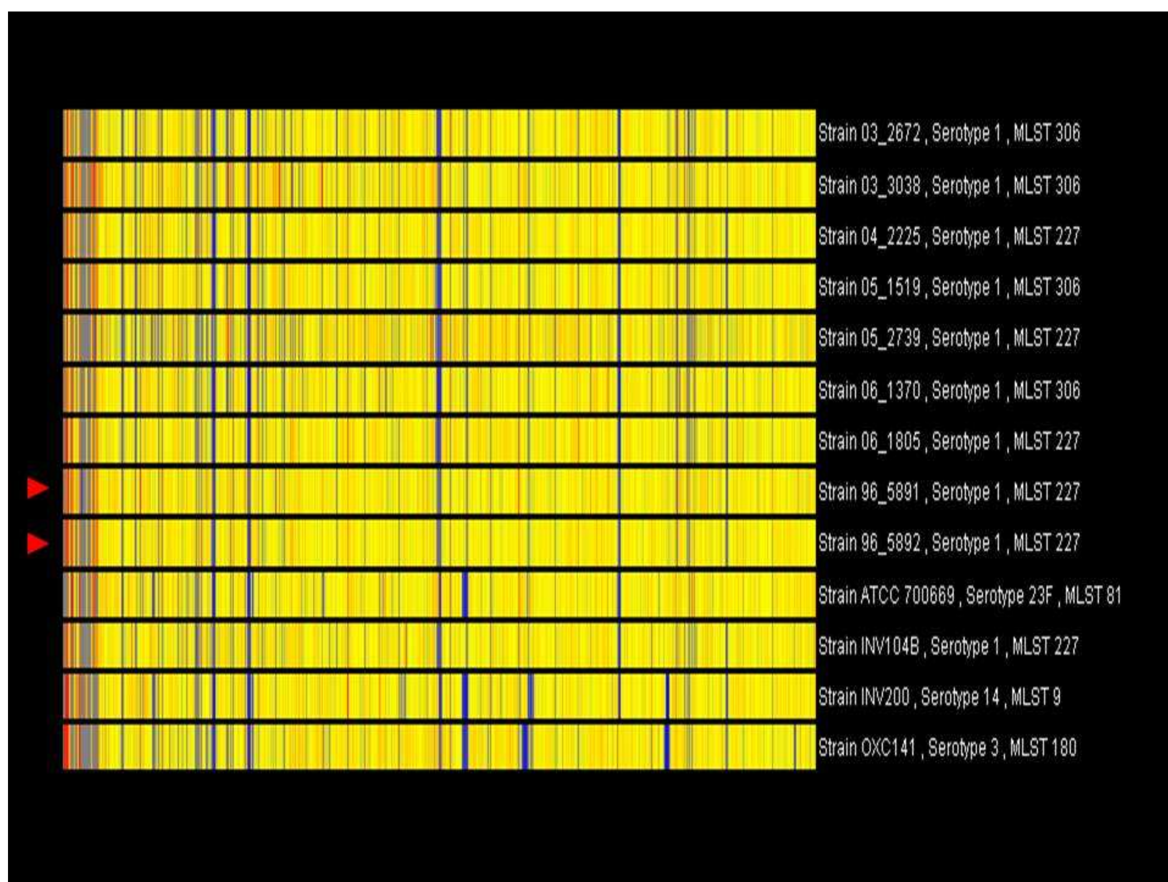


Figure 8-1 Comparison of CGH results for serotype 1 outbreak associated isolates.

Genes are represented consecutively by Genespring GX 7.3.1 from left to right by a coloured line. Yellow indicates that competitive hybridization occurred with both DNA from the test isolate and TIGR4. Blue lines indicate that DNA from TIGR4 hybridized but DNA from the test isolate did not. Grey lines indicate that either no hybridization occurred or that the fluorescence intensity of the hybridization that did occur was too low to pass the threshold value in Bluefuse for Microarrays 3.5 © to be included in the analysis. Red lines indicate that no DNA from TIGR4 hybridized but DNA from the test isolate hybridized – usually to probes from the R6 genome but occasionally because of false negative hybridization of TIGR4 DNA to a TIGR4 gene. Red arrows indicate the two outbreak related serogroup 1, ST227 isolates which can be easily distinguished from isolates of other serogroups which have been sequenced at the Wellcome Sanger Institute (Chapter 3) but which are not easily distinguished from other serogroup 1 isolates (Chapter 10).

Serotype 1 Outbreak	
96-5891	96-5892
SP0067	SP0067
SP0068	SP0068
SP0069	SP0069
SP0070	SP0070
SP0071	SP0071
SP0072	
	SP0074
SP0076	
SP0080	
SP0093	
SP0110	SP0110
SP0111	SP0111
SP0112	SP0112
SP0113	SP0113
	SP0124
SP0126	
SP0244	
	SP0308
SP0350	SP0350
SP0351	SP0351
SP0352	SP0352
SP0353	
SP0354	SP0354
SP0355	SP0355
	SP0356
SP0357	SP0357
SP0358	SP0358
SP0359	SP0359
SP0360	SP0360
SP0452	
SP0460	SP0460
SP0461	SP0461
SP0462	SP0462
SP0463	SP0463
SP0464	SP0464
SP0465	SP0465
SP0466	SP0466
SP0467	SP0467
SP0468	SP0468
SP0509	
SP0535	
SP0548	
SP0560	
	SP0569
SP0570	
SP0574	
SP0696	
	SP0773
SP1036	
SP1037	
SP1047	SP1047
SP1048	SP1048
	SP1049
SP1050	SP1050
SP1051	SP1051
SP1052	SP1052
SP1053	SP1053
SP1054	SP1054
SP1055	SP1055
SP1056	SP1056
SP1057	SP1057
SP1059	SP1059
SP1060	SP1060
SP1061	SP1061
SP1062	SP1062
SP1063	SP1063
SP1064	SP1064
	SP1143
SP1144	SP1144
SP1145	SP1145
	SP1146
	SP1188
	SP1189
SP1221	SP1221
SP1222	SP1222
SP1304	
SP1318	
SP1336	
SP1439	
SP1503	
SP1615	SP1615
SP1616	SP1616
SP1617	SP1617
SP1618	SP1618
SP1619	SP1619
SP1620	SP1620
SP1621	SP1621
SP1622	SP1622
SP1791	
SP1796	SP1796
SP1797	SP1797
SP1798	SP1798
SP1799	SP1799
	SP1819
	SP1835
	SP1866
SP1948	SP1948
SP1949	SP1949
SP1950	SP1950
SP1951	SP1951
SP1952	SP1952
SP1953	SP1953
SP1954	SP1954
SP1955	

Serotype 1 Outbreak	
96-5891	96-5892
	spr0105
spr0111	spr0111
spr0112	spr0112
spr0113	spr0113
spr0114	spr0114
spr0115	spr0115
spr0116	spr0116
spr0117	spr0117
spr0118	spr0118
spr0119	spr0119
spr0225	
	spr0317
spr0320	spr0320
spr0321	spr0321
spr0322	spr0322
spr0323	spr0323
	spr0957
spr0958	spr0958
	spr0959
spr0962	
spr0965	
	spr0972
spr1403	spr1403
spr1404	spr1404
spr1550	spr1550
spr1618	spr1618
spr1619	spr1619
spr1620	spr1620
spr1621	spr1621
spr1818	

Figure 8-2 Genelists created from CGH analysis of dye swap experiments for serotype 1, ST227 isolates 96-5891 and 96-5892.

The list of genes from the TIGR4 genome which were identified as not hybridizing to the array in one or both isolates is on the left and the list of genes from the R6 genome which did not hybridize is on the right. Where results have been highlighted in colour, PCR of the gene was performed using the primers displayed in Appendix 3 used to make the PCR probe for that gene on the microarray (MWG Biotech AG, Germany). Yellow indicates that a PCR product of the correct size could be identified by PCR and gel electrophoresis. Blue indicates that no PCR product was identified.

8.4.2 Discussion of CGH Results for Serogroup 1 Outbreak

These results again demonstrate some limitations of performing CGH on this "spotted" microarray particularly if trying to use the results as a "typing" tool (see also Chapter 3). In some cases when the array identified genes which had not hybridized, PCR corroborated with the array adding confidence to the conclusion that the genes were truly absent from the test genome (highlighted blue in Figure 8-2). However, in a substantial number of instances, PCR identified the gene as present in both isolates 96-5891 and 96-5892 (highlighted yellow in Figure 8-2) indicating that false negative hybridization for these genes had occurred in the microarray dye swap experiments. Possible reasons for this are identified and discussed in Chapter 3.

Many of these false negative hybridizations are occurring within known regions of diversity in the pneumococcal genome (Table 1-2) which suggests that the lack of hybridization may be due to the presence of the genes in the test genome having such a dissimilar sequence to that in TIGR4 that hybridization does not occur with the probe on the microarray. This issue of poor sensitivity does not necessarily mean that microarray based DNA CGH could not be used to group outbreak isolates into a cluster as it can readily distinguish between serogroups but the discrimination between ST227 and ST306 (which are part of the same clonal complex within serogroup 1 and are double locus variants as demonstrated in Figure 8-3) is poorer and may not be sufficient to distinguish a hybridization pattern specific to outbreak related isolates for this serogroup.

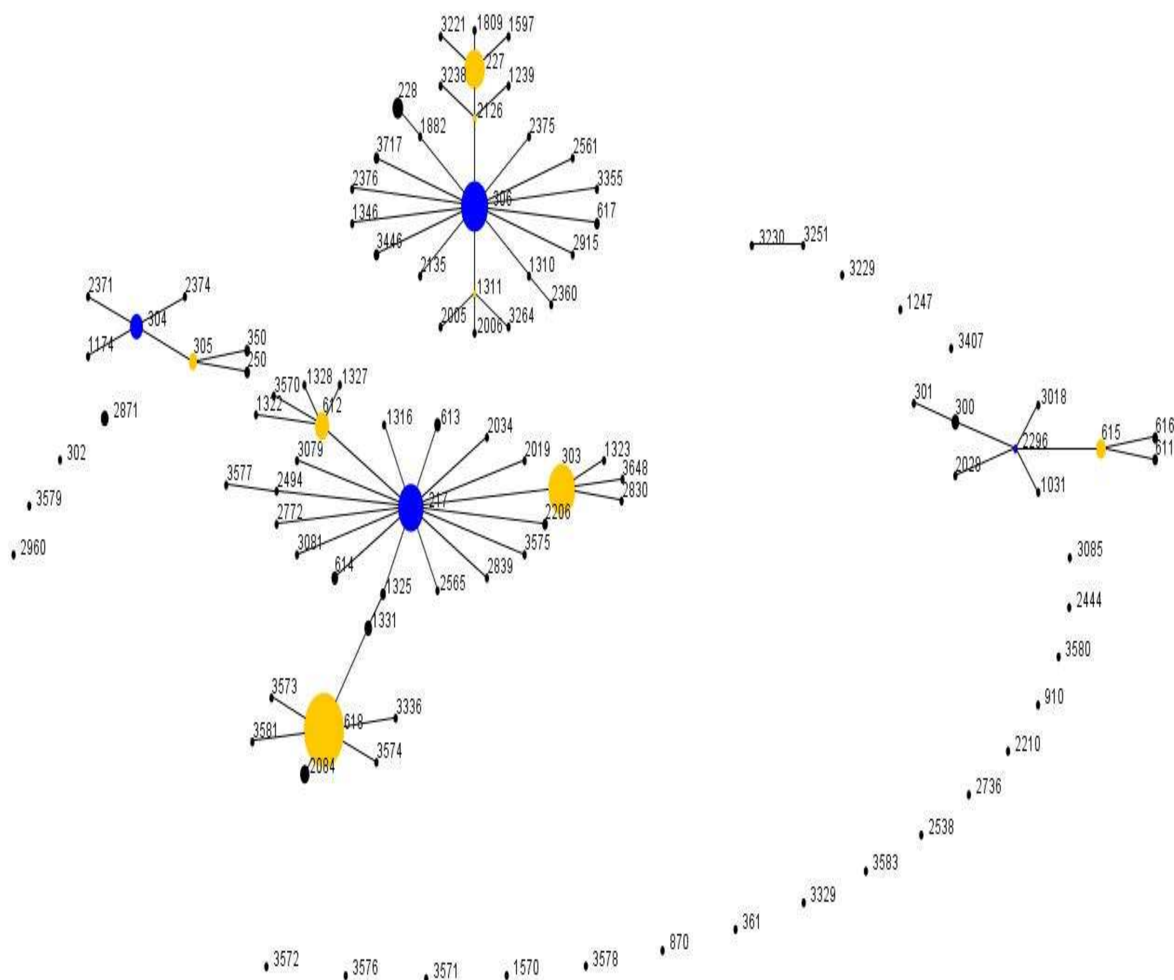


Figure 8-3 Relationships of MLST sequence types constructed using eBURST version 3[9] of all isolates in the MLST database which express serotype 1 capsule.

Blue dots indicate STs which are founders of clones from which single locus variants are demonstrated in black. Yellow dots indicate STs which are subgroup founders. ST306 is a founder clone with ST227 a subgroup founder which is a double locus variant of ST306.

[9] http://spneumoniae.mlst.net/eburst {accessed 20th December 2008}

8.4.3 CGH Results from Serogroup 4 Outbreak

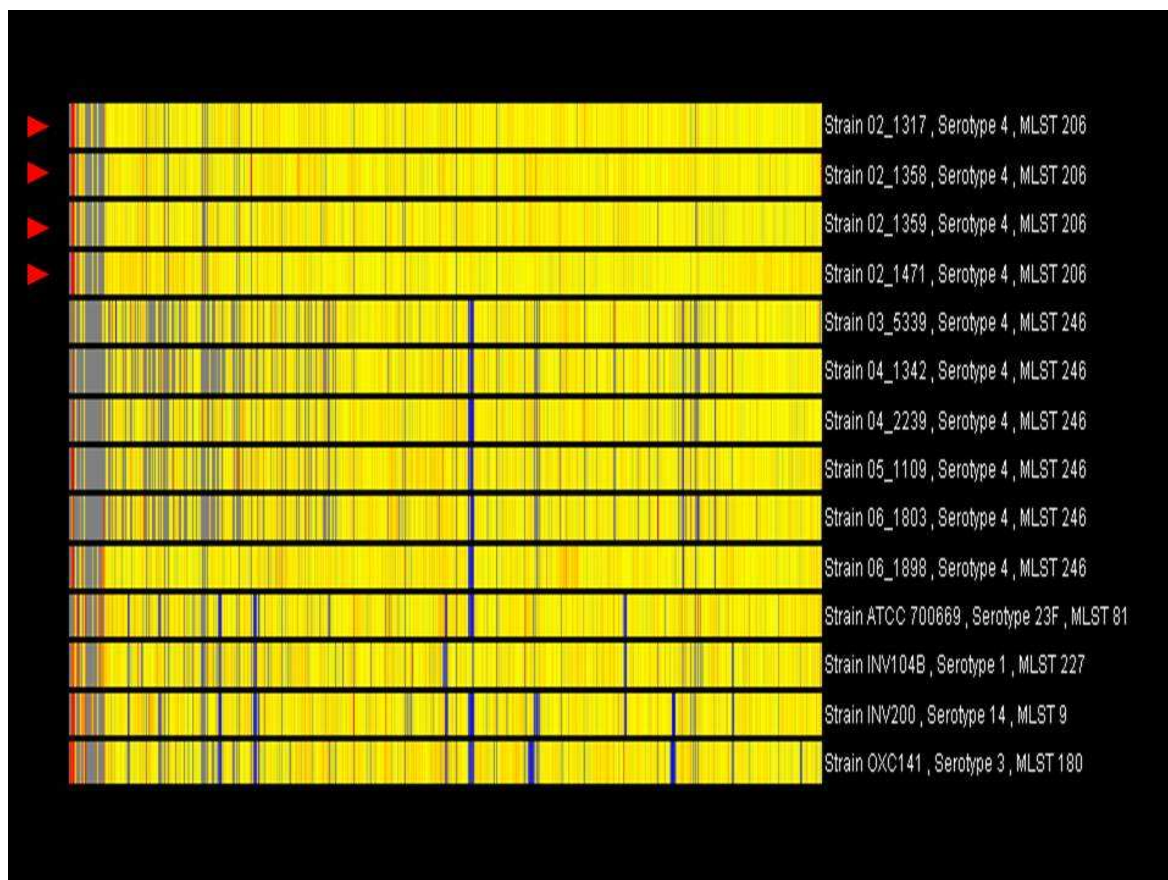


Figure 8-4 Comparison of CGH results for serotype 4 outbreak isolates.

Genes are represented by Genespring GX 7.3.1 consecutively from left to right by a coloured line. Yellow indicates that competitive hybridization occurred with both DNA from the test isolate and TIGR4 hybridizing. Blue lines indicate that DNA from TIGR4 hybridized but DNA from the test isolate did not. Grey lines indicate that either no hybridization occurred or that the fluorescence intensity of the hybridization that did occur was too low to pass the threshold value in Bluefuse for Microarrays 3.5 © to be included in the analysis. Red lines indicate that no DNA from TIGR4 hybridized but DNA from the test isolate hybridized to probes from the R6 genome. Red arrows indicate the four serogroup 4, ST206 isolates which can be easily distinguished from the ST246, serogroup 4 isolates tested (Chapter 7) and isolates of other serogroups which have been sequenced at the Wellcome Trust Sanger Institute (Chapter 3).

Serotype 4 ST206 Outbreak			
02-1471	**02-1317**	**02-1358**	**02-1359**
		SP0059	
SP0076			SP0076
	SP0113		
SP0114			
SP0115	SP0115	SP0115	SP0115
SP0116			
	SP0132		
		SP0191	
			SP0304
SP0431			
		SP0452	
		SP0512	
SP0548			
SP0600			
		SP0683	
SP0815			
	SP0906	SP0906	
	SP0908		
		SP1058	
	SP1579		
		SP1866	
			SP1921
spr0104	spr0104	spr0104	spr0104
spr0105	spr0105	spr0105	spr0105
	spr0107		
spr0112	spr0112	spr0112	spr0112
spr0113	spr0113	spr0113	spr0113
spr0114	spr0114	spr0114	spr0114
spr0115	spr0115	spr0115	spr0115
spr0116	spr0116	spr0116	spr0116
spr0117	spr0117	spr0117	spr0117
spr0118	spr0118	spr0118	spr0118
spr0119	spr0119	spr0119	spr0119
spr0587			
			spr0960
			spr1550
spr1191			
spr1193			

Figure 8-5 Genelists created from CGH analysis of dye swap experiments for serogroup 4, ST206 isolates.

The list of genes from the TIGR4 genome which were identified as not hybridizing to the array in one or both isolates is placed at the top of the list and genes from the R6 genome which did not hybridize are below these. Where results have been highlighted in colour, PCR of the gene was performed using the primers displayed in Appendix 3 used to make the PCR probe for that gene on the microarray. Yellow indicates that a PCR product of the correct size could be identified by gel electrophoresis. Blue indicates that no PCR product was identified.

8.4.4 Discussion of CGH Results from the Serogroup 4 Outbreak

Compared to the results using serogroup 1 isolates, the serogroup 4 ST206 isolates could be readily distinguished from other serogroup 4 isolates and those of other serogroups by their hybridization pattern. In addition, far fewer genes were identified as having discrepant hybridization results between the four outbreak isolates.

This CGH experiment also demonstrated an unforeseen benefit of having more than one pneumococcal genome represented on the microarray as the gene complement of the test isolates was very similar to that of TIGR4 (ST205) as they are part of the same clonal complex and are only single locus variants of each other as determined by eBURST (Figure 8-6). Consequently, the majority of genes showed no differences in hybridization between test isolate genomes and the TIGR4 control genome. Differences were however apparent between the complement of R6 related genes present in the test isolates.

Figure 8-6 e-BURST version 3.0 representation of serogroup 4 related MLST clonal complexes.

This identifies the outbreak strain ST206 as a single locus variant of TIGR4 ST205 with both lying in the same clonal complex. ST246 lies within a completely different clonal complex. Blue dots indicate that an ST is the probable founder of a clonal complex and yellow dots indicate a probable subgroup founder.

8.5 Discussion

8.5.1 Possible role for Microarrays in Public Health Outbreak Investigations

The results of both these outbreak investigations suggest that this microarray shows good discrimination between pneumococcal isolates of different serogroups and between isolates of different clonal complexes within a serogroup but it is not sensitive enough to demonstrate whether outbreak related isolates are identical in terms of genetic content. Undoubtedly such a high degree of resolution would give useful insights into how quickly pneumococcal genomes mutate and diverge from their "parent" genome *in vivo* during the direct human to human transmission usually implicated in pneumococcal outbreaks.

However this high degree of resolution would not necessarily be required for the microarray to be a useful tool to link pneumococcal isolates cultured from proven cases, suspected cases and asymptomatic carriers during an outbreak investigation and discriminate whether they form an outbreak related cluster or not. Even so, it does not offer any significant advantage over MLST for this purpose and should not be considered a replacement for MLST where such facilities exist. A further practical consideration to address regards a comparison of costs for a microarray based investigation compared to an MLST based one. Leaving aside capital costs and labour costs, the cost for a complete microarray CGH dye swap experiment is £200 which compares poorly with £30 for MLST of an isolate (Mathew Diggle, SMPRL, personal communication).

8.5.2 Genomic Diversity of Chosen Outbreak Related Strains

The CGH results for these outbreak isolates also further demonstrate diversity within pneumococcal genomes. The serogroup 4, ST206 isolate CGH results all demonstrate the presence of genes from the region of diversity within the R6 genome between spr0102 and spr0119 which was first described by Bruckner *et al* (Bruckner et al., 2004) and relates to arginine biosynthesis although PCR results for these genes suggest many of these to be false positive hybridizations and the genes to be absent by PCR. The other five regions of diversity in the R6 genome described by Bruckner *et al* appear to be present by CGH hybridization in the serogroup 4, ST206 genomes. The serogroup 1, ST227 genomes appear to contain genes from 5 of the 6 regions of diversity in the R6 genome while having greater diversity of genes identifiable in the TIGR4 genome.

9 Genomic Diversity in a Paediatric Carriage Population in the Bolivian Amazon

9.1 Background

9.1.1 Reasons for this Study

In order to investigate the genomic diversity of a defined carriage population of pneumococci, it was decided to perform a nasopharyngeal carriage study in a paediatric population as access to isolates from existing carriage studies was not possible. Rather than re-examine pneumococcal carriage in a population where the serotypes and sequence types were already well documented, an opportunity was available to integrate this study of pneumococcal carriage into a newly established programme for the prevention of hearing impairment in the Beni region of Bolivia where a substantial amount of acquired deafness in children is due to the sequelae of acute otitis media or bacterial meningitis (at least 39% of deafness in children in Beni, Bolivia results from the sequelae of bacterial meningitis (Santana-Hernandez, 2006)). Both otitis media and meningitis commonly have a pneumococcal aetiology and are potentially preventable using a pneumococcal conjugate vaccine as has been the case in Brazil (Brandileone *et al*., 2003). Consequently the project had the dual aims of providing paediatric carriage isolates of pneumococci for further investigations of their genomic diversity and of eliciting some basic epidemiological data regarding pneumococcal carriage (in a region where none existed) with a view to enabling more informed planning of strategies to prevent morbidity and mortality associated with pneumococcal disease. The mortality rate from pneumococcal meningitis is 47% in Salvador, Brazil (Ko *et al*., 2000), 37% in under five year olds in Guatemala (Asturias *et al*., 2003) and 33% in Paraguayan children (Lovera and Arbo, 2005), a substantial amount of which may be preventable by vaccination.

9.1.2 Association of Pneumococcal Carriage and Acute Otitis Media

Pneumococcal carriage can be as high as 90% in healthy children under the age of three in developing countries (Obaro and Adegbola, 2002). Pneumococci account for 35-40% of cases of acute otitis media (Echaniz-Aviles, 2001). It is estimated that pneumococcal otitis media affects 70% of children at some time and 5-10% of these develop sequelae such as

chronic otitis media with effusion, mucosal granulation, mastoiditis, ossicular erosion and fixation or cholesteatoma (Echaniz-Aviles, 2001). Commenting on the chronic sequelae of acute otitis media in Latin America, Garcia *et al* make the observation that,

> "Chronic otitis and hearing loss are common in developing countries, and reduction of this morbidity would be an important public health contribution. Further research is warranted (Garcia *et al*., 2006)."

Identification of pneumococci being carried in the nasopharynx can be used for surrogate identification of strains responsible for otitis media without having to resort to tympanic aspiration (Harper, 1999). The antimicrobial susceptibility patterns of nasopharyngeal isolates usually reflect the antibiotic susceptibility rates of invasive isolates taken during the same time period and usually demonstrate the same serotypes although the rank order of serotypes may differ between carriage and invasive disease with a smaller subset of serotypes responsible for invasive disease (Kellner *et al*., 1998).

It has been determined that nasopharyngeal carriage of pneumococci results in higher rates of onset of otitis media than those seen in non carriers (Leach *et al*., 1994, Faden *et al*., 1997). There is some suggestion that there are also racial differences in the incidence of pneumococcal otitis media (Klein, 1981).

Once established, acute otitis media can then result in higher rates of pneumococcal nasopharyngeal carriage resulting from impaired local immunity and resulting in a vicious cycle of carriage and disease (Garcia-Rodriguez and Martinez, 2002).

9.1.3 Issues regarding Pneumococcal conjugate vaccination in Latin America for the prevention of otitis media and carriage

It is known that pneumococcal conjugate vaccination (PCV) has many benefits and included among them are a reduction of nasopharyngeal carriage (O'Brien and Dagan, 2003, Obaro and Adegbola, 2002) (which is not achieved by pneumococcal polysaccharide vaccines), promotion of herd immunity (O'Brien and Dagan, 2003, Obaro and Adegbola, 2002) and a reduction in episodes of pneumococcal and non pneumococcal acute otitis media, reduction in cases of recurrent otitis media and in numbers of tympanic tube insertions (Arguedas *et al*., 2005, Echaniz-Aviles, 2001). These results although mainly observed in developed countries of Europe and North America have been replicated in some developing countries (Obaro and Adegbola, 2002, Laval *et al*., 2003).

A working group of the Pan American Health Organisation (PAHO) and the Pneumococcal Vaccines Accelerated Development and Introduction Plan (PneumoADIP) of the Global Alliance for Vaccines and Immunization (GAVI) has determined that countries in Latin America should,

> "strive to introduce the pneumococcal (conjugate) vaccine when it becomes affordable (Garcia *et al.*, 2006)."

Bolivia is eligible for GAVI vaccine fund support (Scott, 2007). The World Health Organization (WHO) also advocates implementation of pneumococcal conjugate vaccine as a priority in developing countries as part of its Global Action Plan for the Prevention and Control of Pneumonia (Greenwood, 2008). As of August 2008, PCV has not been utilised in Bolivia (Anonymous, 2008) although their current paediatric vaccination schedule comprises BCG, Polio, components of a pentavalent vaccine (Diphtheria, Tetanus, Pertussis, Hepatitis B, *Haemophilus influenzae* B), Measles Mumps and Rubella (MMR) and Yellow Fever (PAHO, 2007). Vaccine coverage is around 84% (PAHO, 2007).

However, due to the serotype constitution of the 7 valent pneumococcal conjugate vaccine (PCV-7) coverage of the serotypes causing IPD is less in Latin America than in other regions of the world (Hausdorff *et al.*, 2000a). This poorer PCV-7 coverage demonstrates a further need to elucidate regional pneumococcal epidemiology, particularly in developing countries (Camargos *et al.*, 2006). The success of PCV-7 in Africa and the USA,

> "poses a challenge to public health physicians throughout the world to quantify the burden of pneumococcal disease in their region and estimate the potential benefits of PCV use (Scott, 2007)."

9.1.4 Pneumococcal surveillance in Latin America

Surveillance of IPD has been performed in Latin America through a co-ordinated programme of the PAHO Special Program for Vaccines and Immunization (SVI) and Regional System for Vaccines (SIREVA) which was initiated in 1993 (Di Fabio *et al.*, 1997). Isolates from blood, cerebrospinal, pleural, peritoneal or synovial fluid from cases aged less than 5 years with a diagnosis of pneumococcal meningitis, bacteraemia without a focus, septic arthritis, peritonitis or the WHO clinical criteria for pneumonia are reportable (Di Fabio *et al.*, 1997) although it is recognised that IPD is frequently under-reported in young children even in developed countries (Hausdorff *et al.*, 2000a). Initially limited to Argentina, Brazil, Chile, Colombia, Mexico and Uruguay, the surveillance was extended in

1998 to several other countries in the region including Bolivia (Garcia *et al.*, 2006) with laboratory support from an international external quality assurance program from the National Centre for Streptococcus in Edmonton, Canada (Lovgren *et al.*, 2007).

The results of this surveillance demonstrate that the serotype distribution causing IPD appears to have been stable since 1993, albeit with minor regional differences (Garcia *et al.*, 2006) but that antibiotic resistance is increasing (Di Fabio *et al.*, 1997, Garcia *et al.*, 2006, Kertesz *et al.*, 1998, Di Fabio *et al.*, 2001). Molecular typing is not currently a feature of this surveillance although some smaller scale molecular typing studies have been performed in a handful of countries particularly regarding penicillin resistant pneumococci (Tomasz *et al.*, 1998). The common serotypes causing IPD in Latin America are serotypes 14, 6A, 6B, 5, 1, 23F, 19F, 18C, 19A, 9V, 7F, 3, 9N and 4 (Di Fabio *et al.*, 2001). Little is known about pneumococcal carriage in the region.

9.1.5 Location of Study

Bolivia is the poorest country in Latin America (Santana-Hernandez, 2006) and 64% of the population do not have sufficient income to cover basic needs (PAHO, 2007). The highest percentage of deaths from pneumonia (15-20%) in the under five age group in the whole of Latin America (Fuchs *et al.*, 2005) is found in Bolivia. Bolivian children also have the highest probability of dying before age five in the whole of Latin America (Fuchs *et al.*, 2005). They also have the highest infant mortality rate of around 54 per 1000 live births (Fuchs *et al.*, 2005, PAHO, 2007) and Bolivia has the highest maternal mortality in Latin America at 230 per 100,000 live births (Santana-Hernandez, 2006, PAHO, 2007). Beni region lies in the Eastern lowlands of Bolivia and is home to a dispersed population of around 407,000 people of which around 90,000 live in Trinidad, the provincial capital (Santana-Hernandez, 2006). Bordering the Beni region to the West lies the Tropical Andes region which is the greatest region of biodiversity in the world[10] with many unique species of animals and plants, although it is not known how this affects the diversity of prokaryotic life in the region.

---

[10] http://www.biodiversityhotspots.org/xp/hotspots/andes/Pages/default.aspx {accessed 11th October 2008}

Figure 9-1 Map of Bolivia.[11]

This demonstrates the locations of Trinidad and Beni within the Beni Region (red arrows).

In February 2007, the Beni region of Bolivia experienced the worst flooding in 25 years with extensive displacement of the population of Trinidad many of whom were temporarily housed in refugee camps until flood waters abated. Swabbing for this study began in May 2007 soon after these camps had been closed as people had been able to return home and children returned to school.

9.2 Materials and Methods

The project involved collaboration between Fundacion Totai (a health charity in Trinidad, Bolivia), Laboratorios Altstadt (a private clinical laboratory, Trinidad, Bolivia), SMPRL at Stobhill General Hospital, Glasgow, United Kingdom and the Faculty of Biological and Life Sciences, University of Glasgow, Glasgow, United Kingdom.

[11] www.boliviangeographic.com/boliviamap.htm {accessed 11th October 2008}

9.2.1 Specimen Collection, Storage and Transportation

The design of this carriage study was constructed in accordance with the standard method of the WHO working group (O'Brien and Nohynek, 2003) and an earlier method devised for a Latin American context[12]. Dacron polyester tipped swabs (Medical Wire and Equipment, UK) were couriered from the United Kingdom for nasopharyngeal swabbing as was Skim Milk Tryptone Glucose Glycerin (STGG) broth media (O'Brien *et al.*, 2001) which had been manufactured, sterilized and quality controlled as 1ml aliquots at SMPRL in cryotubes (Sarstedt AG & Co., Germany) to use as a short term transport and storage media at -20°C.

Microbiological media is not readily available in Bolivia so 5% horse blood agar (E & O Media Services Limited, United Kingdom) was couriered from the United Kingdom as were optochin discs (Oxoid, United Kingdom) and Transwabs (TSCswabs, United Kingdom). The use of 5% horse blood rather than blood agar with gentamicin (Converse and Dillon, 1977), colistin-nalidixic acid or colistin-oxolinic acid was a necessary deviation from the published standard method (O'Brien and Nohynek, 2003).

Nasopharyngeal swabs were taken by an experienced otolaryngologist (Dr Santana-Hernandez). If nasopharyngeal swabbing was not tolerated or not possible in younger children, oropharyngeal swabs were performed. The tips of the Dacron polyester swabs were then cut off and stored in STGG and either plated onto 5% horse blood agar on the same day or stored at -20°C until they could be cultured. After plating out and culturing the naso- or oropharyngeal secretions, alpha haemolytic colonies were subcultured onto 5% horse blood agar for optochin sensitivity testing. Optochin sensitive alpha haemolytic organisms were presumed to be pneumococci. Incubation was performed at 37°C in a carbon dioxide enriched atmosphere using candle jars at Laboratorios Altstadt, Trinidad, Bolivia.

Pure cultures of presumed pneumococci were stored at room temperature on Transwabs (TSCswabs, United Kingdom) until ready for transportation abroad. At this point they were subcultured onto 5% horse blood agar to obtain fresh pure cultures which were then inoculated onto new Transwabs for international transportation by air (Inverarity *et al.*, 2007).

---

[12] http://www.paho.org/spanish/ad/ths/ev/LABS-manual-vigilancia-serotipos.pdf {accessed 10th October 2008}

Facilities for serotyping in Latin America are sparse (Camargos *et al*., 2006). Although serotyping and antimicrobial susceptibility testing is possible in La Paz at Laboratorio Nacional de Referencia en Bacteriologia Clinica, MLST is not possible in Bolivia. Arranging international transportation of the pure isolates was fraught with difficulties. Transportation was eventually co-ordinated by a cargo and logistics company Inbolpack S.R.L., Bolivia to the Reference Laboratory for Meningococci, Madrid, Spain who kindly then redirected the isolates from Spain to SMPRL using the courier company Fedex. Transportation of isolates from Trinidad, Bolivia to Glasgow, United Kingdom took 42 days on Transwabs under conditions which were not environmentally controlled.

Blood agar with neomycin (Oxoid, United Kingdom) was used at SMPRL to resuscitate pneumococci by culturing isolates received on Transwabs for 48 hours under anaerobic conditions. Isolates which had survived transportation were further subcultured on 5% horse blood agar and stored at -80° on Protect beads (TSC Ltd, United Kingdom).

9.2.2 Epidemiological Data Collection and Analysis

In May and June 2007, 601 children were assessed as part of the Programa De Prevencion Del Deficit Auditivo En Beni and had nasopharyngeal or oropharyngeal swabs taken during this process after informed consent from a parent or guardian. This was performed at nine educational institutions (six in Trinidad, Beni and three in Riberalta, Beni). Basic epidemiological data was also collected regarding the child’s age, gender, location, number of people living in their house (particularly other inhabitants under two years of age or aged two to five), whether they lived with a person who smoked and whether they had received penicillin in the previous 30 days.

Statistical analysis of epidemiological data was performed by Mr Paul Johnson at the Robertson Centre for Biostatistics, University of Glasgow using R version 2.6.0 for Windows.

9.2.3 Ethical Considerations

Ethical approval for this study was received in the United Kingdom via the NHS Research Ethics Committee approval process (REC Ref 06/S0704/6) from the North Glasgow University Hospitals NHS Division East Office Research Ethics Committee and in Bolivia from the Colegio Medico De Bolivia Filial Beni.

9.3 Results

9.3.1 Epidemiological Data

202 carriers of optochin sensitive, alpha haemolytic organisms were identified demonstrating a carriage rate of 34%. Characteristics of the carriers compared to non-carriers were investigated by Mr Paul Johnson, Robertson Centre for Statistics, University of Glasgow using Odds Ratios of carriage and univariate and multivariate models. These results are displayed below in Table 9-1.

Characteristics		Carrier status, N (%)		Odds ratio (95% CI)	
		Negative (N=399)	Positive (N=202)	Univariate models	Multivariate model
Age in Trinidad (N=493)	0-5	38 (64.4%)	21 (35.6%)	1.0	1.0
	6+	275 (63.4%)	159 (36.6%)	0.6 (0.3, 1.1)	0.7 (0.4, 1.3)
Age in Riberalta (N=107)	0-5	26 (96.3%)	1 (3.7%)	1.0	1.0
	6+	59 (73.8%)	21 (26.2%)	9.3 (1.2, 72.5)	9.4 (1.2, 76.9)
Gender	Female	182 (73.4%)	66 (26.6%)	1.0	1.0
	Male	217 (61.5%)	136 (38.5%)	1.7 (1.1, 2.4)	1.6 (1.1, 2.4)
Household size	2-5	95 (66.9%)	47 (33.1%)	1.0	
	6-9	165 (65.0%)	89 (35.0%)	1.1 (0.7, 1.8)	
	10+	93 (62.8%)	55 (37.2%)	1.0 (0.6, 1.7)	
Children aged 2-5 at home	Yes	154 (62.1%)	94 (37.9%)	1.0	
	No	205 (68.1%)	96 (31.9%)	0.8 (0.6, 1.1)	
Children aged < 2 at home	Yes	130 (64.4%)	72 (35.6%)	1.0	
	No	229 (66.0%)	118 (34.0%)	1.1 (0.7, 1.5)	
Live with a smoker	Yes	163 (65.2%)	87 (34.8%)	1.0	
	No	195 (65.9%)	101 (34.1%)	1.0 (0.7, 1.4)	
Residence	Riberalta	85 (79.4%)	22 (20.6%)	1.0	1.0
	Trinidad	313 (63.5%)	180 (36.5%)	1.9 (0.8, 4.5)	15.2 (1.6, 146.3)

Table 9-1 Odds ratios for pneumococcal carriage risk factors estimated from univariate and multivariate models.

In both univariate and multivariate models the clustering of samples within schools was accounted for by fitting a mixed effects logistic regression model with random intercepts. The multivariate model was selected by backwards deletion of fixed effects that did not significantly improve the model based on a likelihood ratio test. The final model contained fixed effects of age, gender and residence and an interaction between age and residence. Odds ratios are expressed with 95% confidence intervals in brackets.

9.3.2 Antibiotic Resistance

Of the 202 isolates identified as optochin sensitive and alpha haemolytic in Bolivia, only 54 were viable when they arrived in Glasgow after delays in transit *ex vivo* in sub-optimal

conditions. Of these 54 isolates, very little antimicrobial resistance was detected as shown in Table 9-2 below. Further details regarding these isolates are displayed in Appendix 2.

Antimicrobial Agent	Number of Fully Sensitive Isolates	Number of Intermediately Sensitive Isolates	Number of Resistant Isolates	% Sensitivity
Penicillin	50	4	0	92%
Erythromycin	53	1	0	98%
Vancomycin	54	0	0	100%
Chloramphenicol	54	0	0	100%
Tetracycline	52	0	2	96%
Cotrimoxazole	46	8	0	85%

Table 9-2 Antibiotic sensitivity for the 54 optochin sensitive Bolivian isolates.

Of the 4 isolates with intermediate sensitivity to penicillin which were found, 2 were serotype 17F (ST2973 and ST3267), 1 was serotype 24F (ST 3770) and 1 was non-typeable. The non typeable isolate also had intermediate sensitivity to erythromycin, cotrimoxazole and was resistant to tetracycline. The MLST of this non-typeable isolate also proved to be atypical with only one known allele being identifiable. The optochin sensitivity of the isolate was subsequently reviewed and found to be smaller (18mm) when grown in a carbon dioxide enriched atmosphere compared to air (25mm) leading to the conclusion that this isolate taxonomically was a *Streptococcus pseudopneumoniae* rather than a pneumococcus according to the definition suggested by Keith and Murdoch (Keith and Murdoch, 2008).

The highest rate of antibiotic non-susceptibility was to cotrimoxazole which showed intermediate sensitivity in 8 isolates (serotypes 23A (ST2974), 24F (ST3770), 34 (ST1902), 4 (ST332), 6B (ST4015) 16F (ST3771), 16A (ST4016).

9.3.3 Serotyping of Bolivian Isolates

The results of serotyping the 53 surviving pneumococcal isolates are illustrated as a pie chart in Figure 9-2 below.

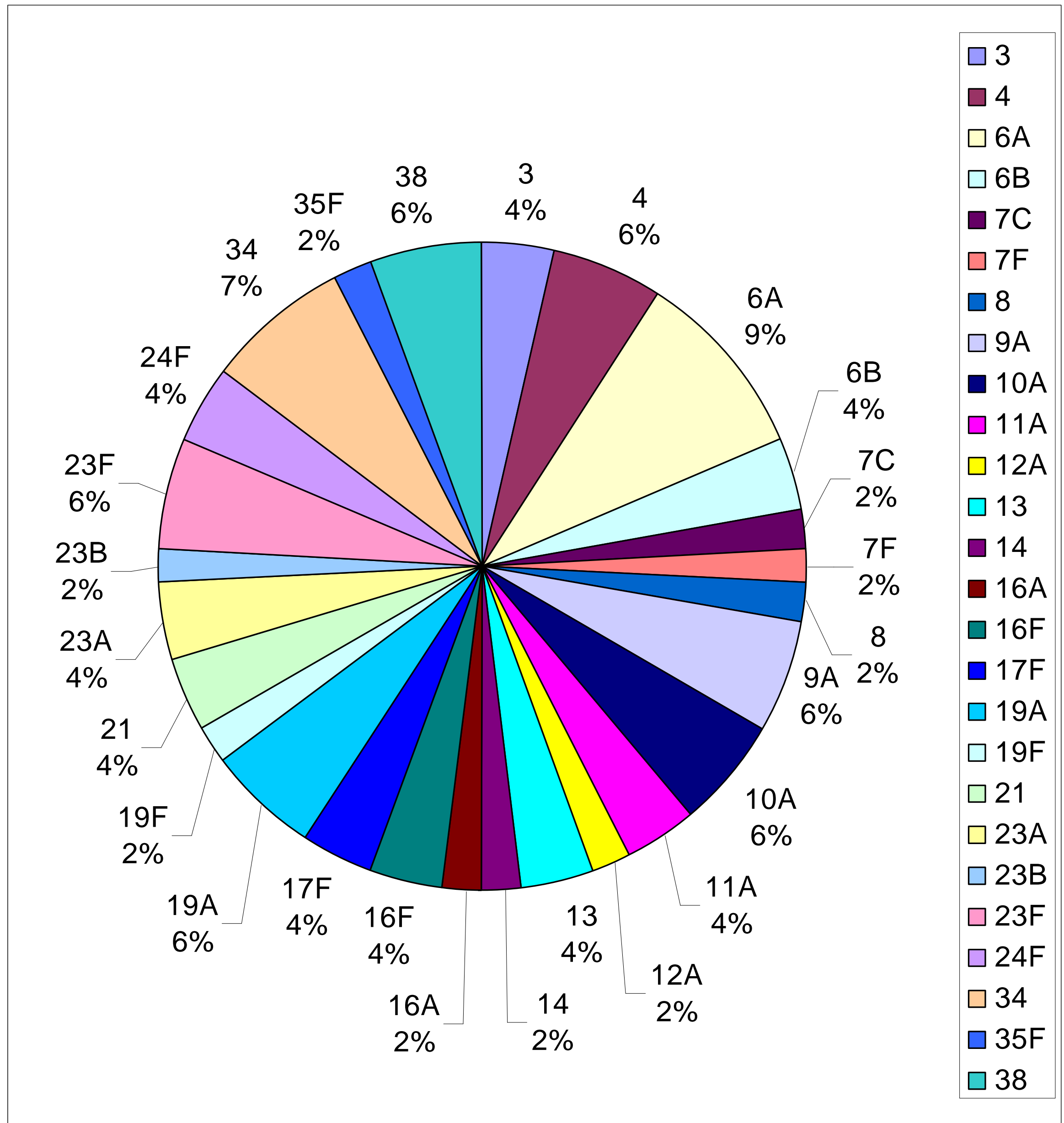


Figure 9-2 Serotype distribution of 53 Bolivian pneumococcal carriage isolates.

9.3.4 MLST of Bolivian Isolates

The majority of the pneumococcal isolates recovered had new, unique MLST profiles. Details of the MLST profiles for the 32 newly discovered sequence types in Bolivia are displayed in Table 9-3.

Number of isolates	Sequence Type	aroe	gdh	gki	recP	spi	xpt	ddl
1	2973	13	5	4	1	8	14	7
3	2974	1	8	2	9	6	4	6
1	2975	2	5	4	1	6	1	1
1	2976	16	5	9	8	8	14	7
1	3267	7	5	4	1	6	25	7
1	3429	7	5	62	5	6	4	14
1	3430	12	5	5	5	6	3	5
1	3431	12	5	87	1	6	1	8
1	3432	7	17	4	16	6	7	17
1	3509	8	5	1	5	8	14	7
1	3534	2	16	89	1	10	79	28
1	3535	16	10	4	32	6	14	6
1	3536	7	13	4	6	3	6	8
1	3537	15	17	4	16	6	26	8
1	3538	7	25	4	5	15	20	18
1	3539	7	5	62	16	6	79	14
1	3540	1	5	5	5	6	3	5
1	3767	5	5	1	5	8	6	7
1	3768	15	5	5	5	1	1	14
1	3769	1	5	4	18	32	48	8
2	3770	7	5	8	6	6	37	8
1	3771	15	5	9	5	6	3	6
1	3852	16	10	4	1	6	14	5
1	3853	1	10	4	5	15	20	18
1	3854	2	16	4	1	10	79	28
1	3855	25	17	4	16	6	96	17
1	3856	7	5	4	5	6	37	1
1	3857	12	5	2	16	6	37	1
1	3858	21	33	62	1	10	28	15
2	4015	7	47	1	2	51	1	14
1	4016	7	47	1	2	51	7	14

Table 9-3 MLST profiles for the newly discovered sequence types currently unique to carriage isolates from Bolivia.

Of the sequence types which were already defined and identifiable on the pneumococcal MLST database, the previously documented serotype and geographical associations are entered in Table 9-4.

MLST	Associated Serotype(s)	Associated Locations	Disease Phenotype
923	13	The Gambia	Not known
180	3	Begium, Brazil, Canada, Denmark, Germany, Italy, Poland, Portugal, South Korea, Spain, Sweden, Taiwan, The Netherlands, UK, USA.	Carriage and Invasive
191	7F	Brazil, Denmark, Italy, Finland, Germany, Hungary, Norway, Poland, Sweden, Switzerland, The Netherlands, UK, Uruguay,	Carriage and Invasive
239	6, 9V, 20	Hungary, Poland, UK	Invasive
280	9V	Brazil, Vietnam	Carriage and Invasive
332	20	Norway	Invasive
387	23F	Brazil, Vietnam	Invasive
404	8	Brazil, Italy, Poland, UK	Carriage and Invasive
776	23F	Argentina	Invasive
1150	6A	France, Portugal	Carriage and Invasive
1902	34	USA	Not known
1989	3	Germany	Invasive
2440	6A	UK	Invasive
2880	19A	Brazil	Invasive

Table 9-4 Serotype and geographical associations of sequence types identified in Bolivia, which have also been identified in other regions[13].

9.4 Discussion

It has been said that molecular epidemiological studies of pneumococcal populations in Latin America are few due to the high costs and logistical difficulties encountered in the region (Tomasz *et al*., 1998, Castanheira *et al*., 2003). This has also been the conclusion of those involved in this study. The unpredictable infrastructure which resulted in prolonged delays of these specimens in transit has undoubtedly influenced these results since about 75% of isolates did not survive international transportation. Nevertheless although this has introduced a selection bias into our sample, this does not appear to have limited the observed extent of genomic and phenotypic diversity which may actually now be under-represented in this remaining collection of isolates. This introduction of bias should be considered when interpreting these results. It is also worth considering that these results from children may not necessarily be extrapolated to an adult population (Borer *et al*., 2001).

[13] http://spneumoniae.mlst.net/ {accessed 9th December 2008}

9.4.1 Epidemiological Data

It is well known that pneumococcal carriage is highest in the under two age group. In Papua New Guinea, carriage rates as high as 60% have been observed in neonates with carriage of the same serotype ranging from 5 to 290 days (Gratten *et al*., 1986). In Latin America, pneumococcal carriage in pre-school children can also reach 60%, falling to 35% in primary school age children and reaching 25% in high school aged children (Echaniz-Aviles, 2001). The majority of the children who were swabbed fall into the primary school age category and so our finding of a carriage rate of 34% is entirely compatible with this previous observation. Although we did not swab adults, it is recognised that pneumococcal carriage in adults without young children is only around 6% whereas the carriage rate in adults with young children is as high as 18-30% (Echaniz-Aviles, 2001). It has also been postulated that pneumococcal transmission tends to be between young siblings and peers rather than from adults to children (Lloyd-Evans *et al*., 1996). It is difficult to account for the higher odds of carriage among boys identified by this study although it may relate to different behaviour or standards of personal hygiene.

In a study of pneumococcal carriage in a Brazilian urban slum, Reis *et al* recently found no association between the number of household members (used as an indicator of overcrowding) and pneumococcal carriage. This was a surprise finding as overcrowding has been considered to be a risk factor for pneumococcal carriage. This study in Trinidad and Riberalta also found no association between number of household members and pneumococcal carriage, despite some very densely populated dwellings (some children living with up to 30 people) resulting from population displacement from flooding. It is possible that overcrowding aids transmission when carriage rates are low and that a threshold may be reached over which it is less of an influence because the density of pneumococcal carriage is already high. However, Reis *et al* did identify school attendance as an independent risk factor for pneumococcal carriage (Reis *et al*., 2008). All the children which we swabbed attended school or kindergarten which might be influencing the carriage rate seen in this study. It has not been possible to adequately account for the higher odds of carriage in older children from Riberalta and this requires further consideration.

9.4.2 Implications for Otitis Media

Serotypes 1, 3, 5, 12F, 19A and 19F are associated with otitis media (Shouval *et al*., 2006). These serotypes are also frequently responsible for IPD in Latin America and

unsurprisingly some of these featured in the nasopharyngeal isolates of this study. It is also unsurprising, that given this combination of factors, otitis media is a significant manifestation of pneumococcal disease in Beni. Two of the children who carried pneumococci and whose isolates survived transportation had ear infections. Isolate 07-2827 (serotype 13, ST923) was associated with acute otitis media. Isolate 07-2839 (serotype 9A, ST239) was associated with chronic otitis media (see Chapter 4).

It has been proposed that early age of infection and multiplicity of bacterial types may contribute to prolonged bacterial carriage and to Eustachian tube damage resulting in otitis media (Leach *et al*., 1994, Faden *et al*., 1997). The results of this study in Bolivia are consistent with this proposition.

Although PCV-7 prevents otitis media due to vaccine included serotypes, there is evidence that there is poorer activity against serotype 19F which can persist in middle ear fluid (Regelmann, 2005). That observation and the finding of serotype 19F in this collection of carriage isolates suggests PCV-7 may not prove very effective in preventing otitis media in Beni.

9.4.3 Antibiotic Resistance

One striking observation regarding this Bolivian pneumococcal strain collection is the low level of antimicrobial non-susceptibility. It is possible that this has resulted from a survival advantage with antimicrobial susceptible pneumococci able to survive for longer *ex vivo,* although such a phenomenon has not previously been described. Another plausible explanation may relate to the relative geographical inaccessibility of Trinidad and Riberalta compared to other Bolivian towns as it is only in the last decade that Trinidad has had a main road connecting it to larger cities while Riberalta remains accessible mainly by the air. These may be barriers to substantial mixing of pneumococcal populations, which may so far have prevented the introduction of antimicrobial resistant clones of pneumococci.

There is some published data relating to antimicrobial resistance among pneumococci causing invasive disease in Bolivia. From 2000-2003 from surveillance of a small sample of IPD isolates (n=45) penicillin resistance was as high as 31%, erythromycin resistance was 13%, chloramphenicol resistance was 13% and cotrimoxazole resistance was 22% (Anonymous, 2004). This is compatible with surveillance of IPD isolates from other Latin American countries as penicillin resistance in IPD isolates varies from 2% in Brazil to

21.1% in Mexico (Camargos *et al*., 2006). In Peru it has been seen that the penicillin resistance rate is higher (42.9%) in IPD isolates than pneumococcal carriage isolates (15.1%) (Cullotta *et al*., 2002) so this could explain the difference in antimicrobial non-susceptibility demonstrable in our carriage isolates compared to higher rates from IPD isolates assessed as part of national surveillance.

Regional differences within countries affecting antimicrobial non-susceptibility rates are well recognised and have been observed in Brazil (Reis *et al*., 2008) and Chile (Inostroza *et al*., 1998). This may also be a factor influencing our results. Bolivia is a country of vast contrasts. The capital, La Paz, where the national clinical microbiology reference laboratory is located is on the Andean Highland Plateau while Beni is in the lowland Amazonian basin. These contrasting geographical locations are mirrored by climatic differences and ethnic differences in the human population all of which may be influencing the pneumococcal population structure. In Vietnam for instance, higher levels of penicillin resistance in pneumococci are associated with living in urban rather than rural settings (Quagliarello *et al*., 2003, Parry *et al*., 2000) which may be the case in Bolivia as Trinidad and Riberalta which are not large towns. (The population of Trinidad is around 90,000 unlike the capital city, La Paz, with a population of around 1,600,000.)

In this collection of strains from Bolivia, the antimicrobial with the highest rates of non-susceptibility was cotrimoxazole. This has been observed in other Latin American countries (Reis *et al*., 2008, Tomasz *et al*., 1998). Interestingly, serotype 6B is associated with cotrimoxazole resistance in Brazil (Brandileone *et al*., 1998), Colombia (Vela *et al*., 2001, Castaneda *et al*., 1998) and Mexico (Echaniz-Aviles *et al*., 1998) but of the eight cotrimoxazole non-susceptible isolates in this Bolivian collection, only one was serotype 6B.

Our results are comparable with the situation in other Latin American countries in the 1980s and early 1990s. When pneumococcal carriage isolates were investigated in Uruguay in the 1980s, intermediate resistance to penicillin was seldom observed but 37.2% were resistant to cotrimioxazole (Mogdasy *et al*., 1992). Again cotrimoxazole resistance in carriage isolates in Colombian children from 1993-94 was 40% but little erythromycin or chloramphenicol resistance was detectable (Leal and Castaneda, 1997). In Brazil from 1988-92, resistance to cotrimoxazole affected 30% of isolates but had been only 1% pre 1988 (Sessegolo *et al*., 1994). Cotrimoxazole resistance remained high in Brazil during the 1990s (Rey *et al*., 2002a, Rey *et al*., 2002b, Mendonca-Souza *et al*., 2004) and is also high in Peru at 56.9% (Cullotta *et al*., 2002). It is possible that socioeconomic factors in Beni,

poor access to antimicrobials, antimicrobial prescribing practices or geographical inaccessibility until the late 1990s have resulted in a pneumococcal population which more readily mirrors that of other Latin American countries twenty years ago rather than currently.

We observed a low rate of penicillin non-susceptibility. It is noteworthy that surveillance in the Brazilian provinces that neighbour Beni demonstrates higher rates of penicillin nonsusceptibility. In North Brazil from 1993-96 pneumococcal penicillin non-susceptibility was 15.8%, in Central West Brazil it was 16.9% and in South Brazil it was 9.5% (Brandileone *et al*., 1998). Much penicillin non-susceptibility in Latin America is due now to a serotype 14 which has been seen in Argentina (Rossi *et al*., 1998, Albarracin Orio *et al*., 2008), Uruguay (Coffey *et al*., 1999) and Colombia (Vela *et al*., 2001). We however did not see penicillin non-susceptibility affecting serotype 14 pneumococci.

In this study, macrolide resistance was rare and this has often been the case for pneumococci in Latin America (Tomasz *et al*., 1998). Erythromycin resistance in non-typeable isolates has been seen in Brazil where it was due to the MLS_B phenotype (Mendonca-Souza *et al*., 2004).

Levels of pneumococcal penicillin resistance in Latin America are increasing (Camargos *et al*., 2006). Four PMEN clones have accounted for much of this resistance – Spain ^{23F}-1, Spain ^{6B}-2, Spain ^{9V}-3 and Czech Republic ^{14}-10 (Camargos *et al*., 2006, Wolf *et al*., 2000, Castanheira *et al*., 2003, Vela *et al*., 2001). Recently the England ^{14}-9 clone has resulted in clonal expansion of macrolide resistance in Brazil (Mendonca-Souza *et al*., 2004) along with the clones Taiwan^{23F}-15 and Colombia^{5}-19 in Colombia (Tamayo *et al*., 1999, Gamboa *et al*., 2002), Mexico (Echaniz-Aviles *et al*., 2008, Gamboa *et al*., 2002) , Brazil and Guatemala (Gamboa *et al*., 2002).

It has been proposed that there is an apparent inverse relationship between the level of penicillin resistance and genetic diversity in Latin American pneumococci (Tomasz *et al*., 1998). As penicillin resistance becomes more prevalent it is usually due to particular clones which begin to dominate the pneumococcal population, reducing its diversity. Our results are compatible with this as, being at a stage where no antimicrobial resistant clones appear to be found in the pneumococcal population in Beni, there is substantial genomic and phenotypic diversity. However, Beni is an exception when compared to other Latin American regions and it is likely that clonal expansion of antimicrobial non-susceptible pneumococci may soon affect the region.

9.4.4 New MLST Profiles

Of the countries in Latin America which have been able to perform MLST on pneumococcal isolates, it is not uncommon to find several sequence types which are unique to an individual country. A search of the *S. pneumoniae* MLST database demonstrates that currently Argentina has identified 21 unique sequence types, Chile has 2 unique sequence types, Columbia has 12 unique sequence types and Uruguay has 18 unique sequence types. Nevertheless, given that there were 3768 known individual sequence types in October 2008, the proportion of previously undescribed sequence types to known sequence types in the study of this relatively small collection of pneumococcal isolates is staggering, albeit not unique. Reis *et al*, 2008 found 11 novel sequence types in carriage isolates from a Brazilian slum (Reis *et al*., 2008) and Zemlickova *et al*, found that 35 of 60 sequence types from IPD isolates from various Latin American countries had new MLST profiles (Zemlickova *et al*., 2005). Four of these accounted for 40% of the isolates showing significant geographical spread of particular clones within Latin America which were not seen in other regions of the world (Zemlickova *et al*., 2005).

It has been noted before that individual housekeeping genes are about ten times more likely to evolve by recombination than by mutation (Feil *et al*., 2000b). In keeping with this, most of the new sequence types identified in this study were new combinations of existing genes while the number of previously unidentified genes were substantially fewer.

The new sequence types which are identified in this study add to the understanding of pneumococcal population dynamics. ST 3540 is newly identified as a subgroup founder in a clonal complex where ST2883 is the existing founder. ST3430 and ST3540 are single locus variants of each other and ST776 is a double locus variant of ST3540. Consequently these three Bolivian associated sequence types (ST776, ST3430 and ST3540) all appear as part of the same serotype 23F associated clonal complex while ST2974 does not (Figure 9-3).

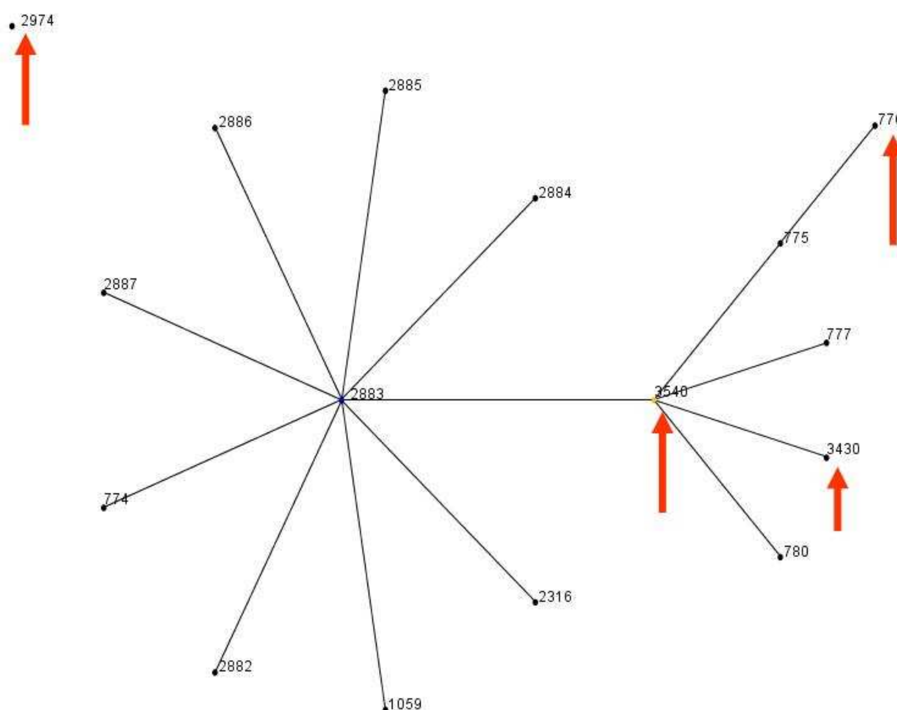


Figure 9-3 New Sequence Types associated with serotype 23F which provide new insight into their relationship within a clonal complex.

New sequence types discovered in this study are indicated by red arrows.

The association of ST3770 (serotypes 7C and 24F) and ST2974 (serotypes 23A, 23F and 38) each with multiple capsular types suggests that they are readily influenced by capsular switching even although there is no vaccine pressure to promote this.

Without virulence studies of these isolates in animal models or MLST surveillance of IPD isolates in Bolivia, it is unknown what the disease causing potential of these new sequence types may be. It is also feasible to propose that since there is such a degree of previously undescribed variation in the relatively stable housekeeping genes utilised in the MLST scheme that there will be even greater diversity in hypervariable genes in these isolates which merits further investigation.

9.4.5 New Insights into Existing Clonal Complexes

Where pre-existing sequence types could be identified in this Bolivian pneumococcal strain collection, some further insights into the existing pneumococcal clonal complexes is possible although the relative absence of MLST data from Latin America can be limiting.

Zemlickova *et al*, performed MLST on 185 IPD isolates from preschoolers in 5 Latin American countries (Zemlickova *et al*., 2005). Their study focussed on serotypes 1, 3, 5, 6B, 7F, 14 and 23F. The greatest genetic diversity was identifiable in serotypes 6B, 14 and 23F. In our series serotype 23F is very diverse with several different STs represented as has been shown in Figure 9-3 above. The association of ST776 with serotype 23F by Zemlickov *et al* was also seen in one of the Bolivian isolates (Zemlickova *et al*., 2005).

Zemlickova *et al* also identified ST191 as a common Latin American sequence type found in all 5 countries assessed which was associated with serotype 7F (Zemlickova *et al*., 2005). This association was also seen with the finding of a serotype 7F, ST191 isolate in the Bolivian strain collection. ST191 was also identifiable in Argentina in a penicillin susceptible non typeable isolate (Albarracin Orio *et al*., 2008).

ST1902 is relatively commonly seen in the Bolivian strain collection and although mainly associated with serotype 34, one isolate was of serotype 38. ST239 is a new association with serotype 9A seen in this study.

A further MLST study of pneumococcal carriage isolates was recently published from Brazil. However, none of the carriage sequence types seen in the slums of Salvador in Brazil apart from ST180 (Reis *et al*., 2008) are seen in Beni despite many of the same serotypes being seen. As reviewed in Chapter 1 and investigated further in Chapters 6 and 10, ST180 is a relatively common sequence type seen internationally associated with serotype 3 in studies of both carriage and invasive pneumocoocal isolates and so it is not a surprise to find it also in Bolivia.

What was surprising was to find ST180 existing concurrently with ST1989 in the nasopharynx of the same child while both genotypes expressed the same phenotype as serotype 3 mucoid isolates and were thereby indistinguishable by any method other than molecular typing. Multiple sequence types within a serotype at the same point in time have not been described before in the same individual. Carriage of dual or triple serotypes concurrently has been documented before (Gratten *et al*., 1989, Charalambous *et al*., 2008, Sa-Leao *et al*., 2002) but molecular analysis demonstrated homogeneous patterns of the same genotype within each serotype (Sa-Leao *et al*., 2002). Likewise, carriage of multiple sequence types at the same time has been demonstrated but it has not been clear if these isolates were from the same serotype (Oriyo *et al*., 2006). Serotype switching of a single nasopharyngeal carriage associated RFEL type has also been demonstrated over time in an individual but not of different RFEL patterns within the same serotype at the same time in

a single individual (Sluijter *et al*., 1998). Demonstrating this phenomenon again with other serotypes may be possible if multiple colony picks were used for MLST – although this would significantly increase the cost of such molecular epidemiological studies. However, it has been recently argued that single colony picks are currently acceptable for such studies (Charalambous *et al*., 2008). Demonstrating co-existent sequence types within the same clonal complex of the same serotype within an individual from a serotype known to readily cause biofilms, greatly adds to our understanding of how conditions occur *in vivo* in the human host to promote genetic exchange.

From our experience with typing this small strain collection from Bolivia though, we would strongly urge caution in extrapolating associations of clonal complexes with invasion or carriage phenotypes calculated from very different pneumococcal and host populations (Brueggemann *et al*., 2003) into geographical regions where both serotype distribution and MLST composition of the pneumococcal population are very different.

9.4.6 Common Serotypes and STs which comprise them

The serotypes which occurred most frequently in this Bolivian strain collection were serotypes 6A, 9A, 34, 23F, 10A, 19A and 38. Often these were associated with novel sequence types. By using eBURST software and the existing pneumococcal MLST database it is possible to see whether these new sequence types add to understanding of pneumococcal clonal complexes. This was the case with serotype 23F as seen in Figure 9-3. However, for the other common serotypes the Bolivian sequence types associated with them did not link to any of the currently known clonal complexes known to be associated with such serotypes. This is illustrated below for serotype 38.

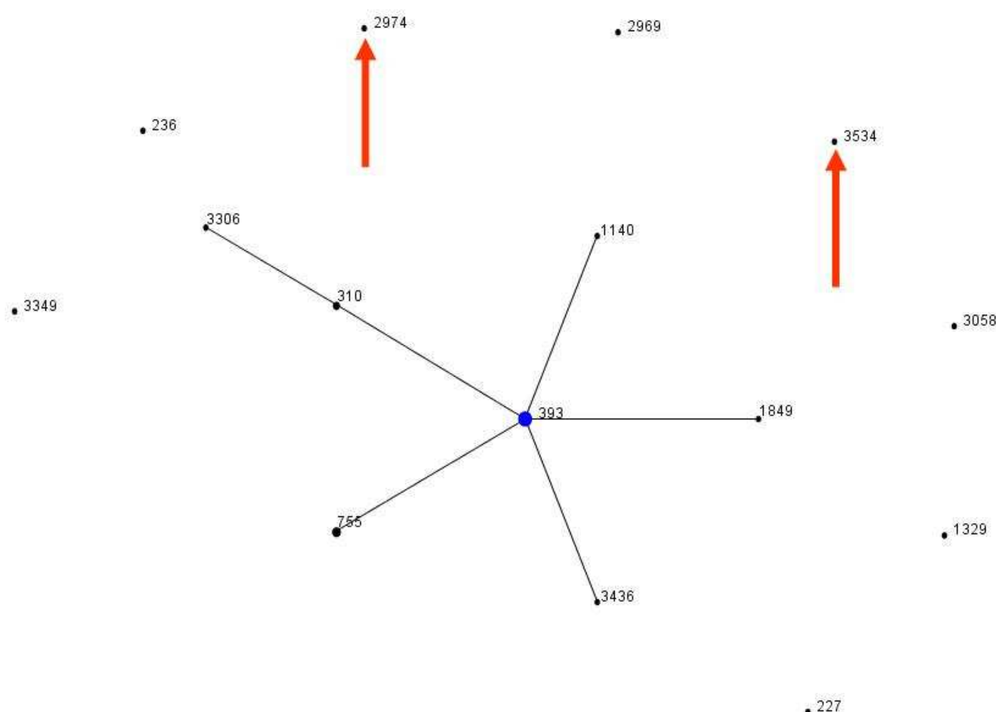


Figure 9-4 e-BURST version 3[14] diagram of serotype 38 pneumococci.

New sequence types from Bolivia are highlighted by red arrows but these do not feature as being closely related to any known clonal complexes.

9.4.7 Serotype Distributions in Latin America and Bolivia

The coverage of pneumococcal serotypes included in vaccines is usually assessed using data from surveillance of isolates obtained from cases of invasive disease as these are of greater immediate risk to health than the more diverse carriage isolates. However, carriage isolates provide a greater pool to allow capsular switching to occur and serotype replacement of vaccine escape serotypes to become more prevalent which limits the effectiveness of the vaccine in the long term. Regarding the risk of serotype replacement, the prevalence of nasopharyngeal carriage in developing countries is considerably higher than developed countries and so the probability of exposure to a non vaccine serotype is greater which facilitates serotype replacement. It should also be appreciated that serotype distribution in younger children < 6 months or 6-29 months may be different which could affect estimates of vaccine coverage depending on the age of the study population which was surveyed (Scott, 2007).

---

[14] http://spneumoniae.mlst.net/eburst {accessed 20th December 2008}

9.4.7.1 Invasive Pneumococcal Disease Serotype Distribution

IPD surveillance data from Bolivia is available from 2000-2003[15] and is displayed below in Figure 9-5.

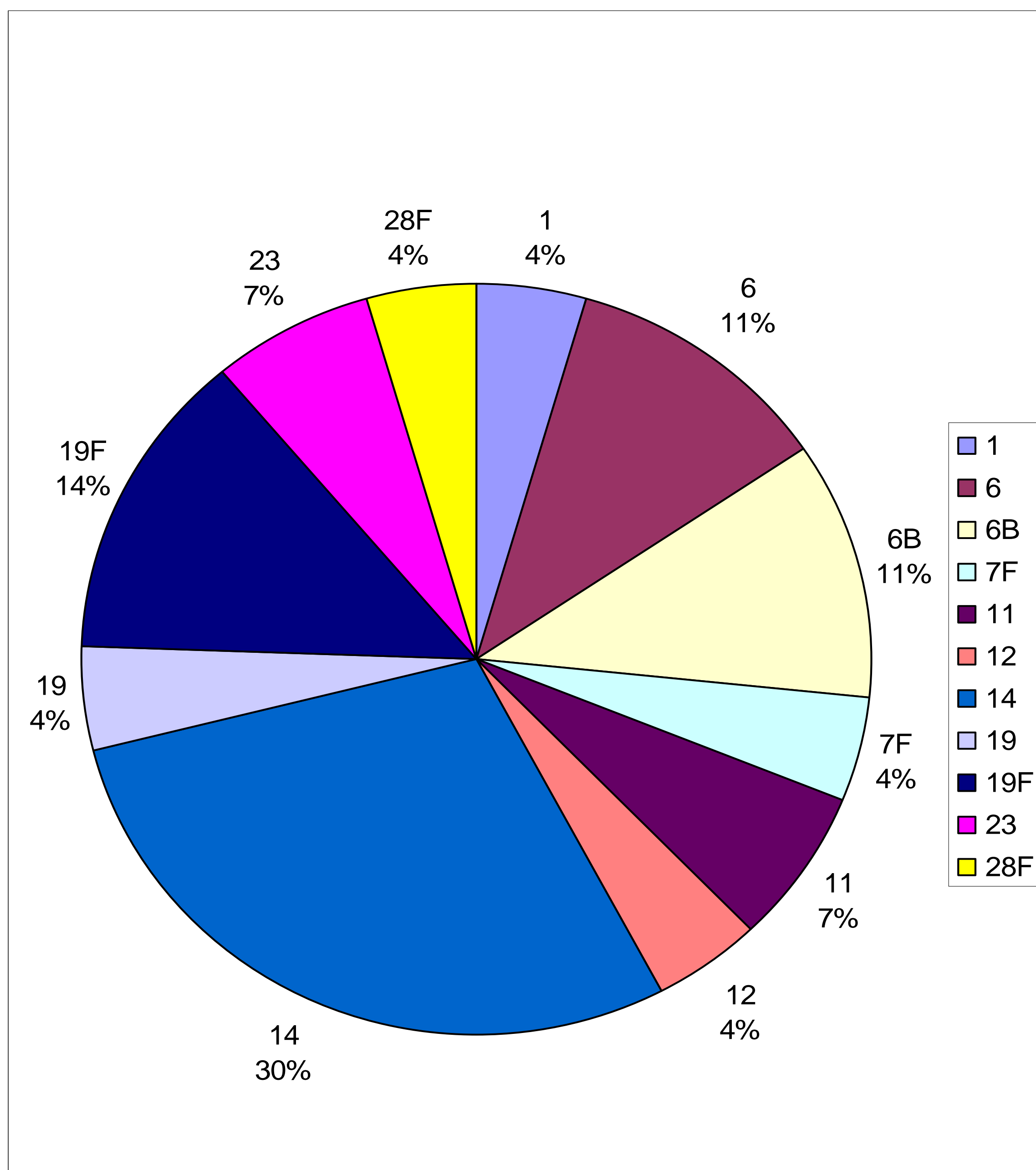


Figure 9-5 Distribution of pneumococcal serotypes causing invasive disease in Bolivia (n=45) from 2000-2003.

[15] http://espanol.geocities.com/bacterioinlasa/patresp.htm {accessed 9th December 2008}

This serotype distribution is similar to that seen previously in Bolivia (Anonymous, 2004) and elsewhere in Latin America. Common serotypes causing IPD from SIREVA surveillance in Latin America are serotypes 14, 5, 1, 6A/B, 23F, 7F, 9V, 19F, 18C, 19A and 9N (Camargos *et al*., 2006). This distribution is reproducible in country and city specific studies of IPD in Latin America although there are inevitably minor changes in the rank order as evident in Table 9-5 below.

Country	IPD Serotype Distribution	Years of Study	Reference
Argentina	14, 5, 6B, 1, 18C, 19F, 9V, 12F, 4, 23F, 7F, 6A, 33F, 9A, 3, 23B, 15, 7B, 20 and 7C	2000-2005	(Tregnaghi *et al*., 2006b, Tregnaghi *et al*., 2006a)
Argentina	14, 5, 1, 6B, 7F, 19A, 9V, 23F, 19F and 18C	1993-2003	(Vescina *et al*., 2006)
Brazil	1, 6B, 18C, 14, 5, 3, 6A, 23F, 19F and 38	1977-1988	(Taunay *et al*., 1990)
Brazil	14, 6B, 23F, 5, 19F, 6A, 1 and 4	1988-1992	(Sessegolo *et al*., 1994)
Brazil	14, 3, 6B, 19F, 6A, 23F, 18C, 4, 8, 10A, 9N and 7F	1995-1999	(Reis *et al*., 2002)
Brazil	1, 5, 6A, 6B, 9V, 14, 18C, 19F and 23F	1977-2000	(Brandileone *et al*., 2003)
Chile	1, 14, 5, 6B	1989-1993	(Levine *et al*., 1998)
Colombia	1, 4, 5, 6B, 8, 14, 19A, 19F and 23F	2006-2007	(Benavides *et al*., 2008)
Mexico	23F, 19F, 6B, 14, 19A, 6A and 9V	1996-2006	(Echaniz-Aviles *et al*., 2008)
Mexico	23F, 6B, 14, 19A, 6A, 9V, 19F, 11A, 15A, 2, 10A, 18C, 42, 16 and 22F	1992-1993	(Echaniz-Aviles *et al*., 1995)

Table 9-5 Serotypes accounting for IPD in various Latin American countries over 3 decades.

There is evidence that the regional pattern of serotype distribution causing IPD in Latin America is different from other parts of the world. Serotype 1 and 5 are among prevalent causes of IPD in Latin America and are less evident elsewhere (Di Fabio *et al*., 2001, Hausdorff *et al*., 2001, Kertesz *et al*., 1998). In Brazil serotypes 1, 5 and 18C are particularly associated with meningitis (Brandileone *et al*., 2003). This has implications for vaccine coverage when vaccine included serotypes are determined by epidemiology from North America and Western Europe. Serotypes 1 and 5 are absent from our carriage study in keeping with the view that they are only carried for short periods (Laval *et al*., 2006) and are more invasive (Smith *et al*., 1993) than other serotypes. Invasive potential does not always equate with disease severity though and it is serotypes 6, 23F, 7F, 8 and 35B which are associated with fatal outcomes from IPD in Colombia (Rios *et al*., 1999).

9.4.7.2 Carriage Serotype Distributions in Latin America

Studies of carriage serotypes in Latin America are not as common as those for IPD isolates. The results of recent examples are summarised in Table 9-6 below.

Country	Carriage Serotype Distribution	Years of Study	Reference
Brazil	14, 6B, 6A, 19F, 10A, 23F and 18C	2000-2001	(Laval *et al.*, 2006)
Brazil	19F, 6A, 23F, 18C, 34, 23B, 11A, 19A, 16F, 22F, 14, 16B, 10A, 21, 15B, 29 and 3	2000-2001	(Reis *et al.*, 2008)
Peru	23F, 6A, 6B, 15B, NT, 19F, 14, 34, 23B, 9V, 19A, 21, 4, 10A, 11A, 16F, 17F, 24F, 35F and 38	2000	(Cullotta *et al.*, 2002)
Venezuela	23F, 6A, 15B, 6B and 19F	2004-2005	(Rivera-Olivero *et al.*, 2007)

Table 9-6 Serotypes associated with pneumococcal carriage in various Latin American countries.

The distribution of serotypes represented by these recent carriage studies show great similarity to that of the Bolivian carriage isolates and there is greater similarity than there is of the Bolivian carriage isolate serotypes with the serotype distribution of IPD isolates of other Latin American countries (Table 9-5). There is also greater similarity between the carriage serotypes in Bolivia and otitis media associated serotypes in Costa Rica (serotypes 19F, 6B, 9V, 16F, 14, 23F, 3 and 6A) than there tends to be with IPD isolates (Arguedas *et al.*, 2005).

This is all in keeping with previous observations that the diversity of nasopharyngeal carriage isolates tends to be greater than that of IPD isolates (Robinson *et al.*, 2001, Takala *et al.*, 1996).

9.4.7.3 Emerging Serotypes and Bolivia

Many serotypes which are being recognised in other parts of the world and often associated with serotype replacement in response to conjugate vaccination are identifiable in this small collection of carriage strains from vaccine naïve children in Bolivia. Serotype 38 is an emerging serotype which can cause fatal meningitis (Baker *et al.*, 2005). Serotype 7F is an emerging serotype in the USA (Byington *et al.*, 2008) and possibly a colonizing serotype of short duration. Serotypes 3 (McEllistrem *et al.*, 2007, McEllistrem *et al.*, 2005), 6B (Carlisle *et al.*, 2001, Syrogiannopoulos *et al.*, 2001, Gherardi *et al.*, 2003) and 19A (Singleton *et al.*, 2007, Pichichero and Casey, 2007) are emerging serotypes and feature in the carriage population in Beni where we have identified entirely novel sequence

types associated with serotype 19A. 19A is not included in 7, 9 or 11 valent PCVs (Camargos *et al*., 2006) but is included in a new 13 valent PCV (Kieninger *et al*., 2008).

9.4.8 Implications for Conjugate Vaccine Implementation

The introduction of conjugate vaccination against *Haemophilus influenzae B* has had a significant positive impact on *Haemophilus* meningitis in Latin America (Laval *et al*., 2003) and so there are hopes that pneumococcal conjugate vaccination will have similar benefits. It is recognised that PCV-7 vaccine coverage in Latin America is poorer than that seen for USA and Europe, Africa and Oceania and this relates to the greater diversity of pneumococcal serotypes causing IPD in Latin America (particularly serotypes 1 and 5) (Hausdorff *et al*., 2000a). The coverage of carriage serotypes is even less than that for IPD isolates (Laval *et al*., 2006).

It is estimated that PCV-7 (which is active against serotypes 4, 6B, 9V, 14, 18C, 19F and 23F) covers 65% of invasive serotypes in Argentina, Brazil, Chile, Colombia, Mexico and Uruguay (Garcia *et al*., 2006, Di Fabio *et al*., 2001). But more regional surveillance studies, such as have been performed in Brazil, have estimated coverage ranging from 58% to as low as 36% (Camargos *et al*., 2006, Laval *et al*., 2006, Reis *et al*., 2008, Brandileone *et al*., 2003). Based on this data from Bolivia, PCV-7 will only cover 18% of carriage serotypes. This could theoretically rise to 28% coverage taking into account cross reactivity against the non-vaccine serotype 6A. Carriage of 6B and 19F can be reduced by 9 valent vaccine (Mbelle *et al*, 1999) and an 11 valent PCV has been under trial in Argentina and Chile (Laval *et al*, 2003). However, increasing the valency of the vaccine and introducing cover against serotypes 1, 5 and 7F only increases the coverage in Beni to 20% (30% if considering cross reactivity with 6A). A new 13 valent PCV covering serotypes 1, 3, 4, 5, 6A, 6B, 7F, 9V, 14, 18C, 19A, 19F and 23F (Kieninger *et al*., 2008) is likely to be the most useful of the PCVs if introduced in Bolivia.

There will undoubtedly be many health benefits to the introduction of PCV if it becomes affordable in Bolivia. However, in Beni, the low coverage of the PCV-7 vaccine included serotypes and the high frequency of serotypes associated with serotype replacement and capsular switching along with emerging serotypes would suggest that the usefulness of the PCV-7 vaccine may be shorter lived than has been the case in North America and Western Europe.

9.4.9 Implications for Protein Vaccines

If the high rate of new MLST sequence types and the finding of new regions of diversity by microarray CGH in Bolivian carriage isolates (Chapters 4 and 6) is an indicator of substantial, as yet unrecognised, genomic diversity in the pneumococcal genome of the isolates from Bolivia then this has implications for other genes coding for proposed protein vaccine candidates. For instance, pneumococcal surface protein A is such a vaccine candidate but the gene which codes for it is already known to show substantial variation in pneumococci from Argentina (Mollerach *et al*., 2004) and Columbia (Coral *et al*., 2001). If the diversity of such genes was also assessed in these Bolivian isolates there may be evidence for even greater diversity than was known before.

9.5 Conclusions

In conclusion, this paediatric carriage study from the Beni region of Bolivia demonstrates not just substantial phenotypic (serotype) diversity but dramatic genomic diversity in this region known for great biodiversity. This has enormous implications for the introduction and design of conjugate vaccines which, although suitable for North America and Western Europe, have much less utility in Latin America. We agree with Levine *et al* that there remains a need for ongoing bacteriological surveillance in the region (Levine *et al*., 2006).

10 Genomic Diversity and Gene Expression in Specific Invasive Pneumococcal Disease Manifestations

10.1 Background

Since the 1990s it has been realised that pneumococcal virulence is a function of both expression of capsular genes and the genotype of the pneumococcus involved (Kelly *et al*., 1994). In the initial phase of lung infection, phenotypic variants of the pneumococcal capsule have been identified by Hammerschmidt *et al* (Hammerschmidt *et al*., 2005) where the ability to adhere to epithelial cells was enhanced by limited or no capsule expression. The ability to survive in different host environments may relate to phase variation and its effect on capsule and surface protein expression (Allegrucci and Sauer, 2008).

Attempts to understand the role of genes other than capsular genes in virulence often focussed on individual genes in an arbitrary *in vitro* environment or a particular *in vivo* animal model. Genome sequencing of the pneumococcus then allowed a rapid expansion of genes identified as having a potential role in virulence and an understanding that virulence genes (microbial specific factors that contribute to the survival and growth of a pathogen *in vivo* (Hava *et al*., 2003)) could be identified,

> "specifically in nasopharyngeal colonization or bacterial sepsis or meningitis, depending on the screening tools used (Hollingshead and Briles, 2001)."

These tools were to include signature tagged mutagenesis in animal models (Hava and Camilli, 2002, Hava *et al*., 2003, Polissi *et al*., 1998, Lau *et al*., 2001), microarray studies (Orihuela *et al*., 2004b), genomic array fingerprinting (which combines microarray and random transposon mutagenesis approaches) (Bootsma *et al*., 2007), differential fluorescence induction (Marra *et al*., 2002) or *in silico* sequence comparisons of orthologous genes (Hiller *et al*., 2007). Such studies have lead to a conclusion that there are host tissue specific and serotype specific virulence factors often relating to transcriptional regulation (Hava *et al*., 2003). This suggests that focusing on serotypes biased towards colonization or invasion may identify key genes involved in virulence (Hava *et al*., 2003). Such an approach using microarrays has been termed "pathotyping" or detecting the presence of groups of several associated virulence factors which may constitute a fundamental attribute of the physiology or pathogenicity of a bacteria (Cassone

et al., 2007, Stabler *et al*., 2006, Korczak *et al*., 2005, Bekal *et al*., 2003, Pannucci *et al*., 2004) or disease state (Ehrlich *et al*., 2008).

With regard to pneumococi, in 1998 a study by Polissi *et al* found evidence that there were virulence factors specific for causing pneumonia which were not required to cause septicaemia in mice (Polissi *et al*., 1998). Using serotype 2 (D39 Xen 7) and serotype 4 (TIGR4) RNA extracted directly from laboratory infected mouse blood, rabbit CSF and tissue culture with Detroit pharyngeal epithelial cells, Orihuela *et al* demonstrated differences in expression could be identified for certain genes for these different environments which represented invasive sites and carriage respectively (Orihuela *et al*., 2004b). Extrapolation of these results to human disease is hampered by the potential confounding introduced by the species differences of the two animal models used and difficulties comparing the expression results from *in vivo* animal models with *in vitro* results from tissue culture. Using genomic subtraction followed by dot blot screening rather than a microarray approach, Pettigrew *et al*, identified two pneumococcal genes more frequently associated with middle ear infections than bloodstream infections, meningitis or carriage (Pettigrew and Fennie, 2005).

This chapter outlines an approach taken using a DNA "spotted" microarray to investigate its utility in elucidating, through identifying differences in gene complement and gene expression, the pathogenesis of serotype 1 associated pneumococcal empyema, serotype 3 associated pneumococcal meningitis and serotype 3 associated cerebral abscesses.

10.2 Serotype 1 Bacteraemic Pneumococcal Pneumonia with Parapneumonic Complications

10.2.1 Background

There has been a substantial recent increase in the incidence of pneumococcal pneumonia complicated by necrotizing pneumonia, parapneumonic effusions or empyema, particularly affecting children and occurring in developed (Byington *et al*., 2005c, Obando *et al*., 2006, Calbo and Garau, 2005, Rees and Spencer, 1997, Fletcher *et al*., 2006, Blanc *et al*., 2007, Lin *et al*., 2006, Thumerelle *et al*., 2005, Obando *et al*., 2008, Eastham *et al*., 2004) and developing countries (Kanungo and Rajalakshmi, 2001). This is in contrast to dramatic declines in the incidence of pneumococcal empyema during the mid-twentieth century (Finland and Barnes, 1978, Nowak, 1939, Weese *et al*., 1973, Bartlett *et al*., 1974,

Bechamps *et al*., 1970, Clagett, 1973, Chonmaitree and Powell, 1983, Taryle *et al*., 1978) and the high incidence at the start of the twentieth century (Glynn and Digby, 1923) which are reviewed in Chapter 1. About 10% of cases of pneumococcal pneumonia can be complicated by empyema in children (Kerem *et al*., 1994). Cases of haemolytic uraemic syndrome are also described as complications of pneumococcal pneumonia with empyema (Lee *et al*., 2006). The clinical manifestations and molecular epidemiological studies of these conditions have been particularly well described in the USA (Tan *et al*., 2002, Byington *et al*., 2002, Gonzalez *et al*., 2004, Byington *et al*., 2006, Hardie *et al*., 1996), England (Eltringham *et al*., 2003, Eastham *et al*., 2004, Fletcher *et al*., 2006, Ramphul *et al*., 2006) and Taiwan (Hsieh *et al*., 2004, Lin *et al*., 2006, Shen *et al*., 2006b). In its severest form it may result in spontaneous discharge through the thoracic wall (empyema necessitates) (Freeman *et al*., 2004) but in most cases drainage requires some form of surgical intervention (Hardie *et al*., 1996). Paediatric empyema can also be a manifestation within a serotype 1 outbreak (Gupta *et al*., 2007).

The majority of cases in the USA and England are caused by *Streptococcus pneumoniae* serotype 1 (Byington *et al*., 2006, Eastham *et al*., 2004, Ramphul *et al*., 2006, Fletcher *et al*., 2006, Byington *et al*., 2002). In addition to serotype 1, serotypes 3, 14, 19A, 5, 6B, 9V, 23F and 4 have been identified as causes of pneumococcal empyema in the United Kingdom (Sheppard *et al*., 2008, Eltringham *et al*., 2003, Ramphul *et al*., 2006, Eastham *et al*., 2002). Serogroups 14, 9, 6, 19, 18, 12 and 29 have been associated with empyema in Utah, USA (Byington *et al*., 2002, Byington *et al*., 2005a). In France, serotypes 1, 3, 5, 6B, 7F, 9V, 14, 18C, 19A and 23F have been observed (Bekri *et al*., 2007) and in Spain serotypes 1, 5, 3, 14, 19A, 7F, 6A and 9V have been identified (Obando *et al*., 2008). Serotypes 3 and 19A have also been identified as vaccine escape serotypes causing empyema in the USA (Byington *et al*., 2006, Byington, 2007, Byington *et al*., 2005a).

In a comparable case series from Taiwan, cases were predominantly caused by *S. pneumoniae* serotype 14 and serotype 1 did not feature at all (Hsieh *et al*., 2004). Similarly, preliminary results from other Asian countries (China, Indonesia, Korea, Taiwan, Thailand and Vietnam) suggested that serotypes 14 and 19 were commoner than serotype 1 from pleural fluid cultures (Kilgore *et al*., 2006) indicating geographical variations in the epidemiology of pneumococcal empyema.

In Scotland, *S. pneumoniae* serotype 1 accounts for 14% of all IPD (Diggle and Edwards, 2006) which includes cases of complicated pneumococcal pneumonia. An increase in incidence in paediatric empyema has also been documented in Scotland although the

molecular epidemiology of the pneumococcal component of this has not yet been elucidated (Roxburgh *et al.*, 2008, Roxburgh and Youngson, 2007).

Unpublished data from England and Wales demonstrate that a mixed population of serotype 1 pneumococci (ST227 and ST306) are responsible for cases of pneumococcal empyema (Robert George, Health Protection Agency, personal communication) and data from Spain (ST228, ST306, ST304 and ST2373) also shows diversity in the empyema causing serotype 1 isolates (Obando *et al.*, 2006, Brueggemann *et al.*, 2006, Obando *et al.*, 2008). In the USA, until 2003, cases were predominantly ST227 (Gonzalez *et al.*, 2004, Byington, 2007, Byington *et al.*, 2005c, Byington *et al.*, 2005b, Brueggemann *et al.*, 2003) but since 2003 serotype 1 associated sequence types have included ST306, ST304 and ST2126 (Byington *et al.*, 2008). The population dynamics of serotype 1 pneumococci causing invasive disease do suggest that it can change rapidly within a geographical region (Lamb *et al.*, 2008, Hedlund *et al.*, 2003, Henriques Normark *et al.*, 2001) and between geographical regions resulting in a diverse global population (Brueggemann *et al.*, 2003).

It is noteworthy that genetic variation in the structure of pneumolysin (a pneumococcal virulence factor) from clinical isolates of ST306 serotype 1 *S. pneumoniae* has recently been shown to be associated with different biological behaviour of the toxin (Kirkham *et al.*, 2006), indicating that genetic differences in the bacteria could account for different disease manifestations with this serotype and sequence type.

Several non-culture molecular (Eltringham *et al.*, 2003, Le Monnier *et al.*, 2006, Lahti *et al.*, 2006, Saglani *et al.*, 2005, Eastham *et al.*, 2002, Menezes-Martins *et al.*, 2005, Poulter *et al.*, 2005) and antigen based methods (Ploton *et al.*, 2006, Boersma *et al.*, 1993, Le Monnier *et al.*, 2006, Porcel *et al.*, 2007) have been used to investigate the cause of culture negative parapneumonic effusion and empyema and molecular techniques may be more useful in surveillance than culture (Sheppard *et al.*, 2008) as pneumococci can be cultured from less than 5-16% of pleural fluid specimens (Kilgore *et al.*, 2006, Obando *et al.*, 2006).

Substantial concern exists because the pneumococcal conjugate vaccine (PCV-7) presently licensed for children, Prevnar® (Wyeth Pharmaceuticals, USA), and introduced into the childhood vaccination schedule in the United Kingdom in 2006 does not prevent invasive diseases due to serotype 1 *S. pneumoniae* (Fletcher *et al.*, 2006).

In this series of experiments representative isolates stored at SMPRL which could be identified as originating from patients with parapneumonic complications of bacteraemic pneumococcal pneumonia were used. Their MLST data was reviewed and the isolates were utilised in microarray experiments using DNA CGH and RNA expression approaches. Serotype 1 isolates from patients with bacteraemic pneumonia without parapneumonic complications were identified for use as controls to establish baseline expression levels for genes which were differentially expressed in cases with parapneumonic complications.

10.2.2 Choice of Isolates for Microarray Studies

Seventeen patients from six hospitals in Central Scotland who had undergone pleurocentesis and who grew serotype 1 pneumococci from an invasive site were identified from the database at SMPRL. Of these seventeen cases, four were aged under eighteen years. MLST results showed 12/17 (71%) were ST306 and 5/17 (29%) were ST227. From these, five isolates were chosen from patients who had documented empyema or effusion complicating bacteraemic pneumonia and two separate isolates (one each of ST306 and ST227) were identified from the database from patients with bacteraemic pneumonia without parapneumonic complications to act as controls. Details of these isolates are outlined below in Table 10-1.

Isolate tested	03-2672	04-2225	03-3038	06-1805	05-2739	05-1519	06-1370
MLST	ST306	ST227	ST306	ST227	ST227	ST306	ST306
Patient Age (Years)	26	51	33	70	7	36	4
Parapneumonic complication	None	None	Empyema	Effusion	Empyema	Effusion	Effusion
Source of Bacteria	Blood	Blood	Pleural Pus	Blood	Blood	Blood	Blood

Table 10-1 Details of isolates used in microarray experiments in Chapter 10.

All 7 isolates were used in DNA CGH dye swap experiments. 5 isolates were used in RNA expression experiments (06-1805 and 05-1519 were not used).

10.2.3 DNA CGH Experiments

The results of the microarray DNA CGH experiments for serotype 1 isolates are displayed in Appendix 14 and Appendix 15.

10.2.4 RNA Expression Experiments

10.2.4.1 Microarray Results

TIGR4 Genes Upregulated in Parapneumonic Complications			
TIGR4 Gene	**Expression Level in Controls**	**Expression Level in Cases**	**Gene Function**
SP0044	1.0	3.9	Phosphoribosylaminoimidazole-succinocarboxamide synthetase
SP0045	0.9	5.5	Phosphoribosylformylglycinamide synthetase
SP0046	1.0	7.2	Amidophosphoribosyl transferase
SP0047	0.9	6.4	Phosphoribosylformylglycinamide cyclo-ligase; Phosphoribosylaminoimidazole synthetase
SP0048	1.0	6.2	Phosphoribosylglycinamide formyltransferase; 5'-phosphoribosylglycinamide transformylase 1
SP0049	1.0	5.7	vanZ protein, putative; Teicoplanin resistance protein
SP0050	1.0	5.5	Phosphoribosylaminoimidazolecarboxamide formyltransferase
SP0051	1.0	4.2	Phosphoribosylamine--glycine ligase; Phosphoribosylglycinamide synthetase
SP0053	1.0	3.2	Phosphoribosylaminoimidazole carboxylase, catalytic subunit
SP0054	1.0	4.5	Phosphoribosylaminoimidazole carboxylase, ATPase subunit; Phosphoribosyl glucinamide formyltransferase
SP0073	1.0	2.2	Conserved hypothetical protein
SP0287	1.0	3.0	Xanthine/uracil permease family protein; Conserved hypothetical protein
SP0288	1.0	2.4	Conserved hypothetical protein
SP0645	0.9	2.0	PTS system IIA component, putative; Phosphotransferase system sugar-specific EII component
SP1229	1.0	2.7	Formate--tetrahydrofolate ligase
SP1326	0.9	5.1	Neuraminidase, putative
SP1327	1.0	5.2	Conserved hypothetical protein
SP1328	0.9	3.3	Sodium:solute symporter family protein
SP1526	0.9	2.8	ABC transporter, ATP-binding protein authentic frameshift
SP1527	0.9	2.8	ABC transporter substrate-binding protein - oligopeptide transport
SP1587	1.0	2.8	Oxalate:Formate Antiporter
TIGR4 Genes Downregulated in Parapneumonic Complications			
TIGR4 Gene	**Expression Level in Controls**	**Expression Level in Cases**	**Gene Function**
SP1884	1.2	0.4	Trehalose PTS system, IIABC components; Phosphotransferase system, trehalose-specific IIBC component
SP2148	0.9	0.2	Arginine deiminase
SP2150	0.9	0.2	Ornithine transcarbamoylase
SP2151	1.0	0.2	Carbamate kinase
SP2152	0.9	0.2	Conserved hypothetical integral membrane protein
SP2153	0.9	0.2	Peptidase, M20/M25/M40 family; Conserved hypothetical protein

Table 10-2 Genes which are significantly up or down-regulated when isolates are grown to midlog under standardised conditions.

$p<0.05$ determined using ANOVA 1 way test with Benjamini and Hochberg corrections. False discovery rate set at 0.05 on data filtered to demonstrate greater than two fold differences from the line of equality (where fluorescence intensity of control and test channels are equal). Data was also initially normalized by sequence type to take account of clustering effects due to MLST. Controls were isolates 03-2672 and 04-2225 as these were not

associated with parapneumonic complications. Cases were all other isolates with parapneumonic complications. Genes highlighted in red have been associated with virulence in a serotype 4 mouse model of pneumonia (Hava and Camilli, 2002).

TIGR4 Gene	Predictive Strength	Gene Function
SP0045	6.64	Phosphoribosylformylglycinamide synthetase
SP0046	7.94	Amidophosphoribosyl transferase
SP0047	6.64	Phosphoribosylaminoimidazole synthetase
SP0048	5.74	Phosphoribosylglycinamide formyltransferase; 5'-phosphoribosylglycinamide transformylase 1
SP0049	7.17	VanZ protein, putative; Teicoplanin resistance protein
SP0051	9.55	Phosphoribosylamine--glycine ligase; Phosphoribosylglycinamide synthetase
SP0053	6.59	Phosphoribosylaminoimidazole carboxylase, catalytic subunit
SP0054	6.64	Phosphoribosyl glucinamide formyltransferase
SP0097	5.74	Conserved domain protein; Hypothetical protein
SP0239	7.23	Conserved hypothetical protein
SP0287	6.64	Xanthine/uracil permease family protein; Conserved hypothetical protein
SP0834	5.74	Hemolysin-related protein; Conserved hypothetical protein, truncation
SP0853	5.74	Hypothetical protein
SP0937	5.74	Conserved hypothetical protein
SP1069	6.64	Conserved hypothetical protein; ABC transporter substrate-binding protein - unknown substrate
SP1229	7.17	Formate--tetrahydrofolate ligase
SP1316	9.18	V-type sodium ATP synthase, subunit B
SP1486	7.23	Degenerate transposase (orf1)
SP1524	6.64	Aminotransferase, class II
SP1587	6.64	Oxalate:Formate Antiporter
SP1965	6.04	Hypothetical protein
SP1978	6.64	Diaminopimelate decarboxylase
SP1979	9.55	Pur operon repressor; Activator of purine biosynthetic genes
SP1988	5.74	Immunity protein, putative; Conserved hypothetical protein
SP2125	6.64	Conserved hypothetical protein
SP2145	7.23	Antigen, cell wall surface anchor family; Conserved hypothetical protein
SP2151	6.64	Carbamate kinase
SP2152	6.64	Conserved hypothetical integral membrane protein
SP2153	6.64	Peptidase, M20/M25/M40 family; Conserved hypothetical protein
SP2156	7.23	SPFH domain/Band 7 family; Conserved hypothetical protein

Table 10-3 Genes identified as being associated with the occurrence of parapneumonic complications when grown to midlog in Brain Heart Infusion.

Highlighted genes have been associated with virulence in a serotype 4 mouse model of pneumonia (Hava and Camilli, 2002).

10.2.4.2 Quantitative RT-PCR Results

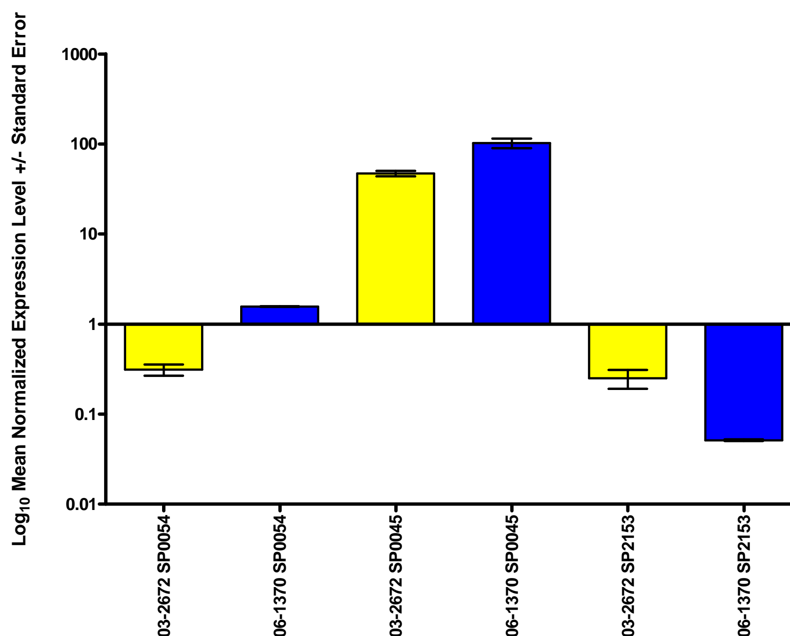


Figure 10-1 Comparison of the Mean Expression Levels for genes SP0054, SP0045 and SP2153 by qRT-PCR.

Mean Expression Levels and standard errors were calculated from 3 experimental replicates for control isolate (03-2672) from serotype 1, ST306 associated bacteraemic pneumonia without parapneumonic complications and test isolate (06-1370) from serotype 1, ST306 associated paediatric empyema.

10.2.5 Discussion Regarding a Genetic Basis for Pneumococcal Parapneumonic Complications

Cases of complicated pneumococcal pneumonia in Scotland result from a mixed population of serotype 1 *S. pneumoniae* (ST227 and ST306) similar to the situation being seen in other European countries and the USA.

Although only two sequence types of *S. pneumoniae* appear to be involved in Scotland, CGH analysis, using microarrays, indicates substantial variation in genes among test isolates of ST227 and ST306 at regions other than the conserved housekeeping genes used in the MLST typing system.

Analysis of bacterial RNA expressed under standardized, predetermined growth conditions shows there to be variation in the regulation of pneumococcal genes which could account for differences in virulence and the clinical presentation and progression of these infections. Using site directed mutagenesis, many of these genes have been shown to code

for virulence factors and have a role in causing pneumonia in a murine model (Hava and Camilli, 2002). However, the growth conditions used in these experiments to grow *S. pneumoniae* for RNA extraction may not accurately mimic those found in the pleural cavity (Byington *et al*., 2002) where other genes may undergo expression and regulation. It is of note that the pleural environment also varies with cases requiring a decortication drainage procedure more likely to have substantially raised lactate dehydrogenase levels (Tan *et al*., 2002, Himelman and Callen, 1986).

However, these experiments do suggest that no single virulence factor will account for the increased virulence in serotype 1 pneumococci causing complicated pneumonia but rather this may be due to a constellation of virulence factors being up and down regulated in the pleural cavity. In keeping with this hypothesis is the fact that penicillin resistance alone does not affect the outcome of paediatric pneumococcal empyema (Paganini *et al*., 2001, Hardie *et al*., 1998). Indeed, subtherapeutic antibiotic levels of susceptible and nonsusceptible antimicrobials in the pleural space (Saglani *et al*., 2005, Giachetto *et al*., 2004) may be contributing to the patterns of expression of virulence genes in empyema formation (Byington *et al*., 2002) (see Chapter 12).

Recently in Taiwan, microarray technology has been used to assess the effect of zinc metalloproteinase B mutations in *S. pneumoniae* isolates causing cases of complicated pneumonia (Hsieh *et al*., 2008). These expression studies were performed with pneumococcal cultures at early log phase (OD600nm of 0.2) with serotype 14 isolates of ST46 and a single locus variant (SLV) of ST328 and the media used for culture was not disclosed. They chose to compare the gene expression of a pleural fluid isolate from a child with pneumococcal empyema (serotype 14, ST46) and a blood culture isolate from a child with bacteraemic necrotizing pneumococcal pneumonia with empyema (serotype 14, ST46) with a “control” blood culture isolate from a child with uncomplicated bacteraemic pneumonia (serotype 14, SLV of ST328). This comparison resulted in the identification of only two genes with a two fold increase in gene expression – SP0664 (zinc metalloprotease) and SP1572 (nonheme iron-containing ferritin).

Neither of these genes appeared in the list of differentially expressed genes identified in this comparison of serotype 1 isolates. This should not be too surprising as this study has used a different serotype, ensured that “control” isolates are matched for sequence type and grown cultures in a defined medium (BHI broth) to mid-logarithmic rather than early-logarithmic phase – all of which will affect which genes are expressed at the point of RNA extraction and which genes are identified during analysis of microarray results. It is not

surprising either that genes identified as associated with virulence may not actually be required in all cases of invasive disease (Blomberg *et al*., 2007).

A significant shortcoming in the use of the microrray to study pneumococcal virulence is its inability to identify genes other than those from the sequenced genomes for TIGR4 and R6. If an increase in virulence in serotype 1 pneumococci has occurred due to the recent acquisition of genes from other bacteria or bacteriophages, microarray analysis will not be able to detect the presence of such genes. By detecting variation (and the absence) of several genes coding for virulence factors in test isolates as illustrated in Appendix 14 and Appendix 15, it may be postulated that these virulence factors do not play a significant role in the development of cases of complicated pneumonia. Included amongst these are genes which code for the first pneumococcal pilus to be described which does not appear to be expressed by any of the serotype 1 isolates analysed on the array. These genes are absent from two serotype 1 ST306 strains associated with parapneumonic complications (06-1370 and 03-3038) which are being sequenced using 454 technology at the Sanger Institute (Nicholas Croucher, Wellcome Trust Sanger Institute, personal communication). However, the genes for a recently described second pneumococcal pilus (Bagnoli *et al*., 2008) identified in the fully sequenced serotype 1, ST227 strain INV104B which are not present in TIGR4 (and not identifiable by microarray DNA CGH) are present in the ST306 isolates 06-1370, 03-3038 and 03-2672 from preliminary 454 sequencing at the Sanger Institute (Nicholas Croucher, Wellcome Trust Sanger Institute, personal communication), consistent with the findings of Bagnoli *et al*, who identified this pilus in serotype 1 ST227 and ST306 isolates. This pilus is involved in adherence to cells of the respiratory tract and so may play a role in the development of bacteraemic pneumonia and parapneumonic complications although this requires further investigation.

Given the concentrations of DNA or RNA (minimum 2µg per 25µl reaction per array) required from test isolates it is unlikely that this microarray analysis could be performed directly on nucleic acid extracted from clinical samples but would require initial culture of the organism or a nucleic acid amplification step which may introduce unacceptable confounding of results. This is unfortunate as the majority of pleural fluid cultures from cases of complicated pneumonia are negative due to prior administration of antibiotics (Byington *et al*., 2002, Eltringham *et al*., 2003). It may be possible to design a suitable microarray to identify isolates likely to result in parapneumonic complications using the genes identified here but using a more sensitive platform such as an oligonucleotide microarray. Such microarrays are already being used to investigate host responses to

respiratory infections and have been designed and used for typing and detecting clinically relevant virulence determinants in *Staphylococcus aureus* isolates (Spence *et al*., 2008).

It is possible using the gene predictor function of Genespring GX 7.3.1 to, *in silico*, create a list of genes implicated in causing parapneumonic complications (Table 10-3). It has been suggested that the presence, by microarray hybridization, of a region SP1050 to SP1053 is associated with highly virulent serotype 1 isolates (Harvey *et al*., 2007). However, our CGH results demonstrate a lack of hybridization for this region and these genes do not feature in the list of those thought to be predictive of parapneumonic complications.

Such complications may resolve without invasive interventions or may require traditional management approaches of thoracocentesis by tube drainage or open surgery (Meyerovitch *et al*., 1985). There is some suggestion that surgery is more likely to be required for serotype 14 pneumococci while tube drainage is more likely to be sufficient alone for serotype 1 (Tan *et al*., 2000). Less invasive strategies are now being utilized (Playfor *et al*., 1997) and microarray analysis of bacterial gene expression could allow more rapid identification of patients at greater risk of empyema or parapneumonic effusion, if these predicted genes are being expressed. Patients infected with isolates expressing these genes may potentially benefit from earlier treatment with intra-pleural administration of fibrinolytic agents (Rosen *et al*., 1993, Playfor *et al*., 1997, Thomson *et al*., 2002, Handman and Reuman, 1993, Sahn, 2007, Kornecki and Sivan, 1997, De Benedictis *et al*., 2000, Kothandapani *et al*., 2006, Campbell, 1995, Hawkins *et al*., 2004, Hamm and Light, 1997, Feola *et al*., 2003) or thoracoscopic decortication (Playfor *et al*., 1997, Sahn, 2007, Kercher *et al*., 2000, Stovroff *et al*., 1995, Gates *et al*., 2004, Wong *et al*., 2005, Hawkins *et al*., 2004, Hamm and Light, 1997, Schultz *et al*., 2004) potentially improving clinical outcome and reducing hospital length of stay (Thomson *et al*., 2002, Hoff *et al*., 1991, Wong *et al*., 2005, Sahn, 2007).

The recent introduction of Prevnar® (Wyeth Pharmaceuticals, USA), into the United Kingdom paediatric vaccination schedule may actually facilitate an even greater increase in incidence of invasive serotype 1 associated disease by providing a niche for its clonal expansion by removing competition from other serotypes. Serotype 1 should be included in future pneumococcal conjugate vaccine formulations (Hanquet *et al*., 2008) such as the 13 valent PCV currently under trial (Kieninger *et al*., 2008).

Orihuela *et al*, demonstrated that the purine synthetic genes SP0045-SP0055 were significantly upregulated in mouse blood (Orihuela *et al*., 2004b). Although bioluminescent imaging was performed on these mice before exanguination, it was not disclosed whether there were parapneumonic complications evident. It is likely that pneumonia was present as the mice had been infected intratracheally. This is important to establish as our data suggests that the upregulation of the purine synthetic genes is not simply a response to their presence in the bloodstream.

Although the majority of isolates which we used for microarray experiments had originated from bloodstream infections, one isolate 03-3038 had been grown from pleural pus. In addition, our analysis considered gene expression of two control isolates grown from the blood cultures of two cases of bacteraemic pneumonia without parapenumonic complications from which we established a baseline gene expression for comparison with isolates from bacteraemic pneumonia cases with parapneumonic complications. This should remove any potential influence of the bloodstream source of isolates on gene expression leaving us to conclude that the genes which we identify as being differentially expressed are due solely to changes which are associated with the presence of parapneumonic complications.

It is also noteworthy that the expression of the purine synthesis gene SP0044 is noted to be temperature dependent, being upregulated above 37°C but downregulated below 37°C (Pandya *et al*., 2005). As patients with empyema often have prolonged, persisting fevers this may be an important survival response by pneumococci to adapt to thermal changes in the host environment.

Alternatively, the change in expression of these purine synthesis genes may not relate directly to the parapneumonic complications but be part of a nonspecific stress response. SP0044 and SP0045 have been found to be downregulated in response to vancomycin stress (Haas *et al*., 2005). It has not been possible to fully review the antibiotic therapy to which these isolates had been exposed in the past although it is known that the patients infected with 03-3038 and 03-2672 did not receive vancomycin. Likewise these isolates have been subcultured and stored frozen in glycerol stocks prior to culture for RNA extraction and so it is unlikely that any preceding antibiotic therapy in the source patient would be influencing these results unless treatment has resulted in a stable mutation.

10.3 Pneumococcal Meningitis

10.3.1 Serotype 3 Association with Meningitis

Serotype 3 is the commonest serotype recovered from cases of meningitis in Poland (Skoczynska and Hryniewicz, 2003) and was recently described as accounting for 18% of adult cases of pneumococcal meningitis in Sweden (Sjostrom *et al*., 2006) although in the 1980s it was the commonest serotype associated with pneumococcal meningitis there (Burman *et al*., 1986). It was also the commonest serotype to be recovered from cerebrospinal fluid in Boston, USA from 1935 to 1972 (Finland and Barnes, 1977a). The proposal by Orihuela *et al* that serotype 3 does not readily enter the bloodstream or CSF because of a large amount of capsule based on a murine model of infection (Orihuela *et al*., 2003) is incompatible with these epidemiological observations in human populations. It is not often that studies have documented which sequence types are present in serotype 3 meningitis associated isolates. However, ST180 serotype 3 is strongly associated with development of meningitis rather than carriage in a Brazilian slum (Reis *et al*., 2008) although the authors do recognise that this association may be more a feature of the serotype 3 capsule than the genotype (Reis *et al*., 2008).

Adjunctive treatments which may disrupt inflammatory cascades as additional treatment strategies (while being treated with appropriate antibiotics) are sought and agents which inhibit CSF cytokines, matrix metalloproteinases and reactive oxygen species show promise (Davis and Greenlee, 2003). Identification of new potential targets for such adjunctive strategies may bring significant breakthroughs.

10.3.2 Serotype 3 ST180 DNA CGH Hybridizations

The results of microarray DNA CGH experiments relating to the isolates 99-4038 and 99-4039 are displayed compared to all other CGH tested ST180 isolates in Figure 6-1 and in Figure 10-2 below.

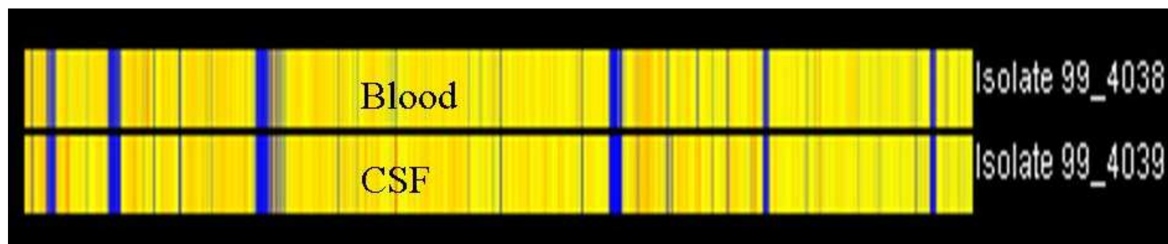


Figure 10-2 CGH comparisons of isolates 99-4038 (cultured from blood) and 99-4039 (cultured from CSF).

Isolates are cultured from specimens taken from the same patient taken on the same day and comparison is generated by Genespring GX 7.3.1. Each individual coloured bar represents a single gene from the TIGR4 or R6 genomes. Shades of yellow and orange indicate competitive hybridization of DNA from both TIGR4 and the test isolates. Blue indicates absence of hybridization for DNA from the test isolate but hybridization by TIGR4 DNA. Red indicates hybridization of test DNA to probes from the R6 genome.

As both isolates were harvested from the same patient on the same day, the discrepant hybridization results between the two isolates for 39 genes was a surprise and a concern. PCR (using TIGR4 as the positive control and PCR grade water as the negative control) was performed for these discrepant genes and a product was obtained for both isolates in each case, indicating that the discrepancies were due to insensitivity of the microarray hybridization rather than differences in the genomes of the isolates. PCR performed at a further 20 genes which did not hybridize for either isolate on the array indicated a product of identical size to that in TIGR4 at all 20 loci which again highlights the lack of sensitivity for DNA hybridizations and limited ability of the microarray to accurately detect true differences in the genetic complement of pneumococcal genomes, as was discussed in Chapter 3.

Gene	Array DNA CGH Result		PCR Result	
	99-4038	99-4039	99-4038	99-4039
SP0068	Blue	Blue	Yellow	Yellow
SP0069	Blue	Blue	Yellow	Yellow
SP0074	Blue	Yellow	Yellow	Yellow
SP0115	Yellow	Blue	Yellow	Yellow
SP0166	Blue	Blue	Yellow	Yellow
SP0168	Blue	Blue	Yellow	Yellow
SP0352	Blue	Blue	Yellow	Yellow
SP0355	Blue	Blue	Yellow	Yellow
SP0467	Blue	Blue	Yellow	Yellow
SP0575	Yellow	Yellow	Yellow	Yellow
SP0697	Yellow	Yellow	Yellow	Yellow
SP1060	Blue	Blue	Yellow	Yellow
SP1064	Yellow	Blue	Yellow	Yellow
SP1132	Blue	Blue	Yellow	Yellow
SP1141	Blue	Blue	Yellow	Yellow
SP1144	Blue	Blue	Yellow	Yellow
SP1318	Blue	Blue	Yellow	Yellow
SP1323	Blue	Blue	Yellow	Yellow
SP1342	Yellow	Yellow	Yellow	Yellow
SP1762	Blue	Blue	Yellow	Yellow
SP1763	Blue	Blue	Yellow	Yellow
SP1765	Blue	Blue	Yellow	Yellow
SP1766	Blue	Blue	Yellow	Yellow
SP1770	Blue	Blue	Yellow	Yellow
SP1771	Blue	Blue	Yellow	Yellow
SP1948	Yellow	Blue	Yellow	Yellow
SP2164	Blue	Blue	Yellow	Yellow

Figure 10-3 Comparison of microarray DNA CGH results for isolates 99-4038 and 99-4039 with PCR results for the same genes using DNA from the same isolates.

Blue indicates an absence of hybridization whereas yellow indicates hybridization in the CGH experiments and the generation of a PCR product of identical size to that in TIGR4 in the PCR experiments.

Although the PCR results highlighted the frequency of false negative hybridizations, there were never any discrepancies in the PCR results for the same gene between the two isolates from blood and CSF. Based on the combination of results from DNA hybridizations and PCR validation of discrepant results, it was concluded that a significant difference in the gene complement of the two isolates in terms of gene deletions was unlikely although the possibility of single nucleotide polymorphisms or insertions of genetic material could not be discounted. To investigate this further, both isolates are undergoing 454 sequencing at the Wellcome Trust Sanger Institute.

10.3.3 Serotype 3 ST180 RNA Expression Experiments

Gene	Normalized Intensity Ratio in Blood	Normalized Intensity Ratio in CSF	Function
Upregulated in CSF			
SP2075	1.1	5.3	ABC transporter, ATP-binding/permease protein
SP2073	1	4.6	ABC transporter, ATP-binding/permease protein
SP0231	0.9	3.8	Adenylate kinase (ATP-AMP transphosphorylase)
SP1869	1.1	3	Iron-compound ABC transporter, permease protein
SP1872	1	2.8	Iron-compound ABC transporter, iron-compound-binding protein
SP0054	0.9	2.8	Phosphoribosylaminoimidazole carboxylase, ATPase subunit
SP1871	1	2.8	Iron-compound ABC transporter, ATP-binding protein
SP2074	1	2.3	Degenerate transposase
SP0507	1	2.2	Type I restriction enzyme EcoKI specificity protein (S protein)
SP1659	1	2.1	Isoleucyl-tRNA synthetase
SP2072	1	2.1	Glutamine amidotransferase, class-I
Downregulated in CSF			
SP2215	1	0.1	30S Ribosomal protein S2
SP2214	1	0.1	Translation elongation factor Ts
SP0496	0.9	0.1	Na/Pi cotransporter II-related protein
SP0488	1	0.1	Conserved hypothetical protein
SP0577	0.9	0.1	PTS system, beta-glucosides-specific IIABC components
SP0578	1.1	0.1	6-phospho-beta-glucosidase
SP1572	1.1	0.1	DNA binding protein starved cells-like peroxide resistance protein
SP0800	1	0.2	Hypothetical protein
SP0517	1	0.2	Class I heat-shock protein (molecular chaperone)
SP0487	1	0.2	Hypothetical protein
SP1626	1.1	0.2	30S Ribosomal protein S15
SP1215	0.9	0.2	Formate-nitrate transporter
SP0489	1	0.2	PAP2 family protein
SP0631	1	0.2	50S Ribosomal protein L1
SP0490	1	0.2	Hypothetical protein
SP1293	1.2	0.2	50S Ribosomal protein L19
SP0630	1.1	0.2	50S Ribosomal protein L11
SP0492	1	0.2	Hypothetical protein
SP0373	1	0.3	Conserved hypothetical protein
SP0649	0.9	0.3	
SP1472	1	0.3	Oxidoreductase, putative
SP1184	0.9	0.3	6-phospho-beta-galactosidase
SP2106	1	0.3	Glycogen phosphorylase family protein
SP0107	1	0.3	LysM domain protein
SP0493	1	0.3	DNA-directed RNA polymerase, delta subunit, putative
SP1027	1	0.3	Conserved hypothetical protein
SP0564	0.9	0.3	Hypothetical protein
SP1185	1	0.3	PTS system, lactose-specific IIBC component
SP1471	1	0.3	Oxidoreductase, putative
SP1739	1	0.4	KH domain protein
SP1449	1	0.4	C3-degrading proteinase
SP1786	1	0.4	Conserved hypothetical protein
SP2226	0.9	0.4	Conserved hypothetical protein
SP2216	1	0.4	General stress protein GSP-781

SP2058	1	0.4	Queuine tRNA-ribosyltransferase
SP0058	1	0.4	Transcriptional regulator, GntR family
SP2063	1.1	0.4	Conserved hypothetical protein
SP0457	1	0.4	Bacitracin resistance protein
SP1674	1	0.4	Phosphosugar-binding transcriptional regulator, putative
SP1975	1	0.4	SpoIIIJ family protein
SP1470	0.9	0.5	Thiamine biosynthesis protein ApbE, putative
SP0375	1	0.5	6-phosphogluconate dehydrogenase, decarboxylating
SP0376	1	0.5	DNA-binding response regulator
SP1111	1	0.5	Conserved hypothetical protein
SP0247	1	0.5	Transcriptional regulator
SP2057	1	0.5	Hypothetical protein
SP2150	1	0.5	Ornithine transcarbamoylase

Table 10-4 Comparison of Normalized Expression Ratios of genes differentially expressed when expression levels in CSF are compared to blood.

Results are for isolates of serotype 3, ST180 cultured from the same patient on the same day using a one tailed ANOVA with significance set at $P<0.05$ and the Benjamini and Hochberg correction used for multiple testing. The expression levels in blood are normalized against themselves to generate a baseline for comparison with expression levels in CSF. Genes highlighted in red have been associated with virulence in a serotype 4 mouse pneumococcal pneumonia model (Hava and Camilli, 2002).

Gene	**Fold Change in CSF (Orihuela *et al*., 2004b)**	**Fold Change in CSF (This study)**
SP0517	-3.1	-5.0
SP1975	-2.2	-2.5
SP1572	-2.4	-11.0
SP2058	-3.1	-2.5
SP0488	-2.9	-10.0
SP0800	+2.1	-5.0

Table 10-5 Comparison of the fold change differences in expression for genes identified as being downregulated in CSF in this study with results from similar work by Orihuela *et al* (Orihuela *et al*., 2004b).

Gene	Fold Change in Blood (Orihuela *et al.*, 2004b)	Expression in CSF compared to blood (This study)
SP0457	+2.5	Downregulated
SP0054	+6.9	Upregulated
SP0231	-3.8	Upregulated
SP1572	-8.4	Downregulated
SP0630	-4.7	Downregulated
SP0631	-5.0	Downregulated
SP1626	-3.5	Downregulated
SP2073	+2.2	Upregulated
SP0107	+2.7	Downregulated
SP0488	-4.7	Downregulated
SP0800	+4.5	Downregulated
SP2063	+4.9	Downregulated

Table 10-6 Comparison of results for genes identified as significantly up or down regulated in CSF with their fold change levels as detected in mouse blood by Orihuela *et al* (Orihuela *et al.*, 2004b).

+ indicates upregulation while – indicates downregulation.

10.3.4 Discussion regarding Serotype 3 Associated Pneumococcal Meningitis

10.3.4.1 Phenotypic variants within Serotype 3

In the serotype 3 pneumococcus, it has been elucidated that mutations in the genome can result in phenotypic variants generated in response to a change in the host environment (particularly within a biofilm environment). These are likely to have an adaptational advantage for survival in a more hostile environment. Until recently, only one such mechanism had been described – phase variation. But now non phase variable colony variants which are generated in response to an environmental change have also been described (Allegrucci and Sauer, 2008, Allegrucci and Sauer, 2007).

10.3.4.1.1 Phase Variants

The phenomenon of phase variation is described in Chapter 1. The transparent phenotype has more cell wall phosphorylcholine, less capsular polysaccharide and has different surface proteins compared to the opaque phenotype (Ring *et al.*, 1998, Weiser and Kapoor, 1999). Although they did not assess any serotype 3 isolates, Ring *et al* found that the transparent phenotype resulted in increased invasion of pneumococci across the blood brain barrier by as much as 6 fold compared to opaque phenotype (Ring *et al.*, 1998) whereas opaque variants survive better than transparent in the bloodstream and are integral

to invasive infection (Overweg *et al*., 2000b). Higher rates of transformation occur in transparent variants compared to opaque so it seems that less capsule (transparent phenotype) enhances transformation (Weiser and Kapoor, 1999). It has been proposed that the change in cell surface components which occurs in phase variation (Overweg *et al*., 2000b) is important for pathogenicity (Bruckner *et al*., 2004). The reversible nature of the capsular mutations which account for these phase variants in serotype 3 is considered in Chapter 6.

10.3.4.1.2 Non revertible variants

In addition to phase variation, serotype 3 pneumococci also generate mucoid and non-mucoid variants by a different mechanism entirely (a large deletion in the *cps3DSU* capsule operon) (Allegrucci and Sauer, 2008). This occurs at high frequency and unlike phase variants, these are not revertible (Allegrucci and Sauer, 2008). It has also been demonstrated that hydrogen peroxide plays a role in their emergence (Allegrucci and Sauer, 2008). Hydrogen peroxide is also known to be mutagenic in several other pneumococcal serotypes, triggering frameshifts resulting from the reversible gain and loss of single bases, deletions resulting from recombination and substitutions of guanine residues – the net result of which can change the pneumococcal phenotype as has been shown by spontaneous mutation to optochin resistance (Pericone *et al*., 2002, Battig and Muhlemann, 2007) or rifampicin resistance (Battig and Muhlemann, 2007).

These mutants which are unable to revert to wild type serotype 3 capsule expression may be at a disadvantage in blood stream infections but it is plausible, although unproven, that they have enhanced ability to penetrate CSF by crossing the blood brain barrier or have a survival advantage in the CSF. It is established that the production of hydrogen peroxide by pneumococci is implicated in the pathogenesis of pneumococcal meningitis where it contributes to neuronal apoptosis (Braun *et al*., 2002), causes ciliary stasis of ependymal cells (Hirst *et al*., 2000), vasodilation and subsequent increased cranial pressure (Hoffman *et al*., 2007). If hydrogen peroxide is a stimulus to mutation and the generation of non-revertible phenotypic variants it may, in the same microenvironment create an altered phenotype able to cross the blood brain barrier by a paracellular route (Koedel *et al*., 2002) precisely in a region whose integrity has been compromised by simultaneous hydrogen peroxide induced vasodilation. This would challenge current beliefs that pneumococci cross the blood brain barrier by a transcellular route (Koedel *et al*., 2002). Intriguingly, a gene which codes for a protein thought to be a peroxide resistance protein or non-heme iron-containing ferritin (SP1572) is downregulated in CSF compared to blood suggesting a need for peroxide resistance in blood but not once past the blood brain barrier – a finding

also documented by Orihuela *et al* (Orihuela *et al*., 2004b). The ability to generate non-revertible phenotypic variants is not unique to serotype 3 but Allegrucci *et al* also have observed this in serogroup 19 pneumococci (Allegrucci and Sauer, 2008) which could explain why some serotypes are more prone to cause meningitis and is entirely compatible with epidemiological observations that serotype 19F is associated with causing meningitis with a high associated mortality (Skoczynska and Hryniewicz, 2003, Burman *et al*., 1986, Reis *et al*., 2002, Urwin *et al*., 1996).

10.3.4.2 Choline binding protein A

It has been shown that choline binding protein A (CbpA or SP2190) is necessary for pneumococci to cross the blood brain barrier (Ring *et al*., 1998, Orihuela *et al*., 2004a, Orihuela *et al*., 2003) by binding to the human polymeric immunoglobulin receptor (pIgR) and crossing the capillary endothelium by transcytosis (Koedel *et al*., 2002). This is related to a transparent phase variant phenotype as the transparent phase variant expresses greater amounts of CbpA (Ring *et al*., 1998). When grown in BHI broth as planktonic cultures, CbpA is expressed in both isolates from blood and CSF at low levels but the fact that it is found to be expressed is compatible with the above hypothesis. Interestingly, Orihuela *et al* did not identify a difference in expression of SP2190 in blood or CSF using serotype 2 (D39 Xen 7) and serotype 4 (TIGR4) pneumococci using microarray experiments either even when bacteria were harvested straight from rabbit CSF (Orihuela *et al*., 2004b).

10.3.4.3 The role of iron

At the start of this chapter it was noted that transcriptional regulators often play a role in virulence. SP0376 is such a gene. Its homolog in R6 nomenclature is spr0336 but it is also known as *RitR* (Repressor of Iron Transport) or the orphan response regulator and is one of the pneumococcal two component systems which can sense changes in the environment and elicit a transcriptional response (Paterson *et al*., 2006).

RitR has been identified as having a role in iron transport in the pneumococcus (Ulijasz *et al*., 2004a). Expression of *RitR* represses iron uptake *in vitro* by downregulating an iron carrier protein *piuA* (Ulijasz *et al*., 2004a). Intriguingly, there is a relationship between iron and hydrogen peroxide as iron catalyses the production of deleterious oxygen free radicals from hydrogen peroxide by means of the Fenton reaction (Ulijasz *et al*., 2004a, Pericone *et al*., 2003). Consequently, the pneumococcus will decrease iron concentrations in the presence of hydrogen peroxide. The results above suggest that SP0376 (*RitR*) is downregulated in the CSF while two of the genes which it regulates SP1869 (Iron-

compound ABC transporter, permease protein) and SP1872 (Iron-compound ABC transporter, iron-compound-binding protein) are upregulated in the CSF as shown in Figure 10-4.

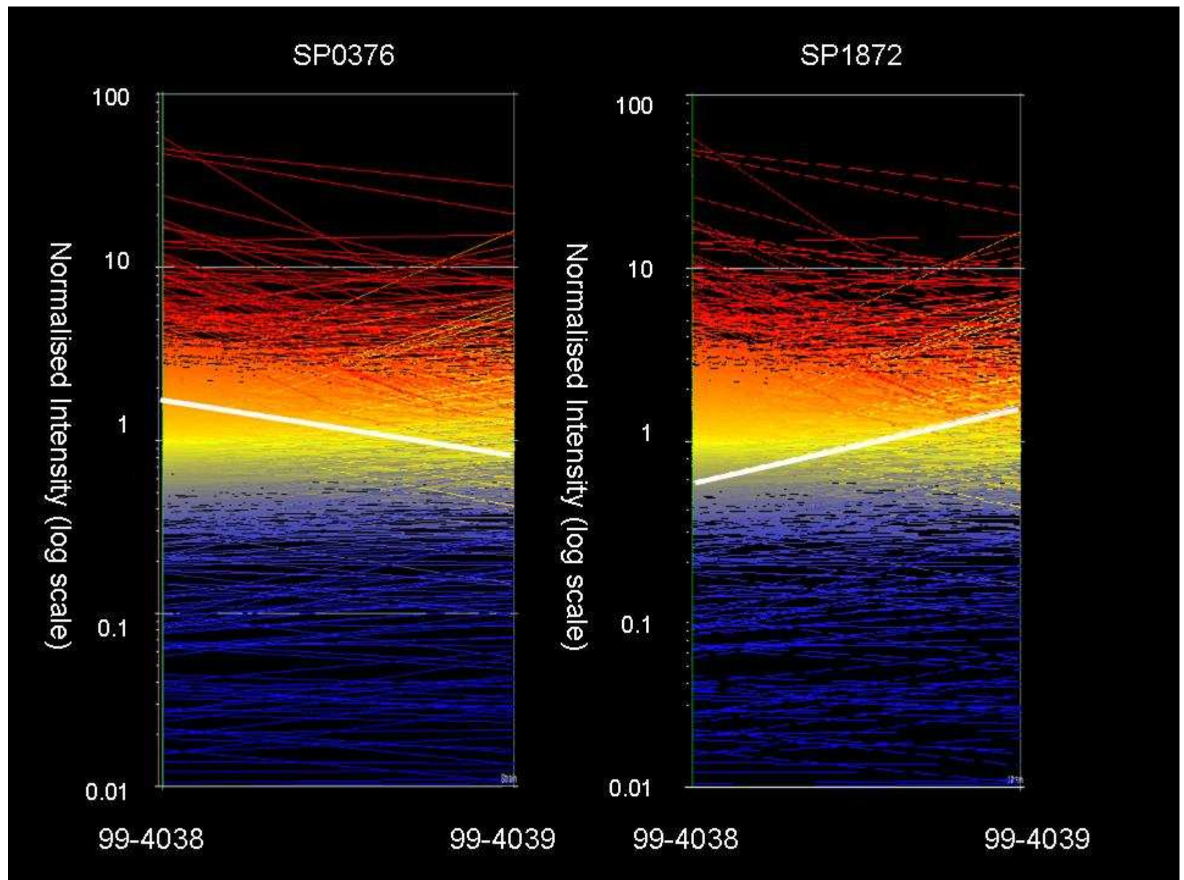


Figure 10-4 Comparison of the significant differences in expression of SP0376 and SP1872 in isolates 99-4038 (blood origin) and 99-4039 (CSF origin)

This figure is generated by Genespring GX 7.3.1. The white lines indicate the downregulation of SP0376 in the CSF isolate and upregulation of SP1872 in the CSF isolate.

These differences in gene expression were also investigated by RT-PCR as demonstrated below in Figure 10-5.

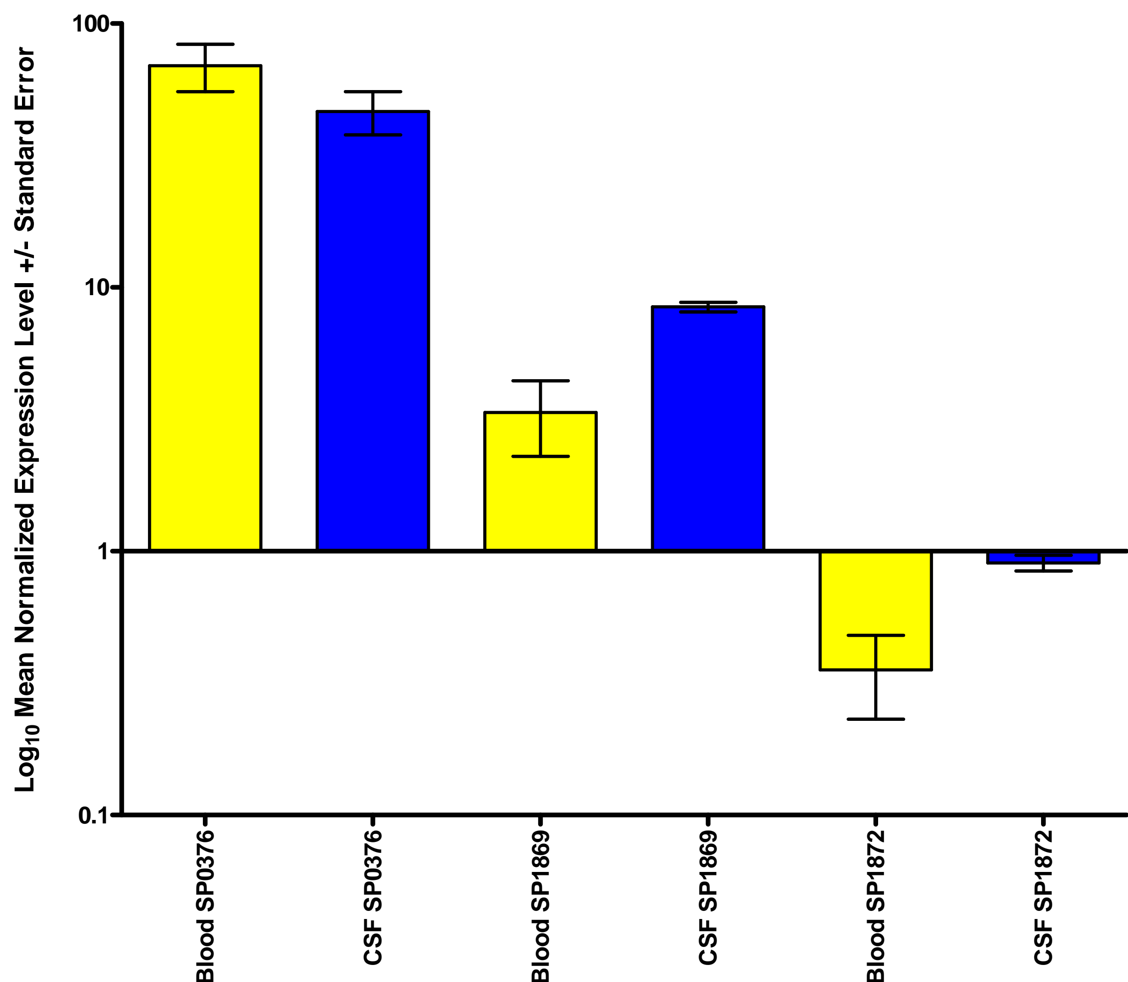


Figure 10-5 qRT-PCR results relating to gene expression of SP0376 (*RitR*) in blood and CSF isolates and the upregulation of SP1869 and SP1872 in a CSF isolate compared to a blood isolate.

P values for the difference in mean normalized expression between blood and CSF isolates were calculated using an unpaired t-test calculated by Q-gene software (Muller *et al.*, 2002) and were for SP0376 (p = 0.121), SP1869 (p = 0.005) and SP1872 (p = 0.014). Error bars indicate standard errors of the mean normalized expression calculated by Q-gene.

It was unexpected that the difference in mean normalized expression for SP0376 (*RitR*) when tested by qRT-PCR did not reach statistical significance whereas the differences in expression of the two genes which SP0376 regulates (SP1869 and SP1872) were significantly different by both microarray and qRT-PCR approaches. This still suggests that in this case of human meningitis, a non revertible mutation or an environmental factor may have affected the expression of *RitR* so that expression levels are higher in the blood than the CSF with a consequent increase in iron uptake in CSF compared to blood. This may be a response to the bloodstream environment being rich in bound iron but the potential interplay with hydrogen peroxide cannot be ignored. If, as proposed above, the pneumococcus generates hydrogen peroxide locally at the blood/endothelial surface after binding via choline binding protein A, the effects of this may be the generation of stable

non revertible mutations and a compromised blood brain barrier resulting from local vasodilation and/or neuronal apotosis at the same site. It is entirely plausible that, within the blood, at this region of local hydrogen peroxide release, iron concentrations are being regulated by *RitR* to remain low locally to aid survival of the pneumococcus by preventing the Fenton reaction. Such a critical role may also explain why neither *RitR* nor the iron compound ABC transporters SP1869-SP1872 are located in a known region of diversity in the pneumococcal genome (Silva *et al*., 2006).

SP1871 (another iron compound ABC transporter) also appeared to be upregulated in CSF. Pandya *et al* found that SP1871 is significantly affected by temperature changes being upregulated at 29°C, 33°C and 40°C when compared to 37°C. It was one of only 8 genes upregulated at 40°C (Pandya *et al*., 2005). Both hypothermia or hyperthermia and body temperature dysregulation may be a host response to intracranial pathology including infection and so the temperature related upregulation of this gene may be important for the survival of the pneumococcus within the CSF once it has passed the blood brain barrier, established infection and is responding to further changes in the host environment.

10.4 Pneumococcal Cerebral Abscess

10.4.1 Background

Pneumococcal cerebral abscesses are uncommon but are associated with significant morbidity (40% of survivors are left with neurological deficits) and high mortality (case fatality rate of 35% in one series) despite appropriate antibiotic therapy and surgery (Grigoriadis and Gold, 1997). Serotype 3 pneumococci have historically been associated with cerebral abscesses (Fincher, 1946, Anonymous, 1970, Colman and Hallas, 1983) and in a rat experimental meningitis model, serotype 3 pneumococcal infections have a preponderance to cause cortical necrosis and abscess formation whereas serotype 1 infections are more associated with cortical haemorrhage and not abscess formation (Ostergaard *et al*., 2004).

The MLST records of pneumococcal isolates grown from cerebral abscess pus in Scotland were identified and serotype 3 isolates chosen in order to perform microarray DNA CGH and RNA expression experiments to identify virulence associated genes in order to gain further understanding of the pathogenesis of this condition.

10.4.2 Serotype and MLST distribution of Cerebral Abscess Associated Pneumococci in Scotland

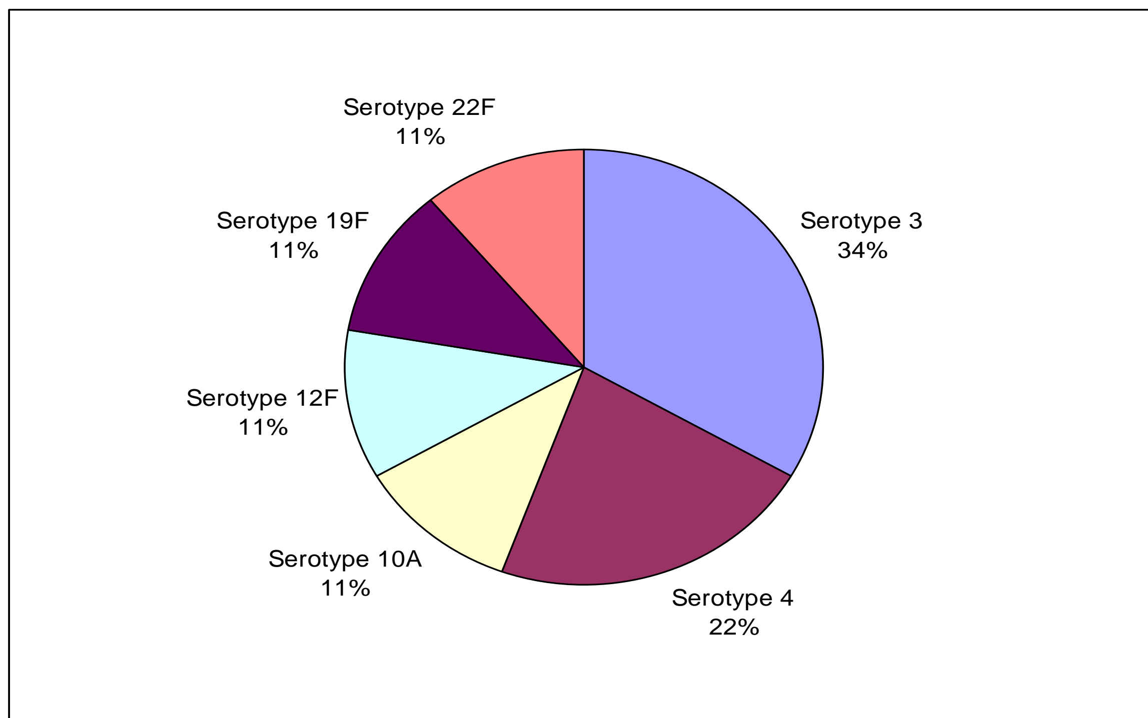


Figure 10-6 Distribution of pneumococcal serotypes associated with cerebral abscesses in Scotland 1993-2007 (n=9).

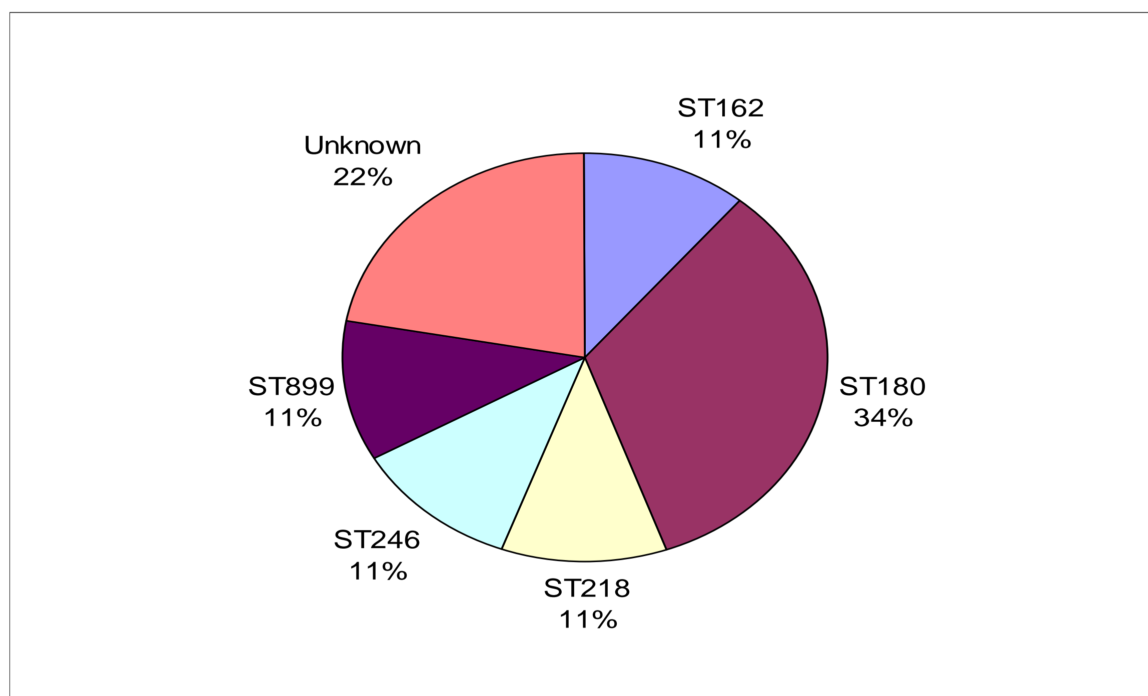


Figure 10-7 Distribution of pneumococcal MLSTs associated with cerebral abscesses in Scotland 1993-2007 (n=9).

10.4.3 Serotype 3 ST180 DNA CGH Hybridizations

Microarray CGH was performed on two cerebral abscess associated isolates (00-3946 and 06-1705). These results feature in Figure 6-1 in Chapter 6.

10.4.4 RNA Expression of Serotype 3 ST180 Cerebral Abscess Associated Isolates

RNA was extracted and utilised in microarray expression experiments according to the procedures outlined in Chapter 2. Data analysis in Genespring GX 7.3.1, involved normalizing the invasive serotype 3, ST180 related data from Figure 6-1 into two categories of cerebral abscess or non-cerebral abscess related isolates and the gene expression levels in these two merged categories were compared (Table 10-7). The gene prediction function of Genespring GX 7.3.1 was used to generate a list of 30 genes whose expression was associated with the presence of a cerebral abscess.

Gene	Normalized Intensity Ratio in non cerebral abscess ST180s	Normalized Intensity Ratio in Cerebral abscess ST180s	Gene Function
Upregulated in cerebral abscess			
spr0957	1.3	102.3	SpR6: Tn5252, relaxase, truncation
SP0110	0.9	6.0	ABC transporter membrane-spanning permease - amino acid transport
SP0111	1.1	3.1	ABC transporter ATP-binding protein - amino acid transport
Downregulated in cerebral abscess			
SP1185	0.9	0.3	PTS system, lactose-specific IIBC component
SP0577	1.2	0.3	PTS system, beta-glucosides-specific IIABC components
SP1686	1.1	0.5	Oxidoreductase, Gfo/Idh/MocA family
SP1688	1.1	0.5	ABC transporter, permease protein

Table 10-7 Genes identified by Genespring GX 7.3.1. as upregulated or downregulated in cerebral abscess associated serotype 3 ST180 isolates compared to non cerebral abscess related serotype 3 ST180 isolates.

Gene	Predictive Strength for Cerebral Abscess	Function
SP0877	27.0	Fructose specific-phosphotransferase system IIBC component
SP1474	20.9	Glycyl-tRNA synthetase beta chain
SP1592	16.6	Conserved domain protein
SP0830	16.6	Hypothetical protein
SP0066	16.1	Aldose-1-epimerase (mutarotase)
SP1097	14.9	Conserved hypothetical protein
SP0831	14.7	Purine nucleoside phosphorylase (inosine phosphorylase)
SP0876	14.7	Fructose-1-phosphate kinase
SP0833	14.7	Hypothetical protein
SP1100	14.6	Phosphate acetyltransferase
SP0834	14.4	Hemolysin-related protein
SP0629	14.1	Conserved hypothetical protein
SP0829	12.8	Phosphopentomutase
SP0828	12.8	Ribose 5-phosphate isomerase
SP1563	12.7	Pyridine nucleotide-disulphide oxidoreductase family protein
SP1701	12.7	Phospho-2-dehydro-3-deoxyheptonate aldolase
SP2096	12.7	Peptidase, M20/M25/M40 family
SP1192	11.4	Galactose-6-phosphate isomerase, LacB subunit
SP1193	11.4	Galactose-6-phosphate isomerase LacA subunit
SP0110	10.7	ABC transporter membrane-spanning permease - amino acid transport
SP0782	9.6	Conserved hypothetical protein
SP0605	5.9	Fructose-bisphosphate aldolase
SP2192	5.9	Sensor histidine kinase
SP0445	5.5	Acetolactate synthase, large subunit, biosynthetic type
SP2002	5.4	Conserved hypothetical protein
SP0015	5.1	IS630-Spn1, transposase Orf1
SP1591	4.7	Proline dipeptidase
SP1988	4.6	Conserved hypothetical protein
SP1417	4.3	Choline binding protein
SP0438	4.3	Glutamyl tRNA-Gln amidotransferase, subunit C

Table 10-8 Genes identified by Genespring GX 7.3.1. as predictive of a cerebral abscess associated phenotype when compared to non cerebral abscess related serotype 3 ST180 isolates.

10.4.5 *Discussion regarding pneumococcal cerebral abscesses*

This is the first case series of pneumococcal cerebral abscesses to be described from Scotland and the first to investigate which sequence types are implicated. These results are consistent with historical case reports and animal model experiments which demonstrate serotype 3 pneumococci as being associated with cerebral abscess formation. Only the cases caused by serotype 4 and 19F were preventable using the PCV-7 conjugate vaccine

formulation with 78% of cases not preventable, highlighting a need for alternative prevention strategies or conjugate vaccination with higher valency vaccines.

Such alternative strategies would benefit greatly from further understanding of the genetic basis of the pathogenesis of this condition. Microarray CGH experiments demonstrate that the cerebral abscess associated serotype 3 isolates show much genomic diversity at 14 known regions of diversity when compared to the fully sequenced serotype 4 pneumococcal isolate, TIGR4 (shown in Appendix 12). When compared to each other, there is much less genetic diversity between the serotype 3 cerebral abscess isolates despite their origins from different patients, in different parts of Scotland and occurring six years apart. Consequently, development of a preventive strategy which focused on utilising these conserved genes or their transcripts could be successful in preventing up to a third of cases.

However, caution is required in interpreting these results as the RT-PCR experiment described in Chapter 6 comparing the gene expression of SP0110 in the cerebral abscess associated isolate (00-3946) and non cerebral abscess related isolate (OXC141) demonstrated that there was no significant difference in their gene expression by that method and that the results generated by Genespring GX 7.3.1 were likely spurious resulting from the normalization procedure used to merge the expression data into two categories. This highlights the need for independent confirmation of data generated from microarray gene expression experiments and the dangers of indiscriminately applying *in silico* results to clinical cases (Chuaqui *et al*., 2002).

10.5 Overall Discussion Regarding Gene Associations with Invasive Pneumococcal Disease Manifestations

This series of investigations using microarray technology to investigate the potential influence that genomic diversity has on determining disease manifestations of IPD demonstrated several factors which require consideration when interpreting the results.

One basic assumption from this approach has been that the disease manifestation relates solely to the genotype and consequent phenotype of the pneumococcus and it does not consider host factors to be contributing. This clearly is not the case as pathology is the result of an interplay between host and pathogen and the 50 to 100 fold higher incidence of IPD in immunocompromised human hosts with HIV (Flannery *et al*., 2006) compared to

the non-HIV infected population or the effect of patient age on the distribution of serotypes causing IPD (Inostroza *et al*., 2001), demonstrates that host factors do contribute to the incidence and manifestations of IPD.

A further assumption has been that microarray technology is a suitable platform by which to accurately compare pneumococcal genomes. The results of DNA based CGH in this chapter are similar to those in previous chapters in that they indicate that lack of hybridization of a gene cannot be assumed to mean that the gene is absent from the genome as lack of hybridization is often due to the gene being present (as demonstrated by PCR) but in a form which is not complementary to the TIGR4 sequence based probe on the microarray. This level of insensitivity makes interpretation of the significance of lack of hybridization impossible without further validation by PCR and/or genome sequencing and so the microarray CGH approach using this "spotted" PCR product microarray alone can only at best be regarded as a screening tool to allow a focus on particular regions where genes may show sequence diversity when compared to the TIGR4 genome.

This is not to say that the microarray does not have a role in the investigation of pathogenesis as the above studies using the microarray to analyse RNA expression have shown. By this approach, significant insights into potential mechanisms of disease for pneumococcal meningitis and empyema have been possible which can now be utilised to develop new diagnostic tests, develop possible pharmacological therapies or target therapeutic interventions. Nevertheless, significant caution is required when extrapolating *in silico* results to a wider population as has been demonstrated with our investigations into cerebral abscess pathogenesis where genes could be identified as significant which clearly are not when investigated by another method (qRT-PCR). This is likely to be the end result of a series of factors such as poorly chosen "control" isolates, study of too small a number of isolates and over-dependance on statistical procedures to merge a diverse bacterial population into two categorical variables of cerebral abscesses being present or absent. This has introduced confounding such as that produced by high expression of a gene (e.g. SP0110) in one isolate when the gene is not actually present in all isolates of that category which may suggest overall significance but biologically cannot be significant if it is not actually present in the genome of some cases.

These results do highlight other issues relating more generally to experimental design when investigating pneumococcal pathogenesis as they add to a growing body of evidence that suggests that with clinical isolates, source tissue or body fluid may influence gene expression and the overall results of an investigation (Ogunniyi *et al*., 2007, Orihuela *et*

al., 2004a) and so documentation of the details of the source isolate should be regarded as essential. There are also implications regarding choice of media for *in vitro* cultures as arbitrarily chosen *in vitro* conditions may not be representative of *in vivo* conditions and alterations in culture conditions do affect virulence gene expression in many bacterial species (Pandya *et al*., 2005). Pneumococci, even when in pure culture grown from a single colony, may not be homogeneous populations but, as stated by Weiser *et al*,

> "an isolate should be considered a mixed population of phenotypes which differ in amounts of capsular polysaccharide, teichoic acid and choline binding proteins (Weiser and Kapoor, 1999)."

As early as 1981, Andersson *et al* saw that pneumococcal isolates cultured from different body sites had different abilities to colonise the nasopharynx and cause invasive disease in a mouse model but were unable at that time to determine a genetic basis for such different behaviour (Andersson *et al*., 1981).

Ideally, investigations of pneumococcal gene expression would be most representative if performed from specimens taken directly from a human patient. Apart from ethical considerations such as informed consent, which would need to be broached in advance of an individual falling ill from an incapacitating severe manifestation of IPD such as meningitis and the confounding effects of antimicrobial therapy which would be unethical to withhold, there are technical difficulties in obtaining sufficient quantities of bacterial RNA directly from blood separated from host human RNA (Shaw and Morrow, 2003) although this has been possible on occasion in mice (Orihuela *et al*., 2004b).

The recent description of non revertible colony variants resulting from stable point mutations triggered by local hydrogen peroxide (Allegrucci and Sauer, 2008) together with existing understanding of the reversible serotype 3 capsular mutations found in opaque and transparent phase variants (Waite *et al*., 2001, McEllistrem *et al*., 2007) is of significance if these results are due to biofilm-derived, non revertible colony variants from different body sites whose particular phenotypes have been selected for adaptation to different host environments. Determining this is likely to be beyond the resolving capacity of a "spotted" microarray approach though, particularly when the genes which are most likely to be involved (the serotype 3 capsular genes) are not represented by probes on this microarray. King *et al*, compared gene expression of opaque and transparent phenotypes of serotype 6A and 6B strains on a "spotted" microarray though and identified 24 genes which were significantly differentially expressed between the two phenotypes (King *et al*., 2004). Of these 24 genes, only SP2150 (ornithine carbomyltransferase) appears in our comparison of

serotype 3 isolates 99-4038 and 99-4039 suggesting that the differences we see in expression of blood and CSF isolates are not due to phase variation. Although the genes involved in phase variation may differ between serotypes and isolates (King *et al*., 2004), a non revertible mutation triggered by hydrogen peroxide rather than phase variation fits more with the current body of evidence implicating hydrogen peroxide in the pathogenesis of cerebral damage and the control of iron metabolism in limiting the toxicity of hydrogen peroxide to the pneumococcus. Given the substantial number of permutations of minor point mutations now documented in the house-keeping genes (chosen as they were not thought to vary substantially) employed in the MLST scheme, it seems rational that point mutations in other essential genes will be occurring with the potential to alter virulence. The observation of this occurrence affecting pneumolysin (Jefferies *et al*., 2007, Kirkham *et al*., 2006), is further supporting evidence to corroborate this hypothesis. Significantly, preliminary data from 454 sequencing of these two strains identifies 31 SNPs which are present in 99-4039 compared to 99-4038 (Nicholas Croucher, Wellcome Trust Sanger Institute). Although none of these occur in the serotype 3 capsular genes or *ritR*, three of these are of substantial interest as they occur within the open reading frames of transcriptional regulators. CcpA and LysR (SP0927) are affected by transpositions and interestingly a deletion is present in the CSF associated ArgR gene. Although no significant differences in expression for these genes were noted in our analysis using the current parameters, there may still be an undetected, biologically significant, differential expression of these mutant genes in blood and CSF sufficient to cause downstream effects which may alter the expression of *ritR* or genes coding for iron binding and this requires further investigation.

Finally, it has been proposed that for any particular pneumococcal isolate it will,

> "probably carry only the particular subsets of genes that permit them to cause the distinct forms of disease associated with each strain (Orihuela *et al*., 2003)."

If that is so, then it should be possible to identify core sets of genes associated with different disease manifestations such as we have attempted to generate using the Genespring GX 7.3.1 gene predictor tool. By comparing such lists, when generated from a large enough sample of relevant isolates, it may be possible to infer whether a particular isolate has the potential for more than one disease manifestation and identify core subsets of genes which could be targeted for therapeutic interventions.

11 Diversity of Pneumococcal Gene Expression in Response to an Antibiotic

11.1 Physicochemical Properties of Clarithromycin

Clarithromycin is a 14-membered macrolide antibiotic which is structurally related to erythromycin through substitution at the 6th carbon atom of the erythronolide ring with a methoxy group (Hardy *et al.*, 1992, Hardy *et al.*, 1988) as shown in Figure 11-1. Unlike erythromycin it is acid stable and undergoes hepatic metabolism to a compound with antimicrobial activity – 14hydroxyclarithromycin (Hardy *et al.*, 1992). It has 55% bioavailability in its oral form (Jain and Danziger, 2004) and reaches peak tissue concentrations 4 hours after oral administration (Hardy *et al.*, 1992) being excreted via urine and bile (Jain and Danziger, 2004). The serum half life is approximately 5 hours (Jain and Danziger, 2004). Tissue (particularly lung and nasal tissue) concentrations are higher than those in blood which accounts for its effectiveness against a range of respiratory bacterial pathogens (Hardy *et al.*, 1992, Jain and Danziger, 2004). It also penetrates well intracellularly into alveolar macrophages (Jain and Danziger, 2004).

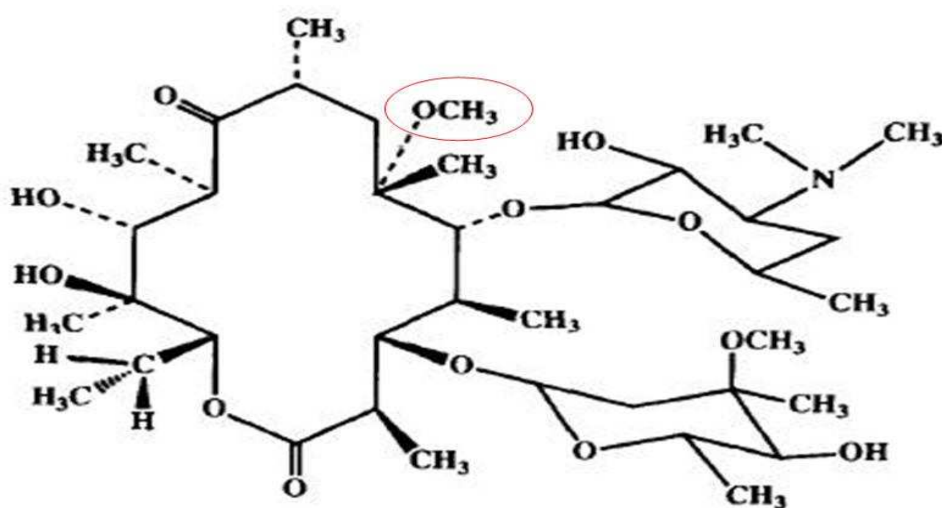


Figure 11-1 Structure of clarithromycin.

Figure is adapted from (Hardy *et al.*, 1992)) with the methoxy group which differentiates clarithromycin from erythromycin circled in red.

11.2 Mechanism of Action of Clarithromycin

Like all macrolides, clarithromycin exerts its antibacterial action by preventing protein synthesis. It does this by binding reversibly to 23S ribosomal RNA in the 50S subunit of bacterial ribosomes (Jain and Danziger, 2004). Two theories exist as to how protein synthesis is disrupted. The first is that peptidyl transfer reactions are inhibited resulting in incomplete peptide chains detaching from the ribosome. The second is that dissociation of peptidyl-tRNA from the ribosome is stimulated, resulting in immature proteins (Jain and Danziger, 2004). Clarithromycin is classed as a bacteriostatic antibiotic but it has been shown to be bactericidal against the pneumococcus (Jain and Danziger, 2004).

11.3 Resistance Mechanisms associated with Clarithromycin

The diverse mechanisms of macrolide resistance are outlined in Chapter 1.

11.4 Putative Effects of Clarithromycin at Subtherapeutic Concentrations

11.4.1 Anti-inflammatory and Immunomodulatory Effects

Shinkai *et al* define immunomodulation as,

> "the activity of suppressing the prolonged activation of the inflammatory and immune system that can lead to adverse effects on the host; but without globally affecting the innate immune responses as would be seen with an immunosuppressive agent (Shinkai *et al*., 2005)."

For instance, in a murine model of pneumococcal pneumonia, a macrolide (HMR 3004) has been shown to inhibit interleukin 6 and nitric oxide synthesis (Chu, 1999).

In their review Shinkai *et al* also note that,

> "macrolides have been reported to regulate prolonged or hyperinflammation by effects on cellular immunity, suppressing the production of pro-inflammatory cytokines and reactive oxygen species, blocking the activation of nuclear transcription factors, inhibiting neutrophil activation and mobilization, accelerating neutrophil apoptosis and improving mucus clearance (Shinkai *et al*., 2005).

Tsai and Standiford review the mechanisms of action of macrolides on pleural inflammation, respiratory tract infection, airways inflammation, pulmonary fibrosis and immune mediated lung diseases (Tsai and Standiford, 2004). Because of these effects, clarithromycin has an established role in the management of diffuse panbronchiolitis and cystic fibrosis because of its immunomodulatory effects rather than antibacterial effects (Amsden, 2005, Shinkai *et al*., 2005).

11.4.2 Anti-neoplastic Effects

Clarithromycin increases the production of interferon-γ and interleukin 4 by T-cells resulting in a decrease in tumour growth in a murine model of primary lung cancer. It also enhances $CD8^+$ T-cell cytotoxicity and Natural Killer cell activity (Tsai and Standiford, 2004, Hamada *et al*., 2000). The net result of the actions has been demonstrated as reduction in primary tumour size and reduction in lung metastases and increased tumour apotosis (Tsai and Standiford, 2004). A further effect of clarithromycin on tumours is to affect the formation of blood vessels (Yatsunami *et al*., 1999). The clinical application of this was to demonstrate a survival benefit in patients with advanced non-small cell lung cancer in patients receiving clarithromycin (Mikasa *et al*., 1997).

11.4.3 Transcriptional modulation

At concentrations below its Minimum Inhibitory Concentrations (MIC), the macrolide antibiotic erythromycin has been shown to influence the transcription of 5-10% of bacterial genes in *Salmonella typhimurium* which, despite altering gene expression in some cases by 10-100 fold, had little effect on growth (Goh *et al*., 2002). Interestingly at concentrations above the MIC, few transcriptional changes occurred. Clarithromycin also alters gene transcription at sub therapeutic concentrations in *S. typhimurium* (Tsui *et al*., 2004). By demonstrating this phenomenon of hormesis (a different response to a molecule at low concentration (transcriptional regulation) compared to higher concentration (growth inhibition), Tsui *et al* postulate that,

> “subinhibitory concentrations identify responses that more accurately reflect antibiotic mode of action and … these effects might represent the “natural” role of antibiotics, since in the environment the concentrations of these molecules rarely attain inhibitory levels (Tsui *et al*., 2004).”

11.4.4 Disruption of Quorum Sensing

Although not demonstrated in the pneumococcus, macrolide antibiotics have the ability to alter bacterial cell to cell signalling (quorum sensing) within colonies of *Pseudomonas aeruginosa* resulting in the regulation of virulence factors which may account for improved clinical outcomes when macrolides are administered to patients with cystic fibrosis and bronchiectasis colonised with biofilms of *P. aeruginosa* against which macrolides have no conventional anti-bacterial activity (Tateda *et al*., 2004). As pneumococci may also form biofilms, it is reasonable to postulate that macrolides may have a role in affecting the quorum sensing between pneumococci.

Tsui *et al* after demonstrating effects of subtherapeutic macrolide antibiotics on *S. typhimurium* suggest (Tsui *et al*., 2004):

> "The transmission of signals from ribosome to RNA polymerase due to subinhibitory macrolides could involve the release of small amounts of incomplete polypeptides, interference with ribosome assembly, induction of translation errors, or possibly interactions of small molecules with RNA. The sequelae of all these events may be low-level stress responses that act through one of the many bacterial sigma factors to activate or repress specific sets of transcripts. These changes might also result in compensating effects on the transcription of nodes of linked metabolic networks."

11.4.5 Effects on Virulence Factors

It has been established that macrolide antibiotics reduce the production of the pneumococcal virulence factor pneumolysin when administered in concentrations both above and below their established breakpoint concentrations and it is postulated that this may account for their beneficial clinical effects even when administered to patients with pneumonia infected with macrolide resistant strains of *S. pneumoniae* (Anderson *et al*., 2007, Fukuda *et al*., 2006).

11.5 Role of Clarithromycin in the Management of IPD

Unlike the USA, macrolide monotherapy is only recommended for the treatment of community acquired pneumonia (CAP) in the United Kingdom in patients who have failed to respond to a course of amoxicillin with non severe CAP although dual therapy of a macrolide with amoxicillin is preferred in this circumstance. A macrolide may be used as an alternative to amoxicillin for oral therapy in non-severe CAP where there is evidence of

penicillin hypersensitivity. Erythromycin or clarithromycin may be used intravenously in combination with ampicillin or benzylpenicillin when oral therapy is contraindicated in non severe CAP or in combination with a beta lactamase stable antibiotic like co-amoxyclav or a second or third generation cephalosporin (BTS, 2001).

There is no role for macrolide antibiotics in the management of pneumococcal meningitis nor are they advised in the empiric treatment of otitis media, empyema or cerebral abscess (Gilbert *et al*., 2008).

11.6 Experimental Design to Assess Pneumococcal Gene Expression in the Presence of Subtherapeutic Clarithromycin Concentrations

11.6.1 Growth of Strain South Africa 2507 for RNA Extraction

Three 50ml aliquots of fresh BHI broth taken from the same batch were each inoculated with 500µl of a freshly thawed glycerol stock of strain South Africa 2507 (total viable count = $\log_{10}$ 8.392 CFU/ml). These were incubated in a water bath at 37°C and the optical density monitored until it reached midlog (Optical Density = 0.6 at 600nm).

At this point, 10ml of each of the three cultures was removed and centrifuged at 5000 rpm at room temperature. The supernatant was decanted and the pellet immediately frozen in liquid nitrogen and stored at -80°C prior to RNA extraction.

The remaining culture was split into paired 18ml aliquots. From one of each pair, clarithromycin was added to a final concentration of 5mg/L while no changes were made to its corresponding culture and both samples were returned to the water bath to incubate at 37°C. After 15 minutes (when the cultures were still in the logarithmic growth phase as shown in Figure 11-4), 10ml of culture from each pair was extracted, centrifuged and frozen in liquid nitrogen as above.

11.6.2 RNA Extraction from Strain South Africa 2507

This was performed according to the protocols described in Chapter 2. The quality of the RNA was assessed using an Agilent 2100 Bioanalyzer (Agilent Technologies, United

Kingdom) which identified all samples as having an RNA Integrity Number between 9.8 and 10.

11.6.3 Microarray Comparative Genomic Hybridization Analysis

Three separate RNA preparations were used for each microarray hybridization experiment from each of 3 categories: Strain South Africa 2507 at midlog prior to the addition of clarithromycin versus TIGR4 at midlog, Strain South Africa 2507 15 minutes after reaching midlog but with the addition of clarithromycin 5mg/L versus TIGR4 at midlog and Strain South Africa 2507 15 minutes after reaching midlog without the addition of clarithromycin versus TIGR4 at midlog.

Hybridization and data analysis was performed according to protocols in Chapter 2.

11.6.4 Quantitative RT-PCR Validation of Microarray Data

Using the three biological replicates of RNA prepared above at midlog and then in the presence and absence of clarithromycin 15 minutes after midlog, cDNA was synthesised according to the protocol in Chapter 2. cDNA concentrations were measured using a NanoDrop® ND-1000 spectrophotometer (NanoDrop® Technologies, USA) and the cDNA was then used in a qRT-PCR reaction using SYBR® Green in a Roche Lightcycler® 480 (Roche Applied Science, United Kingdom). The expression of three genes (whose expression was identified on the microarray as being altered in the presence of 5mg/L clarithromycin (SP0740, SP0800 and SP1631)) was examined by qRT-PCR (Figure 11-5). The primer sequences used were those used in the manufacture of the PCR product amplicons utilised in the manufacture of the microarray displayed in Appendix 3. The expression of the genes pneumolysin and *ErmB* was also investigated using qRT-PCR (Figure 11-6). TIGR4 and GyrA RNA were used as a positive control and negative controls were nucleic acid free water (no cDNA negative control) and the original RNA (no reverse transcription control). Standard curves were constructed and qRT-PCR reaction efficiencies were calculated using Microsoft Office Excel 2003 (Microsoft®, United Kingdom) and comparison of the mean normalized gene expressions in the presence and absence of 5mg/L clarithromycin at the same time point was calculated using Q Gene (Muller *et al*., 2002).

11.7 Strain South Africa 2507

11.7.1 Antibiotic sensitivities

The antibiotic sensitivities of South Africa 2507 have previously been determined and published for erythromycin (MIC>256mg/L), clarithromycin (MIC>256mg/L) and ceftriaxone (MIC = 0.5mg/L)(Anderson *et al.*, 2007).

11.7.2 Calculation of Clarithromycin Minimum Inhibitory Concentration (MIC)

Calculation of a MIC for clarithromycin for isolate South Africa 2507 was performed initially using a clarithromycin E-test (AB Biodisk, Sweden) with confluent colonies of South Africa 2507 grown on Mueller Hinton agar with 5% horse blood (E&O Laboratories Limited, United Kingdom) using TIGR4 as a control pneumococcal isolate with known sensitivity to clarithromycin. This demonstrated an MIC of clarithromycin for isolate South Africa 2507 of >256mg/L while the MIC for clarithromycin for TIGR4 was 0.125mg/L.

To further assess the MIC of clarithromycin for South Africa 2507, the agar dilution method of the British Society of Chemotherapy (Andrews, 2001) was adapted to allow assessment of the growth of South Africa 2507 in the presence of 512mg/L and 1024mg/L clarithromycin using TIGR4 as a control isolate sensitive to clarithromycin. This determined a clarithromycin MIC of 1024mg/L for South Africa 2507 (Figure 11-2).

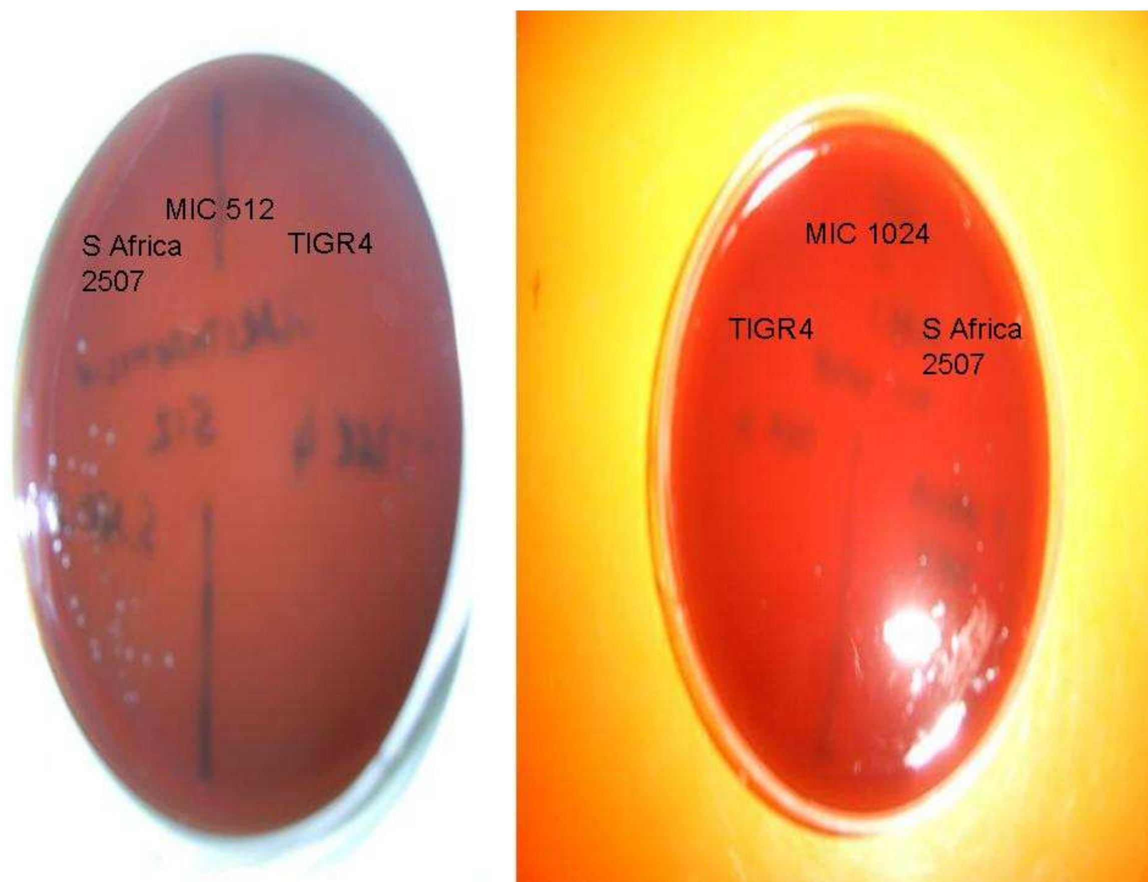


Figure 11-2 Growth of isolate South Africa 2507 to determine MIC to clarithromycin.

Isolates grown on IsoSensitest Agar (Oxoid Limited, United Kingdom) with 5% horse blood (E&O Laboratories Limited, United Kingdom) containing 512 mg/L clarithromycin (Sigma-Aldrich®, United Kingdom) on the left and 1024 mg/L clarithromycin (Sigma-Aldrich®, United Kingdom) on the right. South Africa 2507 colonies are clearly visible in the presence on 512 mg/L clarithromycin while there are no TIGR4 colonies. There were no colonies of either South Africa 2507 or TIGR4 in the presence of clarithromycin 1024 mg/L indicating a clarithromycin MIC for South Africa 2507 of 1024 mg/L.

11.7.3 Growth Curves for Strain South Africa 2507 in Brain Heart Infusion

Isolate South Africa 2507 was grown in BHI broth in order to demonstrate a growth curve and determine when cultures would reach their midlog growth phase for the addition of clarithromycin 5mg/L. Cultures were grown in triplicate.

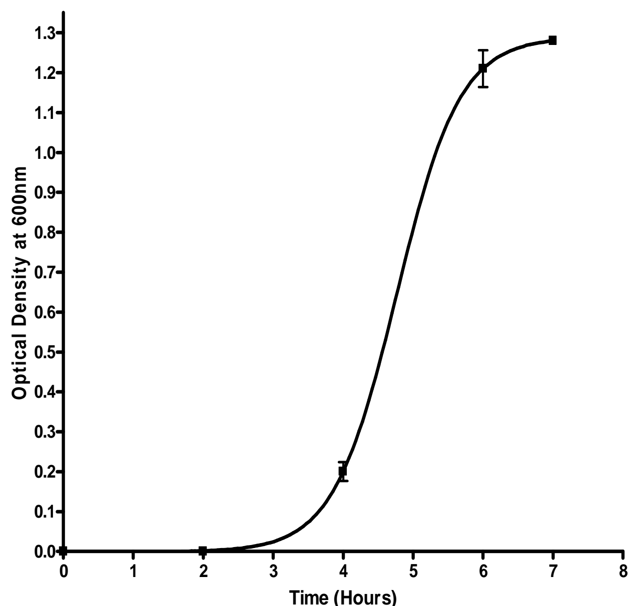


Figure 11-3 Growth curve of isolate South Africa 2507 when grown in Brain Heart Infusion broth.

Optical densities were determined using a WPA Biowave C08000 Cell Density Meter set at 600nm. The curve is constructed using the mean optical density calculated from 3 separate cultures with error bars demonstrating standard errors of the mean optical density.

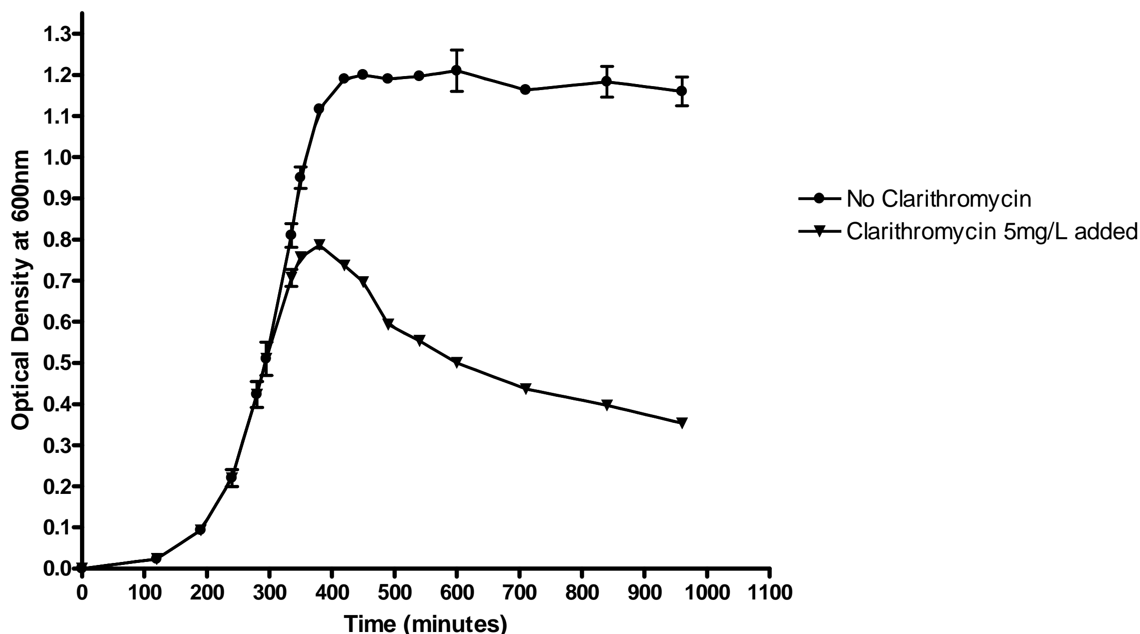


Figure 11-4 Comparison of the growth of South Africa 2507 in Brain Heart Infusion broth without clarithromycin added and the influence of adding clarithromycin.

Clarithromycin was added to a final concentration of 5mg/L when the culture reached an optic density of 0.6 at 600nm. The mean of three data points are shown with error bars indicating standard errors where these were significant enough to be illustrated. In the presence of clarithromycin, there was initially continued growth of the isolate, then growth inhibition and lysis for several hours.

11.7.4 Typing of Strain South Africa 2507

Strain South Africa 2507 has previously been determined as serotype 23F and by BOX-PCR fingerprinting, it is related to the multiply antibiotic resistant pneumococcal Pneumococcal Molecular Epidemiology Network (PMEN) clone Spain^{23F}-1 (Anderson *et al*., 2007). MLST performed at SMPRL found it to be ST81.

11.8 Microarray results

The list of genes identified by Genespring GX 7.3.1 (Agilent Technologies, USA) as significantly up or downregulated are displayed in Appendix 16 (Bonferonni multiple testing correction used) and Appendix 17 (Benjamini and Hochberg multiple testing correction used).

11.9 Quantitative Real Time PCR results

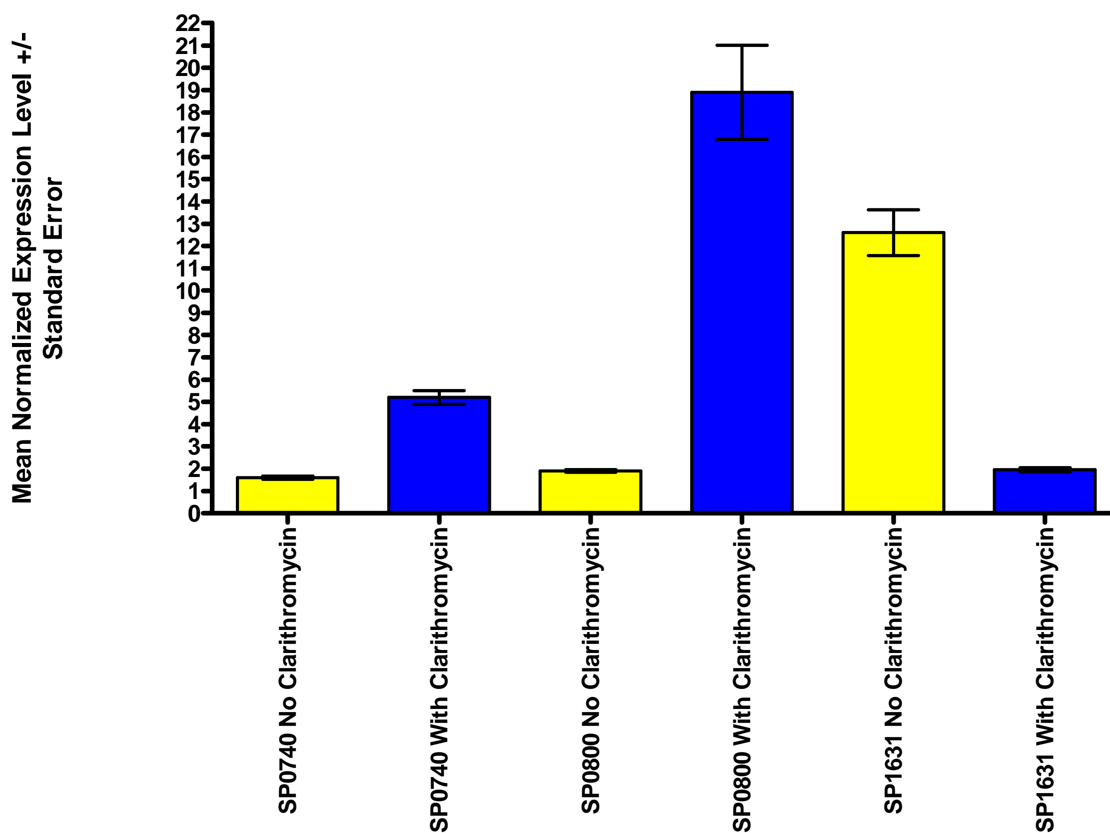


Figure 11-5 Mean Normalized Expression Levels identified by qRT-PCR for genes SP0740, SP0800 and SP1631.

Expression was measured at 15 minutes after midlog in the presence and absence of clarithromycin 5mg/L as calculated by Q-Gene software with standard errors marked. P values were calculated using a one-tailed t-test and were 0.0002, 0.0006 and 0.0002 for SP0740, SP0800 and SP1631 respectively.

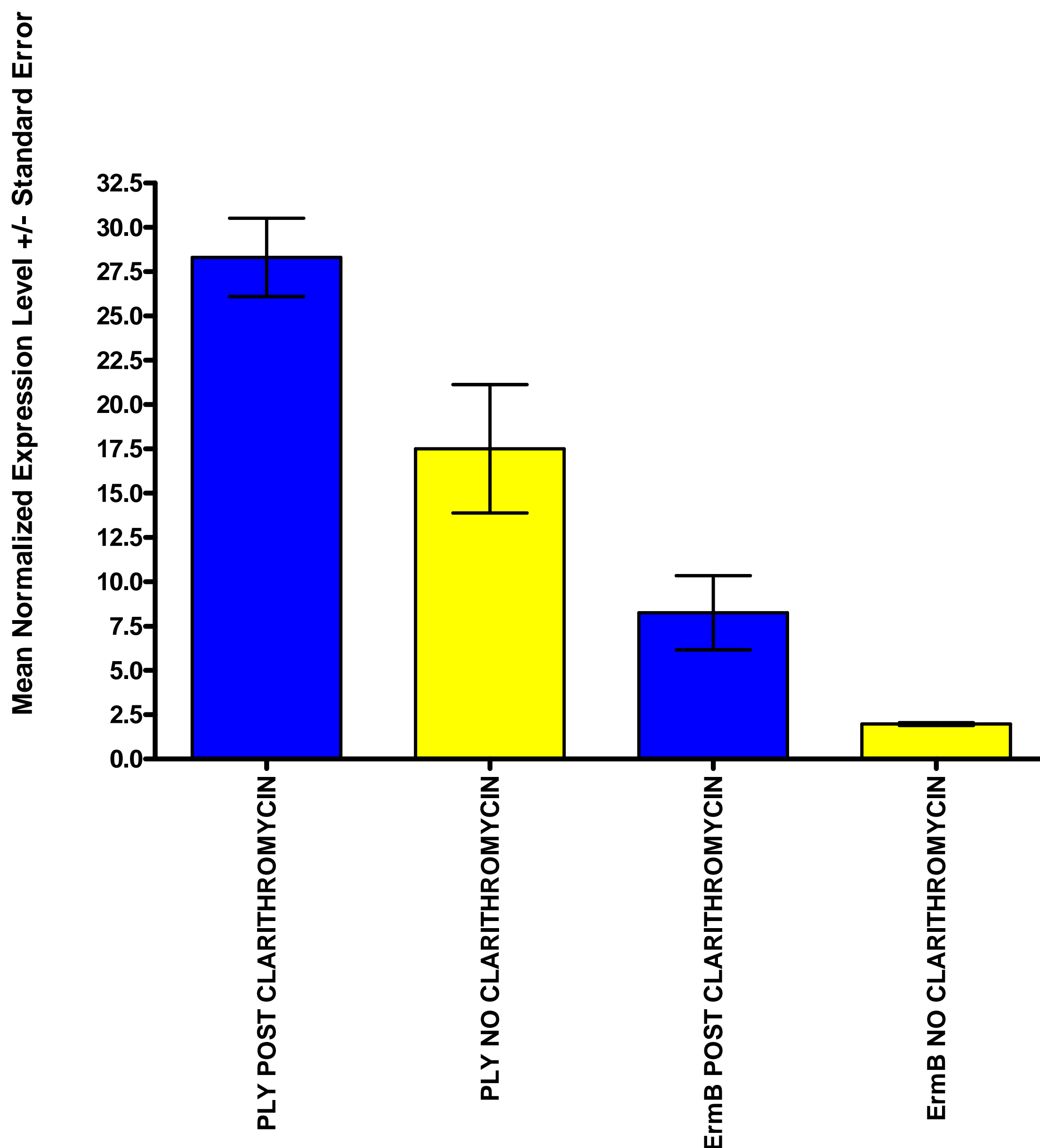


Figure 11-6 Mean Normalized Expression Levels identified by qRT-PCR for genes *ErmB* and pneumolysin (PLY).

Expression was measured at 15 minutes after midlog in the presence and absence of clarithromycin 5mg/L as calculated by Q-Gene with standard errors marked. P values were calculated using a one-tailed t-test and were 0.02 and 0.03 for *ErmB* and pneumolysin respectively.

11.10 Discussion

This series of experiments utilised the SPv1.1 pneumococcal microarray as a tool to examine pneumococcal gene expression and investigate the effect of subtherapeutic, sub-MIC clarithromycin on a multiresistant pneumococcus (South Africa 2507). This highlights the power of this technology when applied to the field of drug discovery as it has an enormous ability to elucidate the mechanisms of action of antimicrobial agents

under predetermined conditions, potentially identifying new targets for antimicrobials, enhancing understanding of resistance mechanisms and aiding in the understanding of the function of, as yet, uncharacterised genes (Shaw *et al.*, 2003). It also allows observation of changes at non target genes or pathways which would have been missed using single gene studies and technologies (Shaw *et al.*, 2003).

In order to achieve this goal, experimental design is of fundamental importance. This required standardisation of every step with the growth of bacteria in parallel in three replicates under identical conditions from the same batch of media and then performing RNA extractions and microarray hybridizations in parallel under identical conditions so as to reduce variation produced by experimental noise rather than gene expression. This included maintaining broth culture at 37°C for as long as possible when removing aliquots for RNA extraction and optical density measurements as temperature change can affect pneumococcal gene expression (Pandya *et al.*, 2005). The benefit of such an approach is also highlighted by Gmuender *et al* in their investigation of the effects of novobiocin and ciprofloxacin on the transcriptome of *Haemophilus influenzae* (Gmuender *et al.*, 2001). Our finding that sampling gene expression from our control culture (without clarithromycin added) at 15 minutes after reaching midlog with the expression at midlog and demonstrating little, if any, change (Figure 11-7) suggests that there have been no significant artefacts of stress responses introduced as result of the manipulation of cultures and the process of splitting them at midlog (Conway and Schoolnik, 2003).

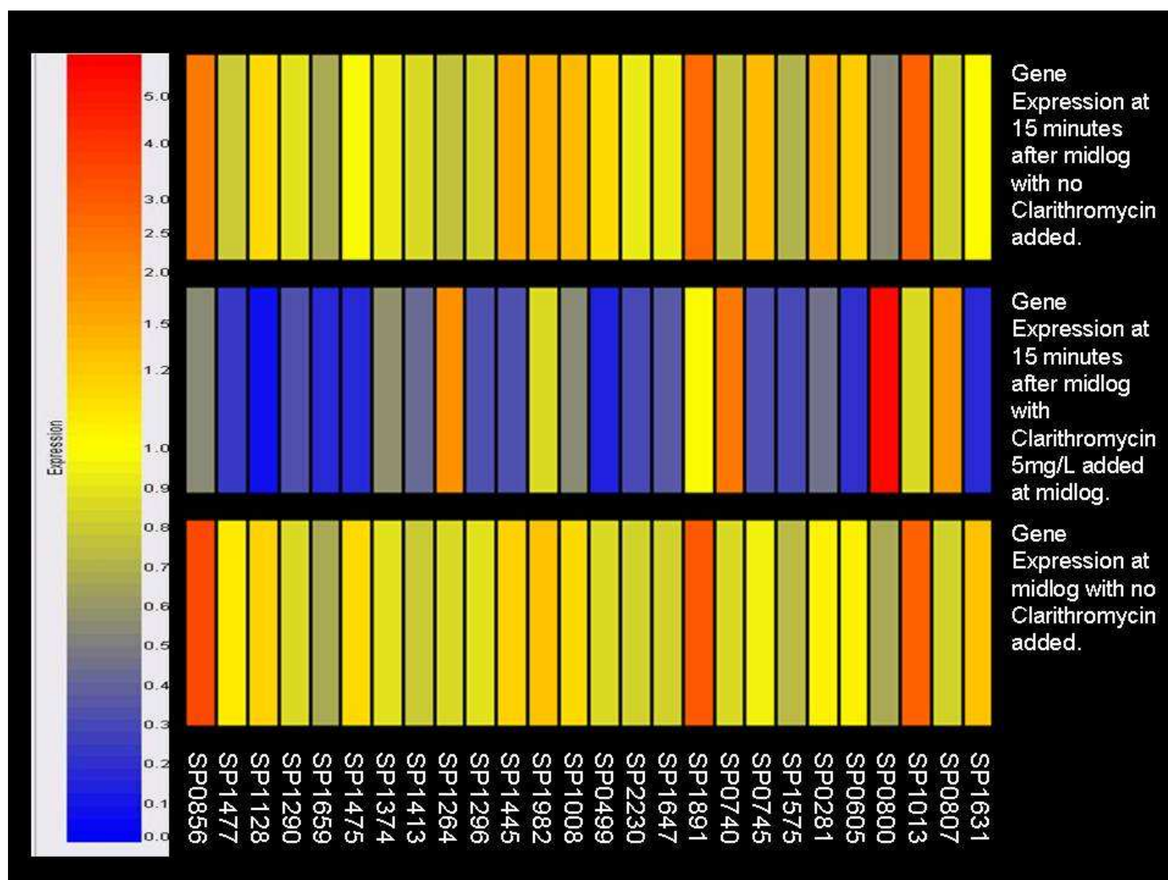


Figure 11-7 Comparison of strain South Africa 2507 gene expression demonstrated by microarray at midlog and 15 minutes later growing in the presence and absence of subtherapeutic clarithromycin (5mg/L).

Genes which were significantly up or down regulated (P<0.05) were identified using Genespring GX 7.3.1 (Agilent Technologies, USA) using a one way ANOVA test with a false discovery rate set at 0.05 and Bonferonni multiple testing correction used. Normalized expression levels at 15 minutes after midlog are colour coded by Genespring GX 7.3.1 according to the spectrum of colours on the left of the diagram with red indicating significant up-regulation when compared to baseline (expression level at 15 minutes after midlog with no clarithromycin added) and blue indicating significant down-regulation compared to this baseline. Comparison of the expression levels of these differentially expressed genes at this baseline versus expression at midlog (prior to exposure to clarithromycin) demonstrates no significant difference in expression indicating that the changes in expression seen in the presence of 5mg/L of clarithromycin are due to the presence of clarithromycin and are not influenced by the more advanced timepoint in the logarithmic growth phase of the bacteria. Genes are identified using the TIGR4 nomenclature.

The choice of antibiotic concentration (5mg/L clarithromycin) had been predetermined to be compatible with previous studies using the South Africa 2507 strain at subtherapeutic concentrations (Anderson *et al*., 2007). There is no consensus regarding the best concentration of antimicrobial to use for such mechanism of action determining transcriptional experiments. Gmuender *et al* advocated using a concentration around that of the MIC (Gmuender *et al*., 2001). Hutter *et al* performed gene expression profiling using a *B. subtilis* spotted microarray with 37 antimicrobials including clarithromycin (Hutter *et al*., 2004). They found that the antimicrobial concentration was crucial for determining the

mechanism of action in microarray expression experiments and that subinhibitory concentrations were best. They also suggested that after one hour of treatment with antimicrobial, the optical density at 600nm of the treated culture should not exceed 15% less than that of its control. Shaw *et al* recommend using low doses and early time points to reduce effects of secondary inhibition by downstream targets (Shaw *et al*., 2003) for their work with *E. coli*. At concentrations below its MIC, erythromycin has been shown to influence the transcription of 5-10% of bacterial genes in *S. typhimurium* which, despite altering gene expression in some cases by 10-100 fold, had little effect on growth (Goh *et al*., 2002). Interestingly at concentrations above the MIC, few transcriptional changes occurred. Freiberg *et al* also make a point of predetermining treatment time and antibiotic dose (Freiberg *et al*., 2004). Brazas and Hancock point out that although most studies suggest using subinhibitory concentrations of antimicrobial and early time points to rule out the complicating effects of secondary targets this may restrict the discovery of effects that contribute to the mechanism of action by nature of them being downstream effects (Brazas and Hancock, 2005). They advocate testing multiple agents under multiple conditions to elucidate direct from indirect effects (Brazas and Hancock, 2005). It is important to appreciate that transcript levels do not reflect all regulatory processes in a cell as they cannot demonstrate the effects of post-transcriptional modification of proteins (Frieberg and Brunner, 2002) and so a proteomic approach (Bandow *et al*., 2003) is also beneficial although more laborious (Freiberg *et al*., 2004).

In this study, it has been possible to analyse our results by ANOVA using Bonferonni multiple testing correction to identify a stringent set of genes whose expression have been altered by the presence of clarithromycin, (Appendix 16). The majority of these are involved in amino acid synthesis (valine, leucine, isoleucine, glycine, serine, threonine, lysine, phenylalanine, tyrosine and tryptophan).

When the ANOVA analysis is performed using the less stringent Benjamini and Hochberg multiple testing correction (Appendix 17) a larger number of significant genes is generated, which gives a broader insight into the multiple effects on the transcriptome. The functions of these genes include ribsosomal proteins, aminoacyl t-RNA synthetases, translation factors as well as amino acid biosynthesis. This is significant as these are all required for fully functioning ribosomes (Jenni and Ban, 2003). It is of particular interest that the genes for ribosomal proteins L4 and L23 are significantly upregulated in the presence of subtherapeutic clarithromycin as these proteins are fundamental to the escape of nascent polypeptides from the ribosomal exit tunnel with L23 mediating the interaction of such polypeptides with cytosolic chaperones and protein targeting factors with L4

known to be a site targeted by macrolide antibiotics (Jenni and Ban, 2003). L23 is involved in protein translation and folding. Mutations of L23 result in the accumulation of misfolded proteins which can be lethal to bacteria (Kramer *et al*., 2002). In addition, L23 is also implicated in the recruitment of the signal recognition particle for the targeting of secretory and transmembrane proteins (Keenan *et al*., 2001). The ribosomal proteins L6, L14 and L7/L12 are also significantly upregulated and these have been identified as part of the structure of a translation factor binding centre in the polypeptide exit channel (Ban *et al*., 1999). Consequently, the upregulation of these genes is consistent with an adaptational response to counter a reduction in function of the ribosome when targeted even by subtherapeutic levels of clarithromycin. The upregulation of ribosomal proteins and aminoacyl t-RNA synthetases in response to exposure to the macrolide erythromycin has also been documented in cultures of *H. influenzae* (Evers *et al*., 2001).

Several of the genes identified by the ANOVA analysis using the Benjamini and Hochberg multiple testing correction (Appendix 17) have been identified as virulence factors in a murine model of serotype 4 pneumococcal pneumonia (Hava and Camilli, 2002). These genes are SP0385, SP0445, SP0622, SP0645, SP0766, SP0797, SP0807, SP0856, SP0943, SP0979, SP0986, SP1032, SP1112, SP1115, SP1154, SP1175, SP1378, SP1591, SP1633, SP1715, SP1780, SP1815, SP1891, SP1970, SP2098, SP2175, SP2176 and SP2239.

As noted above, transcriptional profiling using microarrays has been performed for a variety of micro-organisms with the majority of studies performed on *B. subtilis* and *E. coli* (Frieberg and Brunner, 2002, Freiberg *et al*., 2004). Such studies allow the generation of large databases of expression profiles of the mechanism of action of known antimicrobials against which agents with unknown mechanisms of action can be compared (Hutter *et al*., 2004, Sabina *et al*., 2003, Shaw and Morrow, 2003, Freiberg *et al*., 2005, Freiberg *et al*., 2004). There is however a dearth of data relating to transcriptional profiling of the effects of antimicrobial agents on pneumococci (Freiberg *et al*., 2004).

It has been demonstrated that pneumococci respond to antibiotic stress (from aminoglycosides and fluoroquinolones but not erythromycin or tetracycline) by activating their competence regulatory cascade and increasing their rate of genetic exchange (Prudhomme *et al*., 2006). It is also noteworthy that this study did not see any effect of clarithromycin on the expression of genes involved in competence. This may have been a result of the time point chosen being too early or too late to see the activation of competence genes, as it is characteristically a short lived effect, but the result is compatible with the observation by Prudhomme *et al* that erythromycin had no effect on competence

(Prudhomme *et al.*, 2006). This antibiotic effect on competence has not been profiled using microarrays although they have been used to profile the effect of competence stimulating peptide (CSP) on pneumococci (Peterson *et al.*, 2004) and CSP itself has been shown to have antibacterial activity (Oggioni *et al.*, 2004). Only two microarray based studies have been published which look directly at the transcriptional profile changes of pneumococci directly in response to antimicrobials – the effects of puromycin, tetracycline, chloramphenicol and erythromycin on R6 gene expression (Ng *et al.*, 2003) and the effects of vancomycin on TIGR4 gene expression (Haas *et al.*, 2005).

Ng *et al* performed microarray gene expression studies of *Streptococcus pneumoniae* with different classes of translation inhibiting antibiotics at sublethal concentrations (Ng *et al.*, 2003). Although their experimental methodology differed in many respects to those used here, as they utilised an Affymetrix array based on the genome of the non virulent pneumococcus R6 and exposed their cultures to 10 minutes of four different classes of translation inhibiting antibiotics (puromycin, tetracycline, chloramphenicol and erythromycin), our results with clarithromycin are remarkably similar to those obtained with the structurally similar erythromycin. Both studies identified altered regulation of genes from the following functional categories: ribosomal proteins (80% congruence between studies), aminoacyl tRNA synthetases (100% congruence), translation factors (50% congruence) and amino acid biosynthesis (58% congruence). The table below outlines the similarities between the results from both studies and compares the degree and direction of the change in gene expression.

Functional category and gene	R6 Gene Nomenclature	TIGR4 Gene Nomenclature	Relative Fold Change when exposed to 10 minutes of sublethal erythromycin (Ng *et al,* 2003)	Relative Fold Change when exposed to 15 minutes of 5mg/L clarithromycin.
Ribosomal Proteins				
rpsD	spr0078	SP0085	1.9	4.1
rpsJ	spr0187	SP0208	1.4	1.8
rplC	spr0188	SP0209	1.5	2.2
rplD	spr0189	SP0210	1.6	2.0
rplW	spr0190	SP0211	1.6	2.4
rpsS	spr0192	SP0213	1.7	2.2
rplV	spr0194	SP0214	1.8	1.8
rplP	spr0196	SP0216	1.7	2.6
rpmC	spr0197	SP0217	1.7	2.1
rplX	spr0200	SP0220	1.9	2.0
rplE	spr0201	SP0221	1.7	1.9
rplF	spr0204	SP0225	1.7	2.3
rplR	spr0205	SP0226	1.6	2.1
rpsP	spr0682	SP0775	1.4	1.6
rplJ	spr1212	SP1355	1	2.4
rpsR	spr1394	SP1539	2.3	1.9
Aminoacyl tRNA synthetases				
serS	spr0372	SP0411	-1.9	-3.8
valS	spr0492	SP0568	-3.0	-4.2
pheS	spr0507	SP0579	-13.2	-5.0
pheT	spr0509	SP0581	-2.7	-4.9
glyS	spr1328	SP1474	-2.2	-6.1
glyQ	spr1329	SP1475	-2.5	-6.4
thrS	spr1472	SP1631	-2.3	-6.8
ileS	spr1502	SP1659	-3.6	-4.5
tyrS	spr1910	SP2100	-2.2	-1.9
hisS	spr1931	SP2121	-2.2	-2.9
Translation Factors				
infC	spr0861	SP0959	1.4	1.8
Amino Acid Biosynthesis				
aspC	spr0035	SP0035	-1.8	-1.4
ilvE	spr0758	SP0856	-3.1	-4.3
asd	spr0918	SP1013	-3.0	-3.4
dapA	spr0919	SP1014	-2.1	-2.1
gdhA	spr1181	SP1306	-3.1	-4.4
metA	spr1434	SP1576	-2.1	-2.7
trpC	spr1634	SP1814	-2.2	-4.1
trpD	spr1635	SP1815	-2.4	-5.4
ilvD	spr1935	SP2126	-2.5	-2.0

Table 11-1 Comparison of the effects of sublethal erythromycin and clarithromycin on pneumococcal gene expression.

These similarities suggest there to be a signature pattern of gene expression for erythromycin and clarithromycin as they are structurally similar 14-membered macrolide antibiotics. Ye *et al* note that different classes of compound often generate such a characteristic signature of expression (Ye *et al.*, 2001). Expression profiles generated by different antibiotic classes often differ significantly from each other (Freiberg *et al.*, 2004, Goh *et al.*, 2002) but subtle differences can occur within subclasses of the same antimicrobial class (Gmuender *et al.*, 2001). Also, different classes of compound acting on the same target gene do not always have the same effect (Brazas and Hancock, 2005).

Interestingly, whereas Ng *et al* identified increased amounts of purine biosynthesis in response to all four translation inhibitor classes, particularly affecting the *pur* gene cluster, the only gene from this cluster to be influenced by clarithromycin in South Africa 2507 was *purB* where the effect of clarithromycin was to downregulate expression whereas Ng *et al* found it to be upregulated (Ng *et al.*, 2003). Likewise we did not see any effect of clarithromycin on the putative *PurR* regulator (spr1793) which had been demonstrated by Ng *et al* with erythromycin (Ng *et al.*, 2003), suggesting subtle differences in the influences of different antibiotics from the same class of translation inhibitor. Ng *et al* demonstrated that when sublethal erythromycin influences gene expression in R6, the relative fold change increases with increasing doses of erythromycin. A relationship between degree of gene expression determined by microarray and dose of antibiotic is also documented by Shaw *et al* in *E. coli* to a range of antibiotic classes at a range of doses (Shaw *et al.*, 2003). As only one dose of clarithromycin was used here, it is imposible to comment as to whether this result is reproducible with clarithromycin although it would seem plausible.

In addition to ribosomal effects and translational effect, clarithromycin unexpectedly was found to have effects on the cell membrane in *B. subtilis* (Brazas and Hancock, 2005, Hutter *et al.*, 2004). The effects of erythromycin on the *B. subtilis* transcriptome at several time points have been documented by Lin *et al* with similar effects to this study on ribosomal proteins (Lin *et al.*, 2005).

The genes which feature on our microarray are predominantly taken from the TIGR4 genome. As the TIGR4 isolate is not resistant to macrolide antibiotics, neither the *ermB* gene nor the *mefA* macrolide resistance genes feature on the array. This is an unfortunate limitation of the array as no comment can be made as to the influence of subtherapeutic clarithromycin on the expression levels of these genes in isolate South Africa 2507 by this method. However, by RT-PCR it was possible to demonstrate the upregulation of the *ermB*

and pneumolysin genes (Figure 11-6). The TIGR4 genome sequence does contain two genes which code for putative macrolide efflux proteins – SP1110 and SP0168. SP1110 lies in a region of diversity associated with macrolide efflux by Bruckner *et al* (Bruckner *et al*., 2004). SP1110 is downregulated in the presence of subtherapeutic clarithromycin possibly as survival of the bacteria is not threatened. Expression of SP1110 may be activated to aid survival in an environment where high levels of macrolide compounds are present. The addition of clarithromycin to cultures at 5mg/L does not appear to influence expression of SP0168. It is clear though that subtherapeutic clarithromycin rapidly alters the expression of a constellation of genes in the pneumococcus and not just antibiotic resistance associated genes. This is consistent with the other known effects that clarithromycin has as an immunomodulatory agent (Shinkai *et al*., 2005, Tsai and Standiford, 2004), on suppressing chronic pulmonary inflammation (Chu, 1999), on pneumococcal virulence factors (Speer *et al*., 2003, Anderson *et al*., 2007, Fukuda *et al*., 2006) as well as on transcriptional modulation (Tsui *et al*., 2004). It is also clear that antibiotics do not have single, straight forward targets or effects on bacteria (Brazas and Hancock, 2005, Shaw and Morrow, 2003) and that our understanding of their effects *in vivo* is likely to be far too simplistic. Rather than considering antibiotics acting on single targets, it may be more accurate to regard them as having effects on interconnected networks (Watts and Strogatz, 1998, Conway and Schoolnik, 2003) as has been proposed by Tsui *et al* (Tsui *et al*., 2004) which may explain how seemingly unconnected pathways respond to a single antimicrobial agent. Such a perspective may more readily explain the "natural" role of antimicrobial compounds produced by micro-organisms at low concentrations in the environment (Tsui *et al*., 2004). This should not come as a surprise as it is recognised that subinhibitory antibiotic concentrations have many effects on some bacteria such as disordered growth and altered virulence (Gemmell and Lorian, 1996) which are not as evident at therapeutic concentrations, demonstrating the biphasic adaptive response known as hormesis (Calabrese and Baldwin, 2002).

There are risks to drawing too many conclusions though from this limited two-condition experimental strategy. These results should be placed in an appropriate biological context with corroborating physiological, biochemical and genetic experiments (Conway and Schoolnik, 2003). It would be revealing to expand this work to include further time points and antimicrobial concentrations, different growth media and also different pneumococcal strains (Frieberg and Brunner, 2002). Demonstrating *in vivo* transcriptional responses will be a significant step forward rather than drawing conclusions from exponentially growing planktonic laboratory cultures (Conway and Schoolnik, 2003). This has been attempted for *Vibrio cholerae* with rice water stool from patients with cholera (Merrell *et al*., 2002) and

in tissue cultures of epithelial and endothelial cell lines for *Neisseria meningitidis* (Dietrich *et al*., 2003) but appropriate reproducible strategies have yet to be devised for invasive pneumococcal infections.

12 Concluding Thoughts

Despite the limitations of a CGH approach to genome comparisons due to low sensitivity of hybridization and the realisation that high levels of false negative hybridizations to microarray probes (relating to hypervariable genes) made interpretation of the absence of hybridization unreliable without confirmatory PCR and sequencing, such a CGH approach has allowed documentation of pneumococcal genomic diversity. Further work using this approach, although more expensive, would be better suited to oligonucleotide Affymetrix type arrays since they are able to discriminate single nucleotide polymorphisms and would have fewer false negative hybridizations for hypervariable genes.

In a series of 10 isolates of unrelated serotype and MLSTs, previously undescribed regions of diversity in the pneumococcal genome were discovered. By focusing on isolates of one MLST (ST9), it has been possible to demonstrate less diversity among isolates of the same serotype (serotype 14) than multiple serotypes of the same MLST. Focusing further on isolates matched by serotype and MLST and clinical manifestation it has been possible to show a role that bacteriophages have in generating genomic diversity and to suggest a role that single nucleotide polymorphisms may play in influencing disease manifestations. These studies are compatible with the proposal that pneumococcal populations share and can tap into a “supragenome” which is larger than any genome held by a single isolate. Despite, this high level of genomic diversity at a population level it has also been possible to demonstrate a more clonal structure of pneumococcal populations in the early stages of transmission within an outbreak situation when there has been less time for the many mechanisms which generate diversity to have influenced the pneumococcal genome.

By assessing transcriptional profiles from standardized and reproducible culture and extraction processes, it has been possible to identify genes which are associated with disease phenotypes in serotype 1 associated pneumococcal pneumonia with parapneumonic complications and pneumococcal cerebral abscesses. In addition it has been possible to identify apparently stable genomic differences between serotype 3 isolates from blood and cerebrospinal fluid (CSF) from the same patient and identify differential expression of iron transport proteins in the two body fluids and relate different sugar metabolism pathways in serotype 3 isolates to the presence and expression level or absence of phosphotransferase system genes. Such findings may be important in further understanding the pathogenesis of pneumococcal meningitis or provide a point for therapeutic intervention to prevent invasion into CSF.

Through collecting nasophayngeal isolates from children in a geographical region of substantial biodiversity, 32 new and currently unique MLST profiles have been discovered along with the, previously undescribed, co-incident dual carriage of two different sequence types of the same serotype, providing new insights (when CGH was performed) into the genomic diversity of the pneumococcus and the facilitation that a shared ecological niche and shared serotype may have in gene transfer.

By investigating the effect of a subtherapeutic antibiotic (clarithromycin) on the gene expression of a pneumococcal isolate new insights have been gained into the diversity of genes which respond to an antibiotic rich environment providing insights into how the pneumococcus both responds to this threat to survival and may communicate this response to other pneumococci and allowing a greater understanding of why subtherapeutic antibiotics may have many other clinical effects other than simply bacterial killing.

Such a high degree of genomic diversity in the pneumococcus has undoubtedly contributed to its success as a frequent paediatric nasopharyngeal commensal and a global pathogen (which continues to cause severe manifestations such as pneumonia, meningitis and otitis media) as such diversity benefits the population of bacteria to increase survival in the face of the stresses of a range of host environments and therapeutic interventions. Therefore, these studies of the pneumococcal "supragenome" demonstrate that multiple forces have shaped its evolution such as gene loss/genome reduction, genome rearrangement and acquisition of genetic material through horizontal gene transfer (Fraser-Liggett, 2005) resulting in a dynamic organism and placing it as one of the commonest and most successful pathogens known to man.

13 References

AAKRA, A., NYQUIST, O. L., SNIPEN, L., REIERSEN, T. S. & NES, I. F. (2007) Survey of Genomic Diversity among *Enterococcus faecalis* Strains by Microarray-Based Comparative Genomic Hybridization. *Applied and Environmental Microbiology,* 73**,** 2207-2217.

ADRIAN, P. V., KLUGMAN, K.P. (1997) Mutations in the dihydrofolate reductase gene of trimethoprim-resistant isolates of *Streptococcus pneumoniae*. *Antimicrobial Agents and Chemotherapy,* 41**,** 2406-2413.

AGUIAR, S. I., SERRANO, I., PINTO, F. R., MELO-CRISTINO, J. & RAMIREZ, M. (2008) The presence of the pilus locus is a clonal property among pneumococcal invasive isolates. *BMC Microbiology,* 8**,** 41.

ALBARRACIN ORIO, A. G., CORTES, P. R., TREGNAGHI, M., PINAS, G. E. & ECHENIQUE, J. R. (2008) A new serotype 14 variant of the pneumococcal Spain9V-3 international clone detected in the central region of Argentina. *Journal of Medical Microbiology,* 57**,** 992-999.

ALLEGRUCCI, M. & SAUER, K. (2007) Characterization of colony morphology variants isolated from *Streptococcus pneumoniae* biofilms. *Journal of Bacteriology,* 189**,** 2030-2038.

ALLEGRUCCI, M. & SAUER, K. (2008) The formation of *Streptococcus pneumoniae* non-phase variable colony variants is due to increased mutation frequency present during biofilm growth conditions. *Journal of Bacteriology,* 190**,** 6330-6339.

ALM, R. A., LING, L. L., MOIR, D. T., KING, B. L., BROWN, E. D., DOIG, P. C., SMITH, D. R., NOONAN, B., GUILD, B. C., DEJONGE, B. L., CARMEL, G., TUMMINO, P. J., CARUSO, A., URIA-NICKELSEN, M., MILLS, D. M., IVES, C., GIBSON, R., MERBERG, D., MILLS, S. D., JIANG, Q., TAYLOR, D. E., VOVIS, G. F. & TRUST, T. J. (1999) Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. *Nature,* 397**,** 176-180.

ALSTON, J. M. & STEWART, D. (1930) Incidence of the serological types of the pneumococcus in lobar pneumonia. *British Medical Journal,* 2(3646)**,** 860-862.

ALTSCHUL, S. F., GISH, W., MILLER, W., MYERS, E. W. & LIPMAN, D. J. (1990) Basic Local Alignment Search Tool. *Journal of Molecular Biology,* 215**,** 403-410.

AMEZAGA, M. R., CARTER, P. E., CASH, P. & MCKENZIE, H. (2002) Molecular epidemiology of erythromycin resistance in *Streptococcus pneumoniae* isolates from blood and noninvasive sites. *Journal of Clinical Microbiology,* 40**,** 3313-3318.

AMSDEN, G. W. (2005) Anti-inflammatory effects of macrolides - an underappreciated benefit in the treatment of community-acquired respiratory tract infections and chronic inflammatory pulmonary conditions. *Journal of Antimicrobial Chemotherapy,* 55**,** 10-21.

ANDERSON, R., STEEL, H. C., COCKERAN, R., SMITH, A. M., VON GOTTBERG, A., DE GOUVEIA, L., BRINK, A., KLUGMAN, K. P., MITCHELL, T. J. & FELDMAN, C. (2007) Clarithromycin alone and in combination with ceftriaxone inhibits the production of pneumolysin by both macrolide-susceptible and macrolide-resistant strains of *Streptococcus pneumoniae. Journal of Antimicrobial Chemotherapy,* 59**,** 224-229.

ANDERSSON, B., ERIKSSON, B., FALSEN, E., FOGH, A., HANSON, L. A., NYLEN, O., PETERSON, H. & SVANBORG EDEN, C. (1981) Adhesion of *Streptococcus pneumoniae* to Human Pharyngeal Epithelial Cells *In Vitro*: Differences in Adhesive Capacity Among Strains Isolated from Subjects with Otitis Media, Septicemia, or Meningitis or from Healthy Carriers. *Infection and Immunity,* 32**,** 311-317.

ANDREWS, J. M. (2001) Determination of minimum inhibitory concentrations. *Journal of Antimicrobial Chemotherapy,* 48**,** 5-16.

ANONYMOUS (1970) Clinical Path Conference. Pneumococcal Meningitis and Brain Abscess. *Minnesota Medicine,* 53**,** 431-7.

ANONYMOUS (1992) A nosocomial outbreak of *Streptococcus pneumoniae* infection. *CDR Weekly,* 2**,** 29.

ANONYMOUS (2004) IV Meeting of the Surveillance Network for Emerging and Re-emerging Diseases of the Amazon Region. *Revista de Patologia Tropical,* 33**,** 1-105.

ANONYMOUS (2008) Progress in Introduction of Pneumococcal Conjugate Vaccine Worldwide, 2000-2008. *Morbidity and Mortality Weekly Report,* 57**,** 1148-1151.

ARGUEDAS, A., DAGAN, R., GUEVARA, S., PORAT, N., SOLEY, C., PEREZ, A. & BRILLA, R. (2005) Middle Ear Fluid *Streptococcus pneumoniae* Serotype Distribution in Costa Rican Children with Otitis Media. *Pediatric Infectious Disease Journal,* 24**,** 631-634.

ASTURIAS, E. J., SOTO, M., MENENDEZ, R., RAMIREZ, P. L., RECINOS, F., GORDILLO, R., HOLT, E. & HALSEY, N. A. (2003) Meningitis and pneumonia in Guatemalan children: the importance of *Haemophilus influenzae* type b and *Streptococcus pneumoniae*. *Revista Panamericana de Salud Publica,* 14**,** 377-84.

AUSTRIAN, R. (1957) Pneumococcal Endocarditis, Meningitis, and Rupture of the Aortic Valve. *Archives of Internal Medicine,* 99**,** 539-544.

AVERY, O. T., CHICKERING, H. T., COLE, R. & A.R., D. (1917) Acute Lobar Pneumonia Prevention and Serum Treatment. *Rockefeller Institute Medical Research.,* Mongraph No. 7**,** 1-110.

BAGNOLI, F., MOSCHIONI, M., DONATI, C., DIMITROVSKA, V., FERLENGHI, I., FACCIOTTI, C., MUZZI, A., GIUSTI, F., EMOLO, C., SINISI, A., HILLERINGMANN, M., PANSEGRAU, W., CENSINI, S., RAPPUOLI, R., COVACCI, A., MASIGNANI, V. & BAROCCHI, M. A. (2008) A second pilus type in *Streptococcus pneumoniae* is prevalent in emerging serotypes and mediates adhesion to host cells. *Journal of Bacteriology,* 190**,** 5480-5492.

BAIN, M., AHMAD, N. & ELDER, A. T. (1990) Pneumococcal cross-infection in hospitalized elderly patients. *British Journal of Hospital Medicine,* 44**,** 416.

BAKER, C. I., BARROZO, C. P., RYAN, M. A. K., PEARSE, L. A. & RUSSELL, K. L. (2005) Fatal meningitis in a previously healthy young adult caused by *Streptococcus pneumoniae* serotype 38: an emerging serotype? *BMC Infectious Diseases,* 5**,** 38.

BALTZ, R. H., NORRIS, F. H., MATSUSHIMA, P., DEHOFF, B. S., ROCKEY, P., PORTER, G., BURGETT, S., PEERY, R., HOSKINS, J., BRAVERMAN, L., JENKINS, I., SOLENBERG, P., YOUNG, M., MCHENNEY, M. A., SKATRUD, P. L. & ROSTECK, P. R. (1998) DNA Sequence Sampling of the *Streptococcus pneumoniae* Genome to Identify Novel Targets for Antibiotic Development. *Microbial Drug Resistance,* 4**,** 1-9.

BAN, N., NISSEN, P., HANSEN, J., CAPEL, M., MOORE, P. B. & STEITZ, T. A. (1999) Placement of protein and RNA structures into a 5 A-resolution map of the 50S ribosomal subunit. *Nature,* 400**,** 841-847.

BANDOW, J. E., BROTZ, H., LEICHERT, L. I. O., LABISCHINSKI, H. & HECKER, M. (2003) Proteomic Approach to Understanding Antibiotic Action. *Antimicrobial Agents and Chemotherapy,* 47**,** 948-955.

BARNES, D. M., WHITTIER, S., GILLIGAN, P. H., SOARES, S., TOMASZ, A. & HENDERSON, F. W. (1995) Transmission of multidrug-resistant serotype 23F *Streptococcus pneumoniae* in group day care: evidence suggesting capsular transformation of the resistant strain in vivo. *Journal of Infectious Diseases,* 171**,** 890-6.

BAROCCHI, M. A., RIES, J., ZOGAJ, X., HEMSLEY, C., ALBIGER, B., KANTH, A., DAHLBERG, S., FERNEBRO, J., MOSCHIONI, M., MASIGNANI, V., HULTENBY, K., TADDEI, A. R., BEITER, K., WARTHA, F., VON EULER, A., COVACCI, A., HOLDEN, D. W., NORMARK, S., RAPPUOLI, R. & HENRIQUES-NORMARK, B. (2006) A pneumococcal pilus influences virulence and host inflammatory responses. *Proceedings of the National Academy of Sciences of the United States of America,* 103**,** 2857-62.

BARTLETT, J. G., GORBACH, S. L., THADEPALLI, H. & FINEGOLD, S. M. (1974) Bacteriology of Empyema. *Lancet***,** 338-340.

BASSET, A., TRZCINSKI, K., HERMOS, C., O'BRIEN, K. L., REID, R., SANTOSHAM, M., MCADAM, A. J., LIPSITCH, M. & MALLEY, R. (2007) Association of the pneumococcal pilus with certain capsular serotypes but not with increased virulence. *Journal of Clinical Microbiology,* 45**,** 1684-9.

BATTIG, P. & MUHLEMANN, K. (2007) F-P10 Naturally occurring *Streptococcus pneumoniae* strains with increased mutation rate to optochin and hyperproduction of H_2O_2. *Europneumo 2007. 8th European Meeting on the Molecular Biology of the Pneumococcus.* Oeiras, Portugal.

BECHAMPS, G. J., LYNN, H. B. & WENZL, J. E. (1970) Empyema in children: review of mayo clinic experience. *Mayo Clinic Proceedings,* 45**,** 43-50.

BEHR, M. A., WILSON, M. A., GILL, W. P., SALAMON, H., SCHOOLNIK, G. K., RANE, S. & SMALL, P. M. (1999) Comparative genomics of BCG vaccines by whole-genome DNA microarray. *Science,* 284**,** 1520-1523.

BEKAL, S., BROUSSEAU, R., MASSON, L., PREFONTAINE, G., FAIRBROTHER, J. & HAREL, J. (2003) Rapid identification of *Escherichia coli* pathotypes by virulence gene detection with DNA microarrays. *Journal of Clinical Microbiology,* 41**,** 2113-2125.

BEKRI, H., COHEN, R., VARON, E., MADHI, F., GIRE, R., GUILLOT, F. & DELACOURT, C. (2007) *Streptococcus pneumoniae* serotypes involved in children with pleural empyemas in France. *Archives De Pediatrie,* 14**,** 239-243.

BENAVIDES, J. A., REYES, C., FLITTER, M., MARKOWITZ, J. & RODGERS, G. (2008) P1-037 Epidemiologic Surveillance of Invasive Pneumococcal Disease (IPD) and Chest Radiograph (CXR) - Confirmed Pneumonia (CXR-Pn) in Children in Bogota, Colombia. *6th International Symposium on Pneumococci and Pneumococcal Diseases.* Reykjavik, Iceland.

BENTLEY, S., AANENSEN, D. M., MAVROIDI, A., SAUNDERS, D., RABBINOWITSCH, E., COLLINS, M., DONOHOE, K., HARRIS, D., MURPHY, L., QUAIL, M. A., SAMUEL, G., SKOVSTED, I. C., M.S., K., BARRELL, B., REEVES, P. R., PARKHILL, J. & SPRATT, B. G. (2006) Genetic analysis of the capsular biosynthetic locus from all 90 pneumococcal serotypes. *PLOS Genetics,* 2**,** 262-269.

BERG, S., TROLLFORS, B., PERSSON, E., BACKHAUS, E., LARSSON, P., EK, E., CLAESSON, B. E., JONSSON, L., RADBERG, G., JOHANSSON, S., RIPA, T., KALTOFT, M. S. & KONRADSEN, H. B. (2006) Serotypes of Streptococcus pneumoniae isolated from blood and cerebrospinal fluid related to vaccine serotypes and to clinical characteristics. *Scandinavian Journal of Infectious Diseases,* 38**,** 427-432.

BERK, S. L., GAGE, K. A., HOLTSCLAW-BERK, S. A. & SMITH, J. K. (1985) Type 8 pneumococcal pneumonia: an outbreak on an oncology ward. *Southern Medical Journal,* 78**,** 159-61.

BERRON, S., FENNOLL, A., ORTEGA, M., ARELLANO, N. & CASAL, J. (2005) Analysis of the genetic structure of nontypeable pneumococcal strains isolated from conjunctiva. *Journal of Clinical Microbiology,* 43**,** 1694-1698.

BESCOS, J. G., UGARRIZA, A. V., SANCHO, C. A., ALDANONDO, M. J. A., LAZARO, M. A., ALBERTO, M. J., LOPEZ, W. V. & TOLEDO, M. (2003) *Streptococcus pneumoniae* transmission in a nursing home: analysis of an epidemic outbreak. *Medicina Clinica,* 121**,** 48-52.

BIJLSMA, J. J. E., BURGHOUT, P., KLOOSTERMAN, T. G., BOOTSMA, H. J., DE JONG, A., HERMANS, P. W. & KUIPERS, O. P. (2007) Development of Genomic Array Fingerprinting for Identification of Conditionally Essential Genes in *Streptococcus pneumoniae*. *Applied and Environmental Microbiology,* 73**,** 1514-1524.

BIRTLES, A., MCCARTHY, N., SHEPPARD, C. L., RUTTER, H., GUIVER, M., HAWORTH, E. & GEORGE, R. C. (2005) Multilocus sequence typing directly on

DNA from clinical samples and a cultured isolate to investigate linked fatal pneumococcal disease in residents of a shelter for homeless men. *Journal of Clinical Microbiology,* 43**,** 2004-2008.

BJORKHOLM, B. M., OH, J. D., FALK, P. G., ENGSTRAND, L. G. & GORDON, J. I. (2001) Genomics and proteomics converge on *Helicobacter pylori*. *Current Opinion in Microbiology,* 4**,** 237-245.

BLAKE, F. G. (1931) Observations on pneumococcus Type III pneumonia. *Annals of Internal Medicine,* 5**,** 673-686.

BLANC, P., DUBUS, J. C., BOSDURE, E. & MINODIER, P. (2007) Community-acquired parapneumonic effusion in children: what's new? *Archives De Pediatrie,* 14**,** 64-72.

BLOMBERG, C., DAGERHAMN, J., MORFELDT, E., SJOSTROM, K., BROWALL, S., NORMARK, S. & NORMARK, B. H. (2007) A-O5 Comparative genomics of pneumococcal clonal types with different invasive disease potential. *Europneumo 2007. 8th European Meeting on the Molecular Biology of the Pneumococcus.* Oeiras, Portugal.

BLUE, C. E. & MITCHELL, T. J. (2003) Contribution of a Response Regulator to the Virulence of *Streptococcus pneumoniae* Is Strain Dependent. *Infection and Immunity,* 71**,** 4405-4413.

BOERSMA, W. G., LOWENBERG, A., HOLLOWAY, Y., KUTTSCHRUTTER, H., SNIJDER, J. A. & KOETER, G. H. (1993) Rapid detection of pneumococcal antigen in pleural fluid of patients with community acquired pneumonia. *Thorax,* 48**,** 160-2.

BOGAERT, D., DE GROOT, R. & HERMANS, P. W. M. (2004) *Streptococcus pneumoniae* colonisation: the key to pneumococcal disease. *Lancet Infectious Diseases,* 4**,** 144-154.

BOHR, V., PAULSON, O. B. & RASMUSSEN, N. (1984) Pneumococcal Meningitis. Late Neurologic Sequelae and Features of Prognostic Impact. *Archives of Neurology,* 41**,** 1045-9.

BOLES, B. R., THOENDEL, M. & SINGH, P. K. (2004) Self-generated diversity produces "insurance effects" in biofilm communities. *Proceedings of the National Academy of Sciences,* 101**,** 16630-16635.

BOOTSMA, H. J., BURGHOUT, P., BIJLSMA, J. J. E., KLOOSTERMAN, T. G., KUIPERS, O. P. & HERMANS, P. W. M. (2007) A-O4 Identification of pneumococcal genes essential for infection by genomic array fingerprinting. *Europneumo 2007. 8th European Meeting on the Molecular Biology of the Pneumococcus.* Oeiras, Portugal.

BORER, A., MEIRSON, H., PELED, N., PORAT, N., DAGAN, R., FRASER, D., GILAD, J., ZEHAVI, N. & YAGUPSKY, P. (2001) Antibiotic-Resistant Pneumococci Carried by Young Children Do Not Appear to Disseminate to Adult Members of a Closed Community. *Clinical Infectious Diseases,* 33**,** 436-444.

BRANDILEONE, M. C., DI FABIO, J. L., VIEIRA, V. S. D., ZANELLA, R. C., CASAGRANDE, S. T., PIGNATARI, A. C. & TOMASZ, A. (1998) Geographic Distribution of Penicillin Resistance of *Streptococcus penumoniae* in Brazil: Genetic Relatedness. *Microbial Drug Resistance,* 4**,** 209-217.

BRANDILEONE, M. C. C., DE ANDRADE, A., DI FABIO, J. L., GUERRA, M. & AUSTRIAN, R. (2003) Appropriateness of a pneumococcal conjugate vaccine in Brazil: Potential impact of age and clinical diagnosis, with emphasis on meningitis. *Journal of Infectious Diseases,* 187**,** 1206-1212.

BRAUN, J. S., SUBLETT, J., FREYER, D., MITCHELL, T. J., CLEVELAND, J. L., TUOMANEN, E. I. & WEBER, J. R. (2002) Pneumococcal pneumolysin and H_2O_2 mediate brain cell apoptosis during meningitis. *The Journal of Clinical Investigation,* 109**,** 19-27.

BRAZAS, M. D. & HANCOCK, R. E. W. (2005) Using microarray gene signatures to elucidate mechanisms of antibiotic action and resistance. *Drug Discovery Today: Targets,* 10**,** 1245-1252.

BRAZMA, A., HINGAMP, P., QUACKENBUSH, J., SHERLOCK, G., SPELLMAN, P., STOECKERT, C., AACH, J., ANSORAGE, W., BALL, C. A., CAUSTON, H. C., GAASTERLAND, T., GLENISSON, P., HOLSTEGE, F. C. P., KIM, I. F., MARKOWITZ, V., MATESE, J. C., PARKINSON, H., ROBINSON, A., SARKANS, U., SCHULZE-KREMER, S., STEWART, J., TAYLOR, R., VILO, J. & VINGRON, M. (2001) Minimum information about a microarray experiment (MIAME) - toward standards for microarray data. *Nature Genetics,* 29**,** 365-371.

BRESNITZ, E., GRANT, C., OSTRAWSKI, S., MORRIS, C., CALABRIA, J., REETZ, B. & CLUGSTON, S. (2001) Outbreak of Pneumococcal Pneumonia Among Unvaccinated Residents of a Nursing Home - New Jersey, April 2001. *Morbidity and Mortality Weekly Report,* 50**,** 707-710.

BROOKS-WALTER, A., BRILES, D. E. & HOLLINGSHEAD, S. K. (1999) The *pspC* Gene of *Streptococcus pneumoniae* Encodes a Polymorphic Protein, PsP C, Which Elicits Cross-Reactive Antibodies to PspA and Provides Immunity to Pneumococcal Bacteremia. *Infection and Immunity,* 67**,** 6533-6542.

BROWN, J. S., GILLILAND, S. M., SPRATT, B. G. & HOLDEN, D. W. (2004) A locus contained within a variable region of pneumococcal pathogenicity island 1 contributes to virulence in mice. *Infection and Immunity,* 72**,** 1587-93.

BRUCKNER, R., NUHN, M., REICHMANN, P., WEBER, B. & HAKENBECK, R. (2004) Mosaic genes and mosaic chromosomes-genomic variation in *Streptococcus pneumoniae*. *International Journal of Medical Microbiology,* 294**,** 157-68.

BRUEGGEMANN, A. B., ARROYO, L. A., HAUSDORFF, W. P., SANCHEZ-TATAY, D., MORENO, D., TORRONTERAS, R., MATEOS, I., FENOLL, A. & OBANDO, I. (2006) PO3.38 Pneumococcal epidemiology in Sevilla and Malaga, Spain. *5th International Symposium on Pneumococci and Pneumococcal Diseases.* Alice Springs, Australia.

BRUEGGEMANN, A. B., GRIFFITHS, D. T., MEATS, E., PETO, T., CROOK, D. W. & SPRATT, B. G. (2003) Clonal Relationships between Invasive and Carriage

Streptococcus pneumoniae and Serotype- and Clone-Specific Differences in Invasive Disease Potential. *Journal of Infectious Diseases,* 187**,** 1424-32.

BRUEGGEMANN, A. B., PAI, R., CROOK, D. W. & BEALL, B. (2007) Vaccine Escape Recombinants Emerge after Pneumococcal Vaccination in the United States. *PLOS Pathogens,* 3**,** 1628-1636.

BRUEGGEMANN, A. B., PETO, T. E. A., CROOK, D. W., BUTLER, J. C., KRISTINSSON, K. G. & SPRATT, B. G. (2004) Temporal and Geographic Stability of the Serogroup-Specific Invasive Disease Potential of *Streptococcus pneumoniae* in Children. *Journal of Infectious Diseases,* 190**,** 1203-11.

BRUEGGEMANN, A. B. & SPRATT, B. G. (2003) Geographic Distribution and Clonal Diversity of *Streptococcus pneumoniae* Serotype 1 Isolates. *Journal of Clinical Microbiology,* 41**,** 4966-4970.

BRUYN, G. A. W., THOMPSON, J. & VAN DER MEER, J. W. M. (1990) Pneumococcal Endocarditis in Adult Patients. A Report of Five Cases and Review of the Literature. *Quarterly Journal of Medicine,* 74**,** 33-40.

BRYANT, P., VENTER, D., ROBINS-BROWNE, R. & CURTIS, N. (2004) Chips with everything: DNA microarrays in infectious diseases. *Lancet Infectious Diseases,* 4**,** 100-111.

BTS (2001) British Thoracic Society Guidelines for the Management of Community Acquired Pneumonia in Adults. *Thorax,* 56**,** 1-64.

BUCK, J. M., LEXAU, C., SHAPIRO, M., GLENNEN, A., BOXRUD, D. J., KOZIOL, B., WHITNEY, C. G., BEALL, B., DANILA, R. & LYNFIELD, R. (2006) A community outbreak of conjunctivitis caused by nontypeable Streptococcus pneumoniae in Minnesota. *Pediatric Infectious Disease Journal,* 25**,** 906-911.

BURGHOUT, P., BOOTSMA, H. J., KLOOSTERMAN, T. G., BIJLSMA, J. J., DE JONGH, C. E., KUIPERS, O. P. & HERMANS, P. W. (2007) Search for genes essential for pneumococcal transformation: the RADA DNA repair protein plays a role in genomic recombination of donor DNA. *Journal of Bacteriology,* 189**,** 6540-50.

BURMAN, L. A., TROLLFORS, B., NORRBY, R., FALSEN, E., HAIDL, S. & HENRICHSEN, J. (1986) Serotype Distribution of *Streptococcus pneumoniae* Strains Isolated from Blood and Cerebrospinal Fluid in Sweden. *Scandinavian Journal of Infectious Diseases,* 18**,** 45-48.

BYINGTON, C. (2007) Empyema in children in the USA; the possible impact of introduction of routine pneumococcal vaccination. *European Respiratory Society Annual Congress.* Stockholm, Sweden.

BYINGTON, C., MASON, E., KORGENSKI, K., DALY, J. & HULTEN, K. (2008) P2-015 Molecular Epidemiology of Pediatric Pneumococcal Empyema in Utah 2001-2007. *6th International Symposium on Pneumococci and Pneumococcal Diseases.* Reykjavik, Iceland.

BYINGTON, C. L., KORGENSKI, K., AMPOFO, K., DALY, J., PAVIA, A. T. & MASON, E. O. (2005a) Pediatric Pneumococcal Empyema: Increasing Rates and

Emerging Serotypes. *43rd Annual Meeting of the Infectious Disease Society of America.* San Francisco, USA.

BYINGTON, C. L., KORGENSKI, K., DALY, J., AMPOFO, K., PAVIA, A. & MASON, E. O. (2006) Impact of the pneumococcal conjugate vaccine on pneumococcal parapneumonic empyema. *Pediatric Infectious Disease Journal,* 25**,** 250-4.

BYINGTON, C. L., KORGENSKI, K., DALY, J., SAMORE, M. H. & PAVIA, A. T. (2005b) Reply to Calbo and Garau. *Clinical Infectious Diseases,* 41**,** 1822.

BYINGTON, C. L., SAMORE, M. H., STODDARD, G. J., BARLOW, S., DALY, J., KORGENSKI, K., FIRTH, S., GLOVER, D., JENSEN, J., MASON, E. O., SHUTT, C. K. & PAVIA, A. T. (2005c) Temporal trends of invasive disease due to *Streptococcus pneumoniae* among children in the Intermountain West: Emergence of nonvaccine serogroups. *Clinical Infectious Diseases,* 41**,** 21-29.

BYINGTON, C. L., SPENCER, L. Y., JOHNSON, T. A., PAVIA, A. T., ALLEN, D., MASON, E. O., KAPLAN, S., CARROLL, K. C., DALY, J. A., CHRISTENSON, J. C. & SAMORE, M. H. (2002) An epidemiological investigation of a sustained high rate of pediatric parapneumonic empyema: Risk factors and microbiological associations. *Clinical Infectious Diseases,* 34**,** 434-440.

CALABRESE, E. J. & BALDWIN, L. A. (2002) Defining hormesis. *Human and Experimental Toxicology,* 21**,** 91-97.

CALBO, E. & GARAU, J. (2005) Invasive pneumococcal disease in children: changing serotypes and clinical expression of disease. *Clinical Infectious Diseases,* 41**,** 1821-2.

CALDER, M. A., MCHARDY, V. U. & SCHONELL, M. E. (1970) Importance of Pneumococcal Typing in Pneumonia. *Lancet,* i**,** 5-7.

CALL, D. R. (2005) Challenges and Opportunities for Pathogen Detection Using DNA Microarrays. *Critical Reviews in Microbiology,* 31**,** 91-99.

CALL, D. R., BAKKO, M. K., KRUG, M. J. & ROBERTS, M. C. (2003) Identifying antimicrobial resistance genes with DNA microarrays. *Antimicrobial Agents and Chemotherapy,* 47**,** 3290-3295.

CAMARGOS, P., FISCHER, G. B., MOCELIN, H., DIAS, C. & RUVINSKY, R. (2006) Penicillin resistance and serotyping of *Streptococcus pneumoniae* in Latin America. *Paediatric Respiratory Reviews,* 7**,** 209-214.

CAMPBELL, P. W. (1995) New developments in pediatric pneumonia and empyema. *Current Opinion in Pediatrics,* 7**,** 278-82.

CARLISLE, J. B., GRATTEN, M. & LEACH, A. J. (2001) Molecular epidemiology of multiple drug resistant type 6B *Streptococcus pneumoniae* in the Northern Territory and Queensland, Australia. *Epidemiology and Infection,* 126**,** 25-29.

CARTER, R. J., SORENSEN, G., HEFFERNAN, R., KIEHLBAUCH, J. A., KORNBLUM, J. S., LEGGIADRO, R. J., NIXON, L. J., WERTHEIM, W. A., WHITNEY, C. G. & LAYTON, M. (2005) Failure to control an outbreak, of multidrug-resistant *Streptococcus pneumoniae* in a long-term-care facility:

Emergence and ongoing transmission of a fluoroquinolone-resistant strain. *Infection Control And Hospital Epidemiology,* 26**,** 248-255.

CARTMILL, T. D. & PANIGRAHI, H. (1992) Hospital outbreak of multiresistant *Streptococcus pneumoniae*. *Journal of Hospital Infection,* 20**,** 130-2.

CARVALHO, M. G., GERTZ, R., LI, Z., MOORE, M. & BEALL, B. (2008) P1-017 Population based strain surveillance of invasive serotype 19A pneumococci recovered in the United States: 2006. *6th International Symposium on Pneumococci and Pneumococcal Diseases.* Reykjavik, Iceland.

CARVALHO, M. G., STEIGERWALT, A. G., THOMPSON, T., JACKSON, D. & FACKLAM, R. R. (2003) Confirmation of nontypeable *Streptococcus pneumoniae*-like organisms isolated from outbreaks of epidemic conjunctivitis as *Streptococcus pneumoniae*. *Journal of Clinical Microbiology,* 41**,** 4415-7.

CARVER, T. J., RUTHERFORD, K. M., BERRIMAN, M., RAJANDREAM, M. A., BARRELL, B. G. & PARKHILL, J. (2005) ACT: the Artemis comparison tool. *Bioinformatics,* 21**,** 3422-3423.

CASH, P., ARGO, E., FORD, L., LAWRIE, L. & MCKENZIE, H. (1999) A proteomic analysis of erythromycin resistance in *Streptococcus pneumoniae*. *Electrophoresis,* 20**,** 2259-2268.

CASSONE, M., D'ANDREA, M. M., IANNELLI, F., OGGIONI, M. R., ROSSOLINI, G. M. & POZZI, G. (2006) DNA Microarray for Detection of Macrolide Resistance Genes. *Antimicrobial Agents and Chemotherapy,* 50**,** 2038-2041.

CASSONE, M., GIORDANO, A. & POZZI, G. (2007) Bacterial DNA microarrays for clinical microbiology: the early logarithmic phase. *Frontiers in Bioscience,* 12**,** 2258-2269.

CASTANEDA, E., PENUELA, I., VELA, M. C. & TOMASZ, A. (1998) Penicillin-resistant *Streptococcus pneumoniae* in Colombia: Presence of international epidemic clones. *Microbial Drug Resistance,* 4**,** 233-239.

CASTANHEIRA, M., JONES, R. N., SILBERT, S., BRANDILEONE, M. C. & SADER, H. S. (2003) Typing and molecular characterization of *Streptococcus pneumoniae* with reduced susceptibility to cefotaxime isolated in Latin America. *Microbial Drug Resistance,* 9**,** 345-51.

CECIL, R. L., BALDWIN, H. S. & LARSEN, N. P. (1927) Lobar Pneumonia. A Clinical and Bacteriological Study of Two Thousand Typed Cases. *Archives of Internal Medicine,* 40**,** 253-280.

CHANCEY, S., GUDLAVALLETI, A., BAUGHMAN, W., MCGEE, L., BEALL, B., FARLEY, M., KLUGMAN, K. & STEPHENS, D. (2008) P2-049 Expansion of a Multidrug Resistant Clone of *Streptococcus pneumoniae* Serotype 19A Harboring Dual Mechanisms of Macrolide Resistance. *6th International Symposium on Pneumococci and Pneumococcal Diseases.* Reykjavik, Iceland.

CHARALAMBOUS, B. M., ORIYO, N. M. & GILLESPIE, S. H. (2008) How valid is single colony isolation for surveillance of *Streptococcus pneumoniae* carriage? *Journal of Clinical Microbiology,* 46**,** 2467-2468.

CHERIAN, T., STEINHOFF, M. C., HARRISON, L. H., ROHN, D., MCDOUGAL, L. K. & DICK, J. (1994) A Cluster Of Invasive Pneumococcal Disease In Young-Children In Child-Care. *Journal of the American Medical Association,* 271**,** 695-697.

CHIAVOLINI, D., MEMMI, G., MAGGI, T., IANNELLI, F., POZZI, G. & OGGIONI, M. R. (2003) The three extra-cellular zinc metalloproteinases of *Streptococcus pneumoniae* have a different impact on virulence in mice. *BMC Microbiology,* 3**,** 14.

CHIZHIKOV, V., RASOOLY, A., CHUMAKOV, K. & LEVY, D. D. (2001) Microarray analysis of microbial virulence factors. *Applied and Environmental Microbiology,* 67**,** 3258-3263.

CHIZHIKOV, V., WAGNER, M., IVSHINA, A., HOSHINO, Y., KAPIKIAN, A.Z., CHUMAKOV, K. (2002) Detection and Genotyping of Human Group A Rotaviruses by Oligonucleotide Microarray Hybridisation. *Journal of Clinical Microbiology,* 40**,** 2398-2407.

CHONMAITREE, T. & POWELL, K. R. (1983) Parapneumonic pleural effusion and empyema in children. Review of a 19-year experience, 1962-1980. *Clinical Pediatrics,* 22**,** 414-9.

CHRISTIE, I. M. (1932) Type-Specific Organisms in Acute Pneumonia and in Convalescents and Contacts. *The Lancet,* ii**,** 1173-1176.

CHU, D. T. W. (1999) Recent developments in macrolides and ketolides. *Current Opinion in Microbiology,* 2**,** 467-474.

CHUANG, S. E., DANIELS, D. L. & BLATTNER, F. R. (1993) Global regulation of Gene Expression in *Escherichia coli*. *Journal of Bacteriology,* 175**,** 2026-2036.

CHUAQUI, R. E., BONNER, R. F., BEST, C. J. M., GILLESPIE, J. W., FLAIG, M. J., HEWITT, S. M., PHILLIPS, J. L., KRIZMAN, D. B., TANGREA, M. A., AHRAM, M., LINEHAN, W. M., KNEZEVIC, V. & EMMERT-BUCK, M. R. (2002) Post-analysis follow-up and validation of microarray experiments. *Nature Genetics Supplement,* 32**,** 509-514.

CHURCHILL, G. A. (2002) Fundamentals of experimental design for cDNA microarrays. *Nature Genetics Supplement,* 32**,** 490-495.

CLAGETT, O. T. (1973) Changing aspects of the etiology and treatment of pleural empyema. *Surgical Clinics of North America,* 53**,** 863-6.

CLARKE, S. C. & DIGGLE, M. A. (2002) Automated PCR/ Sequence Template Purification. *Molecular Biotechnology,* 21**,** 221-224.

CLARKE, S. C., LAWRIE, D. & DIGGLE, M. A. (2004a) Genetic relatedness of antibiotic-resistant pneumococci isolated during case clusters. *Journal of Medical Microbiology,* 53**,** 1097-99.

CLARKE, S. C., REID, J. A., THOM, L., DIGGLE, M. A., MCCHLERY, S., SCOTT, K. & EDWARDS, G. F. S. (2005) Erythromycin resistance among invasive

pneumococci in Scotland, 1994-2003. *British Journal of Biomedical Science,* 62**,** 28-30.

CLARKE, S. C., SCOTT, K. J. & MCCHLERY, S. M. (2004b) Erythromycin resistance in invasive serotype 14 pneumococci is highly related to clonal type. *Journal of Medical Microbiology,* 53**,** 1101-1103.

CLARKE, S. C., SCOTT, K. J. & MCCHLERY, S. M. (2004c) Serotypes and Sequence Types of Pneumococci Causing Invasive Disease in Scotland Prior to the Introduction of Pneumococcal Conjugate Polysaccharide Vaccines. *Journal of Clinical Microbiology,* 42**,** 4449-4452.

CLAVERYS, J. P., PRUDHOMME, M., MORTIER-BARRIERE, I., MARTIN, B. (2000) Adaptation to the environment: *Streptococcus pneumoniae*, a paradigm for recombination-mediated genetic plasticity. *Molecular Microbiology,* 35**,** 251-259.

CLEWELL, D. B., FLANNAGAN, S. E. & JAWORSKI, D. D. (1995) Unconstrained bacterial promiscuity: The Tn*916* -Tn*1545* family of conjugative transposons. *Trends in Microbiology,* 3**,** 229-236.

COFFEY, T. J., DANIELS, M., ENRIGHT, M. C. & SPRATT, B. G. (1999) Serotype 14 variants of the Spanish penicillin-resistant serotype 9V clone of *Streptococcus pneumoniae* arose by large recombinational replacements of the cpsA-pbp1a region. *Microbiology,* 145**,** 2023-31.

COFFEY, T. J., DOWSON, C. G., DANIELS, M. & SPRATT, B. G. (1993) Horizontal spread of an altered penicillin-binding protein 2B gene between *Streptococcus pneumoniae* and *Streptococcus oralis*. *FEMS Microbiology Letters,* 110**,** 335-340.

COFFEY, T. J., DOWSON, C. G., DANIELS, M., ZHOU, J., MARTIN, C., SPRATT, B. G. & MUSSER, J. M. (1991) Horizontal transfer of multiple penicillin-binding protein genes, and capsular biosynthetic genes, in natural populations of *Streptococcus pneumoniae*. *Molecular Microbiology,* 5**,** 2255-2260.

COFFEY, T. J., ENRIGHT, M. C., DANIELS, M., MORONA, J. K., MORONA, R., HRYNIEWICZ, W., PATON, J. C. & SPRATT, B. G. (1998a) Recombinational exchanges at the capsular polysaccharide biosynthetic locus lead to frequent serotype changes among natural isolates of *Streptococcus pneumoniae*. *Molecular Microbiology,* 27**,** 73-83.

COFFEY, T. J., ENRIGHT, M. C., DANIELS, M., WILKINSON, P., BERRON, S., FENOLL, A. & SPRATT, B. G. (1998b) Serotype 19A variants of the Spanish serotype 23F multiresistant clone of *Streptococcus pneumoniae*. *Microbial Drug Resistance,* 4**,** 51-5.

COLE, R. & MOORE, H. F. (1917) The Production of Antipneumococcic Serum. *The Journal of Experimental Medicine,* 26**,** 537-561.

COLLINS, K. R. (1905) The Application of the Reaction of Agglutination to the Pneumococcus. *The Journal of Experimental Medicine,* 7**,** 420-429.

COLMAN, G. & HALLAS, G. (1983) Systemic disease caused by pneumococci. *Journal of Infection,* 7**,** 248-55.

CONVERSE, G. M. & DILLON, H. C. (1977) Epidemiological Studies of *Streptococcus pneumoniae* in Infants: Methods of Isolating Pneumococci. *Journal of Clinical Microbiology,* 5**,** 293-296.

CONWAY, T. & SCHOOLNIK, G. K. (2003) Microarray expression profiling: capturing a genome-wide portrait of the transcriptome. *Molecular Microbiology,* 47**,** 879-889.

CORAL, M. C. V., FONSECA, N., CASTANEDA, E., DI FABIO, J. L., HOLLINGSHEAD, S. & BRILES, D. (2001) Pneumococcal Surface Protein A of Invasive *Streptococcus pneumoniae* Isolates from Colombian Children. *Emerging Infectious Diseases,* 7**,** 832-836.

COURVALIN, P. & CARLIER, C. (1987) Tn*1545*: a conjugative shuttle transposon. *Molecular and General Genetics,* 206**,** 259-264.

COWAN, J., HARRINGTON, A. W., CRUIKSHANK, R., CUTHBERTSON, D. P. & FLEMING, J. (1932) Treatment of Lobar Pneumonia by Felton's Serum. *Lancet,* ii**,** 8-11.

CRAIG, A. S., ERWIN, P. C., SCHAFFNER, W., ELLIOTT, J. A., MOORE, W. L., USSERY, X. T., PATTERSON, L., DAKE, A. D., HANNAH, S. G. & BUTLER, J. C. (1999) Carriage of multidrug-resistant *Streptococcus pneumoniae* and impact of chemoprophylaxis during an outbreak of meningitis at a day care center. *Clinical Infectious Diseases,* 29**,** 1257-1264.

CRUM, N. F., BARROZO, C. P., CHAPMAN, F. A., RYAN, M. A. K. & RUSSELL, K. L. (2004) An outbreak of conjunctivitis due to a novel unencapsulated *Streptococcus pneumoniae* among military trainees. *Clinical Infectious Diseases,* 39**,** 1148-1154.

CRUM, N. F., WALLACE, M. R., LAMB, C. R., CONLIN, A. M. S., AMUNDSON, D. E., OLSON, P. E., RYAN, M. A. K., ROBINSON, T. J., GRAY, G. C. & EARHART, K. C. (2003) Halting a pneumococcal pneumonia outbreak among United States Marine Corps trainees. *American Journal Of Preventive Medicine,* 25**,** 107-111.

CULLOTTA, A. R., KALTER, H. D., DELGADO, J., GILMAN, R. H., FACKLAM, R. R., VELAPATINO, B., CORONEL, J., CABRERA, L. & URBINA, M. (2002) Antimicrobial Susceptibilities and Serotype Distribution of *Streptococcus pneumoniae* Isolates from a Low Socioeconomic Area in Lima, Peru. *Clinical and Diagnostic Laboratory Immunology,* 9**,** 1328-1331.

DAGAN, R., GRADSTEIN, S., BELMAKER, I., PORAT, N., SITON, Y., WEBER, G., JANCO, J. & YAGUPSKY, P. (2000) An outbreak of *Streptococcus pneumoniae* serotype 1 in a closed community in southern Israel. *Clinical Infectious Diseases,* 30**,** 319-321.

DAGKESSAMANSKAIA, A., MOSCOSO, M., HENARD, V., GUIRAL, S., OVERWEG, K., REUTER, M., MARTIN, B., WELLS, J. & CLAVERYS, J. P. (2004) Interconnection of competence, stress and CiaR regulons in Streptococcus pneumoniae: competence triggers stationary phase autolysis of ciaR mutant cells. *Molecular Microbiology,* 51**,** 1071-1086.

DAVIDSON, L. S. P. (1925) Proceedings of the Edinburgh Medico-Chirurgical Society. *Edinburgh Medical Journal*, 67.

DAVIES, A. J. & LOCKLEY, M. R. (1987) A prospective survey of hospital cross-infection with *Streptococcus pneumoniae*. *Journal of Hospital Infection,* 9**,** 162-168.

DAVIS, L. E. & GREENLEE, J. E. (2003) Pneumococcal meningitis: antibiotics essential but insufficient. *Brain,* 126**,** 1013-1014.

DAWSON, S., PALLETT, A., DAVIDSON, A. & TUCK, A. (1992) Outbreak of multiresistant pneumococci. *Journal of Hospital Infection,* 22**,** 328-9.

DE BENEDICTIS, F. M., DE GIORGI, G., NICCOLI, A., TROIANI, S., RIZZO, F. & LEMMI, A. (2000) Treatment of Complicated Pleural Effusion With Intracavitary Urokinase in Children *Pediatric Pulmonology,* 29**,** 438-442.

DE GALAN, B. E., VAN TILBURG, P. M. B., SLUIJTER, M., MOL, S. J. M., DE GROOT, R., HERMANS, P. W. M. & JANSZ, A. R. (1999) Hospital-related outbreak of infection with multidrug-resistant *Streptococcus pneumoniae* in the Netherlands. *Journal of Hospital Infection,* 42**,** 185-192.

DE PAOLIS, F., BEGHETTO, E., SPADONI, A., MONTAGNANI, F., FELICI, F., OGGIONI, M. R. & GARGANO, N. (2007) Identification of a human immunodominant B-cell epitope within the immunoglobulin A1 protease of *Streptococcus pneumoniae*. *BMC Microbiology,* 7**,** 113.

DE RISI, J. L., IYER, V. R. & BROWN, P. O. (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science,* 278**,** 680-686.

DE SAIZIEU, A., CERTA, U., WARRINGTON, J., GRAY, C., KECK, W. & MOUS, J. (1998) Bacterial transcript imaging by hybridization of total RNA to oligonucleotide arrays. *Nature Biotechnology,* 16**,** 45-48.

DE SAIZIEU, A., GARDES, C., FLINT, N., WAGNER, C., KAMBER, M., MITCHELL, T. J., KECK, W., AMREIN, K. E. & LANGE, R. (2000) Microarray-Based Identification of a Novel *Streptococcus pneumoniae* Regulon Controlled by an Autoinduced Peptide. *Journal of Bacteriology,* 182**,** 4696-4703.

DELCHER, A. L., HARMON, D., KASIF, S., WHITE, O. & SALZBERG, S. L. (1999) Improved microbial gene identification with GLIMMER. *Nucleic Acids Research,* 27**,** 4636-4641.

DEMARIA, A., BROWNE, K., BERK, S. L., SHERWOOD, E. J. & MCCABE, W. R. (1980) An Outbreak of Type 1 Pneumococcal Pneumonia in a Men's Shelter. *Journal of the American Medical Association,* 244**,** 1446-1449.

DHARMADI, Y. & GONZALEZ, R. (2004) DNA microarrays: experimental issues, data analysis, and application to bacterial systems. *Biotechnology Progress,* 20**,** 1309-1324.

DI FABIO, J. L., CASTANEDA, E., AGUDELO, C. I., DE LA HOZ, F., HORTAL, M., CAMOU, T., ECHANIZ-AVILES, G., BARAJAS, M. N. C., HEITMANN, I., HORMAZABAL, J. C., BRANDILEONE, M. C. C., VIEIRA, V. S. D.,

REGUEIRA, M., RUVINSKI, R., CORSO, A., LOVGREN, M., TALBOT, J. A. & DE QUADROS, C. (2001) Evolution of *Streptococcus pneumoniae* serotypes and penicillin susceptibility in Latin America, Sireva-Vigia Group, 1993 to 1999. *Pediatric Infectious Disease Journal,* 20**,** 959-967.

DI FABIO, J. L., HOMMA, A. & DE QUADROS, C. (1997) Pan American Health Organisation Epidemiological Surveillance Network for *Streptococcus pneumoniae*. *Microbial Drug Resistance,* 3**,** 131-133.

DIAS, R. D. & CANICA, M. (2004) Emergence of invasive erythromycin-resistant *Streptococcus pneumoniae* strains in Portugal: contribution and phylogenetic relatedness of serotype 14. *Journal of Antimicrobial Chemotherapy,* 54**,** 1035-1039.

DIETRICH, G., KURZ, S., HUBNER, C., AEPINUS, C., THEISS, S., GUCKENBERGER, M., PANZER, U., WEBER, J. R. & FROSCH, M. (2003) Transcriptome Analysis of *Neisseria meningitidis* during Infection. *Journal of Bacteriology,* 185**,** 155-164.

DIGGLE, M. A. & CLARKE, S. C. (2005) Truncated xpt gene present in invasive Streptococcus pneumoniae may have implications for MLST schemes. *Journal of Medical Microbiology,* 54**,** 909-912.

DIGGLE, M. A. & EDWARDS, G. S. F. (2006) PO3.62 The Epidemiology of *Streptococcus pneumoniae* in Scotland between 2003 and 2005. *5th International Symposium on Pneumococci and Pneumococcal Diseases.* Alice Springs, Australia.

DOCHEZ, A. R. & AVERY, O. T. (1915) The Occurrence of Carriers of Disease-Producing Types of Pneumococcus. *The Journal of Experimental Medicine,* 22**,** 105-113.

DOCHEZ, A. R. & GILLESPIE, L. J. (1913) A Biologic Classification of Pneumococci by Means of Immunity Reactions. *The Journal of the American Medical Association,* 61**,** 727-732.

DOHERTY, N., TRZCINSKI, K., PICKERILL, P., ZAWADZKI, P. & DOWSON, C. G. (2000) Genetic diversity of the tet (M) gene in tetracycline-resistant clonal lineages of *Streptococcus pneumoniae*. *Antimicrobial Agents and Chemotherapy,* 44**,** 2979-2984.

DOPAZO, J., MENDOZA, A., HERRERO, J., CALDARA, F., HUMBERT, Y., FRIEDLI, L., GUERRIER, M., GRAND-SCHENK, E., GANDIN, C., DE FRANCESCO, M., POLISSI, A., BUELL, G., FEGER, G., GARCIA, E., PEITSCH, M. & GARCIA-BUSTOS, J. F. (2001) Annotated Draft Genomic Sequence from a *Streptococcus pneumoniae* Type 19F Clinical Isolate. *Microbial Drug Resistance,* 7**,** 99-125.

DORRELL, N., HINCHCLIFFE, S. J. & WREN, B. W. (2005) Comparative phylogenomics of pathogenic bacteria by microarray analysis. *Current Opinion in Microbiology,* 8**,** 620-626.

DORRELL, N., MANGAN, J. A., LAING, K. G., HINDS, J., LINTON, D., AL-GHUSEIN, H., BARRELL, B. G., PARKHILL, J., STOKER, N. G., KARLYSHEV, A. V., BUTCHER, P. D. & WREN, B. W. (2001) Whole genome comparison of *Campylobacter jejuni* human isolates using a low cost microarray reveals extensive genetic diversity. *Genome Research,* 11**,** 1706-1715.

DOWSON, C. G. (2004) *What is a Pneumococcus? In The Pneumococcus,* Washington, DC, ASM Press.

DOWSON, C. G., COFFEY, T. J., KELL, C. & WHILEY, R. A. (1993) Evolution of penicillin resistance in *Streptococcus pneumoniae*; the role of *Streptococcus mitis* in the formation of a low affinity PBP 2B in *Streptococcus pneumoniae*. *Molecular Microbiology,* 9**,** 635-43.

DOWSON, C. G., HUTCHISON, A., BRANNIGAN, J. A., GEORGE, R. C., HANSMAN, D., LINARES, J., TOMASZ, A., SMITH, J. M. & SPRATT, B. G. (1989) Horizontal transfer of penicillin-binding protein genes in penicillin-resistant clinical isolates of *Streptococcus pneumoniae*. *Proceedings of the National Academy of Sciences of the United States of America,* 86**,** 8842-6.

DURAND, M. L., CALDERWOOD, S. B., WEBER, D. J., MILLER, S. I., SOUTHWICK, F. S., CAVINESS, V. S. & SWARTZ, M. N. (1993) Acute bacterial meningitis in adults. A review of 493 episodes. *New England Journal of Medicine,* 328**,** 21-28.

EASTHAM, K. M., FREEMAN, R., KEARNS, A., ELTRINGHAM, G., LEEMING, J. P., CLARK, J. & SPENCER, D. A. (2002) S68 Pneumococcal serotypes in children with culture negative empyema. *Thorax,* 57**,** iii21-iii22.

EASTHAM, K. M., FREEMAN, R., KEARNS, A. M., ELTRINGHAM, G., CLARK, J., LEEMING, J. & SPENCER, D. A. (2004) Clinical features, aetiology and outcome of empyema in children in the north east of England. *Thorax,* 59**,** 522-525.

ECHANIZ-AVILES, G., CARNALLA-BARAJAS, M. N., SOTO-NOGUERON, A., VELAZQUEZ-MEZA, M. E., AZAOLA-ESPINOZA, A., SILVA-SANCHEZ, J., CONDE-GONZALEZ, J. C., URIBE-SALAS, F., GABASTOU, J. M. & DI FABIO, J. L. (2008) P2-052 Serotype distribution and antimicrobial susceptibility of invasive pneumococcal isolates from Mexican children. Sireva Network 1996-2006. *6th International Symposium on Pneumococci and Pneumococcal Diseases.* Reykjavik, Iceland.

ECHANIZ-AVILES, G., CARNALLA-BARAJAS, M. N., VELAZQUEZ-MEZA, M. E., SOTO-NOGUERON, A., ESPINOSA DE LOS MONTEROS, L. E. & SOLORZANO-SANTOS, F. (1995) Capsular types of *Streptococcus pneumoniae* causing disease in children from Mexico City. *Pediatric Infectious Disease Journal,* 14**,** 907-9.

ECHANIZ-AVILES, G., VELAZQUEZ-MEZA, M. E., CARNALLA-BARAJAS, M. N., SOTO-NOGUERON, A., DI FABIO, J. L., SOLORZANO-SANTOS, F., JIMENEZ-TAPIA, Y. & TOMASZ, A. (1998) Predominance of the multiresistant 23F international clone of *Streptococcus pneumoniae* among isolates from Mexico. *Microbial Drug Resistance,* 4**,** 241-246.

ECHANIZ-AVILES, I. G. (2001) Meeting the Challenge: Prevention of Pneumococcal Disease with Conjugate Vaccines. *Salud Publica de Mexico,* 43**,** 352-367.

EHRLICH, G. D., HILLER, N. L. & HU, F. Z. (2008) What makes pathogens pathogenic. *Genome Biology,* 9**,** 225.

ELTRINGHAM, G., KEARNS, A., FREEMAN, R., CLARK, J., SPENCER, D., EASTHAM, K., HARWOOD, J. & LEEMING, J. (2003) Culture-negative

childhood empyema is usually due to penicillin-sensitive *Streptococcus pneumoniae* capsular serotype 1. *Journal of Clinical Microbiology,* 41**,** 521-522.

EMBRY, A., HINOJOSA, E. & ORIHUELA, C. J. (2007) Regions of Diversity 8, 9 and 13 contribute to *Streptococcus pneumoniae* virulence. *BMC Microbiology,* 7**,** 80.

ENRIGHT, M. C. & SPRATT, B. G. (1998) A multilocus sequence typing scheme for *Streptococcus pneumoniae*: identification of clones associated with serious invasive disease. *Microbiology,* 144**,** 3049-3060.

ENRIGHT, M. C. & SPRATT, B. G. (1999a) Extensive variation in the *ddl* gene of penicillin-resistant *Streptococcus pneumoniae* results from a hitchhiking effect driven by the penicillin binding protein 2b gene. *Molecular Biology And Evolution,* 16**,** 1687-1695.

ENRIGHT, M. C. & SPRATT, B. G. (1999b) Multilocus sequence typing. *Trends in Microbiology,* 7**,** 482-487.

ERTUGRUL, N., RODRIGUEZ-BARRADAS, M. C., MUSHER, D. M., RYAN, M. A., AGIN, C. S., MURPHY, S. J., SHAYEGANI, M. & WATSON, D. A. (1997) BOX-polymerase chain reaction-based DNA analysis of nonserotypeable *Streptococcus pneumoniae* implicated in outbreaks of conjunctivitis. *Journal Of Infectious Diseases,* 176**,** 1401-5.

ESKOLA, J., KILPI, T. & PALMU, A. (2001) Efficacy of a pneumococcal conjugate vaccine against acute otitis media. *New England Journal of Medicine,* 344**,** 403-409.

EVERS, S., PADOVA, K. D., MEYER, M., LANGEN, H., FOUNTOULAKIS, M., KECK, W. & GRAY, C. P. (2001) Mechanism related changes in the gene transcription and protein synthesis patterns of *Haemophilus influenzae* after treatment with transcriptional and translational inhibitors. *Proteomics,* 1**,** 522-544.

FADEN, H., DUFFY, L., WASIELEWSKI, R., WOLF, J., KRYSTOFIK, D. & TUNG, Y. (1997) Relationship between nasopharyngeal colonization and the development of otitis media in children. *Journal of Infectious Diseases,* 175**,** 1440-1445.

FEIL, E. J., ENRIGHT, M. C. & SPRATT, B. G. (2000a) Estimating the relative contributions of mutation and recombination to clonal diversification: a comparison between *Neisseria meningitidis* and *Streptococcus pneumoniae*. *Research in Microbiology,* 151**,** 465-9.

FEIL, E. J., LI, B. C., AANENSEN, D. M., HANAGE, W. P. & SPRATT, B. G. (2004) eBURST: Inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. *Journal of Bacteriology,* 186**,** 1518-1530.

FEIL, E. J., SMITH, J. M., ENRIGHT, M. C. & SPRATT, B. G. (2000b) Estimating recombinational parameters in *Streptococcus pneumoniae* from multilocus sequence typing data. *Genetics,* 154**,** 1439-50.

FEIL, E. J. & SPRATT, B. G. (2001) Recombination and the Population Structures of Bacterial Pathogens. *Annual Review of Microbiology,* 55**,** 561-90.

FEINGOLD, E. K. (2003) The outbreak of conjunctivitis at Dartmouth. *New England Journal of Medicine,* 348**,** 2577-8.

FEOLA, G. P., SHAW, C. A. & COBURN, L. (2003) Management of Complicated Parapneumonic Effusions in Children. *Techniques in Vascular and Interventional Radiology,* 6**,** 197-204.

FINCHER, E. F. (1946) Craniotomy and total dissection as a method in the treatment of abscess of the brain. *Annals of Surgery,* 123**,** 789-807.

FINKELSTEIN, D., EWING, R., GOLLUB, J., STERKY, F., CHERRY, J. M. & SOMERVILLE, S. (2002) Microarray data quality analysis: lessons from the AFGC project. *Plant Molecular Biology,* 48.

FINLAND, M. & BARNES, M. W. (1970) Changing Etiology of Bacterial Endocarditis in the Antibacterial Era. *Annals of Internal Medicine,* 72**,** 341-348.

FINLAND, M. & BARNES, M. W. (1977a) Acute Bacterial Meningitis at Boston City Hospital during 12 Selected Years, 1935-1972. *Journal of Infectious Diseases,* 136**,** 400-415.

FINLAND, M. & BARNES, M. W. (1978) Changing Ecology of Acute Bacterial Empyema: Occurrence and Mortality at Boston City Hospital during 12 Selected Years from 1935 to 1972. *Journal of Infectious Diseases,* 137**,** 274-291

FINLAND, M. & SUTLIFF, W. D. (1934) Infections with Pneumococcus Type III and Type VIII. *Archives of Internal Medicine,* 53**,** 481-507.

FINLAND, M. F. & BARNES, M. W. (1977b) Changes in Occurrence of Capsular Serotypes of *Streptococcus pneumoniae* at Boston City Hospital During Selected Years Between 1935 and 1974. *Journal of Clinical Microbiology,* 5**,** 154-166.

FIORE, A. E., IVERSON, C., MESSMER, T., ERDMAN, D., LETT, S. M., TALKINGTON, D. F., ANDERSON, L. J., FIELDS, B., CARLONE, G. M., BREIMAN, R. F. & CETRON, M. S. (1998) Outbreak of pneumonia in a long-term care facility: Antecedent human parainfluenza virus 1 infection may predispose to bacterial pneumonia. *Journal of the American Geriatrics Society,* 46**,** 1112-1117.

FLANNERY, B., HEFFERNAN, R. T., HARRISON, L. H., RAY, S. M., REINGOLD, A. L., HADLER, J., SCHAFFNER, W., LYNFIELD, R., THOMAS, A. R., LI, J., CAMPSMITH, M., WHITNEY, C. G. & SCHUCHAT, A. (2006) Changes in Invasive Pneumococcal Disease among HIV-infected Adults Living in the Era of Childhood Pneumococcal Immunization. *Annals of Internal Medicine,* 144**,** 1-9.

FLANNERY, B., SCHRAG, S., BENNETT, N. M., LYNFIELD, R., HARRISON, L. H., REINGOLD, A., CIESLAK, P. R., HADLER, J., FARLEY, M. M., FACKLAM, R. R., ZELL, E. R. & WHITNEY, C. G. (2004) Impact of Childhood Vaccination on Racial Disparities in Invasive *Streptococcus pneumoniae* Infections. *Journal of the American Medical Association,* 291**,** 2197-2203.

FLEISCHMANN, R. D., ADAMS, M. D., WHITE, O., CLAYTON, R. A., KIRKNESS, E. F., KERLAVAGE, A. R., BULT, C. J., TOMB, J. F., DOUGHERTY, B. A., MERRICK, J. M., MCKENNEY, K., SUTTON, G., FITZHUGH, W., FIELDS, C.,

GOCAYNE, J. D., SCOTT, J., SHIRLEY, R., LIU, L. I., GLODEK, A., KELLEY, J. M., WEIDMAN, J. F., PHILLIPS, C. A., SPRIGGS, T., HEDBLOM, E., COTTON, M. D., UTTERBACK, T. R., HANNA, M. C., NGUYEN, D. T., SAUDEK, D. M., BRANDON, R. C., FINE, L. D., FRITCHMAN, J. L., FUHRMANN, J. L., GEOGHAGEN, N. S. M., GNEHM, C. L., MCDONALD, L. A., SMALL, K. V., FRASER, C. M., SMITH, H. O. & VENTER, J. C. (1995) Whole-Genome Random Sequencing And Assembly Of *Haemophilus influenzae* Rd. *Science,* 269**,** 496-512.

FLETCHER, M., LEEMING, J., CARTWRIGHT, K. & FINN, A. (2006) Childhood empyema: Limited potential impact of 7-valent pneumococcal conjugate vaccine. *Pediatric Infectious Disease Journal,* 25**,** 559-560.

FLIPPIN, H. F., MATTEUCCI, W. V., SCHIMMEL, N. H. & BOGER, W. P. (1951) Aureomycin, Chloramphenicol, and Penicillin in the Treatment of Bacterial Pneumonia. *Journal of the American Medical Association,* 147**,** 918.

FOSTER, W. R. & HUBER, R. M. (2002) Current themes in microarray experimental design and analysis. *Drug Discovery Today,* 7**,** 290-292.

FOURNIER, P. E., DRANCOURT, M. & RAOULT, D. (2007) Bacterial genome sequencing and its use in infectious diseases. *Lancet Infectious Diseases,* 7**,** 711-723.

FRAENKEL, A. (1886) Weitere Beitrage zur Lehre von den Mikrococcen der genuinen fibrinosen Pneumonie. *Zeitschrift fur Klinische Medizin,* 11**,** 437-458.

FRASER-LIGGETT, C. M. (2005) Insights on biology and evolution from microbial genome sequencing. *Genome Research,* 15**,** 1603-1610.

FREEMAN, A. F., BEN-AMI, T. & SHULMAN, S. T. (2004) *Streptococcus pneumoniae* empyema necessitatis. *Pediatric Infectious Disease Journal,* 23**,** 177-179.

FREIBERG, C., BROTZ-OESTERHELT, H. & LABISCHINSKI, H. (2004) The impact of transcriptome and proteome analyses on antibiotic drug discovery. *Current Opinion in Microbiology,* 7**,** 451-459.

FREIBERG, C., FISCHER, H. P. & BRUNNER, N. A. (2005) Discovering the Mechanism of Action of Novel Antibacterial Agents through Transcriptional Profiling of Conditional Mutants. *Antimicrobial Agents & Chemotherapy,* 49**,** 749-759.

FRIEBERG, C. & BRUNNER, N. A. (2002) Genome-wide mRNA profiling: impact on compound evaluation and target identification in anti-bacterial research. *Drug Discovery Today: Targets,* 1**,** 20-29.

FRY, A. M., UDEAGU, C. C. N., SORIANO-GABARRO, M., FRIDKIN, S., MUSINSKI, D., LACLAIRE, L., ELLIOTT, J., COOK, D. J. P., KORNBLUM, J., LAYTON, M. & WHITNEY, C. G. (2005) Persistence of fluoroquinolone-resistant, multidrug-resistant *Streptococcus pneumoniae* in a long-term-care facility: Efforts to reduce intrafacility transmission. *Infection Control And Hospital Epidemiology,* 26**,** 239-247.

FUCHS, S. C., FISHER, G. B., BLACK, R. E. & LANATA, C. (2005) The burden of pneumonia in children in Latin America. *Paediatric Respiratory Reviews,* 6**,** 83-87.

FUKUDA, Y., YANAGIHARA, K., HIGASHIYAMA, Y., MIYAZAKI, Y., HIRAKATA, Y., MUKAE, H., TOMONO, K., MIZUTA, Y., TSUKAMOTO, K. & KOHNO, S. (2006) Effects of macrolides on pneumolysin of macrolide-resistant *Streptococcus pneumoniae*. *European Respiratory Journal,* 27**,** 1020-1025.

GALPERIN, M. Y. & KOONIN, E. V. (1999) Searching for drug targets in microbial genomes. *Current opinion in Biotechnology,* 10**,** 571-578.

GAMBOA, L., CAMOU, T., HORTAL, M. & CASTANEDA, E. (2002) Dissemination of *Streptococcus pneumoniae* clone Colombia^5-19 in Latin America. *Journal of Clinical Microbiology,* 40**,** 3942-3950.

GARCIA-RODRIGUEZ, J. A. & MARTINEZ, M. J. F. (2002) Dynamics of nasopharyngeal colonization by potential respiratory pathogens. *Journal of Antimicrobial Chemotherapy,* 50**,** 59-73.

GARCIA, E. & LOPEZ, R. (1995) *Streptococcus pneumoniae* type 3 encodes a protein highly similar to the human glutamate decarboxylase (GAD65). *FEMS Microbiology Letters,* 133**,** 113-118.

GARCIA, S., LEVINE, O. S., CHERIAN, T., GABASTOU, J. M., ANDRUS, J. & MEMBERS, W. G. (2006) Pneumococcal disease and vaccination in the Americas: an agenda for accelerated vaccine introduction. *Pan American Journal of Public Health,* 19**,** 340-8.

GASC, A. M., KAUC, L., BARRAILLE, P., SICARD, M. & GOODGAL, S. H. (1991) Gene Localization, Size, and Physical Map of the Chromosome of *Streptococcus pneumoniae*. *Journal of Bacteriology,* 173**,** 7361-7367.

GATES, R. L., CANIANO, D. A., HAYES, J. & ARCA, M. J. (2004) Does VATS provide optimal treatment of empyema in children? A systematic review. *Journal of Pediatric Surgery,* 39**,** 381-386.

GAY, K. & STEPHENS, D. S. (2001) Structure and dissemination of a chromosomal insertion element encoding macrolide efflux in *Streptococcus pneumoniae*. *Journal of Infectious Diseases,* 284**,** 56-65.

GEMMELL, C. G. & LORIAN, V. (1996) *Effects of Low Concentrations of Antibiotics on Bacterial Ultrastructure, Virulence, and Susceptibility to Immunodefenses: Clinical Significance. In, "Antibiotics in Laboratory Medicine.",* Pennsylvania, USA, Williams and Wilkins.

GEORGE, R. C., GUNGABISSOON, U., SLACK, M. P. E., MARTIN, S., BROUGHTON, K., WAIGHT, P. & MILLER, E. (2006) Invasive pneumococcal disease in England and Wales 1996-2004. *5th International Symposium on Pneumococci and Pneumococcal Diseases.* Alice Springs, Australia.

GHERARDI, G., DEL GROSSO, M., D'ABUSCO, A. S., D'AMBROSIO, F., DICUONZO, G. & PANTOSTI, A. (2003) Phenotypic and genotypic characterization of two penicillin-susceptible serotype 6B *Streptococcus pneumoniae* clones circulating in Italy. *Journal of Clinical Microbiology,* 41**,** 2855-2861.

GHERARDI, G., WHITNEY, C. G., FACKLAM, R. R. & BEALL, B. (2000) Major related sets of antibiotic-resistant pneumococci in the United States as determined by pulsed-field gel electrophoresis and pbp1a-pbp2b-pbp2x-dhf restriction profiles. *Journal of Infectious Diseases,* 181**,** 216-229.

GIACHETTO, G., PIREZ, M. C., NANNI, L., MARTINEZ, A., MONTANO, A., ALGORTA, G., KAPLAN, S. L. & FERRARI, A. M. (2004) Ampicillin and penicillin concentration in serum and pleural fluid of hospitalized children with community-acquired pneumonia. *Pediatric Infectious Disease Journal,* 23**,** 625-629.

GIANFALDONI, C., CENSINI, S., HILLERINGMANN, M., MOSCHIONI, M., FACCIOTTI, C., PANSEGRAU, W., MASIGNANI, V., COVACCI, A., RAPPUOLI, R., BAROCCHI, M. A. & RUGGIERO, P. (2007) *Streptococcus pneumoniae* Pilus Subunits Protect Mice against Lethal Challenge. *Infection and Immunity,* 75**,** 1059-1062.

GILBERT, D. N., MOELLERING, R. C., ELIOPOULOS, G. M. & SANDE, M. A. (2008) *The Sanford Guide to Antimicrobial Therapy*.

GILLESPIE, S. H., MCHUGH, T. D., HUGHES, J. E., DICKENS, A., KYI, M. S. & KELSEY, M. (1997) An outbreak of penicillin resistant *Streptococcus pneumoniae* investigated by a polymerase chain reaction based genotyping method. *Journal of Clinical Pathology,* 50**,** 847-851.

GILMAN, B. B. & ANDERSON, G. W. (1938) A Community Outbreak of Type I Pneumococcus Infection. *American Journal of Hygiene,* 28**,** 345-358.

GINDREAU, E., LOPEZ, R. & GARCIA, P. (2000) MM1, a Temperate Bacteriophage of the Type 23F Spanish/USA Multiresistant Epidemic Clone of *Streptococcus pneumoniae:* Structural Analysis of the Site-Specific Integration System. *Journal of Virology,* 74**,** 7803-7813.

GLEICH, S., MORAD, Y., ECHAGUE, R., MILLER, J. R., KORNBLUM, J., SAMPSON, J. S. & BUTLER, J. C. (2000) *Streptococcus pneumoniae* serotype 4 outbreak in a home for the aged: Report and review of recent outbreaks. *Infection Control And Hospital Epidemiology,* 21**,** 711-717.

GLYNN, E. E. & DIGBY, L. (1923) Bacteriological and Clinical Observations on Pneumonia and Empyemata, with Special Reference to the Pneumococcus and to Serum Treatment. *Medical Research Council, Special Report Series,* 79**,** 1-165.

GMUENDER, H., KURATLI, K., DI PADOVA, K., GRAY, C. P., KECK, W. & EVERS, S. (2001) Gene Expression Changes Triggered by Exposure of *Haemophilus influenzae* to Novobiocin or Ciprofloxacin: Combined Transcription and Translation Analysis. *Genome Research,* 11**,** 28-42.

GOH, E. B., YIM, G., TSUI, W. H. W., MCCLURE, J. E., SURETTE, M. G. & DAVIES, J. (2002) Transcriptional modulation of bacterial gene expression by subinhibitory concentrations of antibiotics. *Proceedings of the National Academy of Sciences,* 99**,** 17025-17030.

GONZALEZ, B. E., HULTEN, K. G., KAPLAN, S. L. & MASON, E. O. (2004) Clonality of *Streptococcus pneumoniae* serotype 1 isolates from pediatric patients in the United States. *Journal of Clinical Microbiology,* 42**,** 2810-2812.

GORDON, J. E. (1921) The Relationship of the Pneumococcus to Acute Infections of the Upper Respiratory Tract in Man. *Journal of Infectious Diseases,* 29**,** 437-461.

GOULD, F. K., MAGEE, J. G. & INGHAM, H. R. (1987) A hospital outbreak of antibiotic-resistant *Streptococcus pneumoniae*. *Journal of Infection,* 15**,** 77-9.

GRABOWSKA, K., HOGBERG, L., PENTTINEN, P., SVENSSON, A. & EKDAHL, K. (2006) Occurrence of invasive pneumococcal disease and number of excess cases due to influenza. *BMC Infectious Diseases,* 6**,** 58.

GRAM, C. (1884) Ueber die isolierte Färbung der Schizomyceten in Schnitt- und Trokenpräparaten. *Fortschritte der Medizin,* 2**,** 185-189.

GRANSDEN, W. R., EYKYN, S. J. & PHILLIPS, I. (1985) Pneumococcal bacteraemia: 325 episodes diagnosed at St Thomas's Hospital. *British Medical Journal,* 290**,** 505-508.

GRANT, M. H. (1922) Pneumonias in Glasgow and their Pneumococcal Types. *Glasgow Medical Journal,* 98**,** 193-195.

GRATTEN, M., GRATTEN, H., POLI, A., CARRAD, E., RAYMER, M. & KOBI, G. (1986) Colonisation of *Haemophilus influenzae* and *Streptococcus pneumoniae* in the Upper Respiratory Tract of Neonates in Papua New Guinea: Primary Acquisition, Duration of Carriage, and Relationship to Carriage in Mothers. *Biology of the Neonate,* 50**,** 114-120.

GRATTEN, M., MONTGOMERY, J., GEREGA, G., GRATTEN, H., SIWI, H., POLI, A. & KOKI, G. (1989) Multiple colonization of the upper respiratory tract of Papua New Guinea children with *Haemophilus influenzae* and *Streptococcus pneumoniae*. *Southeast Asian Journal of Tropical Medicine and Public Health,* 20**,** 501-509.

GRATTEN, M., MOREY, F., DIXON, J., MANNING, K., TORZILLO, P., MATTERS, R., ERLICH, J., HANNA, J., ASCHE, V. & RILEY, I. (1993) An Outbreak Of Serotype 1 *Streptococcus pneumoniae* Infection In Central Australia. *Medical Journal of Australia,* 158**,** 340-342.

GRAY, B. M., CONVERSE, G. M. & DILLON, H. C. (1980) Epidemiologic studies of *Streptococcus pneumoniae* in infants: acquisition, carriage, and infection during the first 24 months of life. *Journal of Infectious Diseases,* 142**,** 923-933.

GREENWOOD, B. (1999) The epidemiology of pneumococcal infection in children in the developing world. *Philosophical Transactions of the Royal Society of London B,* 354**,** 777-785.

GREENWOOD, B. (2008) A global action plan for the prevention and control of pneumonia. *Bulletin of the World Health Organization,* 86**,** 321-416.

GRIGORIADIS, E. & GOLD, W. L. (1997) Pyogenic brain abscess caused by *Streptococcus pneumoniae*: Case report and review. *Clinical Infectious Diseases,* 25**,** 1108-1112.

GRIMM, V., EZAKI, S., SUSA, M., KNABBE, C., SCHMID, R. D. & BACHMANN, T. T. (2004) Use of DNA Microarrays for Rapid Genotyping of TEM Beta-Lactamases That Confer Resistance. *Journal of Clinical Microbiology,* 42**,** 3766-3774.

GRIST, N. R., LANDSMAN J.B., ANDERSON, T., ROWAN, G. & SMITH, A. (1952) Studies in the Aetiology of Pneumonia in Glasgow 1950-51. *The Lancet,* i**,** 640-646.

GUO, Z., GUILFOYLE, R. A., THEIL, A. J., WANG, R. & SMITH, L. M. (1994) Direct fluorescence analysis of genetic polymorphisms by hybridization with oligonucleotide arrays on glass supports. *Nucleic Acids Research,* 22**,** 5456-5465.

GUPTA, A., KHAW, F. M., GORTON, R., STOKLE, L. & SPENCER, D. (2007) An outbreak of *Streptococcus pneumoniae* serotype 1 pneumonia in a UK primary school. *European Respiratory Society Annual Congress 2007.* Stockholm, Sweden.

HAAS, W., KAUSHAL, D., SUBLETT, J., OBERT, C. & TUOMANEN, E. I. (2005) Vancomycin Stress Response in a Sensitive and a Tolerant Strain of *Streptococcus pneumoniae*. *Journal of Bacteriology,* 187**,** 8205-8210.

HAKENBECK, R., BALMELLE, N., WEBER, B., GARDES, C., KECK, W. & DE SAIZIEU, A. (2001) Mosaic genes and mosaic chromosomes: Intra- and interspecies genomic variation of *Streptococcus pneumoniae*. *Infection and Immunity,* 69**,** 2477-2486.

HAKENBECK, R., GREBE, T., ZAHNER, D. & STOCK, J. B. (1999) Beta-lactam resistance in *Streptococcus pneumoniae*: penicillin-binding proteins and non-penicillin-binding proteins. *Molecular Microbiology,* 33**,** 673-8.

HALL-STOODLEY, L., HU, F. Z., GIESEKE, A., NISTICO, L., NGUYEN, D., HAYES, J., FORBES, M., GREENBERG, D. P., DICE, B., BURROWS, A., WACKYM, P. A., STOODLEY, P., POST, J. C., EHRLICH, G. D. & KERSCHNER, J. E. (2006) Direct detection of bacterial Biofilms on the middle-ear mucosa of children with chronic otitis media. *Journal of the American Medical Association,* 296**,** 202-211.

HALL, L. M. C., WHILEY, R. A., DUKE, B., GEORGE, R. C. & EFSTRATIOU, A. (1996) Genetic relatedness within and between Serotypes of *Streptococcus pneumoniae* from the United Kingdom: Analysis of Multilocus Enzyme Electrophoresis, Pulsed-Field Gel Electrophoresis, and Antimicrobial Resistance Patterns. *Journal of Clinical Microbiology,* 34**,** 853-859.

HAMADA, K., MIKASA, K., YUNOU, Y., KURIOKA, T., MAJIMA, T., NARITA, N. & KITA, E. (2000) Adjuvant effect of clarithromycin on chemotherapy for murine lung cancer. *Chemotherapy,* 46**,** 49-61.

HAMM, H. & LIGHT, R. W. (1997) Parapneumonic effusion and empyema. *European Respiratory Journal,* 10**,** 1150-1156.

HAMMERSCHMIDT, S., WOLFF, S., HOCKE, A., ROSSEAU, S., MULLER, E. & ROHDE, M. (2005) Illustration of pneumococcal polysaccharide capsule during adherence and invasion of epithelial cells. *Infection And Immunity,* 73**,** 4653-4667.

HANAGE, W., BISHOP, C., HINRICHSEN, V., HUANG, S., STEVENSON, A., PELTON, S. & FINKELSTEIN, J. (2008) P2-020 Expansion of 19A clones derived through serotype switching in the Massachusetts carriage population leading to increased multidrug resistance. *6th International Symposium on Pneumococci and Pneumococcal Diseases.* Reykjavik, Iceland.

HANAGE, W. P., KAIJALAINEN, T., HERVA, E., SAUKKORIIPI, P. A., SYRJANEN, R. & SPRATT, B. G. (2005) Using Multilocus Sequence Data to Define the Pneumococcus. *Journal of Bacteriology,* 187**,** 6223-6230.

HANDMAN, H. P. & REUMAN, P. D. (1993) The Use of Urokinase for Loculated Thoracic Empyema in Children: A Case Report and Review of the Literature. *Pediatric Infectious Disease Journal,* 12**,** 958-959.

HANQUET, G., KISSLING, E., TARRAGO, D., FENOLL, A., VARON, E., GEORGE, R., HAUSDORFF, W. P., LERNOUT, T. & VERHAEGEN, J. (2008) S01-O4 Dynamic changes of three non-vaccine pneumococcal serotypes in Spain, France, Belgium and England and Wales, 1996-2006. *6th International Symposium on Pneumococci and Pneumococcal Diseases.* Reykjavik, Iceland.

HANSMANN, Y., DOYLE, A., REMY, V., JAULHAC, B., CHRISTMANN, D., LESENS, O. & PERROCHEAU, A. (2006) An outbreak of pneumococcal pneumonia among residents of a retirement home in France during October 2003. *Infection Control And Hospital Epidemiology,* 27**,** 1252-1254.

HARDIE, W., BOKULIC, R., GARCIA, V. F., REISING, S. F. & CHRISTIE, C. D. (1996) Pneumococcal pleural empyemas in children. *Clinical Infectious Diseases,* 22**,** 1057-63.

HARDIE, W. D., ROBERTS, N. E., REISING, S. F. & CHRISTIE, C. D. C. (1998) Complicated parapneumonic effusions in children caused by penicillin-nonsusceptible *Streptococcus pneumoniae*. *Pediatrics,* 101**,** 388-392.

HARDY, D. J., GUAY, D. R. P. & JONES, R. N. (1992) Clarithromycin, a Unique Macrolide A Pharmacokinetic, Microbiological, and Clinical Overview. *Diagnostic Microbiology and Infectious Diseases,* 15**,** 39-53.

HARDY, D. J., HENSEY, D. M., BEYER, J. M., VOJTKO, C., MCDONALD, E. J. & FERNANDES, P. B. (1988) Comparative In Vitro Activities of New 14-,15-, and 16-Membered Macrolides. *Antimicrobial Agents and Chemotherapy,* 32**,** 1710-1719.

HARPER, M. B. (1999) Nasopharyngeal colonization with pathogens causing otitis media: how does this information help us? *Pediatric Infectious Disease Journal,* 18**,** 1120-1124.

HARVEY, R., STROEHER, U. H., OGUNNIYI, A. D., LEACH, A. & PATON, J. C. (2007) A-O6 Identification of potential virulence enhancing genes in *Streptococcus pneumoniae* type 1 isolates. *Europneumo 2007. 8th European Meeting on the Molecular Biology of the Pneumococcus.* Oeiras, Portugal.

HASHSHAM, S. A., WICK, L. M., ROUILLARD, J. M., GULARI, E. & TIEDJE, J. M. (2004) Potential of DNA microarrays for developing parallel detection tools

(PDTs) for microorganisms relevant to biodefense and related research needs. *Biosensors and Bioelectronics,* 20**,** 668-683.

HAUSDORFF, W. P. (2007) The roles of pneumococcal serotypes 1 and 5 in paediatric invasive disease. *Vaccine,* 25**,** 2406-2412.

HAUSDORFF, W. P. (2008) Evolving Epidemiology of Bacterial Pathogens: Impact on Vaccination Strategies. *6th International Symposium on Pneumococci and Pneumococcal Diseases.* Reykjavik, Iceland.

HAUSDORFF, W. P., BRYANT, J., KLOEK, C., PARADISO, P. R. & SIBER, G. R. (2000a) The Contribution of Specific Pneumococcal Serogroups to Different Disease Manifestations: Implications for Conjugate Vaccine Formulation and Use, Part II. *Clinical Infectious Diseases,* 30**,** 122-140.

HAUSDORFF, W. P., BRYANT, J., PARADISO, P. R. & SIBER, G. R. (2000b) Which pneumococcal serogroups cause the most invasive disease: Implications for conjugate vaccine formulation and use, Part I. *Clinical Infectious Diseases,* 30**,** 100-121.

HAUSDORFF, W. P., SIBER, G. & PARADISO, P. R. (2001) Geographical differences in invasive pneumococcal disease rates and serotype frequency in young children. *Lancet,* 357**,** 950-952.

HAUSDORFF, W. P., YOTHERS, G., R., D., KILPI, T., PELTON, S., COHEN, R., JACOBS, M. R., KAPLAN, S. L., LEVY, C., LOPEZ, E. L., MASON, E. O., SYRIOPOULOU, V., WYNNE, B. & BRYANT, J. (2002) Multinational study of pneumococcal serotypes causing otitis media in children. *Pediatric Infectious Disease Journal.,* 21**,** 1008-1016.

HAVA, D. L. & CAMILLI, A. (2002) Large-scale identification of serotype 4 *Streptococcus pneumoniae* virulence factors. *Molecular Microbiology,* 45**,** 1389-406.

HAVA, D. L., LEMIEUX, J. & CAMILLI, A. (2003) From nose to lung: the regulation behind *Streptococcus pneumoniae* virulence factors. *Molecular Microbiology,* 50**,** 1103-1110.

HAWKINS, J. A., SCAIFE, E. S., HILLMAN, N. D. & FEOLA, G. P. (2004) Current Treatment of Pediatric Empyema. *Seminars in Thoracic and Cardiovascular Surgery,* 16**,** 196-200.

HEDLUND, J., SORBERG, M., NORMARK, B. H. & KRONVALL, G. (2003) Capsular types and antibiotic susceptibility of invasive *Streptococcus pneumoniae* among children in Sweden. *Scandinavian Journal of Infectious Diseases,* 35**,** 452-458.

HEEG, C., FRANKEN, C., VAN DER LINDEN, M., AL-LAHHAM, A. & REINERT, R. R. (2007) Genetic diversity of pneumococcal surface protein A of *Streptococcus pneumoniae* meningitis in German children. *Vaccine,* 25**,** 1030-1035.

HENDLEY, J. O., SANDE, M. A., STEWART, P. M. & GWALTNEY, J. M. (1975) Spread of *Streptococcus pneumoniae* in families. I. Carriage rates and distribution of types. *Journal of Infectious Diseases,* 132**,** 55-61.

HENDRIKSEN, W. T., BLUE, C. E., ADRIAN, P. V., DE GROOT, R., MITCHELL, T. J. & HERMANS, P. W. M. (2004) Transcriptome analysis of *Streptococcus pneumoniae* TCS mutant *delta rr09. 4th International Symposium on Pneumococci and Pneumococcal Diseases.* Helsinki, Finland.

HENDRIKSEN, W. T., SILVA, N., BOOTSMA, H. J., BLUE, C. E., PATERSON, G. K., KERR, A. R., DE JONG, A., KUIPERS, O. P., HERMANS, P. W. & MITCHELL, T. J. (2007) Regulation of gene expression in *Streptococcus pneumoniae* by response regulator 09 is strain dependent. *Journal of Bacteriology,* 189**,** 1382-9.

HENNINK, M., ABBAS, Z., MCDONALD, R. R., NAGLE, E., MONTGOMERY, K. L., DIENER, T., HORSMAN, G. B. & LEVETT, P. N. (2006) *Streptococcus pneumoniae* outbreak in a rural Regina community. *Canada Communicable Disease Report,* 32**,** 181-6.

HENRICHSEN, J. (1999) Typing of *Streptococcus pneumoniae*: Past, Present, and Future. *American Journal of Medicine,* 107 (1A)**,** 50S-54S.

HENRIQUES, B., KALIN, M., ORTQVIST, A., LILJEQUIST, B. O., ALMELA, M., MARRIE, T. J., MUFSON, M. A., TORRES, A., WOODHEAD, M. A., SVENSON, S. B. & KALLENIUS, G. (2000) Molecular epidemiology of *Streptococcus pneumoniae* causing invasive disease in 5 countries. *Journal of Infectious Diseases,* 182**,** 833-839.

HENRIQUES NORMARK, B., KALIN, M., ORTQVIST, T., AKERLUND, B., OLSSON LILJEQUIST, B., HEDLUND, J., SVENSON, S. B., ZHOU, J., SPRATT, B. G., NORMARK, S. & KALLENIUS, G. (2001) Dynamics of penicillin-susceptible clones in invasive pneumococcal disease. *Journal of Infectious Diseases,* 184**,** 861-869.

HERMANS, P. W., SLUIJTER, M., ELZENAAR, K., VAN VEEN, A., SCHONKEREN, J. J., NOOREN, F. M., VAN LEEUWEN, W. J., DE NEELING, A. J., VAN KLINGEREN, B., VERBRUGH, H. A. & DE GROOT, R. (1997a) Penicillin-resistant Streptococcus pneumoniae in the Netherlands: results of a 1-year molecular epidemiologic survey. *Journal of Infectious Diseases,* 175**,** 1413-22.

HERMANS, P. W., SLUIJTER, M., HOOGENBOEZEM, T., HEERSMA, H., VAN BELKUM, A. & DE GROOT, R. (1995) Comparative study of five different DNA fingerprint techniques for molecular typing of *Streptococcus pneumoniae* strains. *Journal of Clinical Microbiology,* 33**,** 1606-12.

HERMANS, P. W. M., SLUIJTER, M., DEJSIRILERT, S., LEMMENS, N., ELZENAAR, K., VANVEEN, A., GOESSENS, W. H. F. & DEGROOT, R. (1997b) Molecular epidemiology of drug-resistant pneumococci: Toward an international approach. *Microbial Drug Resistance,* 3**,** 243-251.

HICKS, L. A., HARRISON, L. H., FLANNERY, B., HADLER, J. L., SCHAFFNER, W., CRAIG, A. S., JACKSON, D., THOMAS, A., BEALL, B., LYNFIELD, R., REINGOLD, A., FARLEY, M. M. & WHITNEY, C. G. (2007) Incidence of Pneumococcal Disease Due to Non-Pneumococcal Conjugate Vaccine (PCV7) Serotypes in the United States during the Era of Widespread PCV7 Vaccination, 1998-2004. *Journal of Infectious Diseases,* 196**,** 1346-1354.

HILLER, N. L., JANTO, B., HOGG, J. S., BOISSY, R., YU, S. S., POWELL, E., KEEFE, R., EHRLICH, N. E., SHEN, K., HAYES, J., BARBADORA, K., KLIMKE, W., DERNOVOY, D., TATUSOVA, T., PARKHILL, J., BENTLEY, S. D., POST, J. C., EHRLICH, G. D. & HU, F. Z. (2007) Comparative genomic analyses of seventeen *Streptococcus pneumoniae* strains: Insights into the pneumococcal supragenome. *Journal of Bacteriology,* 189**,** 8186-8195.

HIMELMAN, R. B. & CALLEN, P. W. (1986) The prognostic value of loculations in parapneumonic pleural effusions. *Chest,* 90**,** 852-6.

HINDS, J., GOULD, K., WITNEY, A., LAMBERTSEN, L., ANTONIO, M., AANENSEN, D. & BENTLEY, S. (2008) S04-O3 Molecular serotyping of *Streptococcus pneumoniae*: a microarray-based genomic tool for isolate typing, detection of multiple carriage and surveillance of serotype replacement in vaccine trials. *6th International Symposium on Pneumococci and Pneumococcal Diseases.* Reykjavik, Iceland.

HINDS, J., LAING, K. G., MANGAN, J. A. & BUTCHER, P. D. (2002a) Glass Slide Microarrays for Bacterial Genomes. *Methods in Microbiology,* 33**,** 83-99.

HINDS, J., WITNEY, A. A. & VASS, J. K. (2002b) Microarray design for bacterial genomes. *Methods in Microbiology,* 33**,** 67-82.

HIRSCH, E. F. & MCKINNEY, M. (1919) An Epidemic of Pneumococcus Bronchopneumonia. *Journal of Infectious Diseases,* 24**,** 594-617.

HIRST, R. A., SIKAND, K. S., RUTMAN, A., MITCHEL, T. J., ANDREW, P. W. & O'CALLAGHAN, C. (2000) Relative Roles of Pneumolysin and Hydrogen Peroxide from *Streptococcus pneumoniae* in Inhibition of Ependymal Ciliary Beat Frequency. *Infection and Immunity,* 68**,** 1557-1562.

HODGES, R. G., MACLEOD, C. M. & BERNHARD, W. G. (1946) Epidemic Pneumococcal Pneumonia III. Pneumococcal Carrier Studies. *American Journal of Hygiene,* 44**,** 207-230.

HOFF, S. J., NEBLETT, W. W., EDWARDS, K. M., HELLER, R. M., PIETSCH, J. B., HOLCOMB, G. W. & HOLCOMB, G. W. (1991) Parapneumonic Empyema In Children - Decortication Hastens Recovery In Patients With Severe Pleural Infections. *Pediatric Infectious Disease Journal,* 10**,** 194-199.

HOFFMAN, O. M., BECKER, D. & WEBER, J. R. (2007) Bacterial hydrogen peroxide contributes to cerebral hyperemia during early stages of experimental pneumococcal meningitis. *Journal of Cerebral Blood Flow and Metabolism,* 27**,** 1792-7.

HOGE, C. W., REICHLER, M. R., DOMINGUEZ, E. A., BREMER, J. C., MASTRO, T. D., HENDRICKS, K. A., MUSHER, D. M., ELLIOTT, J. A., FACKLAM, R. R. & BREIMAN, R. F. (1994) An Epidemic Of Pneumococcal Disease In An Overcrowded, Inadequately Ventilated Jail. *New England Journal Of Medicine,* 331**,** 643-648.

HOLLINGSHEAD, S. K., BECKER, R. & BRILES, D. E. (2000) Diversity of PspA: Mosaic genes and evidence for past recombination in *Streptococcus pneumoniae*. *Infection and Immunity,* 68**,** 5889-5900.

HOLLINGSHEAD, S. K. & BRILES, D. E. (2001) *Streptococcus pneumoniae*: new tools for an old pathogen. *Current Opinion in Microbiology,* 4**,** 71-77.

HOLT, K. E., PARKHILL, J., MAZZONI, C. J., ROUMAGNAC, P., WEILL, F. X., GOODHEAD, I., RANCE, R., BAKER, S., MASKELL, D. J., WAIN, J., DOLECEK, C., ACHTMAN, M. & DOUGAN, G. (2008) High-throughput sequencing provides insights into genome variation and evolution in *Salmonella Typhi*. *Nature Genetics,* 40**,** 987-993.

HOSKINS, J., ALBORN, W. E., JR., ARNOLD, J., BLASZCZAK, L. C., BURGETT, S., DEHOFF, B. S., ESTREM, S. T., FRITZ, L., FU, D. J., FULLER, W., GERINGER, C., GILMOUR, R., GLASS, J. S., KHOJA, H., KRAFT, A. R., LAGACE, R. E., LEBLANC, D. J., LEE, L. N., LEFKOWITZ, E. J., LU, J., MATSUSHIMA, P., MCAHREN, S. M., MCHENNEY, M., MCLEASTER, K., MUNDY, C. W., NICAS, T. I., NORRIS, F. H., O'GARA, M., PEERY, R. B., ROBERTSON, G. T., ROCKEY, P., SUN, P. M., WINKLER, M. E., YANG, Y., YOUNG-BELLIDO, M., ZHAO, G., ZOOK, C. A., BALTZ, R. H., JASKUNAS, S. R., ROSTECK, P. R., JR., SKATRUD, P. L. & GLASS, J. I. (2001) Genome of the bacterium *Streptococcus pneumoniae* strain R6. *Journal of Bacteriology,* 183**,** 5709-17.

HOWARD, S. L., GAUNT, M. W., HINDS, J., WITNEY, A., STABLER, R. & WREN, B. W. (2006) Application of Comparative Phylogenomics To Study the Evolution of *Yersinia enterocolitica* and To Identify Genetic Differences Relating to Pathogenicity. *Journal of Bacteriology,* 188**,** 3645-3653.

HSIEH, Y. C., HSUEH, P. R., LU, C. Y., LEE, P. I., LEE, C. Y. & HUANG, L. M. (2004) Clinical manifestations and molecular epidemiology of necrotizing pneumonia and empyema caused by *Streptococcus pneumoniae* in children in Taiwan. *Clinical Infectious Diseases,* 38**,** 830-835.

HSIEH, Y. C., TSAO, P. N., CHEN, C. L., LIN, T. L., LEE, W. S., SHAO, P. L., LEE, C. Y., HSUEH, P. R., HUANG, L. M. & WANG, J. T. (2008) Establishment of a young mouse model and identification of an allelic variation of *zmpB* in complicated pneumonia caused by *Streptococcus pneumoniae*. *Critical Care Medicine,* 36**,** 1248-1255.

HUTTER, B., SCHAAB, C., ALBRECHT, S., BORGMANN, M., BRUNNER, N. A., FREIBERG, C., ZIEGELBAUER, K., ROCK, C. O., IVANOV, I. & LOFERER, H. (2004) Prediction of Mechanisms of Action of Antibacterial Compounds by Gene Expression Profiling. *Antimicrobial Agents and Chemotherapy,* 48**,** 2838-2844.

HYDE, T. B., GAY, K., STEPHENS, D. S., VUGIA, D. J., PASS, M., JOHNSON, S., BARRETT, N. L., SCHAFFNER, W., CIESLAK, P. R., MAUPIN, P. S., ZELL, E. R., JORGENSEN, J. H., FACKLAM, R. R. & WHITNEY, C. (2001) Macrolide Resistance Among Invasive *Streptococcus pneumoniae* Isolates. *Journal of the American Medical Association,* 286**,** 1857-1862.

IANNELLI, F., OGGIONI, M. R. & POZZI, G. (2002) Allelic variation in the highly polymorphic locus *pspC* of *Streptococcus pneumoniae*. *Gene,* 284**,** 63-71.

INOSTROZA, J., TRUCCO, O., PRADO, V., VINET, A. M., RETAMAL, G., OSSA, G., FACKLAM, R. R. & SORENSEN, R. U. (1998) Capsular Serotype and Antibiotic

Resistance of *Streptococcus pneumoniae* Isolates in Two Chilean Cities. *Clinical and Diagnostic Laboratory Immunology,* 5**,** 176-180.

INOSTROZA, J., VINET, A. M., RETAMAL, G., LORCA, P., OSSA, G., FACKLAM, R. R. & SORENSEN, R. U. (2001) Influence of Patient Age on *Streptococcus pneumoniae* Serotypes Causing Invasive Disease. *Clinical And Diagnostic Laboratory Immunology,* 8**,** 556-559.

INVERARITY, D., DIGGLE, M., EDWARDS, G. & MITCHELL, T. (2007) An evaluation of media suitable for the transportation by air of *Streptococcus pneumoniae* isolates. *Journal of Infection,* 55**,** e65-e66.

INVERARITY, D., LAMB, K., DIGGLE, M., ROBERTSON, C., GREENHALGH, D., MITCHELL, T., SMITH, A., JEFFERIES, J., MCMENAMIN, J., CLARKE, S. & EDWARDS, G. (2008) P2-136 Death or Survival from Invasive Pneumococcal Disease in Scotland: Associations With Serogroups and Multilocus Sequence Types. *6th International Symposium on Pneumococci and Pneumococcal Diseases.* Reykjavik, Iceland.

JACOBS, M. R. (2002) *In vitro Veritas*: *In Vitro* Macrolide Resistance in Systemic *Streptococcus pneumoniae* Infections Does Result in Clinical Failure. *Clinical Infectious Diseases,* 35**,** 565-569.

JAIN, R. & DANZIGER, L. H. (2004) The Macrolide Antibiotics: A Pharmacokinetic and Pharmacodynamic Overview. *Current Pharmaceutical Design,* 10**,** 3045-3053.

JEFFERIES, J., SMITH, A., COWAN, G., MCMENAMIN, J., EDWARDS, G., CLARKE, S. & MITCHELL, T. (2008) P2-012 Temporal analysis of invasive pneumococcal clones in Scotland 2001-2006. *6th International Symposium on Pneumococci and Pneumococcal Diseases.* Reykjavik, Iceland.

JEFFERIES, J. M., JOHNSTON, C. H., KIRKHAM, L. A., COWAN, G. J., ROSS, K. S., SMITH, A., CLARKE, S. C., BRUEGGEMANN, A. B., GEORGE, R. C., PICHON, B., PLUSCHKE, G., PFLUGER, V. & MITCHELL, T. J. (2007) Presence of nonhemolytic pneumolysin in serotypes of *Streptococcus pneumoniae* associated with disease outbreaks. *Journal of Infectious Diseases,* 196**,** 936-44.

JEFFERIES, J. M. C., CLARKE, S. C., SMITH, A., DOWSON, C. & MITCHELL, T. J. (2003) Automated pneumococcal MLST using liquid-handling robotics and a capillary DNA sequencer. *Molecular Biotechnology,* 24**,** 303-308.

JEFFERIES, J. M. C., SMITH, A., CLARKE, S. C., DOWSON, C. & MITCHELL, T. J. (2004) Genetic analysis of diverse disease-causing pneumococci indicates high levels of diversity within serotypes and capsule switching. *Journal of Clinical Microbiology,* 42**,** 5681-5688.

JENNI, S. & BAN, N. (2003) The chemistry of protein synthesis and voyage through the ribosomal tunnel. *Current Opinion in Structural Biology,* 13**,** 212-219.

JORGENSEN, G., SINGLETON, R., BUTLER, J., COTTLE, T., HENNESSY, T., HURLBURT, D., RUDOLPH, K. & PARKS, D. (2005) Outbreak of Invasive Pneumococcl Disease - Alaska, 2003-2004. *Morbidity and Mortality Weekly Report,* 54**,** 72-75.

JOYCE, E. A., CHAN, K., SALAMA, N. R. & FALKOW, S. (2002) Redefining bacterial populations: A post-genomic reformation. *Nature Reviews Genetics,* 3**,** 462-473.

KANUNGO, R. & RAJALAKSHMI, B. (2001) Serotype distribution & antimicrobial resistance in *Streptococcus pneumoniae* causing invasive & other infections in south India. *Indian Journal of Medical Research,* 114**,** 127-132.

KAUSHIK, D. K. & SEHGAL, D. (2008) Developing Antibacterial Vaccines in Genomics and Proteomics Era. *Scandinavian Journal of Immunology,* 67**,** 544-552.

KEENAN, R. J., FREYMANN, D. M., STROUD, R. M. & WALTER, P. (2001) The signal recognition particle. *Annual Reviews in Biochemistry,* 70**,** 755-775.

KEITH, E. R. & MURDOCH, D. R. (2008) Antimicrobial Susceptibility Profile of *Streptococcus pseudopneumoniae* Isolated from Sputum. *Antimicrobial Agents and Chemotherapy,* 52**,** 2998.

KELL, C. M., JORDENS, J. Z., DANIELS, M., COFFEY, T. J., BATES, J., PAUL, J., GILKS, C. & SPRATT, B. G. (1993) Molecular Epidemiology of Penicillin-Resistant Pneumococci Isolated in Nairobi, Kenya. *Infection and Immunity,* 61**,** 4382-4391.

KELLNER, J. D., MCGEER, A., CETRON, M. S., LOW, D. E., BUTLER, J. C., MATLOW, A., TALBOT, J. & FORD-JONES, E. L. (1998) The use of *Streptococcus pneumoniae* nasopharyngeal isolates from healthy children to predict invasive disease. *Pediatric Infectious Disease Journal,* 17**,** 279-286.

KELLY, T., DILLARD, J. P. & YOTHER, J. (1994) Effect of genetic switching of capsular type on virulence of *Streptococcus pneumoniae*. *Infection and Immunity,* 62**,** 1813-1819.

KERCHER, K. W., ATTORRI, R. J., HOOVER, D. & MORTON, D. (2000) Thorascopic Decortication as First-Line Therapy for Pediatric Parapneumonic Empyema. A Case Series. *Chest***,** 24-27.

KEREM, E., ZIV, Y. B., RUDENSKI, B., KATZ, S., KLEID, D. & BRANSKI, D. (1994) Bacteremic Necrotizing Pneumococcal Pneumonia in Children. *American Journal of Respiratory and Critical Care Medicine,* 149**,** 242-4.

KERR, A. R., PATERSON, G. K., MCCLUSKEY, J., IANNELLI, F., OGGIONI, M. R., POZZI, G. & MITCHELL, T. J. (2006) The contribution of PspC to pneumococcal virulence varies between strains and is accomplished by both complement evasion and complement-independent mechanisms. *Infection and Immunity,* 74**,** 5319-24.

KERTESZ, D. A., DI FABIO, J. L., BRANDILEONE, M. C. D., CASTANEDA, E., ECHANIZ-AVILES, G., HEITMANN, I., HOMMA, A., HORTAL, M., LOVGREN, M., RUVINSKY, R. O., TALBOT, J. A., WEEKES, J. & SPIKA, J. S. (1998) Invasive *Streptococcus pneumoniae* infection in Latin American children: Results of the Pan American Health Organization surveillance study. *Clinical Infectious Diseases,* 26**,** 1355-1361.

KIENINGER, D. M., KUEPER, K., STEUL, K., JUERGENS, C., AHLERS, N., BAKER, S., GIARDINA, P., GRUBER, W. & SCOTT, D. (2008) G-2117 Safety and Immunologic Non-Inferiority of 13-Valent Pneumococcal Conjugate Vaccine

Compared to 7-Valent Pneumococcal Conjugate Vaccine Given with Routine Vaccines in Healthy Infants *48th Annual Interscience Conference on Antimicrobial Agents and Chemotherapy (ICAAC) and the Infectious Diseases Society of America (IDSA) 46th Annual Meeting.* Washington DC, USA.

KILGORE, P. E., NYAMBAT, B., HAN, S. H., MCINTYRE, P. B., SHEN, X. Z., SOENARTO, Y., YONG, D. E., CHIU, C. H., THARAVICHITKUL, P., ANH, D. D., NGO, T. T. & HAUSDORFF, W. P. (2006) PO3.24 A multi-centre retrospective analysis of empyema and pleural effusion in hospitalised Asian children. *5th International Symposium on Pneumococci and Pneumococcal Diseases.* Alice Springs, Australia.

KING, S. J., HIPPE, K. R., GOULD, J. M., BAE, D., PETERSON, S., CLINE, R. T., FASCHING, C., JANOFF, E. N. & WEISER, J. N. (2004) Phase variable desialylation of host proteins that bind to *Streptococcus pneumoniae* in vivo and protect the airway. *Molecular Microbiology,* 54**,** 159-171.

KIRKHAM, L. A. S., JEFFERIES, J. M. C., KERR, A. R., JING, Y., CLARKE, S. C., SMITH, A. & MITCHELL, T. J. (2006) Identification of invasive serotype 1 pneumococcal isolates that express nonhemolytic pneumolysin. *Journal of Clinical Microbiology,* 44**,** 151-159.

KLEIN, J. O. (1981) The Epidemiology of Pneumococcal Disease in Infants and Children. *Reviews in Infectious Diseases,* 3**,** 246-253.

KLUDT, P., LETT, S. M., DEMARIA, A., CRUTCHER, J. M., CURTIS, N. M., GARVEY, C. W., FRANK, L. L., DANNER, A., BENJAMIN, G. C. & DWYER, D. (1997) Outbreaks of Pneumococcal Pneumonia Among Unvaccinated Residents in Chronic-Care Facilities - Massachusetts, October 1995, Oklahoma, February 1996, and Maryland, May-June 1996. *Morbidity and Mortality Weekly Report,* 46**,** 60-62.

KO, A. I., REIS, J. N., COPPOLA, S. J., GOUVEIA, E. L., CORDEIRO, S. M., LOBO, T. S., PINHEIRO, R. M., SALGADO, K., RIBEIRO DOURADO, C. M., TAVARES-NETO, J., ROCHA, H., GALVAO REIS, M., JOHNSON, W. D., JR. & RILEY, L. W. (2000) Clonally related penicillin-nonsusceptible *Streptococcus pneumoniae* serotype 14 from cases of meningitis in Salvador, Brazil. *Clinical Infectious Diseases,* 30**,** 78-86.

KO, K. S., PARK, S., OH, W. S., SUH, J. Y., OH, T., AHN, S., CHUN, J. & SONG, J. H. (2006) Comparative analysis of growth-phase-dependent gene expression in virulent and avirulent *Streptococcus pneumoniae* using a high-density DNA microarray. *Molecules and Cells,* 21**,** 82-88.

KOECK, J. L., NJANPOP-LAFOURCADE, B. M., CADE, S., VARON, E., SANGARE, L., VALJEVAC, S., VERGNAUD, G. & POURCEL, C. (2005) Evaluation and selection of tandem repeat loci for *Streptococcus pneumoniae* MLVA strain typing. *BMC Microbiology,* 5**,** 66.

KOEDEL, U., SCHELD, W. M. & PFISTER, H. W. (2002) Pathogenesis and Pathophysiology of Pneumococcal Meningitis. *Lancet Infectious Diseases,* 2**,** 721-736.

KOEUTH, T., VERSALOVIC, J. & LUPSKI, J. R. (1995) Differential subsequence conservation of interspersed repetitive *Streptococcus pneumoniae* BOX elements in diverse bacteria. *Genome Research,* 5**,** 408-418.

KOPPEL, E. A., SAELAND, E., DE COOKER, D. J., VAN KOOYK, Y. & GEIJTENBEEK, T. B. (2005) DC-SIGN specifically recognizes *Streptococcus pneumoniae* serotypes 3 and 14. *Immunobiology,* 210**,** 203-10.

KORCZAK, B., FREY, J., SCHRENZEL, J., PLUSCHKE, G., PFISTER, R., EHRICHT, R. & KUHNERT, P. (2005) Use of Diagnostic Microarrays for Determination of Virulence Gene Patterns of *Escherichia coli* K1, a Major Cause of Neonatal Meningitis. *Journal of Clinical Microbiology,* 43**,** 1024-1031.

KORNECKI, A. & SIVAN, Y. (1997) Treatment of Loculated Pleural Effusion With Intrapleural Urokinase in Children. *Journal of Pediatric Surgery,* 32**,** 1473-1475.

KOTHANDAPANI, J. S. G., HEAF, L., SOUTHERN, K., RIORDAN, A., COURIEL, J. & HEAF, D. (2006) P067 Increased incidence of empyema in children - a regional unit experience. *Federation of Infection Societies Conference.* Cardiff, United Kingdom.

KRAMER, G., RAUCH, T., RIST, W., VORDERWULBECKE, S., PATZELT, H., SCHULZE-SPECKING, A., BAN, N., DEUERLING, E. & BUKAU, B. (2002) L23 protein functions as a chaperone docking site on the ribosome. *Nature,* 419**,** 171-174.

KRISTINSSON, K. G. (2008) S05-KS2 Pneumococcal diversity - Success and failure of pneumococcal clones. *6th International Symposium on Pneumococci and Pneumococcal Diseases.* Reykjavik, Iceland.

KUMAR, A., GOEL, G., FEHRENBACH, E., PUNIYA, A. K. & SINGH, K. (2005) Microarrays: The technology, analysis and application. *Engineering in Life Sciences,* 5**,** 215-222.

KYAW, M. H., CHRISTIE, P., CLARKE, S. C., MOONEY, J. D., AHMED, S., JONES, I. G. & CAMPBELL, H. (2003) Invasive pneumococcal disease in Scotland, 1999-2001: Use of record linkeage to explore associations between patients and disease in relation to future vaccination policy. *Clinical Infectious Diseases,* 37**,** 1283-91.

KYAW, M. H., CLARKE, S., JONES, I. G. & CAMPBELL, H. (2002) Non-invasive pneumococcal disease and antimicrobial resistance: vaccine implications. *Epidemiology and Infection,* 128**,** 21-27.

LAHTI, E., MERTSOLA, J., KONTIOKARI, T., EEROLA, E., RUUSKANEN, O. & JALAVA, J. (2006) Pneumolysin polymerase chain reaction for diagnosis of pneumococcal pneumonia and empyema in children. *European Journal of Clinical Microbiology and Infectious Diseases,* 25**,** 783-789.

LAIBLE, G., SPRATT, B.G., HAKENBACK, R. (1991) Interspecies recombinational events during the evolution of altered PBP 2x genes in penicillin-resistant clinical isolates of *Streptococcus pneumoniae. Molecular Microbiology,* 5**,** 1993-2002.

LAMB, F. H. & BRANNIN, E. B. (1919) The Epidemic Respiratory Infection at Camp Cody, N.M. *Journal of the American Medical Association,* 72**,** 1056-1062.

LAMB, K., DIGGLE, M., INVERARITY, D., JEFFERIES, J. M., EDWARDS, G., MCMENAMIN, J., MITCHELL, T. J. & CLARKE, S. C. (2008) P1-012 Trends in Serotypes and Multi Locus Sequence Types (MLST) among cases of Invasive Pneumococcal Disease (IPD) in Scotland. *6th International Symposium on Pneumococci and Pneumococcal Diseases.* Reykavik, Iceland.

LAN, R. & REEVES, P. R. (2001) When does a clone deserve a name? A perspective on bacterial species based on population genetics. *Trends in Microbiology,* 9**,** 419-424.

LANGE, R., WAGNER, C., DE SAIZIEU, A., FLINT, N., MOLNOS, J., STIEGER, M., CASPERS, P., KAMBER, M., KECK, W. & AMREIN, K. E. (1999) Domain organization and molecular characterization of 13 two-component systems identified by genome sequencing of *Streptococcus pneumoniae*. *Gene,* 237**,** 223-34.

LANIE, J. A., NG, W. L., KAZMIERCZAK, K. M., ANDRZEJEWSKI, T. M., DAVIDSEN, T. M., WAYNE, K. J., TETTELIN, H., GLASS, J. I. & WINKLER, M. E. (2007) Genome sequence of Avery's virulent serotype 2 strain D39 of Streptococcus pneumoniae and comparison with that of unencapsulated laboratory strain R6. *Journal of Bacteriology,* 189**,** 38-51.

LAU, G. W., HAATAJA, S., LONETTO, M., KENSIT, S. E., MARRA, A., BRYANT, A. P., MCDEVITT, D., MORRISON, D. A. & HOLDEN, D. W. (2001) A functional genomic analysis of type 3 *Streptococcus pneumoniae* virulence. *Molecular Microbiology,* 40**,** 555-571.

LAVAL, C. A. B., PIMENTA, F. C., DE ANDRADE, J. G., ANDRADE, S. S. & DE ANDRADE, A. L. S. S. (2003) Progress Towards Meningitis Prevention in the Conjugate Vaccines Era. *The Brazilian Journal of Infectious Diseases,* 7**,** 315-324.

LAVAL, C. B., DE ANDRADE, A. L. S. S., PIMENTA, F. C., DE ANDRADE, J. G., DE OLIVEIRA, R. M., SILVA, S. A., DE LIMA, E. C., DI FABIO, J. L., CASAGRANDE, S. T. & BRANDILEONE, M. C. C. (2006) Serotypes of carriage and invasive isolates of *Streptococcus pneumoniae* in Brazilian children in the era of pneumococcal vaccinations. *Clinical Microbiology and Infection,* 12**,** 50-55.

LE HELLO, S., LEVY, M., YVON, J. F., CHARAVAY, F., BROWN, M. & PAGE, S. (2008.) P2-022 Predominance of ST306 serotype 1 among invasive *Streptococcus pneumoniae* in the South Pacific. *6th International Symposium on Pneumococci and Pneumococcal Diseases.* Reykjavik, Iceland.

LE MONNIER, A., CARBONNELLE, E., ZAHAR, J. R., LE BOURGEOIS, M., ABACHIN, E., QUESNE, G., VARON, E., DESCAMPS, P., DE BLIC, J., SCHEINMANN, P., BERCHE, P. & FERRONI, A. S. (2006) Microbiological diagnosis of empyema in children: Comparative evaluations by culture, polymerase chain reaction, and pneumococcal antigen detection in pleural fluids. *Clinical Infectious Diseases,* 42**,** 1135-1140.

LEACH, A. J., BOSWELL, J. B., ASCHE, V., NIENHUYS, T. G. & MATHEWS, J. D. (1994) Bacterial Colonization Of The Nasopharynx Predicts Very Early-Onset And Persistence Of Otitis-Media In Australian Aboriginal Infants. *Pediatric Infectious Disease Journal,* 13**,** 983-989.

LEAL, A. L. & CASTANEDA, E. (1997) Antibiotic susceptibility of *Streptococcus pneumoniae* colonizing the nasopharynx of Colombian children with pneumonia. *Revista Panamericana de Salud Publica,* 1**,** 266-72.

LEE, C. F., LIU, S. C., LUE, K. H., CHEN, J. P. & SHEU, J. N. (2006) Pneumococcal pneumonia with empyema and hemolytic uremic syndrome in children: report of three cases. *Journal of Microbiology, Immunology and Infection,* 39**,** 348-52.

LEFEVRE, J. C., FAUCON, G., SICARD, A. M. & GASC, A. M. (1993) DNA Fingerprinting of *Streptococcus pneumoniae* Strains by Pulsed-Field Gel Electrophoresis. *Journal of Clinical Microbiology,* 31**,** 2724-2728.

LEIBERMAN, A., DAGAN, R., LEIBOVITZ, E., YAGUPSKY, P. & FLISS, D. M. (1999) The bacteriology of the nasopharynx in childhood. *International Journal of Pediatric Otorhinolaryngology,* 49**,** S151-153.

LEIGHTON, C., PIPER, D., GUNDERMAN-KING, J., REA, V., GENSHEIMER, K., RANDOLPH, J., DANFORTH, R., WEBBER, L., PRITCHARD, E., BECKETT, G., SHINDE, V., FACKLAM, R., WHITNEY, C., HAYES, N. & FLANNERY, B. (2003) Pneumococcal Conjunctivitis at an Elementary School. *Morbidity and Mortality Weekly Report,* 52**,** 64-66.

LEIMKUGEL, J., FORGOR A.A., GAGNEUX S., PFLUGER V., FLIERL C., AWINE E., NAEGELI M., DANGY J-P., SMITH T., HODGSON A. & G., P. (2005) An Outbreak of Serotype 1 *Streptococcus pneumoniae* Meningitis in Northern Ghana with Features That Are Characteristic of *Neisseria meningitidis* Meningitis Epidemics. *Journal of Infectious Diseases,* 192**,** 192-198.

LEMIEUX, J., HAVA, D. L., BASSET, A. & CAMILLI, A. (2006) RrgA and RrgB are components of a multisubunit pilus encoded by the *Streptococcus pneumoniae* rlrA pathogenicity islet. *Infection and Immunity,* 74**,** 2453-6.

LEONARD, E. E., TAKATA, T., BLASER, M. J., FALKOW, S., TOMPKINS, L. S. & GAYNOR, E. C. (2003) Use of an open-reading-frame-specific *Campylobacter jejuni* DNA microarray as a new genotyping tool for studying epidemiologically related isolates. *Journal of Infectious Diseases,* 187**,** 691-94.

LEVINE, M. M., LAGOS, R., LEVINE, O. S., HEITMANN, I., ENRIQUEZ, N., PINTO, M. E., ALVAREZ, A. M., WU, E., MAYORGA, C. & REYES, A. (1998) Epidemiology of invasive pneumococcal infections in infants and young children in metropolitan Santiago, Chile, a newly industrializing country. *Pediatric Infectious Disease Journal,* 17**,** 287-293.

LEVINE, O. S., O'BRIEN, K. L., KNOLL, M., ADEGBOLA, R. A., BLACK, S., CHERIAN, T., DAGAN, R., GOLDBLATT, D., GRANGE, A., GREENWOOD, B., HENNESSY, T., KLUGMAN, K. P., MADHI, S. A., MULHOLLAND, K., NOHYNEK, H., SANTOSHAM, M., SAHA, S., SCOTT, J. A., SOW, S., WHITNEY, C. & CUTTS, F. (2006) Pneumococcal vaccination in developing countries. *Lancet,* 367**,** 1880-1882.

LIN, B. C., BLANEY, K. M., MALANOSKI, A. P., LIGLER, A. G., SCHNUR, J. M., METZGAR, D., RUSSELL, K. L. & STENGER, D. A. (2007) Using a resequencing microarray as a multiple respiratory pathogen detection assay. *Journal of Clinical Microbiology,* 45**,** 443-452.

LIN, C. J., CHEN, P. Y., HUANG, F. L., LEE, T., CHI, C. S. & LIN, C. Y. (2006) Radiographic, clinical, and prognostic features of complicated and uncomplicated community-acquired lobar pneumonia in children. *Journal of Microbiology, Immunology and Infection,* 39**,** 489-95.

LIN, J. T., CONNELLY, M. B., AMOLO, C., OTANI, S. & YAVER, D. S. (2005) Global Transcriptional Response of *Bacillus subtilis* to Treatment with Subinhibitory Concentrations of Antibiotics That Inhibit Protein Synthesis *Antimicrobial Agents and Chemotherapy,* 49**,** 1915-1926.

LINDBERG, J. & FANGEL, S. (1999) Recurrent Endocarditis Caused by *Streptococcus pneumonaie*. *Scandinavian Journal of Infectious Diseases,* 31**,** 409-410.

LINDSAY, J. A., MOORE, C. E., DAY, N. P., PEACOCK, S. J., WITNEY, A., STABLER, R. A., HUSAIN, S. E., BUTCHER, P. D. & HINDS, J. (2006) Microarrays Reveal that each of the Ten Dominant Lineages of *Staphylococcus aureus* Has a Unique Combination of Surface-Associated and Regulatory Genes. *Journal of Bacteriology,* 188**,** 669-676.

LISTER, F. S. (1913) Specific serological reactions with pneumococci from different sources. *South African Institute for Medical Research* Publication No. 2.

LIU, R. H., YANG, J., LENIGK, R., BONANNO, J. & GRODZINSKI, P. (2004) Self-Contained, Fully Integrated Biochip for sample Preparation, Polymerase Chain Reaction, Amplification and DNA microarray detection. *Analytical Chemistry,* 76**,** 1824-1831.

LLOYD-EVANS, N., O'DEMPSEY, T. J. D., BALDEH, I., SECKA, O., DEMBA, E., TODD, J. E., MCARDLE, T. F., BANYA, W. & GREENWOOD, B. M. (1996) Nasopharyngeal carriage of pneumococci in Gambian children and in their families. *Pediatric Infectious Disease Journal,* 15**,** 866-871.

LLULL, D., LOPEZ, R. & GARCIA, E. (2000) Clonal Origin of the Type 37 *Streptococcus pneumoniae*. *Microbial Drug Resistance,* 6**,** 269-275.

LLULL, D., MUNOZ, R., LOPEZ, R. & GARCIA, E. (1999) A single gene (*tts*) located outside the cap locus directs the formation of *Streptococcus pneumoniae* type 37 capsular polysaccharide. Type 37 pneumococci are natural, genetically binary strains. *Journal of Experimental Medicine,* 190**,** 241-251.

LOMHOLT, H. (1995) Evidence of Recombination and an Antigenically Diverse Immunoglobulin-A1 Protease among Strains of *Streptococcus pneumoniae*. *Infection and Immunity,* 63**,** 4238-4243.

LOPEZ, R. (2006) Pneumococcus: the sugar-coated bacteria. *International Microbiology,* 9**,** 179-190.

LOPEZ, R. & GARCIA, E. (2004) Recent trends on the molecular biology of pneumococcal capsules, lytic enzymes, and bacteriophage. *FEMS Microbiology Reviews,* 28**,** 553-580.

LOPEZ, R., GONZALEZ, M. P., GARCIA, E., GARCIA, J. L. & GARCIA, P. (2000) Biological roles of two new murein hydrolases of *Streptococcus pneumoniae*

representing examples of module shuffling. *Research in Microbiology,* 151**,** 437-443.

LOVERA, D. & ARBO, A. (2005) Risk factors for mortality in Paraguayan children with pneumococcal bacterial meningitis. *Tropical Medicine and International Health,* 10**,** 1235-41.

LOVGREN, M., TALBOT, J. A., BRANDILEONE, M. C., CASAGRANDE, S., AGUDELO, C. I., CASTANEDA, E., REGUEIRA, M., CORSO, A., HEITMANN, I., MALDONADO, A., ECHANIZ-AVILES, G., SOTO-NOGUERON, A., HORTAL, M., CAMOU, T., GABASTOU, J. M. & DI FABIO, J. L. (2007) Evolution of an International External Quality Assurance Model To Support Laboratory Investiagtion of *Streptococcus penumoniae*, Developed for the SIREVA Project in Latin America, from 1993 to 2005. *Journal of Clinical Microbiology,* 45**,** 3184-3190.

MACKENZIE, G. M., MCKEE, T. M. & TEPPERMAN, J. (1940) Epidemiology of an outbreak of pneumococcal pneumonia in a rural community. *Transactions of the Association of American Physicians,* 55**,** 199-208.

MAIDEN, M. C. J. (1998) Horizontal genetic exchange, evolution, and spread of antibiotic resistance in bacteria. *Clinical Infectious Diseases,* 27**,** S12-S20.

MAIDEN, M. C. J., BYGRAVES, J. A., FEIL, E. J., MORELLI, G., RUSSELL, J. E., URWIN, R., ZHANG, Q., ZHOU, J., ZURITH, K., CAUGANT, D. A., FEAVERS, I. M., ACHTMAN, M. & SPRATT, B. G. (1998) Multilocus sequence typing: A portable approach to the identification of clones within populations of pathogenic microorganisms. *Proceedings of the National Academy of Sciences of the United States of America,* 95**,** 3140-3145.

MALLOCH, A. (1922) A Study of Pneumonia and Types of Pneumococci. *Quarterly Journal of Medicine,* 15**,** 103-113.

MANDIGERS, C. M. P. W., DIEPERSLOOT, R. J. A., DESSENS, M., MOL, S. J. M. & VANKLINGEREN, B. (1994) A Hospital Outbreak Of Penicillin-Resistant Pneumococci In The Netherlands. *European Respiratory Journal,* 7**,** 1635-1639.

MANN, B., ORIHUELA, C., ANTIKAINEN, J., GAO, G., SUBLETT, J., KORHONEN, T. K. & TUOMANEN, E. (2006) Multifunctional Role of Choline Binding Protein G in Pneumococcal Pathogenesis. *Infection and Immunity,* 74**,** 821-829.

MARGULIES, M., EGHOLM, M., ALTMAN, W. E., ATTIYA, S., BADER, J. S., BEMBEN, L. A., BERKA, J., BRAVERMAN, M. S., CHEN, Y. J., CHEN, Z., DEWELL, S. B., DU, L., FIERRO, J. M., GOMES, X. V., GODWIN, B. C., HE, W., HELGESEN, S., HO, C. H., IRZYK, G. P., JANDO, S. C., ALENQUER, M. L., JARVIE, T. P., JIRAGE, K. B., KIM, J. B., KNIGHT, J. R., LANZA, J. R., LEAMON, J. H., LEFKOWITZ, S. M., LEI, M., LI, J., LOHMAN, K. L., LU, H., MAKHIJANI, V. B., MCDADE, K. E., MCKENNA, M. P., MYERS, E. W., NICKERSON, E., NOBILE, J. R., PLANT, R., PUC, B. P., RONAN, M. T., ROTH, G. T., SARKIS, G. J., SIMONS, J. F., SIMPSON, J. W., SRINIVASAN, M., TARTARO, K. R., TOMASZ, A., VOGT, K. A., VOLKMER, G. A., WANG, S. H., WANG, Y., WEINER, M. P., YU, P., BEGLEY, R. F. & ROTHBERG, J. M. (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature,* 437**,** 376-80.

MARRA, A., ASUNDI, J., BARTILSON, M., LAWSON, S., FANG, F., CHRISTINE, J., WIESNER, C., BRIGHAM, D., SCHNEIDER, W. P. & HROMOCKYJ, A. E. (2002) Differential Fluorescence Induction Analysis of *Streptococcus pneumoniae* Identifies Genes Involved in Pathogenesis. *Infection and Immunity,* 70**,** 1422-1433.

MARTENS, P., WORM, S. W., LUNDGREN, B., KONDRADSEN, H. B. & BENFIELD, T. (2004) Serotype-specific mortality from invasive *Streptococcus pneumoniae* disease revisited. *BMC Infectious Diseases,* 4**,** 21.

MARTIN-GALIANO, A. J., WELLS, J. M. & DE LA CAMPA, A. G. (2004) Relationship between codon biased genes, microarray expression values and physiological characteristics of *Streptococcus pneumoniae*. *Microbiology,* 150**,** 2313-2325.

MARTIN, D. R. & BRETT, M. S. (1996) Pneumococci causing invasive disease in New Zealand, 1987-94: serogroup and serotype coverage and antibiotic resistances. *New Zealand Medical Journal,* 108**,** 288-290.

MARTIN, M., TURCO, J. H., ZEGANS, M. E., FACKLAM, R. R., SODHA, S., ELLIOTT, J. A., PRYOR, J. H., BEALL, B., ERDMAN, D. D., BAUMGARTNER, Y. Y., SANCHEZ, P. A., SCHWARTZMAN, J. D., MONTERO, J., SCHUCHAT, A. & WHITNEY, C. G. (2003a) An Outbreak of Conjunctivitis Due to Atypical *Streptococcus pneumoniae*. *New England Journal of Medicine,* 348**,** 1112-21.

MARTIN, M. J., HERRERO, J., MATEOS, A. & DOPAZO, J. (2003b) Comparing Bacterial Genomes Through Conservation Profiles. *Genome Research,* 13**,** 991-998.

MASCHER, T., HEINTZ, M., ZAHNER, D., MERAI, M. & HAKENBECK, R. (2006) The CiaRH system of *Streptococcus pneumoniae* prevents lysis during stress induced by treatment with cell wall inhibitors and by mutations in pbp2x involved in beta-lactam resistance. *Journal of Bacteriology,* 188**,** 1959-68.

MASCHER, T., ZAHNER, D., MERAI, M., BALMELLE, N., DE SAIZIEU, A. B. & HAKENBECK, R. (2003) The *Streptococcus pneumoniae* cia regulon: CiaR target sites and transcription profile analysis. *Journal of Bacteriology,* 185**,** 60-70.

MAVROIDI, A., GODOY, D., AANENSEN, D. M., ROBINSON, D. A., HOLLINGSHEAD, S. K. & SPRATT, B. G. (2004) Evolutionary genetics of the capsular locus of serogroup 6 pneumococci. *Journal of Bacteriology,* 186**,** 8181-8192.

MCCLUSKEY, J., DOWSON, C. G. & MITCHELL, T. J. (2002) The use of microarray technology for the analysis of *Streptococcus pneumoniae*. *Comparative and Functional Genomics,* 3**,** 366-368.

MCDANIEL, L. S., THORNTON, J. A. & MCDANIEL, D. O. (2004) Use of cDNA microarrays to analyze responses to pneumococcal virulence factors. *Indian Journal of Medical Research,* 119**,** 99-103.

MCDOUGAL, L. K., RASHEED, J. K., BIDDLE, J. W. & TENOVER, F. C. (1995) Identification of Multiple Clones of Extended-Spectrum Cephalosporin-Resistant *Streptococcus pneumoniae* Isolates in the United States. *Antimicrobial Agents and Chemotherapy,* 39**,** 2282-2288.

MCDOUGAL, L. K., TENOVER, F. C., LEE, L. N., RASHEED, J. K., PATTERSON, J. E., JORGENSEN, J. H. & LEBLANC, D. J. (1998) Detection of Tn*917*-like sequences within a Tn*916*-like conjugative transposon (Tn*3872*) in erythromycin resistant isolates of *Streptococcus pneumoniae*. *Antimicrobial Agents and Chemotherapy,* 42**,** 2312-2318.

MCELLISTREM, M. C., ADAMS, J. M., PATEL, K., MENDELSOHN, A. B., KAPLAN, S. L., BRADLEY, J. S., SCHUTZE, G. E., KIM, K. S., MASON, E. O. & WALD, E. R. (2005) Acute otitis media due to penicillin-nonsusceptible *Streptococcus pneumoniae* before and after the introduction of the pneumococcal conjugate vaccine. *Clinical Infectious Diseases,* 40**,** 1738-1744.

MCELLISTREM, M. C., RANSFORD, J. V. & KHAN, S. A. (2007) Characterization of *in vitro* biofilm-associated pneumococcal phase variants of a clinically relevant serotype 3 clone. *Journal of Clinical Microbiology,* 45**,** 97-101.

MCGEE, L., MCDOUGAL, L., ZHOU, J., SPRATT, B. G., TENOVER, F. C., GEORGE, R., HAKENBECK, R., HRYNIEWICZ, W., LEFEVRE, J. C., TOMASZ, A. & KLUGMAN, K. P. (2001a) Nomenclature of major antimicrobial-resistant clones of *Streptococcus pneumoniae* defined by the pneumococcal molecular epidemiology network. *Journal of Clinical Microbiology,* 39**,** 2565-2571.

MCGEE, L., WANG, H., WASAS, A., HUEBNER, R., CHEN, M. & KLUGMAN, K. P. (2001b) Prevalence of serotypes and molecular epidemiology of *Streptococcus pneumoniae* strains isolated from children in Beijing, China: Identification of two novel multiply-resistant clones. *Microbial Drug Resistance,* 7**,** 55-63.

MCINTYRE, P. B., BERKEV, C. S., KING, S. M., SCHAAD, U. B., KILOI, T. & KANRA, G. Y. (1997) Dexamethasone as adjunctive therapy in bacterial meningitis. A meta analysis of randomized clinical trials since 1988. *Journal of the American Medical Association,* 278**,** 925-931.

MCLELLAND, J. E. (1918) Importance of Blood Cultures in Pneumonia. *Journal of the American Medical Association,* 71**,** 1299-1301.

MEDEIROS, M. I. C., NEME, S. N., DA SILVA, P., SILVA, J. O., CARNEIRO, A. M. M., CARLONI, M. C. & BRANDILEONE, M. C. (1998) *Streptococcus pneumoniae* and *Haemophilus influenzae* as etiological agents of conjunctivitis outbreaks in the region of Ribeirao Preto, SP, Brazil. *Revista do Instituto de Medicina Tropical de Sao Paulo,* 40**,** 7-10.

MELAMED, R., GREENBERG, D., LANDAU, D., KHVATSKIN, S., SHANY, E. & DAGAN, R. (2002) Neonatal Nosocomial Pneumococcal Infections Acquired by Patient-to-patient Transmission. *Scandinavian Journal of Infectious Diseases,* 34**,** 385-386.

MENDONCA-SOUZA, C. R., CARVALHO, M. D. A. G., BARROS, R. R., DIAS, C. A., SAMPAIO, J. L., CASTRO, A. C., FACKLAM, R. R. & TEIXEIRA, L. M. (2004) Occurrence and Characteristics of Erythromycin-Resistant *Streptococcus pneumoniae* Strains Isolated in Three Major Brazilian States. *Microbial Drug Resistance,* 10**,** 313-20.

MENEZES-MARTINS, L. F., MENEZES-MARTINS, J. J., MICHAELSEN, V. S., AGUIAR, B. B., ERMEL, T. & MACHADO, D. C. (2005) Diagnosis of

parapneumonic pleural effusion by polymerase chain reaction in children. *Journal of Pediatric Surgery,* 40**,** 1106-1110.

MERCAT, A., NGUYEN, J. & DAUTZENBERG, B. (1991) An Outbreak Of Pneumococcal Pneumonia In 2 Mens Shelters. *Chest,* 99**,** 147-151.

MERRELL, D. S., BUTLER, S. M. & QADRI, F. (2002) Host induced epidemic spread of the cholera bacterium. *Nature,* 417**,** 642-45.

METZKER, M. L. (2005) Emerging technologies in DNA sequencing *Genome Research,* 15**,** 1767-1776.

MEYER, J. (1920) Types of pneumococci in the throats of one hundred normal persons. *Journal of the American Medical Association,* 75**,** 1268.

MEYEROVITCH, J., SHOHET, I. & RUBINSTEIN, E. (1985) Analysis of thirty-seven cases of pleural empyema. *European Journal of Clinical Microbiology,* 4**,** 337-9.

MIERNYK, K., RUDOLPH, K., REASONOVER, A., HARKER-JONES, M., BOYD-HUMMEL, K., ZULZ, T., BRUCE, M. & HENNESSY, T. (2008) P2-005 Molecular characterization of pneumococcal isolates of serotypes 3,7F,8, and 12F causing increased invasive disease after introduction of the 7-valent pneumococcal conjugate vaccine in Alaska. *6th International Symposium on Pneumococci and Pneumococcal Diseases.* Reykjavik, Iceland.

MIKASA, K., SAWAKI, M., KITA, E., HAMADA, K., TERAMOTO, S., SAKAMOTO, M., MAEDA, K., KONISHI, M. & NARITA, N. (1997) Significant survival benefit to patients with advanced non small-cell lung cancer from treatment with clarithromycin. *Chemotherapy,* 43**,** 288-296.

MILLAR, M. R., BROWN, N. M., TOBIN, G. W., MURPHY, P. J., WINDSOR, A. C. M. & SPELLER, D. C. E. (1994) Outbreak Of Infection With Penicillin-Resistant *Streptococcus pneumoniae* In A Hospital For The Elderly. *Journal of Hospital Infection,* 27**,** 99-104.

MIZRACHI-NEBENZAHL, Y., PORAT, N., LIFSHITZ, S., NOVICK, S., LEVI, A., LING, E., LIRON, O., MORDECHAI, S., SAHU, R. K. & DAGAN, R. (2004) Virulence of *Streptococcus pneumoniae* may be determined independently of capsular polysaccharide. *FEMS Microbiology Letters,* 233**,** 147-152.

MOGDASY, M. C., CAMOU, T., FAJARDO, C. & HORTAL, M. (1992) Colonizing and invasive strains of *Streptococcus pneumoniae* in Uruguayan children: type distribution and patterns of antibiotic resistance. *Pediatric Infectious Disease Journal,* 11**,** 648-652.

MOINE, P., VERCKEN, J. B., CHEVRET, S. & GAJDOS, P. (1995) Severe community-acquired pneumococcal pneumonia. The French Study Group of Community-Acquired Pneumonia in ICU. *Scandinavian Journal of Infectious Diseases,* 27**,** 201-6.

MOLIN, S. & TOLKER-NIELSEN, T. (2003) Gene transfer occurs with enhanced efficiency in biofilms and induces enhanced stabilisation of the biofilm structure. *Current Opinion in Biotechnology,* 14**,** 255-261.

MOLLERACH, M., REGUEIRA, M., BONOFIGLIO, L., CALLEJO, R., PACE, J., DI FABIO, J. L., HOLLINGSHEAD, S. & BRILES, D. (2004) Invasive *Streptococcus pneumoniae* isolates from Argentinian children: serotypes, families of pneumococcal surface protein A (PspA) and genetic diversity. *Epidemiology and Infection,* 132**,** 177-184.

MORGAN, A. D., BUIRSKI G., CALDER M.A., WIGHTMAN A.J., LEITCH A.G. & D.C., F. (1983) Streptococcus pneumoniae serotypes causing pneumonia in Edinburgh and Nottingham. *The Lancet,* 321**,** 68-69.

MOSCHIONI, M., DONATI, C., MUZZI, A., MASIGNANI, V., CENSINI, S., HANAGE, W. P., BISHOP, C. J., REIS, J. N., NORMARK, S., HENRIQUES NORMARK, B., COVACCI, A., RAPPUOLI, R. & BAROCCHI, M. A. (2008) *Streptococcus pneumoniae* contains 3 *rlrA* Pilus Variants That Are Clonally Related. *Journal of Infectious Diseases,* 197**,** 1-9.

MOSCOSO, M., LOPEZ, E., GARCIA, E. & LOPEZ, R. (2005) Implications of physiological studies based on genomic sequences: *Streptococcus pneumoniae* TIGR4 synthesizes a functional LytC lysozyme. *Journal of Bacteriology,* 187**,** 6238-6241.

MULLER-GRAF, C. D., WHATMORE, A. M., KING, S. J., TRZCINSKI, K., PICKERILL, A. P., DOHERTY, N., PAUL, J., GRIFFITHS, D., CROOK, D. & DOWSON, C. G. (1999) Population biology of *Streptococcus pneumoniae* isolated from oropharyngeal carriage and invasive disease. *Microbiology,* 145**,** 3283-93.

MULLER, P. Y., JANOVJAK, H., MISEREZ, A. R. & DOBBIE, Z. (2002) Processing of Gene Expression Data Generated by Quantitative Real-Time RT-PCR. *BioTechniques,* 32**,** 1372-1379.

MUNFORD, R. S. (2005) *Sepsis, Severe Sepsis and Septic Shock in Mandell, Douglas and Bennett's Principles and Practice of Infectious Diseases,* Philadelphia, Elsevier Churchill Livingstone.

MUNOZ, R., COFFEY, T. J., DANIELS, M., DOWSON, C. G., LAIBLE, G., CASAL, J., HACKENBECK, R., JACOBS, M., MUSSER, J. M., SPRATT, B. G. & TOMASZ, A. (1991) Intercontinental spread of a multiresistant clone of serotype 23F *Streptococcus pneumoniae*. *Journal of Infectious Diseases,* 164**,** 302-6.

MUSHER, D. M. (2004) *A Pathogenetic Categorization of Clinical Syndromes Caused by Streptococcus pneumoniae in The Pneumococcus,* Washington, DC., ASM Press.

MUSHER, D. M. (2005) *Streptococcus pneumoniae in Mandell, Douglas and Bennett's Principles and Practice of Infectious Diseases,* Philadelphia, Elsevier Churchill Livingstone.

MUSHER, D. M., ALEXANDRAKI, I., GRAVISS, E. A., YANBEIY, N., INDERIAS, A., LUZMIN, A. & PHAN, H. M. (2000) Bacteremic and Nonbacteremic Pneumococcal Pneumonia: A Prospective Study. *Medicine (Baltimore).* 79**,** 210-221.

MUSHER, D. M., DOWELL, M. E., SHORTRIDGE, V. D., FLAMM, R. K., JORGENSEN, J. H., LE MAGUERES, P. & KRAUSE, K. L. (2002) Emergence of

macrolide resistance during treatment of pneumococcal pneumonia. *New England Journal of Medicine,* 346**,** 630-1.

MUSHER, D. M., GROOVER, J. E., REICHLER, M. R., RIEDO, F. X., SCHWARTZ, B., WATSON, D. A., BAUGHN, R. E. & BREIMAN, R. F. (1997) Emergence of Antibody to Capsular Polysaccharides of *Streptococcus pneumoniae* during Outbreaks of Pneumonia: Association with Nasopharyngeal Colonization. *Clinical Infectious Diseases,* 24**,** 441-6.

NEELEMAN, C., KLAASSEN, C. H., DE VALK, H. A., DE RUITER, M. T. & MOUTON, J. W. (2004a) Amplified fragment length polymorphism fingerprinting is an effective technique to distinguish *Streptococcus pneumoniae* from other Streptococci and an efficient alternative to pulsed-field gel electrophoresis for molecular typing of pneumococci. *Journal of Clinical Microbiology,* 42**,** 369-71.

NEELEMAN, C., KLAASSEN, C. H. W., KLOMBERG, D. M., DE VALK, H. A. & MOUTON, J. W. (2004b) Pneumolysin Is a key Factor in Misidentification of Macrolide-Resistant *Streptococcus pneumoniae* and Is a Putative Virulence Factor of *S. mitis* and Other Streptococci. *Journal of Clinical Microbiology,* 42**,** 4355-4357.

NEUFELD, F. (1902) Ueber die Agglutination der Pneumokokken und űber die Theorieen der Agglutination. *Zeitschrift fur Hygiene und Infektionskrankheiten,* 40**,** 54-72.

NEUFELD, F. & ETINGER-TULCZYNSKA, R. (1932) Nasale Pneumokokkeninfectionen and Pneumokokkenkeimtrager im Tierversuch. *Zeitschrift fur Hygiene und Infektionskrankheiten,* 112**,** 492-526.

NEUFELD, F. & HANDEL, L. (1909) Ueber Herstellung und Prufung von antipneumokokken-Serum und uber die Aussichten einer spezifischen Behandlung der Pneumonie. *Ztschr. f. Immunitatsforsch. u. exper. Therap., Originale,* 3**,** 159.

NEUFELD, F. & HANDEL, L. (1910) Weitere Untersuchungen uber Pneumokokken Heilsera III. Mitteilung. Uber Vorkommen und Bedeutung atypischer Varietaten des Pneumokokkus. *Arbeiten aus dem Kaiserlichen Gsundheitsamte,* 34**,** 293-304.

NG, W. L., KAZMIERCZAK, K. M., ROBERTSON, G. T., GILMOUR, R. & WINKLER, M. E. (2003) Transcriptional regulation and signature patterns revealed by microarray analyses of *Streptococcus pneumoniae* R6 challenged with sublethal concentrations of translation inhibitors. *Journal of Bacteriology,* 185**,** 359-370.

NOVAK, R., CAUWELS, A., CHARPENTIER, E. & TUOMANEN, E. (1999) Identification of a *Streptococcus pneumoniae* gene locus encoding proteins of an ABC phosphate transporter and a two-component regulatory system. *Journal of Bacteriology,* 181**,** 1126-33.

NOWAK, S. J. G. (1939) Empyema Thoracis: An Analytical Study of 500 Cases With General Remarks. *Medical Clinics of North America,* 23**,** 1355-1369.

NUNES, S., SA-LEAO, R., PEREIRA, L. C. & DE LENCASTRE, H. (2007) Emergence of a serotype 1 *Streptococcus pneumoniae* lineage colonising healthy children in Portugal in the seven-valent conjugate vaccination era. *Clinical Microbiology and Infection,* 14**,** 82-84.

NUORTI, J. P., BUTLER, J. C., CRUTCHER, J. M., GUEVARA, R., WELCH, D., HOLDER, P. & ELLIOTT, J. A. (1998) An outbreak of multidrug-resistant pneumococcal pneumonia and bacteremia among unvaccinated nursing home residents. *New England Journal Of Medicine,* 338**,** 1861-1868.

O'BRIEN, K. L. (2008) S01-KS1 Global Pneumococcal Disease Epidemiology. *6th International Symposium on Pneumococci and Pneumococcal Diseases.* Reykjavik, Iceland.

O'BRIEN, K. L., BRONSDON, M. A., DAGAN, R., YAGUPSKY, P., J., J., ELLIOT, J., WHITNEY, C. G., YANG, Y. H., ROBINSON, L. G. E., B., S. & CARLONE, G. M. (2001) Evaluation of a Medium (STGG) for Transport and Optimal recovery of *Streptococcus pneumoniae* from Nasopharyngeal Secretions Collected during Field Studies. *Journal of Clinical Microbiology,* 39**,** 1021-1024.

O'BRIEN, K. L. & DAGAN, R. (2003) The potential indirect effect of conjugate pneumococcal vaccines. *Vaccine,* 21**,** 1815-1825.

O'BRIEN, K. L. & NOHYNEK, H. (2003) Report from a WHO Working Group: standard method for detecting upper respiratory carriage of *Streptococcus pneumoniae*. *Pediatric Infectious Disease Journal,* 22**,** e1-11.

O'BRIEN, K. L., WALTERS, M. I., SELLMAN, J., QUINLISK, P., REGNERY, H., SCHWARTZ, B. & DOWELL, S. F. (2000) Severe pneumococcal pneumonia in previously healthy children: the role of preceding influenza infection. *Clinical Infectious Diseases,* 30**,** 784-9.

OBANDO, I., ARROYO, L. A., SANCHEZ-TATAY, D., MORENO, D., HAUSDORFF, W. P. & BRUEGGEMANN, A. B. (2006) Molecular typing of pneumococci causing parapneumonic empyema in Spanish children using multilocus sequence typing directly on pleural fluid samples. *Pediatric Infectious Disease Journal,* 25**,** 962-963.

OBANDO, I., MUNOZ-ALMAGRO, C., ARROYO, L. A., TARRAGO, D., SANCHEZ-TATAY, D., MORENO-PEREZ, D., DHILLON, S. S., ESTEVA, C., HERNANDEZ-BOU, S., GARCIA-GARCIA, J. J., HAUSDORFF, W. P. & BRUEGGEMANN, A. B. (2008) Pediatric Parapenumonic Empyema, Spain. *Emerging Infectious Diseases,* 14**,** 1390-1397.

OBARO, S. K. & ADEGBOLA, R. A. (2002) The pneumococcus: carriage, disease and conjugate vaccines. *Journal of Medical Microbiology,* 51**,** 98-104.

OBERT, C., SUBLETT, J., KAUSHAL, D., HINOJOSA, E., BARTON, T., TUOMANEN, E. I. & ORIHUELA, C. J. (2006) Identification of a candidate *Streptococcus pneumoniae* core genome and regions of diversity correlated with invasive pneumococcal disease. *Infection and Immunity,* 74**,** 4766-4777.

OBERT, C. A., GAO, G., SUBLETT, J., TUOMANEN, E. I. & ORIHUELA, C. J. (2007) Assessment of molecular typing methods to determine invasiveness and to differentiate clones of *Streptococcus pneumoniae*. *Infection, Genetics and Evolution,* 7**,** 708-716.

OBREGON, V., GARCIA, J. L., GARCIA, E., LOPEZ, R. & GARCIA, P. (2003) Genome Organization and Molecular Analysis of the Temperate Bacteriophage MM1 of *Streptococcus pneumoniae. Journal of Bacteriology,* 185**,** 2362-2368.

OGBURN, E. & DOWSON, C. (2004) Gene alterations occurring in *Streptococcus pneumoniae* as a result of biofilm growth. *4th International Symposium on Pneumococci and Pneumococcal Diseases.* Helsinki, Finland.

OGGIONI, M. R., IANNELLI, F., RICCI, S., CHIAVOLINI, D., PARIGI, R., TRAPPETTI, C., CLAVERYS, J. P. & POZZI, G. (2004) Antibacterial Activity of a Competence-Stimulating Peptide in Experimental Sepsis Caused by *Streptococcus pneumoniae*. *Antimicrobial Agents and Chemotherapy,* 48**,** 4725-4732.

OGGIONI, M. R., MEMMI, G., MAGGI, T., CHIAVOLINI, D., LANNELLI, F. & POZZI, G. (2003) Pneumococcal zinc metalloproteinase ZmpC cleaves human matrix metalloproteinase 9 and is a virulence factor in experimental pneumonia. *Molecular Microbiology,* 49**,** 795-805.

OGGIONI, M. R., MULAS, L., TRAPPETTI, C., DAVIDSEN, T., TETTELIN, H., HAKENBECK, R. & POZZI, G. (2008) A single phosphotransferase system accounts for almost all phenotypic carbohydrate fermentation diversity between D39, TIGR4 and the type 19F strain G54. *6th International Symposium on Pneumococci and Pneumococcal Diseases.* Reykjavik, Iceland.

OGUNNIYI, A. D., LEMESSURIER, K. S., MAHDI, L. K. & PATON, J. C. (2007) B-O3 Pneumococcal virulence-associated genes are differentially expressed during progression of disease. *Europneumo 2007. 8th European Meeting on the Molecular Biology of the Pneumococcus.* Oeiras, Portugal.

OKITSU, N., KAIEDA, S., YANO, H., NAKANO, R., HOSAKA, Y., OKAMOTO, R., KOBAYASHI, T. & INOUE, M. (2005) Characterization of *ermB* Gene Transposition by Tn*1545* and Tn*917* in Macrolide-Resistant *Streptococcus pneumoniae* Isolates. *Journal of Clinical Microbiology,* 43**,** 168-173.

ORIHUELA, C. J., GAO, G., FRANCIS, K. P., YU, J. & TUOMANEN, E. I. (2004a) Tissue-specific contributions of pneumococcal virulence factors to pathogenesis. *Journal of Infectious Diseases,* 190**,** 1661-9.

ORIHUELA, C. J., GAO, G., MCGEE, M., YU, J., FRANCIS, K. P. & TUOMANEN, E. (2003) Organ-specific models of *Streptococcus pneumoniae* disease. *Scandinavian Journal of Infectious Diseases,* 35**,** 647-52.

ORIHUELA, C. J., RADIN, J. N., SUBLETT, J. E., GAO, G. L., KAUSHAL, D. & TUOMANEN, E. I. (2004b) Microarray analysis of pneumococcal gene expression during invasive disease. *Infection and Immunity,* 72**,** 5582-5596.

ORIYO, N., SAM, N., GILLESPIE, S. H. & CHARALAMBOUS, B. M. (2006) Longitudinal carriage study of *Streptococcus pneumoniae* serotype or group in Tanzanian children. *5th International Symposium on Pneumococci and Pneumococcal Diseases.* Alice Springs, Australia.

OSTERGAARD, C., BRANDT, C., KONRADSEN, H. B. & SAMUELSSON, S. (2004) Differences in survival, brain damage, and cerebrospinal fluid cytokine kinetics due

to meningitis caused by 3 different *Streptococcus pneumoniae* serotypes: Evaluation in humans and in 2 experimental models. *Journal of Infectious Diseases,* 190**,** 1212-1220.

OVERWEG, K., BOGAERT, D., SLUIJTER, C., YOTHER, J., DANKERT, J., DE GROOT, R. & HERMANS, P. W. M. (2000a) Genetic relatedness within serotypes of penicillin-susceptible *Streptococcus pneumoniae* isolates. *Journal of Clinical Microbiology,* 38**,** 4548-4553.

OVERWEG, K., PERICONE, C. D., VERHOEF, G. G., WEISER, J. N., MEIRING, H. D., DE JONG, A. P., DE GROOT, R. & HERMANS, P. W. (2000b) Differential protein expression in phenotypic variants of *Streptococcus pneumoniae*. *Infection and Immunity,* 68**,** 4604-10.

PAGANINI, H., GUINAZU, J. R., HERNANDEZ, C., LOPARDO, H., GONZALEZ, F. & BERBERIAN, G. (2001) Comparative analysis of outcome and clinical features in children with pleural empyema caused by penicillin-nonsusceptible and penicillin-susceptible *Streptococcus pneumoniae*. *International Journal of Infectious Diseases,* 5**,** 86-8.

PAHO (2007) *Bolivia in Heath in the Americas (PAHO Scientific and Technical Publication No.622),* Washington DC.

PAI, R., MOORE, M. R., PILISHVILI, T., GERTZ, R. E., WHITNEY, C. G. & BEALL, B. (2005) Postvaccine Genetic Structure of *Streptococcus pneumoniae* Serotype 19A from Children in the United States. *Journal of Infectious Diseases,* 192**,** 1988-95.

PALACIOS, G., QUAN, P. L., JABADO, O. J., CONLAN, S., HIRSCHBERG, D. L., LIU, Y., ZHAI, J., RENWICK, N., HUI, J., HEGYI, H., GROLLA, A., STRONG, J. E., TOWNER, J. S., GEISBERG, T. W., P.B., J., BUCHEN-OMOND, C., ELLERBROK, H., SANCHEZ-SECO, M. P., LUSSIER, Y., FORMENTY, P., NICHOL, S. T., FELDMANN, H., BRIESE, T. & LIPKIN, W. I. (2007) Panmicrobial Oligonucleotide Array for Diagnosis of Infectious Diseases. *Emerging Infectious Diseases,* 13**,** 73-80.

PANDYA, U., ALLEN, C. A., WATSON, D. A. & NIESEL, D. W. (2005) Global profiling of *Streptococcus pneumoniae* gene expression at different growth temperatures. *Gene,* 360**,** 45-54.

PANNUCCI, J., CAI, H., PARDINGTON, P. E., WILLIAMS, E., OKINAKA, R. T., KUSKE, C. R. & CARY, R. B. (2004) Virulence signatures: microarray-based approaches to discovery and analysis. *Biosensors and Bioelectronics,* 20**,** 706-718.

PARK, I. H., PRITCHARD, D. G., CARTEE, R. T., BRANDAO, A., BRANDILEONE, M. C. & NAHM, M. H. (2007) Discovery of a New Capsular Serotype 6C within Serogroup 6 of *Streptococcus pneumoniae*. *Journal of Clinical Microbiology,* 45**,** 1225-1233.

PARK, J. H. & CHICKERING, H. T. (1919) Type I Pneumococcus Lobar Pneumonia Among Porto Rican Laborers at Camp Jackson, South Carolina. *Journal of the American Medical Association,* 73**,** 183-186.

PARK, W. H., WILLIAMS, A. W., OPPENHEIMER, A., BOLDUAN, C., BERRY, J. L., ASSERSEN, M. A., LOWDEN, M. & VAN GIESEN, I. (1905) A Study of Pneumococci: A comparison between the pneumococci found in the throat secretions of healthy persons living in both city and country and those obtained from pneumonic exudates and diseased mucous membranes. *The Journal of Experimental Medicine,* 7**,** 403-419.

PARKHILL, J. (2002) Annotation of Microbial Genomes. *Methods in Microbiology,* 33**,** 3-26.

PARRY, C. M., DIEP, T. S., WAIN, J., HOA, N. T. T., GAINSBOROUGH, M., NGA, D., DAVIES, C., PHU, N. H., HIEN, T. T., WHITE, N. J. & FARRAR, J. J. (2000) Nasal Carriage in Vietnamese Children of *Streptococcus pneumoniae* Resistant to Multiple Antimicrobial Agents. *Antimicrobial Agents and Chemotherapy,* 44**,** 484-488.

PASTEUR, L. (1881) Note sur la maladie nouvelle provoquée par la salive d'un enfant mort de la rage. *Comptes Rendus de la Societe de Biologie,* 92**,** 159-165.

PATERSON, G. K., BLUE, C. E. & MITCHELL, T. J. (2006) Role of two-component systems in the virulence of *Streptococcus pneumoniae*. *Journal of Medical Microbiology,* 55**,** 355-363.

PATERSON, G. K. & MITCHELL, T. J. (2006) The role of *Streptococcus pneumoniae* sortase A in colonisation and pathogenesis. *Microbes and Infection,* 8**,** 145-53.

PEPLIES, J., GLOCKNER, F. O. & AMANN, R. (2003) Optimization strategies for DNA microarray based detection of bacteria with 16S rRNA targeting oligonucleotide probes. *Applied and Environmental Microbiology,* 69**,** 1397-1407.

PERICONE, C. D., BAE, D., SHCHEPETOV, M., MCCOOL, T. & WEISER, J. N. (2002) Short-sequence tandem and nontandem DNA repeats and endogenous hydrogen peroxide production contribute to genetic instability of *Streptococcus pneumoniae*. *Journal of Bacteriology,* 184**,** 4392-9.

PERICONE, C. D., PARK, S., IMLAY, J. A. & WEISER, J. N. (2003) Factors contributing to hydrogen peroxide resistance in *Streptococcus pneumoniae* include pyruvate oxidase (SpxB) and avoidance of the toxic effects of the Fenton reaction. *Journal of Bacteriology,* 185**,** 6815-25.

PETERSON, S., CLINE, R. T., TETTELIN, H., SHAROV, V. & MORRISON, D. A. (2000) Gene expression analysis of the *Streptococcus pneumoniae* competence regulons by use of DNA microarrays. *Journal of Bacteriology,* 182**,** 6192-202.

PETERSON, S. N., SUNG, C. K., CLINE, R., DESAI, B. V., SNESRUD, E. C., LUO, P., WALLING, J., LI, H. Y., MINTZ, M., TSEGAYE, G., BURR, P. C., DO, Y., AHN, S., GILBERT, J., FLEISCHMANN, R. D. & MORRISON, D. A. (2004) Identification of competence pheromone responsive genes in *Streptococcus pneumoniae* by use of DNA microarrays. *Molecular Microbiology,* 51**,** 1051-1070.

PETTIGREW, M. M. & FENNIE, K. P. (2005) Genomic Subtraction Followed by Dot Blot Screening of *Streptococcus pneumoniae* Clinical and Carriage Isolates Identifies Genetic Differences Associated with Strains That Cause Otitis Media. *Infection and Immunity,* 73**,** 2805-2811.

PETTIGREW, M. M., FENNIE, K. P., YORK, M. P., DANIELS, J. & GHAFFAR, F. (2006) Variation in the Presence of Neuraminidase Genes among *Streptococcus pneumoniae* Isolates with Identical Sequence Types. *Infection and Immunity,* 74**,** 3360-3365.

PICHICHERO, M. E. & CASEY, J. R. (2007) Emergence of a Multiresistant Serotype 19A Pneumococcal Strain Not Included in the 7-valent Conjugate Vaccine as an Otopathogen in Children. *Journal of the American Medical Association,* 298**,** 1772-1778.

PLAYFOR, S. D., SMYTH, A. R. & STEWART, R. J. (1997) Increase in incidence of childhood empyema. *Thorax,* 52**,** 932.

PLOTON, C., FREYDIERE, A. M., BENITO, Y., BENDRIDI, N., MAZZOCCHI, C., BELLON, G. & VANDENESCH, F. (2006) *Streptococcus pneumoniae* thoracic empyema in children: rapid diagnosis by using the Binax NOW immunochromatographic membrane test in pleural fluids. *Pathologie Biologie,* 54**,** 498-501.

POLISSI, A., PONTIGGIA, A., FEGER, G., ALTIERI, M., MOTTL, H., FERRARI, L. & SIMON, D. (1998) Large-Scale Identification of Virulence Genes from *Streptococcus pneumoniae*. *Infection and Immunity,* 66**,** 5620-5629.

PORAT, N., GREENBERG, D., GIVON-LAVI, N., SHUVAL, D. S., TREFLER, R., SEGEV, O., HANAGE, W. P. & DAGAN, R. (2006) The Important Role of Nontypeable *Streptococcus pneumoniae* International Clones in Acute Conjunctivitis. *Journal of Infectious Diseases,* 194**,** 689-96.

PORCEL, J. M., RUIZ-GONZALEZ, A., FALGUERA, M., NOGUES, A., GALINDO, C., CARRATALA, J. & ESQUERDA, A. (2007) Contribution of a pleural antigen assay (Binax NOW) to the diagnosis of pneumococcal pneumonia. *Chest,* 131**,** 1442-1447.

POULSEN, K., REINHOLDT, J., JESPERSGAARD, C., BOYE, K., BROWN, T. A., HAUGE, M. & KILIAN, M. (1998) A comprehensive genetic study of streptococcal immunoglobulin A1 proteases: Evidence for recombination within and between species. *Infection and Immunity,* 66**,** 181-190.

POULTER, M. D., DEVILLE, J. G., CHERRY, J. D., SHAPIRO, A. M. & MCAVIN, J. C. (2005) Real-time fluorescence polymerase chain reaction (PCR) identification of *Streptococcus pneumoniae* from pleural fluid and tissue. *Scandinavian Journal of Infectious Diseases,* 37**,** 391-392.

POZZI, G., MASALA, L., IANNELLI, F., MANGANELLI, R., HAVARSTEIN, L. S., PICCOLI, L., SIMON, D. & MORRISON, D. A. (1996) Competence for genetic transformation in encapsulated strains of *Streptococcus pneumoniae*: two allelic variants of the peptide pheromone. *Journal of Bacteriology,* 178**,** 6087-90.

PROULX, J. F., DERY, S., JETTE, L. P., ISMAEL, J., LIBMAN, M. & DE WALS, P. (2002) Pneumonia epidemic caused by a virulent strain of *Streptococcus pneumoniae* serotype 1 in Nunavik, Quebec. *Canada Communicable Disease Report,* 28**,** 129-31.

PRUDHOMME, M., ATTAIECH, L., SANCHEZ, G., MARTIN, B. & CLAVERYS, J. P. (2006) Antibiotic stress induces genetic transformability in the human pathogen *Streptococcus pneumoniae. Science,* 313**,** 89-92.

QUAGLIARELLO, A. B., PARRY, C. M., HIEN, T. T. & FARRAR, J. J. (2003) Factors Associated with Carriage of Penicillin-resistant *Streptococcus pneumoniae* among Vietnamese Children: A Rural-urban Divide. *Journal of Health Population and Nutrition,* 21**,** 316-324.

QUICK, R. E., HOGE, C. W., HAMILTON, D. J., WHITNEY, C. J., BORGES, M. & KOBAYASHI, J. M. (1993) Underutilization Of Pneumococcal Vaccine In Nursing-Homes In Washington-State - Report Of A Serotype-Specific Outbreak And A Survey. *American Journal Of Medicine,* 94**,** 149-152.

RADETSKY, M. S., ISTRE, G. R., JOHANSEN, T. L., PARMELEE, S. W., LAUER, B. A., WIESENTHAL, A. M. & GLODE, M. P. (1981) Multiply resistant pneumococcus causing meningitis: its epidemiology within a day-care centre. *Lancet,* 2**,** 771-3.

RAHAV, G., TOLEDANO, Y., ENGELHARD, D., SIMHON, A., MOSES, A. E., SACKS, T. & SHAPIRO, M. (1997) Invasive pneumococcal infections. A comparison between adults and children. *Medicine,* 76**,** 295-303.

RAMIREZ, M., SEVERINA, E. & TOMASZ, A. (1999) A high incidence of prophage carriage among natural isolates of *Streptococcus pneumoniae*. *Journal of Bacteriology,* 181**,** 3618-3625.

RAMIREZ, M. & TOMASZ, A. (1999) Acquisition of new capsular genes among clinical isolates of antibiotic-resistant *Streptococcus pneumoniae*. *Microbial Drug Resistance,* 5**,** 241-6.

RAMPHUL, N., EASTHAM, K. M., FREEMAN, R., ELTRINGHAM, G., KEARNS, A. M., LEEMING, J. P., HASAN, A., HAMILTON, L. J. R. & SPENCER, D. A. (2006) Cavitatory lung disease complicating empyema in children. *Pediatric Pulmonology,* 41**,** 750-753.

REES, J. H. M. & SPENCER, D. A. (1997) Increase in incidence of childhood empyema in West Midlands, UK. *Lancet,* 349**,** 402.

REGELMANN, W. E. (2005) A pain in the Ear: What Has the 7-Valent Conjugated Pneumococcal Vaccine Done to Reduce the Incidence of Acute Otitis Media. *Clinical Infectious Diseases,* 40**,** 1745-7.

REICHMANN, P. & HAKENBECK, R. (2000) Allelic variation in a peptide-inducible two-component system of *Streptococcus pneumoniae*. *FEMS Microbiology Letters,* 190**,** 231-6.

REIS, J. N., CORDEIRO, S. M., COPPOLA, S. J., SALGADO, K., CARVALHO, M. G., TEIXEIRA, L. M., THOMPSON, T. A., FACKLAM, R. R., REIS, M. G. & KO, A. I. (2002) Population-based survey of antimicrobial susceptibility and serotype distribution of *Streptococcus pneumoniae* from meningitis patients in Salvador, Brazil. *Journal of Clinical Microbiology,* 40**,** 275-7.

REIS, J. N., PALMA, T., RIBEIRO, G. S., PINHEIRO, R. M., RIBEIRO, C. T., CORDEIRO, S. M., DA SILVA FILHO, H. P., MOSCHIONI, M., THOMPSON, T. A., SPRATT, B., RILEY, L. W., BAROCCHI, M. A., REIS, M. G. & KO, A. I. (2008) Transmission of *Streptococcus pneumoniae* in an urban slum community. *Journal of Infection,* 57**,** 204-213.

REN, B., SZALAI, A. J., THOMAS, O., HOLLINGSHEAD, S. K. & BRILES, D. E. (2003) Both family 1 and family 2 PspA proteins can inhibit complement deposition and confer virulence to a capsular serotype 3 strain of *Streptococcus pneumoniae*. *Infection and Immunity,* 71**,** 75-85.

RETSEMA, J. & FU, W. (2001) Macrolides: structures and microbial targets. *International Journal of Antimicrobial Agents,* 18**,** S3-S10.

REY, L. C., WOLF, B., MOREIRA, J. L., MILATOVIC, D., VERHOEF, J. & FARHAT, C. K. (2002a) Antimicrobial susceptibility and serotypes of nasopharyngeal Streptococcus pneumoniae in children with pneumonia and in children attending day-care centres in Fortaleza, Brazil. *International Journal of Antimicrobial Agents,* 20**,** 86-92.

REY, L. C., WOLF, B., MOREIRA, J. L. B., VERHOEF, J. & FARHAT, C. K. (2002b) Nasopharyngeal isolates of *S. pneumoniae* from healthy carriers and children with pneumonia: colonization rates and antimicrobial susceptibility. *Jornal de Pediatria,* 78**,** 105-12.

RICHMOND, C. S., GLASNER, J. D., MAU, R., JIN, H. & BLATTNER, F. R. (1999) Genome wide expression profiling in *Escherichia coli* K-12. *Nucleic Acids Research,* 27**,** 3821-3835.

RIMINI, R., JANSSON, B., FEGER, G., ROBERTS, T. C., DE FRANCESCO, M., GOZZI, A., FAGGIONI, F., DOMENICI, E., WALLACE, D. M., FRANDSEN, N. & POLISSI, A. (2000) Global analysis of transcription kinetics during competence development in *Streptococcus pneumoniae* using high density DNA arrays. *Molecular Microbiology,* 36**,** 1279-1292.

RING, A., WEISER, J. N. & TUOMANEN, E. (1998) Pneumococcal Trafficking across the Blood-Brain Barrier. Molecular Analysis of a Novel Bidirectional Pathway. *Journal of Clinical Investigation,* 102**,** 347-360.

RIOS, A. M., DE LA HOZ, F., LEAL, A. L., CASTILLO, O. & CASTANEDA, E. (1999) The impact of antimicrobial resistance and *Streptococcus pneumoniae* serotype distribution on the mortality of children under 5 years of age with invasive disease. *Pan American Journal of Public Health,* 5**,** 69-76.

RIVERA-OLIVERO, I. A., BOGAERT, D., BELLO, T., DEL NOGAL, B., SLUIJTER, M., HERMANS, P. W. M. & DE WAARD, J. H. (2007) Pneumococcal Carriage among Indigenous Warao Children in Venezuela: Serotypes, Susceptibility Patterns, and Molecular Epidemiology. *Clinical Infectious Diseases,* 45**,** 1427-34.

ROBINSON, D. A., BRILES, D. E., CRAIN, M. J. & HOLLINGSHEAD, S. K. (2002) Evolution and virulence of serogroup 6 pneumococci on a global scale. *Journal of Bacteriology,* 184**,** 6367-6375.

ROBINSON, D. A., EDWARDS, K. M., WAITES, K. B., BRILES, D. E., CRAIN, M. J. & HOLLINGSHEAD, S. K. (2001) Clones of *Streptococcus pneumoniae* isolated from nasopharyngeal carriage and invasive disease in young children in central Tennessee. *Journal of Infectious Diseases,* 183**,** 1501-1507.

ROBINSON, D. A., HOLLINGSHEAD, S. K., MUSSER, J. M., PARKINSON, A. J., BRILES, D. E. & CRAIN, M. J. (1998) The IS1167 insertion sequence is a phylogenetically informative marker among isolates of serotype 6B *Streptococcus pneumoniae*. *Journal of Molecular Evolution,* 47**,** 222-229.

ROBINSON, D. A., TURNER, J. S., FACKLAM, R. R., PARKINSON, A. J., BREIMAN, R. F., GRATTEN, M., STEINHOFF, M. C., HOLLINGSHEAD, S. K., BRILES, D. E. & CRAIN, M. J. (1999) Molecular characterization of a globally distributed lineage of serotype 12F *Streptococcus pneumoniae* causing invasive disease. *Journal of Infectious Diseases,* 179**,** 414-22.

ROCHE, H., REN, B., MCDANIEL, L. S., HAKANSSON, A. & BRILES, D. E. (2003) Relative roles of genetic background and variation in PspA in the ability of antibodies to PspA to protect against capsular type 3 and 4 strains of *Streptococcus pneumoniae*. *Infection and Immunity,* 71**,** 4498-4505.

RODRIQUEZ-BARRADAS, M. C., THARAPEL, R. A., GROOVER, J. E., GIRON, K. P., LACKE, C. E., HOUSTON, E. D., HAMILL, R. J., STEINHOFF, M. C. & MUSHER, D. M. (1997) Colonization by *Streptococcus pneumoniae* among human immunodeficiency virus-infected adults: Prevalence of antibiotic resistance, impact of immunization, and characterisation by polymerase chain reaction with BOX primers of isolates from persistent *S. pneumoniae* carriers. *Journal of Infectious Diseases,* 175**,** 590-597.

ROSEN, H., NADKARNI, V., THEROUX, M., PADMAN, R. & KLEIN, J. I. (1993) Intrapleural streptokinase as adjunctive treatment for persistent empyema in pediatric patients. *Chest,* 103**,** 1190.

ROSSI, A., CORSO, A., PACE, J., REGUEIRA, M. & TOMASZ, A. (1998) Penicillin-resistant *Streptococcus pneumoniae* in Argentina: Frequent occurrence of an internationally spread serotype 14 clone. *Microbial Drug Resistance,* 4**,** 225-231.

ROXBURGH, C. S. D. & YOUNGSON, G. G. (2007) Childhood Empyema in North-East Scotland over the Past 15 Years. *Scottish Medical Journal,* 52**,** 25-27.

ROXBURGH, C. S. D., YOUNGSON, G. G., TOWNEND, J. A. & TURNER, S. W. (2008) Trends in pneumonia and empyema in Scottish children in the past 25 years. *Archives of Disease In Childhood,* 93**,** 316-318.

RUCKINGER, S., VON KRIES, R., REINERT, R. R., VAN DER LINDEN, M. & SIEDLER, A. (2008) Childhood invasive pneumococcal disease in Germany between 1997 and 2003: Variability in incidence and serotype distribution in absence of general pneumococcal conjugate vaccination. *Vaccine,* 26**,** 3984-3986.

RYAN, D., RAHIMI, M., LUND, J., MEHTA, J. & PARVIZ, B. A. (2007) Toward nanoscale genome sequencing. *Trends in Biotechnology,* 25**,** 385-389.

SA-LEAO, R., TOMASZ, A., SANCHES, I. S. & DE LENCASTRE, H. (2002) Pilot study of the genetic diversity of the pneumococcal nasopharyngeal flora among children attending day care centers. *Journal of Clinical Microbiology,* 40**,** 3577-3585.

SABINA, J., DOVER, N., TEMPLETON, L. J., SMULSKI, D. R., SOLL, D. & LAROSSA, R. A. (2003) Interfering with Different Steps of Protein Synthesis Explored by Transcriptional Profiling of *Escherichia coli* K-12. *Journal of Bacteriology,* 185**,** 6158-6170.

SAGLANI, S., HARRIS, K. A., WALLIS, C. & HARTLEY, J. C. (2005) Empyema: the use of broad range 16S rDNA PCR for pathogen detection. *Archives Of Disease In Childhood,* 90**,** 70-73.

SAHN, S. A. (2007) Diagnosis and Management of Parapneumonic Effusions and Empyema. *Clinical Infectious Diseases,* 45**,** 1480-1486.

SAILER, J., HALL M.W., WILSON, R. L. & MCCOY, C. (1919) A Study of Pneumococcus Carriers. *Archives of Internal Medicine,* 24**,** 600-610.

SANDGREN, A., ALBIGER, B., ORIHUELA, C. J., TUOMANEN, E., NORMARK, S. & HENRIQUES NORMARK, B. (2005) Virulence in Mice of Pneumococcal Clonal Types with Known Invasive Disease Potential in Humans *Journal of Infectious Diseases,* 192**,** 791-800.

SANDGREN, A., SJOSTROM, K., OLSSON-LILJEQUIST, B. O., CHRISTENSSON, B., SAMUELSSON, A., KRONVALL, G. & NORMARK, B. H. (2004) Effect of clonal and serotype-specific properties on the invasive capacity of *Streptococcus pneumoniae*. *Journal of Infectious Diseases,* 189**,** 785-796.

SANTAGATI, M., IANNELLI, F., OGGIONI, M. R., STEFANI, S. & POZZI, G. (2000) Characterization of a genetic element carrying the macrolide efflux gene *mef(A)* in *Streptococcus pneumoniae*. *Antimicrobial Agents and Chemotherapy,* 44**,** 2585-2587.

SANTANA-HERNANDEZ, D. J. (2006) The need of a programme for the prevention of hearing impairment in Beni state, Bolivia. *Community Ear and Hearing Health,* 3**,** 17-32.

SCHOTTMULLER, H. (1903) Die Artunterscheidung der fur den Menschen pathogenen Streptokokken durch Blutagar. *Munchen. med. Wchnschr,* 50**,** 908.

SCHRODER, M. C. & COOPER, G. (1930) An epidemic of colds, bronchitis and pneumonia due to Type V pneumococci. *Journal of Infectious Diseases,* 46**,** 384-392.

SCHULTZ, K. D., FAN, L. L., PINSKY, J., OCHOA, L., SMITH, E. O., KAPLAN, S. L. & BRANDT, M. L. (2004) The changing face of pleural empyemas in children: Epidemiology and management. *Pediatrics,* 113**,** 1735-1740.

SCOTT, J. A. G. (2007) The preventable burden of pneumococcal disease in the developing world. *Vaccine,* 25**,** 2398-2405.

SCOTT, J. A. G., HALL, A. J., DAGAN, R., DIXON, J. M. S., EYKYN, S. J., FENOLL, A., HORTAL, M., JETTE, L. P., JORGENSEN, J. H., LAMOTHE, F., LATORRE,

C., MACFARLANE, J. T., SHLAES, D. M., SMART, L. E. & TAUNAY, A. (1996) Serogroup-specific epidemiology of *Streptococcus pneumoniae*: Associations with age, sex, and geography in 7,000 episodes of invasive disease. *Clinical Infectious Diseases,* 22**,** 973-981.

SELANDER, R. K., CAUGANT, D. A., OCHMAN, H., MUSSER, J. M., GILMOUR, M. N. & WHITTHAM, T. S. (1986) Methods of multilocus enzyme electrophoresis for bacterial genetics and systematics. *Applied and Environmental Microbiology,* 51**,** 873-884.

SERRUTO, D., ADU-BOBIE, J., CAPECCHI, B., RAPPUOLI, R., PIZZA, M. & MASIGNANI, V. (2004) Biotechnology and vaccines: application of functional genomics to Neisseria meningitides and other bacterial pathogens. *Journal of Biotechnology,* 113**,** 15-32.

SESSEGOLO, J. F., LEVIN, A. S., LEVY, C. E., ASENSI, M., FACKLAM, R. R. & TEIXEIRA, L. M. (1994) Distribution of serotypes and antimicrobial resistance of *Streptococcus pneumoniae* strains isolated in Brazil from 1988 to 1992. *Journal of Clinical Microbiology,* 32**,** 906-11.

SHAPIRO, E. D. & AUSTRIAN, R. (1994) Serotypes responsible for invasive *Streptococcus pneumoniae* infections among children in Connecticut. *Journal of Infectious Diseases,* 169**,** 212-214.

SHAW, K. J., MILLER, N., LIU, X., LERNER, D., WAN, J., BITTNER, A. & MORROW, B. J. (2003) Comparison of the changes in global gene expression of *Escherichia coli* induced by four bactericidal agents. *Journal of Molecular Microbiology and Biotechnology,* 5**,** 105-122.

SHAW, K. J. & MORROW, B. J. (2003) Transcriptional profiling and drug discovery. *Current opinion in Pharmacology,* 3**,** 508-512.

SHEN, K., ANTALIS, P., GLADITZ, J., SAYEED, S., AHMED, A., YU, S. J., HAYES, J., JOHNSON, S., DICE, B., DOPICO, R., KEEFE, R., JANTO, B., CHONG, W., GOODWIN, J., WADOWSKY, R. M., ERDOS, G., POST, C. J., EHRLICH, G. D. & HU, F. Z. (2005) Identification, Distribution, and Expression of Novel Genes in 10 Clinical Isolates of Nontypeable *Haemophilus influenzae*. *Infection and Immunity,* 73**,** 3479-3491.

SHEN, K., GLADITZ, J., ANTALIS, P., DICE, B., JANTO, B., KEEFE, R., HAYES, J., AHMED, A., DOPICO, R., EHRLICH, N., JOCZ, J., KROPP, L., YU, S., NISTICO, L., GREENBERG, D. P., BARBADORA, K., PRESTON, R. A., POST, J. C., EHRLICH, G. D. & HU, F. Z. (2006a) Characterization, Distribution, and Expression of Novel Genes among Eight Clinical Isolates of *Streptococcus pneumoniae*. *Infection and Immunity,* 74**,** 321-330.

SHEN, Y. H., HWANG, K. P. & NIU, C. K. (2006b) Complicated parapneumonic effusion and empyema in children. *Journal of Microbiology, Immunology and Infection,* 39**,** 483-8.

SHEPPARD, C., GUIVER, M., SPENCER, D., CLIFF, D., GEORGE, R. & KAYE, P. (2008) P1-092 Non-culture surveillance of *Streptococcus pneumoniae* serotypes causing pneumococcal empyema in UK children, post conjugate vaccine

introduction. *6th International Symposium on Pneumococci and Pneumococcal Diseases.* Reykavik, Iceland.

SHINKAI, M., PARK, C. S. & RUBIN, B. K. (2005) Immunomodulatory Effects of Macrolide Antibiotics. *Clinical Pulmonary Medicine,* 12**,** 341-348.

SHOUVAL, D. S., GREENBERG, D., GIVON-LAVI, N., PORAT, N. & DAGAN, R. (2006) Site-specific disease potential of individual *Streptococcus pneumoniae* serotypes in pediatric invasive disease, acute otitis media and acute conjunctivitis. *Pediatric Infectious Disease Journal,* 25**,** 602-607.

SIBOLD, C., WANG, J. F., HENRICHSEN, J. & HACKENBECK, R. (1992) Genetic relationships of penicillin-susceptible and -resistant *Streptococcus pneumoniae* strains isolated on different continents. *Infection and Immunity,* 60**,** 4119-4126.

SILVA, N. A., MCCLUSKEY, J., JEFFERIES, J. M. C., HINDS, J., SMITH, A., CLARKE, S. C., MITCHELL, T. J. & PATERSON, G. K. (2006) Genomic diversity between strains of the same serotype and multilocus sequence type among pneumococcal clinical isolates. *Infection and Immunity,* 74**,** 3513-3518.

SINGLETON, R., HOLMAN, R., YORITA, K., WENGER, J., ZULZ, T., BOYD-HUMMEL, K., STEINER, C., PAISANO, E. & CHEEK, J. (2008) P1-004 Pneumococcal Empyema and Pleural Effusion among Alaska Native Children less than 10 years of age. *6th International Symposium on Pneumococci and Pneumococcal Diseases.* Reykjavik, Iceland.

SINGLETON, R. V., HENNESSY, T. W., BULKOW, L. R., HAMMITT, L. L., ZULZ, T., HURLBURT, D. A., BUTLER, J. C., RUDOLPH, K. & PARKINSON, A. (2007) Invasive Pneumococcal Disease Caused by Nonvaccine Serotypes Among Alaska Native Children With High Levels of 7-Valent Pneumococcal Conjugate Vaccine Coverage. *Journal of the American Medical Association,* 297**,** 1784-1792.

SINIGAR, H. (1903) The Variability in Virulence of the Pneumococcus. *Lancet,* i**,** 169-170.

SJOSTROM, K., BLOMBERG, C., FERNEBRO, J., DAGERHAMN, J., MORFELDT, E., BAROCCHI, M. A., BROWALL, S., MOSCHIONI, M., ANDERSSON, M., HENRIQUES, F., ALBIGER, B., RAPPUOLI, R., NORMARK, S. & HENRIQUES-NORMARK, B. (2007) Clonal success of piliated penicillin nonsusceptible pneumococci. *Proceedings of the National Academy of Sciences of the United States of America,* 104**,** 12907-12.

SJOSTROM, K., SPINDLER, C., ORTQVIST, A., KALIN, M., SANDGREN, A., KUHLMANN-BERENZON, S. & HENRIQUES NORMARK, B. (2006) Clonal and Capsular Types Decide Whether Pneumococci Will Act as a Primary or Opportunistic Pathogen. *Clinical Infectious Diseases,* 42**,** 451-9.

SKOCZYNSKA, A. & HRYNIEWICZ, W. (2003) Genetic relatedness, antibiotic susceptibility, and serotype distribution of *Streptococcus pneumoniae* responsible for meningitis in Poland, 1997-2001. *Microbial Drug Resistance,* 9**,** 175-182.

SLOAS, M. M., BARRETT, F. F., CHESNEY, P. J., ENGLISH, B. K., HILL, B. C., TENOVER, F. C. & LEGGIADRO, R. J. (1992) Cephalosporin treatment failure in

penicillin- and cephalosporin-resistant *Streptococcus pneumoniae* meningitis. *Pediatric Infectious Disease Journal,* 11**,** 662-666.

SLUIJTER, M., FADEN, H., DE GROOT, R., LEMMENS, N., GOESSENS, W. H., VAN BELKUM, A. & HERMANS, P. W. (1998) Molecular characterization of pneumococcal nasopharynx isolates collected from children during their first 2 years of life. *Journal of Clinical Microbiology,* 36**,** 2248-53.

SMART, L. E. (1986) Serotyping of *Streptococcus pneumoniae* by coagglutination. *Journal of Clinical Pathology,* 39**,** 328-331.

SMART, L. E. (1987) The Serology of Pneumococcal Infection. A Study of the Laboratory Diagnosis of Pneumococcal Infection and the Distribution of Pneumococcal Types. Glasgow, University of Glasgow.

SMEALL, J. T. (1931) Types of Pneumococci in Relation to Disease. *British Medical Journal,* i**,** 661-663.

SMILLIE, W. G., WARNOCK, G. H. & WHITE, H. J. (1938) A Study of a Type I Pneumococcus Epidemic at the State Hospital at Worcester, Mass. *American Journal of Public Health,* 28**,** 293-302.

SMITH, A. J., JEFFERIES, J., CLARKE, S. C., DOWSON, C., EDWARDS, G. F. S. & MITCHELL, T. J. (2006) Distribution of epidemic antibiotic-resistant pneumococcal clones in Scottish pneumococcal isolates analysed by multilocus sequence typing. *Microbiology,* 152**,** 361-365.

SMITH, T., LEHMANN, D., MONTGOMERY, J., GRATTEN, M., RILEY, I. D. & ALPERS, M. P. (1993) Acquisition and invasiveness of different serotypes of *Streptococcus pneumoniae* in young children. *Epidemiology and Infection,* 111**,** 27-39.

SMOOT, J. C., BARBIAN, K. D., VAN GOMPEL, J. J., SMOOT, L. M., CHAUSSEE, M. S., SYLVA, G. L., STURDEVANT, D. E., RICKLEFS, S. M., PORCELLA, S. F., PARKINS, L. D., BERES, S. B., CAMPBELL, D. S., SMITH, T. M., ZHANG, Q., KAPUR, V., DALY, J. A., VEASY, L. G. & MUSSER, J. M. (2002) Genome sequence and comparative microarray analysis of serotype M18 group A *Streptococcus* strains associated with acute rheumatic fever outbreaks. *Proceedings of the National Academy of Sciences of the United States of America,* 99**,** 4668-73.

SPEER, A., KERSTAN, H., BOTTCHER, T., GERBER, J., SIEMER, A., ZYSK, G., MITCHELL, T. J., EIFFERT, H. & NAU, R. (2003) Reduced Release of Pneumolysin by *Streptococcus pneumoniae* In Vitro and In vivo after Treatment with Nonbacteriolytic Antibiotics in Comparison to Ceftriaxone. *Antimicrobial Agents and Chemotherapy,* 47**,** 2649-2654.

SPENCE, R. P., WRIGHT, V., ALA-ALDEEN, D. A. A., TURNER, D. P., WOOLDRIDGE, K. G. & JAMES, R. (2008) Validation of Virulence and Epidemiology DNA Microarray for Identification and Characterization of *Staphylococcus aureus* Isolates. *Journal of Clinical Microbiology,* 46**,** 1620-1627.

SPRATT, B. G. & GREENWOOD, B. M. (2000) Prevention of pneumococcal disease by vaccination: does serotype replacement matter? *Lancet,* 356**,** 1210-1211.

STABLER, R. A., GERDING, D. N., SONGER, J. G., DRUDY, D., BRAZIER, J. S., TRINH, H. T., WITNEY, A. A., HINDS, J. & WREN, B. W. (2006) Comparative Phylogenomics of *Clostridium difficile* Reveals Clade Specificity and Microevolution of Hypervirulent Strains. *Journal of Bacteriology,* 188**,** 7297-7305.

STANDISH, A. J., STROEHER, U. H. & PATON, J. C. (2007) The pneumococcal two-component signal transduction system RR/HK06 regulates CbpA and PspA by two distinct mechanisms. *Journal of Bacteriology,* 189**,** 5591-600.

STANEK, R. J. & MUFSON, M. A. (1999) A 20 Year Epidemiological Study of Pneumococcal Meningitis. *Clinical Infectious Diseases,* 28**,** 1265-72.

STERNBERG, G. M. (1881) A fatal form of septicaemia in the rabbit, produced by the subcutaneous injection of human saliva. An experimental research. *National Board of Health Bulletin,* 2**,** 781-783.

STEWART, M. J. & FLINT, H. L. (1919) A Case of Rapidly Fatal Ulcerative Endocarditis. *Lancet,* ii**,** 1114.

STILLMAN, E. G. (1916) A Contribution to the Epidemiology of Lobar Pneumonia. *Journal of Experimental Medicine,* 24**,** 651-670.

STILLMAN, E. G. (1917) Further Studies on the Epidemiology of Lobar Pneumonia. *Journal of Experimental Medicine,* 26**,** 513-535.

STOVROFF, M., TEAGUE, G., HEISS, K. F., PARKER, P. & RICKETTS, R. R. (1995) Thoracoscopy In The Management Of Pediatric Empyema. *Journal of Pediatric Surgery,* 30**,** 1211-1215.

STRAUS, A. L. & HAMBURGER, M. (1966) Pneumococcal Endocarditis in the Penicillin Era. *Archives Of Internal Medicine,* 118**,** 190-198.

STROM, A. (1932) An Epidemic of Croupous Pneumonia Caused by Pneumococcus Type I. *Journal of Infectious Diseases,* 50**,** 430-436.

SUBLETT, J., HAAS, W. & TUOMANEN, E. (2004) Mutational analysis of the *vex123-vncRS* locus in *S.pneumoniae* T4. *4th International Symposium on Pneumococci and Pneumococcal Diseases,*. Helsinki, Finland.

SUBRAMANIAN, D., SANDOE, J. A. T., KEER, V. & WILCOX, M. H. (2003) Rapid spread of penicillin-resistant *Streptococcus pneumoniae* among high-risk hospital inpatients and the role of molecular typing in outbreak confirmation. *Journal of Hospital Infection,* 54**,** 99-103.

SULLIVAN, C. B., DIGGLE, M. A. & CLARKE, S. C. (2005) Multilocus Sequence Typing Data Analysis in Clinical Microbiology and Public Health. *Molecular Biotechnology,* 29**,** 245-254.

SWIATLO, E., CRAIN, M. J., MCDANIEL, L. S., BROOKS-WALTER, A., COFFEY, T. J., SPRATT, B. G., MORRISON, D. A. & BRILES, D. E. (1996) DNA polymorphisms and variant penicillin-binding proteins as evidence that relatively penicillin-resistant pneumococci in western Canada are clonally related. *Journal of Infectious Diseases,* 174**,** 884-8.

SWIDEREK, H., CLAUS, H., FROSCH, M. & VOGEL, U. (2005) Evaluation of custom-made DNA microarrays for mulilocus sequence typing of *Neisseria meningitidis*. *International Journal of Medical Microbiology,* 295**,** 39-45.

SYROGIANNOPOULOS, G. A., BOGAERT, D., GRIVEA, I. N., BERATIS, N. G., DE GROOT, R. & HERMANS, P. W. M. (2001) Molecular epidemiology of penicillin-susceptible, multidrug-resistant serotype 6B pneumococci isolated from children in Greece. *Journal of Clinical Microbiology,* 39**,** 581-585.

TAKALA, A. K., VUOPIO-VARKILA, J., TARKKA, E., LEINONEN, M. & MUSSER, J. M. (1996) Subtyping of Common Pediatric Pneumococcal Serotypes from Invasive Disease and Pharyngeal Carriage in Finland. *Journal of Infectious Diseases,* 173**,** 128-35.

TAKAMATSU, D., BENSING, B. A. & SULLAM, P. M. (2004) Genes in the accessory *sec* locus of *Streptococcus gordonii* have three functionally distinct effects on the expression of the platelet-binding protein GspB. *Molecular Microbiology,* 52**,** 189-203.

TAMAYO, M., SA-LEAO, R., SANCHES, I. S., CASTANEDA, E. & DE LENCASTRE, H. (1999) Dissemination of a Chloramphenicol- and Tetracycline-Resistant but Penicillin-Susceptible Invasive Clone of Serotype 5 *Streptococcus pneumoniae* in Colombia. *Journal of Clinical Microbiology,* 37**,** 2337-2342.

TAN, C. G., OSTRAWSKI, S. & BRESNITZ, E. A. (2003) A preventable outbreak of pneumococcal pneumonia among unvaccinated nursing home residents in New Jersey during 2001. *Infection Control And Hospital Epidemiology,* 24**,** 848-852.

TAN, T. Q., MASON, E. O., BARSON, W. J., WALD, E. R., SCHUTZE, G. E., BRADLEY, J. S., GIVNER, L. B., YOGEV, R., KIM, K. S. & KAPLAN, S. L. (2000) Clinical and Pleural Fluid Parameters in Children with Complicated Pneumococcal Pneumonia. *Pediatric Research,* 47**,** 278A.

TAN, T. Q., MASON, E. O., WALD, E. R., BARSON, E. J., SCHUTZE, G. E., BRADLEY, J. S., GIVNER, L. B., YOGEV, R., KIM, K. S. & KAPLAN, S. L. (2002) Clinical characteristics of children with complicated pneumonia caused by *Streptococcus pneumoniae*. *Pediatrics,* 110**,** 1-6.

TANG, C. M. & MOXON, E. R. (2001) The Impact of Microbial Genomics On Antimicrobial Drug Development. *Annual Review of Genomics and Human Genetics,* 2**,** 259-269.

TARYLE, D. A., POTTS, D. E. & SAHN, S. A. (1978) The Incidence and Clinical Correlates of Parapneumonic Effusions in Pneumococcal Pneumonia. *Chest,* 74**,** 170-173.

TATEDA, K., STANDIFORD, T. J., PECHERE, J. C. & YAMAGUCHI, K. (2004) Regulatory effects of macrolides on bacterial virulence: potential role as quorum-sensing inhibitors. *Current Pharmaceutical Design,* 2004**,** 3055-3065.

TAUNAY, A. E., AUSTRIAN, R., LANDGRAF, I. M., VIEIRA, M. F. & MELLES, C. E. (1990) Serotypes of *Streptococcus pneumoniae* isolated from cerebrospinal fluid in 1977-1988 in Sao Paulo City, Brazil. *Revista do Instituto de Medicina Tropical de Sao Paulo,* 32**,** 11-5.

TETTELIN, H., NELSON, K. E., PAULSEN, I. T., EISEN, J. A., READ, T. D., PETERSON, S., HEIDELBERG, J., DEBOY, R. T., HAFT, D. H., DODSON, R. J., DURKIN, A. S., GWINN, M., KOLONAY, J. F., NELSON, W. C., PETERSON, J. D., UMAYAM, L. A., WHITE, O., SALZBERG, S. L., LEWIS, M. R., RADUNE, D., HOLTZAPPLE, E., KHOURI, H., WOLF, A. M., UTTERBACK, T. R., HANSEN, C. L., MCDONALD, L. A., FELDBLYUM, T. V., ANGIUOLI, S., DICKINSON, T., HICKEY, E. K., HOLT, I. E., LOFTUS, B. J., YANG, F., SMITH, H. O., VENTER, J. C., DOUGHERTY, B. A., MORRISON, D. A., HOLLINGSHEAD, S. K. & FRASER, C. M. (2001) Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*. *Science,* 293**,** 498-506.

THAKKER, B., GORMAN, L. J., MCNAMEE S., JONES, B. L. & CURRAN, E. T. (1998) An Outbreak of *Streptococcus pneumoniae* serotype 1 in a geriatric ward. *38th Interscience Conference on Antimicrobial Agents and Chemotherapy.* San Diego, CA.

THOMPSON, L. J., MERRELL, D. S., NEILAN, B. A., MITCHELL, H., LEE, A. & FALKOW, S. (2003) Gene Expression Profiling of *Helicobacter pylori* Reveals a Growth-Phase-Dependent Switch in Virulence Gene Expression. *Infection and Immunity,* 71**,** 2643-2655.

THOMSON, A. H., HULL, J., KUMAR, M. R., WALLIS, C. & BALFOUR LYNN, I. M. (2002) Randomised trial of intrapleural urokinase in the treatment of childhood empyema. *Thorax,* 57**,** 343-347.

THUMERELLE, C., SANTOS, C., MORILLON, S., BOTT, L., POUESSEL, G. & DESCHILDRE, A. (2005) Risk factors of parapneumonic effusions in children. *Archives De Pediatrie,* 12**,** 827-829.

TOMASZ, A., CORSO, A., SEVERINA, E. P., ECHANIZ-AVILES, G., BRANDILEONE, M. C. D., CAMOU, T., CASTANEDA, E., FIGUEROA, O., ROSSI, A. & DI FABIO, J. L. (1998) Molecular epidemiologic characterization of penicillin-resistant *Streptococcus pneumoniae* invasive pediatric isolates recovered in six Latin-American countries: An overview. *Microbial Drug Resistance,* 4**,** 195-207.

TREGNAGHI, M., CEBALLOS, A., RUTTIMANN, R., HAUSDORFF, W., USSHER, J. & YUDOUSKI, S. (2006a) PO3.61 Evolution of pneumococcal serotype distribution over a 5 year period in 1-35 month old children in the city of Cordoba, Argentina. *5th International Symposium on Pneumococci and Pneumococcal Diseases.* Alice Springs, Australia.

TREGNAGHI, M., CEBALLOS, A., RUTTIMANN, R., USSHER, J., TREGNAGHI, P., PEETERS, P. & HAUSDORFF, W. P. (2006b) Active epidemiologic surveillance of pneumonia and invasive pneumococcal disease in ambulatory and hospitalized infants in Cordoba, Argentina. *Pediatric Infectious Disease Journal,* 25**,** 370-2.

TRZCINSKI, K., THOMPSON, C. M. & LIPSITCH, M. (2004) Single-step capsular transformation and acquisition of penicillin resistance in *Streptococcus pneumoniae*. *Journal of Bacteriology,* 186**,** 3447-3452.

TSAI, W. C. & STANDIFORD, T. J. (2004) Immunomodulatory Effects of Macrolides in the Lung: Lessons from *In-vitro* and *In-vivo* Models. *Current Pharmaceutical Design,* 10**,** 3081-3093.

TSOLIA, M. N., STAMOS, G., IOANNIDOU, S., TREFLER, R., FOUSTOUKOU, M., KAFETZIS, D. & PORAT, N. (2002) Genetic relatedness of resistant and multiresistant *Streptococcus pneumoniae* strains, recovered in the Athens area, to international clones. *Microbial Drug Resistance,* 8**,** 219-226.

TSUI, W. H. W., YIM, G., WANG, H. H., MCCLURE, J. E., SURETTE, M. G. & DAVIES, J. (2004) Dual Effects of MLS Antibiotics: Transcriptional Modulation and Interactions on the Ribosome. *Chemistry and Biology,* 11**,** 1307-1316.

ULIJASZ, A. T., ANDES, D. R., GLASNER, J. D. & WEISBLUM, B. (2004a) Regulation of iron transport in *Streptococcus pneumoniae* by RitR, an orphan response regulator. *Journal of Bacteriology,* 186**,** 8123-8136.

ULIJASZ, A. T., ANDES, D. R., GLASNER, J. D. & WEISBLUM, B. (2004b) Tissue-specific regulation of iron uptake and oxidative stress responses in *Streptococcus pneumoniae* by two-component signal transduction. *4th International Symposium on Pneumococci and Pneumococcal Diseases.* Helsinki, Finland

URQUHART, A. L. (1921) Serological and Morphological Characteristics of the Pneumococcus. *The Lancet,* ii**,** 1313-1317.

URWIN, G., YUAN, M. F., HALL, L. M. C., BROWN, K. & EFSTRATIOU, A. (1996) Pneumococcal meningitis in the North East Thames Region UK: epidemiology and molecular analysis of isolates. *Epidemiology and Infection,* 117**,** 95-102.

VAN BELKUM, A., SLUIJTER, M., DE GROOT, R., VERBRUGH, H. A. & HERMANS, P. W. (1996) Novel BOX repeat PCR assay for high-resolution typing of *Streptococcus pneumoniae* strains. *Journal of Clinical Microbiology,* 34**,** 1176-1179.

VAN DE BEEK, D., DE GANS, J., MCINTYRE, P. & PRASAD, K. (2004) Steroids in adults with acute bacterial meningitis: a systematic review. *The Lancet Infectious Diseases,* 4**,** 139-143.

VAN STEENBERGEN, T. J. M., COLLOMS, S. D., HERMANS, P. W. M., DE GRAAF, J. & PLASTERK, R. H. A. (1995) Genomic DNA fingerprinting by restriction fragment end labeling (RFEL). *Proceedings of the National Academy of Sciences of the United States of America,* 92**,** 5572-5576.

VEENHOVEN, R. H., BOGAERT, D. & UITERWAAL, C. (2003) Effect of pneumococcal conjugate vaccine followed by polsaccharide pneumococcal vaccine on recurrent acute otitis media: a randomised study. *Lancet,* 361**,** 2189-2195.

VELA, M. C., FONSECA, N., DI FABIO, J. L. & CASTANEDA, E. (2001) Presence of international multiresistant clones of *Streptococcus pneumoniae* in Colombia. *Microbial Drug Resistance,* 7**,** 153-164.

VERNET, G., JAY, C., RODRIGUE, M. & TROESCH, A. (2004) Species differentiation and antibiotic susceptibility testing with DNA microarrays. *Journal of Applied Microbiology,* 96**,** 59-68.

VESCINA, C. M., REGUEIRA, M., GONZALEZ AVALA, S. E., GATTI, B., CECCHINI, D. M. & AGOSTI, M. R. (2006) PO3.20 *Streptococcus pneumoniae* serotypes isolated from Argentinian children with invasive disease. *5th International Symposium on Pneumococci and Pneumococcal Diseases.* Alice Springs, Australia.

WAITE, R. D., STRUTHERS, J. K. & DOWSON, C. G. (2001) Spontaneous sequence duplication within an open reading frame of the pneumococcal type 3 capsule locus causes high-frequency phase variation. *Molecular Microbiology,* 42**,** 1223-32.

WALTMAN, W. D., GRAY, B. M., SVANBORG, C., FACKLAM, R. & BRILES, D. E. (1991) Epidemiologic Studies of Group 9 Pneumococci in Terms of Protein Type and 9N versus 9V Capsular Type. *The Journal of Infectious Diseases,* 163**,** 812-818.

WANG, D. G., FAN, J. B. F., SIAO, C. J., BERNO, A., YOUNG, P., SAPOLSKY, R., GHANDOUR, G., PERKINS, N., WINCHESTER, E., SPENCER, J., KRUGLYAK, L., STEIN, L., HSIE, L., TOPALOGLOU, T., HUBBELL, E., ROBINSON, E., MITTMANN, M., MORRIS, M. S., SHEN, N., KILBURN, D., RIOUX, J., NUSBAUM, C., ROZEN, S., HUDSON, T. J., LIPSHULTZ, R., CHEE, M. & LANDER, E. S. (1998) Large scale identification, mapping and genotyping of single nucleotide polymorphisms in the human genome. *Science,* 280**,** 1077-1082.

WANG, Q., WANG, M., KONG, F., GILBERT, G. L., CAO, B., WANG, L. & FENG, L. (2007) Development of a DNA microarray to identify the *Streptococcus pneumoniae* serotypes contained in the 23-valent pneumococcal polysaccharide vaccine and closely related serotypes. *Journal of Microbiological Methods,* 68**,** 128-136.

WATANABE, H., ASOH, N., HOSHINO, K., WATANABE, K., OISHI, K., KOSITSAKULCHAI, W., SANCHAI, T., KUNSUIKMENGRAI, K., KAHINTAPONG, S., KHANTAWA, B., THARAVICHITKUL, P., SIRISANTHANA, T. & NAGATAKE, T. (2003) Antimicrobial Susceptibility and Serotype Distribution of *Streptococcus pneumoniae* and Molecular Characterization of Multidrug-Resistant Serotype 19F, 6B, and 23F Pneumococci in Northern Thailand. *Journal of Clinical Microbiology,* 41**,** 4178-4183.

WATERS, A. M., KERECUK, L., LUK, D., HAQ, M. R., FITZPATRICK, M. M., GILBERT, R. D., INWARD, C., JONES, C., PICHON, B., REID, C., SLACK, M. P. E., VAN'T HOFF, W., DILLON, M. J., TAYLOR, C. M. & TULLUS, K. (2007) Hemolytic uremic syndrome associated with invasive pneumococcal disease: The United Kingdom experience. *Journal of Pediatrics,* 151**,** 140-144.

WATTS, D. J. & STROGATZ, S. H. (1998) Collective dynamics of "small-world" networks. *Nature,* 393**,** 440-442.

WEATHERHOLTZ, R., MILLAR, E., MOULTON, L., PERILLA, M., REID, R., PARKINSON, A., SANTOSHAM, M. & O'BRIEN, K. (2008) S11-O3 No replacement invasive pneumococcal disease 11 years after introduction of PCV in a population at high risk for IPD: the Navajo experience. *6th International Symposium on Pneumococci and Pneumococcal Diseases.* Reykjavik, Iceland.

WEBER, J. R. (2004) *Pathogenesis of Pneumococcal Meningitis in The Pneumococcus,* Washington DC, ASM Press.

WEESE, W. C., SHINDLER, E. R., SMITH, I. M. & RABINOVICH, S. (1973) Empyema of the Thorax Then and Now. A Study of 122 Cases Over Four Decades. *Archives of Internal Medicine,* 131**,** 516-520.

WEISER, J. N., AUSTRIAN, R., SREENIVASAN, P. K. & MASURE, H. R. (1994) Phase Variation In Pneumococcal Opacity - Relationship Between Colonial Morphology And Nasopharyngeal Colonization. *Infection and Immunity,* 62**,** 2582-2589.

WEISER, J. N. & KAPOOR, M. (1999) Effect of intrastrain variation in the amount of capsular polysaccharide on genetic transformation of *Streptococcus pneumoniae*: implications for virulence studies of encapsulated strains. *Infection and Immunity,* 67**,** 3690-2.

WEISS, K., RESTIERI, C., GAUTHIER, R., LAVERDIERE, M., MCGEER, A., DAVIDSON, R. J., KILBURN, L., BAST, D. J., DE AZAVEDO, J. & LOW, D. E. (2001) A nosocomial outbreak of fluoroquinolone-resistant *Streptococcus pneumoniae*. *Clinical Infectious Diseases,* 33**,** 517-522.

WHATMORE, A. M., BARCUS, V. A. & DOWSON, C. G. (1999) Genetic diversity of the streptococcal competence (com) gene locus. *Journal of Bacteriology,* 181**,** 3144-3154.

WHATMORE, A. M. & DOWSON, C. G. (1999) The autolysin-encoding gene (lytA) of *Streptococcus pneumoniae* displays restricted allelic variation despite localized recombination events with genes of pneumococcal bacteriophage encoding cell wall lytic enzymes. *Infection and Immunity,* 67**,** 4551-4556.

WHATMORE, A. M., EFSTRATIOU, A., PICKERILL, A. P., BROUGHTON, K., WOODARD, G., STURGEON, D., GEORGE, R. & DOWSON, C. G. (2000) Genetic relationships between clinical isolates of *Streptococcus pneumoniae*, *Streptococcus oralis*, and *Streptococcus mitis*: Characterization of "atypical" pneumococci and organisms allied to *S. mitis* harboring *S. pneumoniae* virulence factor-encoding genes. *Infection and Immunity,* 68**,** 1374-1382.

WHITTLE, C. H. (1929) Further Observations on Pneumococcal Infections. *British Medical Journal***,** 898-900.

WHO (2007) Pneumococcal conjugate vaccine for childhood immunization - WHO position paper. *Weekly Epidemiological Record,* 82**,** 93-104.

WILSON, W. J., STROUT, C. L., DESANTIS, T. Z., STILWELL, J. L., CARRANO, A. V. & ANDERSEN, G. L. (2002) Sequence-specific identification of 18 pathogenic microorganisms using microarray technology. *Molecular and Cellular Probes,* 16**,** 119-127.

WITNEY, A. A., MARSDEN, G. L., HOLDEN, T. G. H., STABLER, R. A., HUSAIN, S. E., VASS, J. K., BUTCHER, P. D., HINDS, J. & LINDSAY, J. A. (2005) Design, Vaildation, and Application of a Seven-Strain *Staphylococcus aureus* PCR Product Microarray for Comparative Genomics. *Applied and Environmental Microbiology,* 71**,** 7504-7514.

WIZEMANN, T. M., HEINRICHS, J. H., ADAMOU, J. E., ERWIN, A. L., KUNSCH, C., CHOI, G. H., BARASH, S. C., ROSEN, C. A., MASURE, H. R., TUOMANEN, E., GAYLE, A., BREWAH, Y. A., WALSH, W., BARREN, P., LATHIGRA, R., HANSON, M., LANGERMANN, S., JOHNSON, S. & KOENIG, S. (2001) Use of a Whole Genome Approach To Identify Vaccine Molecules Affording Protection Against *Streptococcus pneumoniae* Infection. *Infection and Immunity,* 69**,** 1593-1598.

WODICKA, L., DONG, H., MITTMANN, M., HO, M. H. & LOCKHART, D. J. (1997) Genome-wide expression monitoring in *Saccharomyces cerevisiae*. *Nature Biotechnology,* 15**,** 1359-1367.

WOLF, B., REY, L. C., BRISSE, S., MOREIRA, L. B., MILATOVIC, D., FLEER, A., ROORD, J. J. & VERHOEF, J. (2000) Molecular epidemiology of penicillin-resistant *Streptococcus pneumoniae* colonizing children with community-acquired pneumonia and children attending day-care centres in Fortaleza, Brazil. *Journal of Antimicrobial Chemotherapy,* 46**,** 757-765.

WONG, K. S., LIN, T. Y., HUANG, Y. C., CHANG, L. Y. & LAI, S. H. (2005) Scoring system for empyema thoracis and help in management. *Indian Journal of Pediatrics,* 72**,** 1025-8.

WOOD, W. B. & SMITH, M. R. (1949) The inhibition of surface phagocytosis by the capsular "slime layer" of pneumococcus Type III. *Journal of Experimental Medicine,* 90**,** 85-96.

YARO, S. Y., LOURD, M., TRAORE, Y., NJANPOP-LAFOURCADE, B.-M., SAWADOGO, A., SANGARE, L., HIEN, A., OUEDRAOGO, M. S., SANOU, O., DU CHATELET, I. P., KOECK, J.-L. & GESSNER, B. D. (2006) Epidemiological and molecular characteristics of a highly lethal pneumococcal meningitis epidemic in Burkina Faso. *Clinical Infectious Diseases,* 43**,** 693-700.

YATSUNAMI, J., FUKUNO, Y., NAGATA, M., TOMINAGA, M., AOKI, S., TSURUTA, N., KAWASHIMA, M., TANIGUCHI, S. & HAYASHI, S. (1999) Antiangiogenic and antitumor effcts of 14-membered ring macrolides on mouse B16 melanoma cells. *Clinical and Experimental Metastasis,* 17**,** 359-365.

YE, R. W., WANG, T., BEDZYK, L. & CROKER, K. M. (2001) Applications of DNA microarrays in microbial systems. *Journal of Microbiological Methods,* 47**,** 257-272.

YIN, D. Z., FOX, B., LONETTO, M. L., ETBERTON, M. R., PAYNE, D. J., HOLMES, D. J., ROSENBERG, M. & JI, Y. D. (2004) Identification of antimicrobial targets using a comprehensive genomic approach. *Pharmacogenomics,* 5**,** 101-113.

ZEMLICKOVA, H., CRISOSTOMO, M. I., BRANDILEONE, M. C., CAMOU, T., CASTANEDA, E., CORSO, A., ECHANIZ-AVILES, G., PASZTOR, M. & TOMASZ, A. (2005) Serotypes and clonal types of penicillin-susceptible *Streptococcus pneumoniae* causing invasive disease in children in five Latin American countries. *Microbial Drug Resistance,* 11**,** 195-204.

ZHANG, W., JAYARAO, B. M. & KNABEL, S. J. (2004) Multi-virulence-locus sequence typing of *Listeria monocytogenes. Applied and Environmental Microbiology,* 70**,** 913-920.

ZHU, L. X., ZHANG, Z. W., WANG, C., YANG, H. W., JIANG, D., ZHANG, Q., MITCHELSON, K. & CHENG, J. (2007) Use of a DNA Microarray for Simultaneous Detection of Antibiotic Resistance Genes among Staphylococcal Clinical Isolates. *Journal of Clinical Microbiology,* 45**,** 3514-3521.

14 List of Publications

Abstracts:

Inverarity D, Diggle M, Edwards G, Mitchell T. An evaluation of media suitable for the transportation by air of *Streptococcus pneumoniae* isolates. 2007. Journal of Infection. **55**(**3**); e65-e66.

Oral Presentations:

"Pneumococcal carriage in Beni State, Bolivia: implications for pneumococcal conjugate vaccine introduction," was presented to the Royal Society of Tropical Medicine and Hygiene on 19th December 2007.

"Paediatric Empyema in Scotland," was presented to British Thoracic Society Satellite Meeting of Collaborators in Enhanced Surveillance of Pneumococcal Empyema in UK Children on 4th December 2007.

"Pneumococcal carriage up the Amazon: Bugs, Bolivia and boldly doing what no-one has done before," was presented at the Scottish Microbiology Association Autumn Meeting, Dunblane on 10th November 2007.

"Pneumococcal Virulence Factors," was presented at the European Respiratory Society Annual Congress, Stockholm, Sweden on 18th September 2007.

Poster Presentations:

6th International Symposium on Pneumococci and Pneumococcal Diseases, Reykjavik, Iceland 8th-12th June 2008

Inverarity D, Hinds J, Diggle M, Edwards G, Mitchell T. Microarray analysis of serotype 1 *Streptococcus pneumoniae* isolates from cases of bacteraemic pneumonia complicated by empyema and parapneumonic effusion in Scotland.

Inverarity D, Hinds J, Diggle M, Edwards G, MacConnachie A, Mitchell T. Microarray analysis of the genomic diversity of serotype 4 *Streptococcus pneumoniae* isolates (ST246) correlated to 6 clinical cases of bacteraemic pneumonia.

Inverarity D, Lamb K, Diggle M, Robertson C, Greenhalgh D, Mitchell T, Smith A, Jefferies J, McMenamin J, Clarke S, Edwards G. Death or survival from invasive pneumococcal disease in Scotland: Associations with serogroups and Multi Locus Sequence Types.

Lamb K, Diggle M, **Inverarity D**, Jefferies JM, Smith A, Edwards GFS, McMenamin J, Mitchell TJ, Clarke SC. Trends in serotypes and Multi Locus Sequence Types (MLST) among cases of Invasive Pneumococcal Disease (IPD) in Scotland.

Khandavilli S, **Inverarity D,** Homer K, Mitchell T, Brown J. Differential contribution of lipoprotein signal peptidase to *S. pneumoniae* phenotype and virulence in two serotype backgrounds.

18th European Congress of Clinical Microbiology and Infectious Diseases, Barcelona, Spain 19th-22nd April 2008

Inverarity DJ, Hinds J, Diggle M, Edwards G, Mitchell TJ. Molecular epidemiology and pathogenesis of pneumococcal cerebral abscesses in Scotland.

Medical Research Society, Academy of MedicalSciences and Royal College of Physicians, London, United Kingdom 28th February 2008

Inverarity DJ, Diggle MA, Hinds J, Gould K, Edwards G, Mitchell TJ. Genomic analysis of Serotype 1 *Streptococcus pneumoniae* causing complicated pneumonia.

8th European Meeting on the Molecular Biology of the Pneumococcus, Oeiras, Portugal 14-17th April 2007

Saville AM, Bentley SD, Diggle MA, **Inverarity DJ**, Mitchell AM, Mitchell TJ. Sequence variation in a putative glycosylation locus from *Streptococcus pneumoniae*. Poster A-P2.

Appendix 1: *Streptococcus pneumoniae* isolates used in microarray experiments

Age of Patient	Isolated From	Clindamycin	Clarithromycin	Ciprofloxacin	Ampicillin	Penicillin	Erythromycin	Oxacillin	Serotype	MLST	Strain Name
1	B	S	R	R	R	R	NT	R	14	ST9	00_1724
1	C	S	R	S	S	S	NT	S	14	ST9	01_5710
59	B	S	R	S	S	S	NT	S	18	ST9	02_1309
5	B	S	R	S	S	S	NT	S	14	ST9	02_2445
66	B	S	S	I	R	R	NT	R	19A	ST9	03_1051
87	B	S	R	S	S	S	NT	S	8	ST9	03_2105
72	B	S	S	S	S	S	NT	S	23F	ST9	04_1168
69	B	S	S	I	S	S	NT	S	19F	ST9	04_1548
65	B	S	R	S	S	S	NT	S	14	ST9	04_1870
12	B	S	S	I	S	S	NT	S	20	ST568	05_1271
70	B	S	S	S	S	S	NT	S	31	ST568	05_2166
85	B	S	S	I	R	R	NT	R	14	ST156	03_2662
34	B	S	S	S	S	S	NT	S	6B	ST156	05_1356
63	B	S	S	I	R	R	NT	R	9V	ST156	05_1821
30	B	S	S	S	S	S	NT	S	12B	ST218	05_2746
79	B	S	S	S	S	S	NT	S	12F	ST218	05_2565
U	U	S	S	I	S	S	NT	S	4	ST205	TIGR4
U	U	S	S	I	S	S	NT	S	1	ST227	INV104B
U	U	R	R	R	R	R	NT	R	19A	ST41	ATCC 700904
U	U	R	R	S	R	R	NT	R	14	ST67	ATCC BAA340
U	U	I	R	I	R	R	NT	R	6A	ST376	ATCC BAA659
U	U	S	R	I	R	R	NT	R	23F	ST37	ATCC 51916
U	U	R	R	S	R	R	NT	R	14	ST67	ATCC BAA340
U	U	S	R	I	R	R	NT	R	35B	ST377	ATCC BAA660
49	B	NT	NT	NT	NT	S	S	S	4	ST246	06_1803
64	B	NT	NT	NT	NT	S	S	S	4	ST246	06_1898
56	B	NT	NT	NT	NT	S	S	S	4	ST246	05_1109
42	B	NT	NT	NT	NT	S	S	S	4	ST246	04_2239
37	B	NT	NT	NT	NT	S	S	S	4	ST246	03_5339
53	B	NT	NT	NT	NT	S	S	S	4	ST246	04_1342
33	P	NT	NT	NT	NT	S	S	S	1	ST306	03_3038
51	B	NT	NT	NT	NT	S	S	S	1	ST227	04_2225
7	B	NT	NT	NT	NT	S	S	S	1	ST227	05_2739
70	B	NT	NT	NT	NT	S	S	S	1	ST227	06_1805
36	B	NT	NT	NT	NT	S	S	S	1	ST306	05_1519
26	B	NT	NT	NT	NT	S	S	S	1	ST306	03_2672
4	B	NT	NT	NT	NT	S	S	S	1	ST306	06_1370
60	B	NT	NT	NT	NT	S	R	S	3	ST180	00-3946
15	B	NT	NT	NT	NT	S	S	S	3	ST180	06_1705
0	B	NT	NT	NT	NT	S	R	S	3	ST180	02_1198
31	B	NT	NT	NT	NT	S	R	S	3	ST180	99_4038
9	NP	NT	NT	NT	S	S	S	S	3	ST180	07_2838
31	C	NT	NT	NT	NT	S	R	S	3	ST180	99_4039
U	U	S	R	NT	S	S	R	S	14	ST9	INV200
U	U	S	S	NT	R	R	S	R	23F	ST81	ATCC700669
9	NP	NT	NT	NT	S	S	S	S	9A	ST239	07_2839

Appendix 1

Details of isolates used on for microarray CGH experiments including source body fluid of isolate, age of source, antibiotic sensitivities, serotypes and MLST. R = Resistant, S = Sensitive, I = intermediate, NT = Not tested, B = Blood culture, C = CSF culture, P = Pleural Fluid culture, U = Unknown.

Appendix 2: Antibiotic Sensitivities and Typing of *Streptococcus pneumoniae* isolates from Bolivia.

Isolate	Serotype	MLST	Antibiotic Sensitivity					
			Penicillin	Erythromycin	Chloramphenicol	Co-trimoxazole	Tetracycline	Vancomycin
07-2801	17F	2973	0.125	0.16	3	0.125	0.19	0.75
07-2802	21	3852	0.023	0.047	3	0.125	0.125	0.75
07-2803	11A	280	0.016	0.032	3	0.125	0.19	0.75
07-2804	4	3767	0.016	0.032	3	0.125	0.25	0.75
07-2805	34	1902	0.023	0.032	2	0.19	0.5	0.75
07-2806	23A	2974	0.012	0.032	3	0.5	0.25	0.5
07-2807	19A	2975	0.032	0.032	4	0.125	0.25	0.75
07-2808	38	2974	0.003	0.047	3	0.5	0.25	0.75
07-2809	19F	2976	0.016	0.023	3	0.19	0.19	0.5
07-2810	23A	2974	0.016	0.047	3	0.75	0.25	0.75
07-2811	34	1902	0.016	0.023	1.5	0.094	0.19	0.75
07-2812	24F	3770	0.016	0.016	2	2	0.19	0.75
07-2813	14	387	0.19	0.016	2	0.064	0.125	0.75
07-2814	6A	3853	0.023	0.047	3	0.5	0.38	0.75
07-2815	34	1902	0.016	0.032	2	0.25	0.125	0.75
07-2816	16F	3768	0.012	0.016	2	0.25	0.19	0.75
07-2817	6A	1150	0.047	0.016	1.5	0.125	0.19	0.75
07-2818	8	404	0.012	0.047	2	0.125	0.25	0.38
07-2819	4	332	0.016	0.016	3	1.5	1.5	0.38
07-2820	38	3534	0.008	0.023	2	0.125	0.125	0.38
07-2821	23F	776	0.016	0.047	2	0.094	0.25	0.75
07-2822	6A	2440	0.016	0.023	2	0.38	0.19	0.75
07-2823	19A	2880	0.032	0.016	1	0.25	0.125	0.75
07-2824	34	1902	0.016	0.047	2	1.5	0.19	0.5
07-2825	4	3509	0.016	0.016	1.5	0.125	0.19	0.75
07-2826	7F	191	0.008	0.016	2	0.064	0.19	0.75
07-2827	13	923	0.016	0.047	1.5	0.125	0.19	0.5
07-2828	10A	3535	0.012	0.047	3	0.25	0.25	0.75
07-2829	23B	3536	0.125	0.016	1.5	0.5	0.25	0.5
07-2830	38	1902	0.016	0.032	2	0.125	0.25	0.38
07-2831	6B	4015	0.047	0.032	2	2	0.19	0.5
07-2832	6B	4015	0.023	0.032	1.5	0.38	0.25	0.5
07-2833	9A	3537	0.023	0.016	2	0.094	0.25	0.5
07-2834	12A	3538	0.012	0.032	2	0.38	0.25	0.75
07-2835	16F	3771	0.012	0.047	2	1	0.19	0.75
07-2836	10A	3539	0.016	0.032	1.5	0.125	0.25	0.75
07-2837	16A	4016	0.064	0.032	1.5	1.5	0.25	0.38
07-2838	3	180/1989	0.008	0.032	3	0.094	0.75	0.75
07-2839	9A	239	0.023	0.032	2	0.125	0.25	0.5
07-2840	7C	3769	0.012	0.032	1.5	0.125	0.19	0.75
07-2841	23F	3540	0.023	0.032	1.5	0.25	0.25	0.75
07-2843	13	784	0.023	0.032	2	0.125	0.25	0.75

Isolate	Serotype	MLST	Antibiotic Sensitivity					
			Penicillin	Erythromycin	Chloramphenicol	Co-trimoxazole	Tetracycline	Vancomycin
07-2844	6A	3854	0.012	0.047	3	0.125	0.25	0.75
07-2845	9A	3855	0.064	0.032	2	0.125	0.25	0.75
07-2846	6A	3856	0.008	0.032	2	0.38	0.19	0.75
07-2847	24F	3770	0.064	0.032	1.5	0.25	0.125	0.5
07-2848	21	3857	0.016	0.064	2	0.094	0.25	0.5
07-2849	17F	3267	0.064	0.047	2	0.094	0.125	0.75
07-2850	19A	3858	0.016	0.032	2	0.094	0.125	0.5
07-2851	10A	3429	0.016	0.047	3	0.125	0.25	0.75
07-2852	23F	3430	0.023	0.032	3	0.25	0.38	0.5
07-2853	35F	3431	0.016	0.023	2	0.094	16	0.5
07-2854	11A	3432	0.008	0.047	3	0.125	0.25	0.75

Appendix 2

Antibiotic sensitivities and Typing of 53 Nasopharyngeal pneumococcal isolates retrieved from school children in the Beni region of Bolivia. Minimum Inhibitory Concentrations for stated antibiotics are documented as tested by E-test and should be interpreted as fully sensitive, intermediate sensitivity or resistant using the following criteria set by the manufacturers AB Biodisk (Sweden):

Antibiotic	Sensitive MIC	Intermediate Sensitivity MIC	Resistant MIC
Penicillin	**<0.06**	**0.12 - 1**	**>2**
Erythromycin	**<1**	**2**	**>4**
Chloramphenicol	**<4**	**Not Applicable**	**>8**
Co-trimoxazole	**<0.5**	**1 - 2**	**>4**
Tetracycline	**<2**	**4**	**>8**
Vancomycin	**<1**		

Appendix 3

Primer	Sequence	Primer	Sequence
SP0031_f	TACTCTCCATGCCTTACTCGGT	SP0031_r	TTGATTAACCTGACTGCACAGG
SP0067_f	CTGGAGCAGTTTTGACAAATGA	SP0067_r	GCCACAAATTCTCCTTTGATTC
SP0068_f	GGAGTGCCTTTCCAACTTCTAA	SP0068_r	CATAGTGACCAGATTTCCCGTT
SP0069_f	TCAGAGGAATTGGCTGGTAGAT	SP0069_r	AGTACCAGTGTGGTCCGACTTT
SP0072_f	ATGCAGATTGCAGGAATCATCT	SP0072_r	TCAAATCAACATTTTCTGTGGG
SP0074_f	CAGGAGAGTTTTACCGTCCATC	SP0074_r	ATTGCGTTCCTGTGGATCTAGT
SP0104_f	TGGAACCCTCGTTGATAGTTCT	SP0104_r	AGATCTGCTTGCTCTCCAAATC
SP0111_f	GATGATGAAACGGTCGAGTACA	SP0111_r	CGATAATGACTGTTTTCCCCTC
SP0113_f	GGTTTGGACACATCAGTTGCTA	SP0113_r	GCTCCTATCTGATGCGCTATTT
SP0115_f	AGATGATGTATCTTGATGGGGG	SP0115_r	ATATGCGATTGCTCCACCTAGT
SP0165_f	GTTCGGTAGGTGCCAGTAATGT	SP0165_r	TTGCTTTACGCAATCCTACCTC
SP0166_f	CTTCAACAATTTTCCCCAAAAG	SP0166_r	ATAACCTGCTGCTCATTCACCT
SP0168_f	TCATTCGCATCATTTTTACAGG	SP0168_r	ACCAAGAAGTTGATGTGGAGGT
SP0278_f	GAATTGGCACATCTAATCGTGA	SP0278_r	AAGTGACATAACCATCTGCACG
SP0309_f	AGATTGAACAACAACGTCGAGA	SP0309_r	TGGGGAAAATAATAATAGGCGA
SP0350_f	TTGATTCAGAAGGTCCGGTTAT	SP0350_r	GCCATCTAACTCCGTCTTCATC
SP0352_f	GAATATCGTTTCTTCAGACGGC	SP0352_r	CAAAGCCAAACAATCTTTTTCC
SP0355_f	TATGTTACATGGTGCAACGGAT	SP0355_r	GATTCCGCAAATTTTTCAAGAG
SP0356_f	GGAATTGGGAATTTAGTTGCTG	SP0356_r	AACGGCTAATCCATAACACCAC
SP0410_f	TTTTAGATACGTTGCCGGTTCT	SP0410_r	CTGTGATAATGGTTGGGAAGGT
SP0458_f	ACAAGTCTGTGGGACTCCAGAT	SP0458_r	GAGCAAGACTTTTTCGTTTGGT
SP0461_f	AATTTCCACTTGAGTTCCCTGA	SP0461_r	AAGGCAATCGTCAAAAAGTGAT
SP0462_f	CTGGGACATACACCTTGACAGA	SP0462_r	ATTCAATCGCTTTCCGTTTTTA
SP0463_f	AGAATGACTGAAGGTTTGGCAT	SP0463_r	ATTACAAATTCTGCCCCAGCTA
SP0464_f	TGGAGAACTATCAGGAGGTGGT	SP0464_r	TTTCATGACTTTGAACATTGCC
SP0465_f	ACCATTTTTATCAGCATCCCTTT	SP0465_r	ATGGATGGTGAATCAGAAAGAAA
SP0466_f	CGAGTGGAATCAAATCAACAAA	SP0466_r	TGCTGCAATAAATTCTTCCTCA
SP0467_f	GAAAAAGAAAGGCGTCTCAGAA	SP0467_r	GTCGATTACGATACACGCGATA
SP0468_f	CGGAGGGATATGAGGTCAATTA	SP0468_r	TAAACGTGCTAGCTTCCACAAA
SP0509_f	AGTTCTTGACCCAAAACTTGGA	SP0509_r	AACTCCAACACGACCATTTTCT
SP0514_f	GCAGGTTCTTGTTCGATTCTTC	SP0514_r	CAAACAACAAGACCACCTGAGA
SP0532_f	TATGGATACTGAAATGCTTGCG	SP0532_r	AGTCCCAATTTATCACCAACCA
SP0568_f	TGTCAAAATCGAAAAACGTGTC	SP0568_r	ACCCAGAAGAAGATGATGTCGT
SP0570_f	GATTTTCCTAGAGAAGCCGTGA	SP0570_r	TATTCTTTTCCGTTGCCATCTT
SP0573_f	CAAGGAGGTGACTCTTATGGCT	SP0573_r	CAAGAAAAACTCACTATCCCCA
SP0574_f	GCTCTCAGTTCGAGGAGATGTT	SP0574_r	ACGACCTTCAACTCCTTATCCA
SP0575_f	TCGTGAAAAGGTAATCGGCTAT	SP0575_r	TATGAATCCGACCTGCATACTG
SP0697_f	GCGGTGAATTAAAGAGAATTGG	SP0697_r	GGCTCACATGAATGAATTGAAA
SP0740_f	CAGAAGAGAGGCAGTCGAATTT	SP0740_r	CAAATCATAAAGCCAGTCGTCA
SP0764_f	AGAAGAGGTCAAAAACTCAGCG	SP0764_r	TTGATAAAAGGCGTGAACATTG
SP0800_f	GAATTAAACCAGCAGCTTCCTC	SP0800_r	CAAAGGCAAGAGCAGTTGATAA
SP0949_f	AGCTCATTATGACCCTCCGTTA	SP0949_r	GTTTTCTGCGGTTACGATAAGG
SP1018_f	AGCAAGGAAAAGGTGTTGTGAT	SP1018_r	ATCTGGATCTGTTCTCCATCGT
SP1048_f	TATGGAATGCTAAGCCGGATAC	SP1048_r	TTCCATTGTATTTCCTTGTTGG
SP1049_f	GTGGAGTGTTGCTACAACGAAA	SP1049_r	GGATTTTTATTGTCAAAATTGCCT
SP1053_f	AGATGTGAAAGCTCAACGTCAA	SP1053_r	TTGAAGGAATCTTTTCCTCCCT
SP1059_f	GAAAATTTGGAATTGTCGGAAG	SP1059_r	AATGGCAAGTGCTACTATGTGC
SP1060_f	GGCCGATATGAAAAATAAATACG	SP1060_r	GCCAAACAATTTAAAGCGTGAT
SP1063_f	TGATCAATTAAGGCACAAATCG	SP1063_r	AGGGATTTAACTGACCAATCCA
SP1064_f	ACCCAAATATCGTCGTCAAATC	SP1064_r	TATTTCGCCAGTAAACGGATCT
SP1129_f	AATCAAAAAGAATGGCCAAAGA	SP1129_r	TTATTGATGCTGATAACACCGC
SP1130_f	CATGGCGACTCTATGGAAAAAT	SP1130_r	AAACAATTTCTGAAGGTGCGAT

SP1131_f	TAAAGAAAACAAGATGCGTGGA	SP1131_r	TGCCTTTGCTTGAGTTTTGTTA
SP1132_f	GCTACAACCCTTGATAACTGGG	SP1132_r	ACACATCTGACGGACGATAGAA
SP1134_f	GAACGAGCAACAGGCTACTTTT	SP1134_r	GGTACTGCTCCAACTCCTCACT
SP1135_f	TTAGCGAAGAAAATGGCCTAAG	SP1135_r	CATTGCCTCCTCAAACTTCTCT
SP1136_f	GATGAGGTAATGGCTGAAATCC	SP1136_r	CCAATCTTCATAGCTGGTAGGC
SP1137_f	CCTATTGTGAACAATGCCAAGA	SP1137_r	AATACGGCTATAATTCCGCTCA
SP1138_f	GGTTAGTTTGCCACACTTGGTT	SP1138_r	ACACTTTTCATGGGGTGAGAAA
SP1139_f	TGATGGAGAAAGGCTTTGTTTT	SP1139_r	TGAATTGCTCCATGTCAATTTC
SP1140_f	AGACAAAGCATGTCAAAAAGCA	SP1140_r	ACTTCCTCAATTTCCAACCTCA
SP1141_f	TGACTTAGGGGAATTTTTCCAA	SP1141_r	TATTTCCGTTTCATCCTCGTTT
SP1142_f	CTCAAGTGGTATTTCATGCGAG	SP1142_r	ATAATATCCGCTGACAAGCTCC
SP1143_f	TTGAGTTGTTAAGCCAACATGG	SP1143_r	GGGTGTTTTCTGTGTCCTTTTC
SP1144_f	GAGGAGATGAGCGGTGTTAGTC	SP1144_r	AACCTGCTCATGCTCTAATGGT
SP1145_f	TTTTAATAGCGCCCGAAAATTA	SP1145_r	ACTGATGATACTTGCTCTGCGA
SP1158_f	GAAATTGATTTGATTGTCCTGA	SP1158_r	CATGAAAAAGAAAACAATAGCAA
SP1181_f	TTATCACCTCGTTTCCTACTACT	SP1181_r	TTGTTTTTGTATTTGAATGATGAGC
SP1189_f	TTGTTTTGAATGTTGATGGGAA	SP1189_r	AGGACCAAGATTGAAACGAAAA
SP1254_f	AGCCAAAGAACGTAAGCGTAAC	SP1254_r	AGAATGCTTCAATTTCTGGGTT
SP1323_f	AGCTGTTTCCCAAGAAGAATTG	SP1323_r	TTTGCATCATTTTCCTTGATTG
SP1336_f	AGTGCTTGGGATTTTGTGAGAT	SP1336_r	TCTGGTGTCATACAAGGAATCG
SP1342_f	GTATTCTGGGTTGGTCAGCAAT	SP1342_r	TCCGTTATTATCATGTTCAGCG
SP1343_f	GGACGGGTTTTGTTAGACACTC	SP1343_r	ATTTGTTTAACGCCTTTCCTGA
SP1350_f	GCGTTATTATAAGCTCCCCAAA	SP1350_r	CAAATAAATAGCCCCATCCTCA
SP1353_f	CCAGTCTTTGAGAGCGATAAGC	SP1353_r	AAAACAAAAATCCTGCCAAAGA
SP1381_f	GTCAGTATCATTTCACCCAGCA	SP1381_r	ACTGTCTTCCCTTTTTGTGCAT
SP1696_f	TCAAAAAGAAACGAGGACTCTCA	SP1696_r	TTTTTGCTTCCTTCTTTTGGAA
SP1718_f	TGTTGTCTGAAACTATTAGCCTTTT	SP1718_r	TTTGAGTAACTCCTTTTTCCTCG
SP1762_f	TCCGTTTTTGATGTTATGCAAG	SP1762_r	TTGATGCATTCGTTCAGGATAC
SP1763_f	GATTCAATCCTTGGCAGTTAGC	SP1763_r	TCAAACAAAACAAAAAGCATGG
SP1765_f	ATTGGGGAGAAAAATGGAAGTT	SP1765_r	CAGCTCCTCAGAACAATCACAC
SP1766_f	CTTCAATGCTGGTGTTCTCTTG	SP1766_r	CCCACCAAACTTCTCTTAGACG
SP1770_f	TATTGCGTCAGAGTGGTTTTTG	SP1770_r	CATGCTCCATCTCACAACTAGC
SP1771_f	TTATGTGACCTTTGTGGACTCG	SP1771_r	AATCCATTCATTTGGAAAATCG
SP1793_f	TTACTGCTAGCTGTGCGTCTTC	SP1793_r	AGCGCATCTGCTAAAGAATACC
SP1794_f	TTGGAGGAACTGATGAAGAACA	SP1794_r	CTCCCCTTTGTAAGCTCTCTCA
SP1796_f	ATGGGCACATTTACTCATTTCC	SP1796_r	TATTTTGCACAGATTGGAGTGG
SP1797_f	TGCTCGTAAAAGAATTGGGAAT	SP1797_r	TACAAGTTGCAATGGTTCCAAG
SP1799_f	CGTCAGGAATTCCCTTTGTTAG	SP1799_r	ATCAAACAAAATACGAACCGCT
SP1839_f	ACCCTTTATACCATGTTGGTGC	SP1839_r	TAGACGATTTCCCTGAACCTGT
SP1895_f	GGATCGGTTCTGATTTTAGTGC	SP1895_r	GAAGTCATTCCAGAACCAAAGG
SP1896_f	TTCCCAGCTGTTTTATCTGGTT	SP1896_r	GCATAACCGAATTGGTTGTTTT
SP1897_f	GCTATACCGATGTTATCGGAGC	SP1897_r	TTGACCATGCCTTGTTTATCAC
SP1948_f	TCAAACGAAAAATTTTGTGGAAA	SP1948_r	TAATAGGAGTAGGCGTCCAACG
SP1949_f	ACAGATGGTGCTGACCCTAGAT	SP1949_r	TTTTTCTTGTTGTCTTTGCCAC
SP1975_f	GGCTGTGTCAATGTCGATAAAA	SP1975_r	ACACTAGCTGGTGAGAAGAGGG
SP2142_f	GGTATGAGGCGCTTAGCTCTTA	SP2142_r	ACCGTGTATTCTTCGTAGGCAT
SP2164_f	GCAGAAATGTTTGTTGGTGAGA	SP2164_r	CCGTGACAAAGCCACATTATTA

Appendix 3 Primer sequences used in the manufacture of the *S. pneumoniae* microarray SPv1.1 and used for confirmatory PCR. Forward primers are on the left and reverse primers are on the right of the table.

Appendix 4

Gene	Sequence (5'-3')
SP0045 f	AGTTGTGGACAAGGATGTC
SP0045 r	CAAAGGTACCAGACATGGA
SP0054 f	CACCGCAACAATATCCTGTCTA
SP0054 r	GCATGGAGTTTGATGACTGGTA
SP0110 f	TTGGTTTTCAAGATAGGGCTGT
SP0110 r	TTCGTTTTTGCTGCTCTTCATA
SP0314 f	TTACAATGGCGATTTGAGTCAC
SP0314 r	CCCAAACTCCTTGTTTAGCATC
SP0376 f	CCTCTCCATGGCTCTTC
SP0376 r	CATCAAGCTTGCTCCGTA
SP0740 f	CAGAAGAGAGGCAGTCGAATTT
SP0740 r	CAAATCATAAAGCCAGTCGTCA
SP0800 f	GAATTAAACCAGCAGCTTCCTC
SP0800 r	CAAAGGCAAGAGCAGTTGATAA
SP1342 f	GTATTCTGGGTTGGTCAGCAAT
SP1342 r	TCCGTTATTATCATGTTCAGCG
SP1343 f	GGACGGGTTTTGTTAGACACTC
SP1343 r	ATTTGTTTAACGCCTTTCCTGA
SP1350 f	GCGTTATTATAAGCTCCCCAAA
SP1350 r	CAAATAAATAGCCCCATCCTCA
SP1631 f	ATTCGAATCTGGCGTAACAACT
SP1631 r	GACGATAGATAGTCAAACCGCC
SP1869 f	CTTTCCATCGGCTAGTCT
SP1869 r	TGCGACCATGAGTTTGG
SP1872 f	GACCTAGTGGGAACTGTC
SP1872 r	CACCGATGGCAAGGGTA
SP2153 f	GCTACCATTCTTGCTCCT
SP2153 r	GGCTTCGGCATAAATATCC
SP1923f	TGCAGAGCGTCCTTTGGTC
SP1923r	CTCTTACTCGTGGTTTCCAACTTGA
SP1219f	TATGGGGTTTGTCTGGGGTC
SP1219r	GCGCGAGCTCTTCCTGATGT
ErmBf	GAAAAGGTACTCAACCAAATA
ErmBr	AGTAACGGTACTTAAATTGTTTAC

Appendix 4

Primer sequences used for qRT-PCR experiments.

Appendix 5

	Microarray CGH Results				Sanger Sequencing Results			
	INV104B	INV200	OXC141	ATCC700669	INV104B	INV200	OXC141	ATCC700669
SP0067								
SP0068								
SP0069								
SP0070								
SP0071								
SP0072								
SP0074								
SP0076								
SP0094								
SP0100								
SP0110								
SP0111								
SP0112								
SP0116								
SP0132								
SP0163								
SP0164								
SP0165								
SP0166								
SP0167								
SP0168								
SP0196								
SP0301								
SP0303								
SP0305								
SP0306								
SP0307								
SP0310								
SP0348								
SP0349								
SP0350								
SP0351								
SP0352								
SP0353								
SP0354								
SP0355								
SP0356								
SP0357								
SP0358								
SP0359								
SP0360								
SP0396								
SP0460								
SP0461								
SP0462								
SP0463								
SP0464								
SP0465								
SP0466								
SP0467								
SP0468								
SP0473								
SP0474								
SP0475								
SP0476								
SP0477								
SP0478								
SP0505								
SP0506								
SP0511								
SP0534								
SP0570								
SP0574								
SP0575								
SP0692								
SP0695								
SP0697								
SP0738								
SP0739								
SP0740								
SP0773								
SP1031								

	Microarray CGH Results				Sanger Sequencing Results			
	INV104B	INV200	OXC141	ATCC700669	INV104B	INV200	OXC141	ATCC700669
SP1047								
SP1048								
SP1049								
SP1050								
SP1051								
SP1052								
SP1053								
SP1054								
SP1055								
SP1056								
SP1057								
SP1059								
SP1060								
SP1061								
SP1062								
SP1063								
SP1064								
SP1129								
SP1130								
SP1131								
SP1132								
SP1133								
SP1134								
SP1135								
SP1136								
SP1137								
SP1138								
SP1139								
SP1140								
SP1141								
SP1142								
SP1143								
SP1144								
SP1145								
SP1147								
SP1154								
SP1181								
SP1221								
SP1222								
SP1315								
SP1316								
SP1317								
SP1318								
SP1319								
SP1320								
SP1321								
SP1322								
SP1323								
SP1324								
SP1325								
SP1326								
SP1327								
SP1328								
SP1329								
SP1330								
SP1331								
SP1335								
SP1336								
SP1339								
SP1340								
SP1341								
SP1342								
SP1343								
SP1344								
SP1345								
SP1348								
SP1349								
SP1350								
SP1611								
SP1615								
SP1616								
SP1617								

	Microarray CGH Results				Sanger Sequencing Results			
	INV104B	INV200	OXC141	ATCC700669	INV104B	INV200	OXC141	ATCC700669
SP1618	Blue	Blue	Yellow	Blue	Blue	Blue	Yellow	Blue
SP1619	Blue	Blue	Yellow	Blue	Blue	Blue	Yellow	Blue
SP1620	Blue	Blue	Yellow	Blue	Blue	Blue	Yellow	Blue
SP1621	Blue	Blue	Yellow	Blue	Blue	Blue	Yellow	Blue
SP1622	Blue	Yellow	Yellow	Blue	Blue	Blue	Blue	Blue
SP1758	Yellow	Yellow	Blue	Yellow	Yellow	Yellow	Blue	Yellow
SP1759	Yellow	Yellow	Blue	Yellow	Yellow	Yellow	Blue	Yellow
SP1760	Yellow	Blue	Blue	Yellow	Yellow	Blue	Blue	Yellow
SP1761	Yellow	Blue	Blue	Yellow	Yellow	Blue	Blue	Yellow
SP1762	Yellow	Blue	Blue	Yellow	Yellow	Blue	Blue	Yellow
SP1763	Yellow	Blue	Blue	Yellow	Yellow	Blue	Blue	Yellow
SP1764	Yellow	Blue	Blue	Yellow	Yellow	Blue	Blue	Yellow
SP1765	Yellow	Blue	Blue	Yellow	Yellow	Blue	Blue	Yellow
SP1766	Yellow	Blue	Blue	Yellow	Yellow	Blue	Blue	Yellow
SP1767	Yellow	Blue	Blue	Yellow	Yellow	Blue	Blue	Yellow
SP1768	Yellow	Blue	Yellow	Yellow	Yellow	Blue	Blue	Yellow
SP1769	Yellow	Blue	Blue	Yellow	Yellow	Blue	Blue	Yellow
SP1770	Yellow	Blue	Blue	Yellow	Yellow	Blue	Blue	Yellow
SP1771	Yellow	Blue	Blue	Yellow	Yellow	Blue	Blue	Yellow
SP1772	Yellow	Yellow	Blue	Yellow	Yellow	Blue	Blue	Yellow
SP1773	Yellow	Yellow	Blue	Yellow	Blue	Blue	Blue	Yellow
SP1796	Blue	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow
SP1797	Blue	Yellow	Yellow	Yellow	Blue	Yellow	Yellow	Yellow
SP1798	Blue	Yellow	Yellow	Yellow	Blue	Yellow	Yellow	Yellow
SP1799	Blue	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow
SP1829	Yellow	Blue	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow
SP1834	Yellow	Blue	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow
SP1843	Yellow	Yellow	Yellow	Blue	Yellow	Yellow	Yellow	Yellow
SP1866	Yellow	Yellow	Yellow	Blue	Yellow	Yellow	Yellow	Yellow
SP1948	Blue	Yellow	Blue	Yellow	Blue	Blue	Blue	Blue
SP1949	Blue	Yellow	Blue	Yellow	Blue	Blue	Blue	Blue
SP1950	Blue	Blue	Blue	Yellow	Blue	Blue	Blue	Yellow
SP1951	Blue	Blue	Blue	Yellow	Blue	Blue	Blue	Yellow
SP1952	Blue	Blue	Blue	Yellow	Blue	Blue	Blue	Yellow
SP1953	Blue	Blue	Blue	Yellow	Blue	Blue	Blue	Yellow
SP1954	Blue	Blue	Blue	Yellow	Blue	Blue	Blue	Yellow
SP1955	Yellow	Yellow	Blue	Yellow	Blue	Blue	Blue	Yellow
SP2159	Yellow	Yellow	Blue	Yellow	Yellow	Yellow	Blue	Yellow
SP2160	Yellow	Yellow	Blue	Yellow	Yellow	Yellow	Blue	Yellow
SP2161	Yellow	Yellow	Blue	Yellow	Yellow	Yellow	Blue	Yellow
SP2162	Yellow	Yellow	Blue	Yellow	Yellow	Yellow	Blue	Yellow
SP2163	Yellow	Yellow	Blue	Yellow	Yellow	Yellow	Blue	Yellow
SP2164	Yellow	Yellow	Blue	Yellow	Yellow	Yellow	Blue	Yellow

Appendix 5

Comparison of microarray DNA CGH results with genome sequence data.

Blue indicates absence of hybridization of the gene while yellow indicates a positive hybridization by DNA from the test isolate. In the sequencing results, blue indicates the absence of the gene and yellow indicates the presence of the gene. Genes are identifiable using the TIGR4 genome nomenclature. Where these genes are indicated in red, they have been identified by signature tagged mutagenesis in a murine pneumonia model as being required for virulence (Hava and Camilli, 2002).

Appendix 6

Gene	Function
SPN23F00250	Putative Phage Integrase (pseudogene)
SPN23F00260	Putative type 1 Restriction Enzyme Related Protein
SPN23F00280	Putative Phage Protein
SPN23F00290	Plasmid Stabilisation System Protein
SPN23F00300	Putative Uncharacterized Protein
SPN23F00310	Putative Phage DNA Binding Protein
SPN23F00320	Putative Uncharacterized Protein
SPN23F00330	Putative DNA Binding Protein
SPN23F00340	Putative Uncharacterized Protein
SPN23F00350	Putative Uncharacterized Protein
SPN23F00360	Putative Peptidoglycan Hydrolytic Amidase (pseudogene)
SPN23F00370	Putative Membrane Protein
SPN23F00380	Putative DNA Binding Protein
SPN23F00710	Putative ATP/GTP Binding Protein
SPN23F00830	Conserved Hypothetical Protein
SPN23F00840	Putative Uncharacterized Protein
SPN23F00850	Conserved Hypothetical Protein
SPN23F00860	Putative Gene Fragment
SPN23F01260	Putative Membrane Protein
SPN23F01270	Putative Membrane Protein
SPN23F01840	Regulatory Protein (Orthologue of SP0189 in TIGR4)
SPN23F03210	Capsule Biosynthesis Tyrosine Protein Kinase (Orthologue of SPD_0318 in D39)
SPN23F03220	Undecaprenylphosphate glucosephosphotransferase (Orthologue of SPD_0319 in D39)
SPN23F03230	Putative Rhamnosyl Transferase (Orthologue of SPD_0320 in D39)
SPN23F03250	Oligosaccharide Repeat Unit Polymerase Unit wzy (Orthologue of SPD_0323 in D39)
SPN23F03260	Putative Glycosyltransferase WchA
SPN23F03270	Putative Glycosyltransferase WchW
SPN23F03280	Capsule Biosynthesis Repeating Unit Flippase (Orthologue of SPD_0325 in D39)
SPN23F03290	Putative Glycerol Phosphotransferase WchX
SPN23F03300	Putative Glycerol-2-Phosphate Dehydrogenase WchY
SPN23F03310	Putative Nucleotidyl Transferase WchZ
SPN23F03320	Putative Phosphotransferase
SPN23F03330	Glucose-1-Phosphate Thymidyl Transferase (Orthologue of SPD_0328 in D39)
SPN23F03340	dTDP-4-keto-6-deoxyglucose-3,5-epimerase RmlC (Orthologue of SPD_0329 in D39)
SPN23F03350	dTDP-glucose-4,6-dehydratase RmlB (Orthologue of SPD_0330 in D39)
SPN23F03360	dTDP-4-dehydrorhamnose reductase RmlD (Orthologue of SPD_0331 in D39)
SPN23F03370	Putative Transposase
SPN23F06010	Putative Pneumococcal Surface Protein
SPN23F06180	Putative Uncharacterized Protein
SPN23F07060	Putative Uncharacterized Protein
SPN23F07070	ABC-type Antimicrobial Peptide Transporter, Permease Component, Putative
SPN23F07080	Putative ABC Transporter ATP-binding Protein
SPN23F07090	Putative Exported Protein
SPN23F09740	Tn5252 orf10 protein (Orthologue of SPD_0934 in D39)

SPN23F09750	Tn5252 orf9 protein (Orthologue of SPD_0935 in D39)
SPN23F09780	Tn5252 Relaxase
SPN23F09800	Putative Lantibiotic Modifying Enzyme
SPN23F09840	Putative Membrane Protein
SPN23F09850	Putative Membrane Protein
SPN23F10590	Zinc Metalloproteinase ZmpD (Orthologue of SPD_0577 in D39)
SPN23F12240	Putative Phosphosugar-Binding Transcriptional Regulator
SPN23F12250	Putative Membrane Protein
SPN23F12270	Putative IS861 Transposase Orf1
SPN23F12410	Integrase
SPN23F12420	Conserved Hypothetical Protein
SPN23F12430	Relaxase (Orthologue of SP1056 in TIGR4)
SPN23F12440	Putative Mobilisation Protein (Orthologue of SP1055 in TIGR4)
SPN23F12450	Putative Mobilisation Protein (Orthologue of SP1054 in TIGR4)
SPN23F12460	Putative Uncharacterized Protein
SPN23F12470	Putative DNA Helicase II, UvrD
SPN23F12480	Hypothetical Protein
SPN23F12490	Putative Uncharacterized Protein
SPN23F12500	Putative Uncharacterized Protein
SPN23F12510	Putative NTPase Protein
SPN23F12520	Hypothetical Protein
SPN23F12530	Putative Phosphoserine Phosphatase
SPN23F12540	Hypothetical Protein
SPN23F12550	Hypothetical Protein
SPN23F12560	Hypothetical Protein
SPN23F12570	Putative Uncharacterized Protein
SPN23F12580	Replication Protein
SPN23F12590	Putative Chloramphenicol Acetyltransferase
SPN23F12600	Putative Uncharacterized Protein
SPN23F12610	Putative Uncharacterized Protein
SPN23F12620	Zeta Toxin
SPN23F12630	Putative Epsilon Antitoxin (Orthologue of SP1050 in TIGR4)
SPN23F12640	Putative Uncharacterized Protein
SPN23F12650	Putative Uncharacterized Protein
SPN23F12660	Putative Uncharacterized Protein
SPN23F12670	Putative Uncharacterized Protein
SPN23F12680	Putative Uncharacterized Protein
SPN23F12690	Putative Lantibiotic Transport/Processing ATP Binding Protein
SPN23F12700	Putative Lantibiotic Synthetase
SPN23F12701	Putative Lantibiotic Precursor
SPN23F12710	Putative Lantibiotic ABC Transporter
SPN23F12720	Putative Lantibiotic ABC Transporter
SPN23F12730	Putative Lantibiotic ABC Transporter
SPN23F12740	Putative Membrane Protein
SPN23F12750	Transcriptional Regulator
SPN23F12760	Putative Membrane Protein
SPN23F12770	Putative Uncharacterized Protein
SPN23F12780	Putative Conjugative Transposon DNA Recombination Protein
SPN23F12790	Putative Group II Intron Reverse Transcriptase/Maturase
SPN23F12820	Putative Uncharacterized Protein
SPN23F12830	Putative Uncharacterized Protein
SPN23F12840	Putative Conjugal Transfer Protein
SPN23F12850	Putative Conjugal Transfer Protein
SPN23F12860	Putative Uncharacterized Protein
SPN23F12870	Putative Uncharacterized Protein
SPN23F12880	Putative Uncharacterized Protein
SPN23F12890	Putative Conjugal Transfer Protein TraG

SPN23F12900	Putative Uncharacterized Protein
SPN23F12910	Putative Uncharacterized Protein
SPN23F12920	Conserved Hypothetical Protein
SPN23F12930	Conserved Hypothetical Protein
SPN23F12940	Conjugative Transposon Protein
SPN23F12950	Conjugative Transposon Protein
SPN23F12960	Conjugative Transposon FtsK/SpoIIIE-family protein
SPN23F12970	Putative Conjugative Transposon Replication Initiation Factor
SPN23F12980	Conjugative Transposon Protein
SPN23F12990	Conjugative Transposon Protein
SPN23F13000	Putative Conjugative Transposon Membrane Protein
SPN23F13030	Putative Cell Wall Hydrolase
SPN23F13040	Putative Conjugative Transposon Exported Protein
SPN23F13050	Conjugative Transposon Tetracycline Resistance Protein
SPN23F13060	Putative Conjugative Transposon Regulatory Protein
SPN23F13061	Putative Uncharacterized Protein
SPN23F13070	Putative Conjugative Transposon Regulatory Protein
SPN23F13080	Excisionase
SPN23F13090	Putative Integrase
SPN23F13160	Putative Replication Initiator Protein
SPN23F13170	Putative Uncharacterized Protein
SPN23F14470	Putative Transposase
SPN23F14540	Putative IS1239 Putative Transposase
SPN23F15110	Putative Collagen-like Surface Anchored Protein
SPN23F15120	Putative Mga-like Regulatory Protein (Ortholog of SPD_1377 in D39)
SPN23F15300	Antiholin
SPN23F15310	Holin
SPN23F15320	Putatuve Uncharacterized Prophage Protein
SPN23F15330	Phage Structural Protein
SPN23F15340	Putative Platelet Binding Phage Protein
SPN23F15350	Phage Tail Protein
SPN23F15360	Putative Phage Minor Tail Protein
SPN23F15370	Putative Phage Gp15 Protein
SPN23F15380	Hypothetical Phage Protein
SPN23F15390	Hypothetical Phage Protein
SPN23F15400	Hypothetical Phage Protein
SPN23F15410	Hypothetical Phage Protein
SPN23F15420	Hypothetical Phage Protein
SPN23F15430	Hypothetical Phage Protein
SPN23F15440	Hypothetical Phage Protein
SPN23F15450	Putative Phage Capsid Protein
SPN23F15460	Putative Phage Scaffolding Protein
SPN23F15470	Minor Capsid Protein
SPN23F15490	Putative Minor Capsid Protein
SPN23F15500	Putative Phage Terminase Large Subunit
SPN23F15510	Hypothetical Phage Protein
SPN23F15520	Putative Phage Protein
SPN23F15530	Putative Phage Protein
SPN23F15540	Putative Phage Protein
SPN23F15550	Hypothetical Phage Protein
SPN23F15560	Hypothetical Phage Protein
SPN23F15570	Putative Phage Protein
SPN23F15580	Putative Phage Protein
SPN23F15590	Hypothetical Phage Protein
SPN23F15600	Putative Phage Protein
SPN23F15610	Putative Phage Protein
SPN23F15620	Putative Phage Holliday Junction Resolvase

SPN23F15630	Phage Protein
SPN23F15640	Phage Protein
SPN23F15650	Phage Protein
SPN23F15660	Putative Phage Protein
SPN23F15670	Putative DNA Methylase
SPN23F15680	Putative Single Strand DNA-binding protein
SPN23F15700	Putative Phage Protein
SPN23F15710	Phage Single Strand DNA Binding Protein
SPN23F15720	Putative Phage Protein
SPN23F15730	Putative Phage Protein
SPN23F15740	Putative Phage Protein
SPN23F15750	Hypothetical Phage Protein
SPN23F15760	Putative Phage DNA Binding Protein
SPN23F15770	Putative DNA Binding Protein
SPN23F15780	Hypothetical Phage Protein
SPN23F15790	Phage Protein
SPN23F15800	Phage Integrase
SPN23F17430	Site Specific Recombinase
SPN23F17440	Hypothetical Protein
SPN23F17450	Putative Membrane Protein
SPN23F17460	Conserved Hypothetical Protein
SPN23F17760	Putative Glycosyl Transferase
SPN23F17960	Putative Otitis Media Associated p41
SPN23F17970	Conserved Hypothetical Protein
SPN23F17980	Putative ATPase
SPN23F18180	Putative Membrane Protein
SPN23F18190	Conserved Hypothetical Protein
SPN23F18200	Putative Choline Sulfatase
SPN23F18210	Sugar Phosphotransferase System (PTS) IIC Component
SPN23F18220	Putative Lactose/Cellobiose-specific Phosphotransferase System (PTS) IIA Component
SPN23F18230	Sugar Phosphotransferase System (PTS) Lactose/Cellobiose-specific Family IIB Component
SPN23F18240	Conserved Hypothetical Protein
SPN23F18250	Sugar Phosphotransferase (PTS) IIC Component
SPN23F18260	Putative ROK Family Repressor Protein
SPN23F18640	Putative Restriction Enzyme
SPN23F18650	Putative DNA Modification Methylase
SPN23F18990	ABC Transporter ATP-Binding Protein
SPN23F19000	Transport System Permease Protein
SPN23F19010	Putative Substrate Binding Protein
SPN23F19700	Conserved Domain Protein
SPN23F19710	Putative Uncharacterized Protein
SPN23F20090	Putative Exported Protein
SPN23F21700	IS1381 Transposase
SPN23F21701	IS1381 Transposase

Appendix 6

Genes identified by ACT as present in the ATCC700669 genome but not present in the TIGR4 genome.

Appendix 7

Gene	Function
SPNOXC00180	Unknown Function
SPNOXC00190	Unknown Function
SPNOXC00200	Unknown Function (Orthologue of SPN23F00280)
SPNOXC00210	Unknown Function (Orthologue of SPN23F15770)
SPNOXC00220	Unknown Function
SPNOXC00230	Unknown Function
SPNOXC00240	Unknown Function
SPNOXC00250	Unknown Function
SPNOXC00260	Unknown Function
SPNOXC00270	Unknown Function
SPNOXC00280	Unknown Function
SPNOXC00290	Unknown Function
SPNOXC00300	Unknown Function
SPNOXC00310	Unknown Function
SPNOXC00320	Unknown Function
SPNOXC00330	DNA Replication Protein (Orthologue of SP1137 in TIGR4)
SPNOXC00340	Unknown Function
SPNOXC00350	Unknown Function
SPNOXC00360	Unknown Function (Orthologue of SPN23F15600)
SPNOXC00370	Unknown Function
SPNOXC00380	Transcriptional Activator
SPNOXC00390	Site Specific Recombinase
SPNOXC00400	Unknown Function
SPNOXC00410	Unknown Function
SPNOXC00420	Unknown Function
SPNOXC00430	Unknown Function
SPNOXC00440	Unknown Function
SPNOXC00450	Unknown Function
SPNOXC00460	Capsid Protein
SPNOXC00470	Unknown Function
SPNOXC00480	Unknown Function
SPNOXC00490	Unknown Function
SPNOXC00500	Unknown Function
SPNOXC00510	Unknown Function
SPNOXC00520	Tail Protein
SPNOXC00530	Unknown Function
SPNOXC00540	Unknown Function
SPNOXC00550	Unknown Function
SPNOXC00560	Unknown Function
SPNOXC00570	Unknown Function
SPNOXC00580	Unknown Function
SPNOXC00590	Unknown Function
SPNOXC00600	Unknown Function (Orthologue of SPN23F15310)
SPNOXC01370	Unknown Function
SPNOXC01440	Hypothetical Protein
SPNOXC01480	Conserved Hypothetical
SPNOXC01500	Unknown Function (Orthologue of SPD_0114 in D39)
SPNOXC01510	Unknown Function (Orthologue of SPD_0115 in D39)
SPNOXC01520	Unknown Function (Orthologue of SPD_0117 in D39)
SPNOXC01530	Unknown Function
SPNOXC01540	Unknown Function

SPNOXC01550	Unknown Function (Orthologue of SPD_0119 in D39)
SPNOXC01560	Unknown Function (Orthologue of SPD_0120 in D39)
SPNOXC01570	Unknown Function
SPNOXC01580	Unknown Function (Orthologue of SPD_0122 in D39)
SPNOXC01590	Putative Membrane Protein (Orthologue of SPN23F01260)
SPNOXC01600	Putative Membrane Protein (Orthologue of SPN23F01270)
SPNOXC01610	Hypothetical Protein (Orthologue of SPN23F01280, SPD_2086 in D39 and SP0116 in TIGR4)
SPNOXC01850	Hypothetical Protein (Orthologue of SPN23F01520)
SPNOXC02670	Putative AraC-family Transcriptional Regulator (Orthologue of SPN23F02341, SPD_0228 in D39)
SPNOXC03530	Putative IS630-Spn 1 Transposase (Pseudogene)
SPNOXC03570	Tyrosine kinase Wzc
SPNOXC03580	IS1548 Transposase
SPNOXC03590	UDP-glucose-6-dehydrogenase Ugd
SPNOXC03600	Serotype 3 Capsule Synthase
SPNOXC04440	Putative Uncharacterised Protein (Orthologue of SPN23F04330 and SPD_2116 in D39)
SPNOXC04780	Putative Membrane Protein
SPNOXC05230	Putative Uncharacterised Protein (Orthologue of SPN23F05150 and SPD_0495 in D39)
SPNOXC05350	Putative IS1239 Transposase
SPNOXC05460	Putative Transposase (Orthologue of SPN23F05390, SPD_0520 in D39 and SP2301 in TIGR4)
SPNOXC05760	IS1381 Transposase orfA
SPNOXC05770	IS1381 Transposase orfB
SPNOXC05790	50S Ribosomal Protein L11 (Orthologue of SPN23F05690 and SPD_0550 in D39)
SPNOXC05810	Putative Uncharacterized Protein (Orthologue of SPN23F05710 and SP2306 in TIGR4)
SPNOXC06450	Putative L-lactate Oxidase-Related Protein (Orthologue of SPN23F06370, SPD_0619 in D39 and SP0712 in TIGR4)
SPNOXC07350	Putative Transposase (Orthologue of SPN23F07370, SP2472 in TIGR4)
SPNOXC08130	Putative Uncharacterized Protein
SPNOXC08970	Putative Uncharacterized Protein (Orthologue of SPN23F09220, SPD_0883 in D39 and SP0997 in TIGR4)
SPNOXC09530	Tn5252 orf10 Protein (Orthologue of SPN23F09740, SPD_0934 in D39 and SP1054 in TIGR4)
SPNOXC09550	Tn5252 relaxase (Orthologue of SPN23F09780, SPD_2181 in D39 and SP1056 in TIGR4)
SPNOXC09560	D-Ala D-Ala Ligase A
SPNOXC09570	Enolase
SPNOXC09580	Branched Chain Amino Acid Aminotransferase
SPNOXC09590	Transketolase
SPNOXC09600	Transketolase
SPNOXC09610	Acetylornithine Aminotransferase
SPNOXC09620	Nucleoside Diphosphate Kinase
SPNOXC09630	2-Isopropylmalate Synthetase
SPNOXC09640	Unknown Function
SPNOXC09650	Unknown Function
SPNOXC09660	Phosphonopyruvate decarboxylase
SPNOXC09670	tRNA Synthetase
SPNOXC09680	Unknown Function
SPNOXC09690	Unknown Function
SPNOXC09700	Unknown Function
SPNOXC09730	ABC Transporter
SPNOXC09740	Unknown Function
SPNOXC09750	Site Specific Recombinase

SPNOXC09760	Unknown Function
SPNOXC09770	Unknown Function
SPNOXC10350	Putative IS630-Spn 1 Transposase
SPNOXC10920	Ion Channel Transport Protein
SPNOXC11710	Conserved Hypothetical Protein
SPNOXC11720	Putative Uncharacterized Protein
SPNOXC12560	Putative ABC Transporter ATP-binding protein
SPNOXC15800	Putative Transposase
SPNOXC15810	Putative Membrane Protein (Orthologue of SPN23F18180)
SPNOXC15820	Conserved Hypothetical Protein (Orthologue of SPN23F18190)
SPNOXC15830	Putative Choline Sulfatase (Orthologue of SPN23F18200)
SPNOXC15840	Sugar Phosphotransferase System (PTS) IIC Component (Orthologue of SPN23F18210)
SPNOXC15850	Putative Lactose/Cellobiose-specific Phosphotransferase System (PTS) IIA Component
SPNOXC15860	Sugar Phosphotransferase System (PTS), Lactose/Cellobiose-specific family IIB Component
SPNOXC15870	Conserved Hypothetical Protein (Orthologue of SPN23F18240)
SPNOXC15880	Sugar Phosphotransferase System (PTS), IIC component (Orthologue of SPN23F18250, SPD_0424 in D39 and SP0474 in TIGR4)
SPNOXC16750	Transcriptional Regulator
SPNOXC16760	Unknown Function
SPNOXC16770	Unknown Function
SPNOXC17150	Putative Secreted Protein (Orthologue of SPN23F19690, SPD_1746 in D39 and SP1947 in TIGR4)
SPNOXC18280	Degenerate Transposase (Orthologue of SPN23F20990, SPD_2259 in D39 and SP2459 in TIGR4)
SPNOXC19050	Glycosyl hydrolase
SPNOXC19070	Unknown Function
SPNOXC19100	ABC Transporter Permease
SPNOXC19110	ABC Transporter

Appendix 7

Genes identified by ACT as present in the OXC141 genome but not present in the TIGR4 genome.

Appendix 8

BAA659	BAA340	05_1271	ATCC700904	BAA660	05_2565	05_1821	ATCC51916	07_2839	OXC141
Serotype 6A	Serotype 14	Serotype 20	Serotype 19A	Serotype 35B	Serotype 12F	Serotype 9V	Serotype 23F	Serotype 9A	Serotype 3
ST376	ST67	ST568	ST41	ST377	ST218	ST156	ST37	ST239	ST180
USA	USA	Scotland	South Africa	USA	Scotland	Scotland	USA	Bolivia	England
						SP0040			
	SP0052								
SP0067	SP0067	SP0067	SP0067		SP0067	SP0067	SP0067	SP0067	SP0067
SP0068	SP0068	SP0068	SP0068	SP0068	SP0068	SP0068	SP0068	SP0068	SP0068
SP0069	SP0069	SP0069	SP0069	SP0069	SP0069	SP0069	SP0069	SP0069	SP0069
SP0070	SP0070	SP0070	SP0070	SP0070	SP0070	SP0070	SP0070	SP0070	SP0070
SP0071	SP0071	SP0071	SP0071	SP0071	SP0071	SP0071		SP0071	SP0071
SP0072			SP0072	SP0072	SP0076	SP0072		SP0072	
		SP0076							
								SP0093	
SP0109									
		SP0110		SP0110					
		SP0111		SP0111					
		SP0112		SP0112					SP0116
				SP0113					
								SP0163	SP0163
SP0164								SP0164	
SP0165							SP0165	SP0165	SP0165
SP0166				SP0166		SP0166	SP0166	SP0166	SP0166
SP0167	SP0167			SP0167	SP0167	SP0167	SP0167	SP0167	SP0167
SP0168	SP0168			SP0168	SP0168	SP0168	SP0168	SP0168	SP0168
	SP0171						SP0171		
							SP0172		
					SP0179				
					SP0183				
									SP0196
			SP0244						
					SP0270				
	SP0303		SP0303		SP0303				
			SP0304			SP0304			
	SP0305		SP0305	SP0305	SP0305	SP0305			
	SP0306		SP0306	SP0306	SP0306	SP0306			
	SP0307		SP0307	SP0307	SP0307	SP0307			
					SP0308				
			SP0309	SP0309					
			SP0310		SP0310	SP0310			
						SP0311			
		SP0331							
SP0348		SP0348	SP0348	SP0348	SP0348	SP0348	SP0348	SP0348	
SP0349		SP0349	SP0349	SP0349		SP0349	SP0349	SP0349	
SP0350	SP0350		SP0350	SP0350	SP0350	SP0350	SP0350	SP0350	SP0350
SP0351	SP0351	SP0351	SP0351	SP0351		SP0351	SP0351	SP0351	SP0351
SP0352	SP0352	SP0352	SP0352	SP0352		SP0352	SP0352	SP0352	SP0352
SP0353	SP0353	SP0353	SP0353	SP0353		SP0353	SP0353	SP0353	SP0353
SP0354	SP0354	SP0354	SP0354	SP0354	SP0354	SP0354	SP0354	SP0354	SP0354
SP0355	SP0355	SP0355	SP0355	SP0355	SP0355	SP0355	SP0355	SP0355	SP0355
SP0356	SP0356		SP0356	SP0356			SP0356	SP0356	SP0356
SP0357	SP0357	SP0357	SP0357	SP0357		SP0357	SP0357	SP0357	SP0357
SP0358	SP0358	SP0358	SP0358	SP0358		SP0358	SP0358	SP0358	SP0358
SP0359	SP0359	SP0359	SP0359	SP0359		SP0359	SP0359	SP0359	SP0359
SP0360	SP0360	SP0360	SP0360	SP0360		SP0360	SP0360	SP0360	SP0360
		SP0379							
				SP0394		SP0394		SP0394	
				SP0395	SP0395	SP0395		SP0395	
				SP0396	SP0396	SP0396		SP0396	SP0396
				SP0397	SP0397	SP0397		SP0397	
	SP0460	SP0460	SP0460		SP0460		SP0460	SP0460	SP0460
	SP0461	SP0461	SP0461		SP0461		SP0461	SP0461	SP0461
	SP0462	SP0462	SP0462		SP0462		SP0462	SP0462	SP0462
	SP0463	SP0463	SP0463		SP0463		SP0463	SP0463	SP0463
	SP0464	SP0464	SP0464		SP0464		SP0464	SP0464	SP0464
	SP0465	SP0465	SP0465		SP0465		SP0465	SP0465	SP0465
	SP0466	SP0466	SP0466		SP0466		SP0466	SP0466	SP0466
	SP0467	SP0467	SP0467		SP0467		SP0467	SP0467	SP0467
	SP0468	SP0468	SP0468		SP0468		SP0468	SP0468	SP0468
	SP0473		SP0473					SP0473	SP0473
	SP0474		SP0474					SP0474	SP0474
	SP0475		SP0475					SP0475	SP0475
	SP0476		SP0476					SP0476	SP0476
	SP0477		SP0477					SP0477	SP0477
	SP0478		SP0478					SP0478	SP0478
	SP0511								
		SP0512							
						SP0535			
SP0571									
					SP0574				

BAA659	BAA340	05_1271	ATCC700904	BAA660	05_2565	05_1821	ATCC51916	07_2839	OXC141
Serotype 6A	Serotype 14	Serotype 20	Serotype 19A	Serotype 35B	Serotype 12F	Serotype 9V	Serotype 23F	Serotype 9A	Serotype 3
ST376	ST67	ST568	ST41	ST377	ST218	ST156	ST37	ST239	ST180
USA	USA	Scotland	South Africa	USA	Scotland	Scotland	USA	Bolivia	England
									SP0575
	SP0610								
					SP0641				
			SP0647						
			SP0648						
SP0684									
						SP0693			
						SP0694		SP0694	
						SP0695		SP0695	
						SP0696		SP0696	
						SP0697		SP0697	
		SP0698						SP0698	
					SP0699				
	SP0724								
			SP0726						
			SP0727						
			SP0728						
			SP0729						
			SP0730						
			SP0731						
			SP0773						
		SP0886							
					SP0888				
		SP0907							
SP0934									
									SP1031
			SP1050						
	SP1051		SP1051						
								SP1053	
SP1054				SP1054	SP1054	SP1054	SP1054	SP1054	
SP1055				SP1055	SP1055	SP1055	SP1055	SP1055	
SP1056				SP1056	SP1056	SP1056	SP1056	SP1056	
SP1057				SP1057	SP1057	SP1057	SP1057	SP1057	SP1057
SP1059		SP1059		SP1059	SP1059	SP1059	SP1059	SP1059	SP1059
SP1060				SP1060	SP1060	SP1060	SP1060	SP1060	SP1060
SP1061	SP1061		SP1061	SP1061	SP1061	SP1061	SP1061	SP1061	SP1061
SP1062	SP1062		SP1062	SP1062	SP1062	SP1062	SP1062	SP1062	SP1062
SP1063	SP1063			SP1063	SP1063	SP1063	SP1063	SP1063	
SP1064	SP1064				SP1064	SP1064	SP1064	SP1064	
SP1065								SP1065	
SP1129	SP1129	SP1129	SP1129	SP1129	SP1129	SP1129		SP1129	SP1129
SP1130	SP1130	SP1130	SP1130	SP1130	SP1130	SP1130	SP1130	SP1130	SP1130
SP1131	SP1131	SP1131	SP1131	SP1131	SP1131	SP1131	SP1131	SP1131	SP1131
SP1132	SP1132	SP1132	SP1132	SP1132	SP1132	SP1132	SP1132	SP1132	SP1132
SP1133	SP1133	SP1133	SP1133	SP1133	SP1133	SP1133	SP1133	SP1133	SP1133
SP1134	SP1134	SP1134	SP1134	SP1134	SP1134	SP1134	SP1134		SP1134
SP1135	SP1135	SP1135		SP1135	SP1135	SP1135	SP1135		SP1135
SP1136	SP1136	SP1136	SP1136	SP1136	SP1136	SP1136	SP1136	SP1136	SP1136
SP1137	SP1137	SP1137	SP1137	SP1137	SP1137	SP1137	SP1137	SP1137	SP1137
SP1138			SP1138	SP1138		SP1138		SP1138	
SP1139	SP1139	SP1139	SP1139	SP1139	SP1139	SP1139	SP1139	SP1139	SP1139
SP1140	SP1140	SP1140	SP1140	SP1140	SP1140	SP1140	SP1140	SP1140	SP1140
SP1141	SP1141	SP1141	SP1141	SP1141	SP1141	SP1141	SP1141	SP1141	SP1141
SP1142	SP1142	SP1142	SP1142	SP1142	SP1142	SP1142	SP1142	SP1142	SP1142
SP1143	SP1143	SP1143	SP1143	SP1143	SP1143	SP1143	SP1143	SP1143	SP1143
SP1144	SP1144	SP1144	SP1144	SP1144	SP1144	SP1144	SP1144	SP1144	SP1144
SP1145	SP1145		SP1145	SP1145	SP1145	SP1145	SP1145	SP1145	SP1145
SP1146									
SP1154	SP1154		SP1154	SP1154	SP1154	SP1154		SP1154	
					SP1188				
			SP1189						
					SP1216				
SP1221	SP1221		SP1221	SP1221	SP1221	SP1221		SP1221	
SP1222	SP1222	SP1222	SP1222	SP1222	SP1222	SP1222	SP1222	SP1222	
			SP1234						
	SP1300								
		SP1315		SP1315		SP1315	SP1315	SP1315	SP1315
		SP1316		SP1316		SP1316	SP1316	SP1316	SP1316
		SP1317		SP1317		SP1317	SP1317	SP1317	SP1317
		SP1318		SP1318		SP1318	SP1318	SP1318	SP1318
		SP1319		SP1319		SP1319	SP1319	SP1319	SP1319
		SP1320		SP1320		SP1320	SP1320	SP1320	SP1320
		SP1321		SP1321		SP1321	SP1321	SP1321	SP1321
		SP1322		SP1322		SP1322	SP1322	SP1322	SP1322
		SP1323		SP1323		SP1323	SP1323	SP1323	SP1323
		SP1324		SP1324		SP1324	SP1324	SP1324	SP1324

BAA659	BAA340	05_1271	ATCC700904	BAA660	05_2565	05_1821	ATCC51916	07_2839	OXC141
Serotype 6A	Serotype 14	Serotype 20	Serotype 19A	Serotype 35B	Serotype 12F	Serotype 9V	Serotype 23F	Serotype 9A	Serotype 3
ST376	ST67	ST568	ST41	ST377	ST218	ST156	ST37	ST239	ST180
USA	USA	Scotland	South Africa	USA	Scotland	Scotland	USA	Bolivia	England
		SP1325		SP1325		SP1325	SP1325	SP1325	SP1325
		SP1326		SP1326		SP1326	SP1326	SP1326	SP1326
		SP1327		SP1327		SP1327	SP1327	SP1327	SP1327
		SP1328		SP1328		SP1328	SP1328	SP1328	SP1328
		SP1329		SP1329		SP1329	SP1329	SP1329	SP1329
		SP1330		SP1330		SP1330	SP1330	SP1330	SP1330
		SP1331		SP1331		SP1331	SP1331	SP1331	SP1331
		SP1332		SP1332		SP1332	SP1332		
		SP1333		SP1333		SP1333	SP1333		
		SP1334		SP1334		SP1334	SP1334		
		SP1335				SP1335	SP1335		
				SP1336		SP1336	SP1336	SP1336	
				SP1337		SP1343			
				SP1338			SP1338		
		SP1339		SP1339	SP1339		SP1339		
		SP1340		SP1340	SP1340		SP1340		
		SP1341		SP1341	SP1341		SP1341		
SP1342		SP1342		SP1342	SP1342		SP1342		
SP1343		SP1343	SP1343	SP1343	SP1343		SP1343	SP1343	SP1343
		SP1344		SP1344	SP1344		SP1344		
				SP1345	SP1345		SP1345		
		SP1346		SP1346			SP1346		
		SP1347		SP1347			SP1347		
		SP1348		SP1348			SP1348		
				SP1349			SP1349		
				SP1350			SP1350		
			SP1351	SP1351					
		SP1353							
SP1433								SP1433	
SP1434			SP1434				SP1434	SP1434	
SP1435			SP1435				SP1435	SP1435	
SP1436			SP1436				SP1436	SP1436	
SP1437			SP1437				SP1437	SP1437	
SP1438			SP1438				SP1438	SP1438	
					SP1454				
	SP1611								SP1611
SP1612									
SP1615	SP1615		SP1615		SP1615	SP1615	SP1615	SP1615	
SP1616	SP1616				SP1616	SP1616	SP1616	SP1616	
SP1617	SP1617		SP1617		SP1617	SP1617	SP1617	SP1617	
SP1618	SP1618		SP1618		SP1618	SP1618	SP1618	SP1618	
SP1619	SP1619		SP1619		SP1619	SP1619	SP1619	SP1619	
SP1620	SP1620		SP1620		SP1620	SP1620	SP1620	SP1620	
SP1621	SP1621		SP1621		SP1621	SP1621	SP1621	SP1621	
SP1622	SP1622				SP1622	SP1622		SP1622	
					SP1718				
	SP1755	SP1755							
SP1756	SP1756			SP1756					
SP1757				SP1757		SP1757			
SP1758	SP1758	SP1758		SP1758	SP1758	SP1758	SP1758		SP1758
SP1759	SP1759	SP1759		SP1759	SP1759	SP1759	SP1759		SP1759
SP1760	SP1760	SP1760		SP1760	SP1760	SP1760	SP1760		SP1760
SP1761	SP1761	SP1761		SP1761	SP1761	SP1761	SP1761		SP1761
SP1762	SP1762	SP1762		SP1762	SP1762	SP1762	SP1762		SP1762
SP1763	SP1763	SP1763		SP1763	SP1763	SP1763	SP1763		SP1763
SP1764	SP1764	SP1764		SP1764	SP1764	SP1764	SP1764		SP1764
SP1765	SP1765	SP1765		SP1765	SP1765	SP1765	SP1765		SP1765
SP1766	SP1766	SP1766		SP1766	SP1766	SP1766	SP1766		SP1766
SP1767	SP1767	SP1767		SP1767	SP1767	SP1767	SP1767		SP1767
SP1768				SP1768					
SP1769	SP1769	SP1769		SP1769	SP1769	SP1769	SP1769		SP1769
SP1770	SP1770	SP1770		SP1770	SP1770	SP1770	SP1770		SP1770
SP1771	SP1771	SP1771		SP1771	SP1771	SP1771	SP1771		SP1771
SP1772	SP1772	SP1772		SP1772	SP1772	SP1772	SP1772		SP1772
SP1773	SP1773	SP1773		SP1773	SP1773	SP1773	SP1773		SP1773
								SP1829	
	SP1836								
						SP1935			
			SP1948			SP1948		SP1948	SP1948
			SP1949			SP1949		SP1949	SP1949
		SP1950	SP1950			SP1950	SP1950	SP1950	SP1950
		SP1951	SP1951			SP1951	SP1951	SP1951	SP1951
		SP1952	SP1952			SP1952	SP1952	SP1952	SP1952
		SP1953	SP1953			SP1953		SP1953	SP1953
			SP1954			SP1954		SP1954	SP1954
			SP1955						SP1955

BAA659	BAA340	05_1271	ATCC700904	BAA660	05_2565	05_1821	ATCC51916	07_2839	OXC141
Serotype 6A	Serotype 14	Serotype 20	Serotype 19A	Serotype 35B	Serotype 12F	Serotype 9V	Serotype 23F	Serotype 9A	Serotype 3
ST376	ST67	ST568	ST41	ST377	ST218	ST156	ST37	ST239	ST180
USA	USA	Scotland	South Africa	USA	Scotland	Scotland	USA	Bolivia	England
				SP2159					SP2159
				SP2160					SP2160
				SP2161					SP2161
				SP2162					SP2162
				SP2163					SP2163
				SP2164					SP2164
				SP2165					
				SP2166					
								SP2180	
								SP2181	
								SP2182	
SP2237									

Appendix 8 List of genes from the TIGR4 genome which are present or non hybridizing in 10 strains of diverse serotype and diverse MLST from diverse geographical regions.

Gene lists generated by Genespring GX 7.3.1 for each strain indicating regions of differential hybridization between the test strains and TIGR4. Yellow indicates hybridization by both TIGR4 DNA and the test strain DNA. Blue indicates the absence of hybridization by test strain DNA and unopposed hybridization by TIGR4 DNA. Genes whose names appear in white have been identified as required for virulence in a mouse bacteraemic pneumonia model (Hava and Camilli, 2002).

Appendix 9

BAA659	BAA340	05_1271	ATCC700904	BAA660	05_2565	05_1821	ATCC51916	07_2839	OXC141
Serotype 6A	Serotype 14	Serotype 20	Serotype 19A	Serotype 35B	Serotype 12F	Serotype 9V	Serotype 23F	Serotype 9A	Serotype 3
ST376	ST67	ST568	ST41	ST377	ST218	ST156	ST37	ST239	ST180
USA	USA	Scotland	South Africa	USA	Scotland	Scotland	USA	Bolivia	England
		spr0067					spr0067		
	spr0102		spr0102					spr0102	
	spr0103		spr0103			spr0103		spr0103	
	spr0104		spr0104			spr0104		spr0104	spr0104
spr0105	spr0105		spr0105		spr0105	spr0105		spr0105	spr0105
			spr0106		spr0106	spr0106		spr0106	spr0106
	spr0107		spr0107		spr0107	spr0107		spr0107	spr0107
	spr0108		spr0108		spr0108	spr0108		spr0108	spr0108
			spr0111		spr0111		spr0111	spr0111	spr0111
spr0112			spr0112	spr0112	spr0112	spr0112	spr0112	spr0112	spr0112
spr0113			spr0113	spr0113	spr0113	spr0113	spr0113	spr0113	spr0113
spr0114			spr0114	spr0114		spr0114		spr0114	
spr0115		spr0115	spr0115	spr0115	spr0115	spr0115	spr0115	spr0115	spr0115
spr0116			spr0116	spr0116	spr0116	spr0116	spr0116	spr0116	spr0116
spr0117	spr0117		spr0117	spr0117	spr0117	spr0117	spr0117	spr0117	spr0117
spr0118	spr0118		spr0118	spr0118	spr0118	spr0118	spr0118	spr0118	spr0118
spr0119	spr0119		spr0119	spr0119	spr0119	spr0119	spr0119	spr0119	spr0119
									spr0225
		spr0320	spr0320				spr0320		
spr0321		spr0321	spr0321				spr0321	spr0321	
spr0322		spr0322					spr0322		
spr0323		spr0323					spr0323		
							spr0491		spr0491
	spr0493								
	spr0955		spr0955		spr0955	spr0955		spr0955	spr0955
	spr0956	spr0956	spr0956		spr0956	spr0956	spr0956	spr0956	spr0956
	spr0957	spr0957	spr0957		spr0957	spr0957	spr0957	spr0957	spr0957
	spr0958		spr0958		spr0958	spr0958		spr0958	spr0958
	spr0959	spr0959			spr0959		spr0959		spr0959
	spr0960		spr0960		spr0960		spr0960		
spr0964									
	spr0968								
spr0970									
					spr1026				
spr1114							spr1114		spr1114
							spr1179		
								spr1189	
								spr1190	
								spr1191	
								spr1192	
								spr1193	
								spr1194	
								spr1195	
						spr1199			
	spr1403			spr1403			spr1403	spr1403	
	spr1404					spr1404	spr1404		
								spr1478	
	spr1484	spr1484							
			spr1549						
			spr1550						
							spr1618		
					spr1619		spr1619	spr1619	
					spr1620		spr1620		
					spr1984			spr1984	

Appendix 9 List of genes from the R6 genome which are present or non hybridizing in 10 strains of diverse serotype and diverse MLST from diverse geographical regions.

Gene lists generated by Genespring GX 7.3.1 for each strain indicating regions where genes present in both the test isolates and the R6 genome have hybridized. Red indicates hybridization by the test strain DNA.

Appendix 10

ST9 Serotype 14					Other Serotype ST9s				
INV 200	02_2445	00_1724	01_5710	04_1870	02_1309	04_1168	04_1548	03_1051	03_2105
						SP0040			
SP0067	SP0067	SP0067		SP0067	SP0067		SP0067		SP0067
SP0068	SP0068	SP0068	SP0068	SP0068	SP0068	SP0068	SP0068	SP0068	SP0068
SP0069	SP0069	SP0069	SP0069	SP0069	SP0069	SP0069	SP0069	SP0069	SP0069
SP0070	SP0070	SP0070	SP0070	SP0070	SP0070	SP0070	SP0070	SP0070	SP0070
SP0071	SP0071	SP0071	SP0071	SP0071	SP0071	SP0071	SP0071	SP0071	SP0071
	SP0072	SP0072				SP0072		SP0072	
	SP0073	SP0073						SP0073	
	SP0074	SP0074						SP0074	
		SP0080							
					SP0089				
									SP0093
SP0094			SP0094						
								SP0110	
								SP0111	
								SP0112	
								SP0113	
								SP0128	
									SP0132
SP0163		SP0163	SP0163	SP0163	SP0163				
SP0164	SP0164								
SP0165	SP0165	SP0165	SP0165	SP0165			SP0165	SP0165	
SP0166	SP0166	SP0166	SP0166	SP0166		SP0166	SP0166	SP0166	
SP0167	SP0167	SP0167	SP0167	SP0167	SP0167	SP0167	SP0167	SP0167	SP0167
SP0168	SP0168	SP0168	SP0168	SP0168	SP0168	SP0168	SP0168	SP0168	SP0168
		SP0171		SP0171			SP0171		
			SP0190						
			SP0233						
			SP0270						
SP0303	SP0303	SP0303		SP0303			SP0303	SP0303	
SP0305	SP0305	SP0305	SP0305	SP0305	SP0305		SP0305	SP0305	SP0305
SP0306	SP0306			SP0306			SP0306		
	SP0307	SP0307		SP0307	SP0307		SP0307	SP0307	
			SP0309				SP0309		
	SP0310	SP0310	SP0310				SP0310		
					SP0319				
									SP0335
								SP0348	
								SP0349	
	SP0350	SP0350			SP0350	SP0350	SP0350	SP0350	SP0350
	SP0351	SP0351	SP0351	SP0351		SP0351	SP0351	SP0351	SP0351
	SP0352	SP0352	SP0352	SP0352		SP0352	SP0352	SP0352	SP0352
SP0353	SP0353	SP0353		SP0353	SP0353	SP0353	SP0353	SP0353	SP0353
SP0354	SP0354	SP0354	SP0354	SP0354	SP0354	SP0354	SP0354	SP0354	SP0354
	SP0355	SP0355	SP0355	SP0355	SP0355	SP0355	SP0355	SP0355	SP0355
SP0356	SP0356	SP0356		SP0356		SP0356	SP0356	SP0356	SP0356
	SP0357	SP0357	SP0357	SP0357	SP0357	SP0357	SP0357	SP0357	SP0357
	SP0358	SP0358	SP0358	SP0358	SP0358	SP0358	SP0358	SP0358	SP0358
	SP0359	SP0359	SP0359	SP0359	SP0359	SP0359	SP0359	SP0359	SP0359
	SP0360	SP0360	SP0360	SP0360	SP0360	SP0360	SP0360	SP0360	SP0360
						SP0361			
							SP0380		
							SP0394	SP0394	
							SP0395	SP0395	
							SP0396	SP0396	
							SP0397	SP0397	
				SP0398					
			SP0431		SP0431				
	SP0460	SP0460	SP0460	SP0460	SP0460	SP0460		SP0460	SP0460
	SP0461	SP0461	SP0461	SP0461	SP0461	SP0461		SP0461	SP0461
	SP0462	SP0462	SP0462	SP0462	SP0462	SP0462		SP0462	SP0462
	SP0463	SP0463	SP0463	SP0463	SP0463	SP0463		SP0463	SP0463
SP0464	SP0464	SP0464	SP0464	SP0464	SP0464	SP0464		SP0464	SP0464
SP0465	SP0465	SP0465	SP0465	SP0465	SP0465	SP0465		SP0465	SP0465
SP0466	SP0466	SP0466	SP0466	SP0466	SP0466	SP0466		SP0466	SP0466
SP0467	SP0467	SP0467	SP0467	SP0467	SP0467	SP0467		SP0467	SP0467
	SP0468	SP0468	SP0468	SP0468	SP0468	SP0468		SP0468	SP0468
								SP0469	
SP0473	SP0473	SP0473	SP0473	SP0473	SP0473			SP0473	SP0473
SP0474	SP0474	SP0474	SP0474	SP0474	SP0474			SP0474	SP0474
SP0475	SP0475	SP0475	SP0475	SP0475	SP0475			SP0475	SP0475
SP0476	SP0476	SP0476	SP0476	SP0476	SP0476			SP0476	SP0476
SP0477	SP0477	SP0477	SP0477	SP0477	SP0477			SP0477	SP0477
SP0478	SP0478	SP0478		SP0478					
			SP0511						
					SP0512				SP0512
					SP0534				
					SP0535				
					SP0548				SP0548
			SP0560						
SP0570									
					SP0573				
							SP0574		
							SP0575		
					SP0584				
						SP0600			
					SP0602				
									SP0696
			SP0716						
					SP0752				
SP0773									
			SP0808						
					SP0853				
								SP0907	
									SP0949
	SP0951	SP0951							
					SP0952				SP0952
		SP0957							
							SP1052		
							SP1053		

ST9 Serotype 14					Other Serotype ST9s				
INV 200	02_2445	00_1724	01_5710	04_1870	02_1309	04_1168	04_1548	03_1051	03_2105
SP1054	SP1054	SP1054	SP1054	SP1054	SP1054	SP1054	SP1054	SP1054	SP1054
SP1055	SP1055	SP1055	SP1055	SP1055	SP1055	SP1055	SP1055	SP1055	SP1055
SP1056	SP1056	SP1056	SP1056	SP1056		SP1056	SP1056	SP1056	SP1056
SP1057	SP1057	SP1057	SP1057	SP1057	SP1057	SP1057	SP1057	SP1057	SP1057
SP1059	SP1059	SP1059	SP1059	SP1059	SP1059	SP1059	SP1059	SP1059	SP1059
SP1060	SP1060	SP1060	SP1060	SP1060	SP1060	SP1060	SP1060	SP1060	SP1060
SP1061	SP1061	SP1061	SP1061	SP1061	SP1061	SP1061	SP1061	SP1061	SP1061
SP1062	SP1062	SP1062	SP1062	SP1062	SP1062	SP1062	SP1062	SP1062	SP1062
SP1063	SP1063	SP1063	SP1063	SP1063	SP1063	SP1063	SP1063	SP1063	SP1063
	SP1064	SP1064		SP1064		SP1064	SP1064	SP1064	SP1064
		SP1065					SP1065		
			SP1104						
SP1129	SP1129	SP1129	SP1129	SP1129	SP1129	SP1129	SP1129	SP1129	SP1129
SP1130	SP1130	SP1130	SP1130	SP1130	SP1130	SP1130	SP1130	SP1130	SP1130
SP1131	SP1131	SP1131	SP1131	SP1131	SP1131	SP1131	SP1131	SP1131	SP1131
SP1132	SP1132	SP1132	SP1132	SP1132	SP1132	SP1132	SP1132	SP1132	SP1132
SP1133	SP1133	SP1133	SP1133	SP1133		SP1133	SP1133	SP1133	SP1133
SP1134	SP1134	SP1134	SP1134	SP1134	SP1134	SP1134		SP1134	SP1134
SP1135	SP1135	SP1135	SP1135	SP1135	SP1135	SP1135		SP1135	SP1135
SP1136	SP1136	SP1136	SP1136	SP1136	SP1136	SP1136	SP1136	SP1136	SP1136
SP1137	SP1137	SP1137	SP1137	SP1137	SP1137	SP1137	SP1137	SP1137	SP1137
	SP1138	SP1138				SP1138	SP1138	SP1138	
SP1139	SP1139	SP1139	SP1139	SP1139	SP1139	SP1139	SP1139	SP1139	SP1139
SP1140	SP1140	SP1140	SP1140	SP1140	SP1140	SP1140	SP1140	SP1140	SP1140
SP1141	SP1141	SP1141	SP1141	SP1141	SP1141	SP1141	SP1141	SP1141	SP1141
SP1142	SP1142	SP1142	SP1142	SP1142	SP1142	SP1142	SP1142	SP1142	SP1142
SP1143	SP1143	SP1143	SP1143	SP1143	SP1143	SP1143	SP1143	SP1143	SP1143
SP1144	SP1144	SP1144	SP1144	SP1144	SP1144	SP1144	SP1144	SP1144	SP1144
SP1145	SP1145	SP1145	SP1145	SP1145	SP1145	SP1145	SP1145	SP1145	SP1145
						SP1146			SP1146
SP1147	SP1147	SP1147	SP1147	SP1147					SP1147
SP1154	SP1154		SP1154	SP1154	SP1154		SP1154	SP1154	SP1154
SP1181									SP1181
			SP1211						
SP1221	SP1221	SP1221	SP1221	SP1221	SP1221	SP1221	SP1221		SP1221
SP1222	SP1222	SP1222	SP1222	SP1222	SP1222	SP1222	SP1222		SP1222
			SP1265						
		SP1307							
		SP1308							
						SP1315		SP1315	
						SP1316		SP1316	
						SP1317		SP1317	
						SP1318		SP1318	
						SP1319		SP1319	
						SP1320		SP1320	
						SP1321		SP1321	
						SP1322		SP1322	
						SP1323		SP1323	
						SP1324		SP1324	
						SP1325		SP1325	
						SP1326		SP1326	
						SP1327		SP1327	
						SP1328		SP1328	
						SP1329		SP1329	
						SP1330		SP1330	
						SP1331		SP1331	
						SP1332		SP1332	
						SP1333		SP1333	
		SP1334				SP1334		SP1334	
SP1335	SP1335	SP1335	SP1335	SP1335		SP1335		SP1335	SP1335
SP1336	SP1336	SP1336	SP1336	SP1336	SP1336	SP1336		SP1336	SP1336
						SP1337			
	SP1338	SP1338	SP1338		SP1338	SP1338			
SP1339	SP1339	SP1339	SP1339	SP1339	SP1339	SP1339			SP1339
SP1340	SP1340	SP1340	SP1340	SP1340	SP1340	SP1340			SP1340
SP1341	SP1341	SP1341	SP1341	SP1341	SP1341	SP1341			SP1341
SP1342	SP1342	SP1342	SP1342	SP1342	SP1342	SP1342			SP1342
SP1343	SP1343	SP1343	SP1343	SP1343	SP1343	SP1343			SP1343
SP1344	SP1344	SP1344	SP1344	SP1344	SP1344	SP1344			SP1344
SP1345	SP1345	SP1345	SP1345	SP1345		SP1345			
	SP1346	SP1346				SP1346			
	SP1347	SP1347				SP1347		SP1347	
SP1348	SP1348	SP1348				SP1348			
SP1349	SP1349	SP1349				SP1349			
SP1350	SP1350	SP1350			SP1350	SP1350			SP1350
		SP1351							
						SP1434			
						SP1435			
						SP1436			
						SP1437			
						SP1438			
					SP1573				
			SP1581						
						SP1611			
							SP1612		
SP1615	SP1615	SP1615	SP1615	SP1615	SP1615	SP1615	SP1615	SP1615	SP1615
SP1616	SP1616	SP1616	SP1616	SP1616	SP1616	SP1616	SP1616	SP1616	SP1616
SP1617	SP1617	SP1617	SP1617	SP1617	SP1617	SP1617	SP1617	SP1617	SP1617
SP1618	SP1618	SP1618	SP1618	SP1618	SP1618	SP1618	SP1618	SP1618	SP1618
SP1619	SP1619	SP1619	SP1619	SP1619	SP1619	SP1619	SP1619	SP1619	SP1619
SP1620	SP1620	SP1620	SP1620	SP1620	SP1620	SP1620	SP1620	SP1620	SP1620
SP1621	SP1621	SP1621	SP1621	SP1621	SP1621	SP1621	SP1621	SP1621	SP1621
	SP1622	SP1622	SP1622	SP1622	SP1622		SP1622	SP1622	SP1622
							SP1640		
									SP1722
			SP1754						
							SP1756		
						SP1757	SP1757		
						SP1758	SP1758		
						SP1759	SP1759		

ST9 Serotype 14					Other Serotype ST9s				
INV 200	02_2445	00_1724	01_5710	04_1870	02_1309	04_1168	04_1548	03_1051	03_2105
SP1760	SP1760	SP1760	SP1760	SP1760	SP1760	SP1760	SP1760		SP1760
SP1761	SP1761	SP1761	SP1761	SP1761	SP1761	SP1761	SP1761		SP1761
SP1762	SP1762	SP1762	SP1762	SP1762	SP1762	SP1762	SP1762		SP1762
SP1763	SP1763	SP1763	SP1763	SP1763	SP1763	SP1763	SP1763		SP1763
SP1764	SP1764	SP1764	SP1764	SP1764	SP1764	SP1764	SP1764		SP1764
SP1765	SP1765	SP1765	SP1765	SP1765	SP1766	SP1765	SP1765		SP1765
SP1766	SP1766	SP1766	SP1766	SP1766		SP1766	SP1766		SP1766
SP1767	SP1767	SP1767	SP1767	SP1767	SP1767	SP1767	SP1767		SP1767
SP1768				SP1768			SP1768		SP1768
SP1769	SP1769	SP1769	SP1769	SP1769	SP1769	SP1769	SP1769		SP1769
SP1770	SP1770	SP1770	SP1770	SP1770	SP1770	SP1770	SP1770		SP1770
SP1771	SP1771	SP1771	SP1771	SP1771	SP1771	SP1771	SP1771		SP1771
						SP1772	SP1772		
		SP1793							
		SP1794							
						SP1818			
		SP1828					SP1828		
SP1829		SP1829							
		SP1830							
			SP1831						
SP1834				SP1834					
								SP1835	
			SP1886						
							SP1931		
							SP1933		
							SP1934		
							SP1935		
							SP1936		
	SP1948	SP1948	SP1948	SP1948			SP1948		SP1948
	SP1949	SP1949	SP1949	SP1949			SP1949	SP1949	SP1949
SP1950	SP1950	SP1950	SP1950	SP1950	SP1950	SP1950	SP1950	SP1950	SP1950
SP1951	SP1951	SP1951	SP1951	SP1951	SP1951	SP1951	SP1951	SP1951	SP1951
SP1952	SP1952	SP1952	SP1952	SP1952	SP1952	SP1952	SP1952	SP1952	SP1952
SP1953	SP1953	SP1953	SP1953	SP1953	SP1953	SP1953	SP1953	SP1953	SP1953
SP1954	SP1954	SP1954	SP1954	SP1954	SP1954		SP1954	SP1954	SP1954
					SP1955			SP1955	
				SP2018					
					SP2115				
			SP2124						

Appendix 10 List of genes from the TIGR4 genome which are present or non hybridizing in 10 ST9 strains.

Gene lists generated by Genespring GX 7.3.1 for each strain indicating regions of differential hybridisation between the test strains and TIGR4. Yellow indicates hybridisation by both TIGR4 DNA and the test strain DNA. Blue indicates the absence of hybridization by test strain DNA and unopposed hybridization by TIGR4 DNA. Genes whose names appear in white have been identified as required for virulence in a mouse bacteraemic pneumonia model (Hava and Camilli, 2002).

Appendix 11

ST9 Serotype 14					Other Serotype ST9s				
INV 200	02_2445	00_1724	01_5710	04_1870	02_1309	04_1168	04_1548	03_1051	03_2105
			spr0015						
								spr0067	
			spr0104	spr0104					
spr0105	spr0105	spr0105	spr0105	spr0105	spr0105		spr0105		spr0105
							spr0106		
							spr0107		
spr0111	spr0111	spr0111	spr0111	spr0111	spr0111				spr0111
spr0112	spr0112	spr0112		spr0112		spr0112	spr0112	spr0112	spr0112
spr0113	spr0113	spr0113		spr0113		spr0113	spr0113		spr0113
spr0114	spr0114	spr0114		spr0114		spr0114	spr0114		spr0114
spr0115	spr0115	spr0115	spr0115	spr0115	spr0115	spr0115	spr0115	spr0115	spr0115
spr0116	spr0116	spr0116	spr0116	spr0116	spr0116	spr0116	spr0116	spr0116	spr0116
spr0117	spr0117	spr0117		spr0117		spr0117	spr0117		
spr0118	spr0118	spr0118	spr0118	spr0118	spr0118	spr0118	spr0118	spr0118	spr0118
spr0119	spr0119		spr0119	spr0119		spr0119	spr0119	spr0119	spr0119
								spr0225	
						spr0320		spr0320	
						spr0321	spr0321		
						spr0322	spr0322		
						spr0323		spr0323	
spr0416								spr0416	
	spr0470								
								spr0491	
								spr0493	
							spr0681		
								spr0703	
spr0800			spr0800				spr0800		
spr0956	spr0956	spr0956	spr0956	spr0956		spr0956	spr0956	spr0956	spr0956
spr0957	spr0957	spr0957	spr0957	spr0957	spr0957	spr0957	spr0957	spr0957	spr0957
spr0958	spr0958	spr0958		spr0958			spr0958		spr0958
						spr0959		spr0959	
						spr0960		spr0960	
						spr0972			
									spr1042
spr1093				spr1093				spr1093	
								spr1114	
								spr1178	
spr1179	spr1179							spr1179	spr1179
						spr1193			
					spr1283				
						spr1403		spr1403	
			spr1404			spr1404		spr1404	
			spr1481						
								spr1484	
								spr1549	
					spr1550			spr1550	
						spr1619			spr1619
						spr1620			
							spr1752		
	spr1755								
								spr1817	
	spr1818					spr1818			

Appendix 11 List of genes from the R6 genome which are present or non hybridizing in 10 ST9 strains.

Gene lists generated by Genespring GX 7.3.1 for each strain indicating regions where genes present in both the test isolates and the R6 genome have hybridized. Red indicates hybridization by the test strain DNA.

Appendix 12

Invasive ST180 Serotype 3					Carriage ST180 Serotype 3					
99-4038	99-4039	00-3946	06-1705	02-1198	OXC141	03-4183	03-4155	03-4156	03-4185	07-2838
										SP0016
										SP0017
SP0040										SP0040
SP0067	SP0067		SP0067		SP0067	SP0067	SP0067	SP0067	SP0067	
SP0068	SP0068	SP0068	SP0068	SP0068	SP0068	SP0068	SP0068	SP0068	SP0068	SP0068
SP0069	SP0069	SP0069	SP0069	SP0069	SP0069	SP0069	SP0069	SP0069	SP0069	SP0069
SP0070	SP0070	SP0070	SP0070	SP0070	SP0070	SP0070	SP0070	SP0070	SP0070	
SP0071	SP0071	SP0071	SP0071	SP0071	SP0071	SP0071	SP0071	SP0071	SP0071	SP0071
SP0072						SP0072				
SP0073										
SP0074										
									SP0076	
SP0080										
								SP0094		
SP0110	SP0110		SP0110						SP0110	
SP0111	SP0111		SP0111						SP0111	
SP0112	SP0112		SP0112	SP0112					SP0112	
SP0113	SP0113		SP0113	SP0113					SP0113	
	SP0114							SP0114		
	SP0115							SP0115		
					SP0116					
	SP0143									
				SP0160						
				SP0161						
				SP0162						
	SP0163			SP0163	SP0163	SP0163				
SP0164	SP0164			SP0164		SP0164				
SP0165	SP0165	SP0165		SP0165	SP0165	SP0165		SP0165	SP0165	
SP0166	SP0166	SP0166		SP0166	SP0166	SP0166	SP0166	SP0166	SP0166	
SP0167	SP0167	SP0167	SP0167	SP0167	SP0167	SP0167	SP0167	SP0167	SP0167	SP0167
SP0168	SP0168	SP0168	SP0168	SP0168	SP0168	SP0168	SP0168	SP0168	SP0168	SP0168
	SP0171									SP0171
										SP0177
			SP0189							
										SP0192
									SP0193	
					SP0196					
	SP0220									
				SP0223						
							SP0244			
									SP0304	
										SP0331
	SP0335									
	SP0347									SP0347
	SP0350	SP0350			SP0350	SP0350	SP0350	SP0350		
SP0351	SP0351	SP0351		SP0351	SP0351	SP0351	SP0351	SP0351	SP0351	
SP0352	SP0352	SP0352	SP0352	SP0352	SP0352	SP0352	SP0352	SP0352	SP0352	
SP0353	SP0353	SP0353	SP0353	SP0353	SP0353	SP0353	SP0353	SP0353	SP0353	
SP0354	SP0354	SP0354	SP0354	SP0354	SP0354	SP0354	SP0354	SP0354	SP0354	
SP0355	SP0355	SP0355	SP0355	SP0355	SP0355	SP0355	SP0355	SP0355	SP0355	
SP0356	SP0356	SP0356	SP0356	SP0356	SP0356	SP0356	SP0356	SP0356	SP0356	
SP0357	SP0357	SP0357	SP0357	SP0357	SP0357	SP0357	SP0357	SP0357	SP0357	
SP0358	SP0358	SP0358	SP0358	SP0358	SP0358	SP0358	SP0358	SP0358	SP0358	
SP0359	SP0359	SP0359	SP0359	SP0359	SP0359	SP0359	SP0359	SP0359	SP0359	SP0359
SP0360	SP0360	SP0360	SP0360	SP0360	SP0360	SP0360	SP0360	SP0360	SP0360	SP0360
	SP0361									
	SP0394					SP0394	SP0394	SP0394		SP0394
	SP0395	SP0395	SP0395	SP0395		SP0395	SP0395	SP0395	SP0395	SP0395
SP0396	SP0396	SP0396	SP0396	SP0396	SP0396	SP0396	SP0396	SP0396	SP0396	SP0396
SP0397	SP0397					SP0397		SP0397	SP0397	
SP0460	SP0460	SP0460	SP0460	SP0460	SP0460	SP0460	SP0460	SP0460	SP0460	SP0460
SP0461	SP0461	SP0461	SP0461	SP0461	SP0461	SP0461	SP0461	SP0461	SP0461	SP0461
SP0462	SP0462	SP0462	SP0462	SP0462	SP0462	SP0462	SP0462	SP0462	SP0462	SP0462
SP0463	SP0463	SP0463	SP0463	SP0463	SP0463	SP0463	SP0463	SP0463	SP0463	SP0463
SP0464	SP0464	SP0464	SP0464	SP0464	SP0464	SP0464	SP0464	SP0464	SP0464	SP0464
SP0465	SP0465	SP0465	SP0465	SP0465	SP0465	SP0465	SP0465	SP0465	SP0465	
SP0466	SP0466	SP0466	SP0466	SP0466	SP0466	SP0466	SP0466	SP0466	SP0466	SP0466
SP0467	SP0467	SP0467	SP0467	SP0467	SP0467	SP0467	SP0467	SP0467	SP0467	
SP0468	SP0468	SP0468	SP0468	SP0468	SP0468	SP0468	SP0468	SP0468	SP0468	SP0468
SP0473	SP0473	SP0473			SP0473	SP0473	SP0473		SP0473	
SP0474	SP0474	SP0474	SP0474	SP0474	SP0474	SP0474	SP0474		SP0474	
SP0475	SP0475	SP0475	SP0475	SP0475	SP0475	SP0475	SP0475		SP0475	
SP0476	SP0476	SP0476	SP0476	SP0476	SP0476	SP0476	SP0476		SP0476	
SP0477	SP0477	SP0477	SP0477	SP0477	SP0477	SP0477	SP0477		SP0477	
SP0478	SP0478	SP0478		SP0478	SP0478	SP0478	SP0478			
										SP0497
										SP0512
	SP0534									
										SP0541
	SP0573									
		SP0574						SP0574	SP0574	SP0574
					SP0575	SP0575				SP0575
			SP0594							
SP0640										
		SP0684								
	SP0694									
								SP0697		
	SP0698			SP0698						
	SP0749		SP0749							
		SP0752								
									SP0773	
	SP0821									
										SP0853
										SP0866
		SP0871								
		SP0908								
										SP0913
	SP0915									
			SP1005							
										SP1025

Invasive ST180 Serotype 3					Carriage ST180 Serotype 3					
99-4038	99-4039	00-3946	06-1705	02-1198	OXC141	03-4183	03-4155	03-4156	03-4185	07-2838
					SP1031					
							SP1037			
	SP1054	SP1054		SP1054			SP1054	SP1054		
	SP1055	SP1055		SP1055		SP1055	SP1055	SP1055		
SP1056	SP1056	SP1056		SP1056		SP1056	SP1056	SP1056		
SP1057	SP1057	SP1057	SP1057	SP1057	SP1057	SP1057	SP1057	SP1057	SP1057	SP1057
						SP1058				
SP1059	SP1059	SP1059	SP1059	SP1059	SP1059	SP1059	SP1059	SP1059	SP1059	SP1059
SP1060	SP1060	SP1060	SP1060	SP1060	SP1060	SP1060	SP1060	SP1060	SP1060	SP1060
SP1061	SP1061	SP1061	SP1061	SP1061	SP1061	SP1061	SP1061	SP1061	SP1061	SP1061
SP1062	SP1062	SP1062	SP1062	SP1062	SP1062	SP1062	SP1062	SP1062	SP1062	SP1062
SP1063	SP1063	SP1063	SP1063	SP1063		SP1063	SP1063	SP1063	SP1063	SP1063
	SP1064	SP1064		SP1064		SP1064	SP1064	SP1064	SP1064	
			SP1113							
										SP1118
SP1129	SP1129	SP1129	SP1129	SP1129	SP1129	SP1129	SP1129	SP1129	SP1129	SP1129
SP1130	SP1130	SP1130	SP1130	SP1130	SP1130	SP1130	SP1130	SP1130	SP1130	SP1130
SP1131	SP1131	SP1131	SP1131	SP1131	SP1131	SP1131	SP1131	SP1131	SP1131	SP1131
SP1132	SP1132	SP1132		SP1132	SP1132	SP1132	SP1132	SP1132	SP1132	SP1132
SP1133	SP1133	SP1133		SP1133	SP1133	SP1133	SP1133	SP1133	SP1133	SP1133
SP1134	SP1134	SP1134	SP1134	SP1134	SP1134	SP1134	SP1134	SP1134	SP1134	SP1134
SP1135	SP1135	SP1135	SP1135	SP1135	SP1135	SP1135	SP1135	SP1135	SP1135	SP1135
SP1136	SP1136	SP1136	SP1136	SP1136	SP1136	SP1136	SP1136	SP1136	SP1136	SP1136
SP1137	SP1137	SP1137	SP1137	SP1137	SP1137	SP1137	SP1137	SP1137	SP1137	SP1137
SP1138	SP1138	SP1138		SP1138		SP1138			SP1138	
SP1139	SP1139	SP1139	SP1139	SP1139	SP1139	SP1139	SP1139	SP1139	SP1139	SP1139
SP1140	SP1140	SP1140	SP1140	SP1140	SP1140	SP1140	SP1140	SP1140	SP1140	SP1140
SP1141	SP1141	SP1141	SP1141	SP1141	SP1141	SP1141	SP1141	SP1141	SP1141	SP1141
SP1142	SP1142	SP1142	SP1142	SP1142	SP1142	SP1142	SP1142	SP1142	SP1142	SP1142
SP1143	SP1143	SP1143	SP1143	SP1143	SP1143	SP1143	SP1143	SP1143	SP1143	SP1143
SP1144	SP1144	SP1144	SP1144	SP1144	SP1144	SP1144	SP1144	SP1144	SP1144	
SP1145	SP1145	SP1145	SP1145	SP1145	SP1145	SP1145	SP1145	SP1145	SP1145	SP1145
						SP1146				
										SP1150
		SP1158					SP1158			
SP1221	SP1221	SP1221	SP1221	SP1221		SP1221	SP1221	SP1221	SP1221	SP1221
SP1222	SP1222	SP1222	SP1222	SP1222			SP1222	SP1222	SP1222	SP1222
								SP1234		
SP1265							SP1265			
	SP1304									
SP1315	SP1315	SP1315	SP1315	SP1315	SP1315	SP1315	SP1315	SP1315	SP1315	SP1315
SP1316	SP1316	SP1316	SP1316	SP1316	SP1316	SP1316	SP1316	SP1316	SP1316	SP1316
SP1317	SP1317	SP1317	SP1317	SP1317	SP1317	SP1317	SP1317	SP1317	SP1317	SP1317
SP1318	SP1318	SP1318	SP1318	SP1318	SP1318	SP1318	SP1318	SP1318	SP1318	SP1318
SP1319	SP1319	SP1319	SP1319	SP1319	SP1319	SP1319	SP1319	SP1319	SP1319	SP1319
SP1320	SP1320	SP1320	SP1320	SP1320	SP1320	SP1320	SP1320	SP1320	SP1320	SP1320
SP1321	SP1321	SP1321	SP1321	SP1321	SP1321	SP1321	SP1321	SP1321	SP1321	SP1321
SP1322	SP1322	SP1322	SP1322	SP1322	SP1322	SP1322	SP1322	SP1322	SP1322	SP1322
SP1323	SP1323	SP1323	SP1323	SP1324	SP1323	SP1323	SP1323	SP1323	SP1323	SP1323
SP1324	SP1324	SP1324	SP1324	SP1325	SP1324	SP1324	SP1324	SP1324	SP1324	SP1324
SP1325	SP1325	SP1325	SP1325		SP1325	SP1325	SP1325	SP1325	SP1325	SP1325
SP1326	SP1326	SP1326	SP1326	SP1326	SP1326	SP1326	SP1326	SP1326	SP1326	SP1326
SP1327	SP1327	SP1327	SP1327	SP1327	SP1327	SP1327	SP1327	SP1327	SP1327	SP1327
SP1328	SP1328	SP1328	SP1328	SP1328	SP1328	SP1328	SP1328	SP1328	SP1328	SP1328
SP1329	SP1329	SP1329	SP1329	SP1329	SP1329	SP1329	SP1329	SP1329	SP1329	SP1329
SP1330	SP1330	SP1330	SP1330	SP1330	SP1330	SP1330	SP1330	SP1330	SP1330	SP1330
SP1331	SP1331	SP1331	SP1331	SP1331	SP1331	SP1331	SP1331	SP1331	SP1331	SP1331
	SP1332	SP1332				SP1332				
	SP1333	SP1333				SP1333				
	SP1334	SP1334				SP1334				
	SP1335	SP1335	SP1335	SP1335		SP1335				
SP1336	SP1336	SP1336	SP1336	SP1336		SP1336		SP1336		
										SP1340
										SP1341
										SP1342
SP1343	SP1343	SP1343	SP1343	SP1343	SP1343	SP1343	SP1343	SP1343	SP1343	SP1343
										SP1344
									SP1579	
										SP1585
					SP1611					
				SP1718						
		SP1757		SP1757		SP1757				
SP1758	SP1758	SP1758	SP1758	SP1758	SP1758	SP1758	SP1758	SP1758	SP1758	SP1758
SP1759	SP1759	SP1759	SP1759	SP1759	SP1759	SP1759	SP1759	SP1759	SP1759	SP1759
SP1760	SP1760	SP1760	SP1760	SP1760	SP1760	SP1760	SP1760	SP1760	SP1760	SP1760
SP1761	SP1761	SP1761	SP1761	SP1761	SP1761	SP1761	SP1761	SP1761	SP1761	SP1761
SP1762	SP1762	SP1762	SP1762	SP1762	SP1762	SP1762	SP1762	SP1762	SP1762	SP1762
SP1763	SP1763	SP1763	SP1763	SP1763	SP1763	SP1763	SP1763	SP1763	SP1763	SP1763
SP1764	SP1764	SP1764	SP1764	SP1764	SP1764	SP1764	SP1764	SP1764	SP1764	SP1764
SP1765	SP1765	SP1765	SP1765	SP1765	SP1765	SP1765	SP1765	SP1765	SP1765	SP1765
SP1766	SP1766	SP1766	SP1766	SP1766	SP1766	SP1766	SP1766	SP1766	SP1766	SP1766
SP1767	SP1767	SP1767	SP1767	SP1767	SP1767	SP1767	SP1767	SP1767	SP1767	SP1767
SP1768	SP1768	SP1768		SP1768		SP1768			SP1768	SP1768
SP1769	SP1769	SP1769	SP1769	SP1769	SP1769	SP1769	SP1769	SP1769	SP1769	SP1769
SP1770	SP1770	SP1770	SP1770	SP1770	SP1770	SP1770	SP1770	SP1770	SP1770	SP1770
SP1771	SP1771	SP1771	SP1771	SP1771	SP1771	SP1771	SP1771	SP1771	SP1771	SP1771
SP1772	SP1772	SP1772	SP1772	SP1772	SP1772	SP1772	SP1772	SP1772	SP1772	SP1772
					SP1773					
SP1791								SP1791		
										SP1803
										SP1822
								SP1828		
	SP1829									
								SP1919		SP1919
	SP1948	SP1948			SP1948	SP1948		SP1948	SP1948	SP1948
	SP1949	SP1949			SP1949	SP1949		SP1949	SP1949	SP1949
SP1950	SP1950	SP1950	SP1950	SP1950	SP1950	SP1950	SP1950	SP1950	SP1950	SP1950
SP1951	SP1951	SP1951	SP1951	SP1951	SP1951	SP1951	SP1951	SP1951	SP1951	SP1951
SP1952	SP1952	SP1952	SP1952	SP1952	SP1952	SP1952	SP1952	SP1952	SP1952	SP1952
SP1953	SP1953				SP1953	SP1953	SP1953	SP1953	SP1953	SP1953

Invasive ST180 Serotype 3					Carriage ST180 Serotype 3					
99-4038	99-4039	00-3946	06-1705	02-1198	OXC141	03-4183	03-4155	03-4156	03-4185	07-2838
	SP1965									
SP2159	SP2159	SP2159	SP2159	SP2159	SP2159	SP2159	SP2159		SP2159	
SP2160	SP2160	SP2160	SP2160	SP2160	SP2160	SP2160	SP2160		SP2160	
SP2161	SP2161	SP2161	SP2161	SP2161	SP2161	SP2161	SP2161		SP2161	
SP2162	SP2162	SP2162	SP2162	SP2162	SP2162	SP2162	SP2162		SP2162	
SP2163	SP2163	SP2163	SP2163	SP2163	SP2163	SP2163	SP2163		SP2163	
SP2164	SP2164	SP2164	SP2164	SP2164	SP2164	SP2164	SP2164		SP2164	
SP2165	SP2165	SP2165	SP2165	SP2165		SP2165	SP2165		SP2165	
SP2166	SP2166	SP2166	SP2166	SP2166		SP2166	SP2166		SP2166	

Appendix 12 List of genes from TIGR4 genome which are present or non hybridizing in 10 strains of serotype 3, ST180.

Gene lists generated by Genespring GX 7.3.1 for each strain indicating regions of differential hybridization between the test strains and TIGR4. Yellow indicates hybridization by DNA from both TIGR4 and the test strain. Blue indicates hybridization of DNA from TIGR4 and not the test strain. Genes highlighted in white have been identified as essential for virulence in a mouse pneumonia model (Hava and Camilli, 2002).

Appendix 13

Invasive ST180 Serotype 3					Carriage ST180 Serotype 3					
99-4038	99-4039	00-3946	06-1705	02-1198	OXC141	03-4183	03-4155	03-4156	03-4185	07-2838
spr0067										
spr0098								spr0098		
								spr0102		
				spr0103				spr0103		spr0103
spr0104					spr0104	spr0104	spr0104	spr0104	spr0104	
spr0105		spr0105		spr0105	spr0105	spr0105	spr0105		spr0105	
spr0106	spr0106	spr0106	spr0106	spr0106	spr0106	spr0106	spr0106		spr0106	
spr0107	spr0107	spr0107		spr0107	spr0107	spr0107	spr0107			
spr0108	spr0108	spr0108	spr0108	spr0108	spr0108	spr0108	spr0108		spr0108	
spr0111	spr0111	spr0111	spr0111	spr0111	spr0111	spr0111	spr0111		spr0111	
spr0112	spr0112	spr0112	spr0112		spr0112	spr0112	spr0112		spr0112	spr0113
spr0113	spr0113	spr0113	spr0113		spr0113	spr0113	spr0113	spr0113	spr0113	
spr0114	spr0114	spr0114	spr0114	spr0114		spr0114	spr0114	spr0114	spr0114	spr0114
spr0115	spr0115	spr0115	spr0115	spr0115	spr0115	spr0115	spr0115	spr0115	spr0115	spr0115
spr0116	spr0116	spr0116	spr0116	spr0116	spr0116	spr0116	spr0116	spr0116	spr0116	spr0116
spr0117	spr0117	spr0117		spr0117	spr0117	spr0117			spr0117	
spr0118	spr0118	spr0118		spr0118	spr0118	spr0118	spr0118		spr0118	spr0118
spr0119	spr0119	spr0119			spr0119	spr0119	spr0119		spr0119	
										spr0121
					spr0225					
	spr0316									
spr0320								spr0320		
			spr0470	spr0470						
spr0491					spr0491					
spr0493										
	spr0681									
spr0703							spr0703			
	spr0800								spr0800	
spr0955	spr0955	spr0955	spr0955	spr0955	spr0955	spr0955	spr0955	spr0955	spr0955	
spr0956	spr0956	spr0956	spr0956	spr0956	spr0956	spr0956	spr0956	spr0956	spr0956	spr0956
spr0957	spr0957	spr0957	spr0957	spr0957	spr0957	spr0957	spr0957	spr0957	spr0957	spr0957
spr0958	spr0958	spr0958	spr0958	spr0958	spr0958	spr0958	spr0958	spr0958	spr0958	spr0958
spr0959	spr0959	spr0959	spr0959	spr0959	spr0959	spr0959	spr0959	spr0959	spr0959	spr0959
spr1042										
					spr1114					
							spr1132			
				spr1192						
	spr1199	spr1199		spr1199		spr1199				spr1401
			spr1401							
						spr1478				
										spr1489

Appendix 13 List of genes from the R6 genome which are present or non hybridizing in 10 strains of serotype 3, ST180.

Gene lists generated by Genespring GX 7.3.1. Red indicates genes from the R6 genome which are found in the test strains but are not present in TIGR4.

Appendix 14

Serotype 1 ST227			Serotype 1 ST306			
06-1805	05-2739	04-2225	05-1519	03-3038	06-1370	03-2672
				SP0009		
		SP0040				
	SP0052					
SP0067	SP0067	SP0067	SP0067		SP0067	SP0067
SP0068		SP0068	SP0068	SP0068	SP0068	SP0068
SP0069	SP0069	SP0069	SP0069	SP0069	SP0069	SP0069
SP0070		SP0070	SP0070			SP0070
SP0071	SP0071	SP0071	SP0071	SP0071	SP0071	SP0071
						SP0072
	SP0073				SP0073	SP0073
SP0074	SP0074	SP0074	SP0074		SP0074	SP0074
				SP0089		
SP0110	SP0110	SP0110	SP0110		SP0110	SP0110
SP0111	SP0111	SP0111	SP0111	SP0111	SP0111	SP0111
SP0112	SP0112	SP0112	SP0112	SP0112	SP0112	SP0112
SP0113	SP0113	SP0113	SP0113	SP0113	SP0113	SP0113
			SP0115			
				SP0125		
			SP0128			
					SP0132	
	SP0160					
				SP0165		
	SP0172					
	SP0179					
	SP0273					
	SP0295					
		SP0304				
	SP0307					
					SP0328	
						SP0346
					SP0347	SP0347
SP0350	SP0350	SP0350	SP0350		SP0350	SP0350
SP0351		SP0351	SP0351		SP0351	SP0351
SP0352	SP0352	SP0352	SP0352	SP0352	SP0352	SP0352
SP0353	SP0353	SP0353	SP0353	SP0353	SP0353	SP0353
SP0354		SP0354	SP0354	SP0354	SP0354	SP0354
SP0355	SP0355	SP0355	SP0355	SP0355	SP0355	SP0355
SP0356		SP0356	SP0356	SP0356	SP0356	SP0356
SP0357	SP0357	SP0357	SP0357	SP0357	SP0357	SP0357
SP0358	SP0358	SP0358	SP0358	SP0358	SP0358	SP0358
SP0359	SP0359	SP0359	SP0359	SP0359	SP0359	SP0359
SP0360	SP0360	SP0360	SP0360	SP0360	SP0360	SP0360
			SP0394	SP0394	SP0394	SP0394
	SP0395		SP0395	SP0395	SP0395	SP0395
			SP0396	SP0396	SP0396	SP0396
			SP0397	SP0397	SP0397	SP0397
	SP0399			SP0399		
			SP0407			
				SP0431		
SP0460	SP0460	SP0460	SP0460	SP0460	SP0460	SP0460
SP0461	SP0461	SP0461	SP0461	SP0461	SP0461	SP0461
SP0462	SP0462	SP0462	SP0462	SP0462	SP0462	SP0462
SP0463	SP0463	SP0463	SP0463	SP0463	SP0463	SP0463
SP0464	SP0464	SP0464	SP0464	SP0464	SP0464	SP0464
SP0465	SP0465	SP0465	SP0465	SP0465	SP0465	SP0465
SP0466	SP0466	SP0466	SP0466	SP0466	SP0466	SP0466
SP0467	SP0467	SP0467	SP0467		SP0467	SP0467
SP0468	SP0468	SP0468	SP0468		SP0468	SP0468
				SP0491		
SP0505	SP0505	SP0505	SP0505		SP0505	SP0505
		SP0506	SP0506		SP0506	
SP0509	SP0509	SP0509	SP0509		SP0509	SP0509
SP0511				SP0511		
SP0512						
				SP0534		
				SP0535		
	SP0559					
				SP0560		
			SP0573	SP0573		
SP0574		SP0574	SP0574	SP0574	SP0574	SP0574
SP0575	SP0575	SP0575	SP0575		SP0575	SP0575
	SP0596					
				SP0598		
	SP0600					
				SP0683		
			SP0684			SP0684
				SP0693		
SP0714	SP0714				SP0714	
SP0739		SP0738	SP0738		SP0738	SP0738
SP0739	SP0739	SP0739	SP0739		SP0739	SP0739
		SP0740	SP0740		SP0740	SP0740
				SP0752		
SP0773						
				SP0808		
				SP0809		
			SP0816			
	SP0821					
SP0826		SP0826	SP0826		SP0826	SP0826
			SP0889		SP0889	SP0889
			SP0890		SP0890	
				SP0908		
	SP0912					
	SP0926					
	SP0940					
	SP0949					
				SP0956		
SP1005						
	SP1030					
						SP1037
SP1047	SP1047	SP1047	SP1047		SP1047	SP1047
SP1048	SP1048	SP1048	SP1048	SP1048	SP1048	SP1048
SP1049		SP1049	SP1049		SP1049	SP1049
SP1050	SP1050	SP1050	SP1050	SP1050	SP1050	SP1050
SP1051	SP1051	SP1051	SP1051	SP1051	SP1051	SP1051
SP1052	SP1052	SP1052	SP1052		SP1052	SP1052

Serotype 1 ST227			Serotype 1 ST306			
06-1805	05-2739	04-2225	05-1519	03-3038	06-1370	03-2672
SP1053	SP1053	SP1053	SP1053	SP1053	SP1053	SP1053
SP1054	SP1054	SP1054	SP1054	SP1054	SP1054	SP1054
SP1055	SP1055	SP1055	SP1055	SP1055	SP1055	SP1055
SP1056	SP1056	SP1056	SP1056	SP1056	SP1056	
SP1057	SP1057	SP1057	SP1057	SP1057	SP1057	SP1057
SP1059	SP1059	SP1059	SP1059	SP1059	SP1059	SP1059
SP1060	SP1060	SP1060	SP1060	SP1060	SP1060	SP1060
SP1061	SP1061	SP1061	SP1061	SP1061	SP1061	SP1061
SP1062	SP1062	SP1062	SP1062	SP1062	SP1062	SP1062
SP1063	SP1063	SP1063	SP1063	SP1063	SP1063	SP1063
SP1064	SP1064	SP1064	SP1064	SP1064	SP1064	SP1064
	SP1103					
	SP1109			SP1109		
				SP1120		
SP1130	SP1130	SP1130	SP1130		SP1130	SP1130
		SP1133	SP1133	SP1133		
					SP1139	
			SP1140			
SP1141	SP1141	SP1141	SP1141			SP1141
SP1142			SP1142		SP1142	SP1142
SP1143	**SP1143**	**SP1143**	**SP1143**		**SP1143**	**SP1143**
SP1144	SP1144	SP1144	SP1144	SP1144	SP1144	SP1144
SP1145		SP1145	SP1145		SP1145	SP1145
				SP1158		SP1158
				SP1165		
				SP1181	SP1181	
SP1221	SP1221	SP1221	SP1221	SP1221	SP1221	SP1221
SP1222	SP1222	SP1222	SP1222	SP1222	SP1222	SP1222
	SP1253					
				SP1303		
SP1335			SP1335		SP1335	SP1335
SP1336	SP1336	SP1336	SP1336		SP1336	SP1336
	SP1352					
SP1379						
	SP1418					
SP1439	SP1439		SP1439			
	SP1481					
SP1490						
SP1503	SP1503					SP1503
SP1595	SP1595				SP1595	SP1595
					SP1612	
SP1615	SP1615	SP1615	SP1615	SP1615	SP1615	SP1615
SP1616	SP1616	SP1616	SP1616	SP1616	SP1616	SP1616
SP1617	SP1617	SP1617	SP1617	SP1617	SP1617	SP1617
SP1618	SP1618	SP1618	SP1618	SP1618	SP1618	SP1618
SP1619	SP1619	SP1619	SP1619	SP1619	SP1619	SP1619
SP1620	SP1620	SP1620	SP1620	SP1620	SP1620	SP1620
SP1621	SP1621	SP1621	SP1621	SP1621	SP1621	SP1621
SP1622	SP1622	SP1622	SP1622	SP1622	SP1622	SP1622
			SP1718			
	SP1719					
						SP1770
SP1794						
SP1796		SP1796	SP1796		SP1796	SP1796
SP1797	SP1797	SP1797	SP1797			SP1797
SP1798	SP1798	SP1798	SP1798		SP1798	SP1798
SP1799	SP1799	SP1799	SP1799		SP1799	SP1799
SP1828			SP1828			
			SP1829		SP1829	
	SP1830				**SP1830**	**SP1830**
						SP1831
						SP1834
	SP1850					
SP1948	SP1948	SP1948	SP1948		SP1948	SP1948
		SP1949	SP1949		SP1949	
SP1950	SP1950	SP1950	SP1950	SP1950	SP1950	SP1950
SP1951	SP1951	SP1951	SP1951	SP1951	SP1951	SP1951
SP1952	**SP1952**	**SP1952**	**SP1952**	**SP1952**	**SP1952**	**SP1952**
SP1953		SP1953	SP1953	SP1953	SP1953	SP1953
SP1954	SP1954	SP1954	SP1954	SP1954	SP1954	SP1954
SP1955			SP1955		SP1955	SP1955
				SP1962		
	SP1971					
				SP2018		
	SP2179					
			SP2237			

Appendix 14 List of genes from TIGR4 genome which are present or non hybridizing in serotype 1 isolates.

Gene lists generated by Genespring GX 7.3.1 for each strain indicating regions of differential hybridization between the test strains and TIGR4. Yellow indicates hybridization by DNA from both TIGR4 and the test strain. Blue indicates hybridization of DNA from TIGR4 and not the test strain. Genes highlighted in white have been identified as essential for virulence in a mouse pneumonia model (Hava and Camilli, 2002)

Appendix 15

Serotype 1 ST227			Serotype 1 ST306			
06-1805	05-2739	04-2225	05-1519	03-3038	06-1370	03-2672
				spr0098		
				spr0103		
	spr0104					
spr0105	spr0105			spr0105		
spr0111	spr0111		spr0111	spr0111		
spr0112			spr0112	spr0112	spr0112	
spr0113			spr0113	spr0113	spr0113	
spr0114			spr0114	spr0114		spr0114
spr0115	spr0115	spr0115	spr0115	spr0115	spr0115	spr0115
spr0116	spr0116	spr0116	spr0116	spr0116	spr0116	spr0116
spr0117		spr0117	spr0117			
spr0118	spr0118	spr0118	spr0118			
		spr0119	spr0119	spr0119	spr0119	
spr0225						
spr0320	spr0320	spr0320	spr0320	spr0320	spr0320	spr0320
spr0321	spr0321	spr0321	spr0321	spr0321		spr0321
spr0322	spr0322	spr0322	spr0322	spr0322		
spr0323	spr0323	spr0323	spr0323	spr0323		
				spr0345		
			spr0416			
				spr0470		
				spr0493		
spr0958	spr0958	spr0958	spr0958	spr0958		
spr0959	spr0959	spr0959	spr0959	spr0959		
				spr0961		
			spr0966			
spr0972			spr0972	spr0972	spr0972	
	spr1114			spr1114		
					spr1179	
					spr1188	
				spr1210		
				spr1283		
					spr1401	
spr1403	spr1403	spr1403	spr1403	spr1403	spr1403	spr1403
spr1404	spr1404	spr1404	spr1404	spr1404		spr1404
			spr1478	spr1478		
				spr1481		
			spr1484	spr1484		
	spr1549		spr1549			
	spr1614					
spr1618	spr1618		spr1618			
spr1619	spr1619		spr1619		spr1619	spr1619
spr1620	spr1620		spr1620	spr1620		spr1620
spr1621			spr1621			
				spr1752		

Appendix 15 List of genes from the R6 genome which are present or non hybridizing in serotype 1 strains.

Gene lists generated by Genespring GX 7.3.1. Results of Microarray CGH serotype 1 isolates. Red indicates genes from the R6 genome which are found in the test strains but are not present in TIGR4.

Appendix 16

Gene	P-value	No Clarithromycin Normalised Expression Level	Post Clarithromycin Normalised Expression Level	Common Name	Putative Function
			Downregulated		
SP1128	0.0080	0.994	0.062	eno	phosphopyruvate hydratase
SP0499	0.0031	1.022	0.101	pgk	phosphoglycerate kinase
SP0565	0.0152	0.988	0.133	SP_0565	hypothetical protein
SP1631	0.0004	1.021	0.151	thrS	threonyl-tRNA synthetase
SP0605	0.0010	1.018	0.154	fba	fructose-bisphosphate aldolase
SP1475	0.0069	1.050	0.163	glyQ	glycyl-tRNA synthetase subunit alpha
SP1474	0.0455	1.021	0.167	glyS	glycyl-tRNA synthetase subunit beta
SP0580	0.0312	1.027	0.185	SP_0580	acetyltransferase, GNAT family
SP0581	0.0129	0.950	0.193	pheT	phenylalanyl-tRNA synthetase subunit beta
SP1445	0.0045	0.971	0.194	guaA	bifunctional GMP synthase/glutamine amidotransferase protein
SP1298	0.0182	1.021	0.207	SP_1298	DHH subfamily 1 protein
SP1659	0.0069	0.992	0.222	ileS	isoleucyl-tRNA synthetase
SP1960	0.0179	0.973	0.222	rpoC	DNA-directed RNA polymerase subunit beta'
SP0745	0.0016	0.998	0.224	upp	uracil phosphoribosyltransferase
SP1306	0.0459	1.009	0.227	gdhA	glutamate dehydrogenase
SP1961	0.0497	0.959	0.230	rpoB	DNA-directed RNA polymerase subunit beta
SP1220	0.0184	1.016	0.233	ldh	L-lactate dehydrogenase
SP0568	0.0239	0.985	0.236	valS	valyl-tRNA synthetase
SP0856	0.0094	1.038	0.240	ilvE	branched-chain amino acid aminotransferase
SP1630	0.0420	1.035	0.246	SP_1630	hypothetical protein
SP1907	0.0286	1.023	0.256	groES	co-chaperonin GroES
SP0857	0.0219	1.027	0.266	SP_0857	
SP1476	0.0110	1.013	0.269	SP_1476	hypothetical protein
SP1477	0.0089	1.031	0.273	SP_1477	hypothetical protein
SP0411	0.0329	1.061	0.278	serS	seryl-tRNA synthetase
SP0400	0.0176	0.990	0.283	tig	trigger factor
SP0623	0.0276	1.041	0.288	pepV	dipeptidase PepV
SP1384	0.0216	1.001	0.292	SP_1384	hypothetical protein
SP1013	0.0008	0.990	0.294	asd	aspartate-semialdehyde dehydrogenase
SP0797	0.0151	1.008	0.297	pepN	aminopeptidase N
SP0281	0.0012	0.989	0.302	pepC	aminopeptidase C
SP2230	0.0023	1.003	0.307	ABC-NBD	ABC transporter, ATP-binding protein
SP0823	0.0152	1.017	0.327	glnP	amino acid ABC transporter, permease protein
SP1655	0.0199	1.007	0.335	gpmA	phosphoglyceromutase
SP1291	0.0444	1.013	0.343	SP_1291	Cof family protein
SP2121	0.0488	1.010	0.348	hisS	histidyl-tRNA synthetase
SP1290	0.0078	1.000	0.355	SP_1290	hypothetical protein

SP1576	0.0281	0.998	0.365	metA	homoserine O-succinyltransferase
SP1279	0.0111	1.012	0.365	nth	endonuclease III
SP1845	0.0357	0.996	0.365	exoA	exodeoxyribonuclease
SP1891	0.0018	0.994	0.366	amiA	oligopeptide ABC transporter, oligopeptide-binding protein AmiA
SP1647	0.0023	0.998	0.367	pepO	endopeptidase O
SP1008	0.0038	0.990	0.389	pepT	peptidase T
SP1296	0.0045	1.000	0.395	SP_1296	hypothetical protein
SP1574	0.0166	0.975	0.402	tpiA	triosephosphate isomerase
SP1575	0.0014	0.992	0.403	SP_1575	hypothetical protein
SP0436	0.0273	1.001	0.411	gatB	aspartyl/glutamyl-tRNA amidotransferase subunit B
SP1470	0.0485	0.995	0.413	apbE	thiamine biosynthesis protein ApbE, putative
SP1781	0.0196	1.009	0.429	SP_1781	hypothetical protein
SP1743	0.0390	0.966	0.436	SP_1743	hypothetical protein
SP0438	0.0386	1.032	0.442	gatC	aspartyl/glutamyl-tRNA amidotransferase subunit C
SP1591	0.0298	1.010	0.442	pepQ	proline dipeptidase
SP0788	0.0320	0.978	0.462	metG	methionyl-tRNA synthetase
SP1413	0.0054	1.001	0.466	hprK	HPr kinase/phosphorylase
SP2176	0.0252	1.002	0.473	dltA	D-alanine--D-alanyl carrier protein ligase
SP1985	0.0399	1.004	0.498	ksgA	dimethyladenosine transferase
Upregulated					
SP1734	0.0157	1.013	2.071	sunL	rRNA methyltransferase RsmB
SP0217	0.0340	1.013	2.118	rpmC	50S ribosomal protein L29
SP1667	0.0468	1.002	2.155	ftsA	cell division protein FtsA
SP0218	0.0198	0.982	2.212	rpsQ	30S ribosomal protein S17
SP0807	0.0006	0.997	2.218	ezrA	septation ring formation regulator EzrA
SP1968	0.0103	1.005	2.237	coaD	phosphopantetheine adenylyltransferase
SP1540	0.0222	1.008	2.240	ssbA	single-strand DNA-binding protein
SP1739	0.0096	1.001	2.369	SP_1739	hypothetical protein
SP1115	0.0138	1.027	2.406	rggD	transcriptional regulator MutR, putative
SP1602	0.0250	1.035	2.474	phnA	phnA protein
SP1264	0.0051	1.016	2.622	SP_1264	hypothetical protein
SP0215	0.0328	1.019	2.711	rpsC	30S ribosomal protein S3
SP1362	0.0308	0.964	2.956	mecA	adaptor protein
SP0740	0.0018	1.002	3.077	SP_0740	MutT/nudix family protein
SP0026	0.0279	0.960	3.421	SP_0026	hypothetical protein
SP0742	0.0134	1.048	4.889	SP_0742	hypothetical protein
SP0800	0.0009	1.002	10.524	SP_0800	hypothetical protein

Appendix 16 Genes present in strain South Africa 2507 identified as differentially expressed 15 minutes after midlog when growing in BHI broth in the presence and absence of clarithromycin 5mg/L (Bonferonni correction).

The genes which are significantly up or down regulated (P<0.05) were identified using Genespring GX 7.3.1 (Agilent Technologies, USA) using a one way ANOVA t-test with a false discovery rate set at 0.05 and using Bonferonni multiple testing correction. Expression differences of less than 2 fold difference were excluded. The P values resulting from the t-test comparing expression in the presence and absence of clarithromycin for each gene and the putative functions of the genes are noted. Genes involved in amino acid synthesis are

highlighted green, transcriptional regulators in orange and ribosomal proteins are highlighted in yellow.

Appendix 17

Gene	P-value	No Clarithromycin Normalised Expression Level	Post Clarithromycin Normalised Expression Level	Common Name	Putative Function
			Downregulated		
SP1128	0.0004	0.994	0.062	eno	phosphopyruvate hydratase
SP0499	0.0002	1.022	0.101	pgk	phosphoglycerate kinase
SP0565	0.0004	0.988	0.133	SP0565	hypothetical protein
SP1631	0.0002	1.021	0.151	thrS	threonyl-tRNA synthetase
SP0605	0.0002	1.018	0.154	fba	fructose-bisphosphate aldolase
SP1475	0.0003	1.050	0.163	glyQ	glycyl-tRNA synthetase subunit alpha
SP1474	0.0007	1.021	0.167	glyS	glycyl-tRNA synthetase subunit beta
SP0567	0.0037	0.932	0.167	SP0567	hypothetical protein
SP0445	0.0046	1.044	0.168	ilvB	acetolactate synthase catalytic subunit
SP1583	0.0016	0.998	0.185	entB	isochorismatase family protein
SP0580	0.0005	1.027	0.185	SP0580	acetyltransferase, GNAT family
SP1815	0.0007	1.034	0.190	trpD	anthranilate phosphoribosyltransferase
SP0581	0.0004	0.950	0.193	pheT	phenylalanyl-tRNA synthetase subunit beta
SP1445	0.0003	0.971	0.194	guaA	bifunctional GMP synthase/glutamine amidotransferase protein
SP1142	0.0297	0.994	0.198	SP1142	hypothetical protein
SP0579	0.0085	1.000	0.199	pheS	phenylalanyl-tRNA synthetase subunit alpha
SP1812	0.0363	1.000	0.202	trpB	tryptophan synthase subunit beta
SP1584	0.0024	1.023	0.202	codY	transcriptional repressor CodY
SP1057	0.0287	0.976	0.205	SP1057	transcriptional regulator PlcR, putative
SP1298	0.0005	1.021	0.207	SP1298	DHH subfamily 1 protein
SP0875	0.0415	0.986	0.218	fruR	lactose phosphotransferase system repressor
SP0409	0.0324	1.146	0.219	mip	hypothetical protein
SP1489	0.0037	1.021	0.219	tuf	elongation factor Tu
SpR6-0323	0.0103	1.084	0.222	cpsO	
SP1659	0.0003	0.992	0.222	ileS	isoleucyl-tRNA synthetase
SP1960	0.0005	0.973	0.222	rpoC	DNA-directed RNA polymerase subunit beta'
SP0745	0.0002	0.998	0.224	upp	uracil phosphoribosyltransferase
SP1651	0.0032	1.028	0.226	tpx	thiol peroxidase
SP1306	0.0007	1.009	0.227	gdhA	glutamate dehydrogenase
SP1473	0.0009	0.958	0.228	SP1473	hypothetical protein
SP1961	0.0007	0.959	0.230	rpoB	DNA-directed RNA polymerase subunit beta
SP0159	0.0339	1.442	0.231	SP0159	hypothetical protein
SP1753	0.0044	1.014	0.232	dctA	
SP1220	0.0005	1.016	0.233	ldh	L-lactate dehydrogenase
SP2239	0.0075	0.981	0.235	SP2239	serine protease
SP0568	0.0005	0.985	0.236	valS	valyl-tRNA synthetase
SP0856	0.0004	1.038	0.240	ilvE	branched-chain amino acid aminotransferase

SP1814	0.0100	1.002	0.244	trpC	indole-3-glycerol-phosphate synthase
SP1630	0.0006	1.035	0.246	SP1630	hypothetical protein
SP1811	0.0475	0.987	0.250	trpA	tryptophan synthase subunit alpha
SP1907	0.0005	1.023	0.256	groES	co-chaperonin GroES
SP1578	0.0239	0.854	0.259	SP1578	methyltransferase, putative
SP0857	0.0005	1.027	0.266	SP0857	
SP1476	0.0004	1.013	0.269	SP1476	hypothetical protein
SP0328	0.0113	1.000	0.271	SP0328	IS1380-Spn1 transposase
SP1499	0.0020	1.001	0.272	bta	bacterocin transport accessory protein
SP1477	0.0004	1.031	0.273	SP1477	hypothetical protein
SP0411	0.0005	1.061	0.278	serS	seryl-tRNA synthetase
SP2240	0.0021	1.053	0.279	spo0J	spspoJ protein
SP0507	0.0011	0.944	0.279	hsdS	type I restriction-modification system, S subunit, putative
SP0400	0.0005	0.990	0.283	tig	trigger factor
SP2026	0.0059	1.021	0.287	adhE	alcohol dehydrogenase, iron-containing
SP0623	0.0005	1.041	0.288	pepV	dipeptidase PepV
SP1906	0.0109	0.888	0.290	groEL	chaperonin GroEL
SP1384	0.0005	1.001	0.292	SP1384	hypothetical protein
SP1013	0.0002	0.990	0.294	asd	aspartate-semialdehyde dehydrogenase
SP0797	0.0004	1.008	0.297	pepN	aminopeptidase N
SP1064	0.0071	1.090	0.297	SP1064	transposase, IS200 family
SP0281	0.0002	0.989	0.302	pepC	aminopeptidase C
SP2230	0.0002	1.003	0.307	ABC-NBD	ABC transporter, ATP-binding protein
SP0746	0.0013	0.996	0.308	clpP	ATP-dependent Clp protease proteolytic subunit
SP0766	0.0120	0.990	0.318	sodA	superoxide dismutase, manganese-dependent
SP1813	0.0210	1.087	0.321	trpF	N-(5'-phosphoribosyl)anthranilate isomerase
SP0622	0.0058	1.015	0.326	nrd	nitroreductase family protein
SP1498	0.0009	1.058	0.327	pgm	phosphoglucomutase
SP0823	0.0004	1.017	0.327	glnP	amino acid ABC transporter, permease protein
SP0501	0.0059	1.044	0.332	glnR	transcriptional regulator, MerR family
SP1577	0.0024	0.962	0.335	apt	adenine phosphoribosyltransferase
SP1655	0.0005	1.007	0.335	gpmA	phosphoglyceromutase
SP1027	0.0033	1.048	0.338	SP1027	hypothetical protein
SP0459	0.0066	1.073	0.338	pfl	formate acetyltransferase
SP1291	0.0007	1.013	0.343	SP1291	Cof family protein
SP1280	0.0058	0.999	0.344	SP1280	hypothetical protein
SP0502	0.0115	1.061	0.344	glnA	glutamine synthetase, type I
SP0858	0.0038	0.993	0.346	SP0858	hypothetical protein
SP2121	0.0007	1.010	0.348	hisS	histidyl-tRNA synthetase
SP1290	0.0004	1.000	0.355	SP1290	hypothetical protein
SP0055	0.0040	0.999	0.358	SP0055	hypothetical protein
SP1576	0.0005	0.998	0.365	metA	homoserine O-succinyltransferase
SP1279	0.0004	1.012	0.365	nth	endonuclease III
SP1845	0.0006	0.996	0.365	exoA	exodeoxyribonuclease
SP1891	0.0002	0.994	0.366	amiA	oligopeptide ABC transporter,

					oligopeptide-binding protein AmiA
SP1647	0.0002	0.998	0.367	pepO	endopeptidase O
SP1412	0.0009	0.961	0.375	lgt	prolipoprotein diacylglyceryl transferase
SP1780	0.0049	1.001	0.376	pepF	oligoendopeptidase F, putative
SP0798	0.0058	0.996	0.376	ciaR	DNA-binding response regulator CiaR
SP1177	0.0011	1.001	0.378	ptsH	phosphocarrier protein HPr
SP2206	0.0053	0.954	0.378	SP2206	ribosomal subunit interface protein
SP1714	0.0428	1.035	0.382	SP1714	transcriptional regulator, GntR family
SP1145	0.0364	1.181	0.382	SP1145	hypothetical protein
SP1246	0.0221	0.946	0.385	SP1246	Cof family protein
SP0340	0.0321	0.998	0.385	luxS	S-ribosylhomocysteinase
SP1130	0.0180	0.994	0.387	SP1130	transcriptional regulator
SP1008	0.0003	0.990	0.389	pepT	peptidase T
SP1154	0.0136	0.945	0.395	SP1154	immunoglobulin A1 protease
SP1296	0.0003	1.000	0.395	SP1296	hypothetical protein
SP0176	0.0052	0.999	0.395	ribA	3,4-dihydroxy-2-butanone 4-phosphate synthase/GTP cyclohydrolase II
SP1574	0.0005	0.975	0.402	tpiA	triosephosphate isomerase
SP1295	0.0118	1.105	0.403	crcB	crcB protein
SP1575	0.0002	0.992	0.403	SP1575	hypothetical protein
SP0730	0.0012	0.985	0.403	spxB	pyruvate oxidase
SP0408	0.0147	0.910	0.406	dagA	sodium:alanine symporter family protein
SP0204	0.0074	0.999	0.407	SP0204	acetyltransferase, GNAT family
SP0713	0.0051	0.989	0.410	lysS	lysyl-tRNA synthetase
SP0436	0.0005	1.001	0.411	gatB	aspartyl/glutamyl-tRNA amidotransferase subunit B
SP1470	0.0007	0.995	0.413	apbE	thiamine biosynthesis protein ApbE, putative
SP0799	0.0016	1.009	0.416	ciaH	sensor histidine kinase CiaH
SP1715	0.0128	0.969	0.420	ABC-NBD	ABC transporter, ATP-binding protein
SP1415	0.0116	0.964	0.421	nagB	glucosamine-6-phosphate isomerase
SP1176	0.0037	1.012	0.426	ptsI	phosphoenolpyruvate-protein phosphotransferase
SP1781	0.0005	1.009	0.429	SP1781	hypothetical protein
SP0385	0.0123	0.998	0.436	SP0385	hypothetical protein
SP1743	0.0006	0.966	0.436	SP1743	hypothetical protein
SP2125	0.0366	0.981	0.437	SP2125	hypothetical protein
SP0816	0.0066	1.002	0.437	SP0816	hypothetical protein
SP0437	0.0333	0.902	0.440	gatA	aspartyl/glutamyl-tRNA amidotransferase subunit A
SP1069	0.0036	0.990	0.440	ABC-SBP	hypothetical protein
SP0438	0.0006	1.032	0.442	gatC	aspartyl/glutamyl-tRNA amidotransferase subunit C
SP1591	0.0005	1.010	0.442	pepQ	proline dipeptidase
SP1243	0.0022	0.965	0.449	zwf	glucose-6-phosphate 1-dehydrogenase
SP2175	0.0041	1.026	0.449	dltB	dltB protein
SP1014	0.0092	0.940	0.451	dapA	dihydrodipicolinate synthase
SP0867	0.0089	0.998	0.453	ABC-NBD	ABC transporter, ATP-binding protein
SP0516	0.0062	0.961	0.455	grpE	heat shock protein GrpE

SP1970	0.0009	1.019	0.457	asnA	asparagine synthetase AsnA
SP1505	0.0008	0.966	0.457	SP1505	membrane protein
SP0056	0.0040	0.954	0.460	purB	adenylosuccinate lyase
SP0788	0.0005	0.978	0.462	metG	methionyl-tRNA synthetase
SP2069	0.0021	1.008	0.465	gltX	glutamyl-tRNA synthetase
SP1413	0.0003	1.001	0.466	hprK	HPr kinase/phosphorylase
SP0828	0.0099	1.025	0.470	rpiA	ribose-5-phosphate isomerase A
SP0824	0.0280	1.000	0.470	glnQ	amino acid ABC transporter, ATP-binding protein
SP0869	0.0367	1.082	0.470	SP0869	aminotransferase, class-V
SP2058	0.0037	0.961	0.471	tgt	queuine tRNA-ribosyltransferase
SP1112	0.0012	1.011	0.471	SP1112	degV family protein
SP1633	0.0096	1.006	0.471	rr01	DNA-binding response regulator
SP1221	0.0025	1.000	0.471	spnII	type II restriction endonuclease, putative
SP0688	0.0033	1.027	0.472	murD	UDP-N-acetylmuramoyl-L-alanyl-D-glutamate synthetase
SP2176	0.0005	1.002	0.473	dltA	D-alanine--D-alanyl carrier protein ligase
SP0173	0.0053	1.000	0.475	mutL	DNA mismatch repair protein
SP1471	0.0160	1.013	0.475	SP1471	oxidoreductase, putative
SP0979	0.0103	0.997	0.476	pepB	oligoendopeptidase F
SP0981	0.0020	0.985	0.477	prsA	foldase protein PrsA
SP1378	0.0007	0.984	0.478	SP1378	hypothetical protein
SP0784	0.0059	0.989	0.479	gor	glutathione reductase
SP0205	0.0234	0.996	0.483	nrdG	anaerobic ribonucleoside-triphosphate reductase activating protein
SP1175	0.0110	1.019	0.485	phtA	hypothetical protein
SP1244	0.0011	1.002	0.490	ftsY	signal recognition particle-docking protein FtsY
SP1985	0.0006	1.004	0.498	ksgA	dimethyladenosine transferase
SP0682	0.0165	1.083	0.498	SP0682	hypothetical protein
SP1247	0.0007	1.013	0.500	smc	hypothetical protein
SP1068	0.0018	1.032	0.503	ppc	phosphoenolpyruvate carboxylase
SP0868	0.0218	1.068	0.508	SP0868	hypothetical protein
SP1693	0.0203	1.090	0.514	nanA	
SP2126	0.0187	1.052	0.515	ilvD	dihydroxy-acid dehydratase
SP1887	0.0172	1.111	0.519	amiF	oligopeptide ABC transporter, ATP-binding protein AmiF
Upregulated					
SP1422	0.0207	0.916	1.859	SP1422	hypothetical protein
SP0210	0.0089	0.939	1.915	rplD	50S ribosomal protein L4
SP0224	0.0017	0.970	1.948	rpsH	30S ribosomal protein S8
SP1035	0.0045	0.956	1.954	ABC-NBD	iron-compound ABC transporter, ATP-binding protein
SP0984	0.0027	0.963	1.960	gpmB	phosphoglycerate mutase family protein
SP1356	0.0025	0.989	1.989	trzA	chlorohydrolase
SP0739	0.0008	0.976	2.000	mta	transcriptional regulator, MerR family
SP1105	0.0018	0.980	2.025	rplU	50S ribosomal protein L21
SP0220	0.0123	1.000	2.026	rplX	50S ribosomal protein L24
SP0004	0.0007	0.987	2.030	SP0004	translation-associated GTPase
SP0974	0.0146	0.997	2.032	secG	preprotein translocase subunit SecG
SP0943	0.0059	0.971	2.040	gidA	tRNA (uracil-5-)-methyltransferase Gid

SP0991	0.0007	0.982	2.049	pfs	5'-methylthioadenosine/S-adenosylhomocysteine nucleosidase
SP1734	0.0004	1.013	2.071	sunL	rRNA methyltransferase RsmB
SP2008	0.0088	0.958	2.080	secE	preprotein translocase subunit SecE
SP1083	0.0016	1.020	2.084	SP1083	hypothetical protein
SP0209	0.0016	0.963	2.098	rplC	50S ribosomal protein L3
SP0226	0.0017	0.996	2.098	rplR	50S ribosomal protein L18
SP1198	0.0145	0.992	2.106	SP1198	hypothetical protein
SP1992	0.0159	1.002	2.113	SP1992	cell wall surface anchor family protein
SP0217	0.0006	1.013	2.118	rpmC	50S ribosomal protein L29
SpR6-0464	0.0263	0.996	2.119	hk13	sensor histidine kinase BlpH, putative
SP2000	0.0339	1.014	2.122	rr11	DNA-binding response regulator
SP1926	0.0321	0.900	2.132	SP1926	hypothetical protein
SP0779	0.0037	1.013	2.139	trmD	tRNA (guanine-N(1)-)-methyltransferase
SP0002	0.0008	0.997	2.147	dnaN	DNA polymerase III subunit beta
SP0191	0.0311	0.995	2.150	SP0191	hypothetical protein
SP1667	0.0007	1.002	2.155	ftsA	cell division protein FtsA
SP1643	0.0292	0.997	2.164	SP1643	hypothetical protein
SP1034	0.0159	0.999	2.168	ABC-MSP	iron-compound ABC transporter, permease protein
SP1084	0.0036	1.011	2.182	map	methionine aminopeptidase
SP0613	0.0048	1.046	2.185	SP0613	metallo-beta-lactamase superfamily protein
SP0227	0.0037	1.019	2.194	rpsE	30S ribosomal protein S5
SP1429	0.0464	1.065	2.197	SP1429	peptidase, U32 family
SP0213	0.0011	1.016	2.200	rpsS	30S ribosomal protein S19
SP0218	0.0005	0.982	2.212	rpsQ	30S ribosomal protein S17
SP0807	0.0002	0.997	2.218	ezrA	septation ring formation regulator EzrA
SP1170	0.0333	0.992	2.224	SP1170	hypothetical protein
SP1081	0.0021	0.990	2.229	murZ	UDP-N-acetylglucosamine 1-carboxyvinyltransferase
SP1427	0.0364	1.094	2.230	SP1427	peptidase, U32 family
SP1541	0.0044	1.003	2.237	rpsF	30S ribosomal protein S6
SP0225	0.0032	0.981	2.237	rplF	50S ribosomal protein L6
SP1968	0.0004	1.005	2.237	coaD	phosphopantetheine adenylyltransferase
SP1540	0.0005	1.008	2.240	ssbA	single-strand DNA-binding protein
SP0231	0.0076	1.001	2.241	adk	adenylate kinase
SP0964	0.0263	0.917	2.255	pyrD	dihydroorotate dehydrogenase 1B
SP0419	0.0442	0.934	2.273	fabK	enoyl-(acyl-carrier-protein) reductase
SP2214	0.0028	0.964	2.274	tsf	elongation factor Ts
SP0211	0.0020	0.933	2.277	rplW	50S ribosomal protein L23
SP0219	0.0166	0.999	2.281	rplN	50S ribosomal protein L14
SP2152	0.0091	0.968	2.294	SP2152	hypothetical protein
SP0380	0.0411	1.058	2.301	SP0380	hypothetical protein
SP2156	0.0027	1.048	2.303	SP2156	SPFH domain/Band 7 family
SP1032	0.0319	0.999	2.326	ABC-SBP	iron-compound ABC transporter, iron compound-binding protein
SP1354	0.0108	1.028	2.333	rplL	50S ribosomal protein L7/L12
SP0992	0.0067	0.997	2.334	SP0992	hypothetical protein
SP1293	0.0081	1.003	2.362	rplS	50S ribosomal protein L19
SP1601	0.0010	0.993	2.365	SP1601	hypothetical protein

SP1739	0.0004	1.001	2.369	SP1739	hypothetical protein
SP0650	0.0010	0.989	2.378	SP0650	hypothetical protein
SP0741	0.0030	1.023	2.378	SP0741	hypothetical protein
SP1115	0.0004	1.027	2.406	rggD	transcriptional regulator MutR, putative
SP1600	0.0022	1.033	2.456	SP1600	hypothetical protein
SP1666	0.0036	0.995	2.463	ftsZ	cell division protein FtsZ
SP2116	0.0179	0.993	2.471	SP2116	hypothetical protein
SP1033	0.0261	1.000	2.471	ABC-MSP	iron-compound ABC transporter, permease protein
SP1602	0.0005	1.035	2.474	phnA	phnA protein
SP0212	0.0431	0.996	2.493	rplB	50S ribosomal protein L2
SP0024	0.0091	0.974	2.511	SP0024	hypothetical protein
SP1355	0.0010	1.053	2.527	rplJ	50S ribosomal protein L10
SP2007	0.0008	0.950	2.539	nusG	transcription antitermination protein NusG
SP2215	0.0008	1.002	2.553	rpsB	30S ribosomal protein S2
SP0783	0.0037	1.055	2.565	bioY	hypothetical protein
SP0986	0.0122	0.998	2.566	SP0986	hypothetical protein
SP0216	0.0037	0.999	2.568	rplP	50S ribosomal protein L16
SP1208	0.0016	1.027	2.576	udK	uridine kinase
SP0185	0.0061	0.999	2.586	corA	magnesium transporter, CorA family
SP0969	0.0051	0.991	2.609	era	GTP-binding protein Era
SP1264	0.0003	1.016	2.622	SP1264	hypothetical protein
SP0215	0.0005	1.019	2.711	rpsC	30S ribosomal protein S3
SP0420	0.0332	0.917	2.744	fabD	acyl-carrier-protein S-malonyltransferase
SP0645	0.0101	0.987	2.750	PTS-EII	PTS system IIA component, putative
SP0025	0.0085	0.988	2.781	SP0025	hypothetical protein
SP0817	0.0447	1.000	2.795	SP0817	MutT/nudix family protein
SP0111	0.0345	1.084	2.808	ABC-NBD	amino acid ABC transporter, ATP-binding protein, putative
SP1586	0.0037	1.031	2.886	SP1586	ATP-dependent RNA helicase, putative
SP1171	0.0058	0.977	2.940	SP1171	hydrolase, haloacid dehalogenase-like family
SP2117	0.0220	1.043	2.944	SP2117	hypothetical protein
SP1922	0.0017	1.076	2.948	SP1922	hypothetical protein
SP1362	0.0005	0.964	2.956	mecA	adaptor protein
SP2062	0.0089	0.904	2.994	marR	transcriptional regulator, MarR family
SP0740	0.0002	1.002	3.077	SP0740	MutT/nudix family protein
SP0430	0.0033	0.998	3.107	SP0430	hypothetical protein
SP1821	0.0156	1.000	3.283	ccpA	sugar-binding transcriptional regulator, LacI family
SP0423	0.0263	0.846	3.313	accB	acetyl-CoA carboxylase biotin carboxyl carrier protein subunit
SP0421	0.0418	0.928	3.358	fabG	3-ketoacyl-(acyl-carrier-protein) reductase
SP1197	0.0037	0.976	3.397	SP1197	hypothetical protein
SP0026	0.0005	0.960	3.421	SP0026	hypothetical protein
SP0427	0.0200	0.844	3.464	accA	acetyl-CoA carboxylase subunit alpha
SP0426	0.0261	0.933	3.676	accD	acetyl-CoA carboxylase subunit beta
SP0424	0.0269	0.889	3.721	fabZ	(3R)-hydroxymyristoyl-(acyl carrier protein) dehydratase

SP0425	0.0304	0.909	3.793	accC	acetyl-CoA carboxylase biotin carboxylase subunit
SP2098	0.0056	1.005	3.816	SP2098	membrane protein
SP0422	0.0292	0.885	3.935	fabF	3-oxoacyl-(acyl carrier protein) synthase II
SP0085	0.0017	0.983	4.016	rpsD	30S ribosomal protein S4
SP2003	0.0162	0.999	4.164	ABC-NBD	ABC transporter, ATP-binding protein
SP0742	0.0004	1.048	4.889	SP0742	hypothetical protein
SP2216	0.0040	1.011	4.997	gSP781	secreted 45 kd protein
SP0800	0.0002	1.002	10.524	SP0800	hypothetical protein
SP2063	0.0087	0.932	15.672	SP2063	LysM domain-containing protein
SP0107	0.0008	1.088	34.684	SP0107	LysM domain protein

Appendix 17 Genes present in strain South Africa 2507 which are identified as differentially expressed 15 minutes after midlog when growing in BHI broth in the presence and absence of clarithromycin 5mg/L (Benjamini and Hochberg correction).

The genes which are significantly up or down regulated (P<0.05) were identified using Genespring GX 7.3.1 (Agilent Technologies, USA) using a one way ANOVA t-test with a false discovery rate set at 0.05 and Benjamini and Hochberg multiple testing correction used. Expression differences with less than 2 fold differences were excluded. The P values resulting from the t-test comparing expression in the presence and absence of clarithromycin for each gene and the putative functions of the genes are noted. Genes involved in amino acid synthesis are highlighted green, transcriptional regulators in orange and ribosomal proteins are highlighted in yellow. Genes highlighted in red have been identified as part of the pneumococcal stress response to vancomycin (Haas *et al.*, 2005).