8   The evolution of serotype 3 ST180 pneumococci

8.1 Introduction

Serotype 3 was one of the earliest pneumococcal capsule types to be identified, first described in 1913 (Dochez and Gillespie, 1913). Bacterial colonies of this serotype have a characteristic mucoid phenotype when grown on an agar plate, due to the long polymeric chains of cellobiuronic acid units that comprise the thick capsule (Heidelberger and Goebel, 1927). For some time, this serotype was treated as a distinct species, named *Pneumococcus mucosus* (Watson *et al.*, 1993), and whilst such a separation cannot be justified on the basis of genetic divergence it does have distinctive epidemiological traits. Unusually for *S. pneumoniae*, the risk of serotype 3 disease increase with age (Gransden *et al.*, 1985; Scott *et al.*, 1996; Inostroza *et al.*, 2001; Harboe *et al.*, 2009), which may relate to the high immunogenicity of the capsule antigen in young children (Douglas *et al.*, 1983). This is likely to explain why serotype 3 has been found to have a high odds ratio for causing disease when comparing infections in adults to carriage in children (Sandgren *et al.*, 2004), whereas this association was not found in similar studies of paediatric disease (Smith *et al.*, 1993; Brueggemann *et al.*, 2003; Sleeman *et al.*, 2006), although a recent similar study in a community vaccinated with PCV7 found serotype 3 to be one of the most invasive capsular types (Yildirim *et al.*, 2010). Furthermore, serotype 3 disease is associated with an increased relative risk of mortality in humans (Gransden *et al.*, 1985; Henriques *et al.*, 2000; Martens *et al.*, 2004; Harboe *et al.*, 2009), and correspondingly strains of this serotype are amongst the quickest to cause death in a mouse model of bacteraemia (Briles *et al.*, 1992).

It has been suggested that these observations may stem from the high frequency with which serotype 3 isolates cause unusual modes of IPD, such as infections of extra-pulmonary sites and broncho-, rather than pulmonary, pneumonia (Finland and Barnes, 1977). Whether these characteristics are the consequence of the capsule or the genetic background itself is difficult to study, because serotype 3 isolates are unusually clonal when compared to other *S. pneumoniae* capsule types. The thick capsule is hypothesized to inhibit the uptake of exogenous DNA by these

pneumococci (Hsieh *et al.*, 2006), thereby enforcing a degree of genetic isolation upon this population. Of the 323 serotype 3 isolates in the multi-locus sequence typing database (Aanensen and Spratt, 2005), the majority belong to a single clonal complex founded by sequence type ST180. This lineage has also been designated as the Netherlands 3-31, or PMEN31, clone (McGee and Klugman, 2011); although it is not associated with penicillin resistance, (Hsieh *et al.*, 2006), many representatives of the lineage are resistant to macrolides (Isozumi *et al.*, 2008). Furthermore, the lineage is geographically highly widespread, having been found across Europe (Enright and Spratt, 1998), Japan (Isozumi *et al.*, 2008) and North and South America (Enright and Spratt, 1998; Beall *et al.*, 2006; Reis *et al.*, 2008). Additionally, it has been observed to increase in prevalence in the USA following the introduction of the heptavalent conjugate vaccine, which does not protect against serotype 3 pneumococci (Beall *et al.*, 2006).

8.2 Genetics of the serotype 3 ST180 population

8.2.1   The genome of *S. pneumoniae* OXC141

The complete genome of *S. pneumoniae* OXC141, a serotype 3 ST180 carriage isolate from a child in Oxford, was generated using a combination of 454 and capillary sequence data. The chromosome was found to be 2,036,867 bp long and contains 1,974 CDSs (including 150 pseudogenes and 58 IS elements), 122 BOX elements, 106 RUP elements and 29 SPRITE repeats (Figure 8.1). The 34-kb prophage φOXC141 is the only clearly autonomously mobile genetic element detectable in the chromosome, although there is also a 6.3 kb island that could be a small ICE. Two large genomic islands also distinguish *S. pneumoniae* OXC141 from ATCC 700669: an AT-rich ~22 kb region directly upstream of *pspA* that encodes multiple bacteriocin systems, and a ~25 kb section of PPI-1 (see Chapter 1) which contains a cryptic set of metabolic genes.

In order to perform an analysis analogous to that of the PMEN1 population, six further representatives of ST180 were sequenced using a combination of 454 and capillary sequencing, so as to assess the diversity of the accessory genome, and 77 ST180 and ST505 (a double locus variant of ST180) strains were sequenced as

multiplexed libraries on the Illumina platform to study the demographics of the population (Appendix IV: Serotype 3 strains).
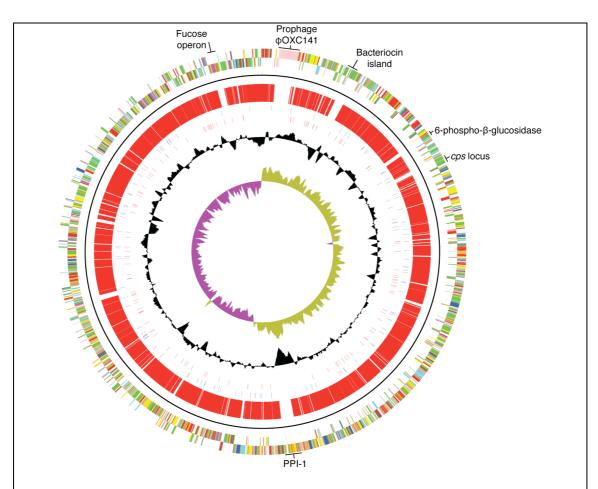

Figure 8.1 Circular diagram representing the *S. pneumoniae* OXC141 genome. The outer rings show the arrangement of coding sequences on the two strands of the genome, coloured according to annotated function (see Chapter 1). The first inner ring shows BLASTN matches to the *S. pneumoniae* ATCC 700669 genome. Moving clockwise from the origin of replication, the major regions of difference are the prophage φOXC141, an island encoding bacteriocin systems, the *cps* locus, PPI-1, and the fucose metabolism operon. Moving inward, the rings show the position of RUP repeats (blue), and BOX elements (red). The black graph indicates sequence GC content, and the innermost green and purple graph shows the GC skew of the sequence.

8.2.2 The phylogeny of ST180

The Illumina sequenced strains had a mean mapping depth of 145-fold coverage, each strain having a mean depth of at least 56-fold coverage. Constructing a phylogeny and identifying recombinations for the isolates, as described in Chapter 4, reveals a very different pattern of evolution from that observed for PMEN1 (Figure 8.2). Within ST180, 8,814 polymorphic sites could be identified, dividing the lineage into two groups (clades I and II), with one isolate (BHN609) in an intermediate position.
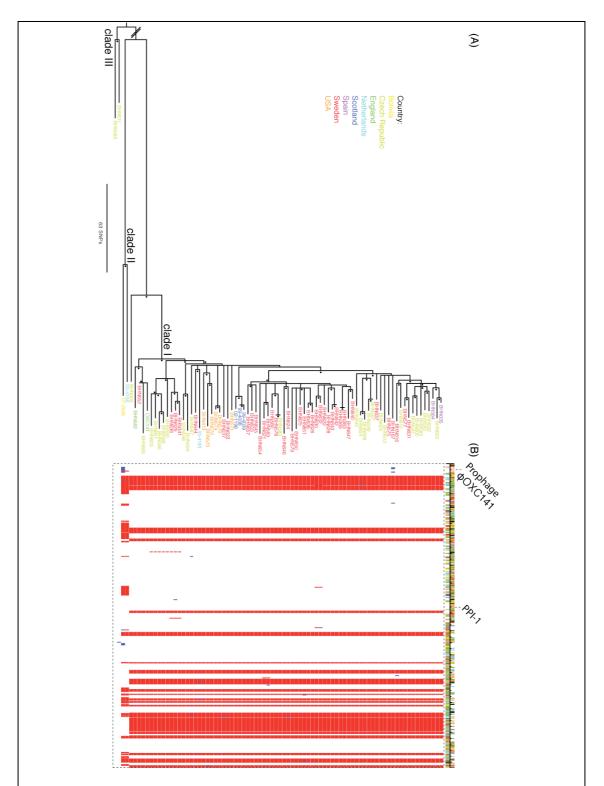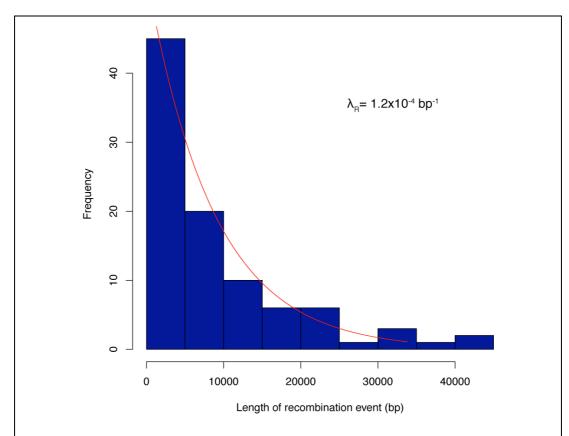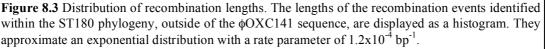
Figure 8.2 Phylogenetic analysis of the serotype 3 isolates. (A) A maximum likelihood phylogeny of the serotype 3 isolates based on the vertically inherited substitutions not introduced through recombination events. The strains are coloured according to their site of geographic origin. Nodes of the tree with a bootstrap support value of greater than 75, from 100 replicates, are marked with a dot. (B) The annotation of the *S. pneumoniae* OXC141 genome is shown, as displayed in Figure 8.1, with prophage φOXC141 and PPI-1 marked. The putative recombination events in each taxon, detected by the algorithm described in Chapter 4, are marked relative to this annotation in the panel underneath outlined by the dotted line. Red blocks indicate recombination events reconstructed as occurring on an internal branch, hence are shared by more than one strain, while blue blocks indicate recombinations unique to one particular strain.
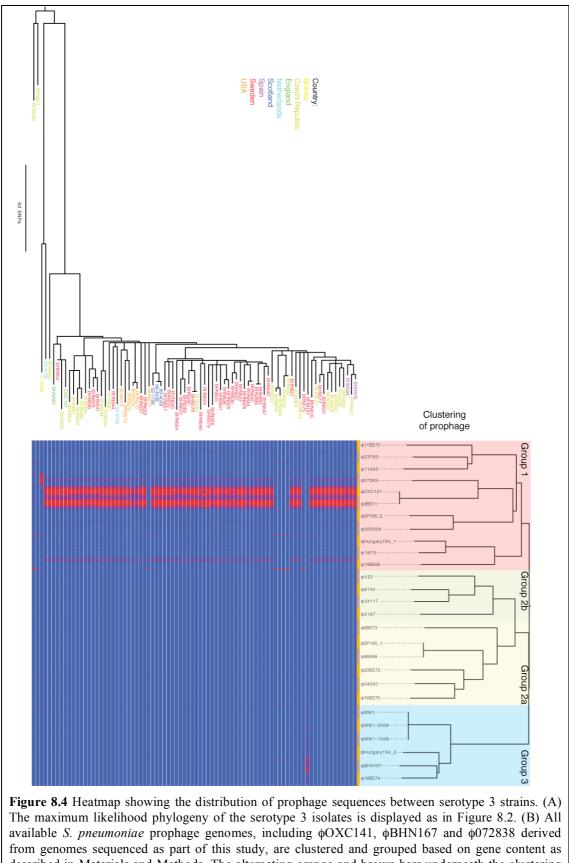
The ST505 outgroup, forming the more distant clade III, predicts that the root of the ST180 phylogeny lies between clades I and II. Reconstructing the history of the lineage suggests that 9,025 SNPs have occurred within ST180, 73% of which have been introduced by 92 recombination events. The ratio of transformation events to point mutations is highly heterogeneous throughout the phylogeny, hence within each clade there is little diversity: the branches separating clades I and II account for 15% of the substitutions occurring within the ST180 tree, but over 84% of the polymorphisms imported by recombination events. Despite their infrequency, the 88 homologous recombinations outside the prophage region appear to have a similar size distribution (λ_R = 1.2×10^{-4} bp^{-1}, 95% confidence interval 9.6×10^{-5}-1.4×10^{-4} bp^{-1}; Figure 8.3) to those identified in PMEN1 (λ_R for PMEN1 = 1.6×10^{-4} bp^{-1}, 95% confidence interval 1.5×10^{-4}-1.7×10^{-4} bp^{-1}). Hence the overall *r/m* for ST180 is 2.6, just over a third of that of PMEN1.

Figure 8.3 Distribution of recombination lengths. The lengths of the recombination events identified within the ST180 phylogeny, outside of the φOXC141 sequence, are displayed as a histogram. They approximate an exponential distribution with a rate parameter of 1.2×10^{-4} bp^{-1}.

The range of sampling dates relative to the age of ST180 was insufficient for root-to-tip analyses to provide precise estimates of the mutation rates or time since the last common ancestor, hence a Bayesian analysis of the phylogeny was required to assess these parameters. The last common ancestor of ST180 was estimated to have existed about 144 years ago (95% credibility interval 51-274 years ago), with the emergence of clade I predicted to be associated with a dramatic increase in population size around 51 years ago (95% credibility interval 24-90 years). This implies a mutation rate, across ST180 and ST505, of 8.2×10^{-7} substitutions per site per year (95% credibility interval 2.6×10^{-7}-1.4×10^{-6} substitutions per site per year), somewhat slower than that estimated for PMEN1 (1.6×10^{-6} substitutions per site per year). This may reflect the greater age of the clone allowing more time for selection to purge older disadvantageous mutations from the genotype.

There is no obvious reason for the reduced rate of transformation relative to PMEN1 evident in the genomes of most of the isolates, implying that the mucoid capsule itself is sufficient to enforce a degree of genetic isolation. However, individual isolates are observed to have frameshift mutations in the *comD* sensor kinase, *comFA* helicase and *comEA* transport protein, suggesting the selection pressure for the retention of the machinery may have been weakened (although a *comFA* frameshift mutation was observed in the PMEN1 population). Overall, 178 mutations causing the truncation of CDSs were observed. Based on the Poisson model of mutation events described in Chapter 4, the only functional CDSs to suffer a significantly high number of such disruptive mutations were 1,551 bp SPNOXC10420, encoding a voltage-gated Ca^{2+} channel (five disruptions, $p = 7.5 \times 10^{-7}$), and 1,245 bp SPNOXC12950, encoding a putative flavin mononucleotide reductase (four disruptions, $p = 9.52 \times 10^{-6}$).

Figure 8.4 Heatmap showing the distribution of prophage sequences between serotype 3 strains. (A) The maximum likelihood phylogeny of the serotype 3 isolates is displayed as in Figure 8.2. (B) All available *S. pneumoniae* prophage genomes, including φOXC141, φBHN167 and φ072838 derived from genomes sequenced as part of this study, are clustered and grouped based on gene content as described in Materials and Methods. The alternating orange and brown bars underneath the clustering indicate the extent of each prophage. (C) A heatmap of Illumina sequence read coverage of the prophage sequences is displayed for each taxon. Blue indicates low coverage, and red indicates high coverage.

Clades I and II also differ quite extensively in their accessory genome content. Notable in clade I is the widespread presence of prophage φOXC141, which appears to have been acquired by an ancestor of the clade and subsequently deleted on three independent occasions in the sampled strains (Figure 8.4). Heightened coverage of this element in the Illumina sequence data suggests it may be capable of active replication (Figure 8.5), which, along with the homoplasic nature of its removal, indicates the element is likely to be detrimental to the host. One of the taxa to have lost φOXC141, BHN167, is the only strain in clade I showing evidence of having acquired a novel prophage. In clade II, *S. pneumoniae* 03-4156 has a prophage sufficiently closely related to φOXC141 that it may represent *in situ* modification of that element, while *S. pneumoniae* 07-2838 has gained a very distinct prophage. Overall, the flux of such elements in the ST180 population appears to be much slower than in PMEN1.

Also found to vary between the two clades are some metabolic and bacteriocin synthesis operons, and the protein antigens PspA and PspC (Figure 8.6). The most parsimonious explanation of the patterns of variation appears to be acquisition of accessory loci during the diversification between the two clades, followed by infrequent deletion of islands over the shorter timescales within each clade, although only clade I is sampled with sufficient density to study this pattern in detail. Just one instance of antibiotic gene acquisition is evident in the population. This is an insertion of Tn*916*, carrying the *tetM* tetracycline resistance gene, into the Bolivian strain *S. pneumoniae* 07-2838. There is no evidence for import of macrolide resistance cassettes, so frequent in PMEN1, or acquisition of other antibiotic resistance genes. Furthermore, the fluoroquinolone resistance polymorphisms within the topoisomerase genes *gyrA*, *gyrB*, *parC* and *parE* that are observed to be homoplasic in the phylogenies of PMEN1 (Chapter 4), *Staph. aureus* ST239 (Harris *et al*., 2010) and *Salmonella* Typhi (Holt *et al*., 2008) are not evident at all in ST180.
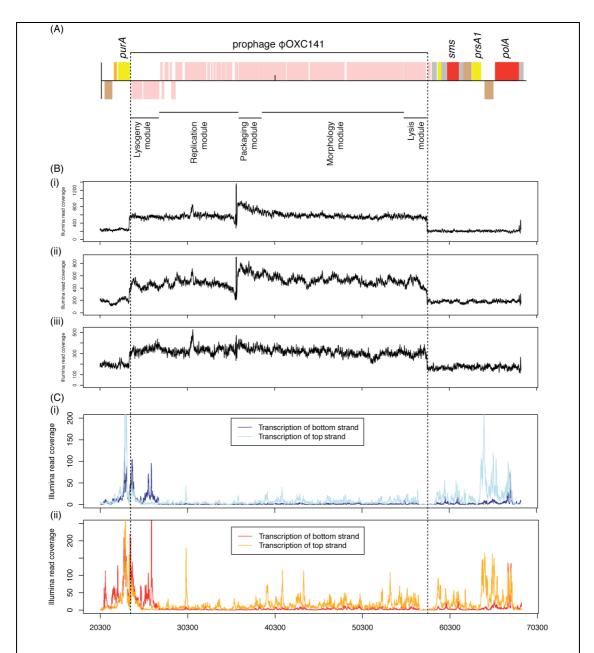
Figure 8.5 Activity of prophage φOXC141. (A) Detailed annotation of φOXC141 and the flanking genomic regions. The modular organisation typical of pneumococcal prophage is marked relative to the viral sequence. (B) Mapping of Illumina genome sequencing data from strains (i) BHN640 (ii) BHN035 and (iii) BHN598, phylogenetically disparate within clade I. The two- to three-fold increased coverage of φOXC141 suggests it is actively replicating. (C) Prophage transcription. The RNA-seq data from (i) 99-4039 and (ii) 99-4038 indicate that in some cells, the lysogeny module is active, keeping the prophage dormant, while in others, the modules required for replication and host cell lysis are expressed.

One SNP that does cause antibiotic resistance, a polymorphism upstream of the ABC-type efflux pump *patAB*, is found distinguishing the closely related sister taxa *S. pneumoniae* 99-4038 and 99-4039. These strains were isolated from a single case of meningitis: 4038 was taken from the bloodstream, and 4039 subsequently from the CSF. The maximum likelihood phylogeny predicts that these are distinguished by just

three polymorphisms, all of which are of the ancestral allele in the isolate from the bloodstream, and of the derived form in the CSF strain. This suggests the mutations have arisen within the patient during, or prior to, the invasive disease. Manual inspection of the assembled sequences confirmed the only other differences between the strains were non-synonymous changes to a LysR-domain regulator and a haloacid dehalogenase-type hydrolase in addition to the mutation upstream of *patAB*.

8.3 Comparison of *S. pneumoniae* 99-4038 and 99-4039

8.3.1 Transcriptional profiles of *S. pneumoniae* 99-4038 and 99-4039

Preliminary work comparing the transcriptional profiles of 99-4038 and 99-4039 using a microarray based on the genome of *S. pneumoniae* TIGR4 was performed by Donald Inverarity of the University of Glasgow (Inverarity, 2009); this revealed significant differences in their patterns of gene expression (Table 8.1). Sequencing of three paired RNA samples from the two strains grown *in vitro* were then used to characterise these differences more precisely. This experiment broadly agreed with the conclusions from the microarray work (Table 8.2, Table 8.3). Analysis of differential gene expression revealed that 53 CDSs had a significantly altered pattern of transcription. These could be grouped into 11 gene clusters and 17 singleton CDSs. Both the microarray and RNA-seq data found the *patAB* genes to be upregulated in *S. pneumoniae* 99-4039, with both CDSs apparently cotranscribed despite the intervening degenerate transposase sequence being encoded on the complementary strand of the genome (Figure 8.7).

The RNA-seq data also showed that the *pur* purine biosynthesis operon, adenylate kinase and guanine synthase were also more highly expressed in *S. pneumoniae* 99-4039, indicating a change in nucleotide metabolism. A similar rise in transcription is observed for the translation initiation protein genes *infA* and *infC*. By contrast, the translation elongation factor Ts gene *tsf* and the cotranscribed *rpsB* ribosomal protein CDS are expressed at a lower level in *S. pneumoniae* 99-4039, according to both RNA-seq and microarray data. RNA-seq data also indicates that the chaperones *dnaK*, *grpE* and *clpL* are transcribed at a lower level in 99-4039. Congruent with the DNA sequence coverage mapping, active transcription of both the ϕOXC141 lysogeny

module, and the prophage's genes for the replication and host cell lysis, was observed in the two strains (Figure 8.5C), suggesting a heterogenous state in which the phage is induced in some cells and dormant in others.

8.3.2 Phenotypic differences between *S. pneumoniae* 99-4038 and 99-4039

Surveys of clinical isolates have found variable levels of *patAB* expression (Garvey *et al.*, 2010). Those strains with elevated levels of expression are observed to have reduced susceptibility to fluoroquinolones and other antimicrobials, including linezolid (Feng *et al.*, 2009), reserpine (Garvey and Piddock, 2008), acriflavine, berberine, ethidium bromide (Robertson *et al.*, 2005) and the dye Hoescht 33352 (Garvey *et al.*, 2010). Expression of the efflux pump is normally induced by DNA damaging agents, including fluoroquinolones and mitomycin C, suggesting they may be part of a wider stress response that includes the induction of competence (El Garch *et al.*, 2010). In order to test whether such phenotypic differences between *S. pneumoniae* 99-4038 and 99-4039, the strains' resistance to a variety of antimicrobial compounds was tested using phenotype microarrays using the Omnilog platform (see Materials and Methods).

Nine arrays were used to test for susceptibility to a range of antimicrobial compounds, each at four different concentrations, and a tenth examined responses to a number of osmolytes. All 21 cases where a statistically significant difference between the strains could be detected resulted from *S. pneumoniae* 99-4039 exhibiting greater resistance to antimicrobial agents (Table 8.4), with the increased expression of *patAB* not appearing to affect the bacterium's membrane integrity, based on the osmolyte assays. The greatest difference was observed in the presence of the purine analogue 6-mercaptopurine, which caused a significantly greater inhibition of *S. pneumoniae* 99-4038's metabolism at three of the concentrations tested. As well as confirming the protection offered against acriflavine, the upregulation of *patAB* also appears to protect against the related compound proflavine and the nucleotide analogue 5,7-dichloro-8-hydroxyquinoline. Other planar, aromatic molecules with nitrogen substitutents causing differential respiration included three uncouplers (pentachlorophenol, crystal violet and 2,4-dinitrophenol) and three drugs prescribed as antidepressants (atropine, orphenadrine and amitriptyline).

Figure 8.6 Distribution of accessory genome loci. (A) The maximum likelihood phylogeny, as displayed in Figure 8.2. (B) The accessory genome loci, identified through alignment of the strains sequenced using 454 and capillary data, are displayed as alternating orange and brown blocks, grouped according to function. Underneath, a heatmap for each taxon indicates the level of Illumina sequence read mapping along these sequences. Blue indicates low levels of mapping, while red suggests higher levels of mapping.

The long chain quaternary ammonium compounds dodecyltrimethyl ammonium bromide, domiphen bromide and cetylpyridinium chloride were also found to cause lower respiration by the blood isolate by the screen. In this context, it is interesting to note that the one CDS highlighted as being co-conserved among genomes with *patAB*, SPNOXC06720 (Szklarczyk *et al*., 2011), which is also upregulated in the CSF strain (Table 8.2), contains a DegV domain. This motif appears to bind fatty acid chains, such as those present in these quaternary ammonium cations, and, from its similarity with related domains, may interact with the transporter. Hence this may explain the structural difference between these molecules and the others found to cause significant differences in metabolism between *S. pneumoniae* 99-4038 and 99-4039.

Figure 8.7 Expression of the *patAB* locus. (A) Annotation of the *patAB* gene cluster, with the intervening IS element remnant on the opposite strand, and the upstream *mutS* gene. The position of the polymorphism distinguishing 99-4038 and 99-4039 is indicated. (B) Graph showing Illumina read coverage from RNA-seq experiments. Each line shows the mean value from three replicates surrounded by a shaded area representing one standard error either side. The red and orange lines represent transcription of the two strands of the genome in 99-4038, and the blue lines represent the transcription of the two strands of the genome in 99-4039.

8.3.3   The mechanism of *patAB* upregulation

Analysis of the region upstream of the *patAB* operon revealed a strong promoter appearing to initiate transcription 69 nt upstream of the *patA* start codon. This 5' untranslated region is predicted to fold into a hairpin followed by a run of uridine residues, indicating that it could function as a terminator (Figure 8.8). This suggests a simple transcriptional attenuation mechanism; any compound that destabilizes this hairpin will increase the transcription of the downstream CDSs. The transcription of these genes is known to be increased by compounds that can intercalate nucleic acids, which have been found to disrupt RNA helices (Berman and Young, 1981). Hence this suggests a simple mechanism whereby any intercalating antimicrobial causes upregulation of an efflux pump capable of removing such planar compounds from the cell.

Figure 8.8 Leader sequence upstream of *patAB*. The graph represents the free energy of the most stable folded form of the 5' region of the *patAB* transcript, as it extends from the promoter, in 99-4038 (black line) and 99-4039 (red line). The positions of the polymorphism distinguishing the two strains, the Shine-Dalgarno sequence and translation initiation site are indicated by the dotted lines. The terminator-like hairpin structures formed by both transcripts as the RNA reaches the point at which translation initiation begins is displayed; it is evident that the mutation in 99-4039 is predicted to greatly weaken the stem of this secondary structure.

The *patAB* upregulatory SNP (PUS) distinguishing *S. pneumoniae* 99-4038 and 99-4039 is predicted to destablise this hairpin (Figure 8.8), thereby reducing any transcriptional attenuation and providing an explanation for the observed difference in expression levels. To test this hypothesis, the ~500 bp region around the PUS was cloned from each of *S. pneumoniae* 99-4038 and 99-4039, and transformed in *S. pneumoniae* TIGR4. These transformations were then grown in BHI broth with no selection, or containing 2 μg mL^{-1} ciprofloxacin. Only the culture transformed with the intergenic region from the CSF isolate generated mutants able to grow in the presence of ciprofloxacin (Figure 8.9). Sequencing the regions upstream of *patAB* in nine ciprofloxacin-resistant colonies, picked from three independent transformations with the leader sequence from the CSF isolate, revealed they had all acquired the PUS, whereas nine colonies picked from the same reactions with no selection all lacked it. The transcriptional profile of a transformant carrying the PUS (*S. pneumoniae* TIGR4^{PUS}) was compared to that of the otherwise isogenic parental strain under the same conditions as the comparison between *S. pneumoniae* 99-4038 and 99-4039. This confirmed the mutation caused the upregulation of the *patAB* operon; however, no other significant differences in expression were observed (Table 8.5).

8.4 Discussion

The broad spectrum of antimicrobial compounds against which PatAB appears to afford protection would seem to indicate the specificity of the efflux pump is relatively relaxed, removing a quite broad range of molecules from the cell. That 6-mercaptopurine appears to be the compound most effectively ejected from the cell suggests purine-like molecules (including fluoroquinolones) are the most efficient substrates of the pump; given the low specificity of its action, it seems likely purines themselves would also be pumped out of the cell. A hypothetical model for explaining the observed differences between *S. pneumoniae* 99-4038 and 99-4039 is outlined in Figure 8.10. Selection against this efflux of crucial metabolites may be the reason *patAB* expression is normally tightly regulated. The variability in expression of these genes observed in clinical isolates may reflect different abilities to adapt to such a detrimental side-effect of expression.

This hypothesis is supported by the apparently different responses of the clade I ST180 genotype and *S. pneumoniae* TIGR4 to the upregulation of *patAB*, although it is important to note that *S. pneumoniae* 99-4038 and 99-4039 are distinguished by a further two SNPs that may also impact on the differences between them. One possible reason for the increased sensitivity of the ST180 genotype to purine efflux may be deduced from Figure 8.5: the transcription of the lysogeny module of the φOXC141 prophage appears to extend across the adjacent *purA* purine synthesis gene in the antisense direction. This may inhibit *purA* expression and thereby limit the maximum possible rate of purine biosynthesis in the ST180 strains. However, in the light of the success of clade I, it would seem unlikely that such a mechanism would affect the fitness of the lineage under typical conditions.

Figure 8.9 The effect of the PUS. A sample of *S. pneumoniae* TIGR4 was transformed with the region upstream of *patAB* from either (A) and (C) *S. pneumoniae* 99-4038 or (B) and (D) *S. pneumoniae* 99-4039. The transformed cultures were then grown in (A) and (B) BHI or (C) and (D) BHI supplemented with 2 µg mL^{-1} ciprofloxacin. The bars show the mean optical density at 600 nm after 20 h growth at 37 °C, with the error bars showing one standard error of the mean.

Such a putative cost of a resistance mechanism exemplifies the potential selective advantage of the continued antibiotic susceptibility of ST180. It seems clear that the mucoid capsule causes some degree of genetic isolation, causing a barrier to transformation. Furthermore, based on comparisons with PMEN1 and the observation

that capsule inhibits phage binding to teichoic acid receptors, the mucoid polysaccharide is likely to block virus and conjugative element acquisition as well. However, rare instances of insertions in this dataset, and the existence of macrolide-resistant ST180 strains (Isozumi *et al*., 2008), suggest the barrier is not impermeable. Furthermore, it seems unlikely that the capsule would be the reason for the paucity of observed fluoroquinolone resistance mutations, given that such substitutions were observed in smaller surveys of the non-transformable species *Salmonella* Typhi (Holt *et al*., 2008) and *Staph. aureus* ST239 (Harris *et al*., 2010). A potentially important caveat is that the sample of ST180 strains is heavily biased towards isolates from Europe, where low levels of fluoroquinolone resistance have been observed (Reinert *et al*., 2005a), and there has been a previous report of an ST180 isolate developing fluoroquinolone resistance through such mutations during treatment (Perez-Trallero *et al*., 2003). However, it is notable that these resistant variants have not displaced their universally susceptible progenitors. Hence it seems that a different response to selection pressures, rather than a lack of opportunity, shapes the contrasting resistance profiles of PMEN1 and ST180.

Figure 8.10 Hypothetical model to explain the changes in gene expression resulting from the PUS. With increased expression of *patAB*, there is predicted to be increased efflux of purines, based on the results of the phenotype microarray results. In turn, this would lead to upregulation of the pathways that replenish the purine levels within the cell, explaining many of the observed results from the RNA-seq comparison between *S. pneumoniae* 99-4038 and 99-4039.

Table 8.1 Microarray analysis of differential expression between *S. pneumoniae* 99-4038 and 99-4039. CDSs found to be expressed at significantly different levels in *S. pneumoniae* 99-4038 and *S. pneumoniae* 99-4039 using a microarray based on *S. pneumoniae* TIGR4. The fold change value represents the ratio of expression level in *S. pneumoniae* 99-4039 to that in *S. pneumoniae* 99-4038.

TIGR4 CDS	OXC141 Orthologue	Gene	Gene Product	Fold Change	Corrected *p* Value
SP_2075	SPNOXC18290	*patB*	ABC transporter ATP-binding membrane protein	5.09	0.00880
SP_2073	SPNOXC18270	*patA*	ABC transporter ATP-binding membrane protein	4.79	0.00880
SP_0493	SPNOXC04590	*rpoE*	putative DNA-directed RNA polymerase, delta subunit	0.38	0.0385
SP_0492	-	-	-	0.30	0.0354
SP_0373	SPNOXC03690		putative RNA methylase family protein	0.28	0.0496
SP_0490	SPNOXC04570		putative uncharacterized protein	0.28	0.0496
SP_0649	-	-	-	0.27	0.0318
SP_0489	SPNOXC04560		PAP2 superfamily protein	0.27	5.62×10^{-5}
SP_1293	SPNOXC11570	*rplS*	50S ribosomal protein L19	0.25	1.00×10^{-4}
SP_0487	-	-	-	0.21	1.21×10^{-4}
SP_0488	SPNOXC04550		putative membrane protein	0.16	2.11×10^{-5}
SP_2214	SPNOXC19560	*tsf*	elongation factor Ts	0.11	2.66×10^{-6}
SP_2215	SPNOXC19570	*rpsB*	30S ribosomal protein S2	0.08	4.62×10^{-6}

Table 8.2 CDSs found to be significantly more highly expressed in *S. pneumoniae* 99-4039 than in *S. pneumoniae* 99-4038 using RNA-seq.

OXC141 ID	TIGR4 Orthologue	Gene	Product	Fold Change	Corrected *p* value
SPNOXC03970	SP_0409	-	putative decarboxylase	2.65	2.29×10^{-15}
SPNOXC18260	SP_2072	-	putative peptidase	2.99	4.54×10^{-7}
SPNOXC00860	SP_0046	*purF*	putative amidophosphoribosyltransferase precursor	2.94	2.81×10^{-6}
SPNOXC00850	SP_0045	-	putative phosphoribosylformylglycinamidine synthase protein	2.68	2.81×10^{-6}
SPNOXC00880	SP_0048	*purN*	phosphoribosylglycinamide formyltransferase	3.00	8.75×10^{-6}
SPNOXC18270	SP_2073	*patA*	ABC transporter ATP-binding membrane protein	1.95	2.40×10^{-5}
SPNOXC00900	SP_0050	*purH*	bifunctional purine biosynthesis protein	2.75	2.04×10^{-4}
SPNOXC05670	SP_0620	-	putative extracellular solute-binding protein	2.70	3.01×10^{-4}
SPNOXC18290	SP_2075	*patB*	ABC transporter ATP-binding membrane protein	2.04	4.01×10^{-4}
SPNOXC02530	SP_0231	*adk*	adenylate kinase	1.69	0.00143
SPNOXC08640	SP_0962	*gloA*	putative lactoylglutathione lyase	1.52	0.00143
SPNOXC03060	SP_0288	-	putative CAAX amino terminal protease family membrane protein	2.90	0.00210
SPNOXC00840	SP_0044	*purC*	phosphoribosylaminoimidazole-succinocarboxamidesynthase	2.98	0.00497
SPNOXC02580	SP_0237	*rplQ*	50S ribosomal protein L17	1.56	0.00497
SPNOXC00890	SP_0049	*vanZ*	putative VanZ-family resistance protein	2.42	0.00794
SPNOXC03050	SP_0287	-	putative permease	1.67	0.0101
SPNOXC11890	SP_1355	*rplJ*	50S ribosomal protein L10	1.17	0.011
SPNOXC08630	SP_0961	*rplT*	50S ribosomal protein L20	1.45	0.011
SPNOXC06720	SP_0742	-	putative fatty-acid binding protein	1.38	0.011
SPNOXC19520	SP_2210	*cysM*	putative cysteine synthase	2.21	0.014
SPNOXC04660	SP_0502	*glnA*	putative glutamine synthetase	1.36	0.017
SPNOXC00930	SP_0054	*purK*	putative phosphoribosylaminoimidazole carboxylase ATPase subunit	1.51	0.018
SPNOXC02540	SP_0232	*infA*	translation initiation factor IF-1	1.27	0.021
SPNOXC12710	SP_1445	*guaA*	GMP synthase [glutamine-hydrolyzing]	1.26	0.022
SPNOXC13390	SP_1527	-	putative extracellular oligopeptide-binding protein	1.28	0.033
SPNOXC08610	SP_0959	*infC*	translation initiation factor IF-3	1.34	0.039
SPNOXC02570	SP_0236	*rpoA*	DNA-directed RNA polymerase alpha chain	1.53	0.044

Table 8.3 CDSs found to be significantly less highly expressed in *S. pneumoniae* 99-4039 than in *S. pneumoniae* 99-4038 using RNA-seq.

OXC141 ID	TIGR4 ID	Gene	Product	Fold Change	Corrected *p* Value
SPNOXC05950	SP_2314	-	putative uncharacterized protein	0.27	1.46×10^{-7}
SPNOXC03690	SP_0373	-	putative RNA methylase family	0.38	1.46×10^{-7}
SPNOXC03700	SP_0374	-	putative membrane protein	0.41	1.46×10^{-7}
SPNOXC19560	SP_2214	*tsf*	elongation factor Ts	0.37	2.81×10^{-6}
SPNOXC04610	SP_0496	-	putative Na^+/Pi-cotransporter protein	0.44	3.10×10^{-6}
SPNOXC19570	SP_2215	*rpsB*	30S ribosomal protein S2	0.27	5.37×10^{-6}
SPNOXC04550	SP_0488	-	putative membrane protein	0.44	5.86×10^{-6}
SPNOXC04560	SP_0489	-	PAP2 superfamily protein	0.42	1.06×10^{-5}
SPNOXC04770	SP_0517	*dnaK*	chaperone protein DnaK (heat shock protein 70)	0.44	3.21×10^{-5}
SPNOXC05940	SP_2313	-	putative uncharacterized protein	0.26	6.60×10^{-5}
SPNOXC16430	SP_1872	-	siderophore uptake periplasmic	0.40	2.25×10^{-4}
SPNOXC16410	SP_1870	-	putative iron compound ABC	0.40	0.00111
SPNOXC04570	SP_0490	-	putative uncharacterized protein	0.47	0.00111
SPNOXC14710	SP_1674	-	putative transcription regulator	0.55	0.00139
SPNOXC16420	SP_1871	-	siderophore uptake ATP-binding	0.40	0.00210
SPNOXC14290	SP_1626	*rpsO*	30S ribosomal protein S15	0.52	0.00210
SPNOXC04780	-	-	putative membrane protein	0.56	0.00210
SPNOXC07270	SP_0800	-	putative membrane protein	0.49	0.00407
SPNOXC18140	SP_2058	*tgt*	queuine tRNA-ribosyltransferase	0.65	0.00448
SPNOXC04580	SP_2283	-	acetyltransferase (GNAT) family	0.47	0.00855
SPNOXC04590	SP_0493	*rpoE*	putative DNA-directed RNA	0.63	0.0101
SPNOXC03470	SP_0338	*clpL*	putative ATP-dependent protease	0.66	0.0111
SPNOXC16320	SP_1859	-	nicotinamide mononucleotide	0.41	0.0165
SPNOXC05800	SP_0631	*rplA*	50S ribosomal protein L1	0.65	0.0216
SPNOXC04760	SP_0516	*grpE*	GrpE protein (HSP-70 cofactor)	0.79	0.0216
SPNOXC11570	SP_1293	*rplS*	50S ribosomal protein L19	0.64	0.0281

Table 8.4 Comparison of *S. pneumoniae* 99-4038 and 99-4039 using phenotype microarrays. Antimicrobial compounds found to result in significantly higher levels of respiration with *S. pneumoniae* 99-4039 than with *S. pneumoniae* 99-4038 on the Omnilog platform.

Antimicrobial	Mechanism	*p* value
Atropine	Acetylcholine receptor agonist	0.015
Aminotriazole	Amino acid analogue	0.011
Dodecyltrimethyl ammonium bromide	Cationic detegent	0.081
Domiphen bromide	Cationic detergent	0.046
Cetylpyridinium chloride	Cationic detergent	0.049
Cefsulodin	Cephalosporin	0.017
Orphenadrine	Cholinergic agonist	0.017
Nordihydroguaiaretic acid	Glutathione depletion	0.011
Acriflavine	Intercalator	0.011
5,7-Dichloro-8-hydroxyquinoline	Intercalator	0.011
Proflavine	Intercalator	0.011
3,4-Dimethoxybenzylalcohol	Oxidising agent	0.011
6-Mercaptopurine	Purine analogue	0.011
6-Mercaptopurine	Purine analogue	0.018
6-Mercaptopurine	Purine analogue	0.018
Amitriptyline	Serotonin reuptake inhibitor	0.018
Boric acid	Toxic anion	0.011
Sodium metaborate	Toxic anion	0.011
Sodium bromate	Toxic anion	0.011
D,L-Methionine hydroxymate	tRNA synthetase inhibitor	0.018
Pentachlorophenol	Uncoupler	0.011
Crystal violet	Uncoupler	0.011
2,4-Dinitrophenol	Uncoupler	0.024

Table 8.5 Microarray analysis of differential expression between *S. pneumoniae* TIGR4 and TIGR4^{PUS}. CDSs found to be significantly more highly expressed in *S. pneumoniae* TIGR4^{PUS} relative to the parental strain.

TIGR4 CDS	OXC141 Orthologue	Gene	Gene Product	Fold Change	Corrected *p* Value
SP_2075	SPNOXC18290	*patB*	ABC transporter ATP-binding membrane protein	9.64	0.00147
SP_2073	SPNOXC18270	*patA*	ABC transporter ATP-binding membrane protein	6.30	0.00147
SP_2074	SPNOXC18280	-		3.60	0.00481
SP_0455	-	-	Hypothetical protein	2.40	0.0291