3 The genome of *S. pneumoniae* ATCC 700669

3.1 Introduction

3.1.1 The spread of the PMEN1 lineage

The PMEN1, or Spain^{23F}-1 lineage, was one of the first multidrug-resistant pneumococcal clones to be recognised as having spread across the globe. Since the early 1980s, serogroup 23 isolates resistant to penicillin, chloramphenicol and tetracycline had been observed in Spain (Linares *et al*., 1983; Latorre *et al*., 1985) and the UK (George *et al*., 1981). The clonal nature of a number of these European isolates, along with a serogroup 19 strain with an identical resistance profile, was initially demonstrated using MLEE (Coffey *et al*., 1991), and, using the same technique, it was found that this lineage had spread to the USA by the late 1980s (Munoz *et al*., 1991). In the 1990s, both MLEE and PFGE were used to identify PMEN1 isolates in South America (Camou *et al*., 1998; Castaneda *et al*., 1998; Echaniz-Aviles *et al*., 1998), Asia (Tarasi *et al*., 1997) and South Africa (Sibold *et al*., 1992). Soon after its introduction, MLST was used to identify this clone, with an ST of 81, in a hospital in Taiwan (Shi *et al*., 1998).

By the late 1990s, PMEN1 was responsible for almost 40% of penicillin-resistant pneumococcal disease in the USA (Corso *et al*., 1998), and about 30% of penicillin-resistant paediatric disease in Latin America (Tomasz *et al*., 1998). During its spread, it expanded its repertoire of antibiotic resistances, being frequently associated with the acquisition of macrolide (Reinert *et al*., 2005b) and fluoroquinolone (Pletz *et al*., 2004) resistance. The lineage has also proved proficient at acquiring different serotypes: serotype 19F (Coffey *et al*., 1991; Coffey *et al*., 1998a), 19A (Coffey *et al*., 1998b) and 14 (Barnes *et al*., 1995) variants were detected in the 1990s. Although serotypes 23F, 19F and 14 have decreased in prevalence since the introduction of PCV7, the increasing incidence of multidrug-resistant serotype 19A disease that has followed the vaccine's use has included switched PMEN1 isolates (Moore *et al*., 2008; Munoz-Almagro *et al*., 2008; Ardanuy *et al*., 2009).

While serotype 23F strains have consistently been found to cause disease at a low level relative to their carriage prevalence, the opposite is true of serotype 4 (Brueggemann *et al*., 2004), and serotype 2, which, although no longer common in most countries, was recently found to be the most common cause of pneumococcal meningitis in Bangladesh (Saha *et al*., 2009). Hence if it is correct that serotype more strongly determines the proclivity of a strain to cause invasive disease than the rest of the bacterial genotype, then differences between the PMEN1 genome and those of TIGR4 (serotype 4) and D39 (serotype 2) may represent adaptations to causing disease at different frequencies, and being carried for different lengths of time (Sleeman *et al*., 2006). However, this may be unlikely, given that isolates with the PMEN1 genotype have been observed expressing serotypes 14 and 19A, both associated with high odds ratios for causing disease (Brueggemann *et al*., 2003; Brueggemann *et al*., 2004).

3.1.2 *S. pneumoniae* ATCC 700669

S. pneumoniae 264 was a serotype 23F isolate from the Hospital de Bellvitge, Barcelona, in 1984 (Coffey *et al*., 1991). Found to be resistant to benzylpenicillin, chloramphenicol and tetracycline, it was submitted to the American Type Culture Collection (ATCC) as the representative type strain of the PMEN1 lineage, thereafter designated *S. pneumoniae* ATCC 700669.

3.2 Description of the genome

3.2.1 Features of the chromosome

The genome of *S. pneumoniae* ATCC 700669 was sequenced using dideoxy terminator technology. The complete circular chromosome of 2,221,315 bp (39.49% GC content; Figure 3.1) contains four rRNA operons and 58 tRNA genes (all but 12 of which are adjacent to rRNA genes). An unusual asymmetry in the GC skew of the chromosome, resulting from the recent integration of a prophage and an ICE into the same replichore, is evident. There are 2,135 predicted coding sequences (CDSs), 144 (6.7%) of which appear to be pseudogenes. In common with other Firmicutes, there is a strong coding bias, with 76% of the CDSs on the leading strand. Overall, 197

(9.2%) of the predicted CDSs in ATCC 700669 are not present in TIGR4 or D39, the majority of these being present on an ICE or prophage-derived sequences. Seventy-nine IS elements can be identified, of which 73% appear to be non-functional due to disruptive mutations. Twenty-one of the IS insertions are not present in the TIGR4 or D39 genomes, the majority of which are due to IS*1167* or IS*1167A*-type elements. Both of these IS families exhibit relatively little sequence diversity when compared with others present in the chromosome, in accordance with the previously proposed hypothesis that IS isotypes spread through short bursts of transposition (Tettelin *et al.*, 2001).

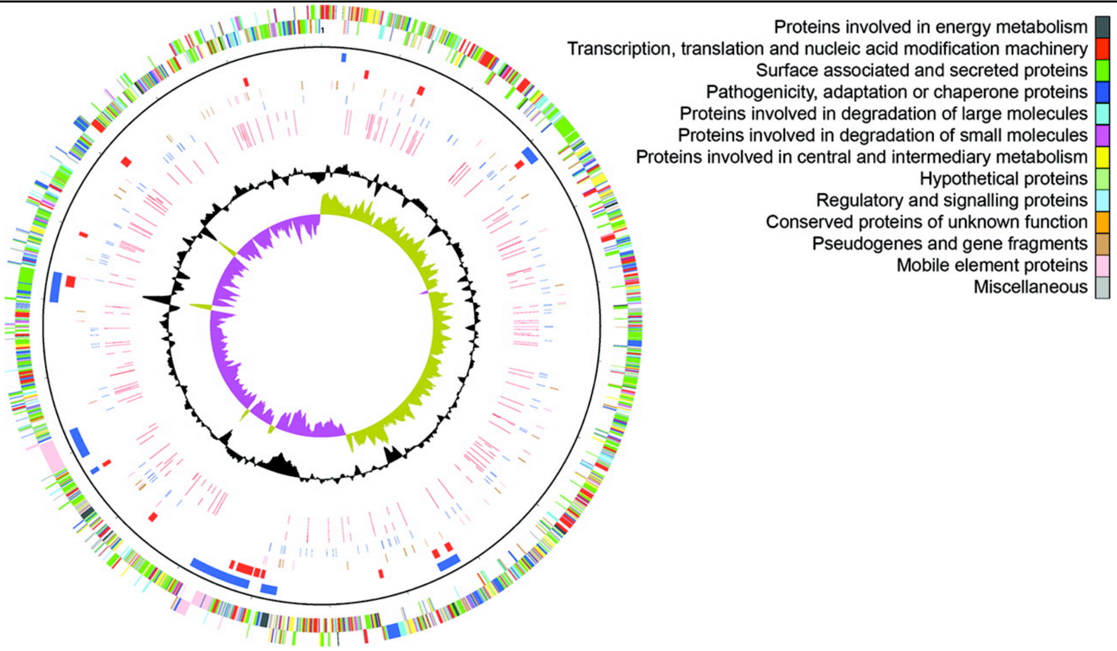
Proteins involved in energy metabolism
Transcription, translation and nucleic acid modification machinery
Surface associated and secreted proteins
Pathogenicity, adaptation or chaperone proteins
Proteins involved in degradation of large molecules
Proteins involved in degradation of small molecules
Proteins involved in central and intermediary metabolism
Hypothetical proteins
Regulatory and signalling proteins
Conserved proteins of unknown function
Pseudogenes and gene fragments
Mobile element proteins
Miscellaneous

Figure 3.1 Circular diagram representing the *S. pneumoniae* ATCC 700669 chromosome, arranged to have the origin of replication at the top, as indicated by the GC deviation (innermost graph). The outer rings show the arrangement of coding sequences on the two strands of the genome, coloured according to annotated function (see key). The first inner ring indicates the major variable regions as blue blocks: moving clockwise from the origin of replication, these represent the prophage remnant, *cps* locus, PPI-1, the Na$^+$-dependent ATP synthase island, ICE*Sp*23FST81, the *pclA* gene cluster, ϕMM1-2008, and the *psrP* gene cluster. The red blocks demarcate loci identified as having atypical nucleotide composition by the Alien Hunter algorithm (Vernikos and Parkhill, 2006). Moving inward, the rings show the position of IS elements (pink if intact, brown if pseudogene), RUP repeats (blue), and BOX A, B, and C repeat modules (red), respectively (see Chapter 7). The black graph indicates sequence GC content, and the innermost green and purple graph shows the GC skew of the sequence.

Of the annotated genes, 29% are predicted to encode surface exposed or secreted proteins. These include 18 phosphotransferase system (PTS) transporters, 17 of which are shared with both TIGR4 and D39, which have four and three PTS transporters not present in ATCC 700669, respectively. The one system unique to ATCC 700669

(SPN23F18210-18250) forms part of a ~10-kb insertion that also includes a putative choline sulphatase. Given the importance of choline to pneumococcal metabolism and pathogenesis, this transporter could represent a novel means of acquiring this nutrient from the host environment. There are also two ABC transporter systems within the ATCC 700669 genome that are absent from both TIGR4 and D39. One of these, present on a ~4-kb insertion along with two putative secreted peptides (SPN23F07060-07090), is similar to systems present in *S. pyogenes* MGAS6180, *S. pyogenes* MGAS10270, and *S. sanguinis* SK36.

3.2.2 Genes implicated in pathogenesis

It has been shown that the *pspA* gene of a serotype 23F strain contained a frameshift mutation that truncated the encoded protein to a secreted, rather than surface-associated, form (McCool *et al.*, 2002). A similar frameshift mutation, caused by variation in the length of the same polyadenosine tract, is observed in this genome. Furthermore, ATCC 700669 lacks the *zmpC* gene, which encodes a metalloprotease that cleaves and activates human matrix metalloprotease 9 and has been implicated in the pulmonary invasion process (Oggioni *et al.*, 2003). However, like *S. pneumoniae* G54, *S. pneumoniae* ATCC 700669 encodes an additional surface-exposed zinc metalloprotease, *zmpD* (Camilli *et al.*, 2006), adjacent to the immunoglobulin A protease gene *zmpA*. Notably absent from the ATCC 700669 genome are the genes encoding two more surface-exposed proteins, choline-binding protein *cbpC* and histidine triad protein *phtB*, which appear to be required for full virulence of TIGR4 in the mouse model, based on signature-tagged mutagenesis data (Hava and Camilli, 2002).

Notably present in the genome are both loci associated with invasive clones by Obert *et al* (Obert *et al.*, 2006): the V-type Na^+-driven ATPase gene cluster and the island encoding *psrP*. Though ATCC 700669 lacks both identified pneumococcal pilus synthesis gene clusters (Barocchi *et al.*, 2006; Bagnoli *et al.*, 2008), *pclA* is present, although the D39 orthologue is ~44% longer due to an expansion in the internal repetitive region of the protein.

3.2.3 Prophage

Although the *S. pneumoniae* G54 and D39 genomes are devoid of prophage, a 10.5-kb phage remnant can be found between the *eno* and *rexB* genes in the chromosome of TIGR4 (Obregon *et al*., 2003). The ATCC 700669 genome contains an intact 39.1-kb prophage, ϕMM1-2008, as well as a smaller prophage remnant. ϕMM1-2008 is more than 97% identical, at the nucleotide sequence level, to both ϕMM1 and ϕMM1-1998. Since ϕMM1 and ϕMM1-2008 are from hosts of the same serotype and sequence type (Obregon *et al*., 2003), they are probably descended from the same insertion event, while the host of ϕMM1-1998 is a penicillin-sensitive serotype 24 strain (Loeffler and Fischetti, 2006); hence, this prophage is likely to be the result of a different infection. All three phage are present in the same locus, between a pyridine nucleotide-disulfide oxidoreductase gene and a CDS of unknown function. The exact insertion site of the phage appears to be within the 3' region of the downstream hypothetical gene; duplication of this 15-bp *att* sequence (Gindreau *et al*., 2000) upon integration maintains the full-length target gene sequence and generates the tandem repeats either side of the prophage.

The 6.4-kb prophage remnant is flanked by genes encoding a putative cytidine deaminase and a deoxyuridine 5' triphosphate nucleotidohydrolase. Along with CDS of phage origin, including integrase and amidase pseudogenes, the prophage appears to carry ‘cargo’ genes that have been retained as the replicative machinery of the virus has degenerated. One of these CDSs encodes a type I restriction endonuclease domain and seems to be a member of a family of uncharacterized genes found in a range of bacterial species. Another is an addiction system toxin gene, which may have inhibited the clearance of the remainder of the prophage from the genome. Although a cognate antitoxin cannot be reliably identified, the overlapping upstream gene is a good candidate, as alignments with intact prophage indicate these CDSs have been acquired as a pair. A complete 37.5-kb long prophage can be found at this locus in the draft genome of *S. pneumoniae* 18-BS74 (Hiller *et al*., 2007), suggesting it may be a common target insertion site for temperate pneumophage.

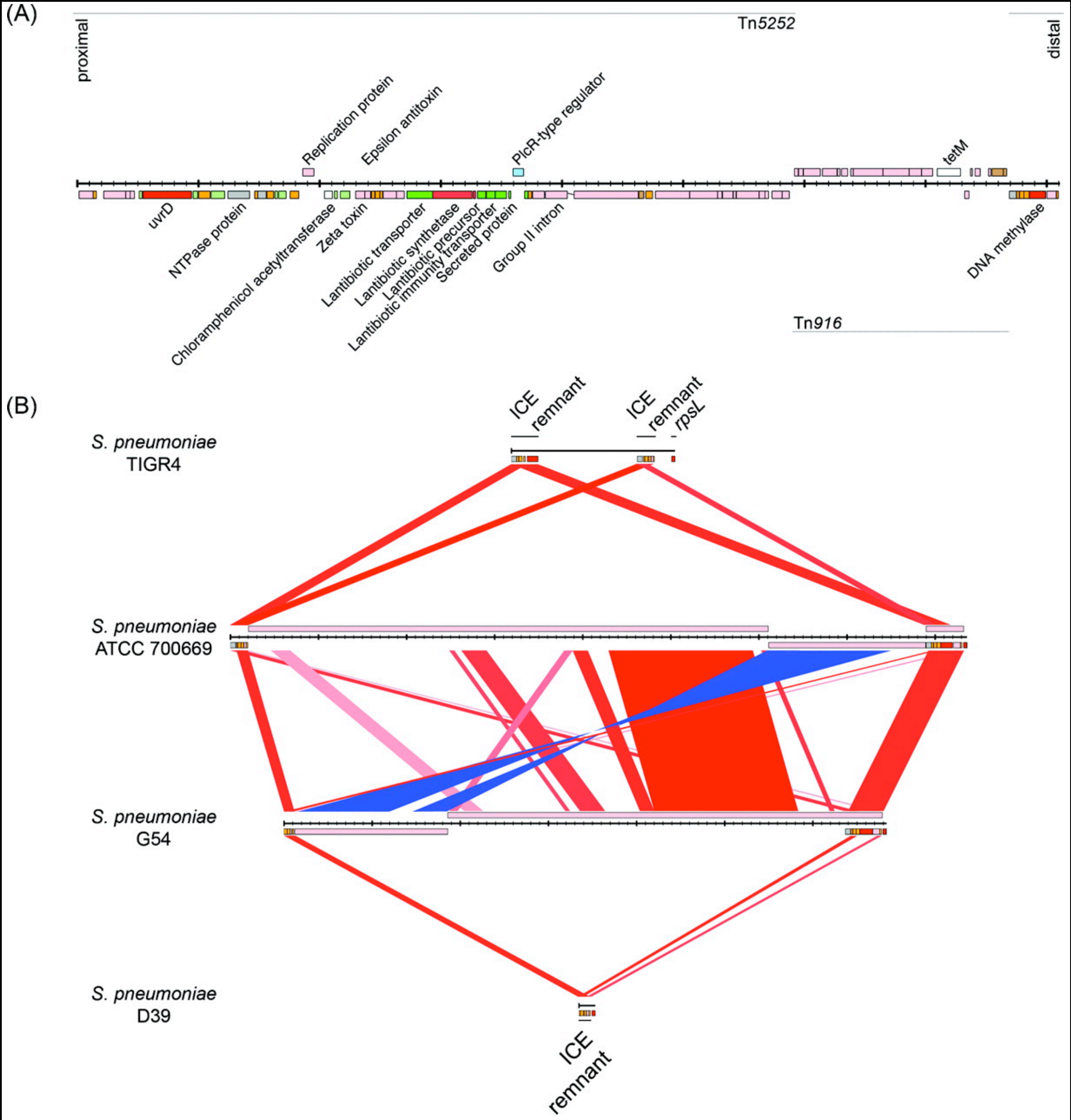

Figure 3.2 (A) Representation of ICE*Sp*23FST81. 'Cargo' genes are coloured according to the scheme detailed in Fig. 3.1 and are labelled with their putative function, where one can be assigned. The division of the ICE into Tn*5252* and Tn*916*-type elements is indicated by the bars at the top and bottom of the diagram. (B) Comparison of the ICE insertion sites in the ATCC 700669 and G54 genomes with the corresponding loci in the TIGR4 and D39 genomes, which lack intact ICEs. The ICEs are represented as pink blocks: Tn*5252*-type elements are represented by blocks above the scale line, while Tn*916*-type elements are represented by blocks below the scale line. Red bands indicate BLAST matches between genomes in the same direction, whereas blue bands indicate matches in opposite directions. The intensity of the band represents the strength of the match. The region of each genome shown is bounded at the 5' end by ICE-derived sequences and at the 3' end by *rplL*, which streptococcal ICEs frequently insert directly upstream of. The ICE remnants in the TIGR4 and D39 genomes, apparently derived from the distal region of the Tn*5252*-type element, are marked. The alignment shows that there is a remnant in D39 directly adjacent to *rplL*, at the point where the elements usually insert, but the two remnants in the TIGR4 genome are further removed upstream.

3.2.4 ICE*Spn*23FST81

The ~81-kb ICE of ATCC 700669, ICE*Sp*23FST81, is a composite element comprising a Tn*916*-like component inserted into a Tn*5252*-like transposon, with the latter consequently being split into a larger proximal region and a smaller distal region (Figure 3.2). This combination of conjugative transposons is common in streptococcal ICEs, although the variation observed in the arrangement of these two elements implies it has arisen independently on a number of occasions (Figure 3.3). This suggests a potential symbiotic advantage between these different transposon types, perhaps resulting from a synergistic combination of the two sets of conjugative machinery or 'cargo' genes. Flanked by a 16-bp tandem duplication, ICE*Sp*23FST81 is inserted near the 3' end of *rplL*, in the same position as that of *S. pneumoniae* G54 and many other streptococcal ICEs, although Tn*5253*, the partially sequenced ICE of *S. pneumoniae* BM6001 (Ayoubi *et al*., 1991), appears to have integrated elsewhere.

Shortly upstream of *rplL* in the TIGR4 and D39 genomes are ~1.8-kb long ICE remnants that are >80% identical, at the nucleotide level, to the distal Tn*5252*-type region of ICE*Sp*23FST81 (Figure 3.2). Similar remnants are also seen in the ATCC 700669 and G54 genomes immediately upstream of their ICE insertions. A second, larger, ICE remnant is also evident in the TIGR4 genome, ~15 kb upstream of *rplL*. This includes a cytosine methyltransferase gene very similar to homologues in the distal region of ICE*Sp*23FST81, on Tn*5253* and in the ϕMM1 phage (85%, 85% and 45% protein sequence identity, respectively). The presence of this gene on conjugative transposons and prophage suggests it may aid the horizontal transfer of both between pneumococci, perhaps through methylating DNA prior to transfer between cells and hence avoiding the recipient's restriction systems.

ICE*Sp*23FST81 is clinically important due to its genetic 'cargo'. The Tn*916*-type component carries a *tetM* gene, responsible for the strain's tetracycline resistance. A similar Tn*916* element, also carrying the *mef(A)* macrolide resistance gene, has been detected in the gammaproteobacterial commensal and emerging nosocomial pathogen *Acinetobacter junii* (Ojo *et al*., 2006). The ~1.2-kb flanking sequences that were determined on either side of the *A. junii* Tn*916* transposon are 99% identical, at the nucleotide level, to the Tn*5252*-type sequences surrounding the Tn*916* transposon on ICE*Sp*23FST81, suggesting these composite elements can transfer between distantly related bacteria. The Tn*5252*-like component carries a gene for chloramphenicol acetyltransferase, which appears to have been acquired through wholesale integration of the pC194 plasmid (Widdowson *et al*., 2000) originally identified in chloramphenicol-resistant *Staph. aureus*.

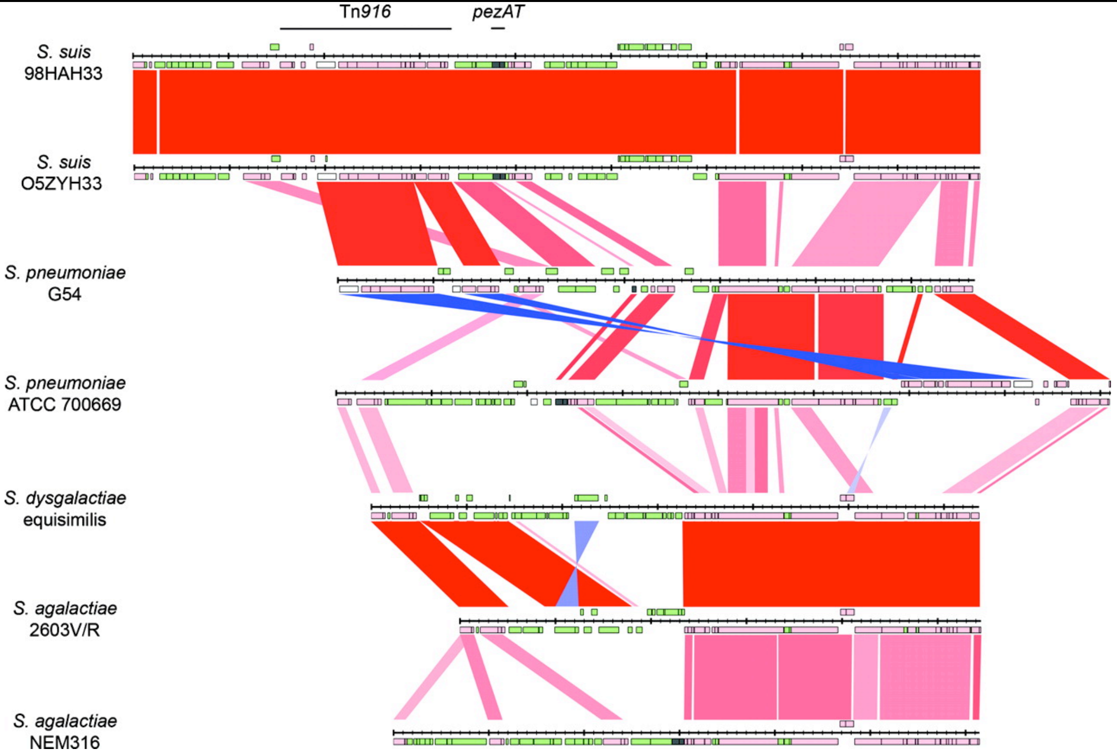

Figure 3.3 Comparison of streptococcal integrative and conjugative elements. Genes likely to be part of the conjugative machinery of the element, on the basis of conservation or functional assignment, are coloured pink. Zeta toxin-epsilon antitoxin systems are coloured grey, antibiotic resistance genes are white, and other 'cargo' genes are green. Bands are coloured as in Fig. 3.2. Unlike the *S. suis* and *S. pneumoniae* elements, the *S. agalactiae* and *S. dysgalactiae* Tn*5252*-like elements do not contain a Tn*916*-type component, which carries the *tetM* gene responsible for tetracycline resistance in the other ICE.

One of the 'cargo' genes found toward the 5' end of the element is a *uvrD* helicase gene, with the closest sequenced homologue being that of the deltaproteobacterium *Geobacter lovleyi* (26% protein sequence identity). A different *uvrD* gene, with a dissimilar sequence (15% protein sequence identity), is present at the equivalent position on the G54 ICE. Streptococci lack an SOS response (Erill *et al*., 2007), so consequently *uvrD* is absent from the *S. pneumoniae* core genome, but horizontal acquisition of this gene could potentially reconstitute the nucleotide excision repair pathway if it were able to act in concert with the *uvrABC* genes shared by all pneumococci. This pathway is important in the repair of peroxidative damage to DNA

(Moller and Wallin, 1998); therefore, given that *S. pneumoniae* is catalase negative and produces hydrogen peroxide, which can function as an antimicrobial (Pericone *et al.*, 2000), the gain of the *uvrD* gene may increase the tolerance of ATCC 700669 to reactive oxygen species and hence aid nasopharyngeal colonization, while also resulting in the ICE maintaining its sequence integrity within the host. The other major branch of the SOS response, also absent from the core genome of pneumococci, is mutagenic lesion repair, which requires a reduction in DNA polymerase III replication fidelity caused by an interaction with the UmuC-UmuD complex. Correspondingly, Tn*5252* carries a *umuCD*-containing operon that was demonstrated to increase the UV tolerance of the host bacterium (Munoz-Najar and Vijayakumar, 1999), suggesting ICE-carried genes can functionally restore at least one aspect of the SOS response.

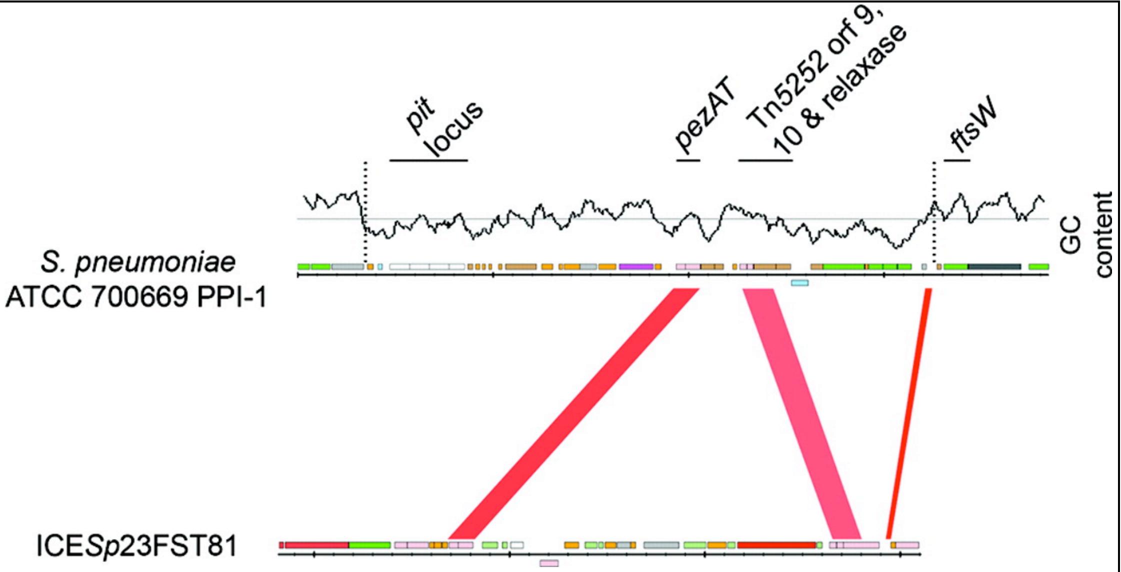

Figure 3.4 Comparison of PPI-1 of ATCC 700669 with ICE*Sp*23FST81 found in the same strain. Three regions of apparently conserved similarity are observed: the zeta toxin-epsilon antitoxin system, a group of Tn*5252* conjugative transfer genes (encoding ORF9, ORF10, and relaxase proteins), and a short stretch of DNA adjacent that forms the 3' border of PPI-1, shortly upstream of the cell division gene *ftsW* in the pneumococcal chromosome. The GC content of this region is shown, with the line across the graph indicating the average for the region (33.71% GC). The vertical dotted lines on the graph delimit the extent of PPI-1.

3.2.5 An ICE-derived genomic island

Genes characteristic of streptococcal ICE are also found in another region of the genome, within the putative Pnuemococcal Pathogenicity Island 1 (PPI-1), as described by Brown *et al.* (Brown *et al.*, 2004). In *S. pneumoniae* ATCC 700669, this

~30-kb region, as defined by its low GC content (SPN23F09511-09860; Figure 3.4), contains the *pezAT* epsilon toxin-zeta antitoxin system, found on ICE*Sp*23FST81, as well as related elements in *S. suis* and *S. agalactiae* strains, and a cluster of three Tn*5252* conjugative machinery genes, including a relaxase and a MobC-domain protein. At the 3' end, coinciding with the edge of the low GC region, there is a further ~200-bp region of similarity (>90% identity at the nucleotide level) with ICE*Sp*23FST81, which is also shared with Tn*5253* and the ICE of the recently sequenced *S. pneumoniae* CGSP14 strain (Ding *et al.*, 2009). A site-specific recombinase, similar to one found in *S. suis* ICEs, is found adjacent to this sequence in some pneumococcal strains, such as *S. pneumoniae* 14-BS69. Hence, it appears likely that this island originated as an ICE insertion that has subsequently degenerated, with the loss of genes required for the element's autonomous mobility. The gene clusters located between these ICE-like regions are very different in many of the strains for which genome data are available (Figure 3.5), suggesting this locus may be able to diversify through exchange of sequences with ICE via homologous recombination in the shared regions.

A further source of variation appears to be the extent to which the conjugative machinery at this locus has degenerated (Figure 3.5), with *S. pneumoniae* 18-BS74 missing all of the ICE-like regions of the island. In contrast, the 5' end of PPI-1, containing the *pit* iron transporter operon crucial for virulence of *S. pneumoniae* (Brown *et al.*, 2001), is conserved among all strains. However, it seems likely the *pit* genes were acquired as part of the original ICE insertion, since they also lie within the low-GC-content region and, despite being ubiquitous among sequenced pneumococci, appear to be absent from other streptococci (Brown *et al.*, 2001).

The ability to exchange sequences with conjugative elements may play a role in allowing this locus to rapidly evolve in response to changing selection pressures. In strains that have retained the variable region, CDSs similar to genes encoding daunorubicin efflux transporters, inhibitor-resistant methionyl tRNA synthetases, macrolide efflux pumps, and apparently truncated aminoglycoside phospho- and acetyltransferases are found. Furthermore, two of the sequenced Pittsburgh disease isolates (9-BS69 and 14-BS69) seem to carry chloramphenicol acetyltransferase genes on the island. This is at least the fourth documented case in which Tn*5252*-type elements appear to have contributed to the transfer of chloramphenicol acetyltransferase genes to pneumococci. Tn*5253* of *S. pneumoniae* BM6001 and ICE*Sp*23FST81, both intact, potentially mobile, elements, carry chloramphenicol acetyltransferase genes at different positions. The PPI-1 locus and the IQ complex of *S. pneumoniae* 529, a genomic island containing Tn*5252*- and Tn*916*-like fragments along with chloramphenicol acetyltransferase and macrolide resistance genes (Mingoia *et al*., 2007), both appear to be fragments of conjugative transposons that have been integrated into the chromosome, suggesting that exchange of DNA between ICE and the pneumococcal genome is likely to be an important mechanism in the dissemination of antibiotic resistance throughout the *S. pneumoniae* population.

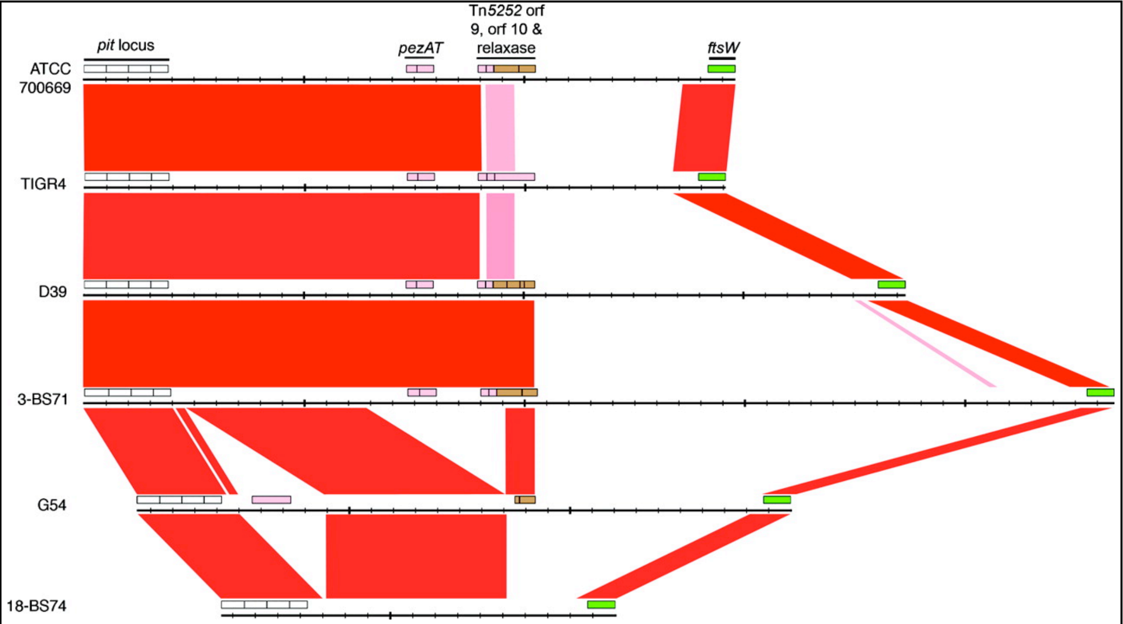

Figure 3.5 Alignment of PPI-1 from complete and draft pneumococcal genome data. Variation in the regions intervening between the relaxase and 3' end of PPI-1 (these boundaries both have sequence homology with Tn*5252*-type ICE) appears to be due to horizontal gene transfer, indicating that conjugative elements may contribute to the diversity found within this island through homologous recombination-mediated exchange. There is also variation putatively resulting from degeneration of the original conjugative element insertion: the PPI-1 of *S. pneumoniae* G54 has lost the *pezAT* toxin-antitoxin system and much of the cluster of conjugative transfer genes (although it has retained a CDS encoding a MobA-domain protein, which are typically associated with autonomously mobile elements, indicated in pink), while that of 18-BS74 appears to have lost all vestiges of the original element's conjugative machinery.

In *S. pneumoniae* ATCC 700669, PPI-1 contains an apparently incomplete lantibiotic synthesis operon. Lantibiotics (lanthionine-containing antibiotics) are small, secreted cyclic peptides containing lanthionine rings formed by the stereospecific intramolecular addition of cysteine to dehydrated serine or threonine residues (Willey and van der Donk, 2007). They frequently function as bacteriocins, with different types hypothesized to operate through inhibiting peptidoglycan transglycoslyation or forming pores in cell membranes, but are also known to act as biosurfactants and phospholipase A2 inhibitors (Willey and van der Donk, 2007). The gene cluster present in the ATCC 700669 PPI-1 lacks a structural prepeptide gene but retains the CDS necessary for immunity. Comparison with the same locus in the serogroup 23 Pittsburgh isolate genome reveals a ~6-kb deletion in the ATCC 700669 gene cluster (Figure 3.6). The sequence absent in ATCC 700669 is flanked by thymidine dinucleotides and encodes two lantibiotic structural genes and fragments of two dehydratases. The putative product of the intact locus is a novel dimeric lantibiotic (Figure 3.6) likely to be similar to those produced by Bacillus and Lactococcus species (Willey and van der Donk, 2007).

3.2.6 Pneumocidins

In ATCC 700669, the bacteriocin-producing *blp* locus has undergone a rearrangement relative to that in TIGR4 (Figure 3.7), resulting in the deletion of *blpM* and *blpN*, both of which are required for *blp*-encoded bactericidal activity (Dawid *et al*., 2007; Lux *et al*., 2007). Given the high level of nasopharyngeal carriage of PMEN1 strains, it seems likely that loci elsewhere in the genome are able to compensate for the loss of this important pneumocidin. Consistent with this suggestion, bacteriocin production by PMEN1 strains has been observed to inhibit the growth of a larger number of indicator strains than a penicillin-sensitive serotype 23F isolate (Lux *et al*., 2007). In addition to the defunct operon in the ATCC 700669 PPI-1, ICE*Sp*23FST81 itself carries an intact lantibiotic synthesis gene cluster. The structural gene appears to be a novel group II lantibiotic, most closely related to mersacidin and lichenicidin (SPN23F12701; Figure 3.6), on the basis of its probable ring structure, diglycine cleavage motif, and net neutral charge. Adjacent to this operon is the probable transcriptional regulator, which is similar to *plcR* of Bacillus species. Such transcription factors can form a minimal extracellular signalling system in conjunction with peptide autoinducers (Slamti and Lereclus, 2002) and, in keeping with such a role, a small secreted peptide is found between the lantibiotic synthesis genes and the *plcR* homologue.

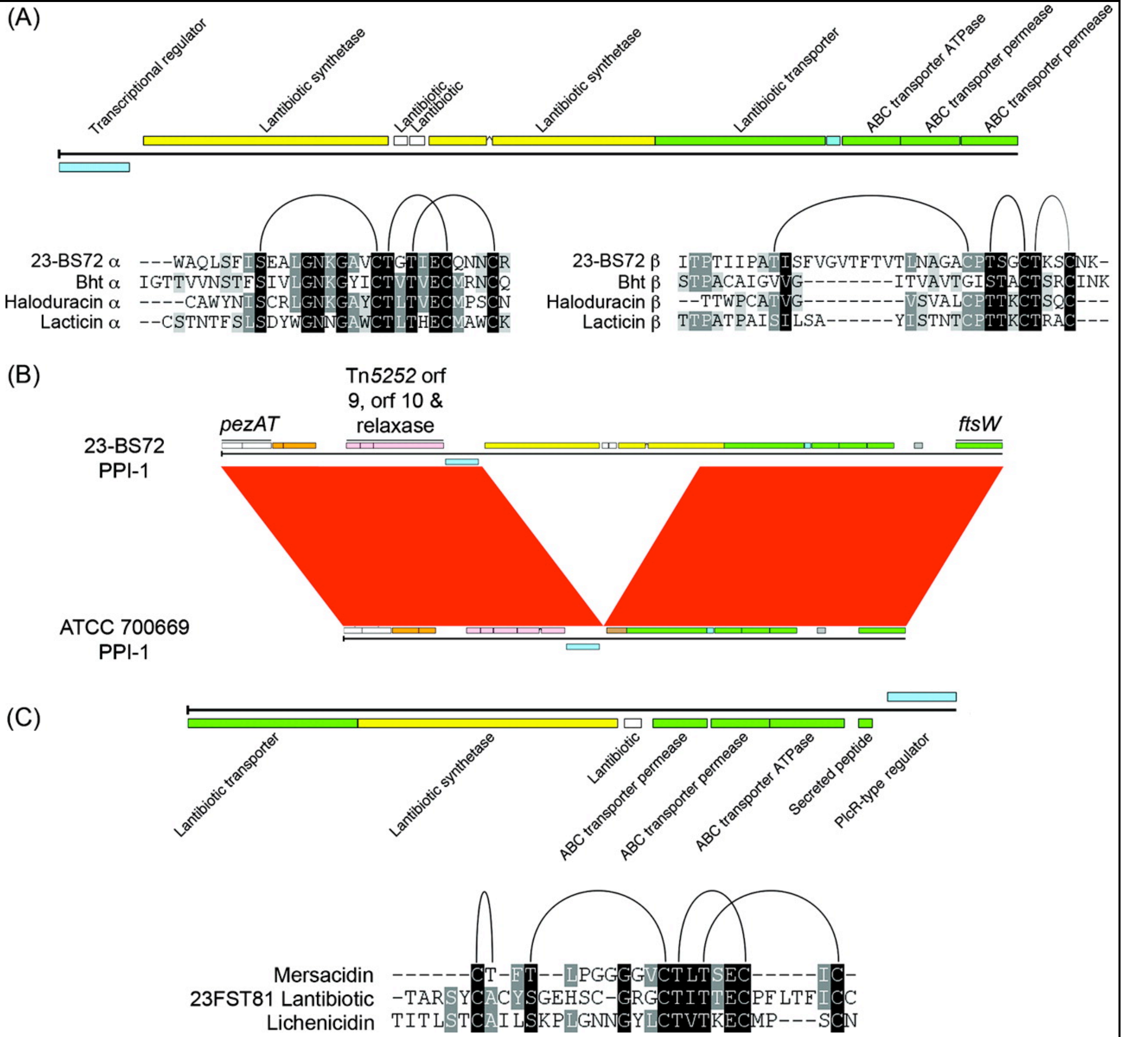
(A)
Transcriptional regulator
Lantibiotic synthetase
Lantibiotic
Lantibiotic
Lantibiotic synthetase
Lantibiotic transporter
ABC transporter ATPase
ABC transporter permease
ABC transporter permease
23-BS72 α ---WAQLSFISEALGNKGAVCTGTIECQNNCR
Bht α IGTTVVNSTFSIVLGNKGYICTVTVECMRNCQ
Haloduracin α ----CAWYNISCRLGNKGAYCTLTVECMPSCN
Lacticin α --CSTNTFSLSDYWGNNGAWCTLTHECMAWCK
23-BS72 β ITPTIIPATISFVGVTFTVTLNAGACPTSGCTKSCNK-
Bht β STPACAIGVVG--------ITVAVTGISTACTSRCINK
Haloduracin β --TTWPCATVG---------VSVALCPTTKCTSQC---
Lacticin β TTPATPAISILSA------YISTNTCPTTKCTRAC---
(B)
Tn*5252* orf 9, orf 10 & relaxase
pezAT
ftsW
23-BS72 PPI-1
ATCC 700669 PPI-1
(C)
Lantibiotic transporter
Lantibiotic synthetase
Lantibiotic
ABC transporter permease
ABC transporter permease
ABC transporter ATPase
Secreted peptide
PlcR-type regulator
Mersacidin -------CT-FT--LPGGGGVCTLTSEC------IC-
23FST81 Lantibiotic -TARSYCACYSGEHSC-GRGCTITTECPFLTFICC
Lichenicidin TITLSTCAILSKPLGNNGYLCTVTKECMP---SCN

Figure 3.6 Lantibiotic synthesis gene clusters and predicted structures. The colour scheme is the same as that used in Figure 3.2. (A) Gene cluster encoded on the PPI-1 of 23-BS72. Each structural gene appears to be associated with its own synthetase. The ABC transporter is thought to be responsible for the self-immunity of the producer. The likely sequences of the mature peptides are shown aligned with other dimeric lantibiotics, with the lanthionine rings (formed by the dehydration of cysteine and serine or threonine residues) conserved between haloduracin and lacticin 3147 indicated by the curved lines above the alignment. The conservation of the functionally most important residues with these other bacteriocins indicates the 23-BS72 lantibiotic is likely to be bactericidal. (B) Alignment of the PPI-1 of *S. pneumoniae* 23-BS72 and ATCC 700669, revealing a ~6-kb deletion in the gene cluster of the latter. (C) Gene cluster encoded on ICE*Sp*23FST81. This is predicted to produce a monomeric lantibiotic, most similar to lichenicidin and mersacidin, produced by Bacillus species. The known ring structure of mersacidin is indicated, again showing that the functional, ring-forming residues are largely conserved in this new lantibiotic. It is the third ring (the most strongly conserved between the three sequences) that is thought to be most important for the bactericidal effects of mersacidin, resulting from the inhibition of peptidoglycan transglycosylation.

A similar *plcR* homologue-secreted peptide combination is present on the G54 ICE, and also adjacent to a different ~10-kb putative lantibiotic synthesis gene cluster

(SPN23F19690-19790) conserved in the chromosomes of ATCC 700669, TIGR4 and D39, next to the *rpoBC* genes. This locus also appears to be a recent addition to the genome on the basis of its atypical nucleotide composition (Vernikos and Parkhill, 2006) and absence from other pneumococcal sequences (*e.g.*, *S. pneumoniae* OXC141, EMBL accession code FQ312027 and *S. pneumoniae* INV104, EMBL accession code FQ312030). In addition, shortly upstream of the ICE*Sp*23FST81 insertion site lies another region (SPN23F12290-12330) in these three genomes that appears to have been recently horizontally acquired, on the basis of its nucleotide composition and flanking tandem repeats, which contains the remains of a set of lantibiotic processing machinery. Hence, the range of bacteriocins that individual pneumococcal strains are able to produce appears to change over time as new gene clusters are gained and old ones degenerate. If these antimicrobial peptides are associated with specific extracellular signalling systems, this could further increase the scope for competition between strains, since differences in signaling and regulation of the bacteriocins would also vary between different genotypes. This is likely to result in strains having a variable secretome that could strongly influence intra- and interspecific competition within the nasopharynx.

3.3 Discussion

Even in a species that exchanges genetic material as readily as *S. pneumoniae*, a small number of clones dominate the population of antibiotic-resistant pneumococci (Klugman, 2002), suggesting either that multiple resistances rarely successfully accumulate in a single strain, or that other factors in the chromosome are responsible for the apparently high fitness of particular ‘pandemic clones’, such as PMEN1. Two events were clearly important in the evolution of the lineage: recombinations affecting the *pbp* genes resulting in penicillin resistance, and the acquisition of ICE*Sp*23FST181, which carries the chloramphenicol and tetracycline resistances. As penicillin resistance has arisen on multiple occasions independently in *S. pneumoniae*, and ICE remnants seem to be quite widespread in the species, indicating such elements are common, it would appear likely that multiple resistances could accumulate in a single genome quite frequently. However, while ICE-carried resistance determinants are widely conserved, it should not be assumed that all penicillin-resistant *pbp* alleles are equally fit, and hence the prevalence of the lineage may simply reflect a minimal cost of β lactam insensitivity.

Alternatively, if such arrangements are common and the associated fitness costs are not too variable, then the rest of the genotype would be likely to play a role in determining the success of the clone. In the case of PMEN1, the variation in the loci that encode bacteriocins, likely to be important in mediating competition within the nasopharynx, suggests one way in which the rest of the genome may contribute to differences in fitness between strains. The loss of the *blpMN* pneumocidin and PPI-1-encoded lantibiotic in ATCC 700669 suggests they may have become redundant during the evolution of the PMEN1 lineage. In both cases, the structural genes were lost, while those required for immunity were retained, ensuring the strain remained nonsusceptible to these compounds secreted by its competitors. A possible explanation is that the pneumocidin on the ICE may have superceded other bacteriocins produced by the strain and assisted nasopharyngeal colonization, and hence the spread, of the Spain^{23F} ST81 clone.

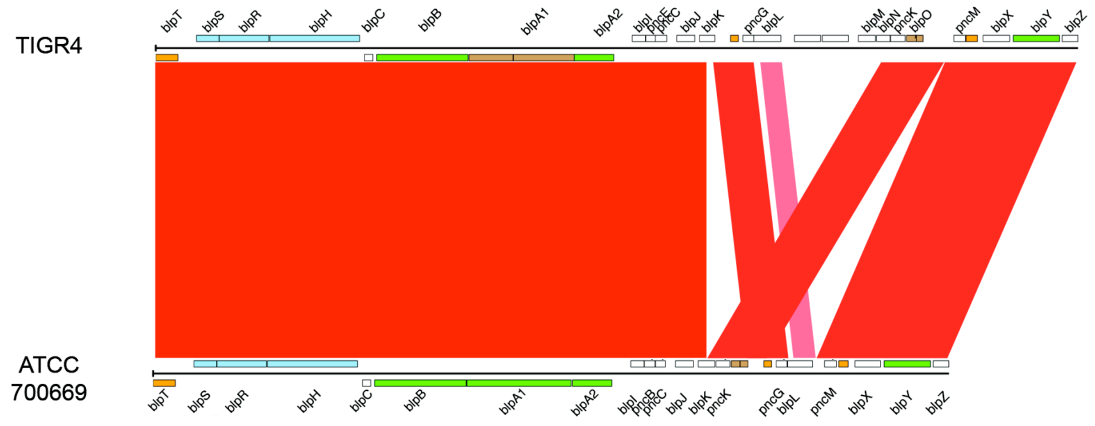

Figure 3.7 Comparison between the *blp* loci of *S. pneumoniae* TIGR4 and ATCC 700669. The *blpMN* genes, which encode a bacteriocin and have been demonstrated to be important in mediating intraspecific competition in a mouse model of pneumococcal colonisation, have been deleted from the locus in ATCC 700669.

Both this lantibiotic synthesis gene cluster and another elsewhere in the chromosome are found adjacent to genes encoding cell-density-dependent *plcR*-type transcriptional regulators and secreted peptides. Such regulation of the ICE-encoded lantibiotic synthesis gene cluster may be advantageous for the mobile element itself, since suppressing production of the antimicrobial compound until the ICE has saturated the available population of potential hosts is likely to facilitate its horizontal transfer. Furthermore, if such quorum-sensing systems are involved in the regulation of bacteriocin production, this would add further layers of complexity onto the intercellular signalling already known to occur in *S. pneumoniae* (the two previously characterized pneumocidins, BlpMN and CibAB, are regulated by the pheromones BlpC and ComC, respectively) (de Saizieu *et al*., 2000; Claverys *et al*., 2007). Hence, just as multilocus epidemiological typing schemes are required to robustly identify lineages from an anthropic perspective, for bacteriocin-mediated competition between pneumococci to be effective several pheromone-controlled bacteriocin systems at different loci are likely to be required, as a system based on single locus with multiple alleles would be too easily confounded by a single recombination event.