The Type I Restriction Modification System *Spn*III of *Streptococcus pneumoniae*

Thesis submitted for the degree of

Doctor of Philosophy

at the University of Leicester

by

Megan De Ste Croix BSc (Hons)

Department of Genetics and Genome Biology

University of Leicester

August 2017

The Type I Restriction Modification System *Spn*III of *Streptococcus pneumoniae*

Megan De Ste Croix

The phase variable expression of alternative specificity subunits by the type I restriction modification system *Spn*III has been proven to differentially regulate important virulence factors. These gene expression differences are envisaged to translate into significant differences in the virulence of strains, dominated by different specificity subunits, in murine models of infection. The presence of this type I system in the core pneumococcal genome hints at a conserved, global mechanism of regulation, including that of virulence.

The use of wildtype strains enriched in alternative specificity subunits has revealed that recombination within the *spnIII* locus is not reciprocal and is highly variable between strains. Despite this variability, methylation of the genome by different SpnIII variants results in differential gene expression and significant differences in models of invasive disease. The recombination within the *spnIII* locus appears to be much more complex than similar, previously reported systems. Recombination is not mediated by the classical homologous recombination pathways but is partially controlled by a site specific tyrosine recombinase.

Investigations into the impact of *Spn*III may be capable of improving our understanding of how this nasopharyngeal coloniser is capable of causing devastating invasive disease.

Acknowledgements

It would not have been possible to complete this thesis without some help, so thanks goes to…

Marco, for making me think

Ana & Leo, for being my teachers

Richard, for knowing where just about everything can be found

Dr Chris Bayliss, for offering a different perspective

The now Dr Jack Aidley, for the coding

Dr Vitor Fernandes, for the mice

Dr Julie Morrissey & Prof. Peter Andrew, for, as my thesis committee, helping me produce better work

The Department of Genetics, for paying me

11.1X Buffer (and Dr Rita Neumann for the recipe), for making most of the work in this thesis possible

Dr Calum Johnston, for so many mutants

And finally

My parents, for asking if I wear a white lab coat, and being proud that I do.

Contents

Figure List

Table List

List of abbreviations

Abbreviation	Meaning	Page
BHI	Brain Heart Infusion Broth	29
BHIA	Brain Heart Infusion Agar	29
CAP	Community Acquired Pneumonia	1
CFU	Colony Forming Units	31
CSP	Competence Stimulating Peptide	4
dsDNA	double stranded DNA	4
EHPC project	Experimental Human Pneumococcal Carriage project	41
Hsd	Host Specificity Determinant	6
INDEL	Insertion/Deletion	34
IPD	Invasive Pneumococcal Disease	1
IR	Inverted repeat	17
KO	Knockout	44
m4C	4-methylcytosine	5
m5C	5-methylcytosine	5
m6A	6-methyladenosine	5
MTase	Methyltransferase	5
OUT	Operational Taxonomic Unit	3
OP	Opaque Colony Phenotype	2
PCR	Polymerase Chain Reaction	34
PCV	Polysaccharide Conjugate Vaccine	1
REase	Restriction endonuclease	7
RM	Restriction Modification	3
RPKM	Reads Per Kilobase Million	55
SAM	S-adenosyl-L-methionine	5
SMRT sequencing	Single Molecule Real Time Sequencing	8
SNP	Single nucleotide polymorphism	34
ssDNA	single stranded DNA	4
SSR	Site specific recombinase	17
ST	sequence types	16
TP	Transparent Colony Phenotype	2
TRD	Target Recognition Domain	8
TSA	Trypticase Soy Agar	29
TSB	Trypticase Soy Broth	29
U/ml	Units per ml	75
WGS	Whole genome sequencing	16
WHO	World Health Organisation	1
WT	Wildtype	39

1 Introduction

Part of this introduction has been adapted for publication in the Journal FEMS Microbiology Reviews, “Phase variable methylation and epigenetic control by Type I Restriction Modification systems” M. De Ste Croix, I. Vacca, M. Kwun, J. Ralph, S. Bentley, R. Haigh, N. Croucher, M.R. Oggioni. (Manuscript submitted).

1.1 *Streptococcus pneumoniae*

Many *Streptococcal* species including *Streptococcus mitits, oralis, pneumoniae* and *pseudopneumoniae* are exclusively adapted to the colonisation of humans and other hominids (1). While many of these closely related species, such as *S. mitis,* are classed as asymptomatic colonisers and often reside in a human host for life (1), *S. pneumoniae* (the pneumococcus) is a globally important human pathogen. It is one of the leading causes of invasive infections worldwide, including pneumonia, septicaemia and meningitis, as well as non-invasive infections such as otitis media and sinusitis (1–4). In 2005 it was estimated by the World Health Organisation (WHO) that *S. pneumoniae* was responsible for more than 1.6 million deaths, of which 0.7-1 million were children under five years of age (5–7). In the UK *S. pneumoniae* accounts for 15-43% of all cases of community acquired pneumonia (CAP) and is a significant cause of invasive pneumococcal disease (IPD) which includes septicaemia and meningitis (8,9).

The virulence of the pneumococcus is closely associated with its polysaccharide capsule (10), of which more than 90 have so far been identified (4,11–13). Each antigenically distinct capsule has a unique order and make up of monosaccharides and side branches (11). The capsule is the main antigenic component of the pneumococcus and generates a strong immune response. As a result, episodes of carriage offer protection against future infections from that, and other closely related, serotypes (9). In 2006 the polysaccharide conjugate vaccine (PCV), targeting the seven most prevalent pneumococcal serotypes, was introduced to the UK’s routine childhood vaccine programme (8). Prior to its introduction, 82% of all IPD cases were caused by the seven serotypes targeted by the vaccine (9). It was subsequently replaced by the PCV13 vaccine in 2010, and cases of IPD have reduced by 56% overall, with a 97% reduction in cases caused by PCV7 strains (8). Due to the specificity of the immune response to

each individual serotype PCV vaccines have led to a reduction in carriage and IPD cases by vaccine serotypes but serotype replacement, where non-vaccine serotypes increase in prevalence, is frequently observed (3,8,14).

The pneumococcal capsule offers the bacterium protection during infection but is metabolically demanding to produce, therefore its production is highly regulated (15,16). Pneumococcal cells are able to switch between two morphological forms, termed opaque (OP) and transparent (TP) (17). Cells displaying the OP phenotype express significantly more capsule and are significantly more virulent than those expressing the TP phenotype (16). The switch between opaque and transparent is phase variable (16), however the exact mechanism controlling the switch from one to the other remains unidentified.

In addition to the capsule there are also a number of surface proteins that are known to influence the host immune system (18). These include; Pneumococcal surface protein A (PspA), Pneumococcal surface protein C (PspC) and the pore-forming toxin pneumolysin (Ply) (19–21). PspA mutants are reported to be less virulent in models of otitis media infection (22) and show a decreased ability to adhere to cells (21). This is the result of PspA's role in preventing the deposition of complement on the cells surface (20). PspC, like PspA, interferes with complement activation, in this case by binding factor H and preventing opsonophagocytosis (19,20). Ply is a cytolytic toxin that forms pores in membranes (21). In experimental models of infection strains not expressing pneumolysin are less virulent and are found in smaller numbers in the nasopharynx (21). Ply is thought to have two main functions, its ability to lyse cells and its ability to interfere with the complement pathway (21).

Despite the high level of morbidity and mortality associated with *S. pneumoniae* (5,13) it frequently colonises the human nasopharynx asymptomatically where it undergoes little genetic variation (23). The pneumococcus predominantly colonises under-fives, as frequent exposure over time allows adults to create sufficient immune memory to limit colonisation (24,25). It is possible this immune memory is protein and not capsule based. As there is a requirement for carriage before disease (25), the burden of pneumococcal infections falls on those who have yet to develop sufficient immunological memory. Investigations into

pneumococcal virulence have been ongoing since the 1930's (26,27) however the mechanisms driving the transition from asymptomatic coloniser to invasive pathogen remain unclear (16). Antibiotic resistance and serotype replacement as a result of polysaccharide based vaccines, mean understanding the basic biology of the pneumococcus is increasingly important.

1.1.1 *Streptococcus pneumoniae* as a lineage of *Streptococcus mitis*

Despite its pathogenic nature, genome sequencing has revealed that *S. pneumoniae* is a phylogenetic lineage of the human commensal *Streptococcus mitis* (1). Based on operational taxonomic unit (OTU) classification using 16S rRNA sequences is has been demonstrated that *S. pneumoniae* is >99% identical to commensals *S. mitis* and *S. oralis* (28). Unlike *S. pneumoniae* which passes between hosts using short episodes of carriage (median 19 days) (29), *S. mitis* forms lifelong association with its host (1). The two species diverged from a common ancestor and have utilised alternative genomic strategies to ensure the maintenance of a host relationship. While some *S. mitis* strains do contain many of the genes associated with pneumococcal virulence, such as autolysin and pneumolysin, they are very rarely associated with disease (1). In addition, it has recently been shown that 74% of *S. mitis* strains (in a panel of 66) also contain a complete *cps* (capsule) locus (30). With the capsule so closely associated with virulence and the immune response this was quite an unexpected discovery (30).

Typically, *S mitis* strains have a genome which is approximately 15% smaller, and considerably more stable than *S. pneumoniae* (1). This is likely to be the result of several major distinctions between the two species. Firstly, the majority of *S. mitis* strains lack the Dpn Restriction-Modification (RM) system that is present in all strains of *S. pneumoniae*, with many acquiring CRISPR sequences for phage protection instead. Dpn is consider part of the competence regulon and aids in the acquisition of exogenous DNA (31). In addition to lacking the Dpn system many *S. mitis* (and *S. oralis)* strains do not have a complete set of genes for the competence pathway, and they do not contain the Type I Restriction-Modification systems *Spn*III and *Spn*IV. Overall, the movement of genes is limited by the lack of a complete competence system in the commensal species, and as

a result genetic transfer is fairly unidirectional from *S. mitis/oralis* strains to *S. pneumoniae* (1).

1.1.2 Natural competence for genetic transformation

The huge genetic diversity in *S. pneumoniae* is largely attributed to its ability to uptake exogenous DNA and incorporate it into its genome via homologous recombination (32–35). This process is known as transformation and was first reported by Griffith in 1928 when he demonstrated that a type I strain could be transformed into a virulent type II strain (36). When mice were subcutaneously infected with a heat killed type II strain and a live type I strain a type II strain could be recovered (36).

In the pneumococcus transformation occurs when cells become competent (32–34). Pneumococcal cultures become competent for a short time, in a synchronised manner and the process can be induced by environmental stressors such as antibiotics (32,37). This occurs through a positive feedback loop initiated when the competence stimulating peptide (CSP), encoded by the gene *comC,* binds its membrane bound receptor ComD, causing it to autophosphorylate (32,34). Autophosphorylated ComD transphosphorylates ComE which is then able to induce expression of *comCDE* and *comX* (32,34,38). Expression of the *comCDE* operon leads to the production and export of more CSP, creating the positive feedback loop observed in pneumococcal cultures (32).

ComX is the competence specific sigma factor which induces the expression of approximately 100 genes, known as the competence regulon (38). The competence regulon includes all proteins known to be associated with natural transformation, with the exception of the nuclease EndA (32). The universal recombinase RecA is essential for transformation (32,34,35). It is now known that RecA is involved in three independent pathways of homologous recombination facilitated by three different DNA loading complexes (39). The RecFOR complex loads RecA onto single stranded DNA (ssDNA) gaps when replication forks stall, while RecBCD (also known as RexAB) loads RecA onto regions of double stranded (dsDNA) breaks. It has been experimentally determined that neither the

RecFOR or RecBCD complexes are essential for transformation (35). Instead the competence specific gene *dprA* has been identified as the dedicated loader of RecA during transformation (32,39).

Transformation begins when exogenous dsDNA binds the transformation pilus ComGC, which subsequently transfers it to the DNA receptor ComEA. Once bound to ComEA one strand is degraded by the EndA nuclease permitting internalisation by the receptor. Once internalised ssDNA is coated by the competence specific single stranded binding protein SsbB. SsbB protects the DNA from degradation by nucleases until it is displaced by DprA (35). Once bound, DprA recruits RecA initiating the search for a homologous region and the subsequent incorporation of the DNA into the genome (32).

1.2 Restriction Modification systems

It is now over 50 years since the enzymatic modification and restriction of bacteriophage and bacterial chromosomal DNA was first described (40). Since then, many families of DNA modification and restriction-modification enzymes have been described, and a wide variety of functions beyond simple defence have been recognised (41). DNA methylation has increasing been shown to be a common feature of prokaryotic genomes, present in more than 90% of species studied (42). It is a chemical modification of either an adenine or cysteine altering the base to 6-methyladenosine (m6A), 4-methylcytosine (m4C) or 5-methylcytosine (m5C). Methylation is commonly used to recognise self DNA and distinguish it from invading foreign DNA. In addition, these modifications alter DNA structure and as a result influence DNA expression. It is now possible to detect these modifications using the Single Molecule Real Time (SMRT) sequencing system developed by Pacific Biosciences (43), with m6A modifications accounting for 75% all of prokaryote methylation (42).

1.2.1 Methyltransferases

Methylation of DNA is a process carried out by a specific class of enzymes; methyltransferases (MTases). Generally these enzymes are responsible for the transfer of a methyl group from a donor molecule of S-adenosyl-L-methionine

(SAM) onto the base to be modified (44). MTases typically exist within restriction modification (RM) systems, which also cleave DNA, however orphan MTases are not uncommon. The addition of a methyl group to a specific DNA sequence allows the DNA to be recognised as self, thereby protecting it from restriction. DNA entering a cell is unlikely to be methylated in the correct pattern, therefore the cell is capable of recognising it as foreign and as a result cleaves it. However, recognition of self DNA is only one purpose of DNA methylation. The addition of methyl groups does more than indicate self DNA, it is also known to regulate gene expression. When a methyl group is added to a base the structure and dynamics of the DNA molecule are altered resulting in changes in DNA-protein interactions (45).

1.2.2 **Type I RM systems**

In terms of genomic structure type I RM systems are the most complex, typically consisting of three *host specificity determinant* (*hsd*) genes, *hsdR, hsdM* and *hsdS* (Fig. 1.1A) (42,46–48). The *hsdR* gene encodes the restriction (R) subunit which is incapable of DNA cleavage unless in a complex with the methylation (M) and specificity (S) subunits, encoded by *hsdM* and *hsdS* respectively (Fig. 1.1B). However, the M and S subunits are able to form an MTase independent of the R subunits, producing an enzyme capable of methylation but not restriction (Fig 1.1B) (42,48). In addition to its role as a DNA cutter the R subunit is also an ATP powered DNA helicase (46,47). Cleavage by the R subunit occurs randomly some distance (up to several kb) from the recognition site, potentially when two enzyme complexes collide or reach a DNA obstruction (42,47,49).

The S subunit is essential for the recognition of a specific bi-partite sequence which is then methylated on both strands (48). The *hsdS* gene encodes two target recognition domains (TRDs), which each recognize one half of the bi-partite recognition site (47). Until recent advances in Single Molecule Real Time (SMRT) sequencing it was thought that all Type I systems methylate adenines, generating 6-methyladenosine (m6A) (48). However, the recent SMRT sequencing of 230 bacterial and archaeal species by Blow et al. revealed Type I RM systems capable of m4C methylation in *Desulfobacca acetoxidans* and *Methanohalophilus mahii* (42). In addition, Morgan et al. identified a similar

system in *Pseudomonas alcaligenes* (50). These m4C MTases are found in systems with two MTases, with the second methylating m6A. One half of the bipartite sequence of both these species contains only G's and C's, therefore m6A methylation is not possible, and the m4C methylation is used to protect this half of the sequence.

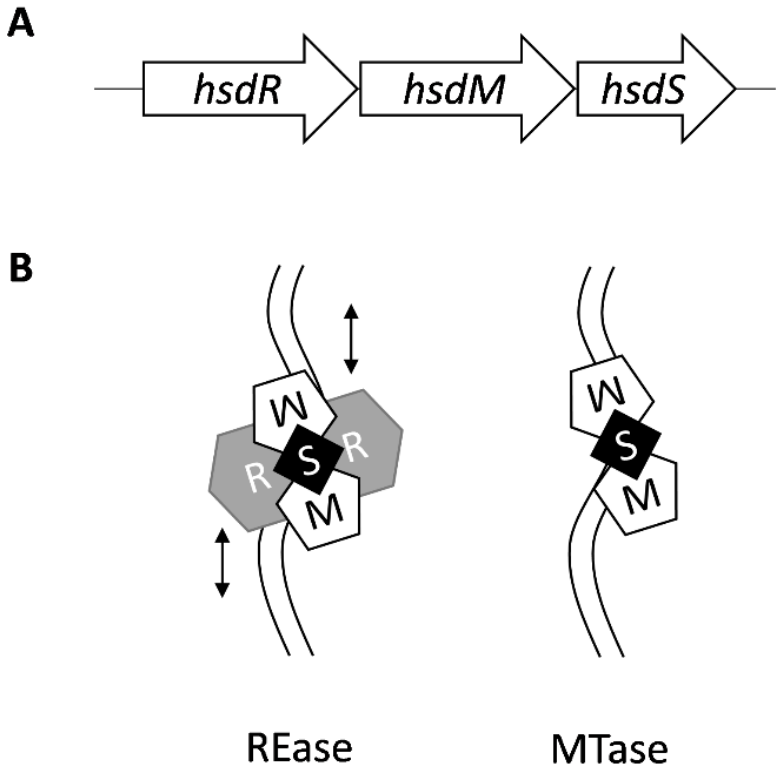


Figure 1.1 – Type I Restriction Modification systems. *Type I Restriction Modification (RM) systems are typically encoded by three separate genes, hsdR, hsdM and hsdS (Panel A). Each of these genes encodes a subunit with a particular function: restriction (R), methylation (M) and sequence recognition (S). The subunits can assemble as either a pentameric REase (2R+2M+S) which is capable of both DNA methylation and cleavage. Alternatively, a trimeric MTase (2M+S) can assemble that is capable of DNA methylation but not cleavage* (46,47,49).

Table 1-1 – General features of Restriction Modification systems

	Type I	Type II(P)	Type III	Type IV
Enzyme structure	2R+2M+1S (REase) 2M+1S (MTase)	2R (REase) 1M (MTase)	1-2R+2M	Variable
Example	EcoKI	EcoRI	EcoP15I	No typical example but enzymes include Mcr and Mrr
Example Recognition sequence (modified base)	A<u>A</u>C-N_6-G<u>T</u>GC	G<u>A</u>ATTC	CAGC<u>A</u>G	Variable McrA recognises (Y>R)CGR
Cleavage	Cleavage by the REase often occurs a significant distance from the recognition site	Cleavage occurs within or very close to the recognition site	Cleavage occurs at a defined site close to (25-27bp) but not within the recognition site.	Variable
Features	S Subunit contains two separate target recognition domains (TRDs) Recognition sequence is bipartite with each TRD recognising one half REase has helicase activity Requires ATP	REase and MTase function as separate enzymes Type II enzymes are split into multiple different categories with EcoRI falling into the most simplistic Type IIP group which recognises a palindromic sequence.	Methylation is single stranded Recognition sites are asymmetric Requires ATP	No single feature unites this class of enzymes Sequence specificity is weak No known MTase activity
Reference(s)	(47,48,51,52)	(47,49)	(47,53,54)	(42,47,49)

The m6A MTase is then used to methylate the half of the sequence that does contain an Adenine (42). Similar systems can be found in REBASE, suggesting this may be a more widespread occurrence, but for the majority the recognition sequence has not yet been identified (42,52).

1.2.3 **Type II RM systems**

Type II RM systems are composed of two genes *res* encoding the REase and *mod* encoding the MTase (48). Of the different classes of restriction enzyme these are the most commonly used in the laboratory due to their specific cleavage at or very close to their recognition site (47). While the group of Type II enzyme represented in table 1.1, Type IIP, is relatively simple there are ten additional sub-classes of Type II enzymes that have unique defining features such as the recognition of an asymmetrical site (Type IIA) or alternative enzyme structure (Type IIH) (47). As Type II enzymes are classified based upon their features and not their relatedness, some enzymes can fall into multiple classes. For example, DpnI of *S. pneumoniae* falls into both the Type IIM and Type IIP categories as it cleaves methylated DNA (Type IIM) within a palindromic recognition sequence (Type IIP) (49,55)

1.2.4 **Type III RM systems**

Like Type II enzymes, Type III RM systems are typically encoded by two genes, *mod* (MTase) and *res* (REase) (48). While Type III enzymes are generally more complex than Type II's there are some other similarities including a defined cleavage site 25-27bp from the recognition site (47,54). As with Type I systems, in order for restriction to occur the R subunits must form a complex with the MTase as this contains the DNA specificity domain required for sequence recognition (42,54). Like Type I enzymes, Type III enzymes have helicase activity and require ATP in order to translocate along DNA molecules (46,48,53,54). It has been shown that Type III enzymes exhibit a preference for recognition sites to be orientated head-to-head for efficient DNA cleavage (47). Cleavage occurs 25-27bp from one of the two sites, although which site appears to be random (47).

1.2.5 Type IV RM systems

Type IV RM systems are a small collection of diverse enzymes that restrict DNA but lack MTase activity (42,56). They are most similar to Type IIM enzymes which recognise and cleave modified DNA, although for now the two remain as separate classifications of enzyme (49). Type IV enzymes recognise a wide variety of DNA modifications including m4C, m5C and m6A methylation as well as the glucosylated 5-hydroxymethylcytosine (glc-HMC) with low sequence specificity (49,56). As they lack any methylation activity Type IV systems are primarily associated with defence against phage and other forms of foreign DNA.

1.3 Model Systems of Prokaryotic DNA Methylation

1.3.1 DAM Methylation

One of the most well studied of all MTases is the DAM system of the γ-proteobacteria (45,57). This orphan Type II MTase is responsible for the methylation of the four base sequence GATC (45,57). Methylation of GATC serves several functions. Firstly, the addition of the methyl group to the *oriC* promotes the binding of the replication initiation complex. In addition, the hemi-methylation that arises during DNA replication allows the recognition of a parent strand from a daughter strand, meaning replication errors can be identified and corrected by the cells mismatch repair machinery (45).

1.3.2 Cell Cycle Regulation in *Caulobacter crescentus*

An example of the tight regulation of the cell cycle by DNA methylation can be seen in *Caulobacter crescentus.* This bacterium has two cell types, stalked and swarmers, and replication is only initiated in stalked cells. Expression of *dnaA* and therefore the initiation of DNA replication occurs when the C*ori* of *C. crescentus* is fully methylated by the MTase CcrM, at the adenine of its recognition sequence GANTC (45). The replication of the genome results in hemi-methylated DNA, preventing further expression of *dnaA,* and activating a cascade of genes that are only expressed when GANTC is hemi-methylated. One of these genes is *ctrA* which binds *dnaA*, as well as activating expression of CcrM

and FtsZ (45). Once CcrM is expressed, it is able to methylate the newly synthesised DNA strands, resulting in complete methylation of the C*ori* allowing a new round of DNA replication only when the one before has been completed. In swarmer cells CtrA remains bound to C*ori* preventing the initiation of DNA replication in this phase of the life cycle (45).

1.3.3 **The Dpn System of *Streptococcus pneumoniae***

Importantly, not all restriction modification systems act as barriers to all forms of incoming DNA. An example of a system devised to rescue incoming DNA is DpnII, found to avoid such post-transformation cell suicide. All *S. pneumoniae* strains contain one of three RM systems, DpnI, DpnII or DpnIII, each of them found at the same position in the genome and, like Dam, all recognising the sequence GATC, but with variation in site methylation (55,58). DpnI, unlike the majority of RM systems, actively targets DNA methylated at G^{m6}ATC, while DpnII restricts unmethylated DNA at the same sequence. These features of this enzyme pair have been used extensively to experimentally verify Dam methylation of target sites. The majority of incoming DNA is not methylated and is therefore not targeted by DpnI. Once the DNA has been incorporated into the genome, it becomes hemi-methylated and is therefore protected from restriction. However, if a replication fork passes over newly incorporated DNA before it has been fully methylated, it could lead to a newly synthesised DNA strand that is completely unmethylated, resulting in restriction of the new DNA (58). This is more likely to occur in regions with a large number of sites to be methylated. To avoid it, the DpnII system, not only encodes a double stranded DNA (dsDNA) MTase, but also a rare ssDNA MTase, DpnA. DpnA is only expressed during competence (58) and ensures that restriction of the newly synthesised ssDNA does not occur. The third Dpn system, DpnIII, is found in the pneumococcal lineage PMEN1. The DpnIII system recognises and methylates the cytosine rather than the adenine of the GATC recognition sequence (31). DpnIII will therefore restrict DNA from strains with either the DpnI or DpnII system which results in a fairly unidirectional movement of genes, with transfer being predominantly from strains with DpnIII to strains with DpnI or DpnII (31).

1.3.4 **Phase Variable DNA methylation**

As previously mentioned the DAM MTase is responsible for the methylation of GATC and is capable of controlling several cellular functions including the initiation of DNA replication and identification of newly synthesised daughter strands via hemi-methylation (45). However, there are also phase variable occurrences of DAM regulation. DAM is known to be responsible for the ON/OFF switching of the *pap* (pyelonephritis-associated pili) operon of uropathogenic *E. coli* strains (59). The locus encodes three genes; *papA, papB* and *papI* and changes in DAM methylation impact transcription of the operon (60). DAM competes with the leucine responsive regulatory protein, Lrp, to bind and methylate the two GATC sites found within the Lrp binding regions in the promoter of the *pap* locus (59,60). There are two Lrp binding regions, known as sites 1-3 and 4-6. If Lrp binds to site 2 it blocks DAM methylation and prevents transcription of *pap* as the RNA polymerase binding site is blocked by Lrp. Alternatively, if DAM is blocked by Lrp from binding site 5 this promotes *pap* transcription. It is thought this occurs through structural changes that result in the RNA polymerase binding site becoming more accessible (59). To switch from an OFF state to an ON state, PapI and Lrp form a complex which has a high affinity for hemi-methylated DNA in sites 4-6, therefore promoting the recruitment of Lrp to site 5 and preventing methylation by DAM (59). The *pap* operon demonstrates a complex, reversible and inheritable method of gene expression.

A phase variable Type IIG RM system has been described in *Campylobacter jejuni* (61). In this system, *cj0031,* an endonuclease and methyltransferase are encoded by a single gene. There is a polyG repeat tract within the gene that, via slipped-strand mispairing, can alter the reading frame and therefore control gene expression. *Cj0031* has been shown to methylate the adenine of both 5’CCCGA and 5’CCTGA resulting in differential gene expression. However, the gene is not universally phase variable as homologues in several other *Campylobacters* were found to lack the polyG tract. *C. jejuni* cells not expressing *cj0031* were found to be less efficient in their adherence to and invasion of Caco-2 cells as well as forming significantly less biofilm, suggesting a generalised mode of action for this RM system (61).

Examples of phase variable RM systems with polyG repeat tracts can also be found in *Helicobacter pylori* (62). The phase variable *modH* gene of a Type III RM system has 17 *mod* alleles, each conferring a different specificity for methylation. Microarray analysis has confirmed that the loss of the system leads to changes in *in vitro* gene expression. The regulation of multiple genes by a phase variable RM system has been termed a “phasevarion” (53). While only a small number of genes are affected (six with a >1.6fold change when compared to a strain with an intact RM system) they include the surface exposed *hopG* and several flagella genes (*flaA* and *fliK*) that are required for motility. These data only represent changes in gene expression from a single *modH* allele, there are potentially 17 different phasevarions in *H. pylori* each altering the expression of a small subset of different genes (62). Phase variation of Type III RM systems by repeat tract changes is not limited to *H. pylori,* similar systems have been identified in *M. haemolytica, H. influenzae, N. meningitidis* and *N. gonorrhoeae* (53).

1.4 **Phase Variable Type I Restriction-Modification Systems**

The generation of new RM systems is possible through the mutation of the gene encoding the specificity region of the enzyme. In Type I systems a single mutation in an S gene can lead to the recognition of a new sequence for DNA methylation (63). In recent years, it has been identified that the flexibility of S subunits is not limited by their ability to permanently mutate their sequence. Instead the presence of multiple *hsdS* genes allows phase variation through changes in genome position (64–67). By moving complete and partial regions of *hsdS* genes in to and out of positions with promoters it is possible to reversibly express different S subunits. This reversible switching results in the continuous generation of diversity within a population.

Phase variable type I restriction modification sequences have been identified in a variety of species. These include but are not limited to: *Mycoplasma pulmonis* (66,68,69)*, Bacteroides fragilis* (67)*, Streptococcus pneumoniae* (23,64,65,70,71)*, Listeria monocytogenes* (72) and *Lactobacillus salivarus* (73)*.* The genomic arrangements of each of these systems varies (Fig 1.2) and each will be discussed individually.

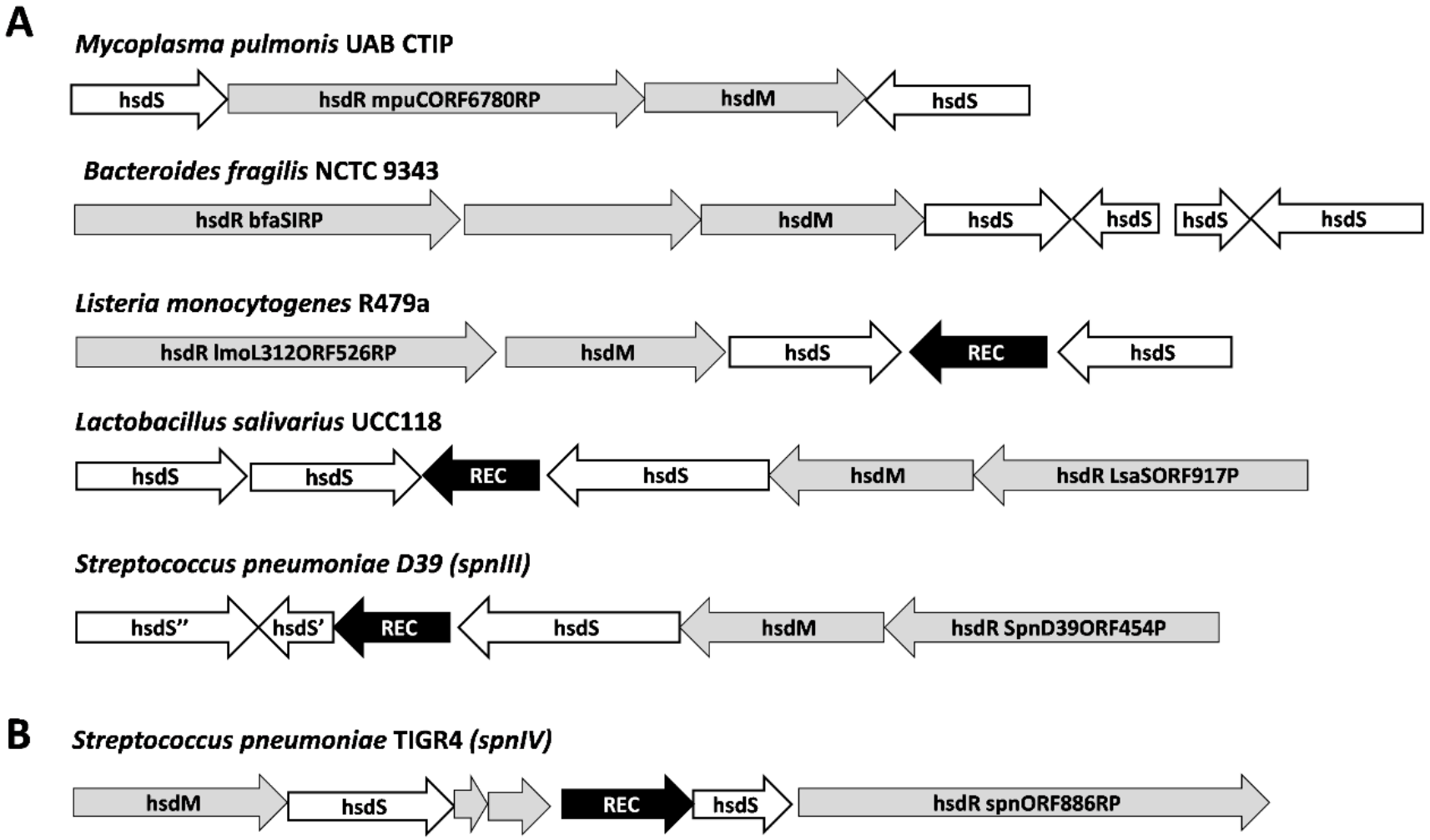

A
Mycoplasma pulmonis UAB CTIP
hsdS
hsdR mpuCORF6780RP
hsdM
hsdS
Bacteroides fragilis NCTC 9343
hsdR bfaSIRP
hsdM
hsdS
hsdS
hsdS
hsdS
Listeria monocytogenes R479a
hsdR lmoL312ORF526RP
hsdM
hsdS
REC
hsdS
Lactobacillus salivarius UCC118
hsdS
hsdS
REC
hsdS
hsdM
hsdR LsaSORF917P
Streptococcus pneumoniae D39 (spnIII)
hsdS''
hsdS'
REC
hsdS
hsdM
hsdR SpnD39ORF454P
B
Streptococcus pneumoniae TIGR4 (spnIV)
hsdM
hsdS
REC
hsdS
hsdR spnORF886RP

Figure 1.2 - Genome arrangements of Type I RM systems with multiple hsdS genes. *The strains have been divided according to the presence of either inverted (A) or direct hsdS repeat (B) sequences. As not all systems are present in all strains of the same species the name of each strain is reported on top of each locus illustration. White arrows indicate hsdS genes, light grey arrows indicate hsdM, hsdR and other genes found within the locus, black arrows indicate site specific recombinase genes. Indicated on each hsdR gene is the associated REBASE designation of each RM system. Listed from the top to the bottom the strains and genes reported in each phase variable locus are: A) Mycoplasma pulmonis UAB CTIP (Genbank accession NC_002771), hsd1 (MYPU_RS02160), hsdR (MYPU_RS02165) hsdM (MYPU_RS02170), hsdS2 (MYPU_RS02175); Bacteroides fragilis NCTC 9343 (GenBank GCA_000025985), hsdR (BF9343_RS08540), hypothetical protein (BF9343_RS08545), hsdM (BF9343_RS08550), hsd1 (BF9343_RS21605), hsd2 (BF9343_RS08560), hsd3 (BF9343_RS08565), hsd4 (BF9343_RS08570); Listeria monocytogenes R479a (Genbank NZ_HG813247), hsdR (LMR479A_0528), hsdM (LMR479A_0529), hsdS1 (LMR479A_0530), integrase (LMR479A_0531), hsdS2 (LMR479A_0532); Lactobacillus salivarius UCC118 (Genbank accession CP000233), hsdS3 (LSL_0915), hsdS2 (LSL_0916), xerC (LSL_0917), hsdS1 (LSL_0917), hsdM (LSL_0919) and hsdR (LSL_0920). Streptococcus pneumoniae D39 (Genbank NC_008533), hsdS'' (SPD_0451), hsdS' (SPD_0450), Cre recombinase (SPD_0452), hsdS (SPD_0453), hsdM (SPD_0454) and hsdR (SPD_0455); B) spnIV which recombines using direct repeat found within its hsdS genes (tvr). The genes reported are for Streptococcus pneumoniae TIGR4 (Genbank NC_003028), hsdM (SP_0886), hsdS1 (SP_0887), hypothetical proteins (SP_0888, SP_0889), integrase (SP_0890), hsd2 (SP_0891), hsdR (SP_0892). Maps not drawn to scale.*

1.4.1 ***Mycoplasma pulmonis***

The earliest identified of these phase variable Type I systems is found in *M. pulmonis,* where there is a 6.8kb invertible locus termed Hsd1 (66) (68). *M. pulmonis* was the first non-enteric bacterium found with a Type I RM system (68). The structure and inverted repeats of the *hsd1* locus (Fig 1.2A) allow the generation of 4 different functional *hsdS* genes through TRD shuffling (68). In addition, there is a second Type I RM locus with high homology termed Hsd2 (66). Hsd1 and Hsd2 appear to be conserved across *M. pulmonis* strains, however a third non-functional type I RM system was identified in the strain UAB CTIP (14). Recombination at the *hsd* loci of *M. pulmonis* has been confirmed by PCR analysis using pairs of primers where one is situated within the inverted region. Inversions result in positive PCR products, and prove that this system is switching its TRDs both *in vitro* (68) and *in vivo* (74).

Using the *Mycoplasma* virus P1 Dybvig *et al.* (75) determined that populations with different active *hsdS* genes also have differing phage susceptibility. By isolating 147 sub-clones of a single laboratory stock, eight individual groups were established. Of these groups one (consisting of 17 subclones) showed no RM activity, suggesting the locus had been inverted so that *hsdR* and *hsdM* were no longer in line with their promoter. This group of cells could be infected with P1 phage propagated in any background (i.e. propagated in a population with the same *hsdS* orientation, propagated in a population with a different *hsdS* orientation, or propagated in a population with no RM activity) (75). Six of the remaining groups were susceptible to phage with the same methylation profile but capable of restricting phage propagated in any other background. The final group showed varying susceptibility depending on the strain used for phage propagation (75). Overall the *hsd* loci of *M. pulmonis* appears to have an essential function beyond its role as a barrier to invading DNA. The authors state that the relatively frequent (17/147 subclones) occurrence of non-restricting cells indicates that the maintenance of the system cannot purely be for phage defence (75).

In vivo work in *M. pulmonis* has also been conducted by Gumulak-Smith *et al.* (74). Rats were intra-nasally challenged with *M. pulmonis* and the bacteria

recovered from the lungs, trachea and nose were analysed for changes in *hsdS* (and *vsa*) orientation. PCR analysis showed that, overall, the *M. pulmonis* populations recovered from the trachea were more variable at both their *hsd* and *vsa* loci when compared to those isolated from the nose. In addition to *hsdS* recombination within the recovered populations, cells were recovered that showed no detectable restriction or methylation activity (74). This finding implies that, in addition to the advantage of being able to express multiple different S subunits, no expression of the RM system may also be advantageous in some environments.

Recombination of the *M. pulmonis hsd* loci is closely linked to the rearrangement of the phase variable *vsa* loci (74,76). This is due to the control of both loci by a single site specific recombinase, HsvR (77). It is not known if the two linkage of the two systems is the result of changes in methylation leading to changes in *vsa* expression, or if the link is more direct with changes in surface antigens requiring a change in S subunit expression.

1.4.2 ***Bacteroides fragilis***

The invertible BF1839 Type I RM systems identified in the non-enterotoxin-producing DNA homology group I *B. fragilis*, strain NCTC 9343 (67) is structurally very similar to the locus of *M. pulmonis* (64,67). In *B. fragilis* (Fig 1.2A), there is a complete *hsd*S gene sitting in line with the *hsdR* and *hsdM* genes, while downstream and in the opposite orientation, there is an inactive *hsd*S gene lacking a start codon. To create even more S subunit diversity there are also two more truncated *hsdS* genes which, with four pairs of inverted repeats, allow the generation of eight different S subunits from six different TRDs (67). The recombination of *hsdS* genes in the *B. fragilis* genome is yet to be experimentally verified.

1.4.3 ***Listeria monocytogenes***

In *L. monocytogenes* there is a type I locus with *hsdR, M* and *S genes* and a second *hsdS* gene situated downstream in the opposite orientation (Fig 1.2A). Each of the two *hsdS* genes contain two TRDs allowing for four possible S

subunit specificities (72). While this system is not present in all *L. monocytogene* sequence types (STs), it appears to be conserved in particular lineages (unpublished data). Genome comparison by Fagerlund *et al.* (72) demonstrated that different active *hsdS* genes are present in different genomes and infer that they might be phase variable. Recombination within *Listeria* strains is ongoing, but is limited to particular environmental conditions (unpublished data, Oggioni Lab).

1.4.4 ***Lactobacillus salivarius***

In 2006 a phase variable Type I RM system (LSL_0915-LSL_0920) was identified in *Lactobacillus salivarius* (73). The system is composed of single *hsdR* and *hsdM* genes alongside three *hsdS* genes (Fig 1.2A). One of these *hsdS* genes is found in the same orientation as *hsdR* and *hsdM*, while the second and third *hsdS* genes are found in the opposite orientation. Located between the first and second *hsdS* genes there is a site specific recombinase (LSL_0917). The locus contains a total of six TRDs and three different inverted repeats (one copy of each repeat is found within each *hsdS* gene) (73). Recombination on the inverted repeats allows the shuffling of the TRDs to generate nine different, active S subunits.

Claesson *et al.* used whole genome sequencing (WGS) reads to measure the proportion each S subunit within a single DNA sample. Sequence reads were mapped to all nine *hsdS* combinations demonstrating that all nine were present within a single sample. In addition, the number of reads was used to quantify their relative abundance in the sample (73).

1.5 **The Phase Variable Type I Restriction-Modification System *Spn*III**

In 2001 a similar Type I system with invertible *hsdS* genes was identified in the pneumococcal serotype 4 strain TIGR4 (71). Later named *Spn*III, this type I RM system is part of the core genome of *S. pneumoniae* (64). The *S. pneumoniae* D39 *spnIII* locus consists of three co-transcribed genes, *hsdR* (SPD_0455), *hsdM* (SPD_0454) and *hsdS* (SPD_0453), along with a tyrosine site specific recombinase (SSR) (SPD_0452) with its own promoter and two truncated *hsdS*

genes. The first of the truncated *hsdS* genes (*hsdS2*, SPD_0450) contains only a single TRD, shown as TRD2.3 (blue) in Fig. 1.3. The second of the truncated *hsdS* genes (*hsdS3*, SPD_0451) contains two TRDs but lacks a ribosome binding site and promoter. In total the locus contains five different TRDs (64,65), which result in six different *hsdS* variants, SpnIIIA-F (Fig 1.3).

Within the three *hsdS* genes are a series of inverted repeats (IRs) which have been proven to facilitate movement of the TRDs (64,65,70). The largest of these inverted repeats (IRs) (α) is 333bp and is located between the two TRDs of *hsdS* and *hsdS3* (Fig. 1.3, checked boxes). Recombination on the α repeat shuffles TRD2 but does not affect TRD1. Partially located within the 333bp is the smallest repeat sequence (β) which is 15bp long. The β repeat is located at the 5' end of TRD2 and will only shuffle TRDs 2.2 and 2.3 (Fig 1.3 horizontal lines). Finally, there is also an 85bp IR (γ) which can be found 5' of TRD1 (Fig 1.3 diagonal lines). Recombination on the 85bp repeat results in a complete inversion of all TRDs and will result in an entirely different active *hsdS* gene.

1.5.1 Phenotypes associated with the phase variable expression of *hsdS* genes

In *S. pneumoniae,* work using "phase-locked" mutants, which are incapable of *hsdS* recombination, has been conducted in mice (64). These phase-locked mutants showed some significantly different *in vivo* phenotypes. Strains expressing SpnIIIA were significantly better at surviving in the blood and significantly more bacteria were recovered from mice intravenously infected with this strain. Alternatively, where mice were intranasally challenged in a carriage model the SpnIIIB strain resulted in significantly higher rates of colonisation. Mice infected with the strain expressing the B specificity showed fewer signs of disease and this strain was deemed to be less virulent (64).

Phenotypic differences in colony opacity have been observed in *S. pneumoniae* strains when phase-locked mutants are generated. Manso *et al*. showed that a D39 strain locked into SpnIIIA is classified as opaque, while those locked into SpnIIIB are >90% transparent (64). The colony opacity phenotypes are classically associated with colonisation (transparent) and invasive disease (opaque) (17). In addition to the work of Manso, Li *et al*. also generated "phase-

locked” mutants (65). While their results differed, indicating that SpnIIIA (*hsdS*$_{A3}$) is transparent in their strain background, this is not entirely surprising. The mechanisms behind the opaque/transparent phenotype are poorly understood and the outcome of colony opacity appears to be highly strain dependent.

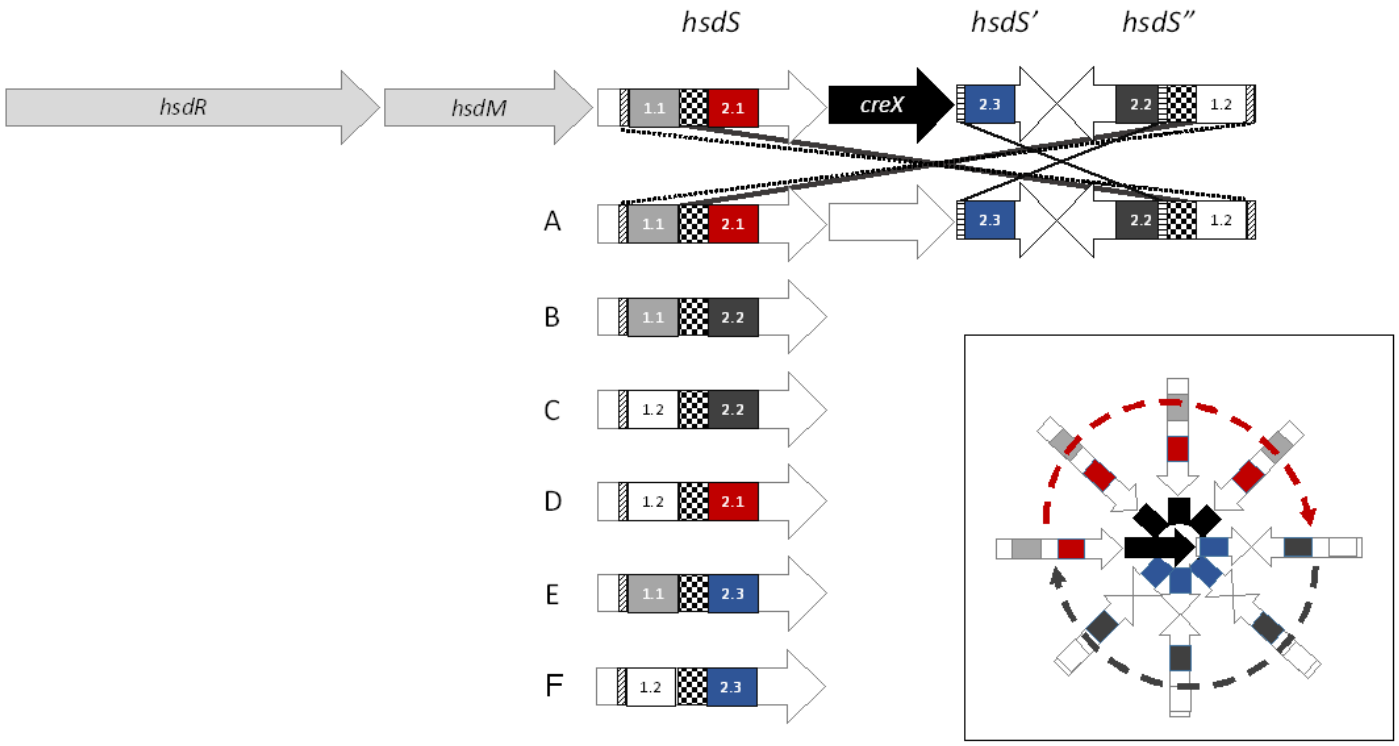


Figure 1.3 - Schematic map of the phase-variable Type I RM locus spnIII. *The SpnIII locus of S. pneumoniae D39* (64,65,71,78)*. The pneumococcal locus includes hsdR (SPD_0455), hsdM (SPD_0454) and hsdS (SPD_0453), a CreX recombinase (phage-type integrase) (SPD_0452), a truncated hsdS2 gene (SPD_0450) with only one variable target recognition domain (TRD) and a further truncated hsdS3 gene (SPD_0451) with two TRDs. There are three independent inverted repeats which allow allele switching and the formation of six different hsdS alleles (named from A to F and shown below the locus map). Inverted repeats are of 85 bp (diagonal lines), 333 bp (checked) and 15 bp (horizontal lines), respectively. Black and dotted lines indicate the direction of recombination occurring at the level of the inverted repeats. The insert represents an illustration of hsdS allele switching and the formation of the six different locus arrangements.*

1.6 Alternative mechanisms of *hsdS* exchange

In addition to recombination on inverted repeats, several other methods of TRD exchange have been investigated. Discussed in more detail below are: *Spn*IV*,* a second phase variable Type I RM system that recombines on direct repeats, a process termed DoMo first identified in *Helicobacter pylori* that shuffles individual protein domains, and finally plasmid mediated exchange of *hsdS* genes in Type I systems.

1.6.1 *Spn*IV

In *S. pneumoniae* there is a second phase variable RM system, *Spn*IV (64,71,78). Unlike *Spn*III, the TRDs of *Spn*IV are shuffled by direct rather than inverted repeats (78). While inverted repeats typically lead to the inversion of DNA fragments by site specific recombinases, direct repeats are associated with the loss of a DNA fragment (79). As a result of this the *Spn*IV RM system is incomplete or inactive in some pneumococcal strains (78) including in the common lab strain D39 (64).

The complete *spnIV* locus consists of a full length *hsdM, hsdR, hsdS* and a tyrosine recombinase, in addition there is a truncated *hsdS* gene with two TRDs and a putative toxin anti-toxin system (Fig 1.4A). In *S. pneumoniae* D39 the deletion includes the complete *hsdS* gene, the recombinase and the toxin anti-toxin system (Fig 1.4B).

There is a site specific recombinase, TrvR, located immediately downstream of the active *hsdS* gene of the *spnIV* locus (Fig 1.2B). It has been shown by PCR analsysis that the loss of the TvrR recombinase prevents any visible reshuffling of the locus (78). Croucher *et al.* propose that TRD shuffling occurs by the excision and re-integration of DNA, and the presence of a putative toxin-antitoxin system within the locus promotes its stability by maximising the likelihood that excised DNA will reintegrate (80).

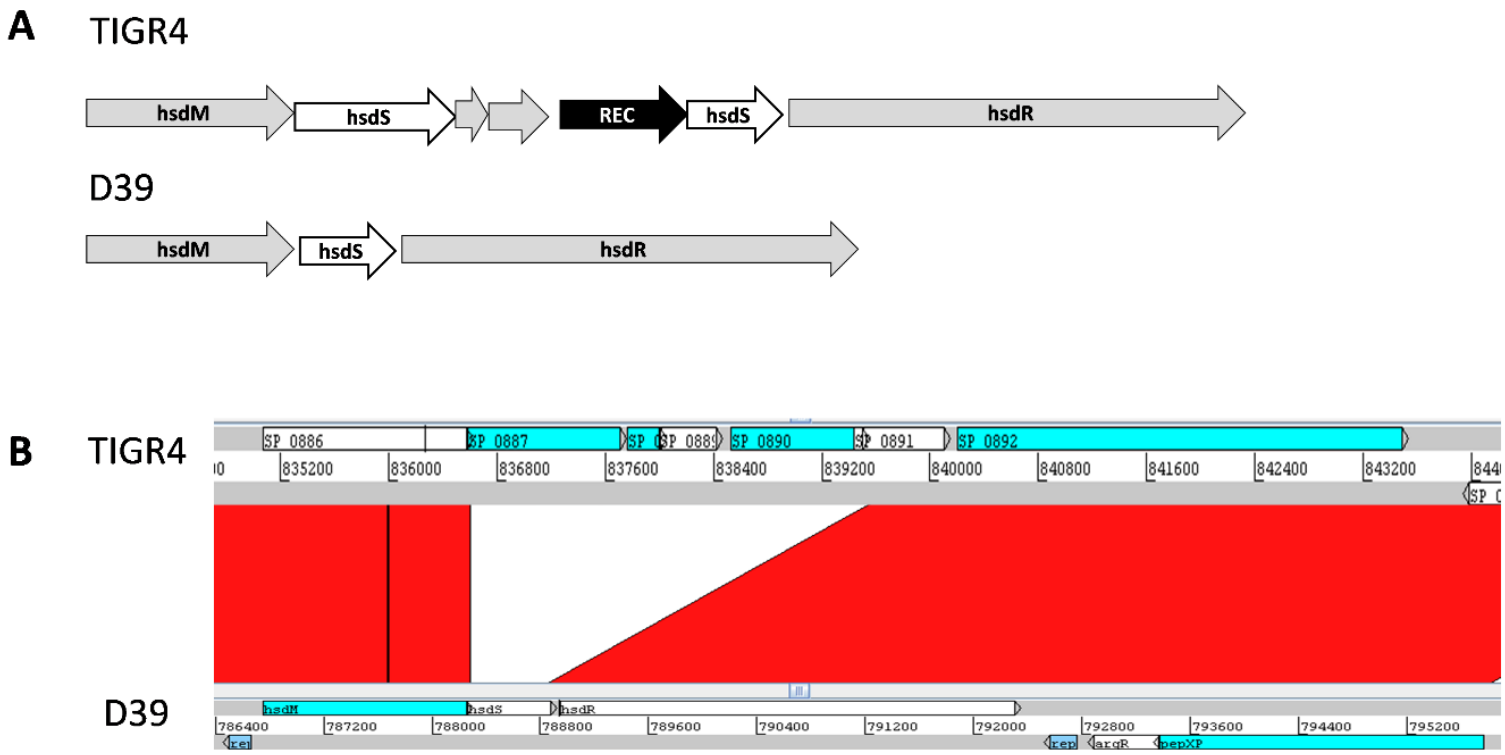


Figure 1.4 - Schematic map and alignment of the phase-variable Type I RM loci SpnIV. *The SpnIV locus of S. pneumoniae TIGR4 and D39* (64,65,71,78)*. Unlike the spnIII loci, spnIV is not found intact in all strains. The complete locus of S. pneumoniae TIGR4 (Panel A) includes the following genes hsdM (SP_0886), a complete hsdS1 (SP_0887), putative toxin anti-toxin system (SP_0888-89), site specific recombinase (SP_0890), a truncated hsdS2 (SP_0891) and hsdR (SP_0892). In S. pneumoniae D39 a significant portion of the locus, including the complete hsdS gene, the putative toxin, antitoxin system and the recombinase, have been lost. There are two pairs of direct repeats in the intact TIGR4 locus that facilitate the movement of four target recognition domains (TRDs). An alignment of TIGR4 and D39 using the Artemis Comparison Tool (ACT) (Panel B) indicates where the deletion in the D39 genome has occurred.*

1.6.2 **Domain Movement**

In *Helicobacter pylori,* a species with large number of RM systems, an alternative process of interchanging TRDs has been identified (81,82). Named Domain Movement, or DoMo, the process relies on flanking direct repeat sequences to move TRDs within and between genes (81,82). There are six different *hsdS* loci in the *H. pylori* genome which can be classified into three distinct groups based on sequence similarity (81). Of these groups one (group 1S) lacks any repeat sequences and does not shuffle TRDs, while the other two (groups 2S and 3S) do. Unlike the invertible systems previously discussed, groups 2S and 3S are capable of swapping TRDs between positions 1 and 2, i.e. in different strains the same TRD may be found encoding TRD1 or TRD2. In addition, TRDs can be swapped across between different loci (81). Group 2S and 3S *hsdS* genes have TRDs flanked by two pairs of direct repeats, the repeats are unique to each *hsdS* group and mediate the movement of TRDs between genes of the same group (81). The initial work conducted by Furuta et al. used genome comparisons to verify DoMo, however more recent work using SMRT sequencing has confirmed that DoMo creates functional HsdS subunits which produce different sequence specificities (82).

In addition to their work in *H. pylori,* Furuta et al. also confirmed that DoMo occurs in *Streptococcus pyogenes* and *Mycoplasma agalactiae* (81). *S. pyogenes* has a single conserved Type I RM system with TRDs that are flanked by two repeats of 37bp and 29bp. A comparison of 13 different genomes proved that TRDs were capable of movement between TRD1 and TRD2 within a single locus (81). In *M. agalactiae* two strains were found to have *hsdS* genes at three different loci, the TRDs were flanked by 37bp and 34bp repeat sequences and TRDs were exchanged between TRD1 and TRD2 in the different strains. In addition, one strain was found to have a single TRD and another to have the structure TRD1 TRD2, TRD2. It is not known if either of these loci encode a functional S subunit (81).

1.6.3 **Plasmid mediated**

A plasmid encoded Type I RM system in *Lactococcus lactis* was one of the first Type I systems identified in Gram positive bacteria (83). In addition to plasmids that carry complete Type I systems, several other plasmids have been identified that carry single *hsdS* genes with no R or M encoded (84–86). Work by Schouler *et al.* demonstrated that a plasmid encoded S subunit conferred enhanced phage resistance through interactions with the chromosomally encoded R and M subunits (86).

Multiple cases of lone *hsdS* genes encoded on plasmids interacting with chromosomally encoded R and M subunits have since been reported (84,85). Madsen *et al.* described an S subunit, S.LlaW12I on the 8 kb plasmid pAW122, capable of interacting with the R and M subunits of the chromosome (or other plasmids) to reduce sensitivity to phage (84). While O'Sullivan *et al.* reported the co-integration of two plasmids, pAH33 which encodes a lone *hsdS* and pAH82 which encodes a complete Type I system (85). The integration of these plasmids, by homologous recombination, occurred at the conserved site between TRDs in the *hsdS* gene. The result was a single, larger plasmid, pAH90, with two new chimeric *hsdS* genes. Cells containing pAH90 were shown to be much more phage resistant as a result of the two new, unique RM specificities conferred by these recombined *hsdS* genes (85).

Despite their ability to methylate, the biological significance of these plasmid mediated systems appears to be limited to a defensive mechanism against phage rather than as a method of global gene regulation.

1.7 **Site Specific Recombination**

Site specific recombination is a process that uses an enzyme to catalyse the cleavage and re-joining of DNA independent of the cells homologous recombination machinery (87). Site specific recombinases (SSRs), the enzymes that facilitate site specific recombination, cut and re-ligate at specific recognition sequences without ATP or the synthesis of any new DNA (88–90). In order for site specific recombination to occur two recombinase molecules must bind, one to each recognition sequence, allowing the sequences to interact (88).

SSRs occur in a wide variety of bacteria and phage and enable the inversion and excision of DNA. When the sequence recognised by the enzyme is inverted (inverted repeats) it facilitates the inversion of the DNA between the recognition sites (91,92). However, when the recognition sites are in the same orientation (direct repeats) it promotes excision from the genome, this is typically associated with the excision of lysogenic phage from the host (79).

Effectively all SSRs fall into one of two phylogenetically unrelated families, tyrosine and serine recombinases. The nomenclature of the enzyme families is based upon the amino acid residue used to bind DNA and facilitate the reaction (88). Each will be discussed separately below.

1.7.1 **Tyrosine recombinases**

The tyrosine family of SSRs excise and invert DNA by the formation of Holliday junction intermediates (88,89,93). Recombination does not require extensive homology, instead there is a small region of homology (the crossover region) flanked by two recombinase binding sites. Cleavage occurs sequentially with one (top) strand being cleaved, exchanged and re-joined before the second (bottom) strand is cleaved (88,93). Upon cleavage of the top strand the tyrosine of the enzymes active site forms a covalent DNA-protein linkage (phosphotyrosine) at the 3' end of the DNA, leaving free hydroxyls at the 5' end. Using the energy transferred from the phosphodiester bond of the DNA the 5' end is now able to attack the 3' phosphotyrosine of the opposite strand forming a Holliday junction intermediate. The process is then repeated on the bottom strand. The resolution of the Holliday junction occurs when the second 5' strand attacks the second 3' phosphotyrosine bond, releasing the recombinase and completing the reaction (88,93)

The related tyrosine recombinases XerC and XerD are some of the most highly conserved and well-studied tyrosine recombinases (89,93). Unlike the majority of SSRs, XerC and XerD function together, and their primary role is to resolve chromosomal dimers (93,94). Resolution occurs when one monomer of each enzyme binds a *dif* (deletion induced filamentation) site separated by a 6-8bp

crossover region (93). This binding will occur in two locations. The two XerC monomers catalyse the cutting and rejoing of the first, or top, strand. Once the top strand has been religated the two XerD monomers will catalyse the cutting and rejoining of the second, or bottom, strand. When both strands have been religated the Holliday junction intermediate is resolved and the recombinase molecules released (95).

Despite the conserved nature of the Xer system, an alternative XerS enzyme has been identified in the *Streptococci* and *Lactococci* which performs the function of both XerC and XerD. XerS is also a tyrosine recombinase but is phylogenetically distinct from XerC/D (94,96).

1.7.2 **Serine recombinases**

The serine family of site specific recombinases are more conserved than the tyrosine recombinases, although there is a wide variation in enzyme size (87,88). Serine recombinases are characterised by two conserved domains which include the serine residue that is used to catalyse recombination (87,88). They are heavily reliant on the Fis (Factor for Inversion stimulation) enhancer protein and its associated enhancer sequence that control the directionality of recombination, often favouring one direction over another.(87). Recombination occurs through the formation of an invertasome complex, where Fis acts as a scaffolding molecule to bring together the inactive recombinase dimers bound to each of the two recognition sites (87). Once the invertasome complex has been assembled the recombinase subunits form an active tetramer that mediates two simultaneous double strand breaks. These breaks occur at the central two base pairs of the recombinase recognition sites (87,88). The 5’ ends of the DNA form covalent phosphoserine linkages with the recombinase subunits, leaving free 3’ OH ends (87,88). One pair of recombinase subunits rotates around the other bringing the DNA into its new recombinant orientation. Once in position the unbound 3’ OH ends are free to attack the phosphoserine bond, ligating the DNA in its new orientation and releasing the enzyme complex (87).

The majority of serine recombinases act on a single invertible DNA region, however there are some exceptions. These include the Min invertase, found on

the *E. coli* plasmid p15B, and the serine recombinases found in the *B fragilis* genome. The 26bp recombination sites of the Min invertase are found in six different locations on the plasmid (87). While the Mpi recombinase of *B. fragilis* catalyses DNA inversion in at least 13 different promoter regions (97).

One of the most well studied serine recombinases is the 190aa Hin recombinase of *Salmonella enterica* (87). The Hin recombinase is found in the invertible 993bp H segment which controls the expression of alternate flagellin genes *fljB* and *fliC* (87,91). Hin catalyses the inversion of the promoter region by binding the *hixL* and *hixR*. When orientated in line with the operon *fljB* is cotranscribed with *fljA*. When expressed FljA binds and degrades *fliC* mRNA, preventing translation, and ensuring expression of the FljB flagellin protein (87,91). Following an inversion event mediated by Hin there is no longer expression of the *fljB* and *fljA* operon (87,91). Without the FljA protein *fliC* mRNA is not degraded and becomes the expressed flagellin protein. This process also required the Fis protein binding the recombinational enchancer locus, without this the invertasome complex does not assemble and recombination is nearly undetectable (87). It has also recently been proven that recombination of the H segment can occur at low levels in the absence of Hin (91). H segment inversion can also be facilitated by Fin, another serine recombinase found 34.5 kb from the H segment. While inversions occur 60-200 times less frequently in the absence of Hin, it is only in a double knockout mutant of both *hin* and *fin* that H segment inversion is completely abolished (91).

1.7.3 **Recombinase Driven Inversions of *hsdS* genes**

Several of the phase variable Type I RM systems described have an associated site specific recombinase.

In ST8 strains of *Listeria monocytogenes* (72). The phase variable Type I RM system utilizes site-specific recombination to switch between the four possible DNA target recognition site specificities, which have been named A, B, C and D. The presence of two pairs of inverted repeats (5’-AGCTTGGGAACAGCGT-3’ and 5’-CTATCGCTCTTCATCAGCGTAAGTTAGAT-3’), which are located in the 5’-end and in the central part of the *hsdS* genes allow recombination to occur. Although direct evidence of its role is still missing, a putative tyrosine

recombinase is encoded between the two hsdS genes and Fagerlund et al. propose that it is most likely responsible for the reported *hsdS* gene inversions (72).

Within the *hsd1* locus of *M. pulmonis* there are two pairs of inverted repeats (5'-CAAAGTGCAATA-3' and 5'-TAATTAAGATTATTGAACCT-3'), which allow the generation of four different active S subunits (77). Inversions at the locus are the result of a single site specific tyrosine recombinase HvsR, however this recombinase is not found within the type I locus (77). Instead it is found near the phase variable *vsa* surface protein genes (98). In what is thought to be a unique situation, the recombinase controlling the inversion at the *vsa* locus also has complete control of *hsdS* gene inversions. Using transposon mutagenesis to generate mutants of HvsR Sitaraman *et al.* (77) were able to prove that, despite the lack of homology, HvsR facilitates recombination at both the *vsa* and *hsdS* loci. When the system was reconstructed within an *E. coli* background it was determined that HvsR alone is capable of facilitating recombination when the longer 20bp inverted repeat sequence is present, however the shorter sequence alone does not appear to be sufficient (77) indicating that another mycoplasmal protein may be required.

In *B. fragilis* the specific recombinase associated with the BF1839 Type I system is not known. However, the genome contains more than 30 site specific recombinases (99) three of which are within close proximity to the locus, BF1833, 1843, and 1845 (67). As yet there is no published evidence to determine which, if any, of these recombinases permits inversions at the Type I locus.

It is of note that these inversions appear to always occur independently of RecA, therefore most likely exclusively through site specific recombination).

1.8 Aims and Objectives

The work of this thesis addresses the hypothesis that *Spn*III dependent methylation has an impact on gene expression and phenotypes in the pneumococcus.

The overall aim of this research project was to identify the molecular mechanisms driving the recombination of the phase variable type I RM system *Spn*III in *S. pneumoniae* and to monitor recombination and selection during experimental carriage and infection. To address these aims the following objectives were addressed:

1. Determine how differential SpnIII expression affects *S. pneumoniae in vitro* through the generation and analysis of strains enriched in a single SpnIII variant.
2. Determine which proteins regulate inversions at the *spnIII* locus through the production and analysis of knockout mutants related to site specific and homologous recombination.
3. Determine if all SpnIII variants are equally capable of nasopharyngeal colonisation and invasive disease and if selection in favour of individual variants is on going *in vivo.*

2 Materials and Methods

2.1 Growth and storage of bacterial strains

S. pneumoniae strains were grown in liquid cultures in either BBL™ Trypticase™ Soy Broth (TSB), casein tryptone broth (CAT), or Brain Heart Infusion (BHI). For growth on solid media BBL™ Trypticase™ Soy Agar (TSA) or Brain Heart Infusion Agar (BHIA) supplemented with 3% v/v Defribinated Horse Blood (ThermoScientific, UK) were used. Where necessary catalase was used as a substitute for horse blood at a concentration of 200U/ml. Where appropriate media was supplemented with antibiotics, the concentrations of which can be found in table 2.1. All strains were stored at -80°C with 10% v/v glycerol or where required 10% v/v sucrose. All media was autoclaved at 121°C for 15 minutes to sterilise.

2.1.1 TSB

12 g of TSB powder (Becton Dickinson, USA) was added to 400 ml of deionised water and autoclaved at 121°C for 15 minutes to sterilise. Broth cultures were incubated at 37°C, 5% CO_2 in a Galaxy B Incubator (Wolf Laboratories, UK).

2.1.2 TSA

12 g of TSB powder (Becton Dickinson, USA) and 6 g of Bioagar (Biogene, UK) were added to 400 ml of deionised water and autoclaved at 121°C for 15 minutes to sterilise. Agar was cooled to 50°C and then supplemented with 3% v/v defribinated horse blood (ThermoScientific, UK) and antibiotics where appropriate. Plates were stored at 4°C for no longer than 4 weeks. Plates were incubated at 37°C, 5% CO_2 in a Galaxy B Incubator (Wolf Laboratories, UK). Where fatty acid supplementation of agar was required 0.1% of filter sterilised Tween 80 was added after sterilisation (100). Where supplementation with CSP1 was required 0.625 µg/mL was added to agar after sterilisation.

2.1.3 BHI

Brain Heart Infusion (Oxoid, UK) was prepared in 200 ml volumes. As stated in the manufacturer's instructions 7.4 g of BHI powder was dissolved in 200 ml of deionised water and autoclaved at 121°C for 15 minutes to sterilise. Broth cultures were incubated at 37°C, 5% CO_2 in a Galaxy B Incubator (Wolf Laboratories, UK).

2.1.4 **BHI-CTM**

For competent cells BHI media was supplemented with 0.1% 1M $CaCl_2$, 0.2% glucose, and 0.4% Bovine Serum Albumin (BSA) (Sigma, UK). Cells were grown in BHI and diluted 1:10 into BHI-CTM and incubated at 37°C, 5% CO_2, for 45-90 minutes. After growth in BHI-CTM cells were either used immediately for transformations or stored at -80°C with 10% v/v glycerol and used for transformations as required.

2.1.5 **BHIA**

Brain Heart Infusion Agar (Oxoid, UK) was purchased from Oxoid and prepared in 400 ml volumes. As stated in the manufacturer's instructions 18.8 g of BHIA powder was dissolved in 400 ml of deionised water and autoclaved at 121°C for 15 minutes to sterilise. Agar was cooled to 50°C and then supplemented with 3% Defribinated horse blood (ThermoScientific, UK) and antibiotics where appropriate. Plates were stored at 4°C for no longer than 4 weeks. Plates were incubated at 37°C, 5% CO_2 in a Galaxy B Incubator (Wolf Laboratories, UK).

2.1.6 **CAT**

4 g Bacto Casitone (Becton Dickinson, USA), 0.4 g Yeast Extract (Oxoid, UK), 4 g Tryptone (Oxoid, UK), and 2 g NaCl were dissolved in 400 mL deionised water and autoclaved at 121°C. CAT media was supplemented with 3% v/v 0.5M K_2HPO_4 and 0.2% v/v glucose. Broth cultures were incubated at 37°C, 5% CO_2 in a Galaxy B Incubator (Wolf Laboratories, UK).

2.1.7 **Low Yeast Extract CAT (Low YE CAT)**

For growth experiments where alternative sugars were required the yeast extract in CAT media was reduced to minimise glucose. 4 g Bacto Casitone (Becton Dickinson, USA), 0.025 g Yeast Extract (Oxoid, UK), 4 g Tryptone (Oxoid, UK), and 2 g NaCl were dissolved in 400 mL deionised water and autoclaved at 121°C for 15 minutes. CAT media was supplemented with 3% v/v 0.5M K_2HPO_4 and sugars as required. Broth cultures were incubated at 37°C, 5% CO_2 in a Galaxy B Incubator (Wolf Laboratories, UK).

2.2 **Growth assays**

For growth assays using different sugars, strains were plated on TSA blood agar plates and incubated at 37°C, 5% CO_2 overnight. Cells were collected using a

soft cotton swab and resuspended in low YE CAT before dilution to a starting OD of ~0.003. Low YE CAT was supplemented with a final concentration of 0.2% of the relevant sugar and 200U/ml of catalase. Growth assays were run at 37°C in an Eon Microplate Spectrophotomer (BioTek, UK) for 24-48 hours depending on the sugar used.

2.2.1 **Generation time calculation**

Exponential Generation time (G) was calculated as follows

B = OD_{590nm} at end point

b = OD_{590nm} at start point

t = Time (hours)

n = (LOG(B) - LOG(b))/(0.301*t)

G in minutes =(1/n)*60

2.2.2 **Plating of serial dilutions**

10 fold serial dilutions were plated using the method previously described by Miles & Misra (101). Samples or cultures were serially diluted 1:10 in TSB up to 10^{-8}. Once diluted 20 µl drops were pipetted onto air dried agar plates in triplicate. Plates were left to dry before being inverted and incubated as previously described (2.1.2). Following 16-20hours incubation colonies in the drops with the greatest number of discrete (un-merged) colonies were counted. Colony forming units per ml (CFU/ml) was calculated as follows:

CFU/ml= Average colonies for dilution counted x 50 x dilution factor.

2.2.3 **UV susceptibility protocol**

Cells were grown to an exponential OD of ~0.2 in 50 ml TSB. Cultures were centrifuged at 4,000rpm for 10 minutes to pellet cells. Pellets were re-suspended in 20 ml of 0.9% NaCl solution. A 10 fold serial dilution of cells was plated prior to UV exposure to determine total cells numbers. Cells in 0.9% NaCl were placed in a sterile petri dish and exposed to a UV source producing 10 J/Min for 45

seconds. A 10 fold serial dilution of UV exposed cells was then plated to determine the number of surviving cells.

2.3 DNA and RNA extraction and purification

2.3.1 Extraction of *S. pneumoniae* genomic DNA from cultured cells

1 ml of exponentially growing cells ($OD600_{nm}$ =~0.2) were centrifuged at 13,000 rpm in an Eppendorf 5424R centrifuge (Eppendorf, USA) at room temperature, for 2 minutes to pellet cells. Cells were then washed in 600 µl cell wash buffer (0.15M NaCl, 0.015M trisodiumcitrate, dH_2O) and centrifuged at 6000 rpm, room temperature, for 5 minutes. After washing cells were resuspended in 450 µl of lysis buffer (0.1% sodium deoxicolate (DOC), 0.01% sodiumdodecylsulfate (SDS), TE pH8) and incubated at 37°C for 10 minutes, or until turbidity had cleared. Once turbidity had cleared 1mg/ml of proteinase K was added and samples were incubated at 60°C for 60 minutes.

The Genomic DNA Clean & Concentrator™ kit (Zymo Research, USA) was then used to extract DNA as follows. A 2:1 volume of ChIP DNA Binding Buffer was added to the sample and mixed by vortexing. Samples were transferred to a spin column and centrifuged at 13,000 rpm for 30 seconds. 200 µl of DNA wash buffer was added to the column and centrifuged at 13,000 rpm for 1 minute. The flow through was discarded and the wash step repeated. The column was centrifuged at 13,000 rpm for 2 minutes to remove any remaining ethanol. Following this the column was placed in a clean microcentrifuge tube and 50 µl of pre-warmed H_2O or TE was added to the column. The sample was incubated at room temperature for 5 mins to elute DNA and then centrifuged at 13,000 rpm for 30 seconds. Where required 1mg/ml RNase was added to the sample and incubated for 20 minutes at 37°C. Following extraction DNA was stored at -20°C.

2.3.2 Sanger sequencing

Sanger sequencing was used to confirm mutants. The region of interest was PCR amplified using the appropriate primers in table 2.3. PCR products were visualised on a 1% (w/v) agarose (SeaKem LE Agarose, Lonza, USA) gel at 100 V for 45 min in 1 x TAE buffer using a 1 kb DNA ladder (HyperladderI). The E.Z.N.A Cycle-Pure Spin kit (Omega Bio-tek, UK) was used to clean up PCR

products according to the manufacturers protocol. Products were eluted in 50 µl of pre-warmed H_2O. The DNA sequence protocol for a single sample was as follows; 2 µl PCR product, 2 µl primer at 1 pmol/µl, 4 µl 1/8 Big Dye (0.5 µl Big Dye, 0.8 µl 5 X Big Dye buffer, 2.7 µl H_2O), 2µl H_2O. This reaction was run in an Eppendorf Mastercycler (Eppendorf, USA) at 95°C for 5mins, followed by 29 cycles of the following; 96°C for 10 secs, 50°C for 10 secs, 60°C for 4 mins. The product of this reaction was heated to 98°C for 5 mins with 0.2% v/v SDS and run through a Performa DTR gel filtration cartridge (EdgeBio, USA) following the manufacturer's instructions. The eluted sample was run on an ABI prism Gene Analyser (Applied Biosystems, USA).

2.3.3 Extraction of *S. pneumoniae* RNA from cultured cells

For gene expression analysis pneumococcal strains were grown to mid-log phase (OD_{590} ~0.15). 5ml of cells were added to 1ml of an ice cold 95% ethanol 5% phenol solution, before centrifugation at 4000rpm for 10mins. The supernatant was removed and the pellets stored at -80°C until processing. RNA was extracted using the Maxwell 16 LEV simplyRNA cells kit (Promega, USA) and the Maxwell 16 LEV instrument (Promega, USA). The manufacturers protocol was followed from step 4, the manufacturers lysis steps (1-3) were replaced with the following protocol to improve cell lysis; pellets were resuspended in 50ul TE with 3 mg/ml lysozyme and incubated at 37°C for 10-20mins. Once extracted samples were stored at -80°C.

2.3.4 RNAseq

S. pneumoniae was grown at 37°C, 5% CO_2 in BHI to OD 0.15-0.18. RNA was extracted as described in method 2.3.3. RNAseq analysis was undertaken at the University of Leicester, UK using a MiSeq Desktop Processor (Illumina, USA) and the ScriptSeq Complete Kit (Bacteria) (CamBio, UK) which includes an rRNA depletion step. Raw data was processed using the bacterial RNAseq specific programme Rockhopper v2.0.3 (102) using D39 as the reference genome.

2.3.5 Whole Genome Sequencing

S. pneumoniae was grown at 37°C, 5% CO_2 in TSB to ~OD 0.2. DNA was extracted as described in method 2.3.1. Genome sequencing was provided by MicrobesNG (http://www.microbesng.uk), which is supported by the BBSRC

(grant number BB/L024209/1). Samples were processed using an Illumina 2500 HiSeq platform to give 250bp paired end reads. SPAdes was used for a de novo assembly of reads and trimmed files were received. VarScan.v2.3.9 was used to call variants and identify SNPs, Tablet v1.16.09.06 was used to identify INDELs. D39 was used as a reference for all analysis (GenBank Accession number: CP000410).

2.4 PCR

2.4.1 Taq polymerase PCR

PCR reactions using Taq polymerase used the following PCR mastermix; 0.5 µM forward and reverse primers (see table 2.3), 0.5 µM dNTPs, 0.05U KapaTaq Polymerase (Kapa Biosystems, UK), 1X KAPA Taq buffer A. All PCR reactions were run on an Eppendorf Mastercycler Pro (Eppendorf, USA) using the following conditions; denaturation at 95°C for 5 mins, followed by 30 cycles of denaturation at 95°C for 30 secs, annealing at 50-62°C for 20 secs and extension at 72°C for 1-5 mins. This was followed by a final extension at 72°C for 5 mins. Annealing temperature used is dependent on optimal primer Tm, and extension was calculated at 1 min/kb amplified. 5 µl of PCR product was run on a 1% (w/v) agarose (SeaKem LE Agarose, Lonza, USA) gel at 100 V for 45 min in 1 x TAE buffer using a 1 kb DNA ladder (HyperladderI).

2.4.2 Phusion polymerase PCR

PCR reactions using the high fidelity Phusion polymerase used the following PCR mastermix; 2.0 µM forward and reverse primers (see table 2.3), 0.2 µM dNTPs, 0.02U Phusion Polymerase (Thermo Scientific, UK), 1X HF buffer. All PCR reactions were run on an Eppendorf Mastercycler Pro (Eppendorf) using the following conditions; denaturation at 94°C for 5 mins, followed by 28 cycles of denaturation at 94°C for 15 secs, annealing at 50-62°C for 15 secs and extension at 72°C for 1-5 mins. This was followed by a final extension at 72°C for 5 mins. Annealing temperature used is dependent on optimal primer Tm, and extension was calculated at 1 min/kb amplified. 5 µl of PCR product was run on a 1% (w/v) agarose (SeaKem LE Agarose, Lonza, USA) gel at 100 V for 45 min in 1 x TAE buffer using a 1 kb DNA ladder (HyperladderI).

2.5 Allele quantification of *spnIII*

The following PCR based allele quantification method has been validated against genomic data by Lees *et al.* 2017 (23). The mapping of active TRDs using whole genome sequencing reads was not significantly different from the PCR based analysis of the same samples. It was also shown that analysis on clinical samples was comparable with analysis of cells cultured from those samples (23).

2.5.1 Allele quantification PCR

To determine SpnIII populations in recovered cells primers FAM-AMRE74 and AMRE59 were used to PCR amplify a 4.2 kb PCR. Amplification was performed in a 25µl reaction consisting of 0.75µl 10mM FAM-AMRE74 forward primer [FAM-GGAAACTGAGATATTTCGTGGTGATGATGGGA], 0.75 µl 10mM AMRE59 reverse primer [CCTGATCGAGCGGAAGAATATTTCTGCCGAGGTTGCC], 19.85 µl dH_2O, 2.25 µl 11.1X buffer (45 mM Tris HCl (pH8.8), 11 mM Ammonium sulphate, 4.5 mM $MgCl_2$, 6.7 mM 2-mercaptoethanol, EDTA (pH8.0), 1mM dNTPs, 113 µg/ml Bovine Serum Albumin)(103), 0.2 µl Kapa Taq (5 U/µl) (Kapa Biosystems, UK), 0.15 µl Tris pH8.8, 0.05 µl PFU (2.5 U/µl) and 1 µl resuspended cells as template. All PCR reactions were performed on a Mastercycler Pro (Eppendorf, UK) as follows: denaturation at 95°C for 5 min, followed by 40 cycles of 1 min denaturation at 95°C, 1 min annealing at 68°C, and 5 min extension at 68°C, with a final extension of 10 min at 68°C. PCR products were examined by agarose gel electrophoresis. 3 µl of PCR product was run on a 1.2% (w/v) agarose (SeaKem LE Agarose, Lonza, USA) gel at 150 V for 30 min in 1 x TAE buffer using a 1 kb DNA ladder (HyperladderI).

2.5.2 Allele Quantification Restriction Digest.

Following PCR amplification, a restriction digest of the 4.2 kb product was performed using restriction enzymes DraI (New England Biolabs, UK) and PleI (New England Biolabs, UK) in a 20 µl reaction. This reaction consists of 0.2 µl DraI (5 U/µl) (New England Biolabs, UK), 0.4 µl PleI (20 U/µl) (New England Biolabs, UK), 1.6 µl CutSmart Buffer (New England Biolabs, UK), 10 µl PCR product, and 7.8 µl dH_2O. Samples were incubated at 37°C for 2 hours followed by heat inactivation of the restriction enzymes at 60°C for 20 minutes. Following digestion each SpnIII variant has a unique size (see table 2.1) that can be

distinguished through capillary electrophoresis on an ABI prism Gene Analyser (Applied Biosystems, USA).

2.5.3 **Capillary Electrophoresis**

Samples were prepared for capillary electrophoresis on the ABI Prism Analyser by loading into a semi-skirted 96 well PCR plate. For each digested sample 1 µl was loaded with 0.5 µl GeneScan™ LIZ1200 Size Standard (Applied Biosystems, USA) and 9 µl formamide (Fisher Scientific, USA). The GeneScan™ 1200 LIZ® Size Standard is suitable for sizing DNA fragments up to 1200 bases. Samples were then run on the ABI Prism Gene Analyser. The data received from the ABI Prism Gene Analyser was analysed using Peak Scanner v1.0 software. The area of an individual peak was then compared to the total peaks in the sample, generating a percentage value. Samples with fewer than 1000 total reads were discarded as a low total peak value can distort results in favour of the most abundant variant.

Table 2-1 – Fragment length of SpnIII PCR products following digestion with DraI and PleI restriction enzymes.

SpnIII *hsdS* Gene	Fragment length following digestion (bp)
SpnIIIA	1188
SpnIIIB	1110
SpnIIIC	1090
SpnIIID	1168
SpnIIIE	1051
SpnIIIF	1032

2.5.4 Preparation of single colonies for allele quantification

For PCR analysis of colonies stocks were streaked to single colonies on TSA with 3% v/v defribinated horse blood. Plates were incubated for 12-16 hours at 37°C, 5% CO_2 unless otherwise specified. Longer incubation times were used for mutants which showed slower generation times. Single colonies were picked using sterile pipette tips and transferred directly to the prepared PCR reaction. Where required additional colonies were picked using a sterile pipette tip and dilutions plated to estimate generation time. Validation of single colony analysis compared to liquid stocks is shown in Fig 2.1.

2.6 Transformation of *S. pneumoniae*

2.6.1 Preparation of competent cells

BHI was inoculated 1:100 using frozen cultures and then incubated at 37°C, 5% CO_2 until an $OD590_{nm}$ 0.03-0.05 was reached. Cells were then diluted 1:10 into BHI-CTM and incubated at 37°C for 60 minutes. Cells were either used immediately for transformation or were stored at -80°C in 10% v/v glycerol until required. When cells were used immediately glycerol was added before use to allow for the freezing of transformed cells.

2.6.2 Transformation of *S. pneumoniae*

200 µl of competent cells, prepared as stated in 2.6.1, were incubated with 5 µl of DNA at a minimum concentration of 10 ng/µl, and either 5 µl CSP1 (25 µg/ml) or CSP2 (100 µg/ml), dependent on strain. Cells were incubated for 45 minutes before plating with appropriate antibiotics (see table 2.2) using BHIA. Depending on antibiotic used cells were plated using one of two methods.

Method 1 (Streptomycin and Kanamycin):

5 ml Agar was added to a sterile petri dish. 100 µl of transformed cells were added to 2 ml BHI with 400 µl defribinated horse blood. The solution containing the cells was then added to 10 ml cooled liquid BHIA which had been supplemented with the appropriate antibiotics (see table 2.2). Once dry, plates were inverted and incubated overnight at 37°C, 5% CO_2 and examined the

following morning for successful transformants. In addition, transformations were stored at -80°C to allow for replating where necessary.

Method 2 (Spectinomycin, Erythromycin and Chloramphenicol):

5 ml Agar was added to a sterile petri dish. 100 µl of transformed cells were added to 2 ml BHI with 400 µl defribinated horse blood. The solution containing the cells was then added to 10 ml cooled liquid BHIA. Once dry, plates were incubated for 90 minutes at 37°C, 5% CO_2 to allow for expression of antibiotic resistance genes. Following incubation, a 5 ml agar layer was added to the plate and allowed to dry. Once dry a final layer of 5 ml, supplemented with the appropriate antibiotics (see table 2.2), was added. When dry plates were inverted and incubated overnight at 37°C, 5% CO_2 and examined the following morning for successful transformants. In addition, transformations were stored at -80°C to allow for replating where necessary.

After overnight incubation successful transformants were picked using a sterile pipette tip and streaked onto two BHIA plates, one supplemented with the appropriate antibiotic and one without antibiotic. If possible a minimum of three colonies were selected. Colonies that grew on both plates were selected as successful transformations and grown to an $OD590_{nm}$ of ~0.2 in liquid media before freezing at -80 with 10% v/v glycerol.

Table 2-2– Antibiotic concentrations

Antibiotic	**Concentration (µg/ml)**
Kanamycin (Sigma, UK)	500
Streptomycin (Sigma, UK)	500
Spectinomycin (Sigma, UK)	200
Chloramphenicol (Sigma, UK)	10
Erythromycin (Sigma, UK)	10

2.7 *In vivo* experiments

2.7.1 Preparation of SpnIII wildtype stocks.

In order to assess how SpnIII recombination impacts virulence, stocks were generated that are enriched in one of the six SpnIII variants. These stocks were

generated through the selection and passaging of *S. pneumoniae* single colonies on Trypticase Soy Agar (TSA) (Becton, Dickinson and Company, USA) supplemented with 3% defribrinated horse blood. *S. pneumoniae* was plated to single colonies, which were then re-passaged to increase cell numbers. The re-passaged cells were resuspended in 100 µl of Trypticase Soy Broth (TSB) (Becton, Dickinson and Company, USA) containing 10% v/v glycerol and stored at -80ºC. To establish wildtype (WT) populations allele quantification was used to identify passaged colonies high in each individual SpnIII allele. Where necessary this process was repeated to obtain stocks enriched in all six SpnIII variants.

Once identified these stocks were then grown to an exponential OD_{590nm} of 0.2-0.3 in TSB at 37ºC, 5% CO_2 and stored at -80ºC with 10% v/v glycerol. Before use in any experiment the SpnIII proportions in each population were confirmed by allele quantification (See method 2.4). The CFU/ml of each stock was determined through serial dilution and adjusted to the appropriate dose for each individual experiment before use.

2.7.2 **Ethical Statement.**

All experiments were conducted in female MF1 or CD1 mice (Charles River, UK, or Harlan Nossan, Italy) at eight-nine weeks of age. A change from MF1 to CD1 mice was made due to supply issues with MF1 mice from Charles River, UK. Mice were housed in groups of five in the University of Leicester Central Research Facility (CRF) on a 12 hour light-dark cycle with free access to food and water. All mice were given a minimum of one week to acclimatise before experiments were undertaken. All work at the university of Leicester was conducted under Project Licence 60/4327 and Personal Licence I1FB642DC, under the supervision of Dr Vitor Fernandes. All procedures were approved by the Home Office and were in compliance with the Animals (Scientific Procedures) Act 1986 and local guidelines. The intracisternal model of infection was conducted at the Dipartimento di Biotechnologie Mediche, Universita di Siena, Siena, Italy under the supervision of Dr Susana Ricci. All procedures adhered to local guidelines.

2.7.3 **Intranasal pneumonia model.**

For each SpnIII WT strain five female nine week old MF1 and CD1 mice (Charles River, UK) were intranasally inoculated with 50 µl of *S. pneumoniae* at $4x10^6$ CFU/ml in TSB media. Mice were anesthetised using 2.5 % (v/v) isoflorane at 1.6-1.8 litre O_2/min in an anaesthetic box. The challenge dose was delivered by inhalation while mice were held in a vertical position. A final challenge dose of $1x10^5$ CFU was delivered to each mouse. Mice were allowed to recover on their backs before being returned to their cage. Each group was housed individually and had free access to food and water. Mice were monitored for signs for disease and culled by cardiac puncture and cervical dislocation 24 hours after challenge. Cardiac puncture was conducted under anaesthetic using 2.5% v/v isoflurane at 1.6-1.8 litre O_2/min delivered in an anaesthetic box. Approximately 100 µl of blood was collected per mouse and immediately diluted in 900 µl TSB with 10% v/v glycerol to protect recovered bacteria and prevent clotting. Death was then confirmed by cervical dislocation, following which the trachea was exposed and a small incision made for the lung lavage. A narrow tipped Pasteur pipette containing 500 µl TSB with 10% v/v glycerol was inserted into the incision. The lungs were washed and the media recovered. This was repeated twice more to maximise bacterial recovery. After three washes the media was transferred to a sterile 2 ml screw capped tube. Following the lung lavage mice were decapitated and an 18 gauge needle was inserted into the nasal passage. Using 500 µl of TSB with 10% v/v glycerol the media was injected through the nasal passage and collected in a sterile tube as it exited the nostrils. This wash was repeated 3 times. For MF1 mice the brain was also collected by dissection of the skull and the removal of the complete brain into 1 ml TSB with 10% v/v glycerol. All samples were stored at -80ºC as soon as possible after collection.

2.7.4 **Intravenous sepsis model.**

For each SpnIII WT strain five female eight week old CD1 mice (Charles River, UK) were intravenously inoculated with 100 µl of *S. pneumoniae* at $1x10^7$ CFU/ml in TSB media (final challenge $1x10^6$ CFU). Mice were placed in a heated box at 37ºC for 10 minutes to stimulate tail vein dilation. Mice were then anesthetised using 2.5 % (v/v) isoflorane at 1.6-1.8 litre O_2/min in an anaesthetic box. To deliver the challenge dose mice were restrained in a rodent restrainer and

injected via the tail vein. Mice were allowed to recover on their backs before being returned to their cage. Each group was housed individually and had free access to food and water. Mice were monitored for signs for disease and culled by cardiac puncture and cervical dislocation 18 hours after challenge. Cardiac puncture was conducted under anaesthetic using 2.5% v/v isoflurane at 1.6-1.8 litre O_2/min delivered in an anaesthetic box. Approximately 100 µl of blood was collected per mouse and immediately diluted in 900 µl TSB with 10% v/v glycerol to protect recovered bacteria and prevent clotting. Death was then confirmed by cervical dislocation. Mice were then dissected to expose and recover the spleen and brain, both of which were stored in 1 ml TSB with 10% v/v glycerol. All samples were stored at -80ºC as soon as possible after collection.

2.7.5 **Intracisternal meningitis model**

For each SpnIII WT strain eight female eight week old MF1 mice (Harlan Nossan, Monza, Italy) were inoculated with 10 µl of *S. pneumoniae* at $1x10^3$ CFU/ml in TSB media. Mice were anesthetised with 200 µl of Zoletil 100 (15 mg/kg) (Milan, Italy) plus Xilor 2% (4 mg/kg) (Bologna,Italy). The challenge was delivered by injection into the cisterna magna while mice were anesthetised. A final challenge dose of 10 CFU was delivered to each mouse. Mice were allowed to recover on their backs before being returned to their cage. Each group was housed individually and had free access to food and water. Mice were monitored for signs for disease and culled by cervical dislocation 18-20 hours after challenge. Approximately 100 µl of blood was collected using a herapin coated syringe and immediately diluted in 900 µl TSB with 10% v/v glycerol to protect recovered bacteria. Mice were then dissected to expose the spleen and brain and both organs were collected and stored in 1ml TSB with 10% v/v glycerol. All samples were stored at -80ºC as soon as possible after collection.

2.7.6 **Plating of recovered bacteria**

All challenge doses and recovered samples were plated using 10 fold serial dilutions on TSA supplemented with 3% Horse Blood. Before plating all spleen and brain samples were passed through a 40µm cell strainer (Falcon, USA) to remove as much tissue as possible. Alternatively, for the meningitis model (See method 2.7.4) spleens and brains were mechanically homogenised using a flat

sterile syringe plunger. Dilutions were plated using the Miles and Misra technique as previously described (101). TSA plates were incubated at 37ºC, 5% CO_2 for 18-24 hours.

For allele quantification of bacteria recovered from mice 10 µl of the sample was plated without dilution onto TSA with 3% horse blood and incubated at 37ºC, 5% CO_2 for 18-24 hours. This ensured only live bacteria were analysed. Where no colonies were recovered allele quantification was not undertaken. All cells that grew from the plated 10 µl were collected using a sterile plastic loop and resuspended in 100 µl of TSB with 10% v/v glycerol. These cells were analysed and the percentage of each SpnIII variant determined using the allele quantification method previously described.

2.7.7 **Processing of bacteria recovered from the EHPC**

For allele quantification of nasal wash samples received from the Experimental Human Pneumococcal Carriage (EHPC) project 10 µl of the sample was plated without dilution onto TSA with 3% horse blood and incubated at 37ºC, 5% CO_2 for 18-24 hours. This ensured only live bacteria were analysed. Where no colonies were recovered allele quantification was not undertaken. All cells that grew from the plated 10 µl were collected using a sterile plastic loop and resuspended in 100 µl of TSB with 10% v/v glycerol. The recovered cells were then analysed using the allele quantification protocol.

2.8 **Mutant strains**

2.8.1 **Marked mutant strains**

All mutant strains in which a cassette was left *in situ* were created by the transformation of PCR generated fragments into competent cells (method 2.5). Approximately 500bp up and downstream of the gene to be deleted were amplified using primers with 20bp tails complementary to either an *aad9* spectinomycin cassette (*aad9,* Aminoglycoside O-nucleotidylyltransferase), Janus kanamycin cassette (*rpsL*, ribosomal protein S12 and *aphIII,* aminoglycoside O-phosphotransferase) (104) or a chloramphenicol cassette (*cat*, Chloramphenicol acetyl transferase) (105). Re-amplification of the flanking regions with the appropriate cassette produced constructs with an antibiotic

marker suitable for transformation. All primers can be found in table 2.3. Successful transformants were confirmed by PCR and Sanger sequencing.

2.8.2 **Unmarked mutants**

To create unmarked mutants (104), the steps in method 2.8.1 were followed using a Janus kanamycin cassette. The construct created was transformed into a D39 strain with a single point mutation in the *rpsL* gene, which confers streptomycin resistance (A167C). A Janus kanamycin cassette provides kanamycin resistance and streptomycin sensitivity by providing a second WT copy of the *rpsL* gene. A successful transformant containing the Janus cassette was selected using 500µg/ml kanamycin and confirmed by sensitivity to 500µg/ml streptomycin. Two overlapping PCRs of approximately 500 bp were used to generate an additional construct to replace the Janus cassette. Primers for the gene of interest included a 20bp overlap that introduced a single, non-synonymous base-pair change. Re-amplification of the flanking regions produces a construct with a single base-pair change. This second construct was transformed into a strain with a Janus kanamycin cassette at the gene of interest. Successful transformants are selected using 500µg/ml streptomycin and confirmed by sensitivity to 500µg/ml kanamycin. All primers used for mutant construction can be found in table 2.3. Successful transformants were confirmed by PCR and Sanger sequencing.

2.9 **Modelling of SpnIII variant percentages *in vivo***

Variation in the genetic composition of bacteria as a result of division and mutations was modelled based the following equation: $X_{n+1} = P * X_n$,

The column vector X_n describes the percentage of different variants within the population after n divisions (n=0 corresponds to the initial distribution of variants)

P describes the transition matrix showing the mutation rates. The matrix P is a Markov (stochastic) matrix where $\sum_{i=1}^{8} p_{i,j}$=1. This signifies that the given variant will recombine or remain. The meaning of each element $p_{i,j}$ is the probability that the variant ‘j’ will mutate and become ‘i’. The coefficients α_i, β_i, γ_i in the matrix P were found by fitting the experimental data.

Table 2-3 – Primers used in this study

Primer Name	**Sequence**	**Use**
AMRE074L	FAM-GGAAACTGAGATATTTCGTGGTGATGATGGGA	Allele Quantification
AMRE059	CCTGATCGAGCGGAAGAATATTTCTGCCGAGGTTGCC	Allele Quantification
IF100	GCTCTAGAACTAGTGGATC	Spectinomycin *aad9* cassette
IF101	TTCCCTTCAAGAGCGATAC	Spectinomycin *aad9* cassette
MD014	GTTTGATTTTTAATGGATAATGTGATATAATCT	Janus cassette
MD015	GGCCCCTTTCCTTATGCTTTTG	Janus cassette
VM017	TTATAAAAGCCAGTCATTAGGCC	Chloramphenicol *cat* cassette
VM018	ATGAAAATTTGTTTGATTTTTAATGG	Chloramphenicol *cat* cassette
MD018	CACAAAAATAGCAGGTAGTCAG	*creX* mutant (amplification from DP1004)
MD019	CCTCTCCTATTGTTCAGTCAGTTATTG	*creX* mutant (amplification from DP1004)
MD001	GGTGTTAGAATTATACGTGGTGG	*creX* Janus cassette, *creX* Unmarked mutant
MD003	TACTGGAACAAGTTATCCTGCAATCAAT	*creX* Janus cassette, *creX* Unmarked mutant
MD013	TCCATTAAAAATCAAACAACATTAAATAGTACCAGTATCTCCG	*creX* Janus cassette
MD037	CATAAGGAAAGGGGCCCAGACCATTGATTACATTTCTGAGC	*creX* Janus cassette
MD038	GCTTTCTTCAAATGTTAATTCAAAATC	*creX* Unmarked mutant
MD039	GATTTTGAATTAACATTTGAAGAAAGC	*creX* Unmarked mutant
MD040	GTCCAGATGTACTATTCTAGTTTC	*xerS* KO
MD041	CTAGTTCTAGAGCTCATATTCCTTAAGGTATTCGTACAGG	*xerS* KO
MD042	CTTGAAGGGAAGAATGCTCTGGATAGTTTATGATTTTG	*xerS* KO
MD043	CGAAATCAATCGCATCAAGGATAC	*xerS* KO
MD066	GGCTCATCCGGAATAGAGGC	SPD_0921 KO
MD074	GATCCACTAGTTCTAGAGCCCTTCAATAGACTTGCCAGAC	SPD_0921 KO
MD075	GTATCGCTCTTGAAGGGAAGGACAATGATGAACAAGTTGAATG	SPD_0921 KO
MD069	GAATATCCTTCGATCACTCTATC	SPD_0921 KO

MD080	CCTAGGTGGGATGCTGGCAG	*clpL* KO
MD081	GATCCACTAGTTCTAGAGCGCAGGAGTAGACTATCTTTAC	*clpL* KO
MD082	GTATCGCTCTTGAAGGGAAGAAGCAGATATGGAAGATGGC	*clpL* KO
MD073	CTATCGAGCACCTCTTAGCC	*clpL* KO
MD091	CGTGCAGGTATCGGGATGGTTG	*clpP* KO
MD092	GATCCACTAGTTCTAGAGCGGCTTGTTTGTTCAATAACTACAGG	*clpP* KO
MD093	GTATCGCTCTTGAAGGGAACCAGGAAACACTTGAATATGGC	*clpP* KO
MD094	CTGCTCCAGACTTCACTTCTCC	*clpP* KO
MD095	CAGGTAACTCTGGCGGCC	*spoJ* KO
MD096	GATCCACTAGTTCTAGAGCCATGATTCTACACTAACACATC	*spoJ* KO
MD097	GTATCGCTCTTGAAGGGAAGCAAACTATCACAAGCCC ATG	*spoJ* KO
MD098	GATAGCAAGAATTTTCCACAAGC	*spoJ* KO
MD099	GAGGTCAATGGTCGTCTTGTC	SPD_1657 KO
MD105	GATCCACTAGTTCTAGAGCCATCCTTATCAGCTCAAGTCTTAC	SPD_1657 KO
MD106	GTATCGCTCTTGAAGGGAAGGAATTGACAGACAAGCCCTG	SPD_1657 KO
MD102	CAAGGAGTTTAGCCGATGATTG	SPD_1657 KO
MD107	CCCAGGATACGGATGGTATC	*recU* KO F
MD108	GTATCGCTCTTGAAGGGAACAAGAAACCTACTTATTGCCGGC	*recU* KO
MD109	GATCCACTAGTTCTAGAGCTGGATAGTTGACCATAATTCTCC	*recU* KO
MD110	CTGCTGGTAATCTCTTAACTGC	*recU* KO
MD121	GATGGAGAAAGCCGAGTGTG	Glycerol operon (D39, TIGR4, 7001)
MD123	CATTATCCATTAAAAATCAAACGCACCGCAGATTTCCTC	Glycerol operon (D39, TIGR4, 7001)
MD124	CAAAAGCATAAGGAAAGGGGCCCCAGCTGTGACAGTTGGTG	Glycerol operon (7001)
MD126	CAAAAGCATAAGGAAAGGGGCCGCTCAGGTTGGATTGTGATTA	Glycerol operon (D39, TIGR4)
MD127	GCAAGATGCTGTGCATGATACG	Glycerol operon (TIGR4, D39, 7001)
MD128	GAGGCGAATCAAGCGCTGG	*recA*

MD129	GGTGGCTGCTACTGCTTCC	*recA*
MD130	CTCGTCAAAATAGGCTAGTTCC	*comC1*
MD131	CTCAGAAGAATTTAGTAAGCAG	*comC1*
MD132	GCGTTTTATTATGGAAATGTTGG	*dprA*
MD133	CAACATCTGACAAAATAGTCGG	*dprA*

2.10 **The Experimental Human Pneumococcal Carriage project**

The EHPC project was set up as a human model of pneumococcal carriage (106). Healthy volunteers are intranasally challenged with a 10^4-10^5 dose of the carriage associated *S. pneumoniae* serotype 6B strain BHN418 (107). 10ml nasal wash samples were collected from volunteers at 2, 7, 14 and 22/24 days' post challenge and pelleted to a total volume of 800µl. A total of 81 samples, from 24 carriage positive volunteers were received on dry ice and stored at -80°C until ready for processing. For each sample a 200 µl aliquot of the nasal wash was supplied for allele quantification. In addition, an aliquot of the corresponding challenge dose was also provided. Each sample with recoverable bacteria, that was PCR positive for *S. pneumoniae,* was subjected to allele quantification. Quantification was conducted on 10 µl of the pellet grown on agar plates without dilution. All growth was collected and stored in 100 µl TSB with 10% (v/v) glycerol. Challenge doses were analysed using the same method. Quantification was conducted as described in method 2.5.

3 The generation and characterisation of wildtype SpnIII strains

3.1 Introduction

Previous research into the phase variable type I RM system *Spn*III has made use of phase-locked mutants (64,65). These mutants, which are unable to recombine at the *spnIII* locus, have been used for phenotypic analysis of colony morphology, murine models of infection and gene expression by RNAseq (64,65). While the results have been revealing, indicating a link between variable S subunit expression and virulence, they do not allow for any investigation of recombination dynamics. Understanding the differential expression of genes associated with *Spn*III is interesting but it does not help determine environmental conditions and stressors that may select one variant over another. As a result, this work has been undertaken to generate a panel of wildtype strains, each enriched in a single SpnIII variant while maintaining a wildtype locus.

These wildtype strains SpnIIIA-F have been used to determine the recombination rates of colonies founded by cells expressing different SpnIII variants. Gene expression analysis has revealed some differences in metabolic and virulence genes, and growth medium does appear to interfere with recombination dynamics in some variants. The inability to maintain a stock enriched in SpnIIIF in *S. pneumoniae* D39 has been extensively investigated. Finally, the lack of self-restriction following recombination events has been explored, by the generation of several different mutants that may facilitate the process of restriction alleviation.

3.2 Aim

The aim of this chapter was to generate and characterise strains that were predominantly expressing a single SpnIII variant but that were free to recombine. The use of wildtype strains allows the investigation of conditions that may influence recombination and select for one SpnIII variant over another.

3.3 Generating wildtype strains

In order to investigate the recombination dynamics of the *Spn*III RM system individual strains were selected, each enriched in a single SpnIII S subunit. These

strains are freely capable of recombination at the *spnIII* locus, and were generated through the selection and re-passaging of single colonies (Fig 3.1). Analysis of colonies offers an advantage as they are founded by a single cell, and as a result of being a single cell express a single SpnIII variant. As the colony grows recombination will decrease the percentage of the founding SpnIII variant. As a result, the founding SpnIII variant of a single colony will remain the predominant variant. Each passaged colony was stored and the SpnIII variant percentages were determined using the previously described quantification methodology (method 2.5) (Fig 3.1). Stocks that contained >70% of a single *spnIII* variant were considered to be enriched. Stocks of a lower percentage were re-passaged to obtain new single colonies founded by the SpnIII variant of interest. This process was repeated until enriched stocks of all SpnIII variants were generated.

Initial work by Manso *et al.* determined that the wildtype D39 strain held within our lab was predominantly expressing SpnIIIE (64), therefore no passaging was required to generate an enriched SpnIIIE strain. Previous work had also isolated enriched SpnIIIB and SpnIIID strains. These three enriched wildtype strains were used for the passaging and selection of SpnIIIA, SpnIIIC and SpnIIIF enriched strains. Stocks enriched in >70% of a single SpnIII variant were generated for SpnIIIA-E, with minimal passaging events (Table 3.1). A strain enriched in SpnIIIF was more difficult to obtain, and eventually a strain of 51.79% was deemed to be the most enriched strain that could be achieved through serial passaging. In total 177 single colony isolates were passaged in order to generate a partially enriched (>50%) SpnIIIF strain.

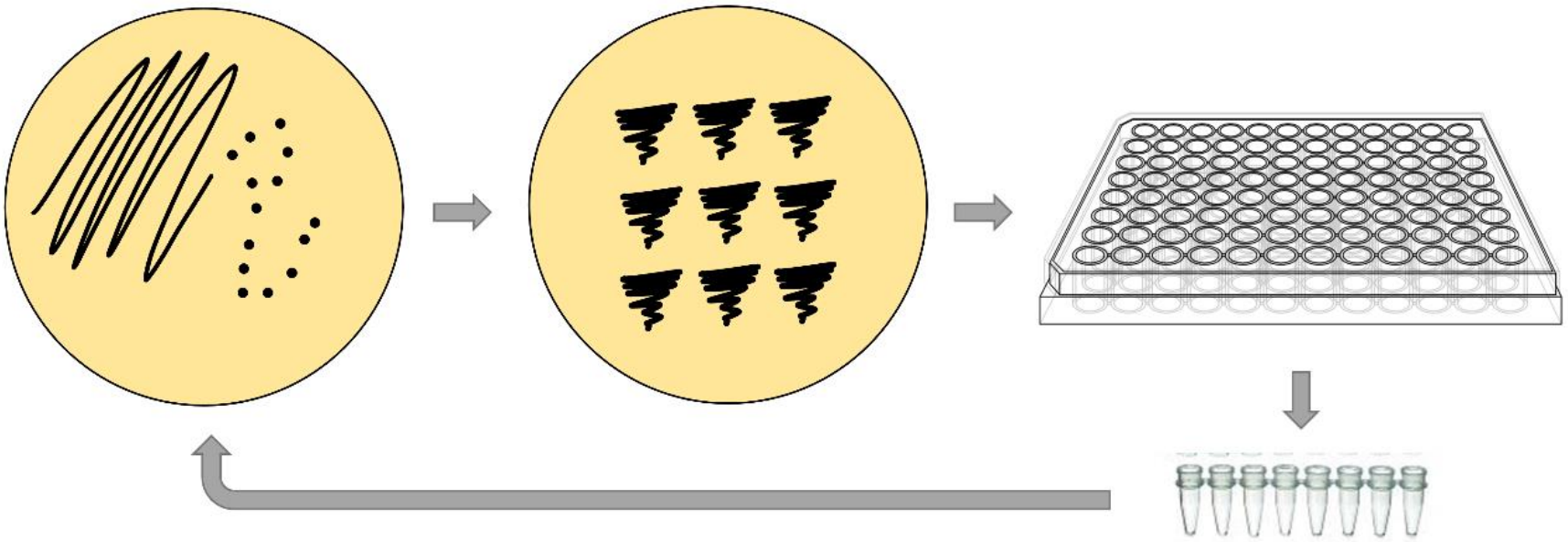


Figure 3.1 – Selection of enriched wildtype SpnIII strains. *In order to generate strains enriched in each SpnIII variant S. pneumoniae D39 was grown on agar plates, single colonies were picked and re-passaged to increase cell density and stored in 10% v/v glycerol at -80°C. PCR based allele quantification was conducted and individual clones with an increased percentage of the desired allele were re-plated and the process repeated until strains enriched in each SpnIII variant were obtained.*

Table 3-1 – initial strain generation, strain descent and % achieved in initial stock generated.

SpnIII variant	Colonies screened	Parent strain	Dominant variant % in selected patch	Dominant variant % in final stock
A	18	SpnIII WT E	91.6	78.5
B	N/A	SpnIII WT E	N/A	73.1
C	22	SpnIII WT D	87.4	70.7
D	N/A	SpnIII WT E	N/A	79.6
E	N/A	SpnIII WT E	N/A	75.4
F	177	SpnIII WT D	88.8	51.8

3.3.1 **12 hour colonies**

In order to use the SpnIII WT strains to investigate conditions affecting recombination it was first necessary to determine if recombination was consistent across all six variants. Each WT strain was grown on blood agar plates for 16 hours (18-20 generations) and the SpnIII variant percentages in single colonies were determined by allele quantification (Fig 3.2). Analysis of single colonies confirmed that recombination is ongoing in all SpnIII WT strains and despite strain enrichment colonies could be founded by non-dominant variants.

Direct comparison of the different enriched strains is challenging unless converted into percentage recombination per repeat. As previously described there are three inverted repeats within the *spnIII* locus, each facilitating a unique recombination event. The outcome of these recombination events is entirely dependent on the *spnIII* variant expressed prior to recombination (Table 3.2). In order to directly compare data from each SpnIII variant, data from all colonies founded by the same SpnIII variant were pooled, regardless of the enriched strain of origin. One colony, from the enriched SpnIIIC strain, was discarded as the near 50:50 split in the colony meant the founder variant could not be determined. Ten

further colonies were re-assigned where their founder SpnIII variant did not match the originator strain. Using the information shown in Table 3.2 the percentage of each SpnIII variant was assigned to a recombination event on either the 333bp, 15bp or 85bp repeat, where more than one recombination event would be required e.g. SpnIIIB → SpnIIIC, this is indicated at >1. Double recombination events cannot be accurately determined as the order of recombination is not certain, however these events typically make up <1% of a colony.

Once recombination events were determined it was possible to compare the median frequency (based on percentage) of recombination on each of the three repeats (Table 3.3). Recombination on the 15bp (β) repeat cannot currently be determined in SpnIIIA and SpnIIID founded cells as it occurs within the silent *hsdS* genes of the locus. The current quantification protocol is limited to the active *hsdS* gene.

Significant differences were observed in the median recombination rates of colonies founded by different SpnIII variants (Table 3.3, Fig. 3.3). SpnIIIA and SpnIIID founded colonies were significantly different from SpnIIIB, C, E and F. These colonies recombined significantly less overall ($P<0.001$ for all comparisons) and this was reflected in lower recombination rates on the 333bp and 85bp repeats (Table 3.4). In addition to the significant differences between SpnIIIA and SpnIIID and other spnIII variants, there are also significant differences in β recombination rates between SpnIIIB/C and SpnIIIE/F colonies (Table 3.4, Fig 3.3). The significant differences in recombination rates appear to group the SpnIII variants by their common TRD2, for example SpnIIIA/D share TRD2.1 (Chapter 1, Fig 1.3).

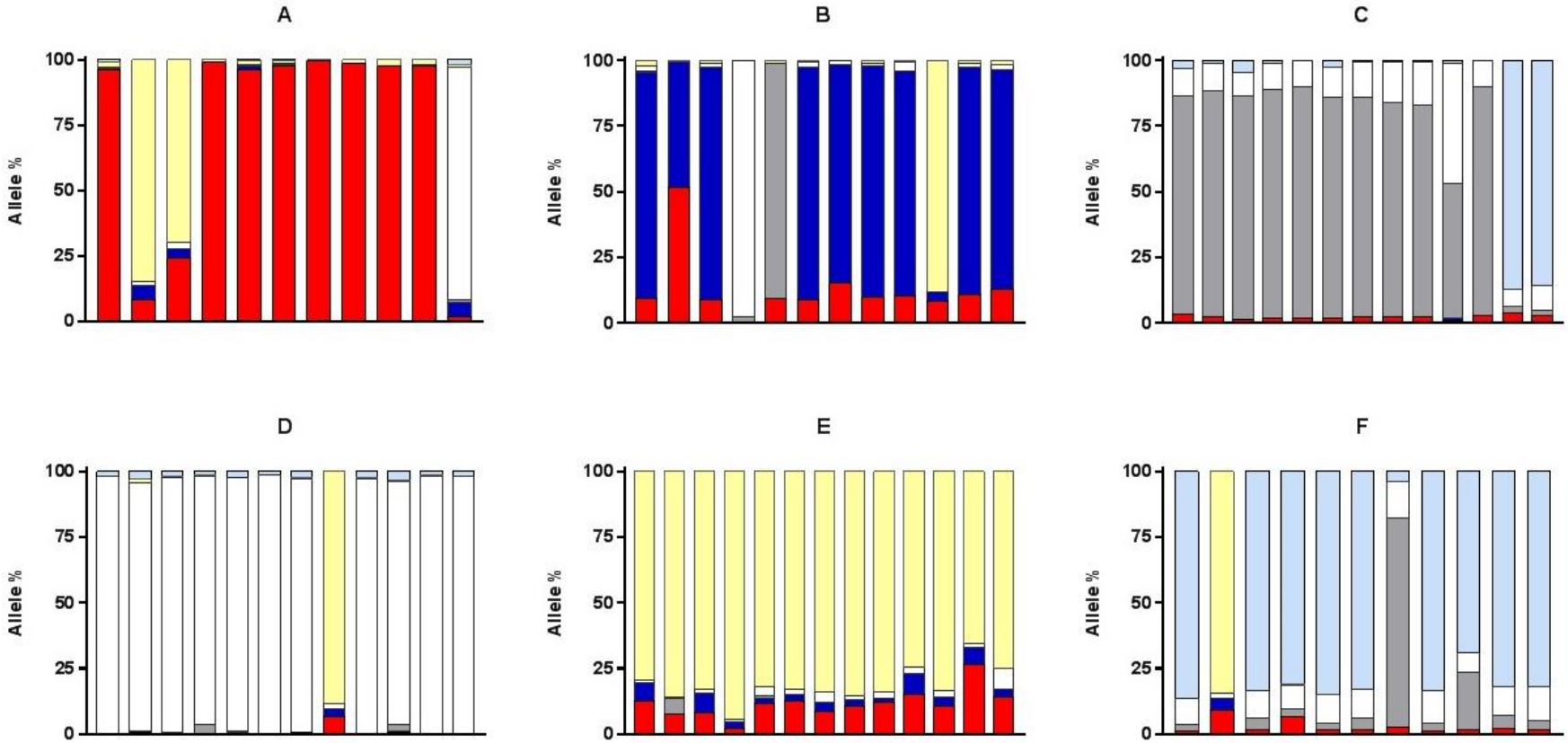


Figure 3.2 – Analysis of single colonies from enriched wildtype SpnIII strains. *Each bar represents a single colony and the colours represent the SpnIII variants detected in that colony (SpnIIIA red, SpnIIIB blue, SpnIIIC grey, SpnIIID white, SpnIIIE yellow and SpnIIIF light blue). Each graph represents the colonies analysed for a single WT strain. Panel A shows the analysis of SpnIII variant percentage in 11 colonies picked form a plate generated from a prevalently SpnIIIA strain (79.5%A, 4.2%B 0.8%%C, 0.3%D, 13.8%E and 2.31%F) see table.3.1 These data show 8/11 colonies yielding mostly the A variant (red), two colonies containing mostly the E variant (yellow), and 1 colony mostly the D variant (white). This is in accordance of the prevalence of SpnIII alleles in the original stock. All colonies show more than one colour. Panels B-F show equivalent data from analysing colonies generated from the strains listed in table 3.1.*

Table 3-2 SpnIII recombination outcomes. *Recombination events at the spnIII locus occur on three independent repeats which are defined by their size; 333bp, 85bp, and 15bp. The founder variant (left hand side of table) influences the outcome of each recombination event i.e. a recombination event on the 333bp repeat will result in SpnIIIA if the founder cell is SpnIIIE, however the same recombination event will result in SpnIIID is the founder cell is SpnIIIF. The breakdown of SpnIII recombination into individual repeat events allows for direct comparison between colonies with different founder variants.*

		SpnIII Variant following recombination event					
		A	**B**	**C**	**D**	**E**	**F**
Founder SpnIII Variant	**A**	-	>1	>1	>1	333bp	85bp
	B	333bp	-	>1	85bp	15bp	>1
	C	85bp	>1	-	333bp	>1	1bp
	D	>1	>1	>1	-	85bp	333bp
	E	333bp	15bp	>1	85bp	-	>1
	F	85bp	>1	15bp	333bp	>1	-

Table 3-3– Median recombination on individual spnIII repeats in each spnIII wildtype variant. *Median percentage calculated using a minimum of 10 individual colonies. Values shown are the frequency of recombination for an individual event, events determined used table 3.2.*

		SpnIII Variant following recombination event					
		A	**B**	**C**	**D**	**E**	**F**
Founder SpnIII Variant	**A**	97.8±1.3	0.0±0.2	0.0±0.2	0.0±0.2	1.6±0.6	0.0±0.3
	B	10.2±2.2	86.0±2.2	0.0±0.2	1.7±0.7	1.0±0.7	0.0±0.2
	C	2.5±2.2	0.0±0.2	84.8±3.1	10.3±4.2	0.0±0.2	0.9±1.7
	D	0.0±0.5	0.0±0.5	0.0±0.5	96.8±2.7	0.4±0.4	1.9±0.9
	E	9.3±3.0	2.8±2.0	0.0±1.1	2.0±1.9	83.9±4.2	0.0±1.1
	F	1.8±1.7	0.0±0.2	3.1±1.1	10.6±1.8	0.0±0.2	83.5±2.1

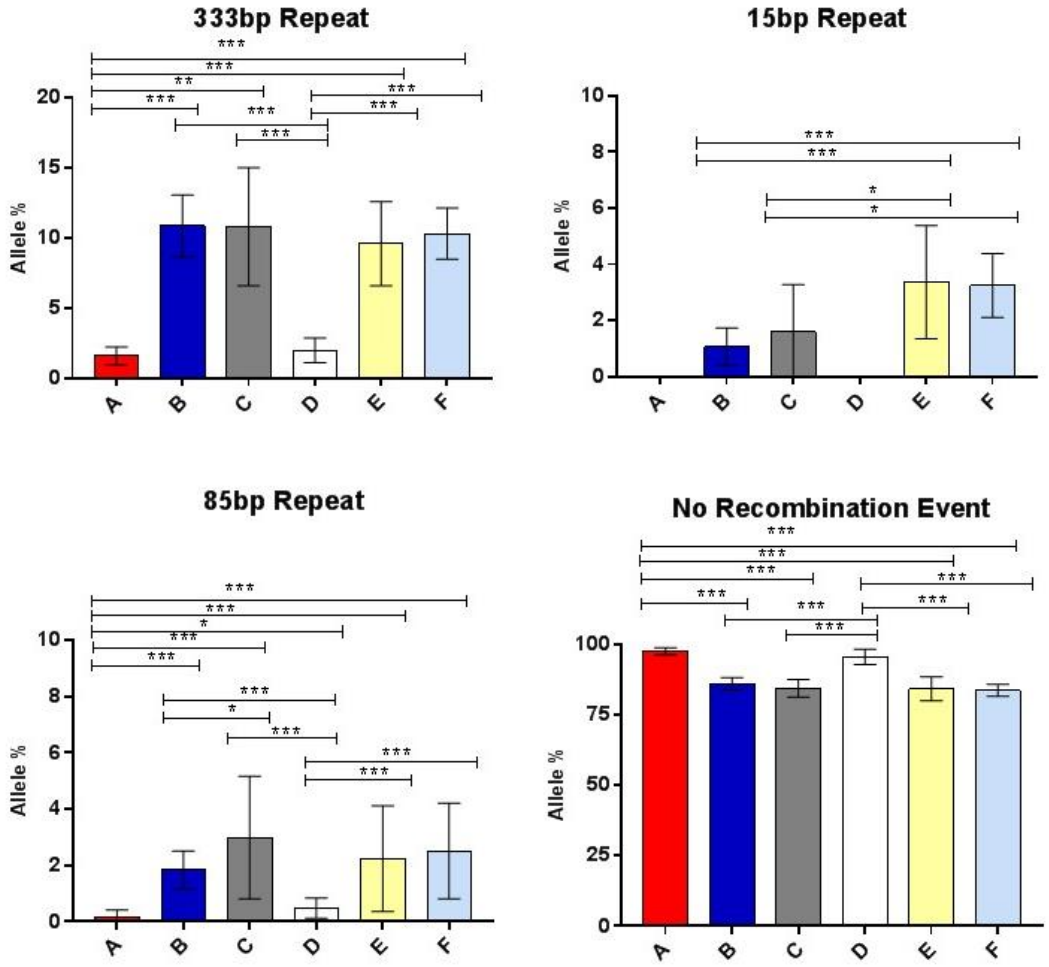


Figure 3.3 – A comparison of individual recombination events in colonies of WT SpnIII strains. *Following analysis of single colonies all those expressing >75% of a single variant were grouped for analysis. Using Fig. 3.3 recombination events were determined and the median per variant per strain was plotted. Significant differences were observed in the recombination rates between colonies predominantly expressing different SpnIII variants. SpnIIIA and D recombined significantly less than SpnIIIB, C, E and F (see table 3.4 for p values). Significant differences were also observed in the recombination rates on the 15bp repeat which were significantly lower in SpnIIIB and C than SpnIIIE and F (see table 3.4 for p values). No 15bp recombination values could be determined for SpnIIIA and D due to a limitation with the quantification method. Significance was tested using a Mann-Whitney test, *p≤0.05, **≤0.01, ***p≤0.001.*

Table 3-4– Significant differences in median recombination rates in colonies founded by different spnIII variants. *Significance tested using a Kruskal-Wallis test.*

		P Value				
Variant 1	Variant 2	No Change	333bp	15bp	85bp	>1 change
A	B	0.049	<0.001	-	0.028	NS
A	C	<0.001	<0.001	-	<0.001	NS
A	D	NS	NS	-	NS	NS
A	E	<0.001	<0.001	-	<0.001	NS
A	F	<0.001	<0.001	-	<0.001	NS
B	D	NS	<0.001	-	NS	<0.01
B	E	NS	NS	0.014	NS	NS
B	F	NS	NS	0.017	NS	NS
C	D	<0.001	<0.001	-	<0.001	<0.001
D	E	<0.001	<0.001	-	<0.001	<0.001
D	F	<0.001	<0.001	-	<0.001	0.001

3.3.2 **Gene expression of WT strains**

RNAseq work in the phase locked SpnIII mutants by Manso *et al.* revealed a number of significant differences in gene expression. The most notable of these were in the SpnIIIB locked mutant which showed reduced expression of the *cps* (capsule) operon, as well as decreased expression of *luxS* and *dexB.* While SpnIIIA showed a >2 fold increase in expression of the DNA chaperone *dnaK* (64). Work is ongoing to investigate if SpnIIIB expressing strains are prone to recombination in the capsule locus which could explain the difference in gene expression observed in the SpnIIIB phase-locked mutant described by Manso *et al.* (64). In order to determine if these changes in gene expression are the result of true epigenetics changes the SpnIII WT strains were analysed by RNAseq.

Table 3-5 – WT B Genes with a >2fold increase in expression (RPKM) when compared alternative WT strains. *Level of expression is shown as a fold change in RPKM. All data shown have been calculated as significantly different by Rockhopper* (102)*.*

Name	Synonym	Product	WT B/ WT C	WT B/ WT D	WT B/ WT E	WT B/ WT F
blpU	SPD_0046	bacteriocin BlpU	0.96	1.76	1.51	2.10
-	SPD_0098	glycosyl transferase, group 2 family protein	1.62	1.62	2.15	1.31
clpL	SPD_0308	ATP-dependent Clp protease, ATP-binding subunit	4.45	1.81	9.80	0.75
-	SPD_0350	conserved hypothetical protein	1.48	1.31	2.85	1.59
-	SPD_0351	sensor histidine kinase, putative	1.44	1.28	2.42	1.28
-	SPD_0352	DNA-binding response regulator	1.56	1.67	3.13	1.92
-	SPD_0353	conserved hypothetical protein	1.38	1.48	3.42	1.81
-	SPD_0355	conserved hypothetical protein	1.25	1.63	2.19	1.63
cbpF	SPD_0357	choline binding protein F	1.25	1.38	2.05	1.47
hrcA	SPD_0458	heat-inducible transcription repressor HrcA	2.24	1.33	4.52	1.78
grpE	SPD_0459	heat shock protein GrpE	2.38	1.15	4.07	1.49
-	sRNA_SN29	-	2.84	1.19	4.72	1.43
dnaK	SPD_0460	chaperone protein DnaK	2.71	1.10	4.03	1.45
-	sRNA_F14	-	2.52	0.91	3.63	1.19
-	sRNA_SN30	-	2.67	1.25	3.92	1.59
dnaJ	SPD_0461	chaperone protein DnaJ	2.41	0.97	3.02	1.38
-	SPD_0537	matrixin family protein	1.47	1.58	2.56	1.94
ciaR	SPD_0701	DNA-binding response regulator CiaR	1.46	1.36	2.46	1.76
ciaH	SPD_0702	sensor histidine kinase CiaH	1.52	1.32	2.32	1.65
-	SPD_0803	conserved hypothetical protein	1.65	1.26	3.55	1.33

-	sRNA_F33	-	1.63	1.19	2.93	1.52
-	SPD_0913	conserved hypothetical protein	1.64	0.93	2.21	1.37
-	SPD_1506	acetyl xylan esterase, putative	1.84	1.35	2.99	1.47
-	sRNA_R06	-	1.07	1.04	1.64	2.06
groL	SPD_1709	chaperonin GroEL	1.55	0.94	2.15	1.10
groES	SPD_1710	chaperonin, 10 kDa	1.45	0.91	2.02	1.08
-	SPD_1865	alcohol dehydrogenase, zinc-containing	1.14	0.98	2.19	0.80
malP	SPD_1932	maltodextrin phosphorylase	1.91	1.39	2.16	1.70
malQ	SPD_1933	4-alpha-glucanotransferase	1.71	1.59	2.23	1.64
-	SPD_2068	serine protease	1.56	1.25	2.58	1.64
-	SPD_2069	SpoJ protein	1.59	1.27	2.73	1.51

Table 3-6 Genes with a >2fold increase in expression (RPKM) when compared to WT-B *Level of expression is shown as a fold change in RPKM. All data shown have been calculated as significantly different by Rockhopper (102).*

Name	Synonym	Product	WT C/ WT B	WT D/ WT B	WT E/ WT B	WT F/ WT B
-	SPD_0084	IS630-Spn1, transposase Orf1	1.04	2.00	2.63	3.00
-	SPD_0093	membrane protein, putative	1.60	1.47	2.52	1.62
-	SPD_0094	conserved hypothetical protein	1.52	1.35	2.39	1.51
-	SPD_0095	conserved hypothetical protein	1.45	1.29	2.06	1.50
sdhB	SPD_0103	L-serine dehydratase, iron-sulfur-dependent, beta subunit	1.39	2.06	1.67	2.11
-	SPD_0104	LysM domain protein	2.24	2.07	0.88	1.49
-	SPD_0262	PTS system, mannose/fructose/sorbose family protein, IID component	1.55	1.07	2.00	1.83
manL	SPD_0264	PTS system, mannose-specific IIAB components	1.64	1.13	2.04	2.03
creX	SPD_0452	integrase/recombinase, phage integrase family protein	1.15	2.92	1.00	1.23
-	SPD_0463	Hit-like protein involved in cell-cycle regulation, putative	1.11	2.06	1.83	1.77
-	sRNA_F21	-	1.13	2.56	4.63	2.81
-	SPD_0716	IS630-Spn1, transposase Orf1	1.33	2.08	2.25	2.50
rpsT	SPD_0732	ribosomal protein S20	1.03	1.55	2.11	1.44
-	SPD_0783	type I restriction-modification system, S subunit, putative	1.19	2.00	1.81	1.95
-	SPD_0785	conserved hypothetical protein	1.24	2.18	1.88	1.88
rplT	SPD_0849	ribosomal protein L20	1.15	1.46	2.01	1.80
-	SPD_0915	iron-compound ABC transporter, iron compound-binding protein	1.23	1.70	2.09	2.15
map	SPD_0970	methionine aminopeptidase, type I	2.10	0.95	1.04	0.96
pcrA	SPD_0973	ATP-dependent DNA helicase PcrA	2.09	0.92	0.98	1.02
-	SPD_0979	aminotransferase, class V	2.01	1.14	1.11	1.22

prs2	SPD_0980	ribose-phosphate pyrophosphokinase	2.00	1.32	1.42	1.40
-	SPD_0982	pyrophosphokinase family protein	2.00	0.97	1.05	1.10
-	SPD_0983	inorganic polyphosphate/ATP-NAD kinase, putative	2.13	0.97	1.14	1.08
pta	SPD_0985	phosphate acetyltransferase	2.12	1.04	1.31	1.23
nrdH	SPD_1041	glutaredoxin-like protein NrdH	1.35	1.71	2.25	1.80
-	sRNA_srn317	-	1.33	2.10	2.33	1.83
-	sRNA_srn395	-	0.80	1.40	3.75	1.93
ply	SPD_1726	pneumolysin	1.05	1.34	2.00	0.99
-	SPD_1729	conserved hypothetical protein	1.08	1.71	2.52	1.19
-	SPD_1864	conserved hypothetical protein	1.43	2.04	2.24	1.98
-	SPD_1874	LysM domain protein	2.96	3.58	0.87	3.02

The SpnIIIA WT strain was removed from the final data set due to the discovery of a SNP in the SPD_0818 gene. The presence of this SNP resulted in a significant change in the expression of genes associated with cysteine metabolism and therefore skewed the dataset. The remaining SpnIII WT strains were analysed by comparison to the WT B strain which showed significant differences in animal models of infection.

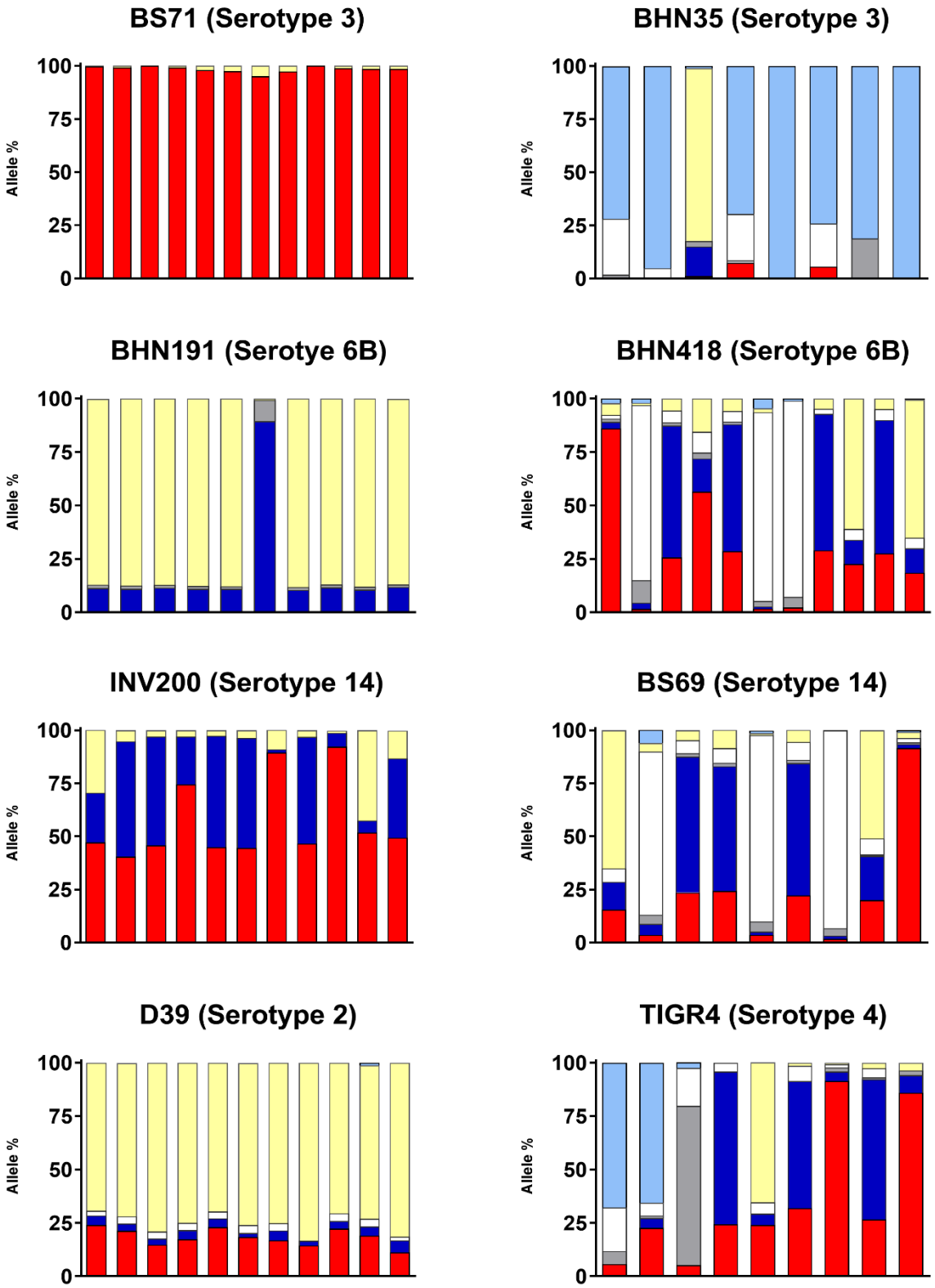
Using a >2 fold increase when compared to the SpnIII WT B strain as a cut off, a range of genes were found to be differentially expressed (Table 3.5 for genes up-regulated in the B strain and Table 3.6 for genes up-regulated in the non-B strains). *dnaK* and *dnaJ* expression was up-regulated in SpnIIIB when compared to SpnIIIE and SpnIIIC. SpnIIIE has a number of PTS and sugar transporter genes up-regulated when compared to SpnIIIB, along with the virulence factor pneumolysin (ply). The differential expression of the pneumolysin operon could have an important impact on overall virulence. The expression of the *spnIII* recombinase, *creX*, is 2.9 fold greater in SpnIIID when compared to SpnIIIB. Two heat shock and heat inducible genes, *grpE* and *hrcA,* were up-regulated in SpnIIIB when compared to the SpnIIIC and SpnIIIE WT strains. However, the greatest increase in expression was seen in the ATP binding subunit of the Clp protease *clpL.*

Table 3-7 – Summary of most notable changes in gene expression.

Name	Synonym	Product	Up-regulated Strain	Comparison strain	Fold change
clpL	SPD_0308	ATP-dependent Clp protease, ATP-binding subunit	B	E, C	9.8, 4.5
creX	SPD_0452	integrase/recombinase, phage integrase family protein	D	B	2.9
grpE	SPD_0459	heat shock protein GrpE	B	E, C	4.1, 2.4
dnaK	SPD_0460	chaperone protein DnaK	B	E, C	4.0, 2.7
dnaJ	SPD_0461	chaperone protein DnaJ	B	E, C	3.0, 2.4
ply	SPD_1726	pneumolysin	E	B	2.0
-	SPD_2069	SpoJ protein	B	E	2.7

Combined these data indicate that there are ongoing global regulatory changes associated with the methylation of different recognition sequences by the RM system *Spn*III. There are likely to be an even greater number of changes, some of which are masked by the use of WT strains that contain multiple SpnIII variants.

3.3.3 *SpnIII* in other strain backgrounds



8

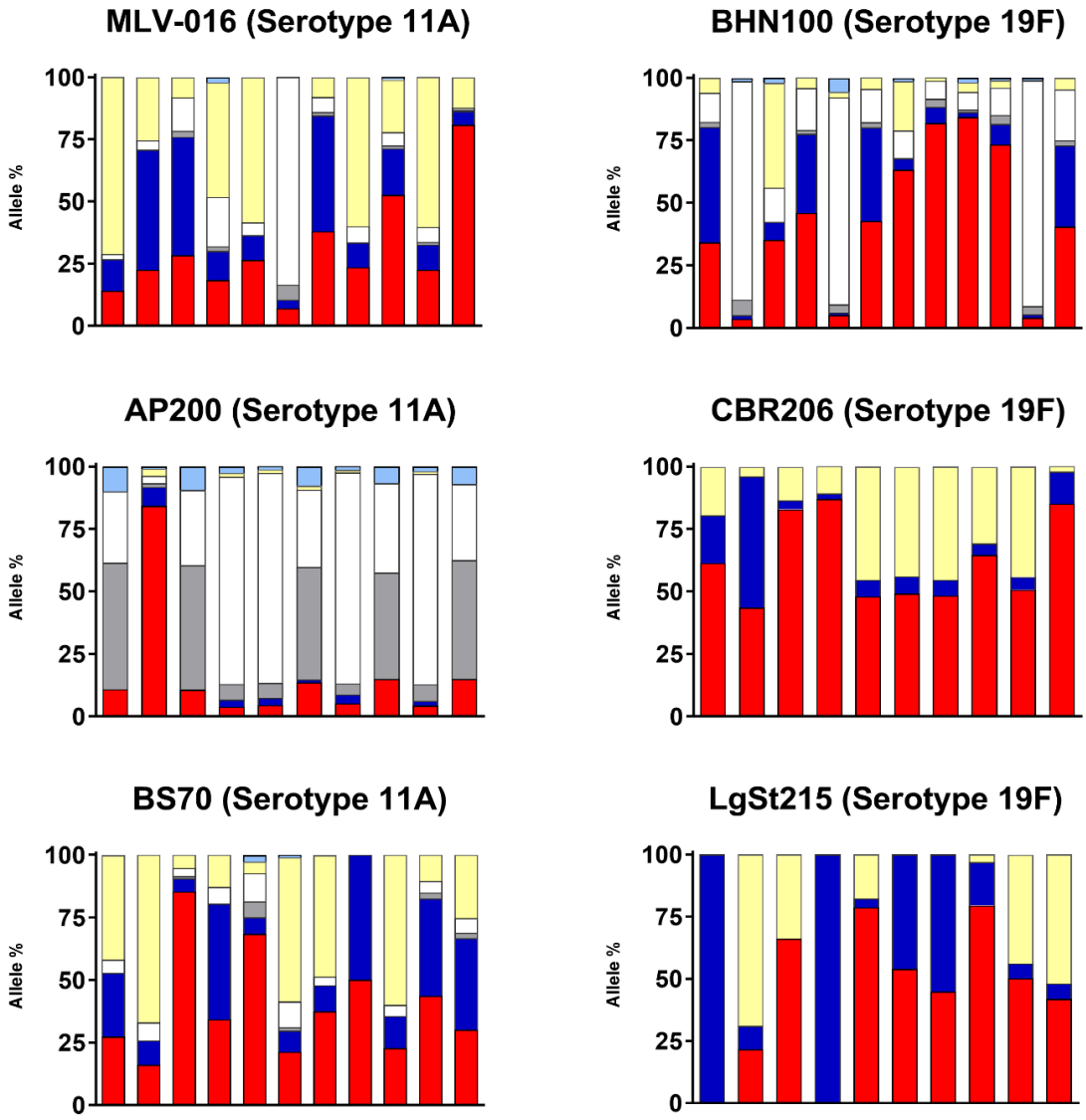


Figure 3.5 – Analysis of single colonies for SpnIII variant proportions in different pneumococcal strains. *Fourteen different strains from seven different serotypes (table 3.6) were analysed using single colonies to determine if SpnIII recombination was consistent across different serotypes and pneumococcal lineages. Each bar represents a single colony and the colours represent the SpnIII variants detected in that colony (SpnIIIA red, SpnIIIB blue, SpnIIIC grey, SpnIIID white, SpnIIIE yellow and SpnIIIF light blue). Each graph represents the colonies analysed for an individual pneumococcal strain with strain name shown above with serotypes indicated in brackets. Recombination is highly variable across strains and serotypes.*

The work of this thesis focuses on recombination of the *spnIII* locus in the serotype 2 strain D39, however the system is part of the core pneumococcal genome (78). In order to determine if recombination at the locus was consistent across different serotypes and lineages single colonies from fourteen strains (Table 3.8) were analysed using the quantification protocol (Fig 3.5). Significant diversity in dominant variants and recombination rates were observed. The highly virulent serotype 3 strain BS71 (108) almost exclusively expressed SpnIIIA, while another serotype 3 strain, BHN35 showed 7/8 colonies were founded by SpnIIIF. While recombination rates were highly variable, recombination events on more than one repeat were still rare. There is no conservation of dominant SpnIII variants within serotypes, however the quantification of a single stock of these strains is unlikely to represent the strain as a whole. As shown by the generation of enriched WT strains in D39, serial passaging within the lab can bottleneck populations.

Table 3-8 – Pneumococcal strains used in this study.

Strain	Serotype	Sequence Type	Reference	Strain origin
D39	2	ST595	(109)	O Avery, The Rockefeller University, New York
SP3-BS71	3	ST180	(110)	FZ Hu, Allegheny-Singer Research Institute, Pittsburgh, Pennsylvania
BHN35	3	ST180	(111)	B Henriques-Normark, Karolinska Institutet, Stockholm
TIGR4	4	ST205	(71)	H Tettelin, University of Maryland, Baltimore, Maryland
BHN191	6B	ST138	(107)	B Henriques-Normark, Karolinska Institutet, Stockholm
BHN418	6B	ST138	(107)	B Henriques-Normark, Karolinska Institutet, Stockholm
MLV-016	11A	ST62	(112)	H Tettelin, University of Maryland, Baltimore, Maryland
AP200	11A	ST62	(113)	A Pantosti, Rome, Italy
SP11-BS70	11A	ST62	(110)	FZ Hu, Allegheny-Singer Research Institute, Pittsburgh, Pennsylvania

INV200	14	ST9	(2)	S Bentley, Sanger Institute, Wellcome Trust, Hinxton
SP14-BS69	14	ST124	(110)	FZ Hu, Allegheny-Singer Research Institute, Pittsburgh, Pennsylvania
BHN100	19F	ST162	(114)	B Henriques-Normark, Karolinska Institutet, Stockholm
CBR206	19F	ST179	(108)	H de Lencastre, Universidade Nova de Lisboa, Oeiras
LgtSt215	19F	ST179	(108)	H de Lencastre, Universidade Nova de Lisboa, Oeiras

3.4 *In vitro* phenotypes of wildtype strains

Due to the proposed epigenetic effects of methylation it was hypothesised that strains predominantly expressing different SpnIII variants could be selected using a variety of environmental conditions. As strains are capable of recombination, selection should also drive recombination at the *spnIII* locus *in* favourable directions.

Wildtype strains were grown in a variety of sugars and their transformation efficiency was compared using their ability to incorporate a single point mutation as a marker.

3.4.1 Growth of wildtype strains in different sugars

SpnIII WT strains were grown in CAT media supplemented with either 0.2% glucose, 0.2% galactose, 0.2% N-Acetyl-D-mannosamine (ManNAc) or 0.2% cellobiose, to determine if differences in gene expression may be reflected in differential growth on different sugars (Fig 3.6). Strains produced similar curves suggesting each was capable of using the sugars present. Growth on galactose, ManNAc and cellobiose was always proceeded by a period of growth on the glucose found in the 0.05% yeast extract within the CAT media. This diauxic growth was accounted for when generation times was calculated.

Using the exponential growth rate on each sugar the generation time in minutes was calculated (Table 3.9). No significant differences in generation time were

observed when the WT strains were grown on 0.2% glucose, however when grown on other sugars several significant differences were noted (Table 3.10). The SpnIIIB WT grew significantly faster in 0.2% galactose than both the SpnIIIC ($p<0.01$) and the SpnIIIF WT's ($p=0.01$). The same trend was observed in the SpnIIID strain (SpnIIIC, $p<0.01$, SpnIIIF, $p=0.01$). SpnIIIB did not display the same advantage in 0.2% ManNAc or 0.2% cellobiose. In ManNAc SpnIIIB grew significantly more slowly than SpnIIIA ($p=0.02$), SpnIIIE ($p<0.01$) and SpnIIIF ($p=0.01$). While in cellobiose growth was significantly slower than SpnIIIA (0.02), SpnIIID (<0.01), SpnIIIE (<0.01) and SpnIIIF ($p<0.01$). SpnIIIA also grew significantly more slowly than SpnIIIF ($p=0.01$).

Table 3-9 – Generation time of SpnIII WTs in different sugars

	Generation time (mins)^			
	0.2% Glucose	0.2% Galactose	0.2% ManNAc	0.2% Cellobiose
SpnIIIA	36±5	34±4	189±15	80±5
SpnIIIB	35±5	33±2	436±72	94±0.5
SpnIIIC	30±5	43±1	178±51	66±16
SpnIIID	32±3	36±1	217±89	64±4
SpnIIIE	40±1	42±11	153±16	70±5
SpnIIIF	36±7	45±1	188±18	63±1

^Mean of three replicas

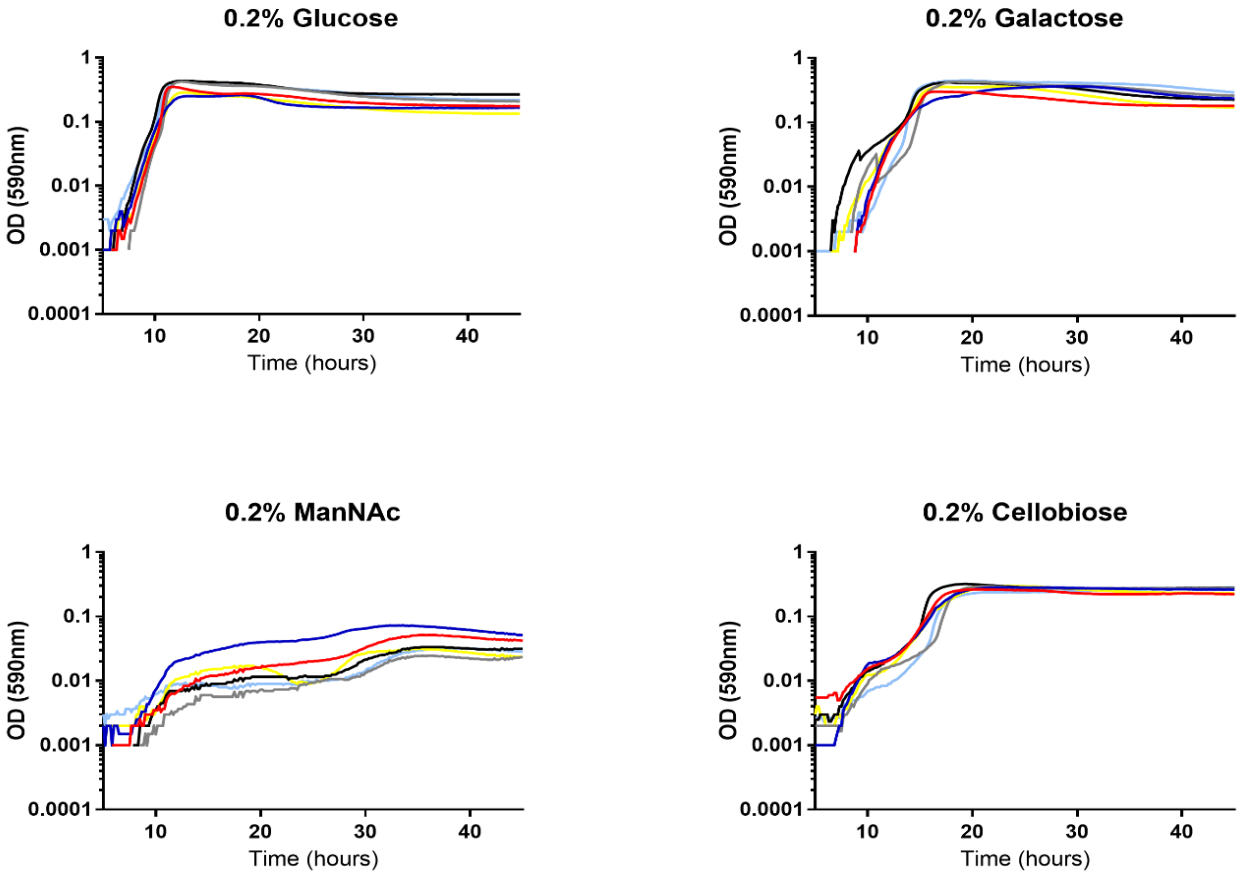


Figure 3.6 – Growth of SpnIII wildtype strains in different sugars. *Each SpnIII WT strain was grown in glucose, galactose, ManNAc and cellobiose. For each graph a single line represents the mean of three replicates of each WT strain (SpnIIIA red, SpnIIIB blue, SpnIIIC grey, SpnIIID white, SpnIIIE yellow and SpnIIIF light blue). No significant differences were observed in growth rates between the strains in glucose, however several were observed in other sugars. SpnIIIB in ManNAc, had a faster growth rate on the small amount of glucose found in the media (8-10hours), but a significantly slower rate of growth on ManNAc once the glucose supply was exhausted (25-30hours). See table 3.9 for generation times and table 3.10 for significant differences in generation time.*

Table 3-10 – Tests of significance in the growth rates of SpnIII WT strains using a Students T test. Significant results shown in bold. *Data not shown for variants with no significant differences in growth rate.*

		P Value^		
Variant 1	Variant 2	0.2% Galactose	0.2% ManNAc	0.2% Cellobiose
A	B	NS	**0.02**	**0.02**
A	F	NS	NS	**0.01**
B	C	**0.01**	NS	NS
B	D	NS	NS	**<0.01**
B	E	NS	**0.01**	**<0.01**
B	F	**0.01**	**0.01**	**<0.01**
C	D	**<0.01**	NS	NS
D	F	**0.01**	NS	NS

^Mean of three replicas

3.4.2 **Transformation efficiency**

Epigenetic differences are regulated by the methylation of specific sites (115). Using the recognition sites of each SpnIII variant, the number of recognition sites in the *comCDE* operon was determined. It was noted that there were seven SpnIIIA sites within the operon but no SpnIIID sites (Table 3.11). As methylation may impact expression of the operon it was hypothesised that there may be differences in the transformation efficiency of strains expressing different SpnIII variants.

This experiment was designed to test transformation efficiency not differences in the natural expression of the competence operon, therefore cells were treated with synthetic CSP1 to induce competence. The standard transformation protocol (see Materials and Methods) was used and cells were transformed with a PCR product that conferred streptomycin resistance via a single point mutant in the *rpsL* gene (A167C). By plating cells with and without 500 µg/ml of streptomycin it was possible to determine total cells numbers and the number of successful

transformants (Fig 3.7A). The mean transformation efficiencies of each WT strain were compared (Fig 3.7B) and no significant differences were observed despite the differences in the number of methylation sites within the *comCDE* operon for each SpnIII variant.

Table 3-11 – SpnIII sites in and around the comCDE operon

SpnIII Variant	Sites in the *comCDE* operon	Comment
A	7	-
B	4	-
C	2 (1)	1 site is located downstream of *comCDE*
D	0	No SpnIIID sites past SPD_2046
E	2	-
F	2	-

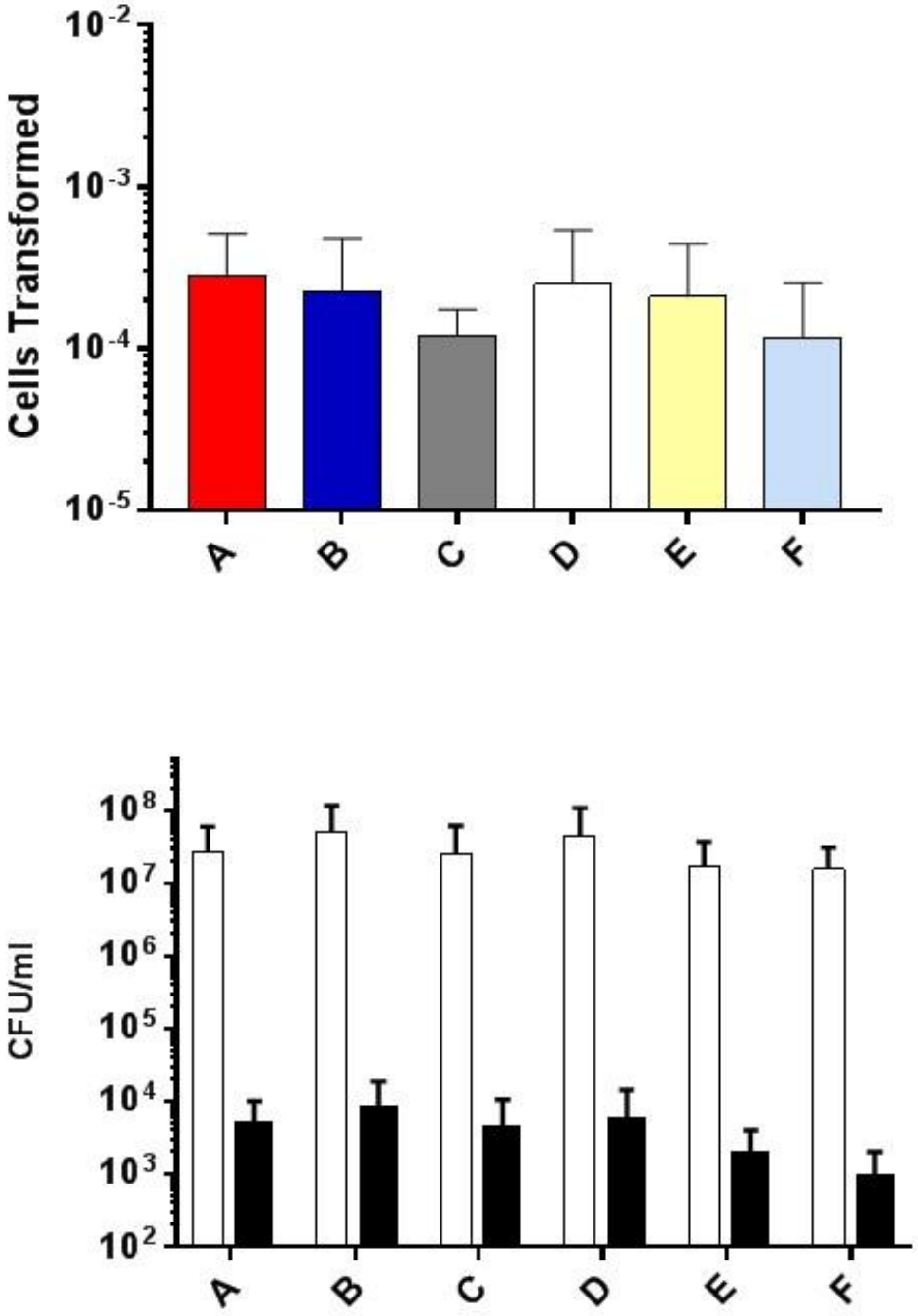


Figure 3.7 – Transformation efficiency of SpnIII WT strains. *Following the induction of competence, SpnIII WT strains were transformed with a single point mutation conferring Streptomycin resistance. Transformed cells were plated with and without 500µg/ml Streptomycin to determine total cell numbers (white bars) and the number of transformed cells (black bars) (Panel A). The transformation efficiency was determined by the proportion of cells transformed compared to the total number of cells (Panel B). No significant differences were observed in the transformation efficiency of the different SpnIII WT strains. Significance was determined using a Student's T test, without assuming a consistent standard deviation.*

3.5 **Modelling of SpnIII WT strains**

In collaboration with Andrew Morozov from the Department of Mathematics, University of Leicester, a model was created that would predict SpnIII recombination in the absence of selection. The model assumes stochastic, or random, recombination accounts for all SpnIII variation and it can therefore be used to determine when outcomes are significantly different from those expected by random switching. The model is based on data obtained from single colonies grown on blood agar plates for a known number of generations (Fig 3.2).

In order to generate the model, data was provided from colonies known to be founded by a single SpnIII variant. These data assumed a 100% starting point and the rate of recombination was taken to be the increase in the percentage of non-founder SpnIII variants. Using the recombination events shown in Table 3.2 a set of parameters were devised that fit the experimental data.

The constructed model is a Markov chain model using a transition matrix which accounts for the frequency of recombination. The elements of the transition matrix were expressed in terms of the probabilities of DNA inversions on three distinct inverted repeats. The parameter values were obtained by fitting modelling results to various sets of experimental data with different initial distributions of genetic variants in the population.

Revision of this model is ongoing, version one was based on three parameters, assuming that recombination rates were consistent across all six SpnIII variants. Subsequent work with the model revealed this was not the case, proving it is capable of detecting deviations from the expected outcome. The current version in development, based on the data found in Table 3.3, includes nine parameters relating to recombination on three repeats in each of the associated pairs (those without significant differences in recombination on any repeat). The model has been used for analysis of samples from the Experimental Human Pneumococcal Carriage project, the results of which can be found in chapter 5. The complete model can be found in appendix 1.

3.6 Investigations into the lower percentage of *SpnIIIF*

Despite no significant differences in the recombination rates in colonies founded by SpnIIIE and SpnIIIF expressing cells (Table 3.3) the generation of an enriched SpnIIIF strain with more than 50-60% was not possible. In the screening of >170 single colony isolates from stocks predominantly expressing spnIIID only three high percentage SpnIIIF were identified following selection and patching. Based on this observation, it was hypothesised that part of the selection method used may be detrimental to cells expressing SpnIIIF.

3.6.1 Whole genome sequencing

Three independent SpnIIIF stocks were generated from the three high percentage SpnIIIF single colony isolates initially identified (Table 3.1). Two of these strains behaved as previous SpnIIIF strains and following culture in liquid media decreased to an SpnIIIF percentage of 52.5% and 53.7% respectively. The third strain maintained an increased SpnIIIF percentage of 63.1%. In order to determine if the increased percentage of SpnIIIF was the result of a mutation all three strains, along with an SpnIIIE strain, were sent for whole genome sequencing.

Analysis of SNPs and INDELs within the three strains revealed a single SNP in the 63% SpnIIIF strain when compared to the two 52-53% SpnIIIF strains. This SNP was located in one of the 333bp α repeats of the *spnIII* locus. A mutation in the locus could explain the increased SpnIIIF percentage as it may reduce the specificity with which a site specific recombinase binds to the locus to facilitate inversions. No other SNPs or INDELS were observed, either between the SpnIIIF strains or when compared to the SpnIIIE strain. The decreased percentage of SpnIIIF is therefore not the result of a SNP or other mutation in a single colony isolate.

3.6.2 SpnIII variant percentages in single colonies compared to patches

The selection of single colony isolates to generate WT SpnIII strains involves the patching of cells (Fig 3.1) to increase total cell numbers. As cells grown for 12 hours on agar plates do not show significantly different recombination rates

between SpnIIIE and SpnIIIF founded cells (Table 3.3) additional analysis of the percentage of each predominant variant within a patch was conducted. Colonies founded by SpnIIIF cells did not contain significantly less of their dominant variant when analysed at 12 hours (Fig 3.8A), however once cells were patched the percentage of SpnIIIF was significantly lower than in all other SpnIII variants (Fig 3.8B). A comparison of the dominant variant percentage in 12 hour colonies and patches also showed significant differences (Fig 3.8C). SpnIIIF patches contained a median percentage of 38.90±15.28, while SpnIIIF 12 hour colonies contained a median percentage of 82.86±4.98 ($p<0.0001$). Significant differences were also found between 12 hour colonies and patches for *SpnIIIA, D* and *E,* $p=0.036$, <0.001 and 0.021 respectively. While SpnIIIA and D saw significant decreases in patches compared to colonies, SpnIIIE showed an increased in patches. Despite the observation of significant differences in other SpnIII variants the greatest decrease was between SpnIIIF colonies and patches (Fig 3.8C). It is not known if this result is due to greater cell density or increased incubation time of patches, however it demonstrates that the mechanism for selecting enriched WT strains may be detrimental to cells expressing SpnIIIF, although the exact reason why is it not yet known.

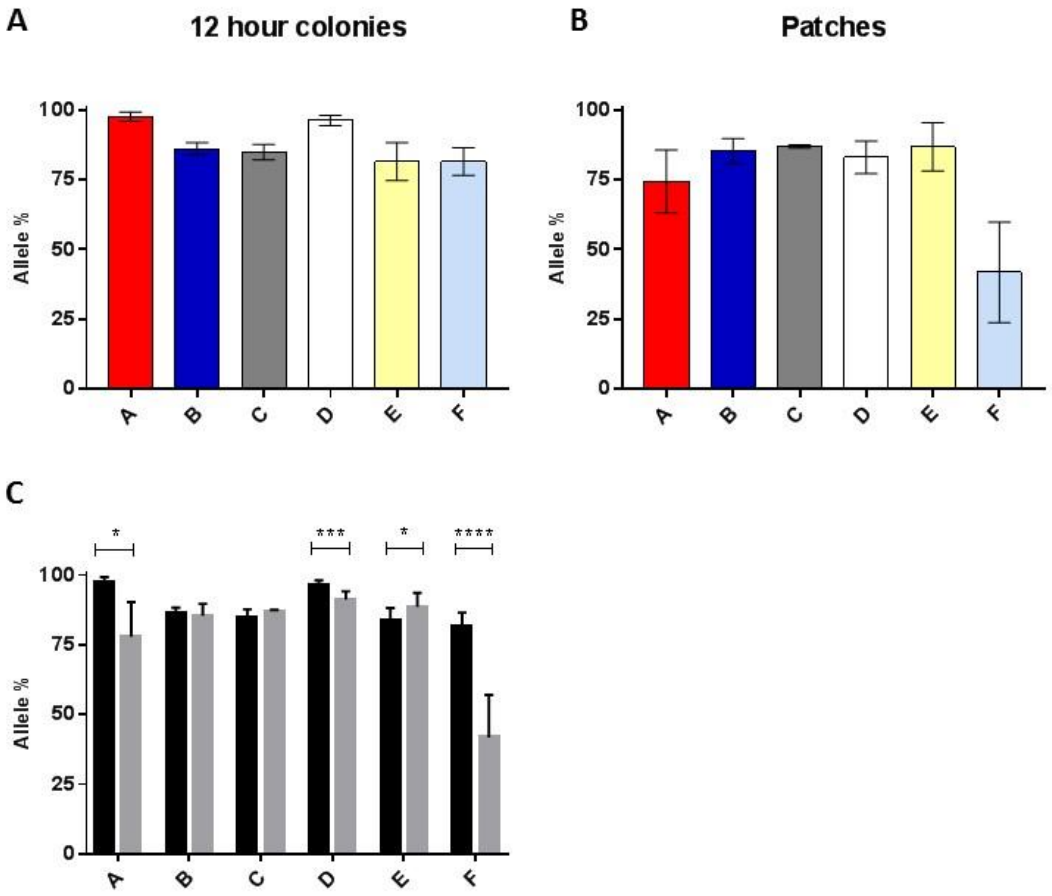


Figure 3.8 – Differences in SpnIII proportions in single colonies grown on solid media for different time periods. *The SpnIII WT strains were grown on agar plates for 12 hours before being collected for PCR quantification and re-passaged (indicated as patches, panel B) as shown in Fig 3.1. The proportion of the dominant allele in single colonies at 12 hours was significantly greater in SpnIIIA and SpnIIID colonies (Panel A) (as previously shown in Table 3.4). This difference was lost when colonies were passaged (Panel B). The median percentage of SpnIIIF dropped significantly in patches when compared to single colonies grown for 12 hours (P<0.0001) (Panel C). The decrease in SpnIIIA and SpnIIID in patches compared to 12 hour colonies was also significant, p=0.036 and p<0.001 respectively (Panel C). A small but significant increase in the median percentage of SpnIIIE in patches compared to 12 hour colonies was also observed (p=0.021) (Panel C). Significance was tested using a Mann-Whitney test, *p≤0.05, **≤0.01, ***p≤0.001, ****p≤0.0001.*

3.6.3 **The Glycerol operon of SpnIIIF**

The pattern observed in D39, that strains cannot maintain a high percentage of SpnIIIF over time, does not appear to be a strain specific phenomenon. Analysis of 79 samples from children with acute otitis media (AOM) received from Markus Hilty at the Institute of Infectious Disease, University of Bern revealed that the only 4/79 clinical isolates showed a high percentage of SpnIIIF variants and importantly that all three serotype 15B/C strains of the collection had high SpnIIIF along with one type 3 strain (see BHN35 above).

A comparison of three genome sequences from serotype 15B/C pneumococcal strains to the genome sequence of D39 revealed that the number of F recognition sites is highly variable between the different serotypes. Site differences were filtered by core genome and number of sites within an operon. This identified clear differences in the number of SpnIIIF sites in the glycerol operon. There are three conserved SpnIIIF sites in the operon of both genomes (Fig 3.9) however, serotype 15B/C strains, represented by the sequence from strain 7001 (GenBank assembly accession: GCA_001148865.2) in Fig 3.9, have an additional three SpnIIIF recognition sites scattered throughout the operon.

In order to determine if the ability to maintain a high proportion of SpnIIIF was a general feature of serotype 15B/C strains single colonies were analysed from two different 15B/C strains. Both serotype 15B/C strains displayed colonies with high SpnIIIF percentages (Fig 3.10). In addition, analysis of frozen cultures demonstrated that this high percentage is maintained when cells are subsequently cultured and is not just a feature of single colonies.

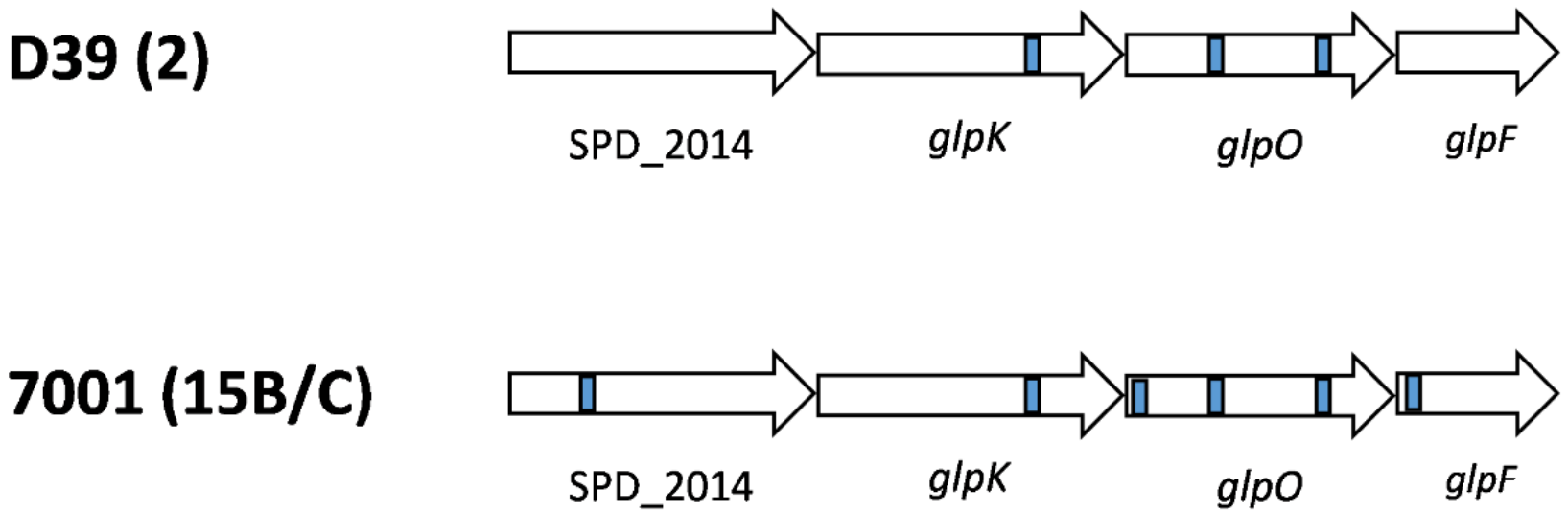


Figure 3.9 – SpnIIIF recognition sites in the glycerol operon of S. pneumoniae D39 and S. pneumoniae 7001. *Previous analysis indicated a significantly greater percentage of SpnIIIF in serotype 15B/C strains when compared to the serotype 2 laboratory strain D39. A comparison of F sites in the D39 genome against three serotype15B/C strains indicated a difference in the number of SpnIIIF recognition sites in the glycerol operon. The different SpnIIIF sites in the glycerol operons of D39 and 7001 (GenBank Accession: GCA_001148865.2). are shown in blue. The three D39 sites appear to be conserved in 7001, which has an additional 3 sites across the operon. Gene numbers shown relate to S. pneumoniae D39. Not drawn to scale. SpnIIIF recognition site: 5'-CAC-N_7-CTT-3'.*

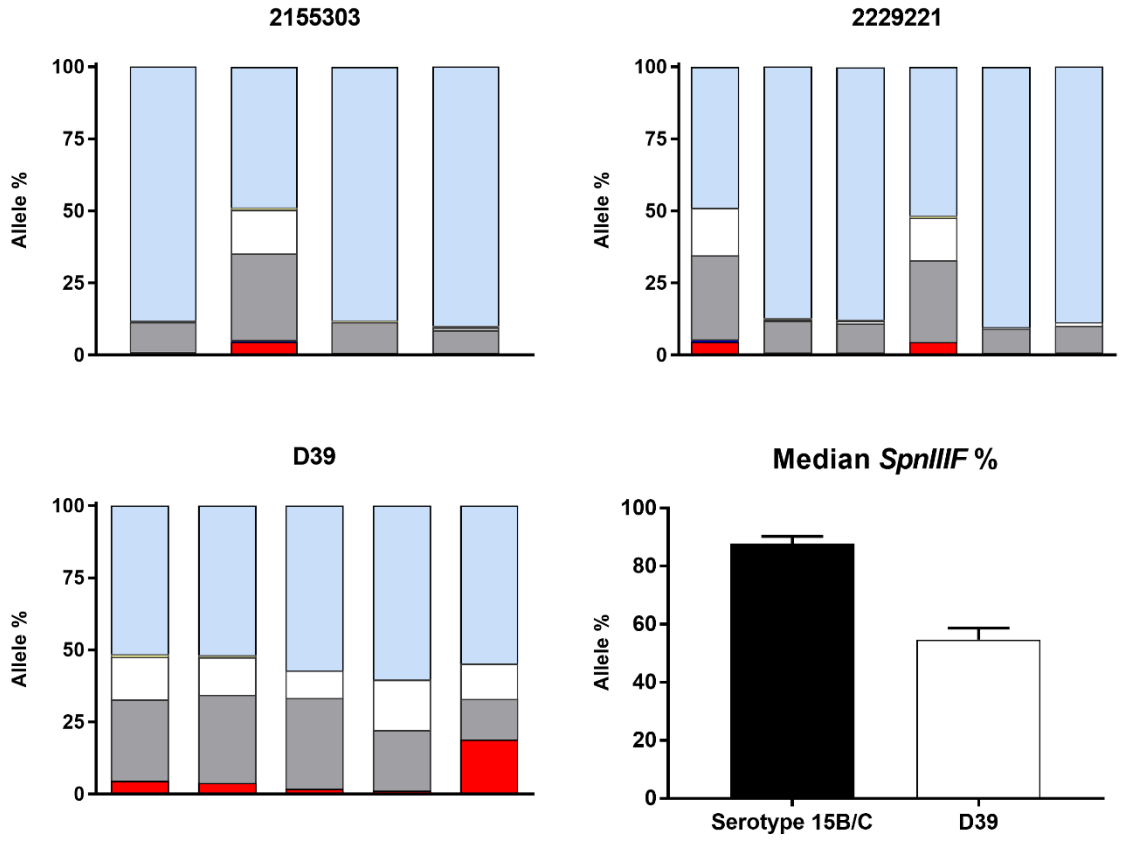


Figure 3.10 – Single colony analysis of serotype 15B/C strains. *Serotype 15B/C strains were identified as potentially enriched in SpnIIIF. Single colonies from two serotype 15B/C strains (2155303 and 2229221) were quantified and compared to single colonies from a D39 stock enriched in SpnIIIF. The median percentage of SpIIIIF in colonies from the two serotypes was not significantly different due to the presence of several ~50% SpnIIIF colonies in the serotype 15B/C strains. Each bar represents a single colony and the colours represent the SpnIII variants detected in that colony (SpnIIIA red, SpnIIIB blue, SpnIIIC grey, SpnIIID white, SpnIIIE yellow and SpnIIIF light blue). Significance was tested using a Mann-Whitney test, *p≤0.05, **≤0.01, ***p≤0.001.*

To investigate whether the number of SpnIIIF sites found within the glycerol operon directly impacted the percentage of SpnIIIF in single colony isolates over time several mutants were generated. Firstly, a glycerol operon knockout was generated by the insertion of a Janus Kan^R cassette at the locus. This knockout mutant was then used to create D39 strains with different glycerol operons. Three D39 strains with non-D39 glycerol operons were generated as well as a complemented strain where the D39 operon was re-inserted. Analysis of single colonies (incubated for >12hours) revealed that D39 strains with alternative glycerol operons did not produce single colonies with high percentages of SpnIIIF (Fig 3.11A). Analysis of the median SpnIIIF percentage in each strain demonstrated that there were no significant differences in strains expressing alternative glycerol operons (Fig 3.11B).

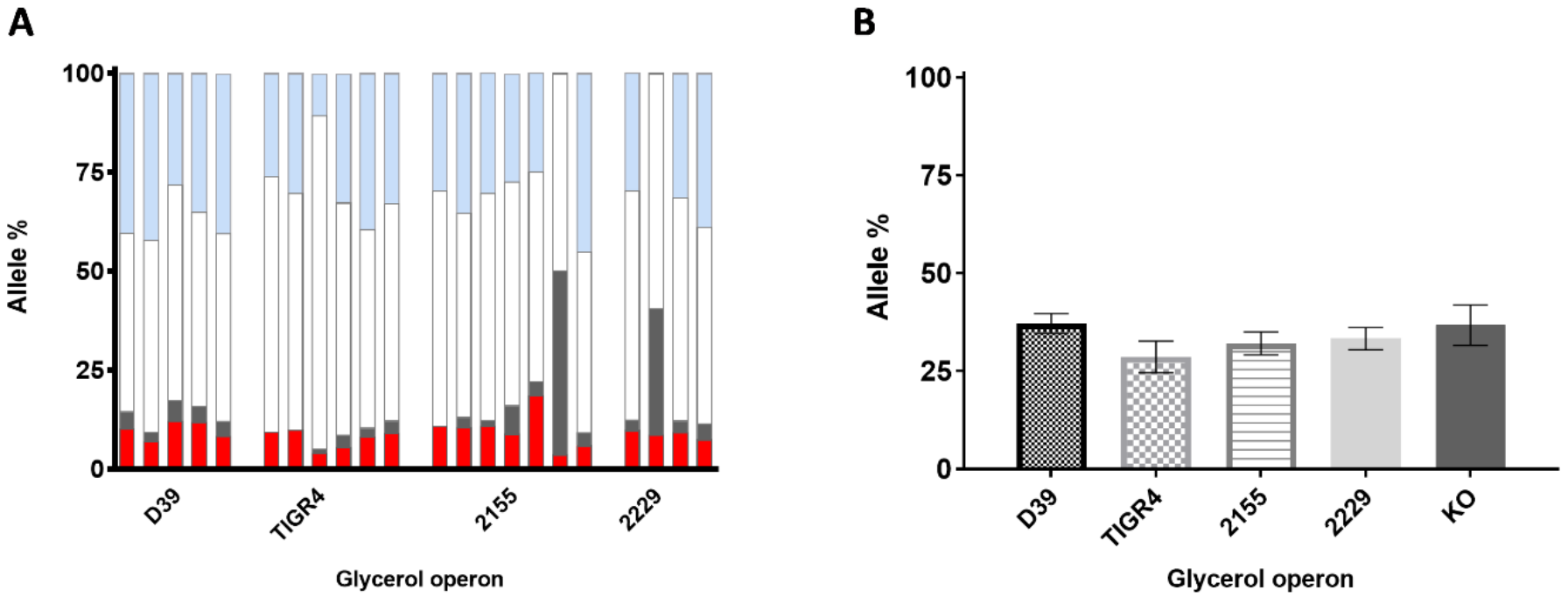


Figure 3.11 – Single colony analysis of D39 with alternative glycerol operons. *D39 mutant strains with different glycerol operons were PCR quantified to determine if the number of SpnIIIF sites in the glycerol operon could influence the percentage of SpnIIIF in colonies. Single colonies were analysed in D39 in which the glycerol operon was knocked out and replaced with either a D39, TIGR4 or serotype 15B/C (2155 and 2229) glycerol operon (Panel A). Each bar represents a single colony and the colours represent the SpnIII variants detected in that colony (SpnIIIA red, SpnIIIB blue, SpnIIIC grey, SpnIIID white, SpnIIIE yellow and SpnIIIF light blue). The median percentage of SpnIIIF in D39 colonies was not significantly increased in mutants with a serotype 15B/C glycerol operon (Panel B). There were also no significant differences in the median percentage of SpnIIIF in D39 without and without a glycerol operon. Significance was tested using a Mann-Whitney test, *p≤0.05, **≤0.01, ***p≤0.001.*

A substantial difference between the quantification of single colonies and patched cells is the storage of patches at -80°C in 10% v/v glycerol. Patched cells are stored as they are required for future use. If the storage in glycerol is detrimental to cells expressing SpnIIIF this could explain the discrepancy in variant percentages in patches when compared to colonies. It is not known if expression of the glycerol operon is required for the successful protection of cells at -80°C, therefore a series of experiments were conducted where strains with different glycerol operons were stored at -80°C in either 10% v/v glycerol or 10% v/v sucrose. A wildtype D39 strain was also stored without cryoprotectant as a control. Significantly fewer cells, 8.04%, survived from the no cryoprotectant group compared to wildtype cells stored in both 10% v/v glycerol ($p= 0.001$) and 10% v/v sucrose ($p<0.001$) (Fig 3.12A-B). All glycerol operon mutants, including the knockout mutant, survived equally well when stored in both glycerol and sucrose (Fig 3.12C). The result from the glycerol operon knockout indicates that expression of the glycerol operon is not required for the effective protection of cells stored at -80ºC.

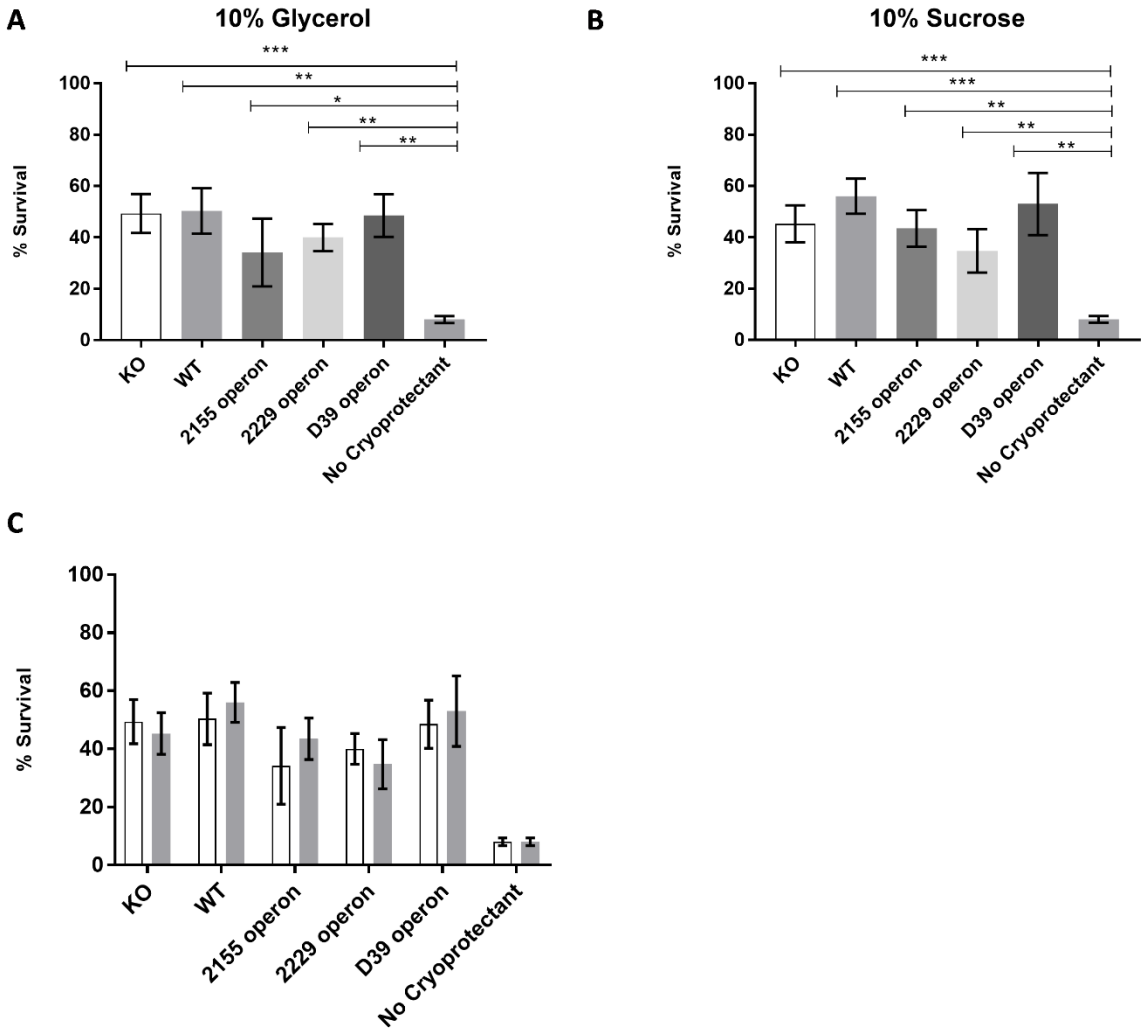


Figure 3.12 – Strain survival in different cryoprotectants. *Mutant strains with different glycerol operons were stored at -80ºC with either 10% v/v glycerol, 10% v/v sucrose or no cryoprotectant. The percentage survival in 10% v/v glycerol (Panel A) and 10% v/v sucrose (Panel B) was calculated by comparison with the CFU/ml of the same stock prior to freezing. In addition, a wildtype strain was stored with no cryoprotectant. No significant differences were observed between strains with different glycerol operons in either glycerol (Panel A) or sucrose (Panel B), the strain with no glycerol operon was also no significantly different in either condition. The only significant difference noted in both conditions was between strains stored with and without cryoprotectant. The percentage survival of each strain in both glycerol and sucrose was also compared (Panel C). No significant differences were observed in the effectiveness of either cryoprotectant. Significance was determined using a Student's T test, without assuming a consistent standard deviation, *p≤0.05, **≤0.01, ***p≤0.001.*

3.6.4 **The impact of fatty acids on cells expressing SpnIIIF**

Differences in the expression of the fatty acid biosynthesis (*fab*) operon were noted in the RNAseq data of the different SpnIII WT strains. The strain with the lowest level of *fab* operon expression was in SpnIIIF WT. This strain showed a >2 fold difference when compared to the SpnIIIE WT strain. In order to determine if supplementation of media with fatty acids, through the addition of 0.1% Tween 80 (100), could rescue the SpnIIIF phenotype the *spnIII* proportions in single colonies grown on media with and without Tween 80 were quantified. In both conditions two colonies with high SpnIIIF percentages were obtained (Fig 3.13A), while the remaining colonies contained between 40-60% SpnIIIF. Obtaining sufficient colonies with >75% SpnIIIF for statistical analysis was not possible, therefore colonies with a lower percentage of SpnIIIF were also analysed. This analysis was independent of those with >70% SpnIIIF. Colonies were grouped as follows: those expressing 40-60% SpnIIIF (low expressing), and those expressing 80-95% SpnIIIF (high expressing). Statistical analysis was conducted between strains within the same group but not across groups. A comparison of cells grown with and without additional fatty acids showed no significant differences in the median percentage of SpnIIIF in either the high or low expressing groups (Fig 3.13B). This result suggests that differential *fab* operon expression is not the reason for the reduced level of SpnIIIF in the WT strain.

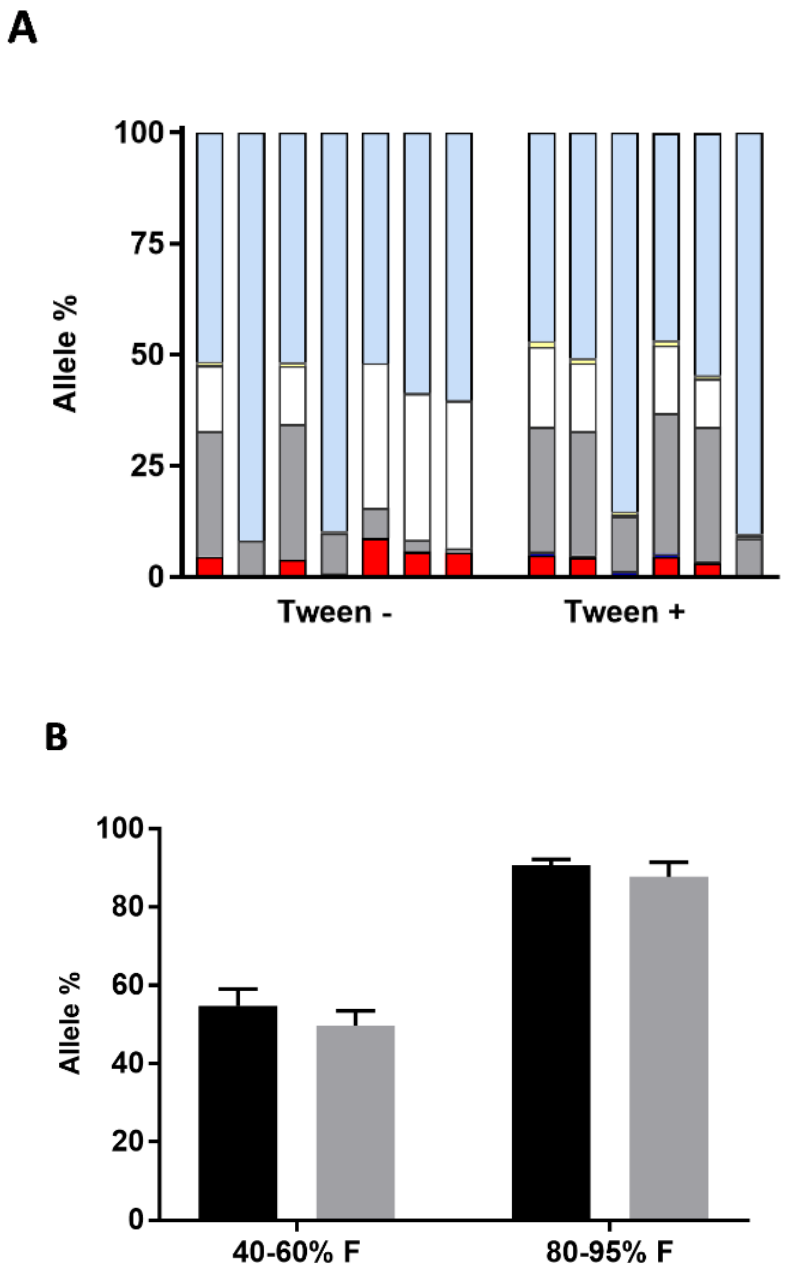


Figure 3.13 – Single colony analysis of strains grown in media with and without Tween 80 as a source of fatty acids. *Differences were observed in the expression of the fatty acid biosynthesis (fab) operon in different SpnIII wildtype strains. As a result, SpnIIIF was plated on agar plates with and without tween 80 as an external source of fatty acids to determine if the percentage of spnIIIF could be increased. Several colonies with a high percentage of SpnIIIF were observed in both the tween – (2/7 colonies) and tween+ (2/6) conditions (Panel A). Each bar represents a single colony and the colours represent the SpnIII variants detected in that colony (SpnIIIA red, SpnIIIB blue, SpnIIIC grey, SpnIIID white, SpnIIIE yellow and SpnIIIF light blue). In order to compare colonies in the tween+ and tween- conditions they were grouped as lower percentage (45-60% SpnIIIF) and higher percentage (80-95% SpnIIIF) colonies. There were no significant differences in the median percentage of SpnIIIF in either high or lower percentage colonies grown on media with and without additional fatty acids. Significance was tested using a Mann-Whitney test, *p≤0.05, **≤0.01, ***p≤0.001.*

3.6.5 **Is SpnIIIF more susceptible to oxidative stress?**

S. pneumoniae is routinely cultured on blood agar to neutralise the hydrogen peroxide produced by SpxB when cells are grown aerobically (116). Alternatively, cells can be grown on plates containing catalase, the enzyme responsible for converting hydrogen peroxide to water and oxygen (116). Pneumococcal growth rates are slower on catalase agar when compared to colonies grown for the same time period on blood agar. Using SpnIIIE colonies to calculate the difference, there were approximately 10 fold more cells in colonies on blood agar (5.00E+6) compared to catalase agar (7.33E+5). To determine if growth on agar plates with different catalase sources could influence SpnIII recombination SpnIIIE and SpnIIIF WT strains were plated on both for equal lengths of time.

As expected the SpnIIIE WT strain produced consistently more SpnIIIE founded colonies on both blood and catalase agar (Fig 3.14A). The SpnIIIF WT strain resulted in approximately half of the colonies having a significant (>50%) percentage of SpnIIIF on both agars (Fig 3.14B). A comparison of the median SpnIII variant percentages on both blood and catalase agar revealed that there is no significant difference in the SpnIIIF colonies (Fig 3.14C). The use of catalase is not detrimental to cells expressing SpnIIIF. Interestingly however a significant difference was found in the SpnIIIE colonies that were used as a control. On catalase agar *SpnIIIE* colonies show a significantly lower median percentage of *SpnIIIB* (0.00±0.00) when compared to colonies grown on blood agar (0.60±1.62) ($p=0.002$), no other significant differences were noted (Fig 3.14C).

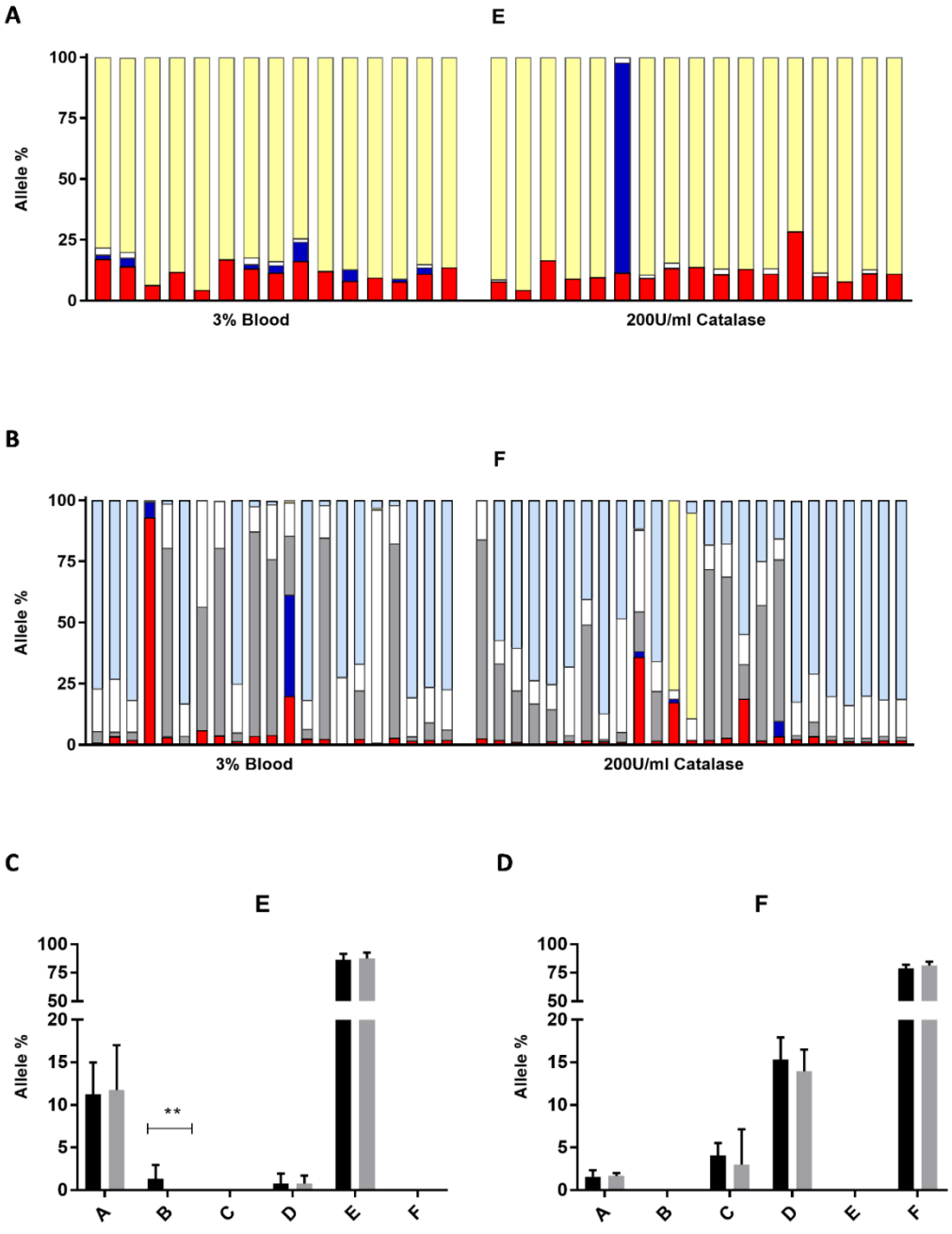


Figure 3.14 – Single colony analysis of strains grown on media supplemented with 3% horse blood or 200U/ml catalase. *SpnIIIE (control) (Panel A) and SpnIIIF (Panel B) WT strains were grown on media supplemented with either 3% horse blood or 200U/ml catalase. Each bar represents a single colony and the colours represent the SpnIII variants detected in that colony (SpnIIIA red, SpnIIIB blue, SpnIIIC grey, SpnIIID white, SpnIIIE yellow and SpnIIIF light blue). To ensure sufficient colonies with >75% SpnIIIF were analysed the number of single colonies tested was increased. Colonies with >75% SpnIIIF showed no significant differences in any SpnIII variant percentage when grown on blood or catalase agar (Panel D). However, a significant difference was observed in colonies with >75% SpnIIIE (Panel C). On agar with catalase the median percentage of SpnIIIB was significantly less than colonies grown on blood supplemented agar (p= 0.002). Despite this a single SpnIIIB colony was obtained from the catalase plate. No other significant differences were observed in median variant percentages in SpnIIIE colonies. Significance was tested using a Mann-Whitney test, *p≤0.05, **≤0.01, ***p≤0.001.*

Due to recombination dynamics the percentage of SpnIIIB is often zero or close to zero in a colony founded by SpnIIIF. As a result, this experiment was repeated using the SpnIIIE and SpnIIIA WT strains (Fig 3.15). Once again the majority of colonies analysed from the SpnIIIE WT strain were founded by SpnIIIE expressing cells (Fig 3.15A). The SpnIIIA WT strain produced several SpnIIIE founded colonies when grown on blood agar (Fig 3.15B), and these were excluded from statistical analysis as no SpnIIIE founded colonies from this strain were observed on catalase agar. There were no significant differences between SpnIIIE founded colonies from either strain on blood agar. Once again when the median variant percentages were analysed SpnIIIB was found at a significantly lower percentage on catalase agar compared to blood agar ($p=0.008$). This was also reflected in a significantly higher median percentage of SpnIIIE on catalase agar ($p=0.017$). No significant differences were observed between SpnIIIA founded colonies on blood or catalase agar. SpnIIIA colonies would be expected to contain a percentage of SpnIIIB, however the lack of significant difference here could be explained by the lower rate of recombination within these colonies (Table 3.4). If recombination is already at almost undetectable levels it may not be possible to accurately analyse differences. It is noteworthy that 2/11 colonies from the SpnIIIE WT were founded by SpnIIIB cells on catalase agar (Fig 3.15A). Growth on catalase therefore does not appear to completely inhibit SpnIIIB, instead it may interfere with recombination dynamics making SpnIIIB cells within an SpnIIIE colony less likely.

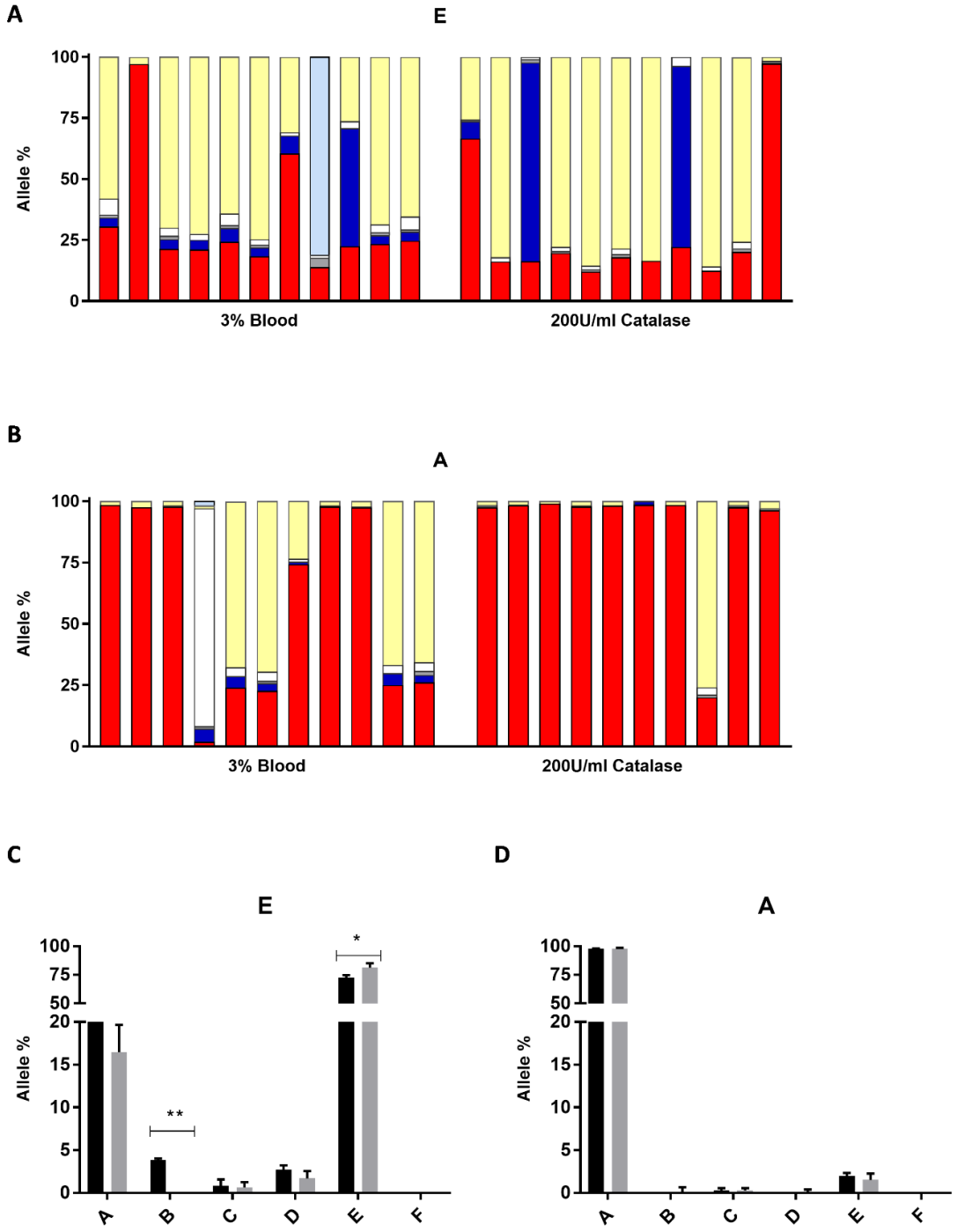


Figure 3.15 – Single colony analysis of strains grown on media supplemented with 3% horse blood or 200U/ml catalase. *SpnIIIE (Panel A) and SpnIIIA (Panel B) WT strains were grown on media supplemented with either 3% horse blood or 200U/ml catalase. Each bar represents a single colony and the colours represent the SpnIII variants detected in that colony (SpnIIIA red, SpnIIIB blue, SpnIIIC grey, SpnIIID white, SpnIIIE yellow and SpnIIIF light blue). Once again colonies with >75% SpnIIIE had a significantly lower median percentage of SpnIIIB when grown on catalase agar compared to blood agar (p= 0.008) (Panel C), this was also reflected in a significant increase in SpnIIIE (p= 0.017). No significant differences were observed in the median variant percentages in SpnIIIA colonies (Panel D). Significance was tested using a Mann-Whitney test, *p≤0.05, **≤0.01, ***p≤0.001.*

3.7 **Restriction alleviation**

Despite an increasing number of phase variable restriction modification systems being identified (64,66,67,73,72), the mechanism by which these cells prevent self-restriction of their own DNA following a change in recognition sequence has not yet been identified.

Restriction alleviation is the process by which bacteria limit DNA restriction when subjected to environmental stresses that can result in DNA damage (117). When significant DNA damage occurs, e.g. following exposure to ultraviolet light, extensive DNA repair must occur. The rapid synthesis of new DNA will result in large regions being temporarily unmethylated. In order to allow time for methylation, the ClpP protease, in complex with chaperone Clp proteins (118), can post-transcriptionally regulate proteins (119). In the case of Type I RM systems this post-transcriptional regulation can occur via the proteolytic degradation of the R subunit by ClpP (117).

The *S. pneumoniae* genome encodes several *clp* genes, including the protease ClpP and the chaperone protein ClpL (119). Gene expression analysis revealed significantly different levels of *clpL* in the different *SpnIII* WT strains (Table 3.5). As a result, it was hypothesised that ClpL was post-transcriptionally regulating expression of the SpnIII R subunit allowing time for the genome to be re-methylated using the new recognition sequence. As levels of *hsdR* mRNA were consistent it is also possible this regulation occurs through sequestering of the HsdR protein and not via degradation. To test this hypothesis *clpP* and *clpL* knockout mutants were generated by the insertion of an *aad9* spectinomycin cassette. If restriction alleviation is ongoing, the loss of the genes regulating the process should result in the self-restriction of cells that recombine. This would therefore be reflected in colonies that no longer display recombination at the *spnIII* locus. To further test this hypothesis single colonies from an *hsdR* frameshift (FS) mutant (provided by R. Haigh, University of Leicester), were also analysed.

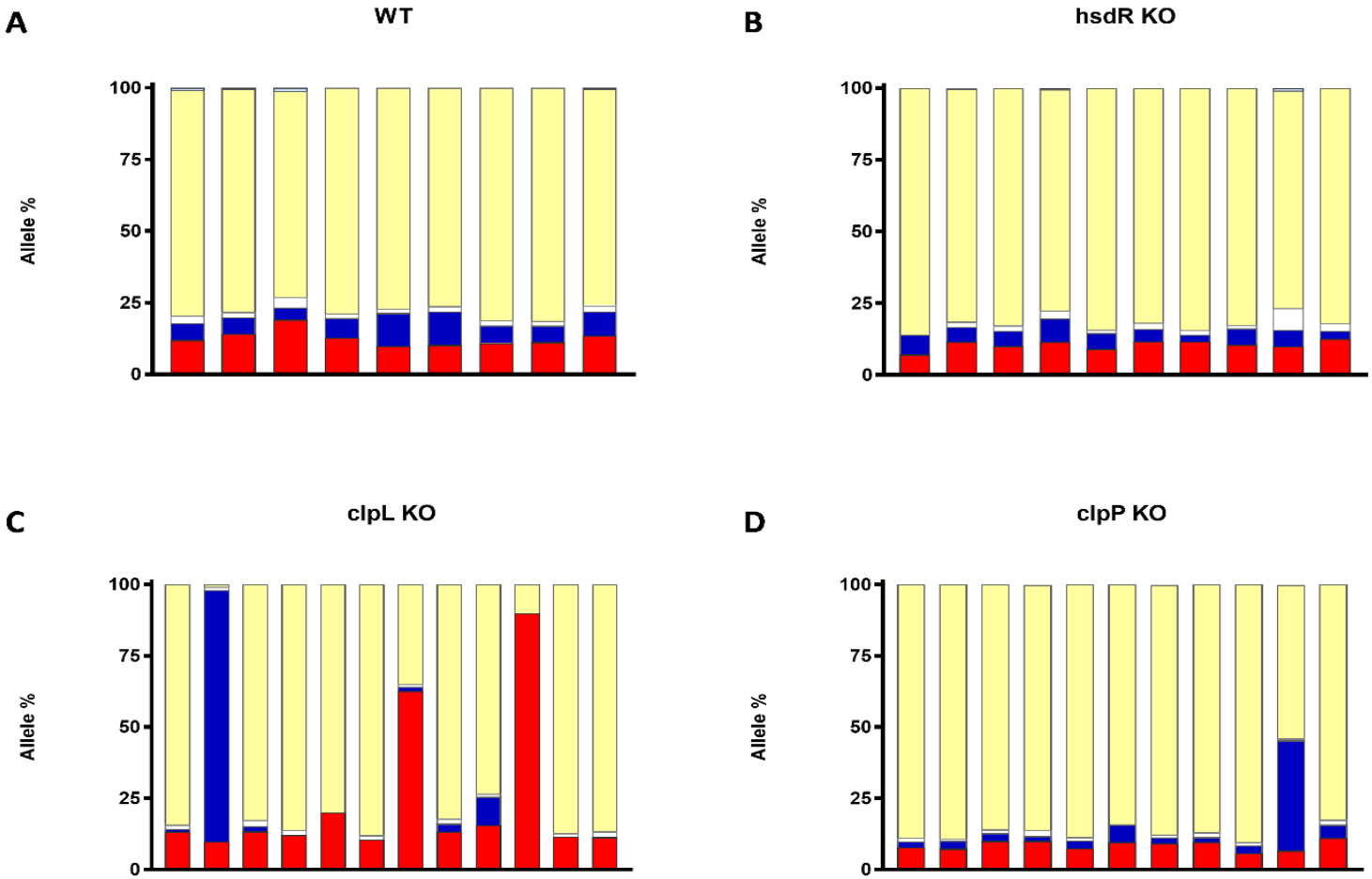


Figure 3.16 – Single colony analysis of mutants associated with DNA restriction and restriction alleviation. *Knockout mutants of clpL and clpP were generated by replacing the genes with an aad(9) spectinomycin cassette, an hsdR unmarked mutant was generated by frameshift mutation at the start of the gene. Each bar represents a single colony and the colours represent each SpnIII variant, (SpnIIIA (red), SpnIIIB (blue), SpnIIIC (grey), SpnIIID (white), SpnIIIE (yellow) and SpnIIIF (light blue)). Within each mutant strain recombination events on all three repeats (α, β and γ) were observed, indicating that the loss of hsdR (Panel B), clpL (Panel C) and clpP (Panel D) genes individually do not result in cell death following recombination.*

Following analysis of single colonies from the *clpP, clpL* and *hsdR* mutants it was determined that recombination is ongoing in all three mutants (Fig 3.16). As with all mutants generated in a WT SpnIIIE background the majority of colonies analysed were founded by SpnIIIE colonies. SpnIIIB and SpnIIIA founded colonies were observed in the *clpL* knockout strain (Fig 3.16C). The abolishment of recombination was not observed in either the *clpL* or *clpP* knockout colonies, suggesting that if restriction alleviation does protect newly recombined cells it is via an alternative protein to those analysed here.

3.8 **Chapter discussion**

The different rates of recombination observed in colonies founded by different SpnIII variants was an unexpected result and is likely to be, at least in part, associated with the increased expression of the CreX recombinase in SpnIIIA and SpnIIID cells. Increased expression of a recombinase would be expected to increase recombination rates, however the opposite is observed experimentally. It has previously been shown that there is an optimal concentration of purified Cre, beyond which increased quantities of enzyme can begin to inhibit recombination by forming DNA aggregates (90,120). It is possible that when CreX is expressed at increased levels in SpnIIIA and SpnIIID DNA-protein aggregates form, inhibiting recombination. It should however be noted that the concentrations of Cre that result in DNA aggregates are not found *in vivo*, and therefore alternative explanations should also be considered. Expression of CreX is fairly uniform in the remaining four variants therefore the differences observed in β repeat recombination are the result of an alternative, as yet unknown, mechanism. The difference in recombination rates between Hin and Fin mediated H-segment inversions in *S. enterica* have been attributed to different specificities of the two enzymes (91). The differences in recombination rates on the three *spnIII* inverted repeats may also arise, in part, from differing specificities for the recombinase(s) involved.

A wide variation was observed in the prevalence of individual SpnIII variants in different strains. These differences were not correlated with either serotype or sequence type suggesting it may be the result of random bottlenecking events

during *in vitro* culture and may not be consistent across the same strain held within different laboratories. A similar observation was made by Dybvig *et al.* where it was demonstrated that strains maintained by different individuals within the same laboratory could display alternative *hsdS* proportions (75).

In addition to random bottlenecking, the lack of conservation of sites across different pneumococcal strains would be likely to heavily impact recombination and expression of different *hsdS* variants. There is no evidence that methylation at one site will result in the same phenotypic effect in different strains. Furthermore, there are differences in the presence of a second type I RM system, known as *Spn*IV or the *tvr* locus, in different pneumococcal strains (78). The presence of an additional SSR within this locus has the potential to impact recombination within the *spnIII* locus. This additional SSR may help explain why some strains show a greater recombination rate, although further study would be required to verify this.

The most notable differences in the SpnIII WT strains were found when gene expression was investigated used RNAseq. Previous work using phase-locked mutants revealed major differences in the expression of the capsule operon (64) which was not replicated in this study. However, a variety of virulence associated genes, included the pneumolysin operon, were found to be differentially regulated. It has previously been demonstrated that strains lacking pneumolysin result in shorter episodes of carriage (121), as well as these strains being cleared from the blood significantly more quickly (20). The implication of these finding is that SpnIII methylation may be regulating genes associated with virulence.

In addition to the differences in expression of the pneumolysin operon it was also noted that the heat shock protein *dnaJ* was 2.15 fold upregulated in the SpnIIIB WT when compared to the SpnIIIE WT. *dnaJ* has been shown to be essential for survival inside macrophages in *Salmonella enterica* (122) and recent work in *S. pneumoniae* has shown that a *dnaJ* knockout is significantly attenuated in invasive models of disease (123). An SpnIIIB phase-locked mutant is also attenuated in invasive models of infection (64). While this appears contradictory at first it has also been shown that inoculation of mice with recombinant DnaJ protein leads to significantly shorter episodes of carriage (124) suggesting that

increased expression of *dnaJ* may aid in colonisation of the nasopharynx, a phenotype previously associated with an SpnIIIB phase-locked mutant (64). Alternatively, increased expression of *dnaJ* may indicate that the SpnIIIB WT strain is under greater stress in the growth conditions used for gene expression analysis. This hypothesis agrees with the observation that the SpnIIIB WT strain can grow significantly faster in a galactose based media, a sugar abundant in the human nasopharynx (112).

A notable result that concurs with previous work in an SpnIIIB phase-locked mutant is the significant decrease in the percentage of SpnIIIB within SpnIIIE founded colonies on agar supplemented with catalase when compared to agar supplemented with blood. SpnIIIB phase-locked mutants have been shown to produce less capsule and show a transparent phenotype compared to mutant's phase-locked into alternative SpnIII variants. It has previously been shown that SpxB is associated with an increased ability to survive in the nasopharynx (125), and that cells displaying the transparent phenotype express significantly more SpxB when compared to opaque cells (126,127). Catalase is used in agar to neutralise the hydrogen peroxide produced by pyruvate oxidase (SpxB) in response to growth in aerobic conditions (116,127). Overtime the quantity of catalase in the agar will decrease, and once exhausted the concentration of hydrogen peroxide will increase. If SpnIIIB expresses SpxB at a higher rate than other variants it will result in the supply of catalase being exhausted at a faster rate. It is therefore possible that, over time, on catalase agar plates, SpnIIIB cells are killed at a significantly faster rate than those expressing an alternative SpnIII variant. Importantly gene expression analysis was conducted in liquid media not on agar plates and differences in SpxB expression may not be obvious in these conditions. Overall while the direct difference in capsule expression seen in phase-locked mutants are not observed in this data-set there are indications that SpnIIIB is more closely associated with colonisation of the nasopharynx than other SpnIII variants.

Extensive analysis has not revealed why *S. pneumoniae* D39 cannot maintain a percentage of SpnIIIF equivalent to the other SpnIII variants. Despite differences in the expression of the fatty acid biosynthesis operon (*fab*) supplementation of media with fatty acids did not improve the percentage of SpnIIIF in colonies. The

lack of meaningful differences in the RNAseq data could be the result of the use of strains expressing 50-60% SpnIIIF. The presence of non SpnIIIF expressing cells within the RNA sequenced will reduce the appearance of gene expression differences.

WGS revealed a single SNP in a strain that could retain a higher SpnIIIF percentage. This SNP was located in the 333bp α inverted repeat located between the two TRD's of an active SpnIIIF S subunit. While the mechanisms controlling *spnIII* recombination are not fully understood, as described in chapter 4, it has long been established that changes to a recombinase binding site can reduce the efficiency of recombination mediated by site specific recombinases (128). It is possible that this SNP has disrupted one or more recombinase binding sites therefore reducing the rate of recombination and allowing a higher percentage of SpnIIIF to be maintained in SpnIIIF founded colonies even if conditions select against this variant.

The differences observed between single colonies grown for 12 hours and cells grown in high cell density patches suggests that SpnIIIF cells are lost at a higher cell density or reduced nutrient levels in the media. The increased cell density could promote cell to cell communication increasing the likelihood of competence induction in patched cells compared to single colonies. It is known that genetically competent pneumococcal cells are capable of killing non-competent siblings in a process known as fratricide (129). While no differences were observed in the ability of the different SpnIII WT strains to incorporate a single point mutation, the timing of competence induction was not investigated. With the higher percentage of non SpnIIIF cells in a patch if there is a delay in the induction of competence in SpnIIIF cells there is a possibility that fratricide may occur with SpnIIIF cells lysed by non-SpnIIIF cells.

No obvious mechanism has been identified to explain how cells survive following recombination events. These events result in a conflict between the methylation pattern of self DNA and the sequence recognise by the REase. In general restriction alleviation, through the Clp family of proteases, is thought to limit REase activity following plasmid acquisition (130), but as yet no evidence has been found that *S. pneumoniae* knockout mutants of the ClpP protease undergo

fewer *spnIII* recombination events. The mechanisms behind the preservation of self-DNA in cases where a new RM system is acquired by a cell are poorly understood. A mechanism of discrimination based on DNA condensation and the presence of non-specific DNA binding proteins has been proposed by Keatch *et al.* and may explain how cells are capable of distinguishing between foreign and self DNA (117). Cells capable of this distinction would be less likely to restrict self-DNA even when its methylation pattern conflicts with REase specificity.

Overall the mechanisms of *spnIII* recombination are complex and diverse. The observation that cells expressing different SpnIII variants show differential regulation of genes, including some associated with virulence, provides evidence in agreement with the work of Manso *et al.* (64). The subsequent work of this thesis will investigate the control of recombination and also the *in vivo* differences between strains pre-dominantly expressing different SpnIII variants.

4 Mechanisms of recombination at the *spnIII* locus

4.1 Introduction

SSRs occur in a wide variety of bacteria and phage to enable the inversion and excision of DNA. The presence of inverted repeat sequences can promote changes in gene expression through mechanisms such as the movement of promoter sequences (91,92). While the presence of direct repeats leads to excision from the genome and is typically associated with the excision of lysogenic phage from the host (79).

Located within the phase variable *spnIII* locus of *S. pneumoniae* there is a site specific tyrosine recombinase named CreX (64,71,78). In addition, there are three inverted repeats within the *spnIII* locus (Table 4.1), each facilitating an alternative recombination event and therefore the generation of a different specificity subunit. Recombination on these repeats is responsible for generating six different specificity (S) subunits. The presence of a recombinase within the locus suggests it may play a role in facilitating ongoing recombination events, in much the same way that the Hin recombinase facilitates inversions of the *fljB* promoter in *Salmonella enterica* (91,92).

4.2 Aim

The aim of this chapter was to determine to what extent CreX is involved in the recombination of the *hsdS* genes at the *spnIII* locus of *S. pneumoniae* D39.

4.3 The Cre recombinase of the *spnIII* locus

In *Mycoplasma pulmonis,* a species with a similar phase variable type I Restriction Modification system, a single site specific tyrosine recombinase (HvsR) is responsible for recombination between *hsdS* genes (77). Using *M. pulmonis* as a model it was investigated whether *hsdS* inversions of *spnIII* were under the control of a single SSR, CreX.

To control for the impact a change in locus size may have two mutants were generated, a deletion mutant where a chloramphenicol cassette was used to replace the gene and an unmarked frameshift mutant where the CreX protein

was truncated at amino acid 6, leaving the rest of the gene in frame. Both mutants were confirmed as having the same phenotype (data not shown) therefore all subsequent work was conducted using the frameshift mutant. In addition, to avoid bias three independent mutant strains (individual transformants) were used in this experiment.

To determine the impact of the loss of the *creX* gene on *spnIII* recombination three independent *creX* mutants were grown for ~18-20 generation on agar plates (16 hours), alongside a WT control. The resulting colonies were analysed using the allele quantification protocol to determine the percentage of each SpnIII variant within single colonies. Single colonies were used to ensure a pure starting population, a single cell founding a colony will express only one of the six SpnIII variants A-F. By selecting single colonies, it is possible to determine the percentage of the population that has undergone recombination events to generate alternative specificity subunits. Had growth been conducted in liquid media the starting population would always contain several, if not all, SpnIII variants, making it impossible to determine the direction of recombination. Single colonies therefore offer the distinct advantage of a known starting point in each experiment conducted.

Table 4-1 – The inverts repeats of the spnIII locus of S. pneumoniae D39.

Repeat name	Repeat Size (bp)	Associated TRDs	Genome Position (D39)	SpnIII variant following inversion on this repeat
α	333	2.1, 2.2, 2.3	458559-458890, 461420-461752	E → A
β	15	2.2, 2.3	458877-458891, 459948-459962	E → B
γ	85	1.1, 1.2, 2.1, 2.2, 2.3	458088-458172, 462158-462242	E → D

Data show that *spnIII* recombination is still occurring within single colonies (Fig. 4.1A), therefore the hypothesis that the locus is under the sole control of the CreX recombinase has been disproven. However, the median percentages of the four SpnIII variants found within these colonies were found to be significantly different

from the WT (Table 4.3). All colonies analysed in this experiment were founded by cells expressing SpnIIIE, therefore it is possible to generate SpnIIIA, B and D via a single recombination event (Table 4.1).

There is a significantly higher median percentage of the founding variant (SpnIIIE) in the *creX* knockout strains (95.70±0.79) compared to the WT (89.30±6.39), $p<0.0001$, suggesting there is significantly less recombination occurring overall within these colonies (Fig 4.1B). The median percentage of SpnIIIA fell from 7.60±5.08 to 4.30±0.73 ($p=0.004$), and while the median percentage of SpnIIID is low in WT colonies (0.61±0.72) it also fell significantly to 0.00±0.14 ($p<0.0001$). Most strikingly the median percentage of SpnIIIB was found to be 0.00±0.00 in *creX* mutant colonies, while it was 2.35±3.64 in WT colonies ($p<0.0001$). In order for a cell to alter its S subunit from SpnIIIE to SpnIIIB it requires a single recombination event on the 15bp (β) repeat.

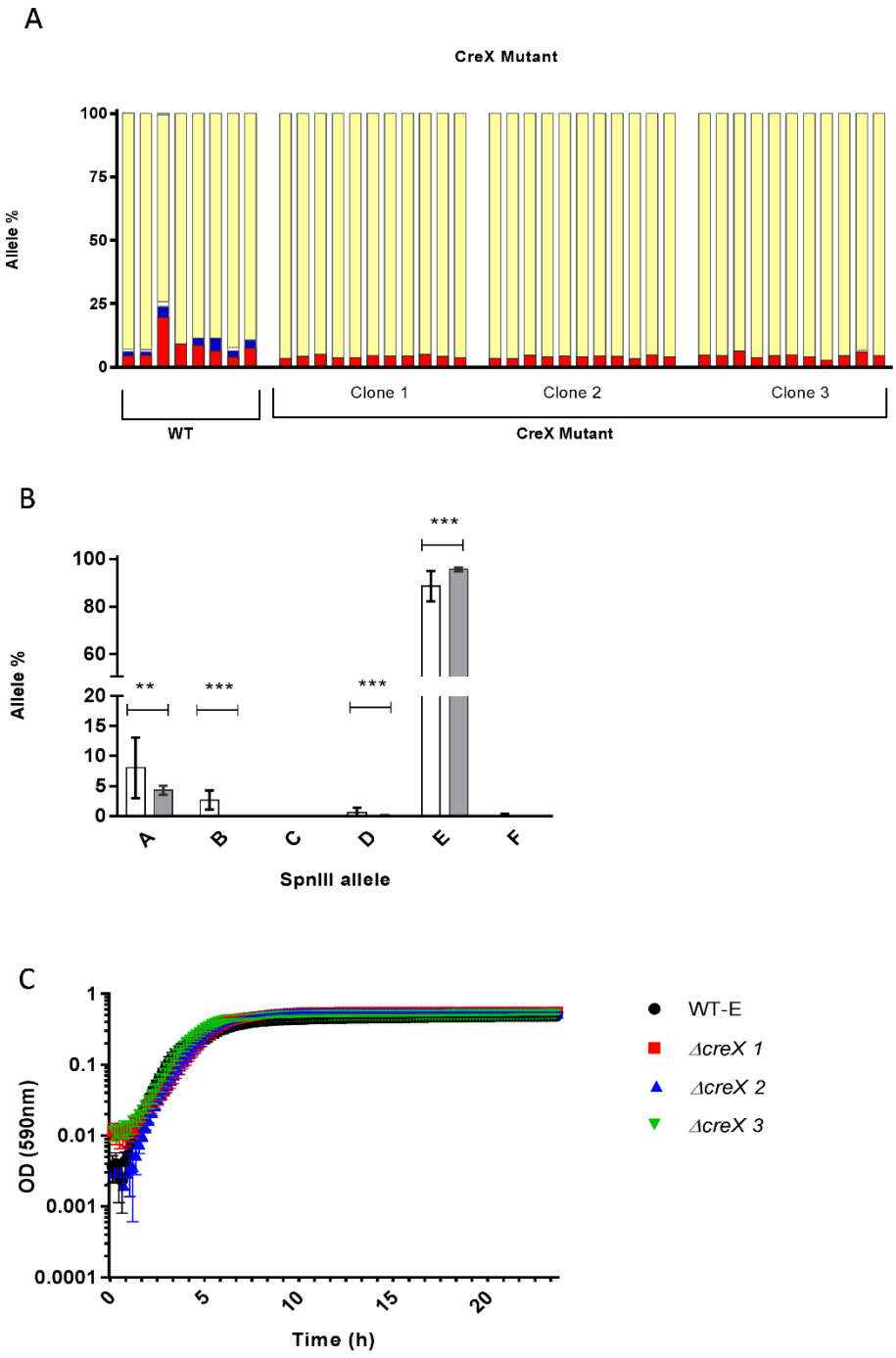


Figure 4.1 – Comparison of strains with and without the creX gene of the spnIII locus. *The site specific recombinase (CreX) found within the spnIII locus was hypothesised to control recombination within the locus. In order to confirm this, multiple frameshift mutants were generated and single colonies (panel were picked for allele quantification analysis. All colonies used were confirmed to be predominantly expressing the same allele (SpnIIIE). In panel A, each bar represents a single colony and the colours represent the SpnIII variants detected in that colony (SpnIIIA red, SpnIIIB blue, SpnIIIC grey, SpnIIID white, SpnIIIE yellow and SpnIIIF light blue). There are statistical significant differences in the SpnIII variant proportions between the wildype and the ΔcreX strains, with significantly less recombination in strains not producing CreX. Recombination events to generate SpnIIIA decreased from a median of 7.60±5.08 to 4.30±0.73 (p=0.0037), SpnIIIB from 2.35±1.60 to 0.00±0.00 (p<0.0001) and SpnIIID from 0.61±0.72 to 0.00±0.14 (p<0.0001) when compared to the WT. The proportion of the founder variant (SpnIIIE) increased significantly from a median of 89.30±6.39 to 95.70±0.79 (p<0.0001) compared to the WT. Significance was tested using a Mann-Whitney test, *p≤0.05, **≤0.01, ***p≤0.001.*

In order to compare the data from individual colonies they were treated as replicas and the median percentage of each allele within colonies was plotted. Colonies from all three *ΔcreX* strains were combined as no statistical differences were observed between them. Statistical analysis demonstrated a significant reduction in recombination on the α and γ repeats. In addition, recombination on the 15bp β repeat was abolished (Fig 4.1B). Changes to SpnIIIC and SpnIIIF could not be determined as each requires two recombination events when a cell initially expresses SpnIIIE. This leads to their percentage within the population being too small to calculate significant changes.

In order to confirm that differences in variant percentage were not a result of slower growth, and therefore fewer generations in which to recombine, each of the mutants was grown for 24 hours in liquid media alongside a WT control (Fig. 4.1C). The mean doubling time of the *ΔcreX* strains was 38.88±2.15 minutes compared to 36.82±2.33 minutes in the WT, which using a Students T test was found not to be significantly different (p=0.188). In addition, to replicate the conditions used in the allele quantification experiment each strain was grown on agar plates. After 12 hrs growth, three colonies from each strain were resuspended in liquid media and diluted using a 10 fold serial dilution. The average number of cells per colony was used to determine the number of generations per colony (assuming logarithm growth). The *creX* mutant strain contained a mean of 2.47E+05 CFU/ml compared to 2.70E+05 CFU/ml in WT colonies, each is equivalent to approximately 18 generations. Using a Student's T test the mean CFU/ml of each strain was found not to be significantly different (p=0.543). It was therefore concluded that differences in *spnIII* recombination between the WT and *creX* mutant strains are not the result of a difference in growth rate.

Overall it can be seen that while expression of *creX* does influence recombination on the three repeats of the *spnIII* locus, it only exclusively controls recombination on the 15bp (β) repeat. Recombination on the two larger repeats (α and γ) is significantly reduced but not abolished suggesting an additional mechanism of recombination is involved in the process. These data are confirmed by the work of Li *et al.* (65), where they also determined that a *creX* knockout is unable to recombine on the smallest 15bp repeat of the locus.

4.4 **The alternative site specific recombinases of *S. pneumoniae* D39.**

As proven by the work of this chapter and also by the work of Li *et al.* (2016), recombination at the *spnIII* locus is partially, but not exclusively, controlled by the CreX recombinase found within the locus. While it is somewhat unusual that recombination at a single locus would be controlled by more than one site specific recombinase (SSR) it has previously been seen in the H segment of *S. enterica.* It has been proven that while the Hin recombinase located in the H segment is primarily responsible for H segment inversions it is not the sole contributor. A second recombinase, Fin, associated with another invertible DNA region, is also able to facilitate H segment inversion, although to a lesser extent than Hin (91). Combined these two SSRs are responsible for the DNA inversions that control the expression of the flagella proteins in *S. enterica.*

There are four other potentially functional SSRs found within the D39 genome, (Fig. 4.2) one or more of which may be linked to recombination on the two larger repeats of the *spnIII* locus. It was therefore hypothesised that recombination of the *spnIII* locus is under the control of CreX and an additional site specific recombinase in a mechanism similar to that of the *hin* locus in *S. enterica* (91).

All SSR's found within the *S. pneumoniae* D39 genome can be found in Fig. 4.2, and the four of interest, along with the CreX recombinase, can be found in Table 4.2. Five of the six recombinases are tyrosine recombinases, while one is a serine recombinase. Each recombinase was evaluated and mutants generated. In addition, the protein sequences of the known SSR's were used to search the D39 genome for any unannotated SSR's, none were found.

Table 4-2 – Site specific recombinases within the S. pneumoniae D39 genome

Gene Name	Gene number	Comment	Reference
creX	SPD_0452	Site specific tyrosine recombinase. Found in the *spnIII* locus.	(71)
SPD_0921	SPD_0921	Site specific serine recombinase. Located next to a partial *hsdS* gene.	(131,132)
SPD_1013	SPD_1013	Site specific tyrosine recombinase. Member of the phage integrase family.	(131,132)
xerS	SPD_1023	Site-specific tyrosine recombinase. Known to perform the function of the XerC/D system.	(94)

xerD	SPD_1657	Site-specific tyrosine recombinase. Responsible for cutting and rejoining DNA molecules	(131,132)

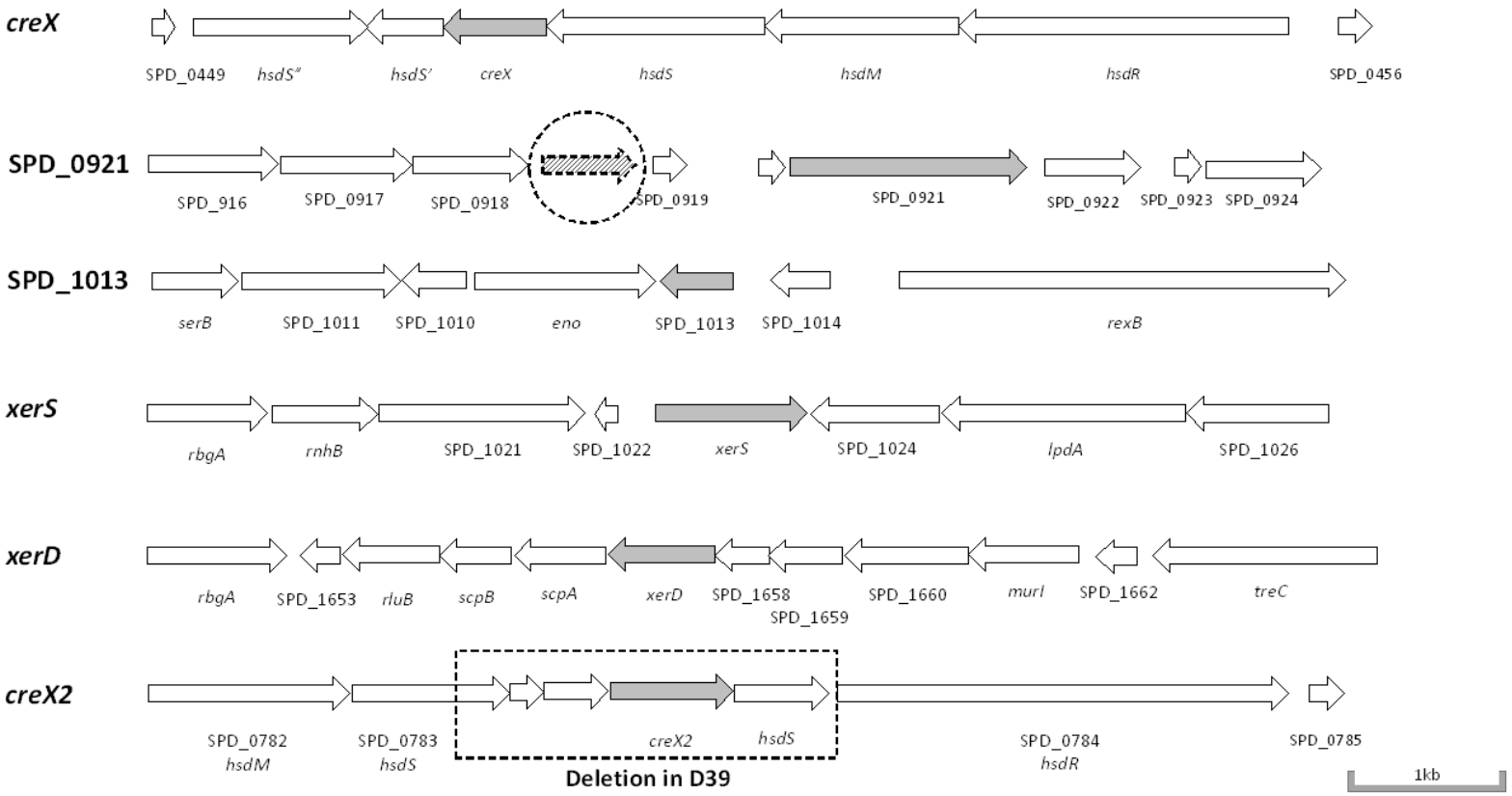


Figure 4.2 – Genome position of all known site specific recombinases in Streptococcus pneumoniae D39. *Within the S. pneumoniae D39 genome there are six different site specific recombinases. The CreX recombinase is found within the spnIII locus. SPD_0921 is the only site specific serine recombinase, it is located next to a non-annotated, truncated hsdS gene (shown within the dashed circle). SPD_1013 is a truncated site specific tyrosine recombinase. XerS is known to be associated with resolution of chromosomal dimers and XerD cuts and rejoins DNA molecules. Finally, there is another recombinase associated to a phase variable type I locus in other pneumococcal strains (creX2), while the complete TIGR4 locus is shown, part of this locus including the recombinase are deleted in the D39 strain (indicated by the dashed box).*

4.4.1 **The site specific recombinase SPD_1013**

In order to determine if the short gene SPD_1013 is a functional site specific recombinase it's complete nucleotide sequence was run through the ExPASy Translate Tool (133). The translation tool showed that no functional, full-length protein was produced from SPD_1013. In addition, an insertion sequence (IS) element, IS_630-SpnI (134), was identified as overlapping SPD_1013 using ISfinder (135). The position of IS_630-SpnI can be seen in green in Fig. 4.3. Although RNAseq data confirm low level SPD_1013 expression (data not shown), gene knock out experiments were not performed as the gene is expected to encode a non-functional protein.

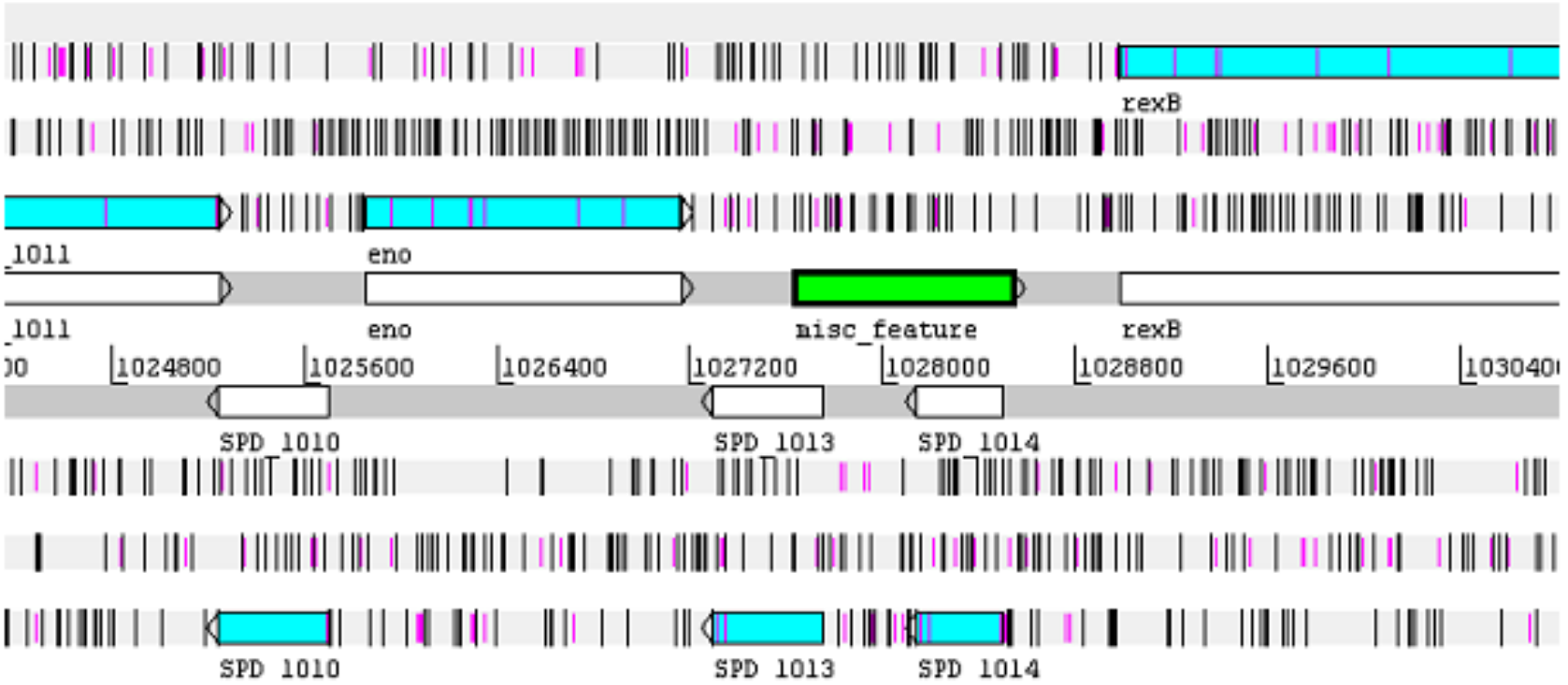


Figure 4.3 – Investigation of SPD_1013 as a functional protein. *The gene SPD_1013 is annotated as a phage integrase in the S. pneumoniae D39 genome however, when translated it is shown to be a non-functional, truncated protein. In addition an IS element, IS-630_Spn1* (134)*, shown in green (mapped with ISfinder,* (135)*) overlaps SPD_1013. This IS element has interrupted the gene confirming that it is non-functional.*

All other SSR's in the D39 genome were deemed to be complete, functional proteins and therefore potential candidates for involvement in *spnIII* recombination. As a result, a series of single and double knockout mutants were generated.

4.4.2 **The site specific recombinase SPD_0921**

SPD_0921 is a site specific serine recombinase, utilising a serine to catalyse the cutting and rejoining of DNA upon the recognition of a specific sequence (89). It is the only serine recombinase identified in *S. pneumoniae* D39. Interestingly, an unannotated partial *hsdS* gene was identified upstream of SPD_0921 (Fig 4.2). It is therefore a strong candidate for potential involvement in *spnIII* recombination. In order to investigate the hypothesis that SPD_0921 is involved in *spnIII* recombination two different mutants were generated, a single knockout mutant disrupting the gene with an *aad9 spectinomycin* cassette, and a double knockout of SPD_0921 (*aad9* cassette) and the *creX* recombinase (unmarked mutant).

To test recombination within the locus, WT and mutant strains were grown for 16 hours on blood agar plates. A WT strain was used as a control for the *ΔSPD_0921* knockout, and a *ΔcreX* strain for the *ΔcreX ΔSPD_0921* double knockout. Single colonies were picked for analysis and all colonies predominantly expressing the same SpnIII variant were treated as replicas.

Three individual knockout mutants of SPD_0921 (clones 1-3) were analysed (Fig. 4.4A). Each of these mutants predominantly expressed SpnIIIE, however several SpnIIIB colonies were identified (Fig. 4.4A) in clones two and three. Using the median percentage of each SpnIII variant as a measure of recombination, data show that there are no significant differences (Fig. 4.4C). The median SpnIIIE percentage in the mutant strain was 75.33±4.82 compared to 77.65±7.27 in the WT ($p=0.154$). No significant differences were observed in the median percentage of any SpnIII variant in the mutant compared to the WT (Table 4.3).

Three individual double knockout mutants of *creX* and SPD_0921 (clones 1-3) were analysed in addition to the single SPD_0921 knockout (Fig. 4.4B). All colonies tested predominantly expressed SpnIIIE. The median percentage of

SpnIIIB was 0.00±0.00 in all colonies of both the parent and mutant strains (Table 4.3). No significant differences in the median percentage of any SpnIII variant were identified between the parent and double knockout mutant (Fig 4.4D). Following this analysis, it was concluded that there is no redundancy and SPD_0921 is not involved in *spnIII* recombination.

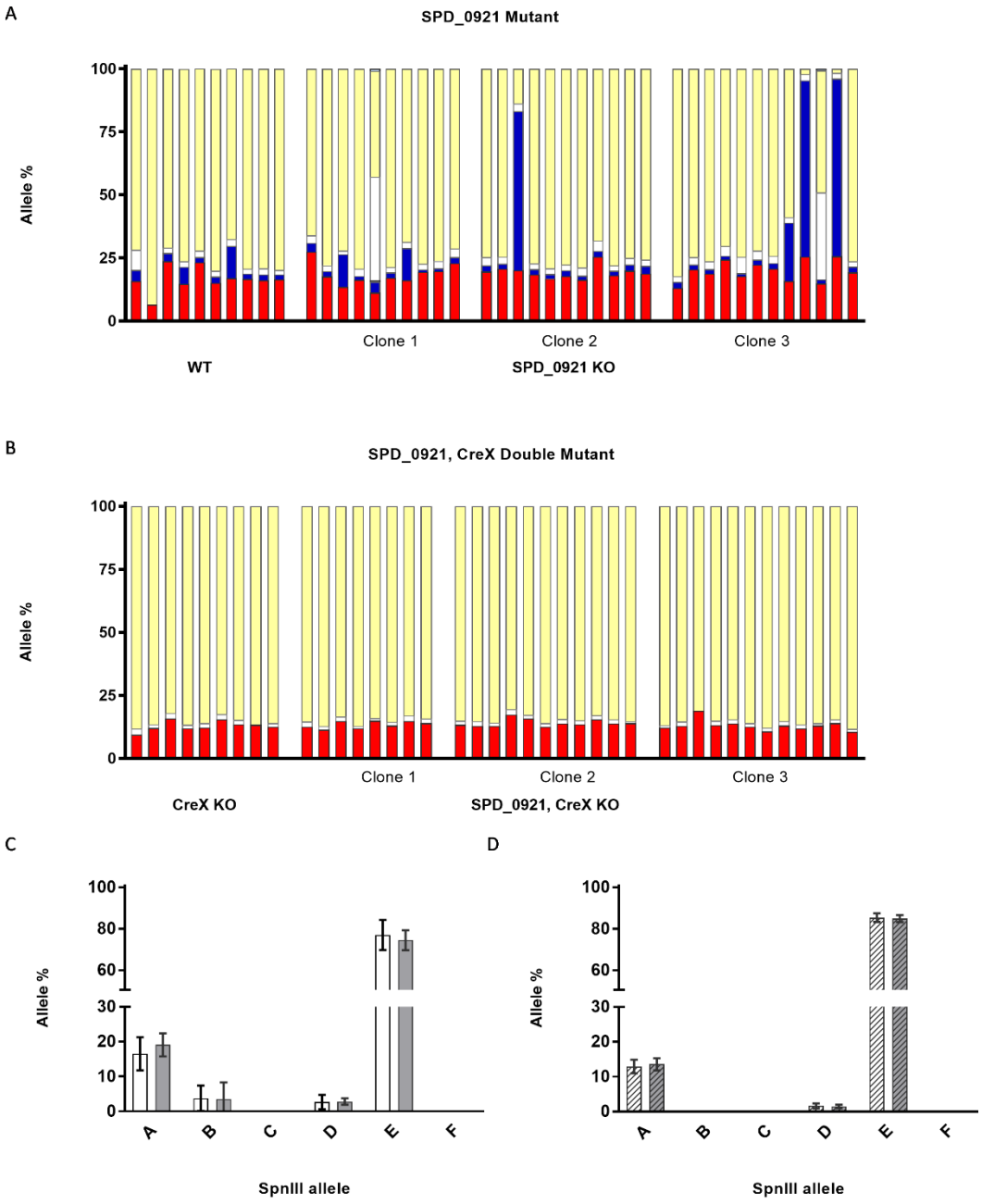


Figure 4.4 – Comparison of single colonies analysed in strains with and without the site specific recombinase SPD_0921. *It was confirmed that recombination within the spnIII locus is not under the sole control of the CreX recombinase found within the locus. In order to confirm if SPD_0921 could facilitate spnIII recombination multiple knockout mutants were generated by replacing the SPD_0921 gene with an aad(9) spectinomycin cassette. Strains were grown for 16 hours on TSA and single colonies were picked for allele quantification analysis. All strains used for the analysis of spnIII recombination were confirmed to be predominantly expressing the same allele (SpnIIIE) and non SpnIIIE colonies were excluded from statistical analysis. In panels A and B, each bar represents a single colony and the colours representing each SpnIII variant, (SpnIIIA (red), SpnIIIB (blue), SpnIIIC (grey), SpnIIID (white), SpnIIIE (yellow) and SpnIIIF (light blue)). The median variant percentage distribution (Panels C and D) is unchanged in the SPD_0921 and SPD_0921, creX double knockouts when compared to their parent strains (Table 4.3). This indicates that recombination is unaffected. All SpnIII variants generated by a single recombination event (SpnIIIA, B and D) are present within the colonies tested. Significance was tested using a Mann-Whitney test, *p≤0.05, **≤0.01, ***p≤0.001.*

4.4.3 **The site specific recombinase XerS (SPD_1023)**

The Xer machinery of prokaryotes is highly conserved across different species. The tyrosine recombinases, XerC and XerD function together to resolve dimers that occur during chromosomal replication (94). However, in both the *Streptococci* and *Lactococci* a phylogenetically distinct tyrosine recombinase, XerS, performs the function of both XerC and XerD (94,96). In addition to the resolution of chromosomal dimers the Xer machinery has been shown to have alternative roles in other species (96) therefore it was hypothesised that it may also be involved in the recombination of the *hsdS* genes of *Spn*III. To investigate this possibility two different *xerS* mutants were generated, a single knockout mutant disrupting the gene with an *aad9 spectinomycin* cassette, and a double knockout of *xerS* (*aad9* cassette) and the *creX* recombinase (unmarked mutant). Despite the conserved nature of this gene there was no observed growth phenotype in these mutants.

The background strain for the two mutants is different as the single *ΔxerS* mutant was difficult to generate. As a result, the unencapsulated D39 derivative DP1004 (136) was used as it is a more easily transformable strain lacking a capsule. The double mutant was generated in the *ΔcreX* background and is therefore a true D39 strain. The difference in strain background has resulted in a different predominant SpnIII variant within the two strains. Comparison of the two strains is complicated by the different variant backgrounds, however it is not necessary as comparison to parents strains is sufficient to determine any differences in recombination.

To test recombination within the locus, WT and mutant strains were grown for 16 hours on blood agar plates. A WT strain was used as a control for the *ΔxerS* knockout, and a *ΔcreX* strain for the *ΔcreX ΔxerS* double knockout. Single colonies were picked for analysis and all colonies predominantly expressing the same SpnIII variant were treated as replicas.

Data show that *spnIII* recombination in the *ΔxerS* strain is not significantly different from recombination in the WT strain (Fig. 4.5A). The median percentage of the dominant allele (SpnIIIA) in the *ΔxerS* knockout strain was 98.92±0.35 compared to 98.45±0.45 in the WT (p=0.051) (Table 4.3). In addition, no

significant differences were found in the median percentage of any SpnIII variant when compared to the WT (Fig. 4.5D) (Table 4.3). This indicates that *xerS* is not involved in *spnIII* recombination.

In order to confirm that *creX* does not compensate in the absence of *xerS*, a double knockout was also tested (Fig. 4.5B). In colonies analysed from the *ΔcreX ΔxerS* strain the dominant variant (SpnIIIA) had a median percentage of 98.94±0.42 compared to 98.81±0.94 in the WT (p=0.669). No significant differences were found in the median percentage of any SpnIII variant in the *ΔcreX ΔxerS* strain when compared to the *ΔcreX* parent strain (Table 4.3). It was therefore concluded there is no redundancy and *xerS*, like SPD_0921, is not involved in *spnIII* recombination.

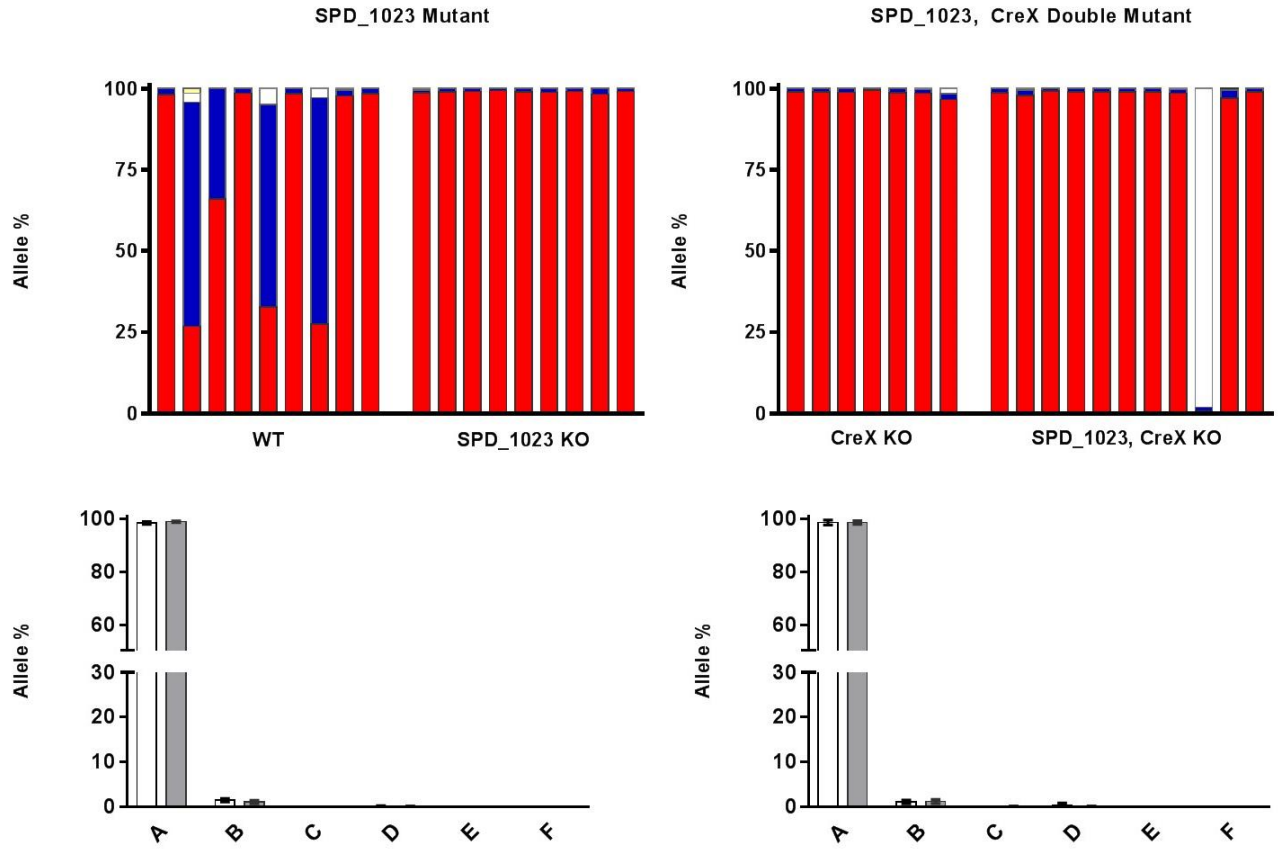


Figure 4.5 – Comparison of single colonies analysed in strains with and without the site specific recombinase SPD_1023 (xerS). *It was confirmed that recombination within the spnIII locus is not under the sole control of the CreX recombinase found within the locus. In order to confirm if XerS could facilitate spnIII recombination knockout mutants were generated by replacing the xerS gene with an aad(9) spectinomycin cassette. Strains were grown for 16 hours on TSA and single colonies were picked for allele quantification analysis. In panels A and B, each bar represents a single colony and the colours representing each SpnIII variant, (SpnIIIA (red), SpnIIIB (blue), SpnIIIC (grey), SpnIIID (white), SpnIIIE (yellow) and SpnIIIF (light blue)). The median variant percentage distribution (Panels C and D) is unchanged in the xerS and xerS, creX double knockouts when compared to their parent strains (Table 4.3). This indicates that recombination is unaffected. Significance was tested using a Mann-Whitney test, *p≤0.05, **≤0.01, ***p≤0.001.*

4.4.4 **The Site Specific Recombinase XerD (SPD_1657)**

In *E. coli* XerD functions alongside XerC, specifically to resolve the Holliday junctions formed by the action of XerC (88). While XerS (SPD_1023) is known to perform the function of XerC and XerD in *Streptococci* (94), a XerD homologue (SPD_1657) has been identified in the genome. Like XerD in other species, this recombinase is responsible for the cutting and re-joining of DNA. As *spnIII* recombination occurs via inverted repeats is it possible that resolution of DNA strand intermediates may be required. In order to determine if this is the case two different *xerD* mutants were generated, a single knockout mutant disrupting the gene with an *aad9 spectinomycin* cassette, and a double knockout of *xerD* (*aad9* cassette) and the *creX* recombinase (unmarked mutant).

To test recombination within the locus, WT and mutant strains were grown for 16 hours on blood agar plates. A WT strain was used as a control for the *ΔxerD* knockout, and a *ΔcreX* strain for the *ΔcreX ΔxerD* double knockout. Single colonies were picked for analysis and all colonies predominantly expressing the same SpnIII variant were treated as replicas.

Single colonies of the WT strain all predominantly expressed SpnIIIE however, of the two different *ΔxerD* knockout mutants one was generated in an SpnIIIB background (Fig 4.6A). Statistical analysis of cells predominantly expressing SpnIIIE show that there are no significant differences in median variant percentage when the *ΔxerD* knockout is compared to the wildtype strain (Fig. 4.6C) (Table 4.3). In colonies analysed from the *ΔxerD* strain the dominant variant (SpnIIIE) had a median percentage of 75.37±12.42 compared to 76.30±7.17 in the WT (p=0.429). No significant differences were found in the median percentage of any SpnIII variant in the *ΔxerD* strain when compared to the WT parent strain (Table 4.3). This indicates that XerD is not involved in *spnIII* recombination.

To confirm that CreX does not compensate in the absence of XerD, a double knockout was also tested (Fig. 4.6B). For the *ΔcreX ΔxerD* knockout both mutants tested predominantly expressed SpnIIIE (Fig 4.6B). The median percentage of SpnIIIE was 80.35±4.35 compared to 82.55±2.11 in the WT (p=0.863). No significant differences were found in the median percentage of any

SpnIII variant in the *ΔcreX ΔxerD* strain when compared to the *ΔcreX* parent strain (Fig. 4.6D) (Table 4.3). It was therefore concluded there is no redundancy and XerD, like SPD_0921 and XerS, is not involved in *spnIII* recombination.

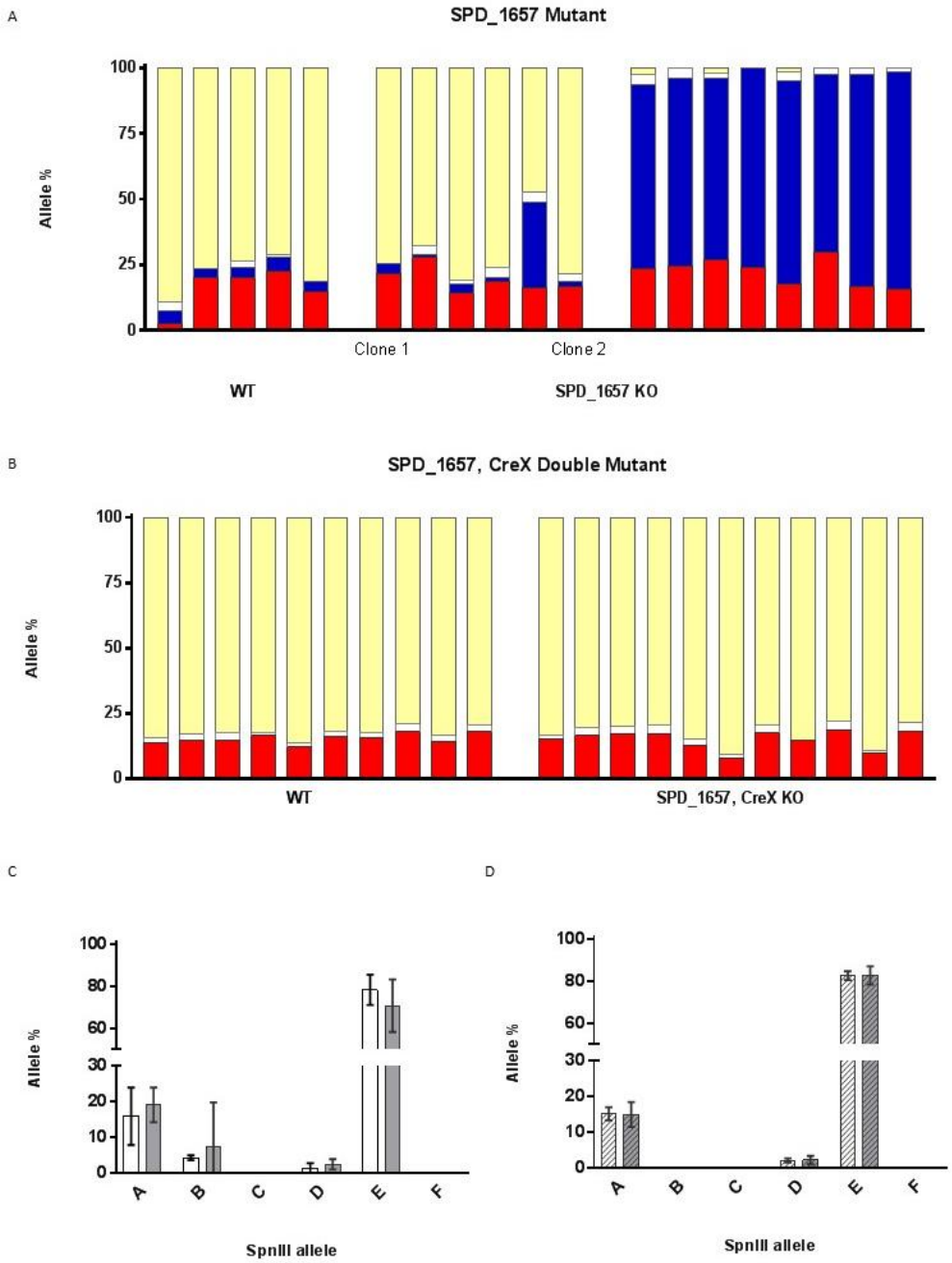


Figure 4.6 – Comparison of single colonies analysed in strains with and without the site specific recombinase SPD_1657 (xerD). *It was confirmed that recombination within the spnIII locus is not under the sole control of the CreX recombinase found within the locus. In order to confirm if SPD_1657 could facilitate spnIII recombination multiple knockout mutants were generated by replacing the SPD_1657 gene with an aad(9) spectinomycin cassette. Strains were grown for 16 hours on TSA and single colonies were picked for allele quantification analysis (Panels A and B). Of the two independent ΔSPD_1657 clones tested one was confirmed to be predominantly expressing SpnIIIB (8/8 colonies analysed) and was therefore excluded from statistical analysis. In panels A and B, each bar represents a single colony and the colours representing each SpnIII variant, (SpnIIIA (red), SpnIIIB (blue), SpnIIIC (grey), SpnIIID (white), SpnIIIE (yellow) and SpnIIIF (light blue)). The lack of SPD_1657 involvement in spnIII recombination is confirmed by the lack any significant difference between the median allele percentages of the wildtype (white bars) and the ΔSPD_1657 (grey bars) strains (panel C) and the ΔcreX (hashed white bars) and double knockout mutant (hashed grey bars) strains (panel D). Significance was tested using a Mann-Whitney test, *p≤0.05, **≤0.01, ***p≤0.001.*

Table 4-3 – Statistical differences in spnIII variant percentage for all site specific recombinase mutants

Strains compared			A			B		
Parent Strain		Mutant Strain	Median Parent	Median Mutant	P Value	Median Parent	Median Mutant	P Value
WT D39 (9)		*ΔcreX* (33)	7.60±5.08	4.30±0.73	0.0037	2.35±1.60	0.00±0.00	<0.001
WT D39 (10)		*ΔSPD_0921* (28)	16.35±4.76	18.78±3.28	NS	2.35±3.64	2.10±4.79	NS
ΔcreX (9)		*ΔcreX ΔSPD_0921* (31)	12.53±1.94	13.16±1.76	NS	0.00±0.00	0.00±0.00	NS
WT D39 (5)		*ΔSPD_1657* (6)	19.9±8.07	17.72±4.88	NS	4.3±0.69	2.71±12.40	NS
ΔcreX (10)		*ΔcreX ΔSPD_1657* (11)	15.13±1.83	16.48±3.48	NS	0.00±0.00	0.00±0.00	NS
WT D39 (5)		*ΔSPD_1023* (7)	98.45±0.45	98.92±0.35	NS	1.56±0.41	1.00±0.39	NS
ΔcreX (7)		*ΔcreX ΔSPD_1023* (8)	98.81±0.94	98.94±0.42	NS	1.19±0.47	1.06±0.46	NS

Strains compared			C			D		
Parent Strain		Mutant Strain	Median Parent	Median Mutant	P Value	Median Parent	Median Mutant	P Value
WT D39 (9)		*ΔcreX* (33)	0.00±0.00	0.00±0.00	NS	0.61±0.72	0.00±0.14	<0.001
WT D39 (10)		*ΔSPD_0921* (28)	0.00±0.00	0.00±0.00	NS	2.25±2.06	2.74±0.91	NS
ΔcreX (9)		*ΔcreX ΔSPD_0921* (31)	0.00±0.00	0.00±0.00	NS	1.75±0.71	1.59±0.46	NS
WT D39 (5)		*ΔSPD_1657* (6)	0.00±0.00	0.00±0.00	NS	1.1±1.53	3.01±1.44	NS
ΔcreX (10)		*ΔcreX ΔSPD_1657* (11)	0.00±0.00	0.00±0.00	NS	2.25±0.59	2.90±1.16	NS
WT D39 (5)		*ΔSPD_1023* (7)	0.00±0.00	0.00±0.00	NS	0.00±0.21	0.00±0.16	NS
ΔcreX (7)		*ΔcreX ΔSPD_1023* (8)	0.00±0.00	0.00±0.00	NS	0.00±0.17	0.00±0.19	NS

Strains compared			E			F		
Parent Strain		Mutant Strain	Median Parent	Median Mutant	P Value	Median Parent	Median Mutant	P Value
WT D39 (9)		*ΔcreX* (33)	89.30±6.39	95.70±0.79	<0.0001	0.00±0.25	0.00±0.00	NS
WT D39 (10)		*ΔSPD_0921* (28)	77.65±7.27	75.33±4.82	NS	0.00±0.00	0.00±0.00	NS

ΔcreX (9)	*ΔcreX ΔSPD_0921* (31)	86.02±2.04	85.07±1.76	NS	0.00±0.00	0.00±0.00	NS
WT D39 (5)	*ΔSPD_1657* (6)	76.3±7.17	75.37±12.42	NS	0.00±0.00	0.00±0.00	NS
ΔcreX (10)	*ΔcreX ΔSPD_1657* (11)	82.55±2.11	80.35±4.35	NS	0.00±0.00	0.00±0.00	NS
WT D39 (5)	*ΔSPD_1023* (7)	0.00±0.00	0.00±0.00	NS	0.00±0.00	0.00±0.00	NS
ΔcreX (7)	*ΔcreX ΔSPD_1023* (8)	0.00±0.00	0.00±0.00	NS	0.00±0.00	0.00±0.00	NS

4.4.5 Site specific recombinase conclusions

Analysis of all known site specific recombinases within the *S. pneumoniae* D39 genome revealed that only the recombinase associated with the *spnIII* locus*,* CreX*,* is involved in the inversion of *hsdS* genes. Data show that in the absence of CreX recombination on the smallest repeat of the locus is abolished, while recombination on the two larger repeats is significantly reduced.

It is important to note that this work has been conducted in *S. pneumoniae* D39 (or derivative strains). All site specific recombinases deemed to be functional were investigated, however alterative recombinases are present in other genomes. Importantly in the *S. pneumoniae* TIGR4 (71) genome there is a site specific recombinase associated with the phase variable type I RM system, *Spn*IV. Several genes are missing from this operon in *S. pneumoniae* D39, including the recombinase. It was therefore concluded that the recombinase of the *spnIV* locus could not be contributing to *spnIII* recombination in D39. It does not however rule out the possibility that this recombinase does contribute to *spnIII* recombination in other strains.

4.5 Homologous recombination and other non-site specific recombination genes.

S. pneumoniae is a naturally competent species known for its ability to take in and incorporate DNA from the environment (32,58,137). The process relies on there being sufficient homology between DNA sequences and the expression of what are known as the early and late competence genes (39).

As shown by the previous work of this chapter, recombination within the *spnIII* locus is partially but not solely controlled by site specific recombination. Due to the presence of inverted repeat sequences of 333bp, 85bp and 15bp in the locus (Table 4.1) there is the potential for an overlap in the function of proteins such as RecA. It was therefore hypothesised that recombination at the *spnIII* locus is occurring via the classical routes of homologous recombination, in addition to site-specific recombination.

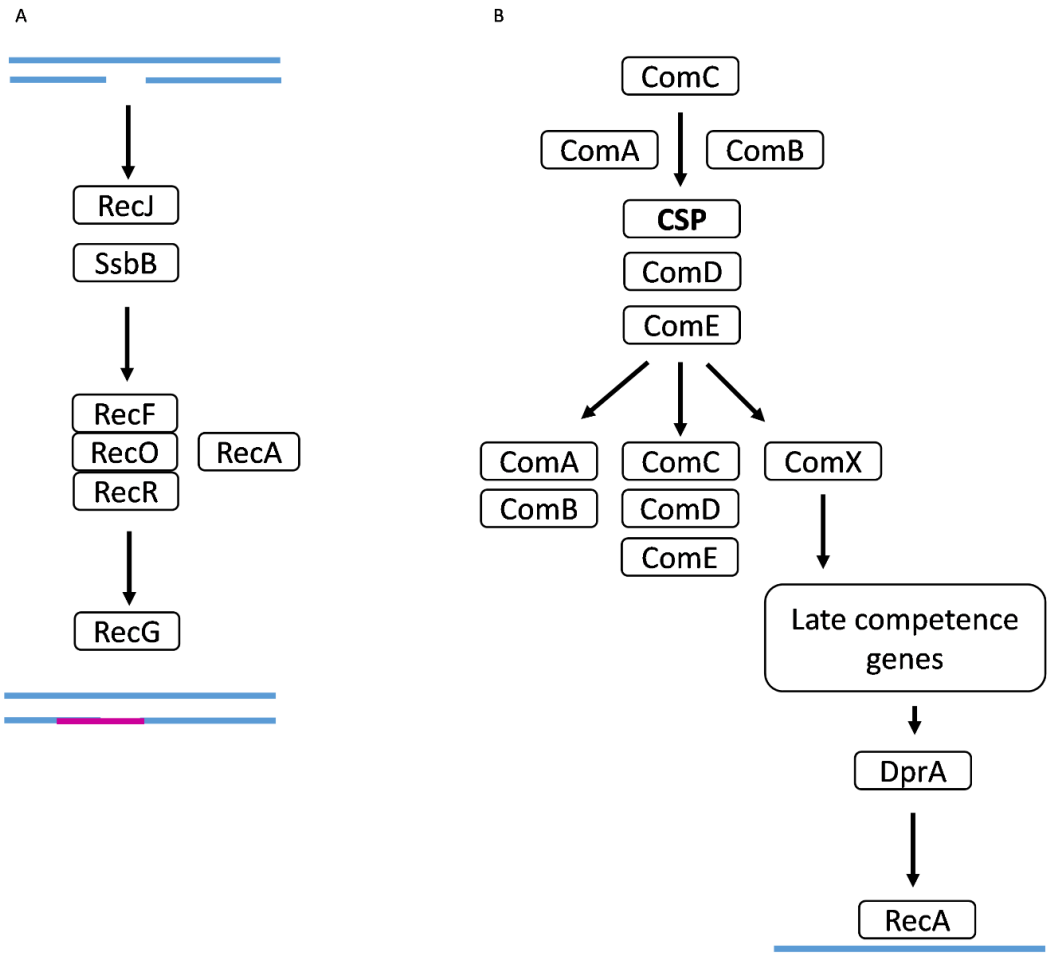


Figure 4.7 – The RecFOR and Competence pathways of Streptococcus pneumoniae. *The RecFOR pathway (Panel A) is primarily responsible for the repair of breaks in ssDNA. Single stranded binding protein (SsbB) protects DNA while the RecFOR complex recruit's recombinase A (RecA). The resulting Holliday junction is resolved by RecG. RecA is also part of the competence pathway (Panel B). Expression of the Competence Stimulating Peptide (CSP) promotes further expression of the com genes, expression of ComX regulates the late competence genes, included DprA* (35,37)*. DprA is responsible for recruiting RecA to incoming ssDNA, initiating the incorporation of DNA into the genome by homologous recombination* (35,39)*.*

The process of natural transformation requires expression of *recA* to incorporate exogenous DNA by homologous recombination. In addition, RecA is part of the RecFOR pathway of DNA repair by homologous recombination. In homologous recombination the RecFOR complex is required to recruit RecA onto ssDNA (35), while in transformation this role is fulfilled by DprA (39). Both of these pathways can be seen in Fig. 4.7. The work of this chapter involved the testing of a panel of mutants at multiple different stages in these pathways as shown in Table 4.4.

4.5.1 Homologous recombination and non-site specific recombination mutants

A panel of mutant strains and genomic DNA were received from Calum Johnston from the lab of Patrice Polard (Table 4.4). For each strain received genomic DNA was extracted and transformed into a WT D39 strain that matched the SpnIII variant background of the other mutants in this study (SpnIIIE). Each of the mutants tested is associated with homologous recombination in some way. In addition to the mutants received from Toulouse, two further candidate genes were identified which may be involved in *spnIII* recombination. These were SPD_2069 (a SpoJ homologue) which was seen to be differentially expressed in phase locked SpnIII mutants (64) and RecU (SPD_0337) a Holliday Junction resolvase with a preference for 3 and 4 strand DNA intermediates (138). As site specific tyrosine recombinases are known to function by the generation of Holliday junctions (138) RecU was considered a promising candidate. An *aad9* spectinomycin cassette was used for both knockout mutants. In addition, a double knockout of *recU* (*aad9* cassette) and *creX* (unmarked mutant) was created. As SPD_2069 is located next to the origin of replication a significant portion of the gene was left in situ to ensure that essential cellular processes were not disrupted.

In order to test recombination within the locus, WT and mutant strains were grown for 16 hours on blood agar plates. A WT strain was used as a control for all single knockout strains, and a *ΔcreX* strain for double knockouts where appropriate. Single colonies were picked for analysis and all colonies predominantly expressing the same SpnIII variant were treated as replicas.

Table 4-4 - Homologous Recombination mutants received from the lab of Patrice Polard

Gene Name	Gene number	Comment	Antibiotic marker	Strain reference
ssbB	SPD_1369	Binds single stranded DNA and facilitates the interaction of other proteins with DNA	Spectinomycin	(139)
recN	SPD_1062	ATPase involved in DNA repair	Spectinomycin	(39)
recR	SPD_1485	DNA repair by homologous recombination, part of the RecFOR complex	Kanamycin	(35)
recF	SPD_2054	DNA repair by homologous recombination, part of the RecFOR complex	Erythromycin	(35)
recG	SPD_1507	Catalyses branch migration in Holliday junction intermediates	Chloramphenicol	(140)
hexA	SPD_0371	Mismatch repair ATPase	Erythromycin	(141)
dprA	SPD_1122	Recombination-mediator protein, responsible for loading RecA onto incoming ssDNA during transformation.	Kanamycin	(39)
recA	SPD_1739	ATP-dependent hybridisation of homologous ssDNA	Chloramphenicol	(137)

4.5.2 **The Holliday junction resolvase RecU**

All colonies analysed for both the WT and *recU* knockout strains predominantly expressed SpnIIIE and contained SpnIII variants A, B and D (Fig. 4.8A). No significant differences were found in the median variant percentages when the *recU* knockout was compared to its wildtype parent strain (Fig. 4.8C). The median percentage of SpnIIIE was 76.00±3.75 in the *recU* knockout compared to 76.30±7.17 in the WT (p=0.721) (Table 4.6).

In addition, colonies of the *recU creX* double knockout were compared to colonies of a *creX* knockout (Fig 4.8B). Of the WT colonies analysed one was found to predominantly express SpnIIIF, while the remaining three expressed SpnIIIE, all

recU creX knockout colonies predominantly expressed SpnIIIE (Fig 4.8B). Statistical analysis of the median variant percentages showed there were no significant differences in any SpnIII variant (Fig 4.8D). The median percentage of SpnIIIE was 86.60±2.78 in the *recU creX* knockout and 86.40±1.62 in the *creX* parent (p=0.885) (Table 4.6). Following this analysis, it was concluded that despite its involvement in the resolution of Holliday junctions, RecU does not play a role in *spnIII* recombination.

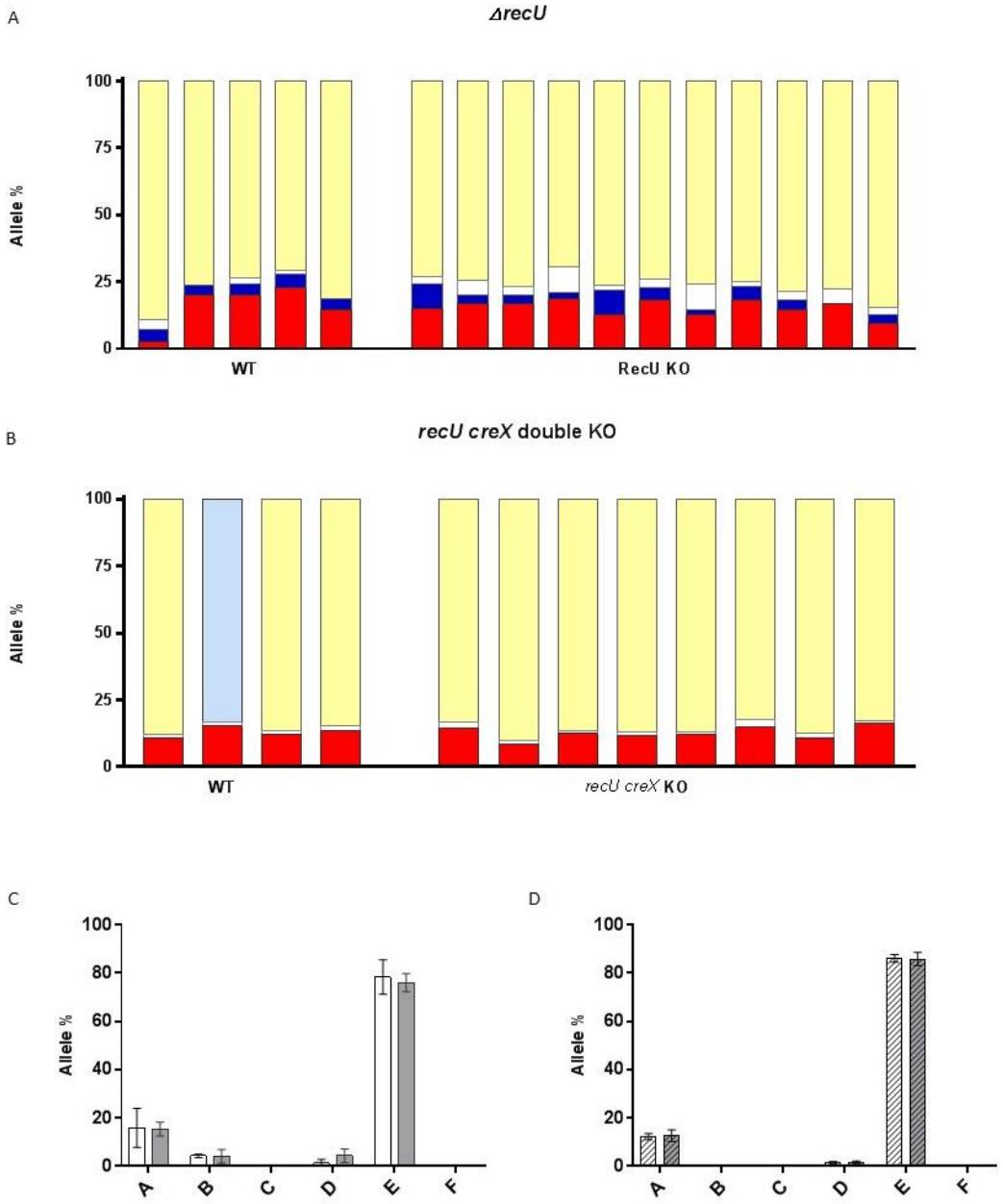


Figure 4.8 – Comparison of single colonies analysed in strains with and without the Holliday junction resolvase RecU. *RecU is Holliday junction resolvase with a preference for binding to 3 and 4 strand DNA intermediates* (138)*, as a site specific tyrosine recombinase (such as CreX) functions by creating DNA intermediates* (88) *the potential role of RecU in spnIII recombination was investigated by the generation and testing of knockout mutants. recU was replaced with an add9 antibiotic resistance cassette in both wildtype (panel A) and ΔcreX (panel B) backgrounds. In panel A, each bar represents a single colony of either a parent (WT) or mutant (ΔrecU) strain. In panel B, each bar represents a single colony of either a parent (ΔcreX) or mutant (ΔrecU, ΔcreX) strain. Each colour represents a single SpnIII variant (SpnIIIA (red), SpnIIIB (blue), SpnIIIC (grey), SpnIIID (white), SpnIIIE (yellow) and SpnIIIF (light blue)). The median variant percentage distribution (Table 4.6) is unchanged in the recU (grey bars, Panel C) and recU, creX (hashed grey bars, Panel D) knockouts when compared to their parent strains (white and hashed white bars respectively), indicating that recombination is unaffected. All SpnIII variants generated by a single recombination event (SpnIIIA, B and D) are present within the colonies tested. Significance was tested using a Mann-Whitney test, *p≤0.05, **≤0.01, ***p≤0.001.*

4.5.3 **The SpoJ homologue SPD_2069**

Following the elimination of *recU* as a candidate for *spnIII* recombination a *ΔspoJ* mutant was tested. Of the eleven colonies tested from the *spoJ* knockout strain, ten were found to predominantly express SpnIIIE, while the remaining colony predominantly expressed SpnIIIA (Fig 4.9A). This single SpnIIIA colony was excluded from statistical analysis. The median variant percentages of the SpnIII variants were found not to be significantly different between the two strains (Fig 4.9B). The median percentage of SpnIIIE in the *spoJ* mutant and the WT was 72.54±4.09 and 73.85±12.31 respectively (Table 4.6). Based on this analysis it was concluded that SpoJ is not involved in *spnIII* recombination.

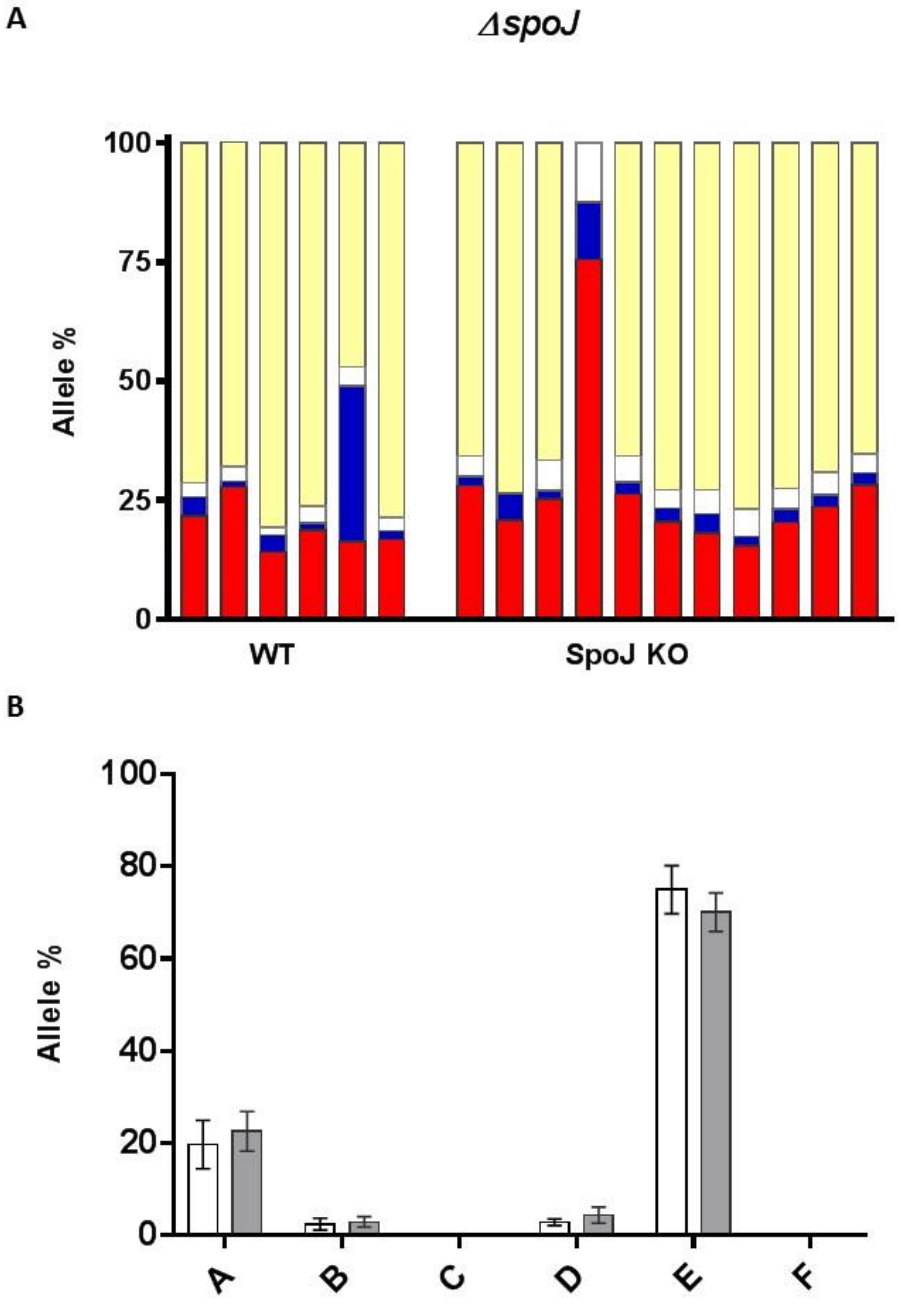


Figure 4.9 – Comparison of single colonies analysed in strains with and without the SpoJ like gene SPD_2069. *SPD_2069 (SpoJ) was identified through RNAseq as differentially expressed in locked SpnIII mutant strains* (64)*. As a result, its impact on spnIII recombination was investigated through the generation of a knockout mutant. SPD_2069 was replaced with a spectinomycin antibiotic resistance cassette and compared to a WT strain (panel A). In panel A, each bar represents a single colony of either a parent (WT) or mutant (ΔspoJ) strain. Each colour represents a single SpnIII variant (SpnIIIA (red), SpnIIIB (blue), SpnIIIC (grey), SpnIIID (white), SpnIIIE (yellow) and SpnIIIF (light blue)). In panel A all SpnIII variants generated by a single recombination event (SpnIIIA, B and D) are present within the colonies tested. The median variant percentage of all colonies (excluding the single SpnIIIA dominant colony) from the parent (white bars) and SPD_2069 knockout (grey bars) were compared (Panel B) and found not to be significantly different (Table 4.6). Significance was tested using a Mann-Whitney test, *p≤0.05, **≤0.01, ***p≤0.001.*

4.5.3.1 Confirmation of the RecA and RecU phenotypes

A UV tolerance assay was conducted to confirm the expected phenotypes of the *recA* and *recU* knockout mutants (Fig 4.10). Cells were grown to an exponential OD before being exposed to approximately 7.5 joules of UV irradiation. In addition to the testing of the *recA* and *recU* knockout mutants several of the SSR knockout mutants were also exposed to the same level of UV irradiation. The loss of a SSR would not be expected to have an impact on cell survival following UV exposure, while it is known that *recA* and *recU* are actively involved in DNA repair following UV damage.

As expected the *recA* and *recU* mutants were significantly less able to survive UV exposure compared to the WT with a survival rate of 0.003% ($p<0.001$) and 0.051% ($p<0.001$) respectively (Table 4.5). Unexpectedly, the *creX* mutant showed a significantly improved rate of survival, 37.35% ($p=0.024$). This was a reproducible result, occurring in multiple replicates of the experiment (data combined). This did not occur in the *creX,* SPD_0921 and *creX,* SPD_1657 double knockouts which had survival rates that were not significantly different from the WT.

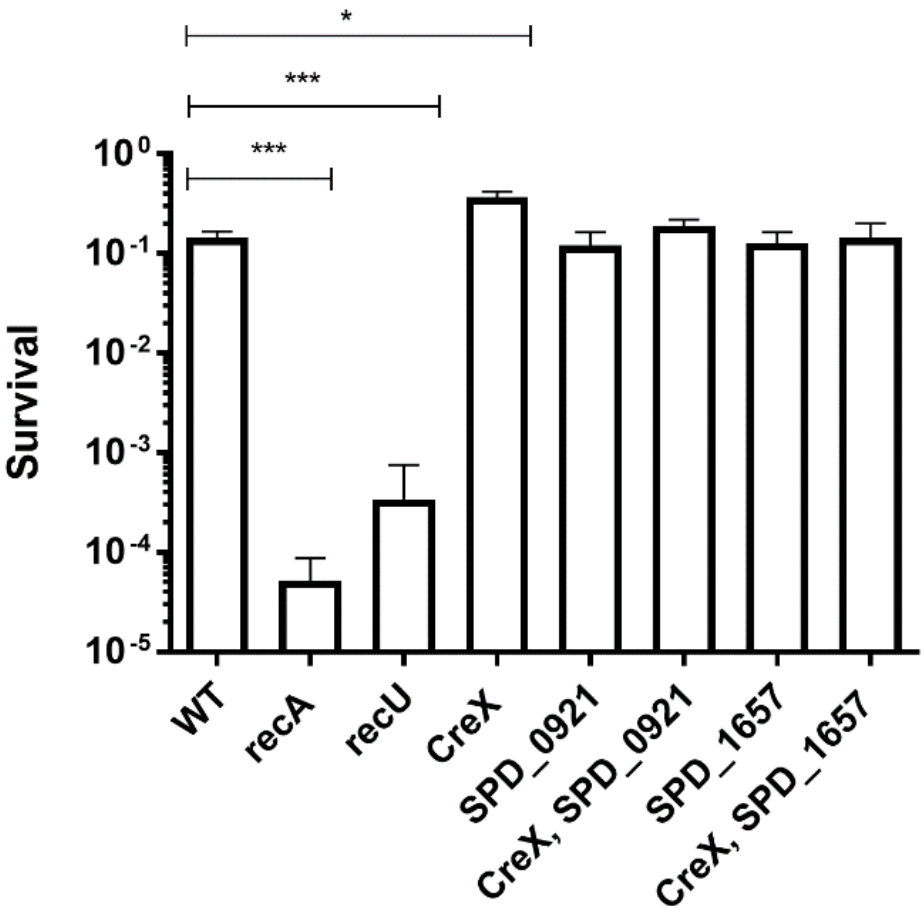


Figure 4.10 – Cell survival following UV exposure. *A panel of mutants were subjected to ~7.5 Joules of UV irradiation and the CFU/ml of cells pre and post UV exposure was used to calculate the proportion of survivors. The mean survival rate for WT cells was 14.37±2.27. There were no significant differences in the proportion of survivors in the SSR mutants (Table 4.5) with the exception of the creX mutant which showed a significantly higher survival rate than the WT (P=0.024). The recA and recU knockouts were significantly less able to survive exposure to UV (p<0.001 for both). Significance was tested using a Students T test, *p≤0.05, **≤0.01, ***p≤0.001.*

Table 4-5 – Cell survival following UV exposure

Strain	Mean CFU/ml without UV exposure[&]	Mean CFU/ml with UV exposure[&]	% Survival	P Value when compared to WT^
WT	2.37e+07	3.37e+06	14.2	-
recA	3.6e+07	1.00e+03	<0.01	<0.01
recU	5.00e+07	2.55e+04	0.05	<0.01
creX	2.23e+07	8.33e+06	37.4	0.02
SPD_0921	3.15e+07	3.83e+06	12.2	0.54
creX, SPD_0921	2.38e+07	5.05e+06	21.2	0.11
SPD_1657	1.30e+07	1.50e+06	11.5	0.54
creX, SPD_1657	2.60e+07	4.67e+06	18.0	0.99

[&]Mean calculated from three replicas.
^Significance calculated using a Student's T test.

4.5.4 **Investigating genes involved in homologous recombination**

Using the genomic DNA of strains supplied by the lab of Patrice Polard the following strains were generated and tested in an *S. pneumoniae* D39 background; *ΔssbB, ΔrecN, ΔrecR, ΔrecF, ΔrecG, ΔhexA, ΔdprA, ΔrecA.* For the individual antibiotic resistance of each mutant, see Table 4.4. While there are other genes associated with the homologous recombination machinery they function in pathways with one or more of the mutants that were tested.

4.5.4.1 SsbB, RecN, RecR, and RecA

The *ΔssbB* mutant showed only colonies initiated by a cells expressing SpnIIIE, within these colonies SpnIIIA, SpnIIIB and SpnIIID were also observed (Fig 4.11A). The median variant percentage of SpnIIIE, A, B and D did not differ significantly from the wildtype strain (Fig. 4.11B) (Table 4.6). The same is true of the *ΔrecN, ΔrecR* and *ΔrecA* strains, none were found to be significantly different to the WT (Table 4.6).

4.5.4.2 RecF

The *ΔrecF* strain proved to be the most diverse. Of the eight colonies tested, four were founded by SpnIIIE, three by SpnIIIA and one by SpnIIIF expressing cells (Fig. 4.11A). For statistical analysis only the SpnIIIE expressing cells were used for comparison to the WT. Despite the small number of replicates, it was sufficient to determine that there were no significant differences in the median variant

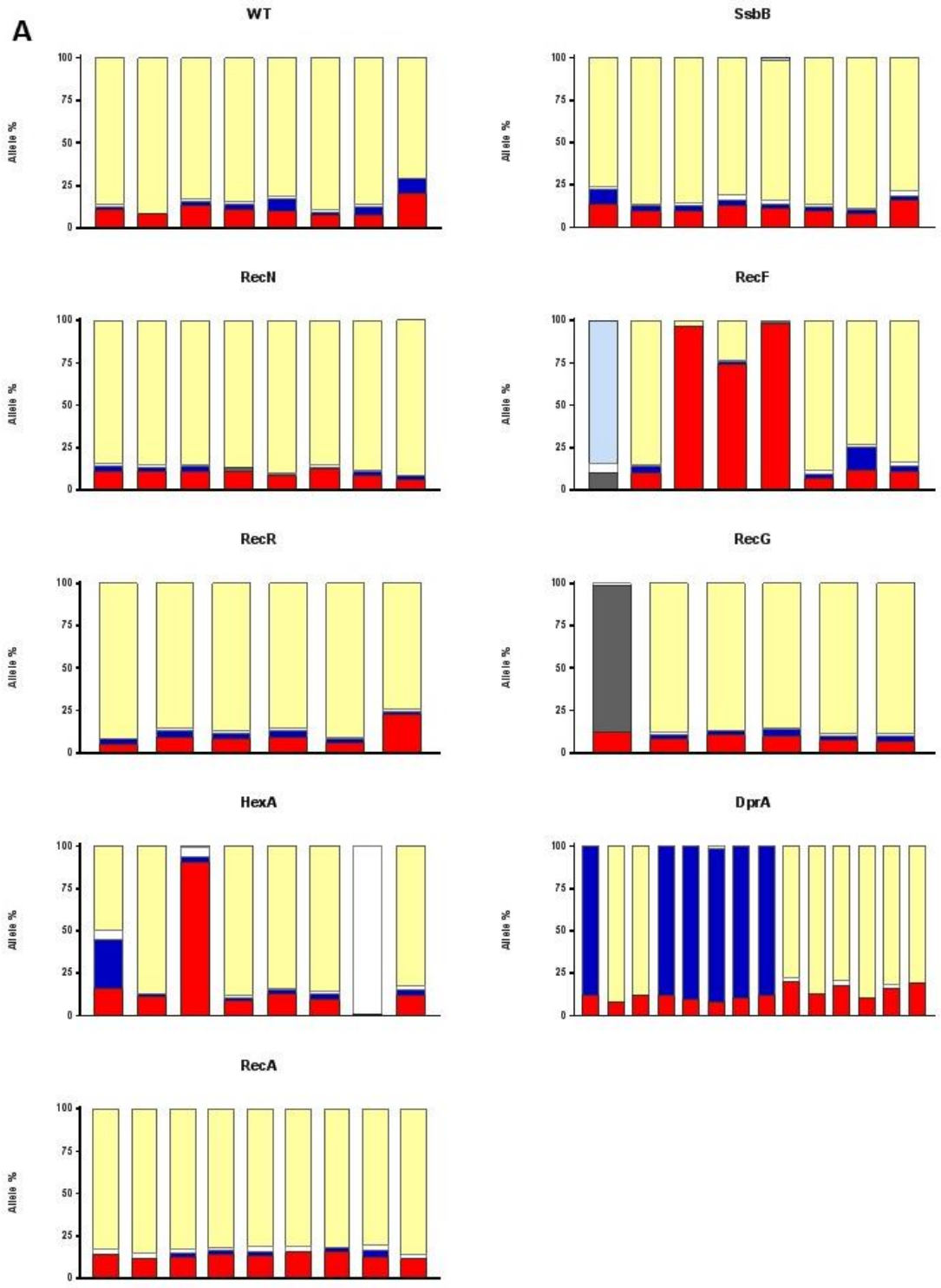
percentages of SpnIIIA, B, D and E between strains with and without *recF* (Fig. 4.11B) (Table 4.6).

4.5.4.3 RecG, and HexA

In the *ΔrecG* strain one colony of six was founded by a non SpnIIIE cell (Fig. 4.11A). This single SpnIIID colony was excluded from statistical analysis (Table 4.6). As with the *ΔssbB, ΔrecN, ΔrecR, ΔrecA and ΔrecF* strains the *spnIII* locus of the *ΔrecG* strain was found to recombine at the same rate at the WT (Fig. 4.11B). The *ΔhexA* strain showed colonies founded by cells expressing SpnIIIE (six), SpnIIIA (one) and SpnIIID (one) (Fig. 4.11A) As with the *ΔrecF* and *ΔrecG* strains only colonies founded by SpnIIIE expressing cells were used for statistical analysis (Table 4.6). When median variant percentages were compared to the WT (Fig. 4.11B), the *ΔhexA* strain was found not to be significantly different.

4.5.4.4 DprA

The final strain tested was a *ΔdprA* strain. In total fourteen colonies from this strain were tested, six of these were founded by SpnIIIB expressing cells, the remaining eight by SpnIIIE cells (Fig. 4.11A) Within SpnIIIE and SpnIIIB founded colonies, SpnIIIA and SpnIIID were observed, however SpnIIIE and SpnIIIB were not present within the same colony. It is a recombination event on the 15bp (β) repeat that results in a switch from SpnIIIE to SpnIIIB and vice-versa. For statistical analysis only SpnIIIE colonies were used and the median percentage of SpnIIIB in *ΔdprA* colonies (0.00±0.00) was found to be significantly different (p=0.001) to that of the WT strain (2.55±2.92) (Fig 4.11B) (Table 4.6). The median percentages of SpnIIIA (14.45±4.33) and SpnIIID (0.00±1.43) were unaffected by the loss of DprA, p=0.152 and 0.761 respectively (Table 4.6). To further investigate a *ΔcreX ΔdprA* strain was also produced and two individual mutants were analysed in comparison to the *ΔcreX* parent strain (Fig. 4.12A). A comparison of the median variant percentages of the *ΔcreX, ΔdprA* and *ΔcreX ΔdprA* strains (Table 4.6) showed there were no significant differences between any of these strains (Fig. 4.12B).


A
WT
SsbB
RecN
RecF
RecR
RecG
HexA
DprA
RecA
Allele %
100
75
50
25
0

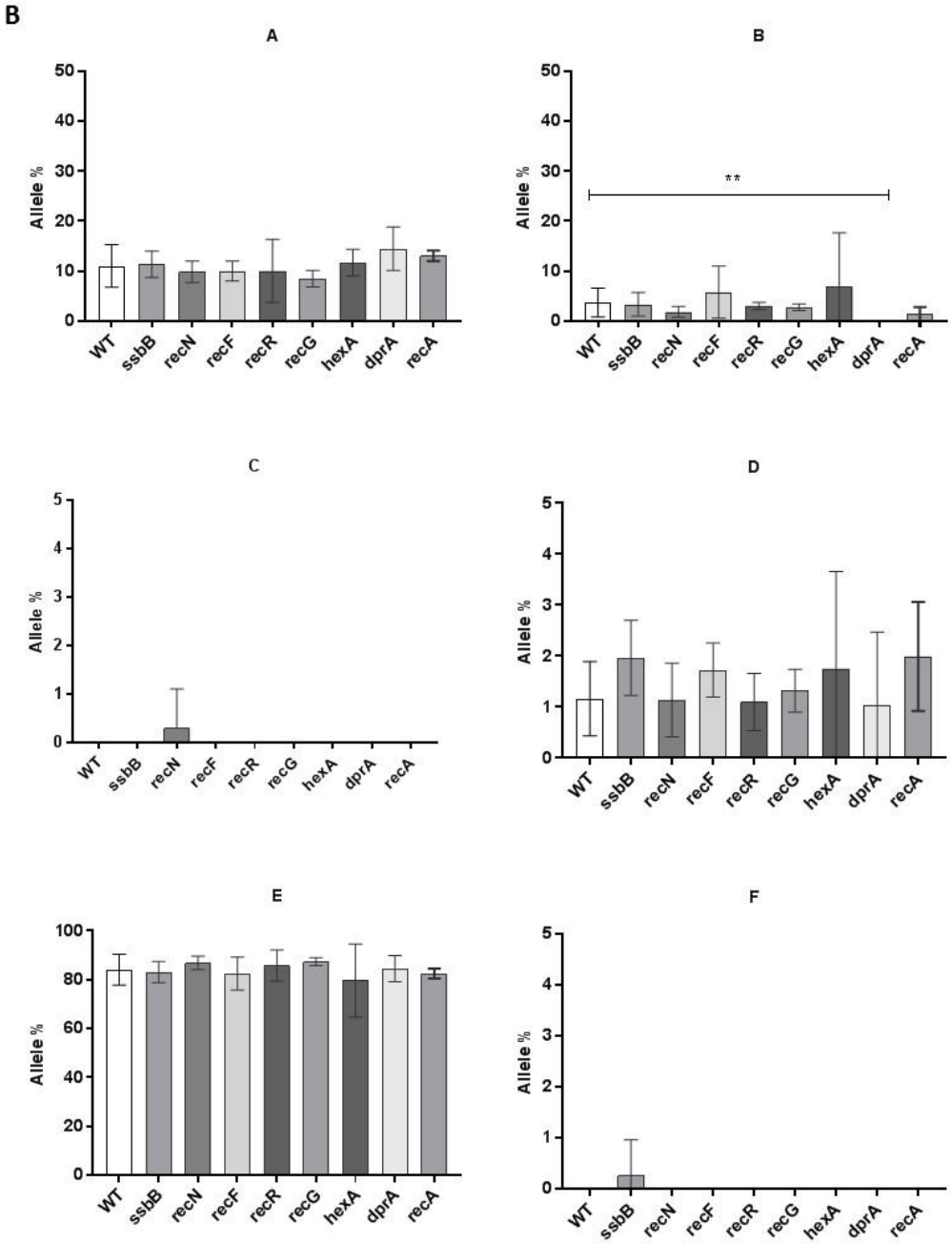


Figure 4.11 – Comparison of single colonies from homologous recombination mutants. *It was confirmed that recombination within the spnIII locus is not under the sole control of the CreX recombinase and is not influenced by any other functional site specific recombinase found within the S. pneumoniae D39 genome. Therefore, it was hypothesised that genes related to the integration of foreign DNA into the host genome could also facilitate spnIII recombination. In Panel Each bar represents a single colony, with each graph representing one mutant strain. Each colour represents a single spnIII variant (SpnIIIA (red), SpnIIIB (blue), SpnIIIC (grey), SpnIIID (white), SpnIIIE (yellow) and SpnIIIF (light blue)). In panel A all SpnIII variants generated by a single recombination event are present within the colonies tested, in all but one strain. The exception to this is found in the dprA knockout. In Panel A the ΔdprA strain generates colonies founded by both SpnIIIE and SpnIIIB, however these do not occur within the same colony. In Panel B, the median variant percentages of all colonies predominantly expressing SpnIIIIE were compared to the WT strain and the percentage of SpnIIIB in the dprA knockout strain (0.00±0.00) was found to be significantly lower than the WT (2.55±2.92), p=0.0014. No other significant differences were observed in any mutants (Table 4.6). Significance was tested using a Mann-Whitney test, *p≤0.05, **≤0.01, ***p≤0.001.*

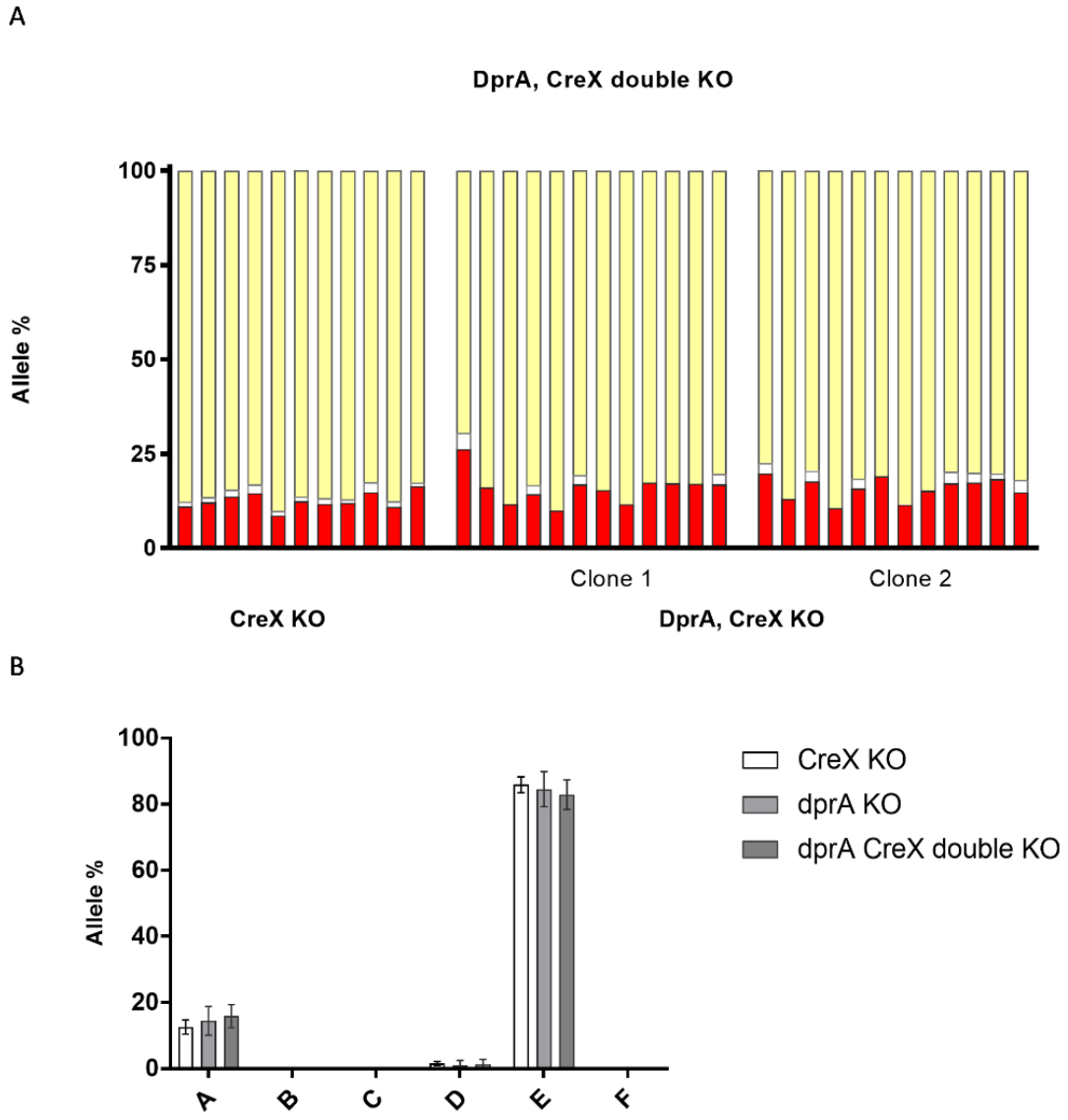


Figure 4.12 – Comparison of strains lacking creX and dprA. *Analyse of mutants demonstrated that both creX and dprA are involved in spnIII recombination, specifically abolishing recombination the 15bp inverted repeat. To determine whether these genes were capable of compensating for one another a double knockout of both genes was constructed. In Panel A, each bar represents a single colony and each colour represents a single SpnIII variant (SpnIIIA (red), SpnIIIB (blue), SpnIIIC (grey), SpnIIID (white), SpnIIIE (yellow) and SpnIIIF (light blue)). In panel A, a comparison of the parent strain (ΔcreX) and the ΔcreX ΔdprA double knockout show that both lack any recombination on the 15bp repeat, indicated by recombination from SpnIIIE (yellow) to SpnIIIB (blue). Statistical analysis of the median variant percentage (Panel B) confirmed there are no significant differences in allele percentage between the ΔcreX, ΔdprA and the double knockout strains. Confirming that, while both genes are involved in recombination on the 15bp repeat, neither is compensating for the other to facilitate recombination on the larger repeats. Significance was tested using a Mann-Whitney test, *p≤0.05, **≤0.01, ***p≤0.001.*

This result indicates an involvement of DprA in the recombination of the *spnIII* locus. However, only recombination on the 15bp repeat is significantly different, the median percentage of SpnIIIB within an SpnIIIE founded colony drops significantly from a median of 2.55±2.92 to 0.00±0.00 (p=0.001) (Table 4.6).

DprA does not appear to be involved in recombination on the 333bp (α) or 85bp (γ) repeats. Despite the lack of β repeat recombination within individual colonies SpnIIIB and SpnIIIE are found within the same liquid culture, as the stocks used to generate colonies for this analysis produced both SpnIIIB and SpnIIIE founded colonies (Fig 4.11A). This therefore suggests that the impact of DprA on *spnIII* recombination is associated with growth on plates but not in liquid media. In order to rule out any polar effects of cassette insertion at the *dprA* locus it was determined that *dprA* is not part of an operon and the presence of a terminator in the antibiotic resistance cassette used (104) means there are unlikely to be downstream effects from the expression of the cassette.

Table 4-6 - Statistical differences in SpnIII variant percentage for all homologous recombination mutants

SpnIII Variant	Median % (D39)	Median % (D39*ΔdprA*)	P Value (Mann Whitney Test)
A	10.3±4.2	14.5±4.3	NS
B	2.6±2.9	0.0±0.0	<0.01
C	0.0±0.0	0.0±0.0	NS
D	1.5±0.7	0.0±1.43	NS
E	85.2±6.2	84.2±5.3	NS
F	0.0±0.0	0.0±0.0	NS

4.5.5 Exploring the involvement of the DNA Processing Protein A (DprA)

A link between the expression of the DNA Processing Protein A (DprA) and *spnIII* recombination was identified through a *dprA* knockout strain. As it is not known when *spnIII* recombination occurs the involvement of DprA suggests a limited time frame (39) as the expression of DprA is known to only occur in a short time

frame following the induction of competence (39). Competence is initiated by the expression of the competence stimulating peptide (CSP) (37), which can come in one of two forms, CSP1 and CSP2. *S. pneumoniae* D39 encodes and responds to CSP1. It was therefore hypothesised that *spnIII* recombination is linked to the induction of the competence pathway.

In order to test this hypothesis, the gene encoding CSP1, *comC1,* was disrupted with a chloramphenicol cassette. A *comC1* knockout was already held within the lab in a DP1004 background. In order to generate comparable data, the cassette was PCR amplified from the DP1004 strain using primers MD130 and MD131 (See Table 2.3 for primer sequences) and transformed into D39. Expression of DprA is thought to be controlled by the expression of CSP1, therefore a *comC1* knockout should not express DprA and would be expected to display the same phenotype.

The *ΔcomC1* and *ΔdprA* strains were grown for 16 hours on agar plates. Single colonies were picked for analysis and all colonies predominantly expressing the same SpnIII variant were treated as replicas for statistical analysis. A total of ten *ΔcomC1* colonies were selected for analysis, all of which were founded by SpnIIIE expressing cells. The pattern observed in the *ΔdprA* strain was not observed in the *ΔcomC1* knockout strain as SpnIIIA, SpnIIIB and SpnIIID were all present within the same colony (Fig. 4.12A). It is thought that DprA is only expressed when cells become competent (39), however the pattern observed in the *ΔcomC1* strain suggests DprA expression without the induction of competence.

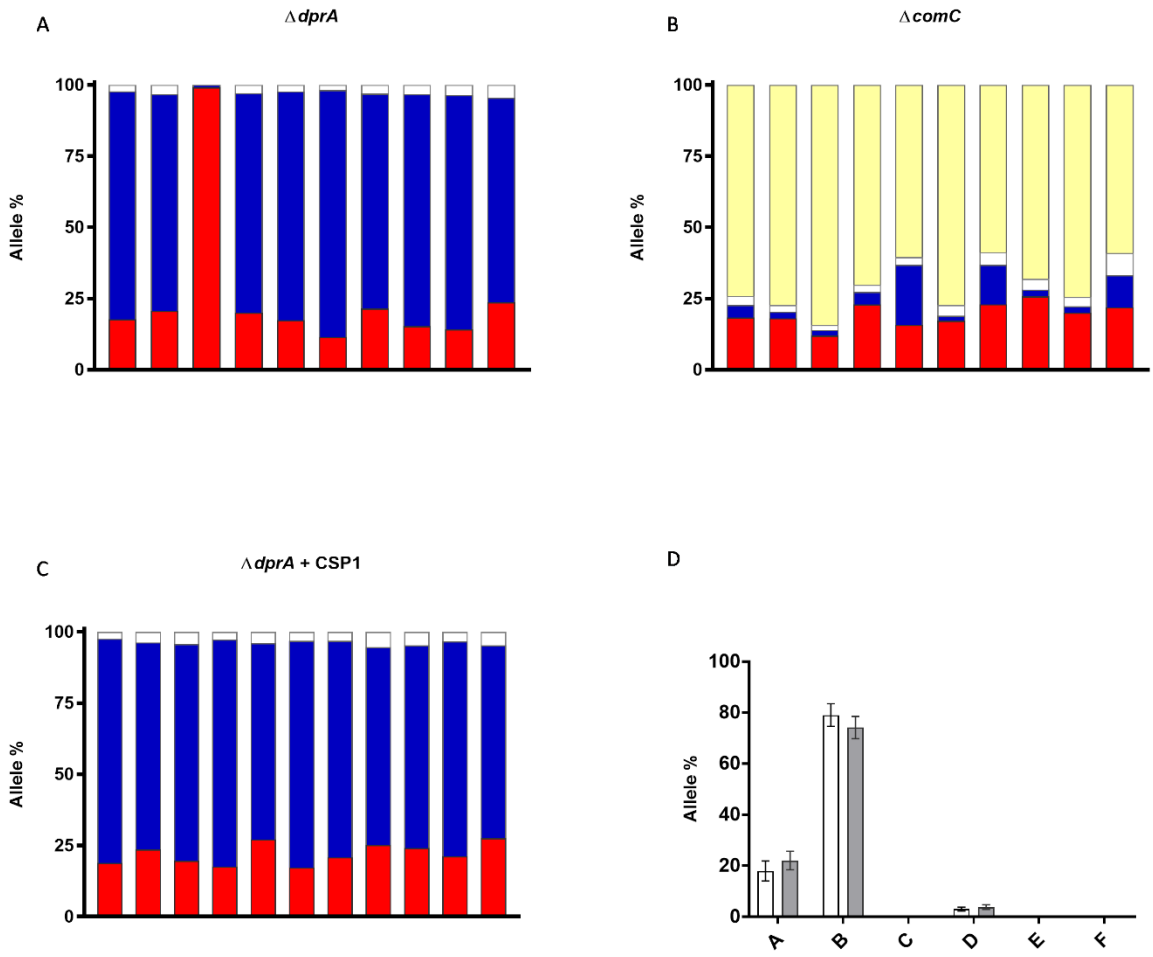


Figure 4.13 – Determining the link between competence and SpnIII recombination. *A ΔcomC1 strain was constructed in S. pneumoniae D39 and compared to the previously generated ΔdprA strain (Panels A and B). Each bar in panels A, B and C represents a single colony and each colour represents a single SpnIII variant (SpnIIIA (red), SpnIIIB (blue), SpnIIIC (grey), SpnIIID (white), SpnIIIE (yellow) and SpnIIIF (light blue)). In a ΔcomC1 strain (Panel B) SpnIIIE (yellow) and SpnIIIB (blue) are present within the same colony. This indicates that recombination on the 15bp repeat is not linked to the expression of competence stimulating peptide 1 (CSP1) in D39 but is linked to the expression of DprA (Panel A). The addition of CSP1 to the media failed to recover recombination on the 15bp repeat in single colonies of a DprA knockout (Panel C). In panel D, it can be seen that there are no significant differences in the median allele percentages in a ΔdprA strain with (grey bar) and without (white bar) stimulation of competence using CSP1. Significance was tested using a Mann-Whitney test, *p≤0.05, **≤0.01, ***p≤0.001.*

In order to determine if *spnIII* recombination is influenced by DprA expression or by another gene of the competence pathway the *ΔdprA* strain was incubated with synthetic CSP1 to induce competence. When stimulated with CSP1 colonies showed no 15bp (β) repeat recombination (Fig 4.13C). A comparison of median variant proportions (Fig. 4.12D) in the *ΔdprA* strain with and without the induction of competence showed no significant differences in the median variant percentages (Table 4.6).

.

4.6 **Chapter discussion**

The work of this chapter initially explored the hypothesis that a single site specific tyrosine recombinase (CreX) was responsible for DNA inversions in the phase variable *spnIII* locus. In *M. pulmonis* the deletion of the *hvsR* gene encoding a site specific recombinase prevents any DNA inversions at the *hsd* loci. The HvsR recombinase is unique as it facilitates inversions of DNA at two distinct and unrelated sites, the *hsd* and *vsa* loci (77). Within the *hsd* loci there are two pairs of inverted repeats, the vip repeat (5'-CAAAGTGCAATA-3') and the hrs repeat (5'-TAATTAAGATTATTGAACCT-3') (77). However, when cloned into *E. coli* only the longer hrs repeat is capable of facilitating recombination in the presence of HvsR, indicating that other *Mycoplasma* specific proteins are required for recombination on the shorter vip repeat. It is possible that CreX functions alongside another protein on the α and γ repeats but is essential to recombination on the β repeat due to its small size. It is possible that alternative proteins allow for CreX independent recombination on the larger α and γ repeats.

The exploration of a second SSR as the additional mechanism of recombination may appear counterintuitive. There are relatively few examples of multiple SSRs acting on the same DNA segment, however such systems have been identified (91). Of particular note is the Hin recombinase of *S. enterica,* where the Hin recombinase functions alongside a second SSR, Fin*,* to enable inversions of the H segment within the *fljB* promoter (91). Hin is located in the invertible region, while Fin is found elsewhere in the genome, interestingly they both bind the same *hix* site, although Hin is thought to bind with greater specificity (91). Despite this no other SSR in the D39 genome was found to significantly impact *spnIII*

recombination but it is important to note that Hin does not function alone. In addition to the two 26bp *hix* sites that flank the H segment and allow the recombinase to bind, the process of H segment inversion also requires the Fis protein to bind a 65bp enhancer sequence (91). Inversions at the *spnIII* locus are large and as a result it may be that an additional facilitator protein is required for the process.

As it has been established that no known SSR other than CreX is involved in *spnIII* recombination the proteins associated with homologous recombination have also been investigated. Many of these genes function as DNA loaders and catalysts for DNA branch migration (35,139).

Homologous recombination is the process by which breaks in DNA are repaired. In prokaryotes a dsDNA break results in the degradation of the broken DNA to produce ssDNA which is then bound and protected by the protein RecA (142). The formation of RecA filaments on ssDNA occurs though a tightly monitored system regulated by the RecBCD compex which, along with SsbB, loads RecA onto the DNA and initiates a search for a homologous region of DNA (142). In addition to the RecBCD pathway the RecFOR pathway initiates the RecA mediated repair of DNA strand gaps (rather than blunt end breaks) which can occur when replication forks stall (142). Until relatively recently it was thought that RecA was essential for all homologous recombination that was not facilitated via a site specific recombinase (142,143). However, it has more recently been proven that mechanisms of RecA independent, non-SSR, homologous recombination can also occur (144). In addition, RecA independent mechanisms can be as efficient as RecA mediated homologous recombination when exonuclease activity, such as that of RecJ, is suppressed (144).

Both RecA and SSR mediated recombination can occur via inverted repeat sequences of DNA (145). In this study recombination was found to be independent of RecA but occurring non-exclusively by the action of a single SSR, CreX. It is clear from the panel of mutants tested that the RecFOR homologous recombination pathway is not involved in *spnIII* recombination. The lack of an observable phenotype in the *hexA* and *recN* knockout mutants suggests the process is also independent of the mismatch repair machinery of the cell. Despite

the testing of many different mutants, the mechanisms behind recombination on the 333bp and 85bp repeats remain unsolved. It is possible that recombination on these repeats is occurring via RecA independent homologous recombination (144). If this is the case it is likely to occur infrequently unless the expression of exonucleases, such as RecJ, are suppressed. The frequency of recombination observed would suggest that recombination, at least on the larger α repeat, is not an infrequent occurrence therefore it may occur during a specific time period where exonucleases are suppressed.

The involvement of the RecA loader DprA in recombination of the 15bp (β) repeat of the locus was unexpected as expression of the protein is thought to be limited to the transient competence window (37). Despite this no evidence was found that induction of the competence pathway impacts *spnIII* recombination. It was however possible to link the loss of DprA to the loss of recombination on the 15bp repeat. In 2013 Mirouze et al. demonstrated that DprA functions beyond it's known role as the loader of RecA. DprA is now known to also interact with ComE to initiate the shutdown of the competence pathway (37). Analysis of *creX* and *dprA* mutants demonstrated that the loss of *creX* had a significant impact on recombination on all three repeats of the *spnIII* locus, while the loss of *dprA* only significantly impacted the smallest β repeat. It is possible that DprA may act as a loader of CreX, as it does for RecA, onto the β repeat but not the α and γ repeats. Further investigation into DNA loading proteins would be necessary to establish whether there are partner protein(s) associated with recombination on the α and γ repeats.

Further work to fully understand recombination at the *spnIII* locus is needed and is likely to include the following strategies; exploration of any remaining genes associated with DNA inversion and homologous recombination, a protein pulldown experiment to determine which proteins can be found bound to *spnIII* repeat regions and finally the transfer of the constructed mutants into an alternative genetic background with a higher rate of *spnIII* recombination. This final strategy will make determining smaller changes in recombination easier to observe.

5 *In vivo* models of infection with SpnIII wildtype strains

5.1 Introduction

Previous work with SpnIII phase-locked mutants revealed clear differences in their ability to cause invasive disease (64). In addition, *Spn*III has been linked to changes in the number of opaque and transparent colonies within a population. While there is some discrepancy in the exact impact of *Spn*III on colony morphology a clear link has been demonstrated in more than one study (64,65). Manso *et al.* reported that a strain locked into the SpnIIIB variant had an improved rate of nasopharyngeal colonisation but was significantly attenuated in invasive disease (64). In contrast a strain locked into the SpnIIIA variant failed to effectively colonise the nasopharynx but proved to be significantly more virulent in invasive disease models (64). The data presented in this chapter explore whether these differences are also observed in murine infection models using SpnIII WT strains and whether any variants are selected for *in vivo.*

In addition to the murine models of infection access was provided to samples from the Experimental Human Pneumococcal Carriage (EHPC) project (3). The current EHPC protocol intranasally challenges healthy human volunteers with the serotype 6B strain BHN418 (24). BHN418 is a carriage associated strain from the lab of Birgitta Henriques-Normark at the Karolinska Institutet, Stockholm, Sweden. The 10ml nasal washes used to confirm successful and ongoing colonisation provide real-time samples of the pneumococcus within the human nasopharynx. Samples from the EHPC provided the opportunity to analyse changes to SpnIII variant percentages over a period of colonisation, by comparing the recovered populations to a known input population.

5.2 Aim

The aim of this chapter was to investigate changes to SpnIII variant percentages *in vivo*. Using murine and human models of infection differences in both invasive disease and carriage have been explored.

5.3 Murine models of infection

Using the same SpnIII WT strains generated in chapter 3, intranasal, intravenous and intracisternal infections were conducted in mice. Based on the data from

phase-locked SpnIII mutant strains it was expected that the strain pre-dominantly expressing SpnIIIB would be significantly less virulent than other variants, but that this strain would be more successful in nasopharyngeal colonisation.

Two intranasal pneumonia models of infection were conducted using two different outbred strains of mouse, MF1 and CD1. All previous work was conducted in MF1 mice, however the supplier (Charles River, UK) opted to discontinue supply of this strain. As a result, the pneumonia model of infection was also conducted in CD1 mice. After no significant differences were detected between MF1 and CD1 mice in the pneumonia model of infection the intravenous experiments were conducted using CD1 mice. The intracisternal model of infection was conducted using MF1 mice at the Dipartimento di Biotechnologie Mediche, Universita di Siena, Siena, Italy.

5.3.1 **Pneumonia models of infection**

To determine the impact of SpnIII expression on the development of a pneumococcal lung infection, groups of mice were intranasally challenged with one of six *S. pneumoniae* WT strains generated and described in chapter 3. Comparable numbers of bacteria were used for all SpnIII WT strains, irrespective of mouse strain. 24 hours after inoculation, samples were collected from the blood, brain, lungs and nasopharynx. Lung and nasopharyngeal samples were collected by lavage.

No significant differences were observed in the number of bacteria recovered from the blood (Fig. 5.1A-B), lungs (Fig. 5.1C-D) and nasopharynx (Fig 5.1E-F) of either MF1 or CD1 mice challenged with different SpnIII WT strains. Brains were only collected from CD1 mice, again no significant differences were observed between mice challenged with different SpnIII WT strains (Fig. 5.1G). Colonisation of the nasopharynx was homogenous (Fig 5.1 E-F), however a trend for the SpnIIIB WT strain to fall below the limit of detection in the blood was observed in both CD1 and MF1 mice (Fig 5.1A-B).

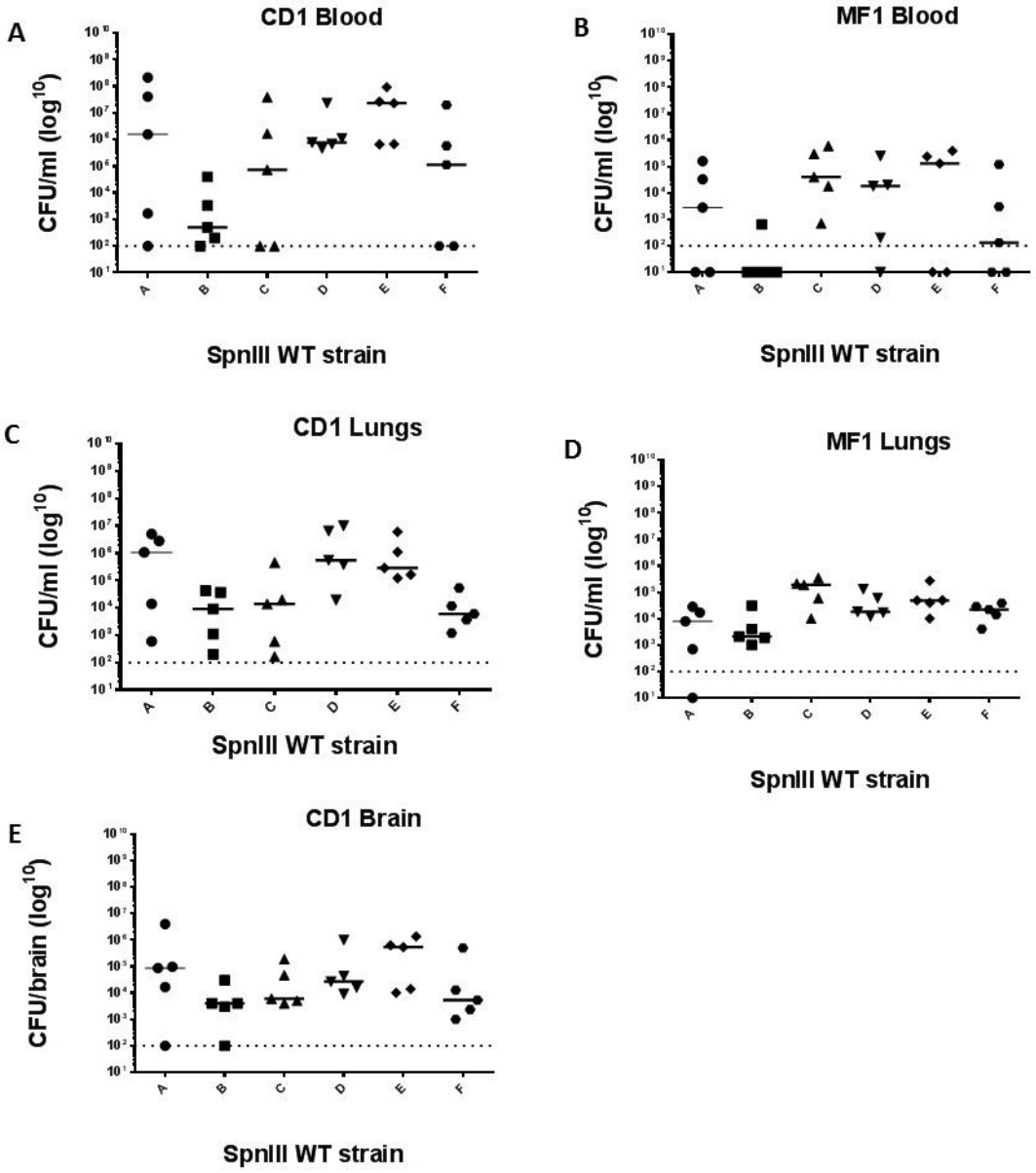


Figure 5.1 – Bacterial cell numbers recovered from the blood (A-B), lungs (C-D), nasopharynx (E-F) and Brain (G) of mice infected intranasally with S. pneumoniae SpnIII WT strains A-F. *When mice were intranasally challenged with WT strains SpnIIIA-F in a pneumonia model of infection no significant differences were observed in the number of bacteria recovered from the blood of CD1 (A) and MF1 (B) mice. Bacteria were collected from the lungs of both CD1 (C) and MF1 (D) mice by a lung lavage. Plating of the cells recovered by lavage showed no significant difference in CFU/ml. Finally, the brain from infected CD1 mice (E) was collected. All bacterial counts were obtained by 10 fold serial dilutions of samples. Significance was tested using a Kruskal-Wallis test, *p≤0.05, **≤0.01, ***p≤0.001.*

To determine whether the lack of significance difference was the result of a small sample size (n=5) the CFU/ml of bacteria recovered from each experimental group were analysed in both mouse strains using a Student's T test. No significant differences were found in the CFU/ml recovered from the blood or lungs of CD1 or MF1 mice challenged with the same SpnIII WT strain. As no significant differences were observed these data were pooled to increase the sample size (Fig 5.2). A Mann-Whitney test of the pooled data revealed that by increasing the sample size, it was possible to identify several significant differences in the number of bacteria recovered from both the lungs and the blood of intranasally infected mice (Table 5.1).

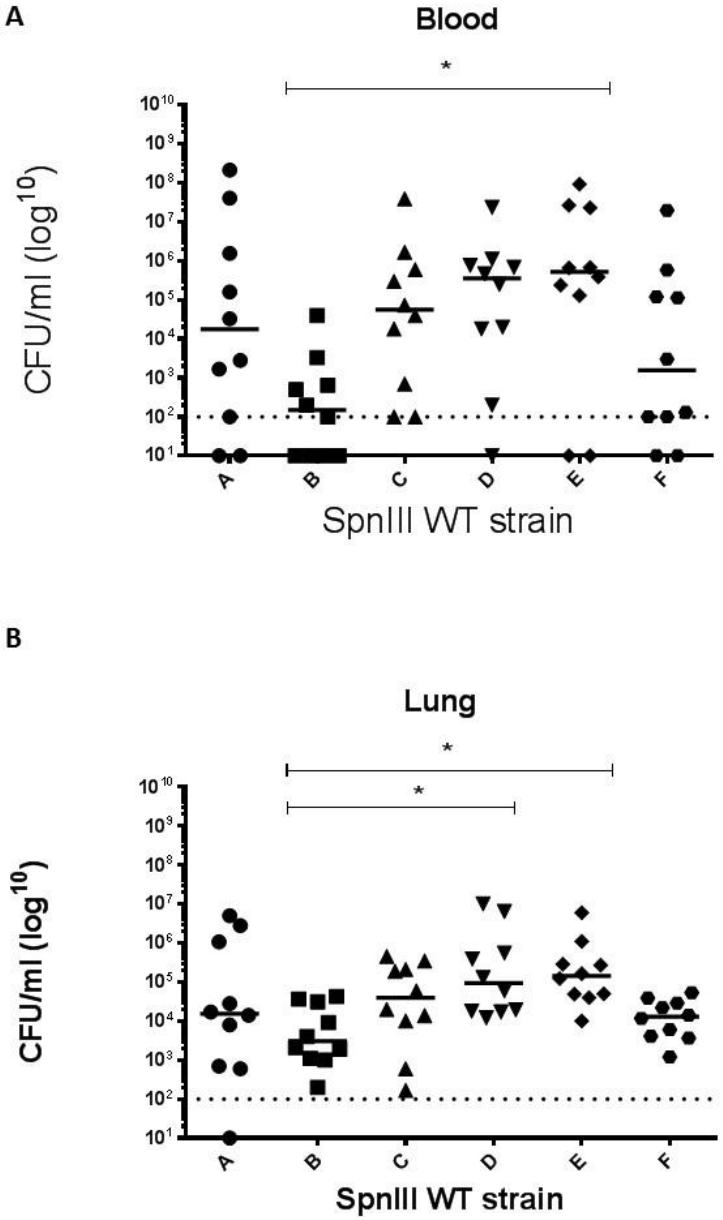


Figure 5.2 - Bacterial cell numbers recovered from the blood (A) and lungs (B) of CD1 and MF1 mice infected intranasally with S. pneumoniae SpnIII WT strains A-F. *The bacterial counts shown in Fig. 5.1A-D were pooled after the same non-significant trend was observed in each. The increased number of replicas improved the statistical power of the Mann-Whitney test and showed that there are significant differences between the SpnIII WT strains in both the blood (A) and lungs (B). In the blood (A) SpnIIIB populations show a significantly lower CFU/ml of recovered bacteria when compared to SpnIIIC, D and E (p=0.005, 0.005 and 0.008 respectively). In the lungs SpnIIIB survival is significantly lower than SpnIIID (p=0.003), and SpnIIIE (p<0.001), as is SpnIIIF survival compared to SpnIIID (p=0.0115) and SpnIIIE (p=0.0011).* Significance was tested using a Kruskal-Wallis test, *p≤0.05, **≤0.01, ***p≤0.001.

Intranasal infection with the SpnIIIB wildtype strain resulted in a significantly lower CFU/ml of blood when compared to SpnIIIC (p=0.005), SpnIIID (p=0.005) and SpnIIIE (p=0.008). SpnIIIB was not significantly different from SpnIIIA (p=0.074) and SpnIIIF (p=0.218) (Fig 5.2).

Significant differences were also observed in the lungs. Once again the CFU/ml (of lung lavage) was significantly lower when mice were infected with the SpnIIIB WT strain when compared to SpnIIID (p=0.003) and SpnIIIE (p<0.001) WT strains. In addition to these differences SpnIIIF was also found to have significantly lower levels of bacteria in the lung when compared to SpnIIID (p=0.012) and SpnIIIE (p=0.001) (Fig 5.2).

Table 5-1 – Average CFU/ml in the blood and lungs of mice intranasally challenged with SpnIII WT strains.

	Mean CFU/ml Blood^	Mean CFU/ml Lungs^
A	$2.6 \times 10^7 \pm 6.7 \times 10^7$	$9.0 \times 10^5 \pm 1.7 \times 10^6$
B	$4.5 \times 10^3 \pm 1.3 \times 10^4$	$1.3 \times 10^4 \pm 1.7 \times 10^4$
C	$4.2 \times 10^6 \pm 1.2 \times 10^7$	$1.3 \times 10^5 \pm 1.6 \times 10^5$
D	$2.7 \times 10^6 \pm 7.4 \times 10^6$	$1.8 \times 10^6 \pm 3.5 \times 10^6$
E	$1.5 \times 10^7 \pm 3.0 \times 10^7$	$8.1 \times 10^5 \pm 1.9 \times 10^6$
F	$2.1 \times 10^6 \pm 6.3 \times 10^6$	$1.9 \times 10^4 \pm 1.7 \times 10^4$

^Mean calculated from 10 mice per group (5 CD1, 5 MF1).

Brains were not collected from MF1 mice, however in CD1 mice similar, non-significant trends were observed (Fig. 5.1G). SpnIIIB and SpnIIIF showed the lowest average CFU/brain which is in line with the significant differences seen in the blood and lungs. It is likely this trend is a reflection of each strain's ability to survive in the blood.

Significant differences were observed in bacteria collected from the lungs and blood of mice when compared to the initial inoculum. Despite this, in a pneumonia model of infection it appears that little selective pressure is exerted on SpnIII WT strains in the nasopharynx.

5.3.2 **Sepsis model of infection**

To determine the impact of SpnIII expression on the development of sepsis, groups of CD1 mice were intravenously challenged with one of six *S. pneumoniae* WT strains generated and described in chapter 3. Using the same stocks prepared for the pneumonia model described in section 5.3.1 comparable numbers of bacteria were used for intravenous inoculation. 18 hours after inoculation, samples were collected from the blood, spleen and brain and bacterial load was determined by 10 fold serial dilution.

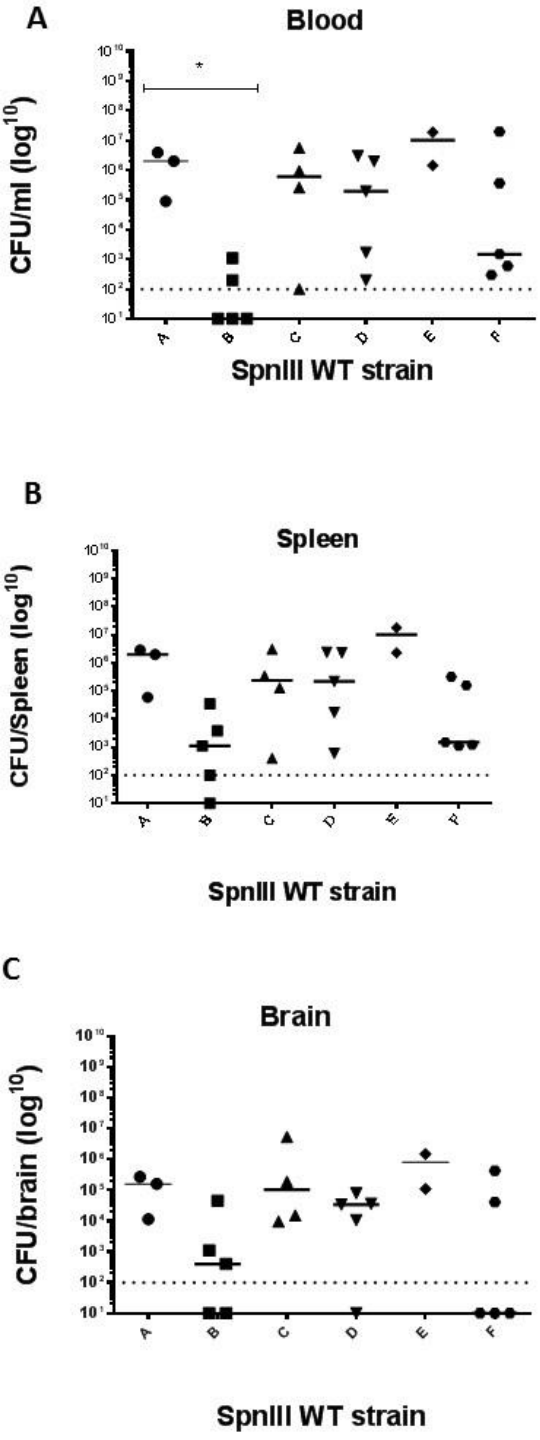


Figure 5.3 – Bacterial cell numbers recovered from the blood (A), spleen (B), and brain (C) of mice infected intravenously with S. pneumoniae SpnIII WT strains A-F. *When mice were intravenously challenged with WT strains SpnIIIA-F in a sepsis model of infection significant differences in the number of bacteria recovered from the blood (A), spleen (B) and brain (C) were observed. Of particular note is the significantly decreased number of colonies recovered from the blood, spleen and brain when mice are infected with an SpnIIIB variant. All CFU counts were obtained by 10fold serial dilutions of samples. Significance was tested using a Kruskal-Wallis test, *p≤0.05, **≤0.01, ***p≤0.001.*

Following completion of the experiment several mice were excluded from the data due to an unexpectedly rapid disease progression. Following intravenous challenge with the SpnIII WT strains six mice were found dead before the 18-hour end point of the experiment. These mice were spread across experimental groups as follows: SpnIIIA (two mice), SpnIIIC (one mouse), and SpnIIIE (three mice). This outcome could not have been anticipated but protocol adjustments have been made to prevent future re-occurrences. Samples were collected from the animals found dead and plated to determine the bacterial load in the blood, spleen and brain. The number of bacteria recovered from these mice was significantly higher than all others within the experiment in the blood, spleen and brain. As it could not be determined whether the high bacterial load was the cause of death or if bacteria continued to increase in number following death these mice were excluded from statistical analysis.

It was confirmed that all SpnIII WT doses were within the range 1.1-2.1 x 10^6, although it should be noted that the SpnIIIA and SpnIIIE groups did receive a slightly increased dose (table 5.2). The SpnIIIC group received a challenge dose comparable to that of SpnIIIB, D and F. This is therefore unlikely to be the cause of the unexpectedly fast disease progression.

Table 5-2 – Challenge dose for mice intravenously challenged with spnIII WT strains.

SpnIII wildtype variant	Dose (CFU)
SpnIIIA	2.1 x 10^6
SpnIIIB	1.1 x 10^6
SpnIIIC	1.1 x 10^6
SpnIIID	1.2 x 10^6
SpnIIIE	2.3 x 10^6
SpnIIIF	1.3 x 10^6

The number of bacteria recovered from the blood of mice challenged with the SpnIIIA and SpnIIIB WT strains were significantly different, with significantly fewer bacteria recovered from the SpnIIIB group (p=0.047). The loss of several mice from some experimental groups makes them too small to draw conclusions, for example the SpnIIIE infected group was reduced to n=2. It is therefore difficult to draw any additional conclusions from these data. It can however be noted that

the trend for SpnIIIB and SpnIIIF wildtype strains to result in a lower a bacterial load, as seen in the pneumonia model of infection, is also present in these data.

The only experimental groups with a bacterial load in the brain below the limit of detection (100 CFU) were those challenged with SpnIIIB and SpnIIIF WT strains and is likely a reflection of the reduced bacterial load in the blood.

Overall these results re-confirm trends seen in the pneumonia model of infection (Fig. 5.1). The SpnIIIB and SpnIIIF WT strains result in a lower bacterial load across multiple organs. The unexpected loss of six mice from the experimental group resulted in a decreased statistical power meaning few conclusions can be drawn. It does however appear that SpnIIIA and SpnIIIE are more capable of causing invasive disease, although additional experiments to increase the number of mice in each group are required to confirm this result.

5.3.3 **Meningitis model of infection**

To determine the impact of SpnIII expression on the development of meningitis, groups of CD1 mice were intracisternally challenged with one of six *S. pneumoniae* WT strains generated and described in chapter 3. A single meningitis model of infection was conducted in MF1 mice using the same prepared stocks of the *S. pneumoniae* WT strains as in the pneumonia and IV models. 18 hours after inoculation, samples were collected from the blood, spleen and brain and bacterial load was determined by 10 fold serial dilution.

Several difference were observed in the CFU/ml of bacteria recovered from the blood of mice intracisternally challenged with different SpnIII WT strains (Fig 5.4). Challenge with the SpnIIIB WT strain resulted in a significantly lower CFU/ml in the blood when compared to all other WT strains ($p<0.01$ for all) (Fig 5.4A). This pattern was also observed in the spleen (Fig 5.4B). Only 4 of the 8 mice challenged with the SpnIIIB WT strain had spleen counts above the limit of detection (100 CFU). Spleen counts were significantly greater for all other WT strains when compared to SpnIIIB ($p<0.01$ for all).

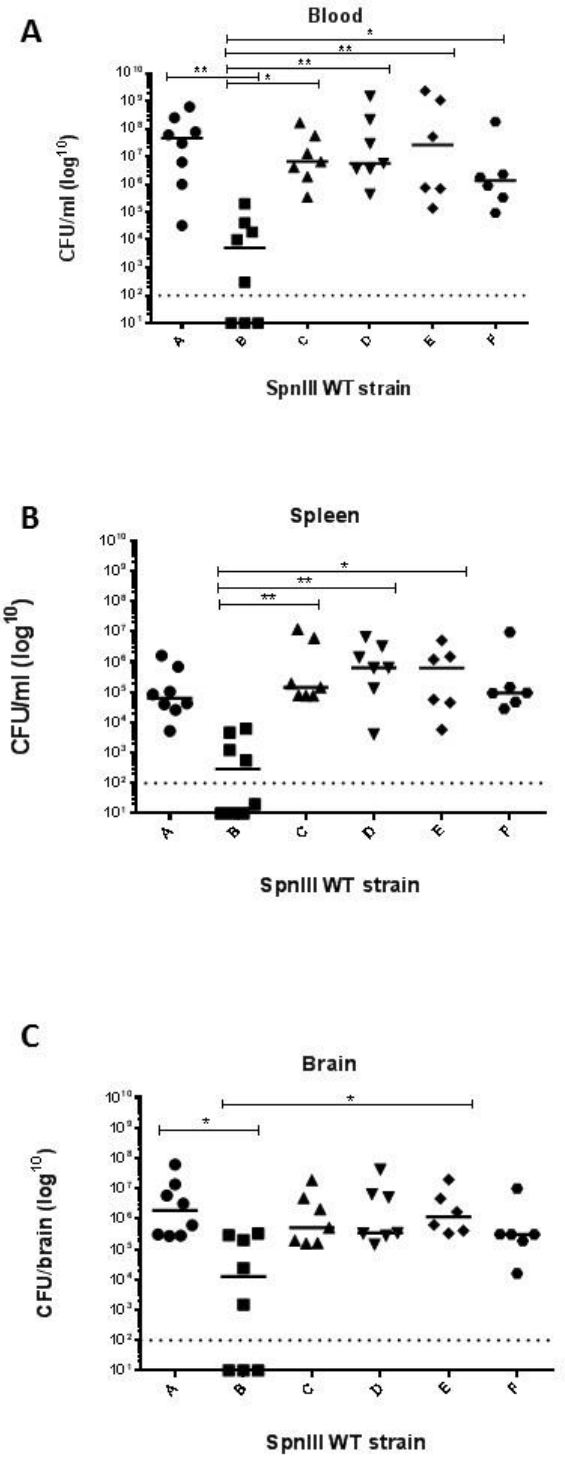


Figure 5.4 – Bacterial cell numbers recovered from the blood (A), spleen (B), and brain (C) of mice infected intracisternally with S. pneumoniae SpnIII WT strains A-F. *When mice were intracisternally challenged with WT strains SpnIIIA-F in a meningitis model of infection significant differences in the number of bacteria recovered from the blood (A), spleen (B) and brain (C) were observed. Significantly fewer bacteria were recovered the blood, spleen and brain of mice challenged with the SpnIIIB WT strain indicating this strain has a reduced capacity for invasive disease. All CFU counts were obtained by 10 fold serial dilutions of samples. Significance was tested using a Kruskal-Walllis test, *p≤0.05, **≤0.01, ***p≤0.001.*

In the brain SpnIIIA, C, D and E WT strains had significantly higher bacterial counts when compared to SpnIIIB ($p<0.05$ for all). SpnIIIF counts were not significantly different from SpnIIIB in the brain. This is an interesting result considering the dose used and route of infection in this model. It is possible that with a lower challenge dose the bacterial load in the blood and spleen for the SpnIIIF wildtype strain would no longer be significantly different from the SpnIIIB strain.

Overall, these results confirm once again that strains predominantly expressing the SpnIIIB variant are impaired during invasive disease. Survival of this strain is significantly lower than all other SpnIII WT strains in the blood, spleen and brain of intracisternally challenged mice. Interestingly the difference previously observed with SpnIIIF (significantly in Fig. 5.2B and as a trend in the sepsis model) is not as apparent here. The bacterial load for SpnIIIF infected mice shows little difference to the other strains (excluding SpnIIIB in the brain).

5.3.4 Overall trends in murine infection models

The infection models conducted demonstrate that there are significant differences in the virulence of strains predominantly expressing different SpnIII specificity subunits. Strains with a high percentage of SpnIIIB are significantly impaired during invasive disease, but colonise the nasopharynx at a similar rate to all other SpnIII WT variants. SpnIIIB cells are routinely cleared from the blood, spleen and brain regardless of route of infection. In addition, SpnIIIF strains also appear to be disadvantaged in some situations, although not as clearly as the SpnIIIB variant. It is possible that SpnIIIA and E expressing strains have an increased capacity for causing invasive disease, however further experiments would be required to confirm this result.

5.4 Trends in SpnIII in bacteria recovered from model infections

To further analyse samples collected from mice intranasally, intravenously and intracisternally challenged with the different *S. pneumoniae* SpnIII WT strains, every sample with recoverable bacteria, i.e. above the limit of detection, was subjected to allele quantification (see method 2.5). This allows for a comparison between the SpnIII proportions of the initial input population (challenge dose) and

the cells recovered. As the SpnIIIB variant, and in some cases the SpnIIIF variant, have been identified as significantly impaired in invasive disease models (Fig 5.2, 5.3 and 5.4) it is predicted that there could be ongoing selection against these variants (SpnIIIB and F) *in vivo.*

In order to identify overall trends results have been separated by SpnIII variant rather than challenge dose. For example, SpnIIIA is found in the SpnIIIA challenge dose, however it also forms approximately 20% of the SpnIIIE challenge dose. This allows for trends to be investigated in multiple populations, reducing the potential for SNPs in a single lineage to impact the result. As WT strains have been selected from single colony isolates any non-dominant SpnIII variants within that population are newly generated and independent of the other WT strains. Using this approach allows for analysis of genuine trends associated with SpnIII *hsdS* expression. In addition, samples have been separated by organ of origin as it is expected that there could be differences between environments, e.g., the blood and nasopharynx.

5.4.1 **Trends in the nasopharynx**

Based on the lack of significant differences in the number of bacteria recovered from the nasopharynx of mice challenged with different SpnIII WT strains (Fig. 5.1E-F) it was expected that little change in SpnIII variant percentages would be observed when the recovered cells were compared to the initial inoculum. In the absence of selection, a population would be expected to increase in less prevalent variants at the expense of the pre-dominant variant through the process of random recombination. The actual result was unexpected as the most prevalent variant in each population was observed to increase (Fig. 5.5). This is not limited by SpnIII variant and occurs in all populations. An increase would be expected when there is a selective advantage, however it's occurrence in all populations indicates that this is not the case. The potential mechanisms behind this unexpected result will be discussed later in this chapter, however it should be noted that while the nasopharynx is taken to represent pneumococcal carriage in this incidence the model used is designed for bacteria to enter the lungs. A genuine model of carriage, when bacteria are not intended to enter the lungs, could result in a different outcome.

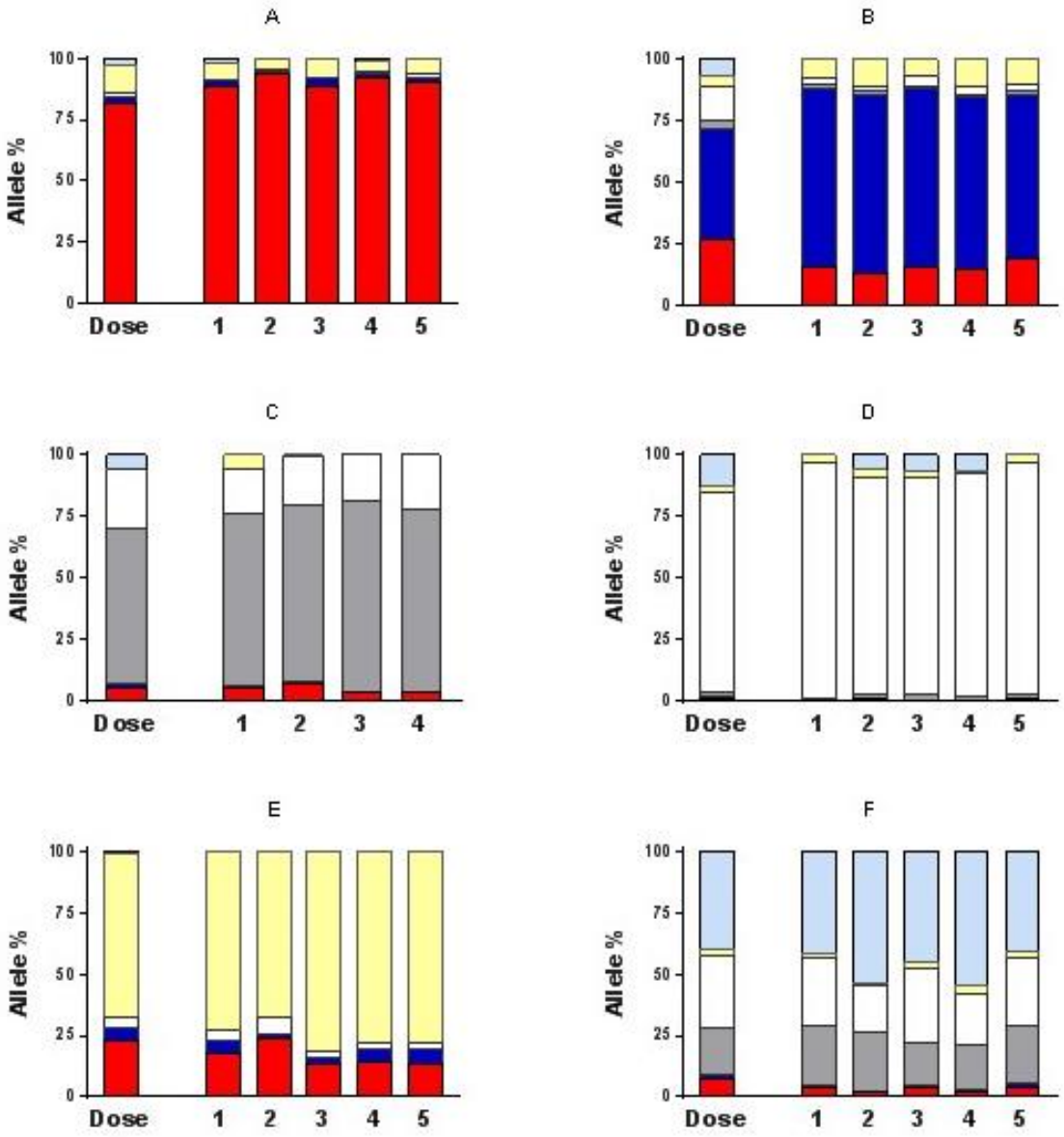


Figure 5.5 – SpnIII population trends within all nasopharyngeal samples obtained from a pneumonia model of infection in MF1 mice. *Each bar represents a single mouse and the colours represent the SpnIII variants detected in live colonies recovered from that mouse (SpnIIIA red, SpnIIIB blue, SpnIIIC grey, SpnIIID white, SpnIIIE yellow and SpnIIIF light blue). Each graph represents a group of mice infected with the same D39 SpnIII WT strain. For each group the initial inoculum is shown (dose).*

To confirm that the unexpected increase in the most prevalent SpnIII variant was not the result of an error in quantification of the initial challenge dose each was quantified at two different concentrations of cells (Fig 5.6A) and in triplicate on two separate occasions (Fig 5.6B). No significant differences were found between any allele percentage on cells plated at either 10^{-3} or 10^{-5}. In addition, no significant differences were found between repeat analysis conducted on separate days (Fig 5.6B). It was therefore concluded that the increase in prevalent SpnIII variants in the nasopharynx was not the result of incorrect quantification of the initial challenge dose.

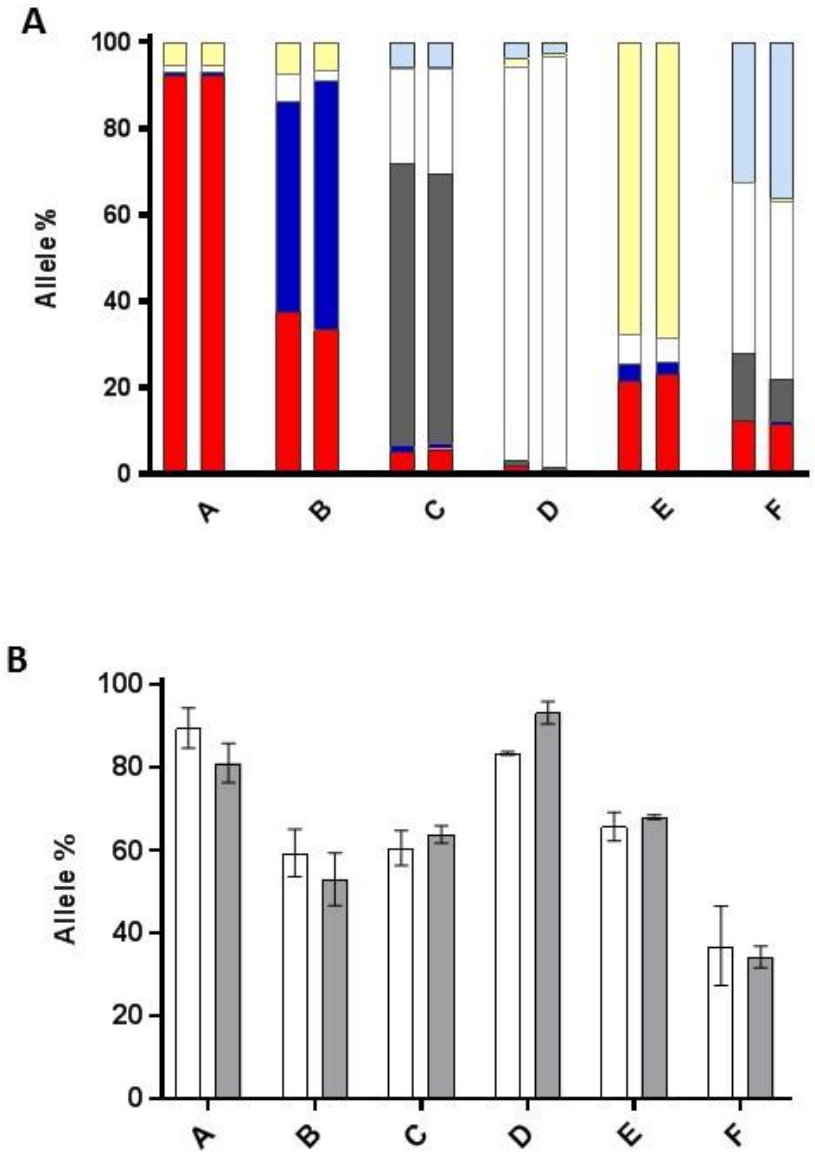


Figure 5.6 – Comparison of allele quantification analysis for SpnIII wildtype strains at different concentrations of cells and different time points. *Each SpnIII WT population used as a challenge dose for in vivo animal infections was quantified multiple times to ensure the reliability of the method and to determine the SpnIII proportions within each stock used. In panel A there is a comparison of 10µl of plated cells from the same stock analysed at two different concentrations of cells (10^{-3} and 10^{-5} CFU/ml). While some small changes are observed these are not statistically significant and no SpnIII variant observed in one dilution was absent in another. It was determined that analysing cells at either concentration was appropriate. Panel B shows the variation observed in allele quantification analysis of the same stock conducted in triplicate on two separate occasions. Only the dominant allele is shown, however no statistically significant differences were observed in the quantification of any SpnIII variant across the repeats regardless of proportion within a population. Significance was tested using a Mann-Whitney test, *p≤0.05, **≤0.01, ***p≤0.001.*

5.4.2 **Trends in the lung**

Trends in SpnIII variant percentages in samples collected from the lungs were analysed using the same approach as described for the nasopharynx (Fig 5.7). SpnIII analysis on bacteria recovered from the lungs showed outcomes were highly variable. The same SpnIII variant may increase in one mouse while decreasing in another. There appears to be a generalised trend for SpnIIIB, C, E and F to increase when they form a small proportion of the population however this is not true for SpnIIIA and D. In one mouse a sharp increase in SpnIIID can be seen at the expense of SpnIIIC. This change is not an overall trend, it is therefore likely to be caused by a significant bottlenecking of the population in an individual mouse. Overall these data indicate that there is no generalised trend relating to SpnIII selection within the lung. No single variant is selected over others despite differences in the number of bacteria recovered from the different infection groups (Fig. 5.2).

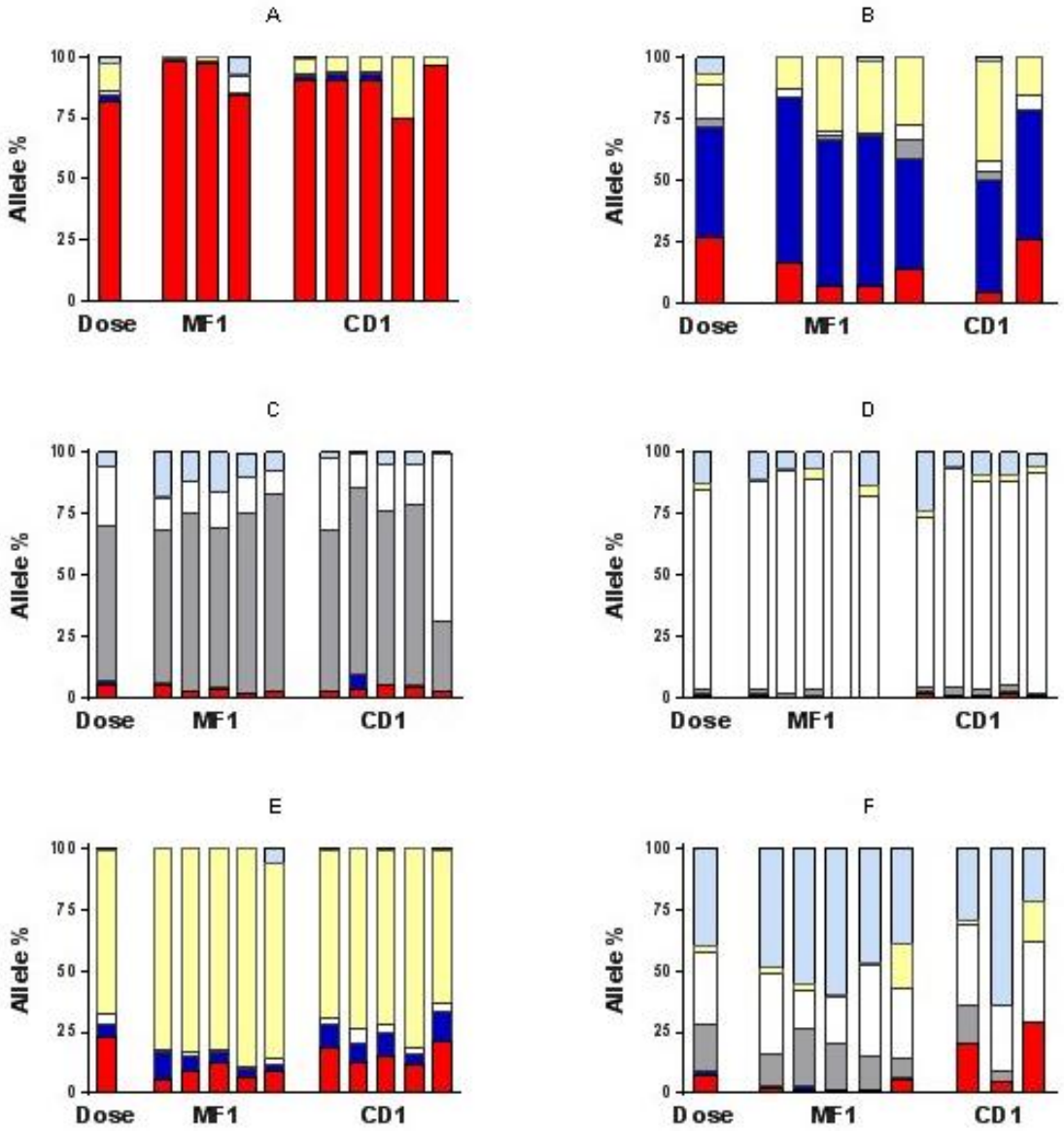


Figure 5.7 – SpnIII population trends within all lung samples obtained from a pneumonia model of infection in MF1 and CD1 mice. *Each bar represents a single mouse and colours represent the SpnIII variants detected in live colonies recovered from that mouse (SpnIIIA red, SpnIIIB blue, SpnIIIC grey, SpnIIID white, SpnIIIE yellow and SpnIIIF light blue). Each graph represents a group of mice infected with the same D39 SpnIII WT strain. For each group the initial inoculum is shown (dose). The outcome of populations is varied, without a specific trend observed, and is the result of the random generation of diversity in individual mice.*

5.4.3 **Trends in the blood**

For all experiments conducted blood was collected, these samples therefore provided the largest set for data for analysis of population dynamics *in vivo* (Fig 5.8). Bacteria recovered from mice challenged with SpnIIIA, C, D and E WT strains showed a high level of variation, with no consistent pattern. In the case of the SpnIIIB strain there are considerably fewer samples to analyse as the method relies on the recovery of live bacteria. It is therefore difficult to draw conclusions about selection against the SpnIIIB population in the blood. It was observed that SpnIIIB does survive as a small percentage of other WT populations, although these numbers are small. There are two potential explanations for this. Firstly it is possible there is an unidentified SNP in the SpnIIIB WT strain. Alternatively there may a critical threshold beyond which expression of SpnIIIB becomes detrimental to the overall population in the blood.

Despite being the least represented of all predominant alleles, SpnIIIF does not show the same spread of outcomes in individual mice. An SpnIIIF percentage >40% is observed in recovered cells from a single mouse. SpnIIIA and D have a tendency to decrease when they are the predominant variant expressed, however this is likely a result of the initial starting population being such a high percentage, there is little room a percentage increase in these samples. With the exception of SpnIIIB and SpnIIIF there is little difference observed in population trends in the blood compared to the lungs.

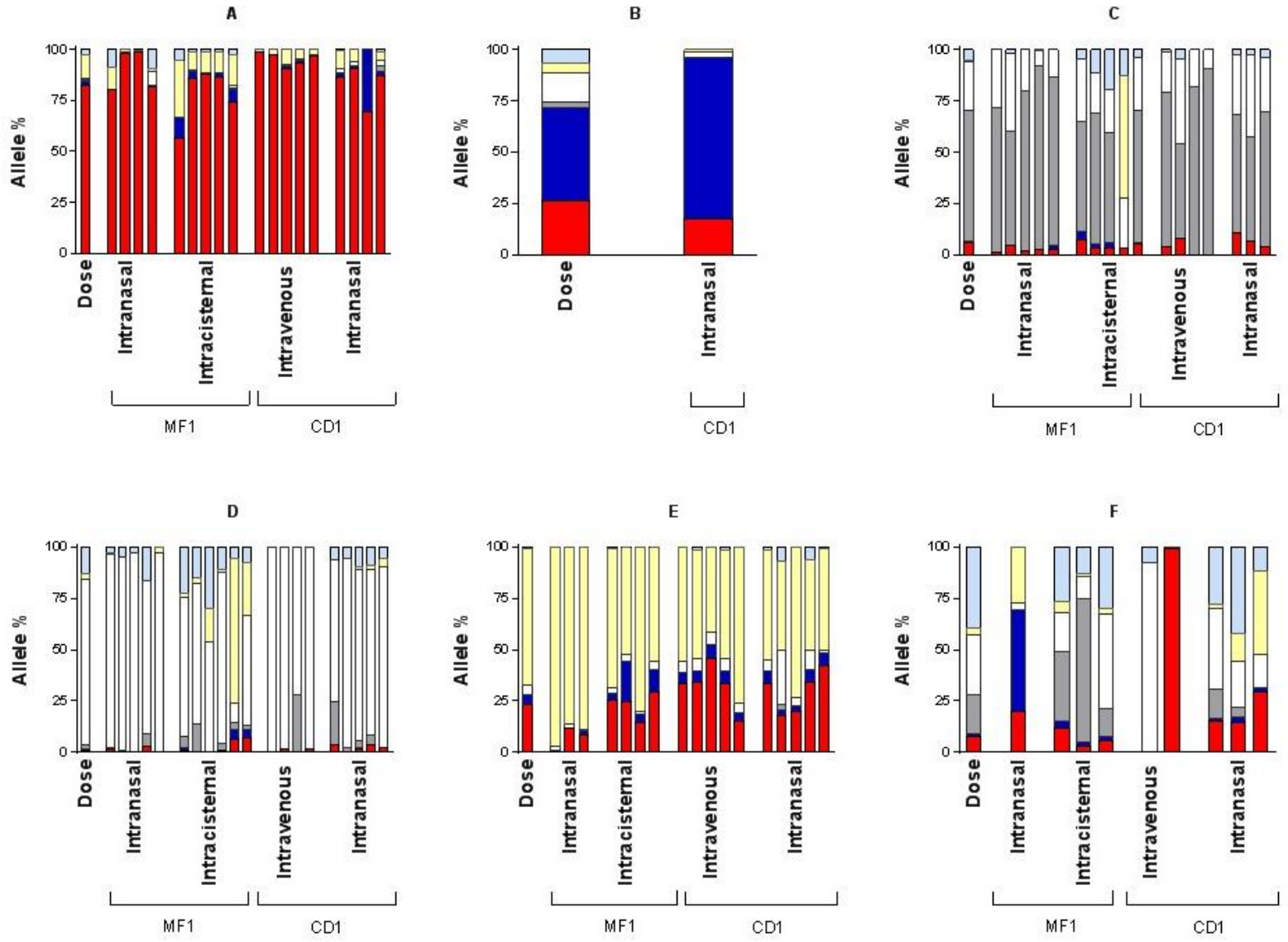

A
B
C
D
E
F
Allele %
100
75
50
25
0
Dose
Intranasal
Intracisternal
Intravenous
Intranasal
MF1
CD1

Figure 5.8 – SpnIII population trends within all blood samples obtained from mouse models of infection. *Each bar represents a single mouse and colours represent the SpnIII variants detected in live colonies recovered from the blood of that mouse (SpnIIIA red, SpnIIIB blue, SpnIIIC grey, SpnIIID white, SpnIIIE yellow and SpnIIIF light blue). Each graph represents a group of mice infected with the same D39 SpnIII WT strain. For each group the initial inoculum is shown (dose). The outcome of populations is varied, with less represented variants increasing in prevalence, e.g. an increase of the A% in mice infected with the WT E strain. SpnIIIB contains considerably fewer samples as the majority of blood samples did not have recoverable bacteria for analysis, however CFU counts indicate this strain is less capable of invasive disease. Very little SpnIIIF is observed, even in mice challenged with the WT F strain, suggesting this variant is also at a disadvantage in the blood.*

5.4.4 **Trends in the spleen**

In addition to analysis of the bacteria recovered from the blood, those recovered from the spleen were also analysed and the population dynamics investigated (Fig 5.9). Once again the same problem can be seen with SpnIIIB, there no samples with live bacteria making analysis impossible in this population. The indication from the number of recovered bacteria is that once again the SpnIIIB WT strain is not capable of surviving in the blood or spleen. There are fewer spleen samples compared to blood but overall a similar picture can be seen in Fig. 5.8 and Fig 5.9. High proportion samples show a wide variety of outcomes, with no generalised trend for any to increase or decrease except in the case of SpnIIIF. Once again this variant can be seen to decrease when it forms the majority of the population used to challenge the animal. There are individual cases when a significant change can be observed in a single mouse, however these do not indicate an overall trend and are likely the result of random bottlenecking events. A bottleneck event would produce a similar outcome to that seen in the generation of the SpnIII WT strains (Chapter 3). The populations is effectively reset to 100% of a single SpnIII variant until sufficient generations have passed to allow the regeneration of diversity through recombination. An example of this type of bottleneck event can be seen in a mouse challenged with an SpnIIIF WT strain where the *S. pneumoniae* population recovered from the spleen was >90% SpnIIID. A similar occurrence can be seen in a second SpnIIIF infected mouse, with the recovered population containing 99% SpnIIIA, while in the challenge dose the SpnIIIA variant made up only 12% of the population.

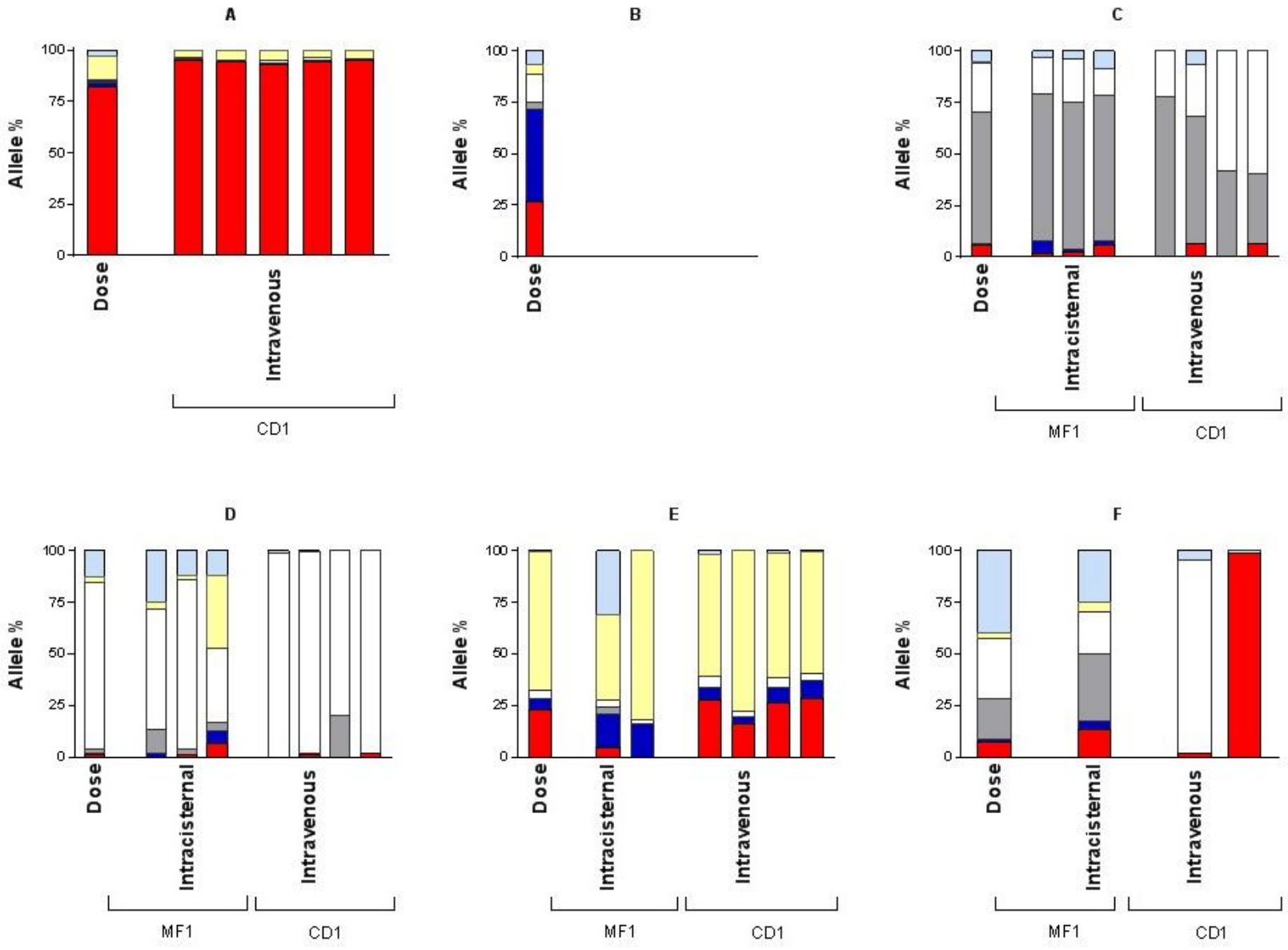

A
B
C
D
E
F
Allele %
100
75
50
25
0
Dose
Intravenous
Intracisternal
CD1
MF1

Figure 5.9 – SpnIII population trends within all spleen samples obtained from mouse models of infection. *Each bar represents a single mouse and colours represent the SpnIII variants detected in live colonies recovered from the spleen of that mouse (SpnIIIA red, SpnIIIB blue, SpnIIIC grey, SpnIIID white, SpnIIIE yellow and SpnIIIF light blue). Each graph represents a group of mice infected with the same D39 SpnIII WT strain. For each group the initial inoculum is shown (dose).The outcome of populations is varied, with SpnIIIF appearing to decrease in the majority of samples (as is also seen in the blood), suggesting a reduced ability to survive within the host. SpnIIIB contains considerably fewer samples, as the majority of spleen samples did not have recoverable bacteria for analysis, however CFU counts indicate this strain is less capable of invasive disease. The wide variations in some samples (e.g., SpnIIID) indicate bottlenecks in individual mice, as these follow no trend it is unlikely this is the result of selection.*

5.4.5 Trends in the brain

Brain samples from the intracisternal model were collected and analysed however some similar problems to the analysis of blood and spleen samples were encountered. There are only two samples from the SpnIIIB WT strain available as the remaining six samples contained no live bacteria at time of analysis. However, the percentage of SpnIIIB did increase in the two available samples. The prevalent alleles decreased in the SpnIIIA, D, E and F populations (Fig 5.10). While there is a possible trend for an increase in SpnIIIC within the brain, there are too few samples to confirm the significance of this result. The intracisternal model of infection is unique in that it is unlikely to create bottlenecks as bacteria directly enter the CFS. It is therefore likely that increases and decreases within these data are more representative of true selection when compared to samples from the lungs, spleen and blood.

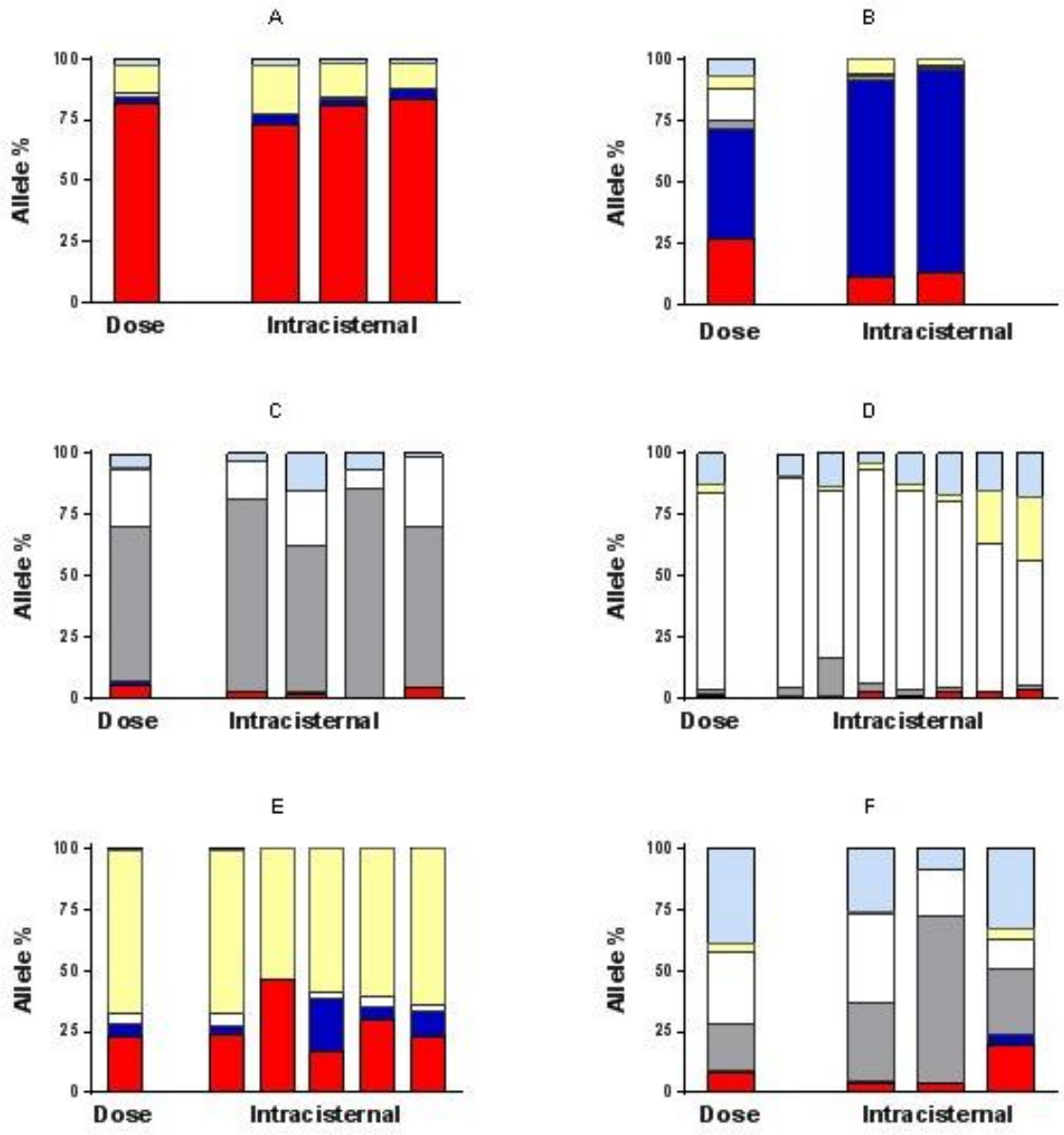


Figure 5.10 – SpnIII population trends within all brain samples obtained from an intracisternal model of infection. *Each bar represents a single mouse and colours represent the SpnIII variants detected in live colonies recovered from the brain of that mouse (SpnIIIA red, SpnIIIB blue, SpnIIIC grey, SpnIIID white, SpnIIIE yellow and SpnIIIF light blue). Each graph represents a group of mice infected with the same D39 SpnIII WT strain. For each group the initial inoculum is shown (dose). The number of samples shown is dependent on how many samples contained live bacteria.*

5.4.6 Overall trends in SpnIII quantification of recovered bacteria

Despite the clear differences observed in bacterial load between the strains there was no observable selection for or against cells expressing particular SpnIII RM specificities. The SpnIIIB variant proved problematic to analyse due to the low number of samples that yielded live bacteria, however as a small proportion of populations dominated by other variants it does not seem to be at an obvious disadvantage. This indicates there may be a threshold at which SpnIIIB expression becomes detrimental to the population as a whole or that the SpnIIIB WT strain contains an unidentified mutation. In addition, SpnIIIF is at a less obvious disadvantage in the blood and spleen, and as a result the brain. Unlike other variants it fails to increase its proportion when it is the dominant allele in any of the mice tested.

5.5 Human model of pneumococcal carriage

While murine models of infection can be informative, the natural niche of the pneumococcus is the human nasopharynx. However, murine models provide control over the input bacterial population that cannot be determined when clinical samples are analysed. The Experimental Human Pneumococcal Carriage (EHPC) project (106) provided an opportunity to analyse pneumococcal samples from carriage episodes with a known input population.

It was hypothesised that over a colonisation period of up to 24 days SpnIII variants more suited to colonisation would be selected for. This would lead to changes in the overall variant percentages over time.

There are several different potential outcomes in the nasopharynx. Firstly, it may be that SpnIII recombination does not occur. In this case changes to variant percentages would indicate selection while a lack of change would indicate no selection. Alternatively, SpnIII recombination is ongoing and is either dependent on or independent of the number of generations. The final possibility is that SpnIII recombination is ongoing and is subject to selection. Using the model described in Chapter 3 the experimental outcome of human carriage has been compared to the outcome predicted by the model. This model assumes no selection and ongoing recombination is dependent on the number of generations.

In total 81 samples from 24 healthy volunteers inoculated with the carriage associated strain BHN418 were received (107). Of the 81 samples received 69 from 20 volunteers were PCR positive for *S. pneumoniae.* The SpnIII variant percentages for each individual were plotted by SpnIII variant (A-F) for all time points for which data was available (Fig 5.11). The input populations varied but were dominated by SpnIIIA (59.00±11.23), SpnIIIB (15.10±10.62) and SpnIIID (17.70±3.24).

No consistent trends where observed when samples were analysed by individual volunteer (Fig 5.11). However analysis of the median variant percentages at days 0 and 22/24 revealed that the percentage of SpnIIIC was significantly increased at day 22/24 compared to day 0 ($p=0.001$) (Fig 5.12A). To determine if this result was due to selection in favour of the SpnIIIC variant the experimental outcome was compared to the outcome predicted by the model. A doubling time of 120 minutes was used, based on *S. pneumoniae* growth on sialic acid which is freely available in the human nasopharynx (112). It was assumed that populations underwent 264 generations (12 generations per day for 22 days) over the course of the experiment. The outcome of each individual challenge dose after 264 generations was determined and the median plotted (Fig 5.12B). Using a Mann Whitney test it was determined that the experimental outcome at day 22/24 was significantly different in SpnIIIC, D, E and F when compared to the outcome predicted by the model. The median percentage of SpnIIIC ($p=0.039$), SpnIIIE ($p=0.001$) and SpnIIIF ($p=0.001$) across all samples PCR positive at day 22/24 was found to be significantly lower than predicted by the model. The median percentage of SpnIIID was found to be significantly greater than predicted ($p=0.001$).

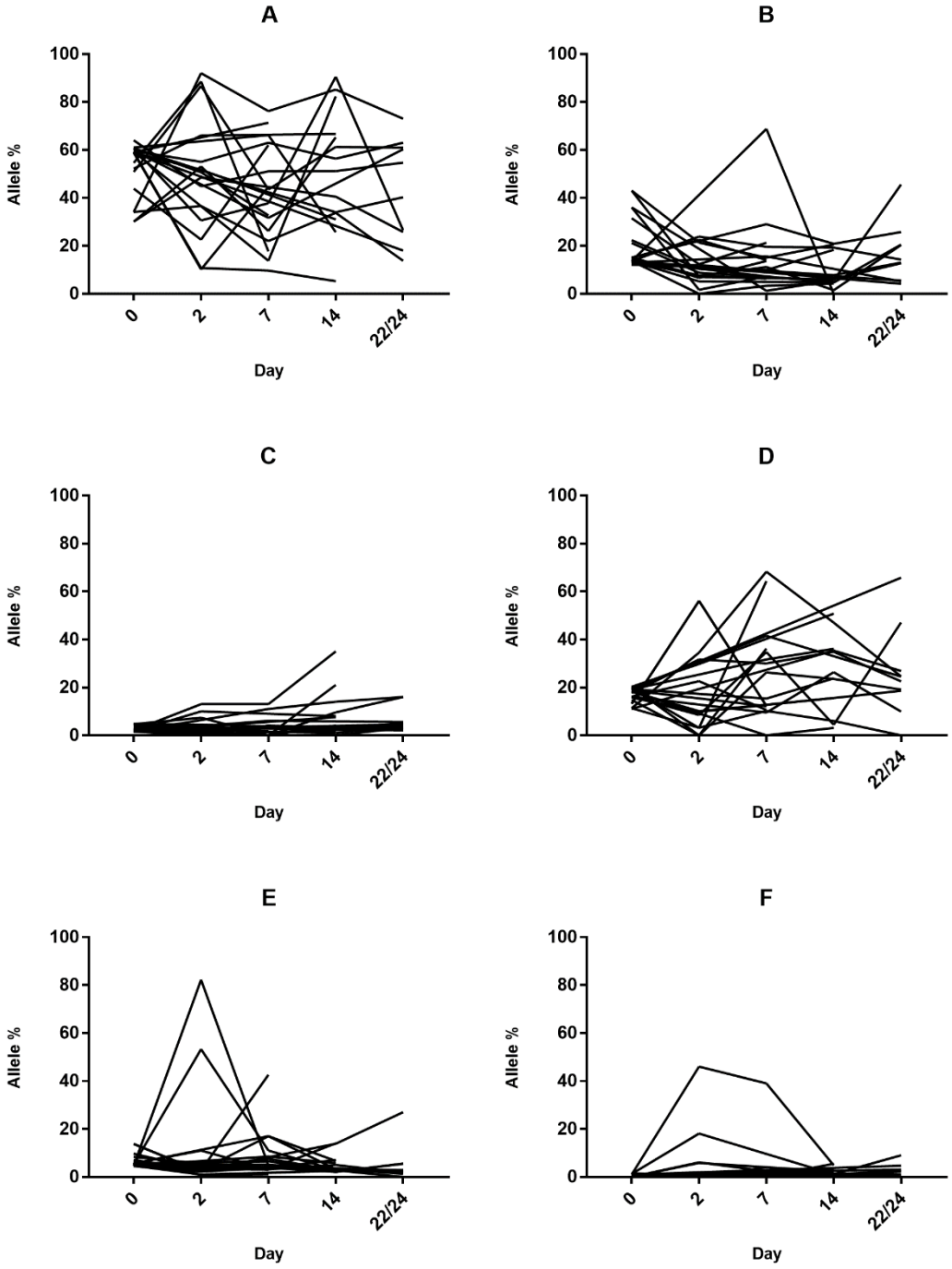


Figure 5.11 – SpnIII population trends within the nasopharynx of 20 healthy human volunteers. *Each nasopharyngeal sample from the Experimental Human Pneumococcal Carriage Project has been analysed with respect to the input population i.e., the challenge dose each volunteer received. Each line represents a single volunteer and each SpnIII variant is represented on an individual graph (A-F). An SpnIII variant's percentage in the challenge dose they received (day 0) was plotted along with the SpnIII variant percentage within the samples at all time points where viable cells were recovered from that volunteer (days 2, 7, 14 and 22/24). It was not possible to analyse samples for all time points for all volunteers. While some peaks can be seen these are correlated with lower bacterial recovery. No overall trends could be observed in the individually analysed samples.*

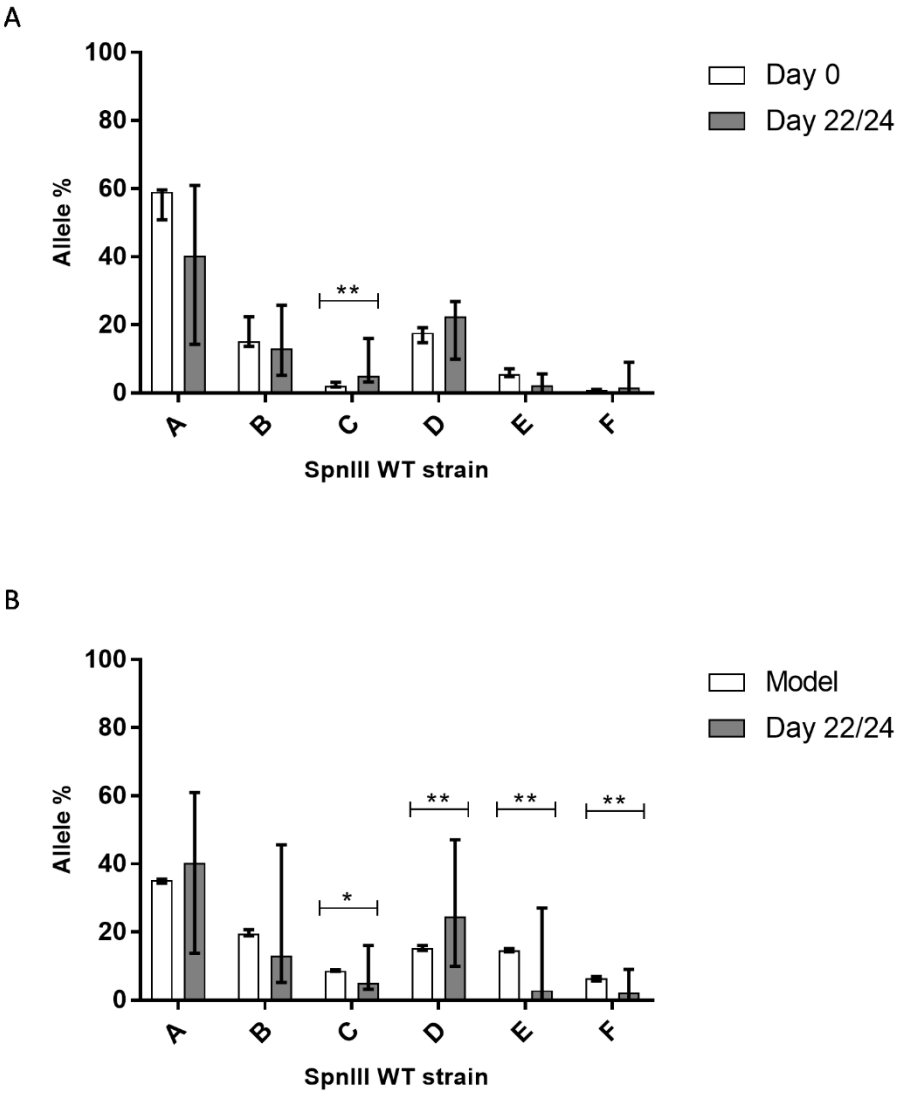


Figure 5.12 – SpnIII population trends within the nasopharynx of 20 healthy human volunteers after 22/24 days of carriage. *Each nasopharyngeal sample from EHPC has been analysed with respect to the input population i.e., the challenge dose each volunteer received. Volunteers who were not carriage positive at day 22/24 were excluded from this analysis. The SpnIII variant percentages for each volunteer carriage positive at day 22/24 were pooled, as were the percentages from all associated challenge doses, and the median plotted. In panel A Input (white bars) and output (grey bars) percentages were compared for each SpnIII variant and significance was determined using a Mann Whitney test. SpnIIIC was significantly increased compared to the input population (p=0.0014). The experimentally derived percentages were compared to the outcome predicted by a model when recombination is stochastic (panel B). Each individual inoculum was entered into a model that uses experimentally defined recombination frequencies for the SpnIII locus (see chapter 3 for additional details). The model was run for 264 generations* (112)*. The experimental and model data were compared using a Mann-Whitney test, and the smaller percentage populations (C, D, E and F) were shown to be significantly different from the outcome predicted by the model at 264 generations. Despite the SpnIIIC percentage being significantly greater than the inoculum (panel A) is it still significantly lower than predicted by the model. SpnIIIE and F also make up significantly less of the day 22/24 population than would be predicted by the model, however the SpnIIID percentage is significantly higher than predicted. SpnIIIA and B do not differ significantly from the outcome predicted by the model. *p≤0.05, **≤0.01, ***p≤0.001.*

The results of individual volunteers are highly variable. Some samples show peaks of a single allele at a single time point, for example volunteer 1069 at day 2 saw a sharp increase in SpnIIIA which is not observed again at day 7. However there is no common SpnIII variant that repeatedly generates peaks. This suggests that these events either occur randomly or are the result of a sampling bias. Although some sharp increases are observed a comparison between variant percentages in the challenge doses and the recovered samples (Fig 5.12A) suggest this is not maintained over time.

5.6 **Chapter discussion**

Murine models of infection have confirmed results previously reported by Manso *et al*. (64) that cells expressing the SpnIIIB variant are less virulent in invasive disease models. However, the previous observation that SpnIIIA variants were less capable of colonising the nasopharynx (64) was not reproduced in WT strains.

Analysis of SpnIII variant percentages in the nasopharynx revealed an unexpected increase in the most prevalent variant, irrespective of which variant was dominant. The predicted outcome of random recombination would be a decrease in the most prevalent variant, with an increase expected to indicate selection. An increase in all variants suggests that this is not the result of selection. Importantly these data have been generated from a pneumonia model of infection, where bacteria are designed to enter the lungs. In contrast a model of colonisation utilises an inoculation protocol designed to minimise bacteria entering the lungs (146). As a result of this the effective population size, i.e. those cells contributing to the next generation (147) within the nasopharynx is likely to be significantly smaller in the nasopharynx than the lung. Furthermore, the effective population size of the pneumococcus in the nasopharynx has been shown to be considerably smaller than the size of the carriage population (147). If the presence of bacteria in the nasopharynx is the result of cells escaping from the lung there is a significant recurring bottleneck. Based on the variant percentages in each WT strain it is most probable that the most prevalent variant will succeed in escape from the lung to the nasopharynx thereby skewing the

population in favour of the most prevalent allele. It should also be noted that the rate of pneumococcal growth in the nasopharynx is difficult to analyse and likely impacted by the sugars available (112). A minimal growth rate, combined with a small effective population size, would explain the apparent increase in SpnIII variant percentages in the nasopharynx.

Significant variation was observed in the lungs, blood and spleen with no consistent pattern of selection in favour of a single SpnIII variant. However, selection against the SpnIIIB variant occurred in both pneumonia and IV models of infection and resulted in significantly lower numbers of bacteria in the blood and lungs. This is in agreement with the observation by Manso *et al.* that a phase-locked SpnIIIB variant was significantly attenuated in invasive disease models (64). It also concurs with the observation in chapter 3 that the SpnIIIB WT strain expresses significantly lower levels of the known virulence factor pneumolysin (20).

Analysis of bacteria recovered from the brains of intracisternally infected mice allowed an investigation of samples that were not affected by random bottlenecking of the population. The demonstration by Moxon *et al.* that infections arising from a small number of cells (100 CFU) result in infections caused by a single bacterium (148) is important in understanding the differences observed in potentially bottlenecked samples (blood and spleen) and non-bottlenecked samples (intracisternally infected brain). The samples collected from the brain showed less variability, with conserved trends such as a decrease in the prevalent variant in the case of SpnIIIA, D, E and F, while SpnIIIC showed a fairly consistent increase, regardless of whether or not it was the predominant variant. This is in contrast to the samples from the blood and spleen where both increases and decreases were seen in all variants. The wide variety of outcomes in these samples is likely the outcome of random bottlenecking reducing the population size and creating a skew that is not representative of selection. The only clear selection observed was against the SpnIIIB WT strain which consistently resulted in lower bacterial counts in the blood, spleen, lungs and brain irrespective of route of infection.

Analysis of samples from the EPHC provided a unique opportunity to build on the work conducted in murine models of infection and investigate changes to SpnIII within the human nasopharynx. Previous work using clinical samples has not identified any strong selection in favour of a single SpnIII variant (23) however this analysis was unable to detect more subtle selection due to the lack of a known input population. As shown in chapter 3, colonies randomly founded by different SpnIII variants result in diverse outcomes, therefore the opportunity to investigate samples with a known input population allowed a more representative investigation into SpnIII selection within the nasopharynx.

The significant differences observed between the outcome predicted by the model and that observed in the experimental samples suggests one of two possible outcomes. If SpnIII recombination is not ongoing *in vivo* the differences could indicate selection against SpnIIIC, E and F and selection in favour of SpnIIID. Alternatively, if SpnIII recombination is generation dependent it may be that an over or under-estimation of the number of generations in the nasopharynx would explain one or more of the significant differences. As with H segment inversion in *S. enterica* it is likely that SpnIII inversion occurs during DNA replication and the level of recombination is therefore dependent on the number of generations (149). If the number of generations has been overestimated this could explain the significantly lower levels of SpnIIIC, E and F, although implies that the significantly higher percentage of SpnIIID is the result of selection in favour of this variant. Alternatively, if generations have been underestimated the result observed in SpnIIIC, E and F would be an under-representation of selection against these variants. If recombination *in vivo* is not ongoing an increase in one variant, a decrease in three variants and no significant difference in the remaining two variants would be highly unlikely. It is a strong possibility that an over-estimation of the number of generation has a significant impact on the conclusions draw from the analysis of the EHPC samples.

6 Discussion

6.1 General discussion

The work of the thesis has explored the *in vivo* and *in vitro* phenotypes associated with differential expression of the *hsdS* subunit of the type I RM system, *Spn*III. The conservation of *Spn*III within the core pneumococcal genome (78) and the previous association with virulence through murine models of infection (64) and differences in colony morphology (64,65) makes it an interesting topic of investigation as a mechanism for global virulence regulation. In addition, the generation of WT *S. pneumoniae* strains enriched in a single *Spn*III variant allowed the investigation of differences in the recombination dynamics driving *spnIII* recombination.

*Spn*III recombination appears to be significantly more complex than initially anticipated. It has previously been observed in *M. pulmonis* that a single site specific recombinase controls all inversions in a similar phase variable type I RM system (68,75,77) therefore it was hypothesised that the site specific tyrosine recombinase found within the *spnIII* locus has sole control over recombination on all three inverted repeats (64,78). This hypothesis was quickly disproven when inactivation of the recombinase, CreX, failed to eliminate recombination. While it is uncommon for two site specific recombinases to function on same region of DNA it is not impossible, as shown by the dual activity of the Hin and Fin recombinases of *S. enterica,* which both function to invert the H segment controlling the phase variable expression of two different flagellin proteins (91,92). Therefore, work was conducted to determine whether or not a second SSR could be functioning alongside CreX to invert regions of the *spnIII* locus. Analysis of all known SSRs in the D39 genome revealed that none were significantly involved in *spnIII* recombination, disproving the hypothesis that two recombinases function simultaneously at the locus.

When the type I RM locus *hsd1* of *M. pulmonis* is expressed in *E. coli,* along with its associated recombinase HvsR, inversions only occur on the larger of the two inverted repeats (77). This suggests that proteins specific to *M. pulmonis* are required for inversions on the smaller *vip* repeat (77). Due to the size of the region inverted at the *spnIII* locus it would not be surprising if additional proteins were involved in facilitating recombination. As a result, a panel of alternative proteins,

known to be associated with DNA binding and homologous recombination were tested for their impact on *spnIII* recombination (35,39,137,139–141). This analysis demonstrated that, like similar systems in *B. fragilis* and *M. pulmonis* (67,77), the process of recombination occurs independently of the universal recombinase RecA. The involvement of both the RecFOR and mismatch repair pathways was ruled out, indicating that the classical routes of homologous recombination are not involved, an expected result as RecA is integral to both pathways. Surprisingly, this analysis did reveal the partial involvement of the competence specific RecA loader DprA on the 15bp repeat of the *spnIII* locus. This was unexpected as DprA is only known to be expressed following the induction of competence (37). The implication of this result is that either *spnIII* recombination is tied to the induction of competence or alternatively DprA function is not limited to the competence pathway. Despite the differences observed in the recombination rates of different SpnIII WT strains, no differences were found in their ability to incorporate exogenous DNA when simulated with synthetic CSP. It is therefore reasonable to hypothesise that DprA may have a function beyond the loading of RecA onto in-coming ssDNA. Mirouze *et al.* have already demonstrated an additional role for DprA, which interacts with ComE to initiate the shutdown of the competence pathway (37). It is therefore not unreasonable to hypothesise a further function unrelated to natural competence.

As inversions at the *spnIII* locus appear to be, at least in part, independent of both site specific and homologous recombination, alternatives must be considered. Dutra *et al.* have demonstrated that not only is RecA independent homologous recombination possible, but it is also efficient in the absence of exonucleases such as RecJ (144). Further work on *spnIII* recombination should explore the possibility of RecA independent homologous recombination as a contributing mechanism. There is also a possibility that *spnIII* recombination is controlled using a methylation sensitive feedback loop to ensure recombination only occurs in appropriate conditions, such as DNA replication or low exonuclease expression. This type of feedback could occur using a mechanism similar to that of the CcrM MTase of *C. crescentus* in which methylation patterns regulate DNA replication in different cells types (45).

The most notable differences in the SpnIII WT strains were found when gene expression was investigated used RNAseq. Previous work using phase-locked mutants revealed major differences in the expression of the capsule operon (64) which was not replicated in this study. However, a variety of virulence associated genes, included the pneumolysin operon, were found to be differentially regulated. It has previously been demonstrated that strains lacking pneumolysin result in shorter episodes of carriage (121), as well as these strains being cleared from the blood significantly more quickly (20). The SpnIIIB WT strain expressed a significantly lower level of pneumolysin than the SpnIIIE WT, possibly explaining why it is rapidly cleared from the blood. Despite this, Manso *et al.* reported that an SpnIIIB phase-locked mutant is significantly more capable of nasopharyngeal colonisation when compared to other phase-locked SpnIII strains (64). In this study using SpnIII WT strains no differences were observed in the CFU/ml recovered from the nasopharynx, suggesting that all WT strains are equally capable of nasopharyngeal colonisation. The increased expression of pneumolysin in the SpnIIIE WT strain offers an explanation for the increased virulence of this strain in the intravenous model of infection.

The reduced pneumolysin expression in the SpnIIIB WT strain may be counteracted by the increased expression of *dnaJ*, another protein that, when knocked out, results is shorter episodes of carriage (123). Of the WT strains SpnIIIB shows the greatest expression of *dnaJ*, which may aid in colonisation of the nasopharynx in the absence of increased pneumolysin. While there was not a significant difference in the expression of *spxB* within the WT strains the significant reduction in the percentage of SpnIIIB within SpnIIIE founded colonies on catalase agar suggests an increased susceptibility to hydrogen peroxide. Hydrogen peroxide is produce by pyruvate oxidase (SpxB) when *S. pneumoniae* grows in aerobic conditions (125). An increased expression of *spxB* has been associated with the ability to survive in the nasopharynx and also the transparent colony phenotype (126,127), both of which have been associated with the SpnIIIB variant (64,65).

Based on gene expression analysis in SpnIII WT strains it appears that differential expression of *hsdS* subunits, and therefore alternative methylation patterns, are

regulating genes known to influence both invasive disease (pneumolysin) and colonisation (DnaJ).

A huge amount of variability was observed in *in vivo* models of infection, particularly in samples collected from a human model of nasopharyngeal colonisation. This is in contrast with nasopharyngeal samples from an intranasal murine infection which showed a consistent pattern of dominant variant increase. There are a variety of differences between the two models that may account for these differences. Importantly the two models do not share the same initial aim, the EHPC model aims to colonise the nasopharynx of approximately 50% of those inoculated (106), whereas the murine model aims to achieve an invasive pneumococcal infection in all mice challenged. There is also a significant difference in the challenge dose used between the two models with the EHPC protocol using 10^4-10^5 CFU/ml of serotype 6B strain BHN418 (106) and murine models typically using a dose in the region of 10^7 CFU/ml of serotype 2 strain D39. Finally, it should be noted that there is a significant difference in total experimental time between the two models. While the EPHC protocol collects samples up to 24 days post-inoculation, the intranasal murine model is completed within 24 hours. The differences in the length of experiments, as well as the intended severity of infection, along with the use of different pneumococcal strains, is likely to account for the differences observed within the nasopharyngeal samples from the two models.

Some significant differences were observed in the model predicted and experimental outcomes of human nasopharyngeal colonisation. This result could indicate selection in favour of variants found at higher than predicted percentages (SpnIIID) or selection against those found at lower than predicted percentages (SpnIIIC, E and F). However it may also be the result of a lower than predicted effective population size (147), as the model assumes that all cells present will contribute to subsequent generations. In addition, the model ties recombination to generations, therefore an over or underestimate of the number of generations has the potential to significantly skew the results. Generation time discrepancies would however only account in either the significantly over or under represented variants, it cannot account for both.

The lack of significant trends in murine models, aside from the reduced virulence of the SpnIIIB WT strain, may be the result of significant population bottlenecks. Bottlenecks within the blood and brain can result in infections initiated by single organisms, regardless of how diverse the input population is (148). As demonstrated in chapter three of this thesis the random selection of single cells can lead to very different SpnIII profiles in the subsequent population, leading to a high level of variability between mice challenged with the same initial dose. In addition to bottlenecks impacting the results observed, the effective population size may also be significantly smaller than the overall population size (147), i.e. not all cells may be contributing to the next generation. The assumption that all cells contribute to the next generation could lead to over-estimations of expected diversity in areas such as the nasopharynx. Interestingly the result observed in the nasopharynx closely mirrors that obtained by Gumulak-Smith *et al.* that *M. pulmonis* cells recovered from the lungs of intranasally challenged rats showed significantly more variability in *hsdS* expression when compared to cells recovered from the nasopharynx (74).

The SpnIII RM system has shown to efficiency methylate 100% of its known recognition sites in phase-locked mutants (64). Despite this it was difficult to identify obvious *in vitro* phenotypes associated with differential methylation. It is possible that this is the result of mixed methylation patterns within WT populations. The proven differences in gene expression are likely to only reveal the most significant differences, as those that are smaller may be masked by the mixed populations. This will be particularly true of the SpnIIIF WT strain which contains a much greater percentage of non-dominant SpnIII variants. Despite these challenges the use of WT strains has proved to be invaluable in determining differences in SpnIII recombination rates and in confirming that strains expressing different SpnIII variants do behave differently in invasive models of infection.

The regulation of important virulence phenotypes by *Spn*III has the potential to hugely impact studies of pneumococcal virulence, both past and present. The ease with which lab strains can be bottlenecked into expressing a single pre-dominant variant demonstrates how the generation of mutants can easily result in strains expressing alternative SpnIII variants. The differences in gene expression or virulence of mutant strains when compared to WT parents could

simply be the result of differential methylation patterns caused by alternative SpnIII variants. Without testing individual strains, it cannot be known whether or not the differences in gene expression and virulence in the literature discussed truly represent mutation of the intended gene or simply differential methylation.

6.2 Future work

Future work on the SpnIII type I RM system should focus on determining the mechanisms controlling recombination, in particular mechanisms of RecA-independent homologous recombination (144). A strategy for investigating this type of recombination would be linked to determining time points where exonucleases are significantly down-regulated (144). In addition, the transfer of previously generated mutants into a background with a higher rate of recombination would provide an opportunity to look more closely at recombination decreases that only occur at a very low rate in D39.

The continued improvement of the mathematical model used for analysis of samples from the EHPC will lead to more refined conclusions, although the difficulty in estimating generation times will remain. Furthermore, the generation of multiple model parameters, accounting for the differing recombination rates of pneumococcal strains would make conclusions more specific to individual strains.

Additional work is also required to determine why *S. pneumoniae* D39 cannot maintain levels of SpnIIIF comparable to all other SpnIII variants. The work completed thus far has helped determine where in the culturing process SpnIIIF percentages decrease, building on this work it would be important to determine if there are any significant gene expression changes associated with increases in cell density.

Finally, the mechanism preserving genomic DNA when methylation patterns are in disagreement with REase specificity should be investigated further. The maintenance of self DNA in these situations is in direct contrast with all that is known about type I RM systems and how they function. The identification of phase variable type I systems in an ever increasing number of different bacterial species suggests that there must be common mechanism to protect self DNA.

6.3 **Concluding remarks**

As systems such as *Spn*III continue to be identified, through ever improving genome sequencing technology, an understanding of how they function will become increasingly important. The proven, reversible, global methylation of genomes suggests a function beyond the typical role of bacteriophage defence usually assigned to RM systems. Despite extensive study, questions about how and when recombination occurs remain unanswered both in *S. pneumoniae* and other species. There are likely to be extensive and complicated mechanisms to regulate recombination and to ensure protection of self DNA following a change in methylation specificity. The control of gene expression through methylation is not a new concept, however the widespread ability to phase variably recognise different methylation patterns offers an exciting new opportunity in the understanding of many bacterial species.

Appendix 1 – Mathematical Model

Model generated by Andrew Morozov from the Department of Mathematics, University of Leicester, using the experimental data produced in this thesis.

A=0.0073; B=0.0015; C=0.0008;

c(1,1)=exp(-A)*exp(-B)*exp(-C)+ A^2*exp(-A)*exp(-B)*exp(-C)/2+B^2*exp(-A)*exp(-B)*exp(-C)/2+C^2*exp(-A)*exp(-B)*exp(-C)/2;
c(2,2)=c(1,1);
c(3,3)=c(1,1);
c(4,4)=c(1,1);
c(5,5)=c(1,1);
c(6,6)=c(1,1);
c(7,7)=c(1,1);
c(8,8)=c(1,1);
c(1,2)=B* exp(-A)* exp(-B)* exp(-C)+A^2*B*exp(-A)* exp(-B)* exp(-C) /2 + C^2*B*exp(-A)* exp(-B)* exp(-C) /2+ B^3 *exp(-A)* exp(-B)* exp(-C)/6;
c(1,3)=A*exp(-A)*exp(-B)*exp(-C)+ A^3*exp(-A)*exp(-B)*exp(-C)/6 + B^2*A*exp(-A)*exp(-B)*exp(-C)/2 + C^2*A*exp(-A)*exp(-B)*exp(-C)/2;
c(1,4)=C*exp(-A)*exp(-B)*exp(-C)+ C^3*exp(-A)*exp(-B)*exp(-C)/6 + A^2*C*exp(-A)*exp(-B)*exp(-C)/2 + B^2*C*exp(-A)*exp(-B)*exp(-C)/2;
c(1,5)=A*C*exp(-A)*exp(-B)*exp(-C);
c(1,6)=A*B*C*exp(-A)*exp(-B)*exp(-C);
c(1,7)=A*B*exp(-A)*exp(-B)*exp(-C);
c(1,8)=B*C*exp(-A)*exp(-B)*exp(-C);

c(2,1)=c(1,2); c(3,1)=c(1,3); c(4,1)=c(1,4); c(5,1)=c(1,5); c(6,1)=c(1,6); c(7,1)=c(1,7);
c(8,1)=c(1,8);
c(2,3)=A*B*exp(-A)*exp(-B)*exp(-C);
c(2,4)=B*C*exp(-A)*exp(-B)*exp(-C);
c(2,5)=A*B*C*exp(-A)*exp(-B)*exp(-C);
c(2,6)=A*C*exp(-A)*exp(-B)*exp(-C);

c(2,7)=A*exp(-A)*exp(-B)*exp(-C)+ A^3*exp(-A)*exp(-B)*exp(-C/6)+C^2*A*exp(-A)*exp(-B)*exp(-C/2)+B^2*A*exp(-A)*exp(-B)*exp(-C)/2;
c(2,8)=C*exp(-A)*exp(-B)*exp(-C) +A^2*C*exp(-A)*exp(-B)*exp(-C)/2 + C^3*exp(-A)*exp(-B)*exp(-C)/6 + B^2*C*exp(-A)*exp(-B)*exp(-C)/2;

c(3,2)=c(2,3); c(4,2)=c(2,4); c(5,2)=c(2,5); c(6,2)=c(2,6); c(7,2)=c(2,7); c(8,2)=c(2,8);
c(3,4)=A*C*exp(-A)*exp(-B)*exp(-C);
c(3,5)= C*exp(-A)*exp(-B)*exp(-C)+A^2*C*exp(-A)*exp(-B)*exp(-C)/2 +B^2*C*exp(-A)*exp(-B)*exp(-C)/2+ C^3 *exp(-A)*exp(-B)*exp(-C/6);
c(3,6)= B*C*exp(-A)*exp(-B)*exp(-C);
c(3,7)=B*exp(-A)*exp(-B)*exp(-C)+A^2*B*exp(-A)*exp(-B)*exp(-C/2)+C^2*B*exp(-A)*exp(-B)*exp(-C)/2+ B^3*exp(-A)*exp(-B)*exp(-C)/6;
c(7,3)=c(3,7);
c(3,8)=A*B*C*exp(-A)*exp(-B)*exp(-C);

c(4,3)=c(3,4);c(5,3)=c(3,5); c(6,3)=c(3,6);c(8,3)=c(3,8);

c(7,8)=A*C*exp(-A)*exp(-B)*exp(-C);
c(8,7)=c(7,8);

c(4,5)=A*exp(-A)* exp(-B)* exp(-C)+ A^3* exp(-A)* exp(-B)* exp(-C)/6 +B^2*A*exp(-A)* exp(-B)* exp(-C)/2 +C^2*A*exp(-A)* exp(-B)* exp(-C) /2;
c(4,6)=A*B*exp(-A)* exp(-B)*exp(-C);
c(4,7)=A*B*C*exp(-A)* exp(-B)*exp(-C);
c(4,8)=B* exp(-A)* exp(-B)* exp(-C)+A^2*B*exp(-A)* exp(-B)* exp(-C) /2 + C^2*B*exp(-A)* exp(-B)* exp(-C) /2+ B^3 *exp(-A)* exp(-B)* exp(-C)/6;

c(5,4)=c(4,5); c(6,4)=c(4,6);c(7,4)=c(4,7); c(8,4)=c(4,8);

c(5,6)= B* exp(-A)* exp(-B)* exp(-C)+A^2*B*exp(-A)* exp(-B)* exp(-C) /2 + C^2*B*exp(-A)* exp(-B)* exp(-C) /2+ B^3 *exp(-A)* exp(-B)* exp(-C)/6;
c(6,5)=c(5,6);

c(5,7)=B*C*exp(-A)*exp(-B)*exp(-C);

c(5,8)=A*B*exp(-A)*exp(-B)*exp(-C);

c(7,5)=c(5,7); c(8,5)=c(5,8);

c(6,7)=C*exp(-A)*exp(-B)*exp(-C)+ C^3*exp(-A)*exp(-B)*exp(-C)/6 + A^2*C*exp(-A)*exp(-B)*exp(-C)/2 + B^2*C*exp(-A)*exp(-B)*exp(-C)/2;
c(6,8)=A*exp(-A)*exp(-B)*exp(-C)+ A^3*exp(-A)*exp(-B)*exp(-C)/6 + B^2*A*exp(-A)*exp(-B)*exp(-C)/2 + C^2*A*exp(-A)*exp(-B)*exp(-C)/2;

```
c(7,6)=c(6,7); c(8,6)=c(6,8);


N0(1)=; %A1
N0(2)=; %A2
N0(3)=; %B
N0(4)=; %C
N0(5)=; %D1
N0(6)=; %D2
N0(7)=; %E
N0(8)= :%F


%1.89   0.00      5.48      84.43    3.48      4.71

n=12;

N=N0';

for i=1:n

NN=2*c*N;

N=NN;

NNN(1:8,i)=N/(sum(N));
end
```

References

1. Kilian M, Riley DR, Jensen A, Brüggemann H, Tettelin H. Parallel evolution of *Streptococcus pneumoniae* and *Streptococcus mitis* to pathogenic and mutualistic lifestyles. MBio. 2014;5(4):e01490–e01414.

2. Donati C, Hiller NL, Tettelin H, Muzzi A, Croucher NJ, Angiuoli S V, *et al.*. Structure and dynamics of the pan-genome of *Streptococcus pneumoniae* and closely related species. Genome Biol. 2010;11(10):R107.

3. Ferreira DM, Jambo KC, Gordon SB. Experimental human pneumococcal carriage models for vaccine research. Trends Microbiol. Elsevier Ltd; 2011;19(9):464–70.

4. Feldman C, Anderson R. Epidemiology, virulence factors and management of the pneumococcus. F1000Research. 2016;5(0):2320.

5. WHO. Pneumococcal conjugate vaccine for childhood immunization – WHO position paper. Vol. 82, Weekly epidemiological record. 2007.

6. Górecki RK, Koryszewska-Bagińska A, Gołebiewski M, Zylińska J, Grynberg M, Bardowski JK. Adaptative potential of the *Lactococcus lactis* IL594 strain encoded in its 7 plasmids. PLoS One. 2011;6(7).

7. Murray J, Agócs M, Serhan F, Singh S, Deloria-Knoll M, O'Brien K, *et al.*. Global Invasive Bacterial Vaccine-Preventable Diseases Surveillance — 2008–2014. MMWR Morb Mortal Wkly Rep. 2014;63(49):1159–62.

8. Waight PA, Andrews NJ, Ladhani SN, Sheppard CL, Slack MPE, Miller E. Effect of the 13-valent pneumococcal conjugate vaccine on invasive pneumococcal disease in England and Wales 4 years after its introduction: An observational cohort study. Lancet Infect Dis. Elsevier Ltd; 2015;15(5):535–43.

9. Melegaro A, Edmunds WJ, Pebody R, Miller E, George R. The current burden of pneumococcal disease in England and Wales. J Infect. 2006;52(1):37–48.

10. Avery OT, MacCleod CM, Mccarty M. Studies on the chemical nature of the substance inducing transformation of pneumococcal types induction. J Exp Med. 1944;79(2):137–58.

11. Hyams C, Camberlein E, Cohen JM, Bax K, Brown JS. The *Streptococcus pneumoniae* capsule inhibits complement activity and neutrophil phagocytosis by multiple mechanisms. Infect Immun. 2010;78(2):704–15.

12. PHE. Pneumococcal disease: guidance, data and analysis. Public Health England. 2014 [cited 2017 Jan 22].

13. WHO. Pneumococcal vaccines. WHO position paper. Wkly Epidemiol Rec Relev épidémiologique Hebd. 2012;87(14):129–44.

14. Chambaud I, Heilig R, Ferris S, Barbe V, Samson D, Galisson F, *et al.*. The

complete genome sequence of the murine respiratory pathogen *Mycoplasma pulmonis*. Nucleic Acids Res. 2001;29(10):2145–53.

15. Geno KA, Gilbert GL, Song JY, Skovsted IC, Klugman KP, Jones C, *et al.*. Pneumococcal capsules and their types: Past, present, and future. Clin Microbiol Rev. 2015;28(3):871–99.

16. Kim JO, Weiser JN. Association of intrastrain phase variation in quantity of capsular polysaccharide and teichoic acid with the virulence of *Streptococcus pneumonia*e. J Infect Dis. 1998;177(2):368–77.

17. Weiser JN, Austrian R, Sreenivasan PK, Masure HR. Phase variation in pneumococcal opacity: Relationship between colonial morphology and nasopharyngeal colonization. Infect Immun. 1994;62(6):2582–9.

18. Feldman C, Anderson R. Recent advances in our understanding of *Streptococcus pneumoniae* infection. F1000Prime Rep. 2014;6(September):82.

19. Palaniappan R, Singh S, Singh UP, Sakthivel SKK, Ades EW, Briles DE, *et al.*. Differential PsaA-, PspA-, PspC-, and PdB-specific immune responses in a mouse model of pneumococcal carriage. Infect Immun. 2005;73(2):1006–13.

20. Quin LR, Moore QC, McDaniel LS. Pneumolysin, PspA, and PspC contribute to pneumococcal evasion of early innate immune responses during bacteremia in mice. Infect Immun. 2007;75(4):2067–70.

21. Mitchell AM, Mitchell TJ. *Streptococcus pneumoniae*: Virulence factors and variation. Clin Microbiol Infect. European Society of Clinical Infectious Diseases; 2010;16(5):411–8.

22. Schachern PA, Tsuprun V, Ferrieri P, Briles DE, Goetz S, Cureoglu S, *et al.*. Pneumococcal PspA and PspC Proteins: Potential vaccine candidates for experimental otitis media. Int J Pediatr Otorhinolaryngol. 2014;78(9):1517–21.

23. Lees J, Kremer PHC, Manso AS, Croucher NJ, Ferwerda B, Serón MV, *et al.*. Large scale genomic analysis shows no evidence for repeated pathogen adaptation during the invasive phase of bacterial meningitis in humans. Microb Genomics. 2017;3(1).

24. Wright AKA, Ferreira DM, Gritzfeld JF, Wright AD, Armitage K, Jambo KC, *et al.*. Human nasal challenge with *Streptococcus pneumoniae* is immunising in the absence of carriage. PLoS Pathog. 2012;8(4):16–9.

25. Wilson R, Cohen J, Jose R, De Vogel C, Baxendale H, Brown J. Protection against *Streptococcus pneumoniae* lung infection after nasopharyngeal colonization requires both humoral and cellular immune responses. Mucosal Immunol. Nature Publishing Group; 2015;8(3):627–39.

26. Avery OT, Dubos R. The protective action of a specific enzyme against type III pneumococcus infection in mice. J Exp Med. 1931;54(1):73–89.

27. Webster LT, Clow AD. Intranasal virulence of pneumococci for Mice. J Exp Med. 1933;58(4):465–83.

28. Kawamura Y, Hou XG, Sultana F, Miura H, Ezaki T. Determination of 16S rRNA sequences of *Streptococcus mitis* and *Streptococcus gordonii* and phylogenetic relationships among members of the genus *Streptococcus*. Int J Syst Bacteriol. 1995;45(2):406–8.

29. Ekdahl K, Hansson HB, Mölstad S, Söderström M, Walder M, Persson K. Limiting the spread of penicillin-resistant *Streptococcus pneumoniae*: experiences from the South Swedish Pneumococcal Intervention Project. Microb Drug Resist. 1997;4(2):99–105.

30. Skov Sørensen UB, Yao K, Yang Y, Tettelin H, Kilian M. Capsular polysaccharide expression in commensal *Streptococcus* species: Genetic and antigenic similarities to *Streptococcus pneumoniae*. MBio. 2016;7(6):e01844-16.

31. Eutsey RA, Powell E, Dordel J, Salter SJ, Clark TA, Korlach J, *et al.*. Genetic stabilization of the drug-resistant PMEN1 pneumococcus lineage by its distinctive DpnIII restriction-modification system. MBio. 2015;6(3):1–12.

32. Johnston C, Campo N, Bergé MJ, Polard P, Claverys JP. *Streptococcus pneumoniae*, le transformiste. Trends Microbiol. 2014;22(3):113–9.

33. Johnston C, Martin B, Granadel C, Polard P, Claverys JP. Programmed protection of foreign DNA from restriction allows pathogenicity island exchange during pneumococcal transformation. PLoS Pathog. 2013;9(2):1–10.

34. Claverys JP, Martin B, Polard P. The genetic transformation machinery: Composition, localization, and mechanism. FEMS Microbiol Rev. 2009;33(3):643–56.

35. Johnston C, Mortier-Barrière I, Granadel C, Polard P, Martin B, Claverys J-P. RecFOR is not required for pneumococcal transformation but together with XerS for resolution of chromosome dimers frequently formed in the process. PLoS Genet. 2015;11(1):e1004934.

36. Griffith BYF. The significance of pneumococcal types. Epidemiol Infect. 1928;27(2):113–57.

37. Mirouze N, Bergé MA, Soulet A-L, Mortier-Barrière I, Quentin Y, Fichant G, *et al.*. Direct involvement of DprA, the transformation-dedicated RecA loader, in the shut-off of pneumococcal competence. Proc Natl Acad Sci U S A. 2013;110(11):E1035-44.

38. Morrison DA, Lee MS. Regulation of competence for genetic transformation in *Streptococcus pneumoniae*: a link between quorum sensing and DNA processing genes. Res Microbiol. 2000;151(6):445–51.

39. Mortier-Barrière I, Velten M, Dupaigne P, Mirouze N, Piétrement O, McGovern S, *et al.*. A key presynaptic role in transformation for a

widespread bacterial protein: DprA conveys incoming ssDNA to RecA. Cell. 2007;130(5):824–36.

40. Arber W, Dussoix D. Host specificity of DNA produced by *Escherichia coli.* I. Host-controlled modification of bacteriophage lambda. J Mol Biol. 1962;5:18–36.

41. Vasu K, Nagaraja V. Diverse functions of restriction-modification systems in addition to cellular defense. Microbiol Mol Biol Rev. 2013;77(1):53–72.

42. Blow MJ, Clark TA, Daum CG, Deutschbauer AM, Fomenkov A, Fries R, *et al.*. The epigenomic landscape of prokaryotes. PLoS Genet. 2016;

43. Clark T a., Murray I a., Morgan RD, Kislyuk AO, Spittle KE, Boitano M, *et al.*. Characterization of DNA methyltransferase specificities using single-molecule, real-time DNA sequencing. Nucleic Acids Res. 2012;40(4).

44. Bheemanaik S, Reddy YVR, Rao DN. Structure, function and mechanism of exocyclic DNA methyltransferases. Biochem J. 2006;399(2):177–90.

45. Marinus MG, Casadesus J. Roles of DNA adenine methylation in host-pathogen interactions: mismatch repair, transcriptional regulation, and more. FEMS Microbiol Rev. 2009;33(3):488–503.

46. Loenen W a M, Dryden DTF, Raleigh E a., Wilson GG. Type I restriction enzymes and their relatives. Nucleic Acids Res. 2014;42(1):20–44.

47. Loenen W a M, Dryden DTF, Raleigh E a., Wilson GG, Murrayy NE. Highlights of the DNA cutters: A short history of the restriction enzymes. Nucleic Acids Res. 2014;42(1):3–19.

48. Murray NE. Type I restriction systems: sophisticated molecular machines (a legacy of Bertani and Weigle). Microbiol Mol Biol Rev. 2000;64(2):412–34.

49. Loenen W a M, Raleigh E a. The other face of restriction: Modification-dependent enzymes. Nucleic Acids Res. 2014;42(1):56–69.

50. Morgan RD, Luyten YA, Johnson SA, Clough EM, Clark TA, Roberts RJ. Novel m4C modification in type I restriction-modification systems. Nucleic Acids Res. 2016;44(19):9413–25.

51. Kan N, Lautenberger J, Edgell M, Hutchison C 3rd. The nucleotide sequence recognized by the *Escherichia coli* K12 restriction and modification enzymes. J Mol Biol. 1979;130(2):191–209.

52. Roberts RJ, Vincze T, Posfai J, Macelis D. REBASE—a database for DNA restriction and modification: enzymes, genes and genomes. Nucleic Acids Res. 2015;43.

53. Srikhanta YN, Fox KL, Jennings MP. The phasevarion: phase variation of type III DNA methyltransferases controls coordinated switching in multiple genes. Nat Rev Microbiol. Nature Publishing Group; 2010;8(3):196–206.

54. Dryden DTF, Edwardson JM, Henderson RM. DNA translocation by type III restriction enzymes: A comparison of current models of their operation derived from ensemble and single-molecule measurements. Nucleic Acids Res. 2011;39(11):4525–31.

55. Lacks S a, Mannarelli BM, Springhorn SS, Greenberg B. Genetic basis of the complementary DpnI and DpnII restriction systems of *S. pneumoniae*: an intercellular cassette mechanism. Cell. 1986;46(7):993–1000.

56. Machnicka MA, Kaminska KH, Dunin-Horkawicz S, Bujnicki JM, Bickle T, Kruger D, *et al.*. Phylogenomics and sequence-structure-function relationships in the GmrSD family of Type IV restriction enzymes. BMC Bioinformatics. 2015;16(1):336.

57. Heusipp G, Fälker S, Alexander Schmidt M. DNA adenine methylation and bacterial pathogenesis. Int J Med Microbiol. 2007;297(1):1–7.

58. Johnston C, Polard P, Claverys J-P. The DpnI/DpnII pneumococcal system, defense against foreign attack without compromising genetic exchange. Mob Genet Elements. 2013;3(4):e25582.

59. Casadesús J, Low D a. Programmed heterogeneity: Epigenetic mechanisms in bacteria. J Biol Chem. 2013;288(20):13929–35.

60. Bayliss CD. Determinants of phase variation rate and the fitness implications of differing rates for bacterial pathogens and commensals. FEMS Microbiol Rev. 2009;33(3):504–20.

61. Anjum A, Brathwaite KJ, Aidley J, Connerton PL, Cummings NJ, Parkhill J, *et al.*. Phase variation of a Type IIG restriction-modification enzyme alters site-specific methylation patterns and gene expression in *Campylobacter jejuni* strain NCTC11168. Nucleic Acids Res. 2016;44(10):4581–94.

62. Srikhanta YN, Gorrell RJ, Steen J a., Gawthorne J a., Kwok T, Grimmond SM, *et al.*. Phasevarion mediated epigenetic gene regulation in *Helicobacter pylori*. PLoS One. 2011;6(12):1–9.

63. Adamczyk-Poplawska M, Lower M, Piekarowicz A. Deletion of one nucleotide within the homonucleotide tract present in the hsdS gene alters the DNA sequence specificity of type I restriction-modification system NgoAV. J Bacteriol. 2011;193(23):6750–9.

64. Manso AS, Chai MH, Atack JM, Furi L, De Ste Croix M, Haigh R, *et al.*. A random six-phase switch regulates pneumococcal virulence via global epigenetic changes. Nat Commun. Nature Publishing Group; 2014;5:5055.

65. Li J, Li JW, Feng Z, Wang J, An H, Liu Y, *et al.*. Epigenetic Switch Driven by DNA Inversions Dictates Phase Variation in *Streptococcus pneumoniae*. PLoS Pathog. 2016;12(7):1–36.

66. Dybvig K, Yu H. Regulation of a restriction and modification system via DNA inversion in *Mycoplasma pulmonis*. Mol Microbiol. 1994;12(4):547–60.

67. Cerdeño-Tárraga AM, Patrick S, Crossman LC, Blakely G, Abratt V, Lennard N, *et al.*. Extensive DNA inversions in the *B. fragilis* genome control variable gene expression. Science. 2005;307(5714):1463–5.

68. Sitaraman R, Dybvig K. The hsd loci of *Mycoplasma pulmonis*: organization, rearrangements and expression of genes. Mol Microbiol. 1997;26(1):109–20.

69. Ron Y, Flitman-Tene R, Dybvig K, Yogev D. Identification and characterization of a site-specific tyrosine recombinase within the variable loci of *Mycoplasma bovis, Mycoplasma pulmonis* and *Mycoplasma agalactiae*. Gene. 2002;292(1–2):205–11.

70. Mostowy R, Croucher NJ, Hanage WP, Harris SR, Bentley S, Fraser C. Heterogeneity in the frequency and characteristics of homologous recombination in pneumococcal evolution. PLoS Genet. 2014;10(5):1–15.

71. Tettelin H, Nelson KE, Paulsen IT, Eisen JA, Read TD, Scott P. Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*. Science. 2001;293(5529):498-506.

72. Fagerlund A, Langsrud S, Schirmer BCT, Møretrø T, Heir E. Genome analysis of *Listeria monocytogenes* sequence type 8 strains persisting in salmon and poultry processing environments and comparison with related strains. PLoS One. 2016;11(3):1–22.

73. Claesson MJ, Li Y, Leahy S, Canchaya C, van Pijkeren JP, Cerdeño-Tárraga AM, *et al.*. Multireplicon genome architecture of *Lactobacillus salivarius*. Proc Natl Acad Sci U S A. 2006;103(17):6718–23.

74. Gumulak-Smith J, Teachman A, Tu AHT, Simecka JW, Russell Lindsey J, Dybvig K. Variations in the surface proteins and restriction enzyme systems of *Mycoplasma pulmonis* in the respiratory tract of infected rats. Mol Microbiol. 2001;40(4):1037–44.

75. Dybvig K, Sitaraman R, French CT. A family of phase-variable restriction enzymes with differing specificities generated by high-frequency gene rearrangements. Proc Natl Acad Sci U S A. 1998;95(23):13923–8.

76. Bhugra B, Voelker LL, Zou N, Yu H, Dybvig K. Mechanisms of antigenic varition in *Mycoplasma pulmonis*. Mol Microbiol. 1995;18:703–714.

77. Sitaraman R, Denison AM, Dybvig K. A unique, bifunctional site-specific DNA recombinase from *Mycoplasma pulmonis*. Mol Microbiol. 2002;46(4):1033–40.

78. Croucher NJ, Coupland PG, Stevenson AE, Callendrello A, Bentley SD, Hanage WP. Diversification of bacterial genome content through distinct mechansims over different timescales. Nat Commun. Nature Publishing Group; 2014;5:1–12.

79. Sadowski P. Site-specific recombinases: changing partners and doing the twist. J Bacteriol. 1986;165(2):341–7.

80. Croucher NJ, Kagedan L, Thompson CM, Parkhill J, Bentley SD, Finkelstein J a., *et al.*. Selective and genetic constraints on pneumococcal serotype switching. PLOS Genet. 2015;11(3):e1005095. 5

81. Furuta Y, Kawai M, Uchiyama I, Kobayashi I. Domain movement within a gene: A novel evolutionary mechanism for protein diversification. PLoS One. 2011;6(4).

82. Furuta Y, Namba-Fukuyo H, Shibata TF, Nishiyama T, Shigenobu S, Suzuki Y, *et al.*. Methylome diversification through changes in DNA methyltransferase sequence specificity. PLoS Genet. 2014;10(4).

83. Schouler C, Clier F, Lerayer AL, Ehrlich SD, Chopin MC. A type IC restriction-modification system in *Lactococcus lactis*. J Bacteriol. 1998;180(2):407–11.

84. Madsen A, Westphal C, Josephsen J. Characterization of a novel plasmid-encoded HsdS subunit, S.LlaW12I, from *Lactococcus lactis* W12. Plasmid. 2000;44(2):196–200.

85. O'Sullivan D, Twomey DP, Coffey A, Hill C, Fitzgerald GF, Ross RP. Novel type I restriction specificities through domain shuffling of HsdS subunits in *Lactococcus lactis.* Mol Microbiol. 2000;36(4):866–75.

86. Schouler C, Gautier M, Ehrlich SD, Chopin MC. Combinational variation of restriction modification specificities in *Lactococcus lactis*. Mol Microbiol. 1998;28(1):169–78.

87. Johnson RC. Site-specific DNA inversion by serine recombinases. Microbiol Spectr. 2015;3(3):1–36.

88. Grindley NDF, Whiteson KL, Rice PA. Mechanisms of site-specific recombination. Annu Rev Biochem. 2006;75:567–605.

89. Deshmukh N., Van Duyne G. Structure and mechanism in site-specific recombination. Curr Opin Struct Biol. 1999;9:14–20.

90. Ghosh K, Van Duyne G. Cre– loxP biochemistry. Methods. 2002;28:374–83.

91. Kutsukake K, Nakashima H, Tominaga A, Abo T. Two DNA invertases contribute to flagellar phase variation in *Salmonella enterica serovar typhimurium* strain LT2. J Bacteriol. 2006;188(3):950–7.

92. Heichman KA, Johnson RC. The Hin Invertasome: Protein-mediated joining of distant recombination sites at the enhancer. Science. 1990;249(4968):511–7.

93. Rajeev L, Malanowska K, Gardner JF. Challenging a paradigm: the role of DNA homology in tyrosine recombinase reactions. Microbiol Mol Biol Rev. 2009;73(2):300–9.

94. Le Bourgeois P, Bugarel M, Campo N, Daveran-Mingot ML, Labonté J, Lanfranchi D, *et al.*. The unconventional xer recombination machinery of

Streptococci/Lactococci. PLoS Genet. 2007;3(7):1190–9.

95. Arciszewska LK, Grainge I, Sherratt DJ. Action of site-specific recombinases XerC and XerD on tethered Holliday junctions. 1997;16(12):3731–43.

96. Leroux M, Jia F, Szatmari G. Characterization of the *Streptococcus suis* XerS recombinase and its unconventional cleavage of the difSL site. FEMS Microbiol Lett. 2011;324(2):135–41.

97. Coyne MJ, Weinacht KG, Krinos CM, Comstock LE. Mpi recombinase globally modulates the surface architecture of a human commensal bacterium. Proc Natl Acad Sci USA. 2003;100(18):10446–51.

98. Shen X, Gumulak J, Yu H, French CT, Zou N, Dybvig K. Gene rearrangements in the vsa locus of *Mycoplasma pulmonis*. J Bacteriol. 2000;182(10):2900–8.

99. Nakayama-Imaohji H, Hirakawa H, Ichimura M, Wakimoto S, Kuhara S, Hayashi T, *et al.*. Identification of the site-specific DNA invertase responsible for the phase variation of SusC/SusD family outer membrane proteins in *Bacteroides fragilis*. J Bacteriol. 2009;191(19):6003–11.

100. Brinster S, Lamberet G, Staels B, Trieu-Cuot P, Gruss A, Poyart C. Type II fatty acid synthesis is not a suitable antibiotic target for Gram-positive pathogens. Nature. Nature Publishing Group; 2009;458(7234):83–6.

101. Miles AA, Misra SS. The estimation of the bactericidal power of the blood. J Hyg (Lond). 1938;38(6):732–49.

102. McClure R, Balasubramanian D, Sun Y, Bobrovskyy M, Sumby P, Genco CA, *et al.*. Computational analysis of bacterial RNA-Seq data. Nucleic Acids Res. 2013;41(14):1–16.

103. Jeffreys AJ, Neumann R, Wilson V. Repeat unit sequence variation in minisatellites: A novel source of DNA polymorphism for studying variation and mutation by single molecule analysis. Cell1. 1990;60(3):473–485.

104. Sung CK, Li H, Claverys JP, Morrison D a. An rpsL Cassette , Janus , for Gene Replacement through Negative Selection in *Streptococcus pneumoniae*. Appl Environ Microbiol. 2001;67(11):5190–6.

105. Claverys JP, Dintilhac A, Pestova E V., Martin B, Morrison DA. Construction and evaluation of new drug-resistance cassettes for gene disruption mutagenesis in *Streptococcus pneumoniae*, using an *ami* test platform. Gene. 1995;164(1):123–8.

106. Gritzfeld JF, Wright AD, Collins AM, Pennington SH, Wright AK, Kadioglu A, *et al.*. Experimental human pneumococcal carriage. J Vis Exp. 2013;(72):1–5.

107. Browall S, Norman M, Tångrot J, Galanis I, Sjöström K, Dagerhamn J, *et al.*. Intraclonal variations among *Streptococcus pneumoniae* isolates influence the likelihood of invasive disease in children. J Infect Dis.

2014;209(3):377–88.

108. Jonczyk MS, Escudero L, Sylvius N, Norman M, Henriques-Normark B, Andrew PW. The variation in inflammatory response during pneumococcal infections is influenced by the host-pathogen interactions but associates with animal survival. Infect Immun. 2016;84(January):IAI.01057-15.

109. Lanie J a., Ng WL, Kazmierczak KM, Andrzejewski TM, Davidsen TM, Wayne KJ, *et al.*. Genome sequence of Avery's virulent serotype 2 strain D39 of *Streptococcus pneumoniae* and comparison with that of unencapsulated laboratory strain R6. J Bacteriol. 2007;189(1):38–51.

110. Hiller NL, Janto B, Hogg JS, Boissy R, Yu S, Powell E, *et al.*. Comparative genomic analyses of seventeen *Streptococcus pneumoniae* strains: Insights into the pneumococcal supragenome. J Bacteriol. 2007;189(22):8186–95.

111. Croucher NJ, Mitchell AM, Gould KA, Inverarity D, Barquist L, Feltwell T, *et al.*. Dominant role of nucleotide substitution in the diversification of serotype 3 pneumococci over decades and during a single infection. PLoS Genet. 2013;9(10).

112. Bidossi A, Mulas L, Decorosi F, Colomba L, Ricci S, Pozzi G, *et al.*. A functional genomics approach to establish the complement of carbohydrate transporters in *Streptococcus pneumoniae*. PLoS One. 2012;7(3).

113. Camilli R, Bonnal RJP, Del Grosso M, Iacono M, Corti G, Rizzi E, *et al.*. Complete genome sequence of a serotype 11A, ST62 *Streptococcus pneumoniae* invasive isolate. BMC Microbiol. BioMed Central Ltd; 2011;11(1):25.

114. Fälker S, Nelson AL, Morfeldt E, Jonas K, Hultenby K, Ries J, *et al.*. Sortase-mediated assembly and surface topology of adhesive pneumococcal pili. Mol Microbiol. 2008;70(3):595–607.

115. Wion D, Casadesús J. N6-methyl-adenine: an epigenetic signal for DNA-protein interactions. Nat Rev Microbiol. 2006;4(3):183–92.

116. Pericone CD, Overweg K, Hermans PW, Weiser JN. Inhibitory and bactericidal effects of hydrogen peroxide production by *Streptococcus pneumoniae* on other inhabitants of the upper respiratory tract. Infect Immun. 2000;68(7):3990–7.

117. Keatch SA, Su T-J, Dryden DTF. Alleviation of restriction by DNA condensation and non-specific DNA binding ligands. Nucleic Acids Res. 2004;32(19):5841–50.

118. Park S-S, Kwon H-Y, Dang-Hien Tran T, Choi M-H, Jung S-H, Lee S, *et al.*. ClpL is a chaperone without auxiliary factors. FEBS J. 2015;282:1352–67.

119. Kwon HY, Ogunniyi a D, Choi MH, Pyo SN, Rhee DK, Paton JC. The ClpP protease of *Streptococcus pneumoniae* modulates virulence gene expression and protects against fatal pneumococcal challenge. Infect Immun. 2004;72(10):5646–53.

120. Loonstra A, Vooijs M, Beverloo HB, Allak BA, van Drunen E, Kanaar R, *et al.*. Growth inhibition and DNA damage induced by Cre recombinase in mammalian cells. Proc Natl Acad Sci USA. 2001;98(16):9209–14.

121. Richards L, Ferreira DM, Miyaji EN, Andrew PW, Kadioglu A. The immunising effect of pneumococcal nasopharyngeal colonisation; protection against future colonisation and fatal invasive disease. Immunobiology. Elsevier; 2010;215(4):251–63.

122. Takaya A, Tomoyasu T, Matsui H, Yamamoto T. The DnaK/DnaJ Chaperone machinery of *Salmonella enterica* serovar Typhimurium is essential for invasion of epithelial cells and survival within macrophages, leading to systemic infection. Infect Immun. 2004;72(3):1364–73.

123. Cui J, Ma C, Ye G, Shi Y, Xu W, Zhong L, *et al.*. DnaJ (hsp40) of *Streptococcus pneumoniae* is involved in bacterial virulence and elicits a strong natural immune reaction via PI3K/JNK. Mol Immunol. Elsevier Ltd; 2017;83:137–46.

124. Cui Y, Zhang X, Gong Y, Niu S, Yin N, Yao R, *et al.*. Immunization with DnaJ (hsp40) could elicit protection against nasopharyngeal colonization and invasive infection caused by different strains of *Streptococcus pneumoniae*. Vaccine. Elsevier Ltd; 2011;29(9):1736–44.

125. Spellerberg B, Cundell DR, Sandros J, Pearce BJ, Idanpaan-Heikkila I, Rosenow C, *et al.*. Pyruvate oxidase, as a determinant of virulence in *Streptococcus pneumoniae*. Mol Microbiol. 1996;19::803–813.

126. Li-Korotky HS, Lo CY, Zeng FR, Lo D, Banks JM. Interaction of phase variation, host and pressure/gas composition: Pneumococcal gene expression of PsaA, SpxB, Ply and LytA in simulated middle ear environments. Int J Pediatr Otorhinolaryngol. 2009;73(10):1417–22.

127. Overweg K, Pericone CD, Verhoef GGC, Weiser JN, Meiring HD, De Jong APJM, *et al.*. Differential protein expression in phenotypic variants of *Streptococcus pneumoniae*. Infect Immun. 2000;68(8):4604–10.

128. Blakely G, Sherratt D. Determinants of selectivity in Xer site-specific recombination. Genes Dev. 1996;10(6):762–73.

129. Claverys J-P, Håvarstein LS. Cannibalism and fratricide: mechanisms and raisons d'être. Nat Rev Microbiol. 2007;5(3):219–29.

130. Wilkins BM. Plasmid promiscuity: Meeting the challenge of DNA immigration control. Environ Microbiol. 2002;4(9):495–500.

131. Reichmann P, Hakenbeck R. A XerD recombinase with unusual active site motifs in *Streptococcus pneumoniae*. J Mol Microbiol Biotechnol. 2002;4(1):101–10.

132. Brückner R, Nuhn M, Reichmann P, Weber B, Hakenbeck R. Mosaic genes and mosaic chromosomes-genomic variation in *Streptococcus pneumoniae*. Int J Med Microbiol. 2004;294(2–3):157–68.

133. Artimo P, Jonnalagedda M, Arnold K, Baratin D, Flegel V, Fortier A, *et al.*. ExPASy : SIB bioinformatics resource portal. Nucleic Acids Res. 2012;40:W597–603.

134. Oggioni MR, Claverys J. Repeated extragenic sequences in prokaryotic genomes : a proposal for the origin and dynamics of the RUP element in *Streptococcus pneumoniae*. Microbiology. 1999;(145):2647–53.

135. Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M. ISfinder : the reference centre for bacterial insertion sequences. 2006;34:32–6.

136. Pozzi G, Musmanno RA, Renzoni EA, Oggioni MR, Cusi MG. Host-vector system for integration of recombinant DNA into chromosomes of transformable and nontransformable *Streptococci*. J Bacteriol. 1988;170(4):1969.

137. Martin B, Garcia P, Castanié MP, Claverys JP. The recA gene of *Streptococcus pneumoniae* is part of a competence-induced operon and controls lysogenic induction. Mol Microbiol. 1995;15(2):367–79.

138. Mcgregor N, Ayora S, Sedelnikova S, Carrasco B, Alonso JC, Thaw P, *et al.*. The structure of *Bacillus subtilis* RecU Holliday junction resolvase and its role in substrate selection and sequence-specific cleavage. 2005;13:1341–51.

139. Attaiech L, Olivier A, Mortier-Barrière I, Soul*et al.*, Granadel C, Martin B, *et al.*. Role of the single-stranded DNA-binding protein ssbB in pneumococcal transformation: Maintenance of a reservoir for genetic plasticity. PLoS Genet. 2011;7(6).

140. Martin B, Sharples GJ, Humbert O, Lloyd RG, Claverys J-P. The *mmsA* locus of *Streptococcus pneumoniae* encodes a RecG-like protein involved in DNA repair and in three-strand recombination. Mol Microbiol. 1996;19(5):1035–45.

141. Prudhomme M, Libante V, Claverys J-P. Homologous recombination at the border: insertion-deletions and the trapping of foreign DNA in *Streptococcus pneumoniae*. Proc Natl Acad Sci USA. 2002;99(4):2100–5.

142. Bell JC, Kowalczykowski SC. RecA: Regulation and mechanism of a molecular search engine. Trends Biochem Sci. 2016;41(6):491–507.

143. Kowalczykowski SC, Dixon DA, Eggleston AK, Lauder SD, Rehrauer WM. Biochemistry of homologous recombination in *Escherichia coli*. Microbiol Rev. 1994;58(3):401–65.

144. Dutra BE, Sutera VA, Lovett ST. RecA-independent recombination is efficient but limited by exonucleases. Proc Natl Acad Sci USA. 2007;104(1):216–21.

145. Bi X, Liu LF. DNA rearrangement mediated by inverted repeats. Proc Natl Acad Sci USA. 1996;93(2):819–23.

146. Wu H-Y, Virolainen A, Mathews B, King J, Russell MW, Briles DE.

Establishment of a *Streptococcus pneumoniae* nasopharyngeal colonization model in adult mice. Microb Pathog. 1997;23:127–37.

147. Li Y, Thompson CM, Trzciński K, Lipsitch M. Within-host selection is limited by an effective population of *Streptococcus pneumoniae* during nasopharyngeal colonization. Infect Immun. 2013;81(12):4534–43.

148. Moxon ER, Murphy PA. *Haemophilus influenzae* bacteremia and meningitis resulting from survival of a single organism. Proc Natl Acad Sci USA. 1978;75(3):1534–6.

149. Lodish H, Berk A, Zipursky S. Molecular Cell Biology. 4th ed. New York: W. H. Freeman; 2000. Functional Rearrangements in Chromosomes.