Recombination of the phase variable spnIII locus is independent of all known pneumococcal site-specific recombinases.

M De Ste Croix, Y Chen, I Vacca, A S Manso, C Johnston, P Polard, M J Kwun, S D Bentley, N J Croucher, C D Bayliss, R D Haigh, M R Oggioni

# Address correspondence to Marco R. Oggioni, mro5@leicester.ac.uk
Abstract

*Streptococcus pneumoniae* is one of the world’s leading bacterial pathogens, causing pneumonia, septicaemia and meningitis. In recent years it has been shown that genetic rearrangements in a type I restriction-modification system (SpnIII) can impact colony morphology and gene expression. By generating a large panel of mutant strains, we have confirmed a previously reported result that the CreX (also known as IvrR and PsrA) recombinase found within the locus is not essential for *hsdS* inversions. In addition, mutants of homologous recombination pathways also undergo *hsdS* inversions. In this work we have shown that these genetic rearrangements, which result in different patterns of genome methylation, occur across a wide variety of serotypes and sequence types including two strains (a 19F and a 6B strain) naturally lacking CreX. Our gene expression analysis, by RNAseq, confirm that the level of creX expression is impacted by these genomic rearrangements. In addition, we have shown that the frequency of *hsdS* recombination is temperature dependent. Most importantly we have demonstrated that the other known pneumococcal site-specific recombinases XerD, XerS and SPD_0921 are not involved in *spnIII* recombination, suggesting a currently unknown mechanism is responsible for the recombination of these phase variable type I systems.
Streptococcus pneumoniae is a leading cause of pneumonia, septicaemia and meningitis. The discovery that genetic rearrangements in a type I restriction modification locus can impact gene regulation and colony morphology have led to a new understanding of how this pathogen switches from harmless coloniser to invasive pathogen. These rearrangements, which alter the DNA specificity of the type I restriction modification enzyme, occur across many different pneumococcal serotypes and sequence types, and in the absence of all known pneumococcal site-specific recombinases. This finding suggests that this is a truly global mechanism of pneumococcal gene regulation and the need for further investigation of mechanisms of site specific recombination.
Introduction

*S. pneumoniae* is a major human pathogen responsible for pneumonia, septicemia and meningitis, along with less severe infections such as otitis media and sinusitis. It is estimated that *S. pneumoniae* colonizes ~50% of under-fives in the UK (1). Although this colonization is largely asymptomatic, pneumococcal infections are the leading cause of lower respiratory infection morbidity and mortality globally (2). Phase variation (PV), the ability to reversibly change protein expression, is often controlled by switching genes on and off stochastically within bacterial populations. PV permits the rapid adaptation to host environments and is therefore likely to contribute to both the asymptomatic spread of *S. pneumoniae* as well as the poorly understood switch between harmless colonizer and invasive pathogen.

In recent years phase variable (PV) type I restriction modification (RM) systems, capable of changing HsdS protein expression through the movement of target recognition domains (TRDs), have been identified in many bacterial species (3). Rather than switch expression of the type I RM system on and off these PV changes, through DNA inversions, permit the expression of multiple *hsdS* genes in a bacterial population. Species with these systems include, but are not limited to, *S. pneumoniae* (4–8), *Listeria monocytogenes* (9), *Bacteriodes fragilis* (10), *Enterococcus faecalis* (11), *Lactobacillus salivarius* (12), and *Streptococcus suis* (13, 14). Our understanding of the biological relevance of these systems is limited despite being first identified in *Mycoplasma pulmonis* more than 20 years ago (15). Interestingly, nearly all of these PV type I RM systems contain an associated site-specific recombinase within their locus (3). This suggests that site-specific recombination may be a key mechanism facilitating the rearrangement of TRDs and therefore PV by methylation in a variety of bacterial pathogens.
SpnIII, also known as the *ivr* (16) and *cod* (6) locus, is a ubiquitous PV type I RM system of the pneumococcus (6, 7, 16, 17). While some strains with modified *spnIII* loci have been detected, no sequenced pneumococcal strain has been found without a *spnIII* locus. SpnIII is encoded by genes SPD_0451-0455 in D39 and genes SP_0505-0510 in TIGR4. PV results in variation through six versions of SpnIII with differing S sub-units and consequent recognition sequences and was observed to occur at rates of greater than 1%. Previous analysis found differential gene expression patterns for each active *hsdS* gene orientation (17). Expression patterns were generated using non-phase-variable mutants where all but two of the TRDs (Target Recognition Domains) had been removed. These mutants also exhibited differences in virulence in murine models of infection (17) and in the opaque/transparent colony morphology, a phenotype previously associated with differences in the colonization and invasive potential of pneumococcal isolates (18). Multiple studies have linked specific *hsdS* expression states to a particular proportion of opaque (OP) to transparent (TP) colonies (6, 17, 19). Thus SpnIII PV has a significant potential to impact the colonization or invasion phenotype during a clonal pneumococcal infection.

PV of SpnIII is driven by site-specific recombination between inverted repeat sequences. Site-specific recombination is mediated by enzymes that catalyze the cleavage and re-joining of DNA fragments independently of RecA and other homologous recombination machinery (20). Site-specific recombinases, the enzymes that facilitate these reactions, cut and re-ligate at specific recognition sequences without ATP or the synthesis of any new DNA (21). These enzymes occur in a wide variety of bacteria and phage and enable the inversion and excision of DNA, depending on the presence of either inverted (inversion) or direct (excision)
repeats (22, 23). Intragenomic recombination in bacteria can also be mediated by the universal recombinase RecA and this process also has the potential to contribute to SpnIII PV. In the pneumococcus, RecA is required for homologous recombination enabling both repair of DNA breaks and integration of incoming DNA during natural transformation (24–26). In the pneumococcus three different DNA loading complexes facilitate the loading of RecA onto DNA as part of three independent pathways of homologous recombination (27). Firstly, the RecFOR complex loads RecA onto single stranded DNA (ssDNA) gaps when replication forks stall (26, 28). Secondly, RexAB (a functional homologue of RecBCD) loads RecA onto regions with double stranded DNA (dsDNA) breaks (29). Finally, the transformation-dedicated recombinase loader dprA loads RecA onto incoming DNA during transformation thereby facilitating incorporation into the chromosome when homologous regions of DNA are present (25, 27).

The mechanisms responsible for SpnIII PV have not been fully defined. Previous work in M. pulmonis demonstrated that a single site-specific recombinase was solely responsible for PV of two loci, including one type I RM system (30). Recently Kwun et al. (31) reported that the direct repeat-mediated shuffling of TRDs at the tvr locus of S. pneumoniae, encoding the type I RM system SpnIV, is enhanced but not exclusively controlled by the site-specific recombinase found within the locus. Similarly variation in the SpnIII locus was found to occur in the absence of the site-specific recombinase (6, 32). Our aim was to determine whether site-specific, homologous recombination, or another process, drives the stochastic movement of TRDs between active and silent positions at the spnIII locus in S. pneumoniae D39. We show that TRD switching is independent of the universal recombinase RecA but
is partially controlled by the site-specific recombinase CreX (also known as IvrR and PsrA), which is encoded within the spnIII locus.
Results

The SpnIII restriction modification system contains the classical type I RM system genes hsdR (restriction), hsdM (methylation) and hsdS (specificity), alongside two additional hsdS genes and a site-specific recombinase (creX) (Fig. 1A) (6, 17). Only one of the hsdS genes found within the locus is a complete gene and is transcribed (17), this gene is termed the active gene. The remaining two hsdS genes lack any transcriptional start sites and act as donors of target recognition domains (TRDs) (17). TRD shuffling is dependent on three IRs within the locus (Fig. 1B), and a on a core conserved target sequence of 10bp (ATTATGGGAA) found within all IRs. We propose that these conserved 10bp are essential to CreX mediated recombination. Recombination frequencies were estimated for wildtype (WT) strain D39 and a wide range of other strains representing range of different serotypes (Fig. S1). Analysis of single colonies enables identification of the active gene present in the founder cell of the colony. As growth of single colonies on an agar plate for 16-18 hours is equal to approximately 20-22 generations, the percentage of cells that have undergone TRD shuffling on each of the three IRs can be determined as a proportion of the total population.

All mutant analysis was conducted in D39 and G54, however to confirm that PV at the spnIII locus is widespread 13 additional WT strains, representing 6 serotypes and 8 sequence types were analysed (Fig. S1; Table S4). TRD shuffling was observed in all strains tested, including strains with an incomplete spnIII locus (G54 and BHN191). The rate of switching was variable between strains, and differed greatly even between strains of the same sequence type (ST) (Fig. S1). The stock of D39 held within our laboratory collection predominantly expresses hsdSE (Fig. S1), therefore all analysis in D39 has been restricted to colonies founded by hsdSE cells.
The use of hsdSE expressing cells allows inversions on all three repeats to be compared within a single colony. As the average number of switches within a single D39 hsdSE colony is dependent on number of generations, colonies left on agar plates for increased periods of time will show greater levels of recombination. In a colony incubated for 16 hours at 37°C there is a mean switching frequency within the colony of 16.8% (SD±5.4) on the 333 bp IR, 1.5% (SD±0.8) (10-fold less) on the 85 bp IR and 4.1% (SD±1.9) (3-fold less) on the 15 bp IR.

To determine the impact of growth conditions on TRD shuffling, D39 was grown at 23°C, 37°C and 42°C with 5% CO₂ (Fig. 1D). TRD shuffling was significantly more frequent at 23°C than at 37°C (p<0.0001) or 42°C (p<0.0001). While growth at 42°C did not eliminate shuffling, inversions on the largest 333bp repeat were reduced as shown by the increase in the proportion of hsdSA in colonies founded by hsdSE expressing cells from 11.8% (SD±6.2) at 42°C to 19.9% (SD±3.9) at 37°C, and further still to 25.7% (SD±6.5) at 23°C.

To investigate the impact on TRD shuffling of the site-specific recombinase CreX, also known as PsrA (32), an RNAseq analysis was performed on WT strains enriched for a single active hsdS. When in the same orientation as the hsdR, hsdM and active hsdS genes, the CreX gene was expressed at a level 2.6-fold greater than when inverted and co-linear with the silent hsdS’ and hsdS” (Fig. 1B-C). Interestingly, the increased creX expression of hsdSA and hsdSD variants was associated with an 80% reduction in recombination at the 333 bp IR: 2.6% (SD±1.4) compared to 11.6% (SD±2.3) in the other four variants. Note that recombination at the short repeat is not quantified but is presumed to be still occurring in these variants (Fig. 1E). These results imply that CreX may block other forms of recombination at the locus.
To examine the specific role of CreX, we generated a frameshift mutant that truncated the CreX protein at amino acid 6 without altering the size of the gene. TRD shuffling in single colonies of the creX mutant (MRO633) were compared to single colonies of the WT D39 strain. In the absence of the CreX recombinase shuffling between the smallest 15 bp IR was no longer observed (i.e. no hsdSB variants were generated in hsdSE founded cells, Fig. 2A). This is in accordance with the hypothesis that site-specific recombination at the conserved 10 bp sequence present in all three IRs (ATTATGGGAA) is strictly CreX dependent. Shuffling to the hsdSA variant was also reduced from 15.8% (SD±5.4) to 12.7% (SD±2.0) (p=0.0007) showing that CreX may be partially but not fully responsible for these events. Shuffling on the 85 bp IR (hsdSE to hsdSD) was unaffected (p=0.9). The lack of impact on the 85 bp IR may be explained by the difference in the 5 bp immediately upstream of the conserved 10 bp sequence (Fig. 2E), in the 333 bp and 15 bp IRs the sequence appears to be conserved as CTCTT but this is not the case for the 85 bp IR (GAAAC). The observation that CreX is not the sole mechanism of control for TRD shuffling is supported by evidence from the analysis of strains G54 (Fig. 1A) and BHN191 (Fig. S1) which both lack the CreX recombinase and at least one TRD. Both G54 and BHN191 display recombination frequencies between the TRDs present that are similar to those observed on the 333bp IR of D39 (Fig. S1). It is unlikely these strains undergo recombination at the 15 bp IR in the absence of the CreX recombinase.

Double knockout mutants of each other known, functional site-specific recombinase within the D39 genome (33) were tested to determine whether multiple site-specific recombinases controlled TRD shuffling at the spnIII locus. All double knockouts contain a creX frameshift along with the knockout of one other site-specific.
recombinase. Analysis of single colonies from double knockout strains showed that site-specific recombinases SPD_0921, XerS and XerD had no impact on spnIII TRD shuffling (Fig 2B), ruling out the possibility of a cooperative interaction of these recombinases with CreX in a dual site-specific recombinase mechanism such as that observed in the Hin locus of Salmonella (23). In addition to SPD_0921, XerS and XerD there is a truncated recombinase (SPD_1013) that is expected to be non-functional, a mutant for this gene has not been tested. The D39 genome was searched for homology to known serine and tyrosine recombinases and no additional genes were found. Single knockouts of these site-specific recombinases were also analysed and shown to have no significant impact on TRD shuffling (Table S5). Due to the impact of environmental conditions such as incubation time and temperature each experimental analysis includes a paired WT or parental strain grown in the same conditions. As a result of these environmental variations multiple strains can be found repeated in table 2 with differing results. Differences in recombination are only determined between strains incubated at the same time under the same conditions.

In addition to testing site-specific recombination, we generated mutants in 10 recombination genes encompassing a variety of homologous recombination pathways (Table 1). No alterations in the rates of recombination were detected for any of the IRs by any of these mutations (Table 2), demonstrating that both RecA and RecFOR-mediated recombination are not involved in spnIII PV. To further investigate the role of RecA, a recA mutant was constructed in strain G54 which has an incomplete spnIII locus lacking TRD’s 1.2 and 2.1 (with a duplicated TRD 1.1) and the CreX recombinase (Fig. 1a). In this strain recombination at the spnIII locus occurs in a CreX-independent manner such that it is only capable of producing active
genes *hsdSB* and *hsdSE* as a result of recombination at one or both of the 85 bp and 333 bp repeats. Thus, this strain provides a useful model for determining the level of non-site-specific recombination. As in D39, no significant differences were observed in the frequency of recombination for the G54 background between active *hsdS* genes in *recA* mutants versus the parental strain (Table 2).

To further test whether RecA or other DNA repair pathways are involved in spnIII TRD shuffling, we tested strains in conditions known to induce DNA damage (Fig. 3). The D39 WT, *creX* frameshift mutant, *recA* mutant, and *creX recA* double mutant were grown with and without a sub-inhibitory concentration (1/2 the MIC value) of ciprofloxacin (CIP) (Fig. 3A). Strains were grown overnight for 16 hours (18-20 generations), *recA* mutants were incubated for 24 hours (16-18 generations) due to their slow growth rate. For colonies grown with and without CIP, no significant differences were observed in the percentages of active *hsdS* genes in any strain, suggesting that CIP has no impact on TRD shuffling at the spnIII locus. Additionally, we tested the impact of UV irradiation on spnIII recombination (Fig. 3B). An exponentially growing culture was split, with one half receiving a sub-lethal UV dose before plating. Approximately 90% of exposed cells were killed following UV exposure (data not shown) but surviving cells showed no significant variation in TRD shuffling when compared to unexposed cells. Both the ciprofloxacin and UV exposure observations confirm that SpnIII PV involves a RecA-independent recombination.

To directly identify proteins with a role in spnIII TRD shuffling, we generated a biotin labelled synthetic 1.2 kb double-stranded DNA fragment containing one 333bp repeat and two 15 bp repeats in an inverted orientation. This fragment was incubated with pneumococcal whole cell lysates from an exponential phase culture stimulated...
with CSP1, followed by purification of DNA-bound proteins using streptavidin-coated beads. Bound proteins were analysed by mass spectrometry. Most of the DNA binding proteins were encoded by essential genes (Table S6), except for the gene for the DNA repair protein UvrA. Based on these data we generated a uvrA mutant, but again inactivation of this protein had no significant impact on spnIII TRD shuffling (Table 2).
Discussion

Our findings indicate that SpnIII TRD shuffling is partially but not wholly controlled by the site-specific recombinase CreX (Fig. 2) however we have demonstrated that the process is independent of other site-specific recombinases in the pneumococcal genome (33). Interestingly, the frequency of inversions is linked to growth temperature, in line with earlier suggestions that PV of the SpnIII locus may offer an advantage in adapting to different environments (6, 7, 16, 17). The presence of multi-
hsdS systems in other pathogenic bacteria (9-14) may also prove to be a temperature regulated method of environmental adaptation.

As the CreX enzyme is encoded within the majority of S. pneumoniae spnIII loci and is highly conserved between strains, our findings strongly suggest that it is a major regulator of SpnIII PV across the pneumococcus genus. Our initial expectation, that all TRD shuffling was controlled by CreX, is in line with previous observations in M. pulmonis (30) and B. fragilis (10). However our finding that this is not the case aligns with recently reported data on the ability of these systems to recombine independently of their locus-encoded site-specific recombinase (6, 31, 32). Our analysis of strains (Fig S1) naturally lacking the CreX recombinase (G54 and BHN191) confirm that a locus-encoded site-specific recombinase is not essential to the process.

Surprisingly, the highest levels of CreX expression (Fig. 1) correspond with the lowest frequency of TRD shuffling on the longer repeats (i.e. in colonies founded by cells with active hsdSA and hsdSD genes), suggesting that it may not simply be the quantity of the CreX recombinase found within the cell that facilitates on-going DNA inversions. One potential explanation for this decrease is that the CreX recombinase may form complexes with the conserved 10 bp target sequence within the repeats,
thereby reducing or preventing the binding of other proteins that may play a part in TRD shuffling on the 85 and 333 bp IRs. The abolition of recombination at the smallest 15 bp IR indicates that the core target site for the CreX recombinase is the conserved core 10 bp sequence found in all three IRs (Fig. 2E), however the 5 bp immediately upstream also appear to play a role, facilitating recombination at the 333 bp but not the 85 bp IR. This is supported by the work of Li 2019 (32) who demonstrated that there was either a complete loss or significant impairment of 15 bp mediated inversions when single nucleotides of the 15 bp sequence were modified. This length of target size is similar to other site-specific recombinases such as Cre and the loxP site which both contain a core 13bp sequence (34).

In the absence of CreX, shuffling of the 15 bp inverted repeats was abrogated but still occurred on the longer repeats (Fig 2A). This complexity suggested the possibility of a dual site-specific recombinase mechanism such as the one seen in the Salmonella flagella locus (23). At this locus the majority of DNA inversions are carried out by the Hin recombinase. A much lower frequency of inversions occurs in the absence of Hin and is mediated by the Fin recombinase which is located elsewhere in the genome (23). Using recombination at the Hin locus as a model, we hypothesized that hsdS inversions may be mediated by a second site-specific recombinase within the pneumococcal genome. All functional site specific recombinases were deleted from the genome as single mutants, as well double knockout mutants where all strains lacked the CreX recombinase to test for redundancy. Our analysis showed that no other site-specific recombinases in the D39 genome (33) have any impact on TRD shuffling at either the short or long repeats of the spnIII locus (Table 2, Table S5).
After ruling out the CreX recombinase as the sole mechanism of control of hsdS inversions, we explored the possibility that the cell’s homologous recombination machinery was partially responsible for TRD shuffling. There are several different homologous recombination pathways in the pneumococcus including RecFOR and the genetic transformation pathway (26, 29, 35, 36). The universal recombinase RecA is crucial to both of these pathways, therefore we explored whether strains lacking RecA were still capable of TRD shuffling. If RecA is required for DNA inversions to occur at the spnIII locus a recA mutant would show a reduced level of hsdS inversions. We found that this was not the case, even when DNA damage was induced to stimulate repair pathways. It is known that ciprofloxacin-induced recombination is RecA-dependent (37). The observation that the D39 WT and the D39 recA mutant show no differences when grown with and without a sub-inhibitory concentration of ciprofloxacin confirms therefore our hypothesis that hsdS inversions are RecA-independent. In addition to testing RecA-mediated pathways a hexA mutant confirmed that the mismatch repair pathway is also not involved.

RecA-independent recombination at large direct repeats has been shown to occur at low but detectable levels in E. coli when sufficient sequence homology is present (38). The frequency of RecA-independent recombination can be significantly increased at shorter homologies (80 and 200 bp) when single-stranded exonucleases are no longer present (42). This provides evidence that in the correct conditions high levels of RecA-independent recombination are not only achievable but a regular occurrence (38) and are likely controlled by the physical proximity of DNA sequences (39). The mechanism behind RecA-independent recombination remains unknown, even in E. coli, however it has been proposed that DNA replication forks are required (38). The E. coli model showed high levels of RecA-
independent recombination at both 80 and 200 bp, and could therefore provide a suitable explanation for how TRD shuffling is controlled at the spnIII locus independently of site-specific recombination.

Recent work by Amarh (40) has shown for *E. coli* that the endonuclease SbcCD is capable of DNA cleavage at specific sites to repair double stranded breaks between sister chromosomes during DNA replication. SbcCD efficiently cuts just one of the sister chromosomes initiating repair of the double stranded break using the intact copy. The subsequent repair process requires the RecABCD complex (40) and hence cannot fully explain the RecA-independent recombination observed at the spnIII locus in *S. pneumoniae*. Nevertheless, there are homologs of these genes in *S. pneumoniae* and so this mechanism offers an attractive avenue for future work with the possibility that exonucleases could play a role in the on-going DNA breakage and repair after inversion.

A key outcome of the complex nature of the control of TRD shuffling at the spnIII locus is that switching between each variant type occurs at different rates and could exhibit differential regulation by DNA replication or non-replicative enzymes. The lack of recombinase cooperation in the on-going site-specific recombination of *hsdS* genes suggests that there are proteins in the pneumococcal genome capable of generating DNA inversions that have yet to be identified and categorised. The temperature dependent phenotype we have observed is likely to play a role in better understanding and working with these multi-*hsdS* systems in the future, and has the potential to explain the adaptability of some bacterial pathogens to a wide range of environmental temperatures. It is clear that there is a multifaceted control mechanism behind *hsdS* inversions in the SpnIII system. This may have phenotypic consequences, in particular where alternative SpnIII methylation patterns in a
bacterial population impact on adaptability and survival in challenging host environments.
Materials and Methods

Growth and storage of Bacterial strains

*S. pneumoniae* strains were grown in liquid cultures in either BBL Trypticase Soy Broth (TSB) (Becton Dickinson, USA) or Brain Heart Infusion (BHI) (Oxoid, UK). 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) was used. Plates were incubated at 37°C with 5% CO₂ for 16-18 hours unless otherwise stated. Where appropriate media was supplemented with antibiotics at the following concentrations; ciprofloxacin 0.25 µg/ml (0.016µg/ml for *recA* mutant strains), kanamycin 500 µg/ml, spectinomycin 200 µg/ml, chloramphenicol 10 µg/ml, streptomycin 500 µg/ml and erythromycin 10 µg/ml.

Transformation of *S. pneumoniae*

BHI broth was inoculated with a 1:100 dilution of a frozen culture and then incubated at 37°C, 5% CO₂ until an OD₅₉₀ 0.03-0.05 was reached. Cells were then diluted 1:10 into BHI supplemented with 0.1% (v/v) 1M CaCl₂, 0.2% (v/v) glucose, and 0.4% (v/v) Bovine Serum Albumin (BSA) (Sigma, UK), known as BHI-CTM, and incubated at 37°C for 60 minutes. 200µl of cells were incubated with a minimum of 50 ng of DNA and 0.625ng/ul of the competence stimulating peptide CSP1 (Inbios, Italy). Cells were incubated for 45 minutes before plating with appropriate antibiotics onto BHI agar containing 3% (v/v) horse blood. Plates were incubated overnight at 37°C, 5% CO₂. All mutants were confirmed by PCR and Sanger sequencing.

Generation of mutants

All mutant strains were created by the transformation of PCR generated fragments or extracted genomic DNA (where strains were supplied by the lab of Patrice Polard,
see table 1). For PCR generated fragments approximately 500 bp up- and down-
stream, of the gene to be deleted, were amplified using primers (see table S1) with
20 bp tails complementary to either an aad9 spectinomycin cassette (aad9,
Aminoglycoside O-nucleotidylyltransferase), Janus kanamycin cassette (rpsL,
ribosomal protein S12 and aphIII, aminoglycoside O-phosphotransferase) (41) or a
chloramphenicol cassette (cat, Chloramphenicol acetyl transferase) (42). Re-
amplification of the flanking regions with the appropriate cassette produced
constructs with an antibiotic marker suitable for transformation. To create unmarked
mutants the two-step process of Sung (41) using a Janus kanamycin cassette was
followed.

hsdS quantification

Primers [6-FAM]-AMRE74L and AMRE59 (table S1) were used to PCR amplify a 4.2
kb fragment from single colonies. Amplification was performed in a 25 µl reaction
consisting of 0.75 µl 10 mM AMRE74L forward, 0.75 µl 10 mM AMRE59 reverse
primer, 19.85 µl dH2O, 2.25 µl 11.1X buffer (for recipe see (43)), 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 or 1 single colony as template. The PCR requires a minimum
100pg DNA (Table S2). All PCR reactions were performed 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. 10-15ul of PCR product was digested following the manufacturer’s instructions
using 1U Dral (New England Biolabs, UK), 2U PleI (New England Biolabs, UK), 1X
CutSmart Buffer (New England Biolabs, UK) in a total volume of 20ul. Following
digestion each FAM labelled SpnIII variant has a unique size (Table S3) that can be
distinguished through capillary electrophoresis on an ABI prism Gene Analyser
(Applied Biosystems, USA). Data received from the ABI Prism Gene Analyser were analysed using Peak Scanner v1.0 software. All experiments to determine PV at the *spnIII* locus used colonies grown overnight on BHIA with 3% horse blood. For each experiment a minimum of 10 single colonies were picked directly into the PCR reaction. All colony data is presented as a mean with standard deviation (SD). The D39 stock maintained by our lab predominantly expresses *hsdSE* as its active *hsdS* gene, therefore colonies not founded by *hsdSE* cells were excluded from analysis.

**Extraction of *S. pneumoniae* genomic DNA**

1 ml of exponentially growing cells (OD600nm =~0.2) were centrifuged at 13,000 rpm, room temperature, for 2 minutes to pellet cells. Cells were then washed in 600 µl wash buffer (0.15M NaCl, 0.015M trisodiumcitrate, dH2O) 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. 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 according to the manufacturer's instructions. 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.

**RNAseq analysis**

For gene expression analysis pneumococcal strains were grown to OD590 0.15-0.18 in triplicate. 5 ml of cells were added to 1 ml of an ice cold 95% ethanol 5% phenol solution, before centrifugation at 4000 rpm for 10 mins. 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
instruments (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 50 μl TE with 3 mg/ml lysozyme and incubated at 37°C for 10-20 mins. RNAseq analysis used a MiSeq Desktop Processor (Illumina, USA) and the ScriptSeq Complete Kit (Bacteria) (CamBio, UK) which includes an rRNA depletion step. Raw data was trimmed using Trimmomatic-0.32 and aligned using BWA-Mem and samtools. Expression data was generated using Rockhopper v2.0.3 (44) with D39 as the reference genome (NC_008533.2).

**UV sensitivity assay**

Cells were grown to an OD$_{690}$ of ~0.2 in 50 ml TSB. Cultures were centrifuged at 4,000rpm for 10 minutes to pellet cells. Cells 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 90mm 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. To determine the impact of UV on SpnIII PV a minimum of 10 single colonies grown for 18 h at 37°C, 5% CO$_2$ were selected and used as the DNA template in the hsdS quantification protocol described above.

**Protein pull-down and mass spectrometry**

D39 was grown as described in the transformation protocol in a 50 ml culture. Bacterial cultures were harvested by centrifugation, washed three times with ultrapure water and frozen at -80°C for a minimum of 1 hour. Cells were freeze thawed a total of 3 times on ice. Cells were resuspended in 1 ml THES buffer (50 mM Tris HCl (pH 7.5), 10 mM EDTA, 20% sucrose, 140 mM NaCl) and sonicated on ice in 10 sec pulses, separated by 1 minute.
To test the binding of pneumococcal proteins to the hsdS repeats, a synthetic DNA region of 357 bp (ThermoFisher) containing two copies of the 15 bp IR and 132 bp of the 333 bp IR was amplified and 5'-biotinylated with primers [BTN]-AMRE05 and IVPD1 (table S1). A PCR amplified fragment of the pncA gene, containing no hsdS repeats, was used as a control to detect non-specific binding. The PCR product was concentrated with Amicon/Microcon centrifugal filter (Millipore) columns so that the final concentration of the probe was 200-450 ng/ul and then incubated 3 times for 20 minutes at room temperature with pre-washed streptavidin magnetic beads (Pierce) to allow binding. The magnetic beads were pre-washed 3 times with 2X Binding/Washing buffer (10 mM Tris-HCl (pH 7.5), 1 mM EDTA, 2 M NaCl). The probe-bead complex was then washed for 3 times with TE and twice with BS/THES (44.3% THES buffer, 20% BS buffer, 35.7% Nuclease free water). To allow binding of the bacterial proteins to the DNA probe, the supernatant of D39 was incubated with the probe-bead complex at room temperature for 1 hour. After five washes with BS/THES, the protein-DNA complex was eluted with SDS-PAGE reducing sample buffer and the proteins separated by SDS/PAGE electrophoresis using Bis-Tris and Tris-Acetate polyacrylamide gels (Invitrogen) and then stained with Coomassie blue, according to standard procedures. The gel was then submitted for mass spectrometry analysis on a LTQ-Orbitrap-Velos-ETD - SN03106B.
Acknowledgements: The work was funded by the BBSRC grant BB/N002903/1 to MRO, NJC and SDB and MRC grant MR/M003078/1 to MRO. NJC is also supported by a Sir Henry Dale Fellowship, jointly funded by Wellcome and the Royal Society (Grant Number 104169/Z/14/Z). The contribution of the sequencing facility Nucleus and the proteomic facility PNACL of the University of Leicester is acknowledged. Parts of this work are included in the PhD thesis of MDSC which can be accessed from the University of Leicester Research Archive.

Competing interest: Authors declare no competing interest with respect to the work performed in the manuscript.
1. Southern J, Andrews N, Sandu P, Sheppard CL, Waight PA, Fry NK, Van Hoek AJ, Miller E. 2018. Pneumococcal carriage in children and their household contacts six years after introduction of the 13-valent pneumococcal conjugate vaccine in England. PLoS One 13:1–14.

2. Troeger C, Blacker B, Khalil IA, Rao PC, Cao J, Zimsen SRM, Albertson SB, Deshpande A, Farag T, Abeze Z, Adetifa IMO, Adhikari TB, Akibu M, Al Lami FH, Al-Eyadhy A, Alvis-Guzman N, Amare AT, Amoako YA, Antonio CAT, Aremu O, Asfaw ET, Asgedom SW, Atey TM, Attia EF, Avokpah EFGA, Ayele HT, Ayuk TB, Balakrishnan K, Barac A, Bassat Q, Behzadifar M, Behzadifar M, Bhaumik S, Bhutta ZA, Bijani A, Brauer M, Brown A, Camargos PM, Castañeda-Orjuela CA, Colombara D, Conti S, Dadi AF, Dandona L, Dandona R, Do HP, Dubljanin E, Edessa D, Elkout H, Endries AY, Fijabi DO, Foreman KJ, Forouzanfar MH, Fullman N, Garcia-Basteiro AL, Gessner BD, Gething PW, Gupta R, Gupta T, Hailu GB, Hassen HY, Hedayati MT, Heidari M, Hibst DT, Horita N, Ilesanmi OS, Jakovljevic MB, Jamal AA, Kahsay A, Kasaeian A, Kassa DH, Khader YS, Khan EA, Khan MN, Khang YH, Kim YJ, Kissoon N, Knibbs LD, Kochhar S, Koula PA, Kumar GA, Lodha R, Magdy Abd El Razek H, Malta DC, Mathew JL, Mengistu DT, Mezgebe HB, Mohammad KA, Mohammed MA, Momeniha F, Murthy S, Nguyen CT, Nielsen KR, Ningrum DNA, Nirayo YL, Oren E, Ortiz JR, PA M, Postma MJ, Qorbani M, Quansah R, Rai RK, Rana SM, Ranabhat CL, Ray SE, Rezai MS, Ruhago GM, Safiri S, Salomon JA, Sartorius B, Savic M, Sawhney M, She J, Sheikh A, Shiferaw MS, Shigematsu M, Singh JA, Somayaji R, Stanaway JD, Sufiyan MB, Taffere GR, Temsah MH, Thompson MJ, Tobe-Gai R, Topor-Madry R,
Tran BX, Tran TT, Tuem KB, Ukwaja KN, Vollset SE, Walson JL, Weldegebreal F, Werdecker A, West TE, Yonemoto N, Zaki MES, Zhou L, Zodpey S, Vos T, Naghavi M, Lim SS, Mokdad AH, Murray CJL, Hay SI, Reiner RC. 2018. Estimates of the global, regional, and national morbidity, mortality, and aetiologies of lower respiratory infections in 195 countries, 1990–2016: a systematic analysis for the Global Burden of Disease Study. Lancet Infect Dis 18:1191–1210.

3. De Ste Croix M, Vacca I, Kwun MJ, Ralph JD, Bentley SD, Haigh R, Croucher NJ, Oggioni MR. 2017. Phase-variable methylation and epigenetic regulation by type I restriction-modification systems. FEMS Microbiol Rev 41:S3–S15.

4. Gerlini A, Colomba L, Furi L, Braccini T, Manso AS, Pammolli A, Wang B, Vivi A, Tassini M, van Rooijen N, Pozzi G, Ricci S, Andrew PW, Koedel U, Moxon ER, Oggioni MR. 2014. The Role of Host and Microbial Factors in the Pathogenesis of Pneumococcal Bacteraemia Arising from a Single Bacterial Cell Bottleneck. PLoS Pathog 10:e1004026.

5. Mostowy R, Croucher NJ, Hanage WP, Harris SR, Bentley S, Fraser C. 2014. Heterogeneity in the Frequency and Characteristics of Homologous Recombination in Pneumococcal Evolution. PLoS Genet 10:e1004300.

6. Li J, Li JW, Feng Z, Wang J, An H, Liu Y, Wang Y, Wang K, Zhang X, Miao Z, Liang W, Sebra R, Wang G, Wang WC, Zhang JR. 2016. Epigenetic Switch Driven by DNA Inversions Dictates Phase Variation in Streptococcus pneumoniae. PLoS Pathog 12:e1005762.

7. Lees J, Kremer PH, Manso AS, Croucher NJ, Ferwerda B, Serón MV, Oggioni MR, Parkhill J, Brouwer MC, van der Ende A, van de Beek D, Bentley SD.
2017. Large scale genomic analysis shows no evidence for repeated pathogen adaptation during the invasive phase of bacterial meningitis in humans. Micro Genomics 3:e000103.

8. Tettelin H, Nelson KE, Paulsen IT, Eisen JA, Read TD, Scott P. 2001. Complete Genome Sequence of a Virulent Isolate of Streptococcus pneumoniae. Science 293:498–506.

9. Furuta Y, Namba-Fukuyo H, Shibata TF, Nishiyama T, Shigenobu S, Suzuki Y, Sugano S, Hasebe M, Kobayashi I. 2014. Methylome Diversification through Changes in DNA Methyltransferase Sequence Specificity. PLoS Genet 10:e1004272.

10. Cerdeño-Tárraga AM, Patrick S, Crossman LC, Blakely G, Abratt V, Lennard N, Poxton I, Duerden B, Harris B, Quail Ma, Barron A, Clark L, Corton C, Doggett J, Holden MTG, Larke N, Line A, Lord A, Norbertczak H, Ormond D, Price C, Rabbinowitsch E, Woodward J, Barrell B, Parkhill J. 2005. Extensive DNA inversions in the B. fragilis genome control variable gene expression. Science 307:1463–1465.

11. Serfiotis-Mitsa D, Roberts GA, Cooper LP, White JH, Nutley M, Cooper A, Blakely GW, Dryden DTF. 2008. The Orf18 Gene Product from Conjugative Transposon Tn916 Is an ArD Antirestriction Protein that Inhibits Type I DNA Restriction–Modification Systems. J Mol Biol 383:970–81.

12. Claesson MJ, Li Y, Leahy S, Canchaya C, van Pijkeren JP, Cerdeño-Tárraga AM, Parkhill J, Flynn S, O’Sullivan GC, Collins JK, Higgins D, Shanahan F, Fitzgerald GF, van Sinderen D, O’Toole PW. 2006. Multireplicon genome architecture of Lactobacillus salivarius. Proc Natl Acad Sci U S A 103:6718–
13. Willemse N, Schultsz C. 2016. Distribution of Type I Restriction–Modification Systems in Streptococcus suis: An Outlook. Pathogens 5:E62.

14. Willemse N, Howell KJ, Weinert LA, Heuvelink A, Pannekoek Y, Wagenaar JA, Smith HE, van der Ende A, Schultsz C. 2016. An emerging zoonotic clone in the Netherlands provides clues to virulence and zoonotic potential of Streptococcus suis. Sci Rep 6:28984.

15. Dybvig K, Yu H. 1994. Regulation of a restriction and modification system via DNA inversion in Mycoplasma pulmonis. Mol Microbiol 12:547–560.

16. Croucher NJ, Coupland PG, Stevenson AE, Callendrello A, Bentley SD, Hanage WP. 2014. Diversification of bacterial genome content through distinct mechanisms over different timescales. Nat Commun 19:5471.

17. Manso AS, Chai MH, Atack JM, Furi L, De Ste Croix M, Haigh R, Trappetti C, Ogunniyi AD, Shewell K, Boitano M, Clark T a., Korfach J, Blades M, Mirkes E, Gorban AN, Paton JC, Jennings MP, Oggioni MR. 2014. A random six-phase switch regulates pneumococcal virulence via global epigenetic changes. Nat Commun 5:5055.

18. Weiser JN, Austrian R, Sreenivasan PK, Masure HR. 1994. Phase variation in pneumococcal opacity: Relationship between colonial morphology and nasopharyngeal colonization. Infect Immun 62:2582–2589.

19. Oliver MB, Roy B, Kumar R, Lefkowitz EJ, Swords WE. 2017. Streptococcus pneumoniae TIGR4 Phase-Locked Opacity Variants Differ in Virulence Phenotypes 2:e00386-17.

20. Grindley NDF, Whiteson KL, Rice PA. 2006. Mechanisms of Site-Specific
21. Ghosh K, Van Duyne G. 2002. Cre–loxP biochemistry. Methods 28:374–83.
22. Heichman KA, Johnson RC. 1990. The Hin Invertasome: Protein-Mediated Joining of Distant Recombination Sites at the Enhancer. Science 249:511–517.
23. Kutsukake K, Nakashima H, Tominaga A, Abo T. 2006. Two DNA invertases contribute to flagellar phase variation in Salmonella enterica serovar typhimurium strain LT2. J Bacteriol 188:950–957.
24. Claverys JP, Martin B, Polard P. 2009. The genetic transformation machinery: Composition, localization, and mechanism. FEMS Microbiol Rev 33:643–656.
25. Johnston C, Campo N, Bergé MJ, Polard P, Claverys JP. 2014. Streptococcus pneumoniae, le transformiste. Trends Microbiol 22:113–119.
26. Johnston C, Mortier-Barrière I, Granadel C, Polard P, Martin B, Claverys J-P. 2015. RecFOR Is Not Required for Pneumococcal Transformation but Together with XerS for Resolution of Chromosome Dimers Frequently Formed in the Process. PLoS Genet 11:e1004934.
27. Mortier-Barrière I, Velten M, Dupaigne P, Mirouze N, Piétrement O, McGovern S, Fichant G, Martin B, Noirot P, Le Cam E, Polard P, Claverys JP. 2007. A Key Presynaptic Role in Transformation for a Widespread Bacterial Protein: DprA Conveys Incoming ssDNA to RecA. Cell 130:824–836.
28. Morimatsu K, Kowalczykowski SC. 2003. RecFOR proteins load RecA protein onto gapped DNA to accelerate DNA strand exchange: A universal step of recombinational repair. Mol Cell 11:1337–1347.
29. Halpern D, Gruss A, Claverys JP, El Karoui M. 2004. rexAB mutants in Streptococcus pneumoniae. Microbiology 150:2409–2414.

30. Sitaraman R, Denison AM, Dybvig K. 2002. A unique, bifunctional site-specific DNA recombinase from Mycoplasma pulmonis. Mol Microbiol 46:1033–1040.

31. Kwun MJ, Oggioni MR, De Ste Croix M, Bentley SD, Croucher NJ. 2018. Excision-reintegration at a pneumococcal phase-variable restriction-modification locus drives within- and between-strain epigenetic differentiation and inhibits gene acquisition. Nucleic Acids Res 46:11438–11453.

32. Li J-W, Li C, Li J, Wang J, Zhang J-R. 2019. Molecular Mechanisms of Hsds Inversions in the Cod Locus of Streptococcus Pneumoniae. J Bacteriol 201:e00581-18.

33. Brückner R, Nuhn M, Reichmann P, Weber B, Hakenbeck R. 2004. Mosaic genes and mosaic chromosomes-genomic variation in Streptococcus pneumoniae. Int J Med Microbiol 294:157–168.

34. Sadowski P. 1986. Site-specific recombinases: changing partners and doing the twist. J Bacteriol 165:341–347.

35. Martin B, Garcia P, Castanié MP, Claverys JP. 1995. The recA gene of Streptococcus pneumoniae is part of a competence-induced operon and controls lysogenic induction. Mol Microbiol 15:367–379.

36. Mirouze N, Bergé MA, Soulet A-L, Mortier-Barrière I, Quentin Y, Fichant G, Granadel C, Noirot-Gros M-F, Noirot P, Polard P, Martin B, Claverys J-P. 2013. Direct involvement of DprA, the transformation-dedicated RecA loader, in the shut-off of pneumococcal competence. Proc Natl Acad Sci U S A 110:E1035-44.
37. López E, Elez M, Matic I, Blázquez J. 2007. Antibiotic-mediated recombination:
Ciprofloxacin stimulates SOS-independent recombination of divergent
sequences in Escherichia coli. Mol Microbiol 64:83–93.

38. Dutra BE, Sutera VA, Lovett ST. 2007. RecA-independent recombination is
efficient but limited by exonucleases. Proc Natl Acad Sci U S A 104:216–221.

39. Lovett ST, Gluckman TJ, Simon PJ, Sutera VA, Drapkin PT. 1994.
Recombination between repeats in Escherichia coli by a recA-independent,
proximity-sensitive mechanism. MGG Mol Gen Genet 245:294–300.

40. Amarh V, White MA, Leach DRF. 2018. Dynamics of RecA-mediated repair of
replication-dependent DNA breaks. J Cell Biol 217:2299–2307.

41. Sung CK, Li H, Claverys JP, Morrison DA. 2001. An rpsL Cassette, Janus,
for Gene Replacement through Negative Selection in Streptococcus
pneumoniae. An rpsL Cassette, Janus, for Gene Replacement through
Negative Selection in Streptococcus pneumoniae. Appl Environ Microbiol
67:5190–5196.

42. Claverys JP, Dintilhac A, Pestova EV, Martin B, Morrison DA. 1995.
Construction and evaluation of new drug-resistance cassettes for gene
disruption mutagenesis in Streptococcus pneumoniae, using an ami test
platform. Gene 164:123–128.

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

44. McClure R, Balasubramanian D, Sun Y, Bobrovskyy M, Sumby P, Genco CA,
Vanderpool CK, Tjaden B. 2013. Computational analysis of bacterial RNA-Seq
45. Attaiech L, Olivier A, Mortier-Barrière I, Soulet AL, Granadel C, Martin B, Polard P, Claverys JP. 2011. Role of the single-stranded dna-binding protein ssbb in pneumococcal transformation: Maintenance of a reservoir for genetic plasticity. PLoS Genet 7:e1002156.

46. Prudhomme M, Libante V, Claverys J-P. 2002. Homologous recombination at the border: insertion-deletions and the trapping of foreign DNA in Streptococcus pneumoniae. Proc Natl Acad Sci U S A 99:2100–2105.

47. Martin B, Sharples GJ, Humbert O, Lloyd RG, Claverys J-P. 1996. The mmsA locus of Streptococcus pneumoniae encodes a RecG-like protein involved in DNA repair and in three-strand recombination. Mol Microbiol 19:1035–1045.

48. Carver T, Harris SR, Berriman M, Parkhill J, McQuillan JA. 2012. Artemis: An integrated platform for visualization and analysis of high-throughput sequence-based experimental data. Bioinformatics 28:464–469.
Table 1: Mutant strains used in this study

| Strain number | Mutant | Gene number | Reference |
|---------------|--------|-------------|-----------|
| MRO633        | D39 creX | SPD_0452    | This work |
| MRO639        | D39 creX SPD_0921::aad9 | SPD_0452, SPD_0921 | This work |
| MRO652        | D39 creX xerD ::aad9 | SPD_0452, SPD_1657 | This work |
| MRO660        | D39 recU ::aad9 | SPD0337 | This work |
| MRO661        | D39 spoJ ::aad9 | SPD_2069 | This work |
| MRO669*       | D39 recA ::cat | SPD_1739 | (35) |
| MRO668        | D39 uvrA ::aphIII | SPD0176 | This work |
| MRO796*       | G54 recA ::cat | SPG_1849 | (35) |
| MRO797        | D39 creX xerS ::aad9 | SPD_0452, SPD_1023 | This work |
| MRO810*       | D39 creX recA ::cat | SPD_0452, SPD_1739 | (35) |
| MRO811*       | D39 ssbB ::aad9 | SPD_1711 | (45) |
| MRO812*       | D39 recN ::aad9 | SPD_1062 | (27) |
| MRO813*       | D39 recF ::ermB | SPD_2054 | (26) |
| MRO814*       | D39 recR ::aphIII | SPD_1485 | (26) |
| MRO815*       | D39 hexA ::ermB | SPD_1903 | (46) |
| MRO816*       | D39 recG ::cat | SPD_1507 | (47) |

*Strains were generated by transformation of genomic DNA from listed reference strain.
|                         | 333bp Repeat |          | 85bp Repeat |          | 15bp Repeat |          |
|-------------------------|--------------|----------|-------------|----------|-------------|----------|
|                         | Mean(SD) %   | P value  | Mean(SD) %  | P value  | Mean(SD) %  | P value  |
| D39                     | 15.8±5.4     | -        | 1.2±0.8     | -        | 4.1±1.9     | -        |
| D39 creX                | 12.7±2.0     | 0.0007   | 1.7±0.6     | NS       | 0.0±0.0     | <0.0001  |
| D39 creX                | 12.3±2.0     | -        | 1.7±0.6     | -        | 0.0±0.0     | -        |
| D39 creX SPD_0921::aad9 | 13.2±2.1     | NS       | 1.6±0.6     | NS       | 0.0±0.0     | NS       |
| D39 creX                | 30.8±8.9     | -        | 8.7±1.2     | -        | 0.0±0.0     | -        |
| D39 creX xerS ::aad9    | 27.4±7.0     | NS       | 8.0±2.7     | NS       | 0.0±0.0     | NS       |
| D39 creX xerD ::aad9    | 29.1±3.0     | NS       | 7.6±1.2     | NS       | 0.0±0.0     | NS       |
| D39                     | 10.8±4.4     | -        | 1.5±0.6     | -        | 3.1±2.7     | -        |
| D39 ssbB ::aad9         | 9.8±2.7      | NS       | 1.6±0.8     | NS       | 2.6±0.4     | NS       |
| D39 recN ::aad9         | 10.9±2.1     | NS       | 1.5±0.7     | NS       | 2.2±1.1     | NS       |
| Strain   | Gene   | Mean | SD   | Mean | SD   | Mean | SD   | p-value |
|----------|--------|------|------|------|------|------|------|---------|
| D39 recF::ermB | 10.4±2.0 | NS   | 1.8±0.5 | NS   | 3.7±5.2 | NS |
| D39 recR::aphIII | 8.8±6.3 | NS   | 1.3±0.6 | NS   | 3.3±0.7 | NS |
| D39 recG::cat | 8.0±1.6 | NS   | 1.4±0.4 | NS   | 2.7±0.7 | NS |
| D39 hexA::ermB | 11.3±1.8 | NS   | 1.2±0.7 | NS   | 2.4±0.8 | NS |
| D39 spoJ::aad9 | 19.7±5.2 | -    | 2.2±1.4 | -    | 2.4±1.3 | -  |
| D39 recU::aad9 | 22.6±4.3 | NS   | 4.4±1.7 | NS   | 2.9±1.1 | NS |
| D39 recA::cat | 16±8.1 | -    | 1.4±1.5 | -    | 4.4±0.7 | -  |
| D39 uvrA::aphIII | 15.4±2.9 | NS   | 3.3±1.3 | NS   | 4.6±2.9 | NS |
| D39 recA::cat | 25.1±7.7 | -    | 8.9±2.2 | -    | 2.8±0.9 | -  |
| D39 uvrA::aphIII | 30.9±7.3 | NS   | 6.8±0.6 | NS   | 1.7±0.9 | NS |
| D39 recA::cat | 18.6±4.6 | -    | 10.1±1.7 | -    | 2.6±1.9 | -  |
| D39 recA::cat | 19.5±10.5 | NS   | 7.6±4.6 | NS   | 2.2±4.0 | NS |
|       |       |       |
|-------|-------|-------|
| G54*  | 36.6±1.9 | -     |
| G54* recA ::cat | 37.4±2.4 | NS |

* This strain lacks TRD's 1.2 and 2.1.
Figure legends

Figure 1 – Active hsdS distribution in *S. pneumoniae* (A) All sequenced pneumococcal isolates have an *spnIII* locus. However, in some strains, including G54 (Accession number NC_011072), the locus appears to have undergone a deletion event, when compared to D39 the G54 locus lacks the truncated hsdS’ gene, the creX recombinase and shows a duplication of TRD 1.1 with no TRD 1.2 (target recognition domain TRD numbering within the arrow of the *hsdS* genes). This results in G54 being limited to active hsdS genes *hsdSB* and *hsdSE* only. (B-C) Due to recombination at the *spnIII* locus on the three inverted repeats shown as hashed, dotted and solid black boxes, the orientation of the creX recombinase can be aligned with *hsdR*, *hsdM* and the active hsdS (panel B), or aligned with the unexpressed hsdS’ and hsdS” genes (B). RNAseq data for an hsdSA (red) expressing strain (B) and an hsdSB (blue) expressing strain (C) were mapped in Artemis (48) using sorted BAM files. (D-E) Bars represent the mean percentage and SD of recombination within a minimum of 10 single colonies. Recombination events are expressed as a percentage of the total population by repeat. Recombination can occur on the 333 bp IR (dashed), 85 bp IR (dotted), 15 bp IR (solid grey) and a small number are generated by >2 recombination events (checked). (D) *S. pneumoniae* D39 was grown at 23°C, 37°C and 42°C to determine if temperature can impact the frequency of TRD shuffling. (E) Distribution of active hsdS genes in single colonies from each lineage (A-F). Statistical analysis was conducted using a 2way ANOVA. * P ≤ 0.05, ** P ≤ 0.01, *** P ≤ 0.001 and *** P ≤ 0.0001.

Figure 2 – Site specific and homologous recombination in SpnIII TRD switching. Bars represent the mean percentage and SD within a minimum of 10
hsdSE single colonies for each strain. Recombination can occur on the 333 bp IR (dashed), 85 bp IR (dotted), 15 bp IR (solid grey) and a small number can be generated by >2 recombination events (checked). (A) The CreX recombinase is responsible for all switching on the smallest inverted repeat within the locus. In addition, there is significantly less recombination at the 333bp IR in this strain (P=0.0007), the 85 bp IR is unaffected by the loss of CreX. (B) Double knockout mutants lacking creX and each other known, functional site-specific recombinases within the D39 genome (33) were tested for their impact on TRD shuffling. No significant differences were found between the parent creX mutant and the double knockout mutants. (C-D) The involvement of homologous recombination to facilitate TRD switching was tested in two different genetic background; D39 (C) and G54 (D). The active gene distribution was not found to be significantly different between WT and recA mutants. G54 has an alternative spnIII loci lacking TRD’s 1.2, 2.1 and 2.1 and the creX recombinase gene. The absence of the creX gene in G54 shows that the absence of RecA does not lead to a significant different in either a creX WT or creX mutant background. Statistical analysis was conducted using a 2way ANOVA. * P ≤ 0.05, ** P ≤ 0.01, *** P ≤ 0.001 and **** P ≤ 0.0001. (E) The conserved 10 bp consensus in the three inverted repeats within the spnIII locus of the site-specific CreX recombinase.

Figure 3 – The impact of DNA damage on spnIII TRD switching. Bars represent the mean percentage and SD within a minimum of 10 hsdSE single colonies for each strain. Recombination can occur on the 333 bp IR (dashed), 85 bp IR (dotted), 15 bp IR (solid grey) and a small number can be generated by >2 recombination events (checked). (A) D39, D39 creX, D39 recA and D39 creX recA were exposed to a sub
MIC dose of ciprofloxacin, after 16-18hr of growth colonies were analysed for recombination events (B) D39 was unexposed to a sub lethal dose of UV and recombination events in 20 single colonies were compared to an unexposed control. Significance was tested using a 2way ANOVA.
Figure 2
