Methylation warfare: interaction of pneumococcal bacteriophages with their host

Leonardo Furi¹, Liam A. Crawford¹, Guillermo Rangel-Pineros¹, Ana S. Manso¹, Megan De Ste Croix¹, Richard D. Haigh¹, Min J. Kwun², Kristine Engelsen Fjelland³, Gregor D. Gilfillan³, Stephen D. Bentley⁴, Nicholas J. Croucher², Martha R. Clokie¹, Marco R. Oggioni¹,∗

¹Department of Genetics and Genome Biology, University of Leicester, Leicester, UK
²MRC Centre for Global Infectious Disease Analysis, Department of Infectious Disease Epidemiology, Imperial College London, London, United Kingdom.
³Department of Medical Genetics, Oslo University Hospital and University of Oslo, Oslo, Norway
⁴Parasites and Microbes, Wellcome Trust Sanger Institute, Hinxton, United Kingdom.

*Correspondence to: Marco R Oggioni. Department of Genetics and Genome Biology, University of Leicester, University Road, Leicester, LE1 7RH, UK. Phone. +44 (0)116 252 2261, Email: mro5@leicester.ac.uk

&Current address: Department of Immunology and Transfusion Medicine, Haugesund Hospital, Haugesund, Norway
Abstract

Virus-host interactions are regulated by complex co-evolutionary dynamics. In *S. pneumoniae* phase-variable Type I restriction modification (R-M) systems are part of the core genome. We hypothesised that the ability of the R-M systems to switch between six target DNA specificities also has a key role in preventing the spread of bacteriophages. Using the streptococcal temperate bacteriophage SpSl1, we showed that the variants of both the SpnIII and SpnIV R-M system were able to restrict invading bacteriophage proportional to the number of target sites in the bacteriophage genome. In addition to restriction of lytic replication SpnIII also led to abortive infection in the majority of host cells. During lytic infection, transcriptional analysis found evidence of phage-host interaction through the strong upregulation of the *nrdR* nucleotide biosynthesis regulon. During lysogery, the phage had less effect on host gene regulation. This research demonstrates a novel combined bacteriophage restriction and abortive infection mechanism, highlighting the importance that the phase-variable Type I R-M systems have in the multi-functional defence against bacteriophage infection in the respiratory pathogen *S. pneumoniae*.

Abstract of Importance

With antimicrobial drug resistance becoming an increasing burden on human health, much attention has been focussed on the potential use of bacteriophages and their enzymes as therapeutics. However, the investigations into the physiology of the complex interactions of bacteriophages with their hosts has attracted far less attention in comparison. This work describes the molecular characterisation of the infectious cycle of a bacteriophage of the important human pathogen *Streptococcus pneumoniae* and explores the intricate relationship between phase variable host defence mechanisms and the virus. This is the first report showing how a phase variable type I restriction modification system is involved in bacteriophage restriction, whilst also providing an additional level of infection control through abortive infection.
Introduction

In almost all ecosystems studied, bacteria are vastly outnumbered by coexisting bacteriophages (9). The survival of both is sustained by an endless state of co-evolutionary cycles of adaptation and counter-adaptation (34). Bacteria have therefore evolved many resistance mechanisms to limit phage infection, including the inhibition of phage attachment to the cell, cleavage of the invading phage genome, and altruistic programmed cell death to abort phage infection. In response, phages evolved the so-called host-range expanding adaptations. This continuous selection for defence and counter-defence traits is often described as an “arms race” (69).

*Streptococcus pneumoniae*, the pneumococcus, is one of the most important human pathogens, being the major cause of community-acquired pneumonia, meningitis, and acute otitis media (16, 30). Interestingly, most sequenced *S. pneumoniae* strains contain prophage genes which sum to approximately 6% of the pneumococcal genome (7), with 30% of all pneumococcal strains harbouring an intact prophage within their genome (7). Despite this apparent pneumococcal phage abundance, only a few bacteriophages have been characterised and studied, including DP-1, CP-1 and the temperate phage MM1 (10, 20, 40, 48). It is only recently that pneumococcal phage biology has again attracted interest with the majority of research focusing on the use of their endolysins as potential therapeutics (10, 20, 40, 48) as opposed to their genomics and their influence on the biology of the pneumococcus (35, 57, 65). However, recent genome wide association studies (GWAS) have indicated phage have a strong effect on their host cells’ epidemiology. One found prophage disrupting the *comYC* gene were asymptomatically carried for shorter durations than expected (38), and the second showed a strong correlation between 30-day patient mortality from *S. pneumoniae*-related sepsis and the phage tail fibre protein gene (*PblB*), a protein shown to bind and activate platelets (76). Despite pneumococcal prophages showing a high degree of genetic heterogeneity (7, 79) this *PblB* tail fibre protein is present in 72% of all sequenced pneumococcal bacteriophages (7).

Phage defence mechanisms in *S. pneumoniae* rely primarily on a panel of restriction modifications (R-M) systems (77). Strains harbour alternatively one of the three allelic variants of the DpnI, DpnII, or DpnIII type II R-M systems (51), one or two further type II R-M systems and two conserved phase variable type I R-M systems (11, 45). Among them, the Type I R-M system SpnIII, is most conserved existing in almost all *S. pneumoniae* isolates (11, 12). The SpnIII operon comprises a contingency locus (49), defined as the “inverting
variable restriction” (ivr) locus (11), in which high-frequency rearrangements occur between
the active hsdS and two other untranscribed hsdS-like genes (11, 45, 71). These
recombination events lead to the alternate formation of six hsdS genes with different DNA
sequence specificities that coexist in a bacterial population (Fig. 1B) (45). The variability of
SpnIII has been hypothesised to have a role in preventing phage transmission in clonally
related bacterial populations (Fig. 1C) (11, 12). The presence of the different enzyme forms
should prevent the spread of phages within populations, even within the same strain. Once a
bacteriophage has infected one variant of the SpnIII system, it’s susceptibility to restriction
from the other five variants would theoretically be unchanged in comparison with a naïve
phage (Fig. 1C). Despite this, experimental demonstrations of functionality of this R-M
system in dsDNA cleavage and in phage resistance are still lacking.

Another well-studied bacteriophage defence mechanism is the induction of altruistic
programmed cell death of a phage-infected cell; this process has been defined as Abortive
infection (Abi) or phage exclusion system. These systems promote death of infected cells in
order to abort phage replication and limit its further spread within the population. Currently,
only a few examples of toxin-antitoxin (TA) systems that protect bacteria from phages have
been described (18, 25, 56, 59, 66) and are considered a subgroup of the Abi systems (67).
Recently, R-M systems have been compared to TA systems (50) as they can also trigger
post-segregational killing (2, 28, 29) and have been shown to be important in bacteriophage
defence. The Type IV McrBC R-M system, previously characterized in Escherichia coli,
specifically recognises the R“C (R = A or G; “C = 5&7C, 5&7C, or 5&75&7C) pattern and cleaves the
dNA between two recognition sites in vitro. Interestingly, McrBC is known to induce Abi upon
bacteriophage infection (42, 70). The McrBC system has also been shown to act as a
bacteriophage defence in S. mitis, reducing the rate of DNA replication of the lytic phage DP-
1 (39).

Here we report the isolation and characterization of a streptococcal temperate
bacteriophage, SpSL1; this was then used as a tool to confirm the methylation and
restriction activities of the phase-variable SpnIII system in S. pneumoniae. Phage-host
interactions were also evaluated by means of RNA-sequencing analysis of bacterial and
phage genome transcriptomes. Finally, we show the SpnIII system is involved in
programmed cell death via an Abi mechanism that inhibits the proliferation of the SpSL1
bacteriophage with a different SpnIII methylation pattern.
**Material and Methods**

**Bacterial strains and growth conditions**

Avery's type 2 *S. pneumoniae* strain D39 (3, 26), its unencapsulated Rx1 derivative DP1004 (26, 58) and the mutants derived from these strains (Table S1) were routinely cultured, where not otherwise specified, in Tryptic Soy Broth (TSB, Becton Dickinson) at 37°C or on Tryptic Soy Agar (TSA) plates with 3% defibrinated horse blood at 37°C in a 5% CO₂ incubator (31, 33, 73). When performing bacteriophage sampling and propagation, or to produce the bacterial lawn for the Double-Layer Agar (D-LA) assay (36), strains were grown instead on CAT-Galactose medium (Bacto Casitone 10 g/l, Bacto Tryptone 10 g/l, Yeast Extract 0.5 g/l, NaCl 5 g/l, K₂HPO₄ 15 mM, and 0.2% D-Galactose) at 32°C up to an OD₆₅₀nm of 0.1 (58). CAT-Galactose Agar medium was supplemented with 250 U/ml of catalase (Sigma, Germany) and plates incubated at 37°C, 5% CO₂.

**Mutant construction**

In both the D39 and DP1004 background, a mutant carrying a deleted *spnIII* locus (*spnIII* deleted) and six mutants expressing only one of the possible *hsdS* variants (*spnD39III-A-F, spnDP1004III-A-F*) (Table S1) were constructed by the gene SOEing technique as previously described (27, 45, 74). In brief, a PCR generated fragment that included an antibiotic selection marker (spectinomycin or kanamycin) and two flanking regions with homology to the surrounding sequence of the genomic locus to be mutated, was transformed into naturally competent pneumococcal cells (21, 74). The synthetic sequences were designed to delete the two non-functional *hsdS* genes (SPD_0450 and SPD_0451) and the creX recombinase (SPD_0452) (13) to prevent any further rearrangement leading to change of the six variants of the enzyme. Primers used to generate such mutants in the DP1004 strain were the same used for D39 and are published elsewhere (45). Primers used to build the PCR products for the deletion of the whole *spnIII* (from SPD_0449 to SPD_0455) and SpnMcrBC systems (SPD_1108-9) are listed in table S1. All mutants were confirmed by Sanger sequencing (Eurofins Genomics, Germany). SpnIV mutants were constructed by Kwun et al (37).

**Sample collection**

Oral swab samples were collected from healthy adult volunteers at the University of Leicester and resuspended in 5 ml of SM buffer (10 mM MgSO₄, 100 mM NaCl, and 50 mM Tris–HCl pH 7.5) and stored at 4°C, protected from light (23, 24). A portion of each of the samples was stored at -80°C with 10% glycerol. Sample collection and storage conditions...
were approved by the Departmental Research Ethics Office of the University of Leicester (authorisation mro5-5d40 21/07/2014). All the experiments were done in accordance with national and institutional guidelines.

**S. pneumoniae bacteriophage isolation method**

The two *spnIII* deleted mutants FP486 and FP470 were used as hosts for propagation. Oral samples were added at 1:100 to mid-exponential phase growing cultures. Overnight enrichments were centrifuged, filtered (0.22 μm membrane) and inoculated into a fresh bacterial culture for three consecutive days. Each day the supernatants were spot assayed on a CAT-Galactose soft medium lawn plate to check for plaques confirming the presence of bacteriophages. Double-Layer Agar (D-LA) assay was then performed using any positive samples. Collection of a single plaque into SM buffer and propagation was repeated several times in order to isolate a single clone of the phage. In order to identify the phage's natural host, the oral swab sample was plated on a TSA-blood plate and single colonies were isolated and propagated. Subsequently each strain underwent PCR screening for the presence of the phage using primers LF_83 and LF_84 (Table S2).

**Lysogen isolation**

Lysogens of FP470 were generated by plating 1x10⁴ CFU on a CAT-Galactose double-layer agar plate with the overlay containing 10⁶ PFU/ml of the bacteriophage. The bacterial clones grown on the plate were propagated for several rounds on TSA. Presence of lysogenic phage and analyses of insertion sites were evaluated by PCR using the primers listed in Table S1.

**Electron microscopy**

A sterile high-titre phage sample (1x10⁹ PFU/ml) was purified by serial centrifugation at 20,000 x g for 60 minutes and resuspended in ammonium acetate solution. The suspension was then adsorbed onto a hydrophilic (freshly glow discharged) carbon-coated Pioloform film coated copper grid (Agar Scientific) and negatively stained with 1% uranyl acetate. Sample visualization was performed on a JEOL 1400 transmission electron microscope (TEM) with an accelerating voltage of 80kV and images were captured using a Mageview III digital camera with iTEM software (Olympus).

**Identification of phage structural proteins**

Phage was purified as described in the electron microscopy methodology. In-gel trypsin digestion of the purified phage SpSL1 followed by liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS; LTQ-Orbitrap-Velos-ETD Mass Spectrometer) was
performed at the local proteomics facility (PNACL, University of Leicester, UK). The resulting peptide sequences were searched (MS/MS Ion Search, Mascot v2.2.04 algorithm, Matrix Science) against UniprotKB-SwissProt and NCBI protein databases.

Phage methodologies
Adsorption was measured by PFU titration from the supernatant at various time points following inoculation of the phage into a FP470 growing culture (OD$_{600nm}$ of 0.1) (MOI of 0.1). To evaluate the restriction and methylation activities of the SpnDP1004III system, D-LA assays were performed using SpnDP1004III-unmethylated phage, derived from a spnDP1004III deleted mutant, at several dilutions to infect all six spnDP1004III-A-F strains. The same experiment was also performed in the presence of 125 ng/ml of CSP (competence stimulating peptide). Plaques were counted, collected into SM buffer and phage propagated in the same mutant used for the D-LA assay in order to obtain $10^9$ PFU/ml titres needed for further experiments. An SpnIII-A-methylated SpSL1 phage was then used to infect each of the spnDP1004III-A-F mutants and the spnDP1004III deleted strain as a control. Prior to phage infection bacterial culture was sampled and analysed with an allele scan protocol (45).

Sequencing and bioinformatics analysis
Bacteriophages were concentrated by centrifugation for 3 hours at 4°C in a TH-641 rotor (Thermo Scientific) at 164,000 x g. Phage DNA purification was performed using the phenol-chloroform extraction technique. Sequencing of the phage genome was performed by GenProbio (Parma, Italy) using a Personal Genome Machine Sequencer (IonTorrent) and by the Norwegian Sequencing Centre (Oslo, Norway) with a MiSeq platform (Illumina). Reads were assembled using both MIRA (ver. 3.9.18) and Velvet (1.2.10) and results combined. Genome edges encompassing the phage cos site were confirmed by Sanger sequencing using primers LF83 and LF84 (Table S1). The complete sequence was annotated with RAST, manually refined using BLASTp (NCBI) and Pfam (EMBL-EBI) and deposited in GenBank with accession number KM882824.

Methylome analysis
To evaluate the activity of the m5C methyltransferase (MTase) carried by SpSL1, DNA was treated with sodium bisulfite, which converts unmethylated cytosine (C) to thymine (T), before Illumina sequencing by Norwegian Sequencing Centre (Norway, Oslo). Reads obtained were aligned against two versions of the phage sequence in which Cs were replaced with Ts and guanines (G) with adenines (A) respectively. Polymorphic changes between Ts and Cs or As and Gs, with a frequency over 50%, were then retrieved using the
Mosaic Aligner suite (The MarthLab, USA) and scored as methylated. Comparison of sequences adjacent to the methylated bases allowed us to obtain the m5C MTase recognition pattern.

**Gene expression analysis**

For gene expression analysis pneumococcal strains were grown in CAT-Galactose to mid-log phase (OD_{590} approximately 0.15) and infected with SpSL1 phage (MOI of 0.2). After 10, 50, and 90 minute time points, 10 ml of cells were added to 2 ml of an ice-cold 95% ethanol 5% phenol solution, before centrifugation at 4000 rpm for 10 minutes. Supernatant was removed and the pellets stored at -80°C until processing. Excluding the infection and time course sampling, the same procedure was followed for the strain with an integrated SpSL1. A non-infected sample was included in the analysis and three independent replicas were collected for each condition. For RNA extraction pellets were resuspended in 50 μl TE (pH8) with 3 mg/ml lysozyme and incubated at 37°C for 20 mins to lyse cells. The Maxwell 16 LEV simplyRNA Cells kit (Promega) was then used along with a Maxwell 16 LEV instrument (Promega) for RNA extraction. RNA samples were processed with the ScriptSeq Complete Kit for bacteria (CamBio), which includes an rRNA depletion step, and sequenced with a MiSeq system (Illumina) at the University of Leicester (UK). RNA-seq fastq files were gently trimmed using Trimmomatic (ver. 0.30). Read mapping to *S. pneumoniae* D39 genome (NC_008533) and to SpSL1 phage (KM882824), transcript abundance quantification and upper-quartile of gene expression normalization, and differential expression analysis were carried out using Rockhopper software (ver. 2.0.2). Differential analysis of the *S. pneumoniae* D39 genome was carried out against a non-infected control. Differentially expressed genes were further filtered by discarding those with a non-statistically significant q-value (>0.01) and values below 2 Log2 fold increase or decrease in transcription (74).

**Allele quantification**

Quantification of the six possible *hsdS* variants in wt strains was performed as previously described (45). In brief, a region common to all of the possible *hsdS* conformations was PCR amplified with one of the two primers being fluorescently tagged. A double digestion of the PCR products, using restriction enzymes Dral and Pfel (New England Biolabs) allowed for the generation of labelled DNA fragments of different lengths, one specifically for each of the *hsdS* variant forms. GeneScan analysis on an ABI prism Gene Analyser was then performed for relative quantification of the six allelic forms.
Results

Isolation of a temperate pneumococcal bacteriophage.

We have previously characterised a phase variable pneumococcal type I restriction system (45). To investigate its impact on bacteriophage control, we used the spnIII deletion mutant as a potential permissive bacteriophage host. Screening of 24 oral swab samples of healthy adult volunteers, using encapsulated and unencapsulated spnIII-deleted mutants as hosts, gave a single positive spot assay result, after three days of enrichment, in the unencapsulated strain only. The phage, named SpSL1, was able to form clear plaques in a D-LA assay (Fig. S1A). A single clone was isolated and propagated to 4x10^8 PFU/ml. To further characterise SpSL1, the phage morphology was determined using transmission electron microscopy (Fig. S1B). The presence of isometric capsid (~50 nm), long non-contractile tail (~160 nm) ending in a single tail fibre (~110 nm) revealed SpSL1 to belong to the *Siphoviridae* family. Phage adsorption to *S. pneumoniae* was found to be extremely rapid. Indeed, the number of free phages in solution decreased by 98% in fewer than 60 seconds when in contact with pneumococcal cells (Fig. S1C). The adsorption rate constant during the first minute of incubation was of 3.9x10^{-6} ml/min.

Phage SpSL1 sequence and proteome.

Sequencing of SpSL1 (GenBank accession code KM882824) revealed a linear genome of 33,756 bp with a GC content of 38.6%. An 11-base single stranded cohesive end (5’-CGGTGTCAATC-3’) required for genome re-circularisation was found at the genome ends. The 50 predicted coding sequences (CDSs) are organized in five operons with packaging, morphogenesis, lysis, lysogeny, and replication functions (Fig. 1A). Genes are all transcribed from the same strand with the exception of those belonging to the lysogeny cluster and *cds50* encoding for an unknown hypothetical protein (Fig. 1A). Packaging and morphology modules are well conserved with respect to other pneumophages including the tail fibre gene (*cds44*), encoding a 1602 aa PblB-like protein. The PblB-like protein is predicted to be the tail fibre of SpSL1 and shows similarity (60-80% amino acid identity) to bacterial platelet-binding proteins and other streptococcal bacteriophage PblB-like proteins. As in other streptococcal phages previously described in a world-wide panel of 482 pneumococcal genomes (7) the lysis gene cluster carrying two holins and an endolysin were present downstream of the packaging genes (Fig. 1A). The lysogeny module, located downstream of the *attP* site, comprises a 382 aa integrase belonging to the Int family (tyrosine recombinases) with an amino acid identity of about 95% compared to other sequenced streptococcal temperate bacteriophages (Fig. 1A). SpSL1 shows a high degree of similarity
to the B2 Cluster of pneumophages previously characterised (7). A transcription regulator cl
and a prophage antirepressor protein are also present on this module (Fig. 1A). Of particular
interest is the position of a second transcriptional regulator, cro (cds25), which is inside the
replication gene cluster (Fig. 1A). The genes in this module mainly encode for proteins
involved in the regulation of phage replication within the host cell. In comparison to the
replication modules found in other referenced pneumophages (7), the replication module in
SpSL1 showed a large degree of rearrangements. In addition, the replication gene cluster
also encodes for two MTases. Indeed, through sequence analysis, cds15 and cds16 are
predicted to encode for a C5-cytosine and N6-adenine MTases. The methylation pattern of
the m5C MTase was identified to be R\textsuperscript{5mCGRC}, but we could not identify any adenine
methylation. Bisulfite methylome analysis showed that 102 out of 113 RCGRC sites were
methylated in the SpSL1 genome. The remaining nine non-methylated sites are reported
(Table 1) and intriguingly these include a site in an inverted repeat in overlapping a possible
promoter for the m5C MTase gene itself.
Proteomic analysis by LC-MS/MS of the SpSL1 particles revealed the presence of 8 out of
the 17 predicted phage structural proteins (Table S2). Except for the hypothetical protein
Cds21, none of the other non-structural proteins yielded significant counts in the MS analysis
(Table S2).

Phage Integration Attachment Site
The phage attachment sequence was found to be 5'-CTTTTCATAATAATCTCCCT-3'. This
sequence, in the S. pneumoniae reference D39 genome, maps to the conserved stretch of
four non-coding small RNAs called csRNA3, 2, 4 and 5 (Fig. S2A). Lysogenic SpSL1 in the
spnDP1004lll deleted mutant was identified to be integrated in both csRNA3 and 2 (Fig.
S2C, D). Despite this, the lysogens were not stable as confirmed by PCR amplification of the
phage genome in its lytic form and intact csRNAs (Fig. S2C, D). When assaying single
colony isolates of the lysogenic clones for integrated phage, SpSL1 was always found
integrated into csRNA3, and in half of the colonies was also integrated into csRNA2 (data
not shown). There was no evidence for specialised transduction.

Gene expression profiling of SpSL1.
Normalised SpSL1 gene expression profiling allowed the identification of five operons (Fig.
2, 3, Table S4). Upper-quartile of gene expression normalisation after was for read
normalisation. Upon infection the transcription of the early replication module, including DNA
methyltransferases and a recombinase (operon 2, cds5 to cds27), was activated
immediately but had then decreased in expression after 50 minutes (Fig. 2, 3). Cds50
(operon 5), a hypothetical protein, is actively expressed throughout the infection but the peak
expression is seen at 10 minutes post infection (Fig. 2, 3). The late replication module
(operon 4, \textit{cds}30-49) containing genes for the virion packaging and structural proteins of the
SpSL1, including the tail fibre protein PblB, showed high levels of expression at 50 and 90
minutes post-infection. Similarly, the operon (operon 3) encoding for a hypothetical protein
(\textit{cds}28) and an endonuclease (\textit{cds}29) increased at the 50 and 90 minutes time points and
are predicted to be the final proteins to be transcribed (Fig. 2, 3). The expression profile of
the lysogenic phage was also evaluated. In this case operon 1, containing the lytic cycle
repressor \textit{cl}, and the integrase, as well as other genes involved in lysogeny maintenance
(\textit{cds}1-4), showed the highest expression. All other phage genes were found to be actively
expressed, albeit at a low level, however it is important to note that there was the presence
of antisense RNA transcription predicted throughout the SpSL1 genome (Fig. 2, 3, S5, Table
S4). The high level of expression of the \textit{cl} repressor allowed us to identify a 5'-UTR of 140
bp. In order to evaluate the transcriptional response of the bacterial host cell to SpSL1
phage lytic infection, RNA-seq data were also mapped onto the pneumococcal genome (Fig.
4; Table S3). Over the 90 minutes time course 164 genes of the \textit{spn}DP1004\textsc{ill} deleted
mutant showed levels of expression significantly altered with respect to those of non-infected
cells (Table 2). The operons responsible for ribonucleoside triphosphate
biosynthesis(SPD_0187-0191, SPD_1041-1043, and SPD_1594) were found to be already
up-regulated at 10 minutes post-infection and were by far the most highly up-regulated host
genes at 50 and 90 minutes. Of relevance, among the genes down-regulated during lytic
infection we found the pneumococcal virulence factor LytB (SPD_0853) and the sodA
manganese-dependant superoxide dismutase (SPD_0667) (Table 2; S3). In the strain with
the integrated prophage the main down-regulated gene clusters were those belonging to the
biosynthesis of thiamine (SPD_0622-4) and pyridine (SPD_0851-3) pathways, and to an
anion ABC importer (SPD_2024-7).

\textbf{Phage restriction by the phase variable RM systems in \textit{S. pneumoniae}.}

SpSL1 was assayed for its interaction with the phase-variable SpnIII and the SpnIV RM
system and its role in bacteriophage restriction. The number of target sites within the
bacteriophage genome is dependent upon the variant of the SpnIII or SpnIV system present
in the host cell that the bacteriophage is infecting, with a range of 4 to 16 sites in the different
SpnIII alleles and 15 and 21 within the SpnIV strains tested here. Infecting alternatively
locked unencapsulated clones clearly showed different levels of restriction (Fig. 5). The
divergence of plaque numbers of \textit{spn}III\textit{DP}1004\textsc{a-d} locked mutant strains with respect to the
\textit{spn}DP1004\textsc{ill} deleted mutant was statistically significant ($P<0.001$). The efficiency of
restriction was approximately proportional to the number of methylation sites for each phase
variable variant present in the SpSL1 genome (Fig. 5A). SpnDP1004\textsc{ill}- unmethylated
phages that were successful in infecting the A-D locked strains were collected and assayed for the presence of methylation in the SpnIII-A-D recognition sites through a subsequent D-LA assay. As shown with SpnIII-A-methylated SpSL1, effective methylation of the phage was demonstrated by abolishing restriction when re-infecting the spnDP1004III-A strain and restriction still occurred when infecting the spnDP1004III-B and spnDP1004III-C strains (Fig. 5B, C). To evaluate the impact of a functional SpnIII RM system in a bacteriophage isolation screening protocol, a DP1004 strain with known spnDP1004III allele composition (85% A, 12% B, 1.6% C, 0.75% D, 0.65% E) was challenged with SpSL1 phage methylated at A or B sites (Fig. 6A, B). The number of plaques was greatly reduced (about 40-fold reduction, P<0.001) when using a SpnIII-B-methylated phage (Fig. 6B), while no reduction was observed with the SpnIII-A-methylated SpSL1 (Fig. 6A). Similarly, we tested mutants in the second pneumococcal phase variable type I RM system SpnIV (37). The SpnIV knock out strain, R6x Δ’vr tvr::Janus strain, which lacks both the SpnIII and SpnIV R-M systems (Table S1), was used as a control for the phase variant R6x Δ’vr hsdS::tvrRMV5ΔtvrR (Table S1), which expresses a functional SpnIV R-M system transferred from strain RMV5 (37). The restriction activity of this strain showed a 10,000-fold reduction in phage activity compared with the control (p<0.001) (Fig. 5C).

Abortive infection mechanism.

A mid-exponential phase wt strain was infected with a SpnIII-A-methylated phage (MOI=0.25) and the variation of the hsdS allele conformation in the bacterial population was evaluated at each hour for the next four hours (Fig. S3). Despite a reduction in viable cells, no significant change in SpnIII allele frequency was identified (Fig. S3). In a similar manner, SpnIII locked mutants were infected with SpnIII-A-methylated and SpnIII-B-methylated phage (Fig. 7A-B). CFU counts, after 55 minutes post-infection, showed a similar reduction of viable cells irrespective of the infecting phage’s methylation pattern (Fig. 7A, C). The involvement of the SpnIII system in the Abi phenotype was evident in that phage recognised as “self” (i.e. SpnIII-A-methylated phage in an SpnIII-A locked mutant, Fig. 7) produced lysis of the host just before two hours (coinciding with virion release), whereas phage recognised as “non-self” (i.e. a different methylation pattern that should be restricted) induced a progressive cell death which started immediately after the first minutes of infection (Fig. 7). The spnDP1004III deletion mutant lacked the Abi phenotype, showing that SpnIII itself is responsible for the cell death (Fig. 7). Mutants deleted for just the SpnIII restriction subunit (spnDP1004ΔhsdR) were also unable to undergo abortive infection. Inhibitory concentrations of chloramphenicol were found not to block the abortive infection, indicative of the need for de novo biosynthesis of proteins (Fig. S4). Mutants of LytA and McrBC still
undergo Abi, providing further evidence that SpnIII is specifically responsible for Abi in *S. pneumoniae* (Fig. 8).

### Discussion

To improve the methodology for isolation of pneumococcal bacteriophages (47, 64, 72) we utilised a knock out mutant of the phase variable Type I R-M system SpnD39III recently described by us and others (45). As D39 and its derivatives are naturally devoid of the other phase variable type I R-M systems (SpnIV, encoded by the *tvr* locus) (37), our new non-encapsulated recipient strain did not contain any of the phase variable R-M systems; this likely contributed to our success in isolating the temperate bacteriophage SpSL1 from a panel of oral swab samples. This siphovirus was then investigated and used to study the underlying bacteriophage-host interactions, involving the phase variable R-M system and abortive infection.

Whole genome analysis showed a functional cluster organization of the SpSL1 genome (Fig. 1A) similar to that of other streptococcal prophages (7, 8, 14, 53, 62, 78). An interesting exception is the *cro* transcriptional regulator that was found within the replication cluster which is far from the lysogeny module. This finding reinforces previous evidence supporting the hypothesis that phage evolutionary exchange can take place at the level of a single gene (15, 52, 62) and contrasts with the theory of modular phage evolution (5). The absence of conserved genes of unknown function, *cg1* and *cg2*, is also noteworthy, as these genes have been described to be present in all previously characterized temperate pneumophages (7, 62). Based upon distribution of genes, integrase sequence homology and *attP* recognition sequence, SpSL1 could be included either in phage group 1 according to *Romero* and colleague’s classification (62) or Cluster B2 as described by *Brueggemann* and colleagues (7). Many regions of the genome including the tail fibre gene, the lytic cluster, and part of the lysogeny and replication modules showed rearrangements when compared with other streptococcal bacteriophages. Due to the differences found between SpSL1 and other pneumococcal prophages the primers described by Romero and colleagues (63), for the identification of temperate *S. pneumoniae* phages and used in a recent study (43), would not have be able to detect SpSL1.

The attachment site sequence is identical to that designated as *att*\_oxc by Romero and colleagues (62). It is present in multiple sites in the *S. pneumoniae* genome (Fig. S2A) and was found to be a conserved sequence belonging to four out of the five cia-dependent small RNAs (csRNAs) (22). These non-coding sRNAs are highly similar to each other showing a predicted secondary structure with inverted repeats at both ends. csRNAs are present in
many streptococcal genomes suggesting a fundamental role in this group of organisms and this likely explains the reason for them being selected as target for phage integration (46). The expression of the csRNAs is regulated by the two-component regulatory system CiaRH (22). In *S. pneumoniae* csRNAs have been shown to modulate stationary phase autolysis (22), to affect virulence during lung infection (44), to be involved in β-lactam resistance (68), and to negatively regulate natural competence development (68, 75). However, despite the identification of some targets of the csRNAs, the molecular mechanism(s) underlying the phenotypes observed is still unknown (6). Interestingly when analysing the SpSL1 attP downstream sequence it was observed to exhibit high nucleotide identity with csRNA2 and its right flanking region over a region of 236 bp (Fig. S2B). Therefore in the case of phage integration at csRNA2 it would be predicted that csRNA2 would remain unaltered. Only two nucleotide changes were found between the csRNA2 and SpSL1 sequences (Fig. S2B); these are located on the loop of the predicted terminator and at the end of the small RNA sequence suggesting an absence of significant secondary structure alterations after phage integration (Fig. S2B). Type 1 temperate phages identified previously in the genome of sequenced *S. pneumoniae* isolates have always been found to be integrated at the csRNA3 site alone (8, 62). The first 30 nucleotides after attOxc in type 1 pneumophages are well conserved, with $\phi$Spn_H_1 being a good example (Fig. S2B). After phage integration the final portion of csRNA3, corresponding to the terminator, is replaced with the end of csRNA2 producing a new chimeric sequence that maintains the secondary structure characteristic of csRNAs and that is therefore likely to be functional. Here we report the first evidence of integration of a temperate phage at the csRNA2 site, however this was never exclusive (Fig. S2 C, D), and integration in csRNA2 was always found to be associated with another phage also integrated at csRNA3, whilst the reverse state was not obligatory. These observations, together with the observation of intact csRNA sites, suggested an active process of phage excision and integration occurring within the same cell. The instability of prophages in the *spnDP1004III* deleted mutant under the growth conditions used, was confirmed by the presence of free phage particles in the culture medium and also by gene expression profiling of the lysogenic strain where genes encoding both lytic- and lysogenic-related proteins were found to be highly expressed (Fig. 2).

Lytic phage infection showed a significant impact on global host transcriptome with the majority of changes occurring transiently in the early stage of infection. The observed variations were typical of a metabolic stress-related response. Unlike previously proposed for the PRD1 phage infecting *E. coli* (60), the early up-regulation of amino acid uptake pathways were not significantly upregulated in *S. pneumoniae*, however the nucleoside synthesis operons were, (Fig. 4A; Table S3), potentially allowing for the greater availability of
nucleosides for bacteriophage genome replication. Conversely the transcriptional down-
regulation of a few other genes (namely the manganese export, ABC transporter,
peptidoglycan biosynthesis operon genes, and LytB) (Table S3) could represent the
inhibition of unnecessary energy-wasting synthetic pathways in order to concentrate the host
biosynthetic machinery exclusively on phage replication. Reduced expression of ABC
transporters and peptidoglycan biosynthesis operons was also observed in *L. lactis* and *P.
aeruginosa* after infection with c2 and PRR1 phage respectively (17, 61). It is noteworthy
that the previous temporal analyses of bacterial gene expression after phage infection
showed the majority of changes to occur at the late stage of phage replication (1, 17, 54, 55,
60). The lysogenic phage was found to be unstable in the condition assayed, and showed
expression of genes associated with both lytic and lysogenic cycles, which most likely reflect
expression data from a mixed population (Fig. 2, 3). It is therefore difficult to distinguish
between the effects on host transcription of the lysogenic phage alone. Of relevance is the
down-regulation of the pyridine biosynthesis operon (Table S3) of which expression was
found to be reduced also in *L. lactis* during mid-late infection of Tuc2009 phage (1). The
gene expression of SpSL1 during lytic infection shows a clear operon structure with early
and late operons as shown in most other lysogenic bacteriophages (Fig. 2, 3). Interestingly
the *cds27-29* genes (operon 3) predicted to be involved with cell lysis show independent
transcriptional regulation to the other traditional late genes suggesting an alternative or
additional regulation (Fig. 2, 3). *Cds50* (operon 5), encoding for a hypothetical protein has
shown independent transcriptional regulation (Fig. 2, 3) and is less expressed than the
hydrolases downstream or the integrase upstream in the lysogen.

Restriction of SpSL1 infection, as previously show with other Type I R-M systems (4), was
confirmed in the four *spnDP1004III-A-D* mutants (Fig. 5A, B) and by testing of the SpnIV
system (Fig. 5C). The efficiency of plating was reduced according to the number of sites
recognised by each *hsdS* conformation in the phage genome (Fig. 5A, B), as was previously
reported for other R-M systems (37, 79). Phages harvested from *spnDP1004III-A* mutants
were found to be methylated and therefore protected in subsequent infections of the same
host strain, yet were still restricted when infecting other locked *spnIII* strains with differing
*hsdS* target sites (Fig., 5B). As previously hypothesised, the ability of the SpnIII system to
switch between 6 active *hsdS* subunits allows the bacteria to increase their defensive
repertoire by recognising several sequence specificities without acquiring new R-M systems
(Fig. 1C) (41). In addition, by use of a wild-type *S. pneumoniae* strain, expressing multiple
forms of the SpnIII enzyme, it was shown that this could lead to the underestimation of
actual phage titre in a D-LA assay. Indeed, clear plaques could only be detected when the
phage methylation matched the R-M system of the prevalent subpopulation of a
heterogeneous wt strain. In contrast, when the phage methylation matched a less prevalent 
R-M system in the wt acceptor strain, fewer plaques than expected were detected (Fig. 6B).
In our example infection of a wt strain (85% SpnIIIA, 12% SpnIIIB) with an SpnIIIB-
methylated phage yielded only plaques with SpnIIIA-methylated phage indicating that all 
these new phages were produced by breaking resistance in spnDP1004III A cells. The lack 
of SpnIIIB-methylated phage plaques is hypothesised to derive from the fact that even if a 
“rare” SpnIIIB cell is infected, the phage progeny cannot spread to the surrounding majority 
of SpnIIIA cells in the soft agar (Fig. 1C). This blocks the generation of a clear plaque even 
in the presence of an initial infection (Fig. 6B). These observations are of relevance 
considering those previous pneumophage isolations where spnIII wt pneumococcal strains 
were used (47, 64, 72). Our data showed the successful isolation of a phage using a wt 
SpnIII bacterial population is influenced by i) free phage titre in the samples, ii) methylation 
state of phage DNA in SpnIII sites and, iii) SpnIII allele composition within the acceptor 
strain. Of course, the last two points could be bypassed using a spnIII deletion mutant as 
shown here.

The Type IV MrBC R-M system, which recognises and cleaves between two R’M°C patterns, 
was previously found to be unable to efficiently restrict a methylated phage genome (19, 32); 
this was demonstrated by MrBC digestion of chromosomal DNA following infection by a 
lambda phage carrying a cloned methylase (19). Our data shows that Abortive infection (Abi) 
occurs in those bacteria that are expected to be able to restrict phage, however our data 
also demonstrates that this was unaffected by the removal of the MrBC system (Fig. 8). In 
contrast the SpnIII knock-out strain, as well as single mutants of the hsdR gene, is unable to 
demonstrate the Abi self-killing phenotype. Unlike previous reports (19), our research shows 
that it is the SpnIII system, rather than the MrBC system, that is the main determinant of Abi 
in S. pneumoniae. We also show the dependency of the SpnIII hsdR restriction enzyme for 
Abi in our system. Reducing the bacterial growth rate by using the bacteriostatic antibiotic 
chloramphenicol reduced Abi, showing that the rate of Abi is also influenced by bacterial 
replication (Fig. S4). Although phase-variable restriction of foreign DNA introduced by 
transformation has previously been shown for both the SpnIII (45) and the SpnIV systems 
(37) this is, to our knowledge, the first phase-variable type I RM system that can induce Abi 
and restrict invading bacteriophages in a phase-dependant manner, thereby making the 
SpnIII system the key population-based armour for S. pneumoniae in the co-evolution war 
against their natural predators.
Acknowledgments
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). This research used the ALICE High Performance Computing and the Advanced Imaging Facility (AIF) at the University of Leicester. We gratefully acknowledge volunteers who contributed saliva samples, and the NUCLEUS Genomic Services at the University of Leicester. High throughput sequencing was performed by the Norwegian Sequencing Centre (www.sequencing.uio.no), a national technology platform hosted by Oslo University Hospital and the University of Oslo and supported by the "Functional Genomics" and "Infrastructure" programs of the Research Council of Norway and the Southeastern Regional Health Authorities.

Supplementary data
Supplementary data are available through the University of Leicester data repository – Figshare (https://leicester.figshare.com/) with the doi:10.25392/leicester.data.8320871.
Figure Legends

Figure 1. The SpSL1 bacteriophage and SpnIII restriction system. In Panel A, the SpSL1 genome displayed in the viral conformation with the cohesive ends flanking the sequence. The colour scheme indicates the operon structure with operon 1 (cds1-4) in orange, operon 2 (cds5-26) in blue, operon 3 (cds27-29) in green, operon 4 (cds30-49) in red, and operon 5 (cds50) in white (GenBank accession code KM882824). Panel B shows the phase-variable type I restriction modification system spnIII, containing the hasDR, hasDM, and the variable hasS genes in addition to the site-specific recombinase gene creX with a schematic representation of the recombination. The hasS gene encodes for and N- and a C-terminal target recognition domain (TRD). The two N-terminal TRDs are in dark blue and light blue, whilst the C-terminal TRDs are in red, orange, and purple. The inverted repeats are shown as grey dotted rectangles. The six different hasS variants are shown below from A to F (45). Panel C represents a cartoon for a population-based bacteriophage defence that arises from a phase-variable restriction modification system, where the bacterial genome is methylated in a specific pattern (shown by the coloured M on the black line). The bacteriophage would only be able to infect and replicate in one variant (blue), whilst unable to infect the other variants with different methylation patterns (orange, red, purple, and green) (12).

Figure 2. SpSL1 phage relative gene expression in lytic and lysogenic stages. A time course of lytic infection of an spnIII deleted strain with a MOI of 0.2 at 10 minutes (Panel A Green), 50 minutes (Panel B Blue), and 90 minutes (Panel C Red) after challenge. Panel D (black) demonstrates the gene expression in the lysogen. All data are shown in terms of normalised read coverage. RNA-Sequencing reads were mapped to the viral conformation of the SpSL1 phage deposited in Genbank (Accession: KM882824) (Panel E), even in the case of the lysogen (Panel D). RNA-Seq Mapping and upper quartile normalisation was performed using Rockhopper. Data are visualised on BAMviewer in Artemis, with the maximum number of 2000 reads (small label on the right of each panel).

Figure 3. Bacteriophage SpSL1 gene expression during the lytic cycle. RNA-Seq data showing the expression of phage SpSL1 at 10, 50, and 90 minutes post-infection of the spnIII negative strain FP470. The transcriptional units are numbered 1 to 5 as previously shown (Fig. 1A). The three early transcriptional units are operon 1 (cds1-4) in orange,
Figure 4. *S. pneumoniae* genes are upregulated in response to SpSL1 infection. RNAseq analysis during SpSL1 infection revealed the upregulation of three *S. pneumoniae* transcripts preceded by an NrdR binding site and encoding for the anaerobic ribonucleoside-triphosphate reductase operon in circles (SPD_0187 to SPD_0191; five gene operon), the ribonucleoside-diphosphate reductase operon in squares (SPD_1041 to SPD_1043; three gene operon), and a hypothetical operon encoding for an unknown transcriptional regulator and a conserved hypothetical protein in triangles (SPD_1594 and SPD_1595; two gene operon). The RNA-Seq data was normalised by upper-quartile gene normalisation and compared with a non-infected control to determine fold change.

Figure 5. Restriction of SpSL1 by phase-variable SpnIII and SpnIV RM systems. Panel A shows plaque assays results using SpnDP1004III-unmethylated SpSL1 phage to infect *spnDP1004A-D* locked strains that express single locked copies of one of the *hsdS* alleles with FP470 used as a control. Panel B displays the differences observed between the control strain deleted for *spnDP1004III* (zero sites recognised) and the other mutants are statistically significant (one-way ANOVA multiple comparisons test, P<0.001) for SpSL1. Panel C shows the restriction of infection of the *spnDP1004III* deleted strain and the *spnDP1004III* locked mutant with SpnDP1004III-A methylated SpSL1 is not statistically significant (one-way ANOVA multiple comparisons test, P>0.05) whereas the infection with *spnDP1004III* and *spnDP1004III* is (one-way ANOVA multiple comparisons test, P<0.001) (panel B). Plaque Assay results using SpSL1 to test the phase-variable SpnIV system show differences between the SpnIV knock out strain and the SpnIV R6x Δirr *hsdS::tvrRΔtvR* recombinant (37) (students t-test, P<0.001).

Figure 6. Phase Variable restriction of SpSL1 by SpnIII in a wild-type population. Plaque assays of a wt strain and *spnDP1004III* locked mutants. Infection of the wt DP1004 strain, harbouring 85% and 12% of *spnDP1004III* and *spnDP1004III* positive cells respectively, were tested for infection by SpnIII-A-methylated phage (panel A) and SpnIII-B-methylated phage (panel B) and compared respectively to infection of *spnDP1004III* locked and *spnDP1004III* locked strains. Panel A shows equally efficient SpnIII-A-methylated phage infection of a wt host with a predominance of SpnIII-A-cells and of a *spnDP1004III* locked mutant. Panel B does show plaque generation of SpnIII-B-methylated phage.
Unrestricted plaque formation in spnDP1004IIIB cells yielded SpnIIIB-methylated phage (data of phage methylation status not shown), while infection of the wt containing 85% of spnDP1004IIIA cells (grey bar in B) yielded less plaques (two tailed t-test, P<0.001) and all phage obtained was SpnIIIA-methylated phage (data not shown).

Figure 7. Abortive Infection by the SpnIII system. Bacterial cell fate and viability are determined by the hsdS allele conformation and by the infecting phage genome methylation status. Panel A shows the spnDP1004IIIA locked mutant infected with SpnIIIA-methylated (blue; MOI=2.5), and SpnIIIA-nonmethylated (orange; MOI=2.5) SpSL1 phages. In panel B the spnDP1004IIIB locked mutant is infected with SpnIIIB-methylated (orange; MOI=2.5), and SpnIIIB-nonmethylated (blue; MOI=2.5) SpSL1 phages. In both the cases the non-restricted phage killed the cells after completion of the lytic cycle, whilst the supposedly restricted phage induced a rapid and progressive lysis. When infecting an spnDP1004IIIl deletion mutant (panel C) the phage underwent a lytic cycle irrespective of its methylation status (SpnIIIA-methylated, blue; SpnIIIB-methylated, orange; MOI=2.5). The same outcome was achieved by inactivating the restriction subunit of spnDP1004III alone (panel D). Uninfected bacterial strains in panels A-E are depicted in green. The one-step growth curves (infecting free viral particles were measured each 30 minutes after infection) in panel E confirmed the production of phage progeny when SpSL1 is not restricted (as with the SpnIIIA-methylated phage infecting the spnDP1004IIIA locked mutant, blue lines; uninfected control is in light green) or the absence of phage replication when SpSL1 is restricted by SpnDP1004III (as with the SpnIIIA-methylated phage infecting the spnDP1004IIIB locked mutant, orange lines; uninfected control is in dark green). SpSL1 burst size is of 20 PFU.

Figure 8. McrBC and LytA are not responsible for abortive infection. SpnMcrBC Type IV R-M system and the autolysin LytA have no effect on the Abi phenotype. Mutants for mcrBC (panel A) and lytA (panel B) were constructed in a spnDP1004IIIA locked background and the pairs of recombinant strains were infected with SpnIIIA-methylated (blue; MOI=2.5), and SpnIIIA-unmethylated (orange; MOI=2.5) SpSL1 phage. Uninfected controls are shown in green. Both mcrBC and lytA mutants showed near immediate lysis upon SpnIIIA-unmethylated phage infection indicative of an unmodified Abi phenotype.
Tables

Table 1. Nonmethylated cytosines in R<sup>m5C</sup>CRGC pattern of the SpSL1 genome

<table>
<thead>
<tr>
<th>Position</th>
<th>Strand</th>
<th>Gene</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>214</td>
<td></td>
<td>&lt;i&gt;cds&lt;/i&gt;30</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>3605</td>
<td>c&lt;sup&gt;a&lt;/sup&gt;</td>
<td>&lt;i&gt;cds&lt;/i&gt;33</td>
<td>prohead maturation protease</td>
</tr>
<tr>
<td>3661</td>
<td>c</td>
<td>&lt;i&gt;cds&lt;/i&gt;33</td>
<td>prohead maturation protease</td>
</tr>
<tr>
<td>18433</td>
<td></td>
<td>&lt;i&gt;cds&lt;/i&gt;49</td>
<td>LytA-like murein hydrolase</td>
</tr>
<tr>
<td>18804</td>
<td>c</td>
<td>&lt;i&gt;cds&lt;/i&gt;49</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>19917</td>
<td>c</td>
<td>&lt;i&gt;cds&lt;/i&gt;01</td>
<td>integrase</td>
</tr>
<tr>
<td>19947</td>
<td>c</td>
<td>&lt;i&gt;cds&lt;/i&gt;01</td>
<td>integrase</td>
</tr>
<tr>
<td>24773</td>
<td>c</td>
<td>&lt;i&gt;cds&lt;/i&gt;11</td>
<td>DNA replication initiation protein (DnaD)</td>
</tr>
<tr>
<td>26775</td>
<td>c</td>
<td>&lt;i&gt;intergenic&lt;/i&gt;</td>
<td>hairpin in front of the m5C MTase (&lt;i&gt;cds&lt;/i&gt;15)</td>
</tr>
</tbody>
</table>

<sup>a</sup>c, complementary strand
References


34. **Koskella B, Brockhurst MA.** 2014. Bacterial-phage coevolution as a driver of ecological and evolutionary processes in microbial communities. FEMS microbiology reviews **38**:916-931.


Mruk I, Kobayashi I. 2014. To be or not to be: regulation of restriction-modification systems and other toxin-antitoxin systems. Nucleic acids research 42:70-86.


