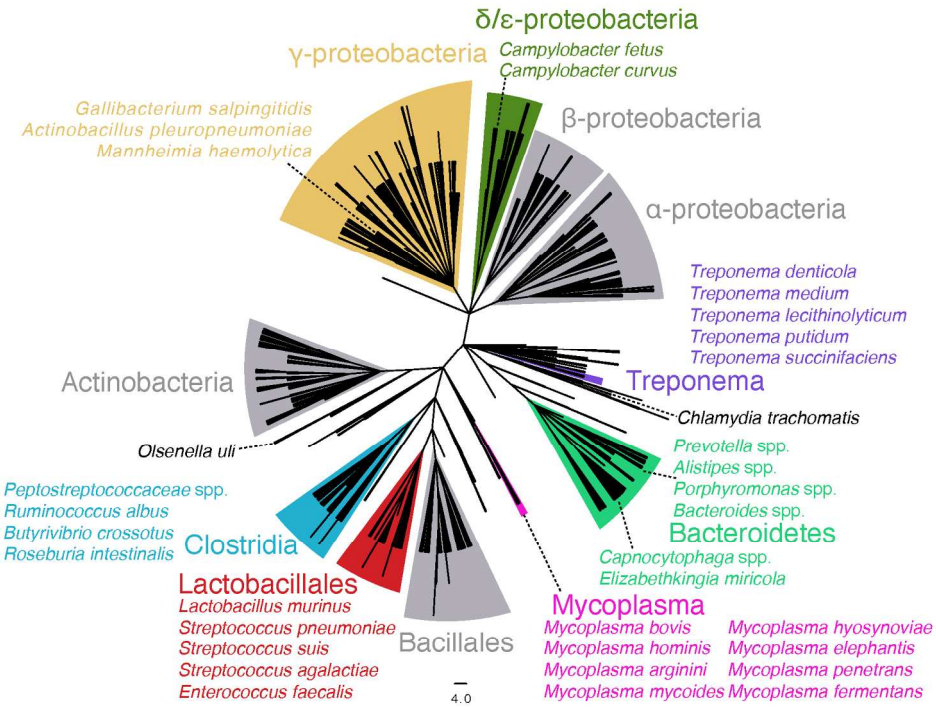
OXFORD UNIVERSITY PRESS | FEMS Microbiology Reviews

http://mc.manuscriptcentral.com/fems

Phase variable methylation and epigenetic control by Type I Restriction Modification systems

Journal:	*FEMS Microbiology Reviews*
Manuscript ID	FEMSRE-17-02-0012.R1
Manuscript Type:	Review Article
Date Submitted by the Author:	n/a
Complete List of Authors:	De Ste Croix, Megan; University of Leicester, Genetics Vacca, Irene; University of Leicester, Genetics Kwun, Min; Imperial College London, Department of Infectious Disease Epidemiology Ralph, Joseph; University of Leicester, Genetics Bentley, Stephen; Wellcome Trust Sanger Institute Haigh, Richard; University of Leicester, Genetics Croucher, Nicholas; Imperial College London, Department of Infectious Disease Epidemiology oggioni, marco; University of Leicester, Genetics
Keywords:	Epigenetics, Restriction-modification, Phase variation, DNA methylation

SCHOLARONE™ Manuscripts



276x213mm (200 x 200 DPI)

eview

Phase variable methylation and epigenetic regulation by Type I restriction modification systems

Megan De Ste Croix[1,*], Irene Vacca[1,*], Min Jung Kwun[2], Joseph Ralph[1], Stephen D. Bentley[3], Richard Haigh[1], Nicholas J Croucher[2], Marco R Oggioni[1,&]

[1] Department of Genetics, University of Leicester, Leicester LE1 7RH, United Kingdom

[2] Department of Infectious Disease Epidemiology, Imperial College London, London W2 1PG, United Kingdom

[3] Infection Genomics, Wellcome Trust Sanger Institute, Hinxton, CB10 1SA, United Kingdom

* Both authors contributed equally

[&] correspondence to Marco R Oggioni, Department of Genetics, University of Leicester, Leicester LE1 7RH, UK. Email mro5@leicester.ac.uk; phone +44 (0)116 252 2261

For Peer Review

Abstract

Epigenetic modifications in bacteria, such as DNA methylation, have been shown to affect gene regulation, thereby generating cells which are isogenic but with distinctly different phenotypes. Restriction-Modification systems contain prototypic methylases which are responsible for much of bacterial DNA methylation. This review focuses on a distinctive group of Type I restriction-modification loci which, through phase variation, can modify their methylation target specificity and can thereby switch bacteria between alternative patterns of DNA methylation. Phase variation occurs at the level of the target recognition domains of the *hsdS* (specificity) gene via reversible recombination processes acting upon multiple *hsdS* alleles. We describe the global distribution of such loci throughout the prokaryotic kingdom and highlight the differences in loci structure across the various bacterial species. Though restriction modification systems are often considered simply as an evolutionary response to bacteriophages, these multi-*hsdS* Type I systems have also shown the capacity to change bacterial phenotypes. The ability of these RM systems to allow bacteria to reversibly switch between different physiological states, combined with the existence of such loci across many species of medical and industrial importance, highlights the potential of phase variable DNA methylation to act as a global regulatory mechanism in bacteria.

For Peer Review

ScholarOne Support 1-434/964-4100

Introduction

It is just over 50 years since enzymatic modification and restriction of both bacteriophage and bacterial chromosomal DNA was first described (Arber and Dussoix 1962). Since then, multiple families of DNA modification and restriction-modification enzymes have been identified, and a wide variety of functions have been characterised in addition to their initially described role as defence mechanisms against foreign DNA (Vasu and Nagaraja 2013). The main aim of this review is to summarise the information currently available about a distinct group of phase variable type I restriction-modification systems in species where there is the potential for epigenetic effects upon bacterial gene expression and complex phenotypes (Dybvig, Sitaraman and French 1998; Manso *et al.* 2014; Li *et al.* 2016). General information on methylation systems, and phase variable methylation systems in particular, has been reviewed in depth elsewhere (Murray 2000; Srikhanta, Fox and Jennings 2010; Loenen *et al.* 2014a, 2014b).

DNA methylation systems

DNA methylation has been shown to be an increasingly common feature of prokaryotic genomes and it is present in more than 90% of species studied (Blow *et al.* 2016). Chemical modification is either at an adenine or a cytosine altering the base to 6-methyladenosine (m6A), 4-methylcytosine (m4C) or 5-methylcytosine (m5C) respectively; of these modifications the m6A account for 75% all observed prokaryote methylation (Blow *et al.* 2016). Such DNA methylation can now be routinely identified during whole genome DNA sequencing using the single molecule real-time (SMRT) sequencing system developed by Pacific Biosciences (Clark *et al.* 2012).

DNA methylation modifications are generated by a methyltransferase enzyme (MTase), which facilitates the transfer of a methyl group from a molecule of S-adenosyl-L-methionine onto the relevant base (Bheemanaik, Reddy and Rao 2006). While orphan MTases are not uncommon, in bacteria MTases are typically found within restriction modification (RM) systems. With such RM systems, the addition of a methyl group to a nucleotide within a specific DNA target sequence then allows the DNA molecule to be recognised as self, thereby protecting it from the restriction function. A random DNA molecule entering a cell is very unlikely to be methylated in the correct pattern and therefore the cell will recognise it as foreign and, as a result, will be capable of cleaving it. While the ability to recognise self-DNA is the primary purpose of DNA methylation for RM systems, the addition of methyl groups to DNA can however have other effects. Indeed when a methyl group is added to a base the structure and dynamics of the DNA molecule are altered which may in turn result in changes in DNA-protein interactions (Marinus and Casadesus 2009); such structural changes are known to be able to regulate gene expression.

ScholarOne Support 1-434/964-4100

For Peer Review

One of the best studied of all the prokaryotic MTases is the DNA adenine methylase (Dam) of the γ-proteobacteria (Heusipp, Fälker and Alexander Schmidt 2007; Marinus and Casadesus 2009). This orphan MTase is responsible for the methylation of the four base sequence GATC (Heusipp, Fälker and Alexander Schmidt 2007; Marinus and Casadesus 2009). Methylation of GATC in bacteria serves several functions. For example, the addition of a methyl group to *oriC* promotes the binding of the replication initiation complex (Wion and Casadesús 2006). Additionally, the hemi-methylation on GATC that arises due to DNA strand replication allows the recognition of the parent strand from the daughter strand, meaning that replication errors can be identified and corrected by the cells mismatch repair machinery (Marinus and Casadesus 2009).

An example of the tight regulation of cell cycle by DNA methylation can be seen in *Caulobacter crescentus.* This bacterium exists in two cell types; swarmer cells, which are motile, and stalked cells, which adhere and are able to replicate. In swarmer cells, the MTase CcrM is active and the DNA is fully methylated on the adenine of GANTC sites, which is essential for activation of the promoter of *dnaA*; however, the regulator CtrA binds to the C*ori* of *C. crescentus* preventing DnaA binding and blocking initiation of DNA replication (Marinus and Casadesus 2009). Over time, CtrA is degraded by protease ClpXP and the DnaA binds to C*ori* initiating replication of the genome, which results in the production of hemi-methylated DNA. This prevents further expression of *dnaA,* and activates a cascade of genes that are only expressed when GANTC is hemi-methylated, which coordinate the swarmer to stalked cell transition. One of these expressed genes is *ctrA,* whose protein again binds to *Cori* preventing DNA initiation, as well as activating the expression of CcrM and FtsZma (Marinus and Casadesus 2009). Once CcrM is expressed, it is able to methylate the newly synthesised DNA strands, resulting in complete methylation of the *dnaA* promoter again thereby allowing for a new round of DNA replication only when the one before has been completed (Marinus and Casadesus 2009).

Restriction-modification systems

Within RM systems, four families have been recognised based upon their mechanism of action, recognition sequence patterns and enzyme structure (Loenen *et al.* 2014a, 2014b). These bifunctional systems are capable of the modification of DNA bases in specific motifs through the addition of methyl groups, as well as the degradation of DNA molecules by cleavage of the phosphodiester backbone. The RM enzymes commonly used in the laboratory are those of the Type II family of enzymes (Loenen *et al.* 2014b) as these enzymes cleave and methylate DNA at, or close to, a fixed recognition site, typically a 4-8 bp palindromic sequence (Pingoud and Jeltsch 2001; Srikhanta, Fox and Jennings 2010). Type II systems are a diverse group of enzymes but usually consist of two separate enzymes for

ScholarOne Support 1-434/964-4100

For Peer Review

restriction and methylation (*res* and *mod*), which recognise the same sequence. Type II enzymes are highly heterogeneous, and include multiple enzyme families that have evolved independently. Type IIP enzymes are the classical enzymes used in molecular biology that recognise as self a palindromic target with methylation on both strands, while cleaving unmethylated DNA within that same target. In contrast, Type IIS enzymes are less common and cleave outside their recognition sequence (Loenen *et al.* 2014b). Lastly, the Type IIG enzymes are large enzymes which contain both of their enzymatic activities within the same protein (Loenen *et al.* 2014b). Similarly to the Type II systems, Type III RM systems are also typically encoded by *mod* and *res* genes, which are co-transcribed (Srikhanta, Fox and Jennings 2010). Methylation by a Type III RM system is single stranded and specific; however these RM enzymes are less commonly used in genetic engineering as the cleavage site occurs 25-27 bp downstream of the 5-6 bp recognition sequence (Srikhanta, Fox and Jennings 2010). Type I RM enzymes differ from the others in that they are encoded by three separate genes: *hsdR, hsdM* and *hsdS*, determining the endonuclease, methyltransferase and specificity subunits, respectively. Type I enzymes also have fewer applications in biotechnology because though the DNA methylation occurs on both strands of a specific asymmetric, bi-partite sequence (Loenen *et al.* 2014a), the DNA cleavage step occurs at random sites some distance away (Loenen *et al.* 2014b). The Type IV RM family is the most diverse. Few enzymes belonging to these systems have been well characterised and they show little common sequence specificity other than their absolute requirement that the DNA site contains one of several possible modifications (Loenen and Raleigh 2014). Cleavage is remote from the target sequence (Loenen *et al.* 2014b).

The nature of the motifs bound by RM enzymes, typically being palindromic or bipartite, reflects the symmetrical interaction of the RM proteins with both strands of the DNA double helix. RM loci are therefore efficient at targeting bacteriophage, infection with which usually occurs by injection of double stranded DNA. By contrast, the mechanism of natural transformation and some conjugation systems involves the transfer of single stranded DNA (ssDNA), which is typically insensitive to cleavage by RM systems, as they can only bind double stranded DNA (Johnston, Polard and Claverys 2013). Conjugative elements can further avoid barriers by encoding anti-restriction proteins, which protect the newly formed unmethylated dsDNA from the actions of RM systems (Wilkins 2002). It has been demonstrated that the import of novel genomic islands by transformation can still be inhibited by typical RM systems; this occurs by RM cleavage of the acquired sequence after it has been integrated into the chromosome and after a complementary, but also unmethylated second strand, has been synthesised.

There is also an example of a system, DpnII, which through a mechanism that seems to have evolved to rescue incoming DNA, does not even act as a complete barrier to

ScholarOne Support 1-434/964-4100

For Peer Review

differentially-methylated dsDNA, and which is found to help avoid such post-transformation cell suicide (Johnston, Polard and Claverys 2013). All *S. pneumoniae* strains contain one of three Type II RM systems, DpnI, DpnII or DpnIII. Each of these is found at the same position in the genome and, like Dam, all recognise the sequence GATC. However they have variation in site methylation and specificity (Johnston, Polard and Claverys 2013). Whereas DpnII is a conventional RM system that restricts DNA at unmethylated GATC motifs, DpnI has the unusual activity of targeting DNA methylated at $G^{m6}ATC$. The complementary features of this pair of enzymes have resulted in them being used extensively to experimentally verify Dam methylation of target sites. DNA from donor cells with a *Dpn*I system is unmethylated at GATC motifs, and hence once incorporated into the genome, can generate hemi-methylated loci, which themselves do not undergo restriction in a *Dpn*II recipient. However, if a replication fork passes over newly incorporated DNA before it has been fully methylated, it could lead to a newly synthesised DNA strand that is completely unmethylated, resulting in restriction of the chromosomal DNA in a *dpnII* cell (Johnston, Polard and Claverys 2013). To avoid cell suicide through this mechanism, the DpnII system not only includes a double stranded DNA (dsDNA) MTase, but also for a rare ssDNA MTase, DpnA, which is only expressed during competence (Johnston, Polard and Claverys 2013) and ensures that restriction of newly acquired loci in the chromosome does not occur. The third Dpn system, DpnIII, is found in a small proportion of the pneumococcal population (Croucher *et al.* 2014), including the multidrug-resistant lineage PMEN1 (Croucher *et al.* 2011). The DpnIII system recognises and methylates the cytosine rather than the adenine of the GATC recognition sequence (Eutsey *et al.* 2015). DpnIII will therefore restrict DNA from strains with either the DpnI or DpnII system. Isolates in which this system was disrupted were observed to undergo extensive acquisition of divergent sequence through recombination, suggesting this RM system may have an important role in inhibiting acquisition of novel loci (Eutsey *et al.* 2015).

It is intriguing that all three of these RM systems recognise the same motif, GATC, albeit with different methylation patterns. This is an effective way of mounting a barrier to reciprocal transfer between genotypes: if the systems targeted different motifs, then dsDNA that included only one of the motifs would be cut when transferring from one donor to a recipient, but not in the reverse direction. By targeting the same sequence, any dsDNA including this motif will be cut when passing between any pair of cells that differ in their *Dpn* system. That the GATC motif is shared with the Dam methylase suggests there may be another function beyond restricting DNA acquisition, although DpnI does not methylate host DNA, thus making a restriction independent function difficult to discern.

Type I Restriction-Modification systems

ScholarOne Support 1-434/964-4100

For Peer Review

Type I RM systems encode enzymes capable of both methylating and cleaving (restricting) host and foreign DNA. These systems consist of three *h*ost *s*pecificity *d*eterminant (*hsd*) genes: *hsdR, hsdM* and hsdS. Each of these genes produces a subunit of the complete enzyme, known as the Restriction (R), Modification (M) and Specificity (S) subunits respectively. The *hsdS* gene encodes a DNA binding protein containing two target recognition domains (TRD), each recognising one of the two half-sites of the bipartite target. These two TRDs are separated by a conserved peptide sequence that determines the number of nucleotides between the two elements of the target sequence. In the example of the EcoKI system of *Escherichia coli*, the N-terminal TRD recognises AAT while the C-terminal TRD recognises GTGC of the bipartite $AACN_6GTGC$ target (Kan *et al.* 1979).

Type I enzyme subunits can form two different complexes: a pentameric enzyme that shows endonuclease activity (RTase) composed of two R's, two M's and an S subunit, or a trimeric (MTase comprising only two M's and an S subunit (Fig. 1A-B) (Murray 2000; Loenen *et al.* 2014a; Blow *et al.* 2016). The full RTase complex is capable of binding to and translocating double stranded DNA as the R-subunits are DNA helicases (Loenen *et al.* 2014a)' (Murray 2000). Type I enzymes bind and methylate at their recognition site; however, unlike Type II RM systems, binding to unmethylated recognition sites does not result in immediate DNA cleavage, but instead initiates translocation of the flanking DNA double strands by the helicase action of the R-subunits. Indeed the endo-nucleolytic cleavage activity only occurs when the helicase activity of the RM complex is prevented from further translocating DNA as would occur by collision with another RM complex or by encountering another form of obstruction, such as supercoiled DNA (Loenen *et al.* 2014a). This means that the restriction event can often occur several kb from the original recognition sequence (Blow *et al.* 2016). While endonucleolytic cleavage is triggered by unmethylated DNA, methylase activity is favoured by hemi-methylated DNA, as would be found after chromosome replication, thereby allowing methylation to be used to distinguish between self and non-self DNA. This ability to discriminate self from non-self means that all RM systems, not just Type I systems, are predominantly viewed as a defence mechanism against invading foreign DNA, but the increasing amount of data relating to a potential epigenetic impact of distinct methylation patterns indicates that these systems play a much larger role (Furuta *et al.* 2014).

The majority of Type I RM systems methylate adenines in both half-targets generating one 6-methyladenine (m6A) residue on each DNA strand. These m6A modifications then function to protect DNA from restriction by the respective R subunit of the system (Murray 2000; Loenen *et al.* 2014a). It has however been recently reported that a 4-methylcytosine (m4C) methylation pattern is also associated with Type I RM systems in *Desulfobacca acetoxidans*, *Methanohalophilus mahii, and Pseudomonas alcaligenes* (Blow *et al.* 2016; Morgan *et al.* 2016). These m4C MTases are found in systems which contain two MTases, and where the

ScholarOne Support 1-434/964-4100

For Peer Review

second MTase methylates at m6A. In cases where one-half of the bipartite target sequence contains only G's and C's and m6A methylation is not possible, m4C methylation is used instead. The m6A MTase is then used to methylate the other half of the sequence, which contains an adenine (Blow *et al.* 2016). Other similar dual MTase systems can be found in REBASE (New England Biolabs online Restriction Enzyme database; http://rebase.neb.com/rebase/rebase.html) suggesting that this may be a more widespread occurrence; however, this remains to be determined as for the majority of them their recognition sequences have not yet been identified (Roberts *et al.* 2015; Blow *et al.* 2016).

"On/off" phase variable DNA methylation

Several types of phase variable mechanisms affecting DNA methylation have been identified across multiple different bacterial species (Bayliss 2009; Srikhanta *et al.* 2011; Casadesús and Low 2013; Atack 2015; Seib *et al.* 2015; Anjum *et al.* 2016). As described previously, the *E. coli* Dam orphan MTase has a role in regulating several cellular functions including the initiation of DNA replication and the identification of newly synthesised daughter strands via hemi-methylation (Marinus and Casadesus 2009). Interestingly, phase variable occurrences of Dam regulation have also been reported; for example, Dam is known to be responsible for the ON/OFF switching of the *pap* (pyelonephritis-associated pili) operon of uropathogenic *E. coli* strains (Casadesús and Low 2013). The locus encodes three genes, *papA, papB* and *papI*, and changes in Dam methylation impact transcription of the operon, showing a complex, inheritable and reversible method of regulating gene expression (Bayliss 2009). Dam competes with the leucine responsive regulatory protein, Lrp, to bind and methylate two GATC sites, known as sites 1-3 and 4-6, which are found within the Lrp binding regions in the promoter of the *pap* locus (Bayliss 2009; Casadesús and Low 2013). If Lrp binds to site 2, it blocks Dam methylation and prevents transcription of *pap*, by also blocking the RNA polymerase binding site. Alternatively, if Dam is blocked by Lrp from binding site 5, then *pap* transcription is promoted. It is thought that these switches occur through structural changes that result in the RNA polymerase binding site becoming more accessible (Casadesús and Low 2013). Overall, to switch from an OFF state to an ON state, PapI and Lrp form a complex that has a high affinity for hemi-methylated DNA in sites 4-6, consequently promoting the recruitment of Lrp to site 5, and preventing methylation by Dam (Casadesús and Low 2013).

There are also instances in which phase variation directly affects the activity of RM methylases thereby influencing global gene expression; such a phase variable Type IIG RM system that has recently been described in *Campylobacter jejuni* (Anjum *et al.* 2016). In this bacterium, the endonuclease and the methyltransferase are encoded by a single gene, *cj0031*. The presence of a poly-guanosine (polyG) repeat tract within the *cj0031* gene can

ScholarOne Support 1-434/964-4100

For Peer Review

change, via slipped-strand mispairing, the reading frame and therefore control the switch between expression of a full length or truncated protein. *Cj0031* has been shown to methylate the adenine of both 5'CCCGA and 5'CCTGA resulting in differential gene expression (Anjum *et al.* 2016). *C. jejuni* cells not expressing *cj0031* were found to be less efficient in their adherence to and invasion of Caco-2 cells, as well as forming significantly less biofilm, suggesting this RM system is required for full expression of cell surface molecules. The gene is not however universally phase variable as homologues in several other *Campylobacter* strains were found to lack the polyG tract.

Phase variation of Type III RM systems through changes in repeat tract length has been reported in many bacterial species such as *Helicobacter pylori* (Srikhanta *et al.* 2011), *Mannheimia haemolytica, Haemophilus influenzae, Neisseria meningitidis* and *Neisseria gonorrhoeae* (Srikhanta, Fox and Jennings 2010). In *H. pylori*, the phase variable *modH* gene of a Type III RM system has 17 *mod* alleles, each conferring a different specificity for methylation. Microarray analysis has confirmed that the loss of the system leads to changes in *in vitro* gene expression (Srikhanta *et al.* 2011). The regulation of multiple genes by a phase variable RM system has been termed a "phasevarion" (Srikhanta, Fox and Jennings 2010). While only a small number of genes are affected (six genes with a >1.6 fold change when compared to a strain with an intact RM system), they include the surface exposed protein HopG and several flagella genes (*flaA* and *fliK*) that are required for motility. These data only represent changes in gene expression seen with a single *modH* allele, however, there are potentially 17 different phasevarions in *H. pylori* each capable of altering the expression of a small subset of different genes (Srikhanta *et al.* 2011).

In 2015 Seib *et al* determined the methylation motifs of three phase variable type III MTases of *N. meningitidis* (*modA11, modA12* and *modD1I)* by SMRT sequencing. When expressed, each of the three *mod* genes recognised and methylated a unique motif and this has been associated with the regulation of a specific set of genes (phasevarion) (Seib *et al.* 2015). Furthermore, Atack *et al* used SMRT sequencing to determine the methylation motifs of the five most clinically prevalent *mod* genes in *H. influenzae.* The reversible ON/OFF switching of *modA2, A4, A5, A9* and *A10*, has been linked to a large number of virulence phenotypes, including antibiotic resistance, immune evasion and biofilm formation (Atack 2015).

Phase variable motif specificity of DNA methylation

Phase variation observed in Type II RM systems typically consists of a reversible switch between an active and an inactive form of the gene, affecting the presence or absence of the methylation enzyme. This is because there is no simple mechanism by which the sequence specificity of both the MTase and RTase can be redirected in a co-ordinated manner. Whilst these on/off systems allow phase variation the one-dimensional nature of the control

ScholarOne Support 1-434/964-4100

For Peer Review

mechanism means it is relatively inflexible. By contrast, phase variation of Type I RM systems can be more complex, and therefore potentially more useful, because they can be easily "reprogrammed" to vary between different DNA target motifs. This flexibility relies on the modular nature of the two TRDs of the S subunit, which each recognise half of the non-palindromic bipartite target sequence (Murray 2000; Loenen *et al.* 2014a). Alterations of the S subunit-encoding gene as small as single nucleotide polymorphisms can lead to recognition of new target sequences for DNA methylation (Adamczyk-Poplawska, Lower and Piekarowicz 2011; Vasu and Nagaraja 2013). But more importantly, the presence in some *hsd* operons of multiple variant *hsdS* genes, or partial genes containing only single TRDs, allows phase variation through TRD "shuffling" by recombination between the *hsdS* genes (Dybvig and Yu 1994; Cerdeño-Tárraga *et al.* 2005; Manso *et al.* 2014; Li *et al.* 2016). By moving complete or partial regions of *hsdS* genes into and out of the actively transcribed *hsd* operon it is possible to reversibly express multiple different S alleles. Moreover, as this reversible switching can occur with high frequency it will result in the continuous generation of diversity within a population. Phase variable, or invertible, type I RM sequences have been identified in a variety of species, including *Mycoplasma pulmonis* (Dybvig and Yu 1994; Sitaraman and Dybvig 1997; Ron *et al.* 2002), *Bacteroides fragilis* (Cerdeño-Tárraga *et al.* 2005), *Streptococcus pneumoniae* (Tettelin *et al.* 2001; Mostowy *et al.* 2014; Li *et al.* 2016; Lees *et al.* 2017), *Streptococcus suis* (Willemse and Schultsz 2016; Willemse *et al.* 2016), *Listeria monocytogenes* (Furuta *et al.* 2014) *and Lactobacillus salivarus* (Claesson *et al.* 2006) (Fig. 2).

The earliest identified of the phase variable Type I systems was that found in *M. pulmonis,* the first non-enteric bacterium to be found with a Type I RM system (Sitaraman and Dybvig 1997), where there is a 6.8 kb invertible locus termed Hsd1 (Dybvig and Yu 1994; Sitaraman and Dybvig 1997). The structure of the Hsd1 locus and the presence of inverted repeats allows for the generation of four different functional *hsdS* genes through TRD shuffling (Sitaraman and Dybvig 1997). In addition, there is a second highly similar type I RM locus termed Hsd2 (Dybvig and Yu 1994). Hsd1 and Hsd2 appear to be conserved across *M. pulmonis* strains, whereas a third non-functional type I RM system has only been identified in the strain UAB CTIP (Chambaud *et al.* 2001). Recombination at the *hsd* loci of *M. pulmonis* has been confirmed by PCR analysis using pairs of primers where just one of them is situated within the inverted region. Inversions resulted in novel positive PCR products and proved that the system was switching its TRDs both *in vitro* (Sitaraman and Dybvig 1997) and *in vivo* (Gumulak-Smith *et al.* 2001).

A comparable invertible Type I system was identified in *Streptococcus pneumoniae* as a hypervariable locus within the first whole genome assembly, that of TIGR4 (Tettelin *et al.* 2001). This RM system, *Spn*III, is encoded at the *ivr* locus and relies on three sets of

ScholarOne Support 1-434/964-4100

For Peer Review

inverted repeats to allow the exchange of five different TRDs in order to generate six different S subunit alleles named A-F (Manso *et al.* 2014). The *ivr* locus was found to be present across all isolates in a diverse population (Croucher *et al.* 2014).

Two other structurally very similar invertible Type I RM systems have also been reported; these are in *B. fragilis* strain NCTC 9343 (Cerdeño-Tárraga *et al.* 2005), and in the ST8 strains of *L. monocytogenes* (Fagerlund *et al.* 2016). In *B. fragilis,* in addition to the complete *hsdS* gene, which sits in line with the *hsdR* and *hsdM* genes, there is an inactive *hsdS* gene lacking a start codon which is downstream and in the opposite orientation. To create even more options for S subunits there are also two additional truncated *hsdS* genes that, via four pairs of inverted repeats, allow the generation of eight alternate active S subunits from the six different TRDs (Cerdeño-Tárraga *et al.* 2005). The recombination of *hsdS* genes in the *B. fragilis* genome has, to the best of our knowledge, not yet been experimentally confirmed. In *L. monocytogenes*, a single non transcribed *hsdS* gene is situated downstream of the *hsdRMS* locus and in the opposite orientation. Each of the two *hsdS* genes contains two TRDs allowing for the generation of four possible S subunit specificities, named A-D (Fagerlund *et al.* 2016). Genome comparison by Fagerlund *et al.* (Fagerlund *et al.* 2016) demonstrated that different active *hsdS* genes were present in different genomes and therefore they inferred that they might be phase variable. We also have unpublished data showing that recombination occurs at inverted *hsdS* genes within *Listeria monocytogenes* strains.

In 2006, another phase variable Type I system (LSL_0915-LSL_0920), controlled by DNA inversion events at intragenic inverted repeat sites, was identified in the *Lactobacillus salivarius* genome (Claesson *et al.* 2006). This Type I RM system is encoded in a region of 9,360 bp that has been defined as a “shufflon”. The shufflon contains two extra complete copies of the *hsdS* gene, downstream and outside of the *hsd* operon, which would potentially allow for the generation of a total of nine possible combinations for the active specificity subunit.

Distribution of inverting phase variable Type I RM systems

The phase variable Type I RM systems are found in multiple species across the diversity of bacteria. As an example, proteins with very high sequence similarity to the *Spn*III RM system’s recombinase IvrR can be found across many taxa, thereby identifying many candidate phase variable type I RM systems (Fig. 4). Like *Mycoplasma* and *Bacteriodetes*, the *Treponema* and *Capnocytophaga* genera are dense with species containing candidate phase variable loci. Other genera, such as *Campylobacter* and *Lactobacillus*, exhibit a more sporadic distribution with candidate phase variable RM systems found in only a few species, although this may be an artefact either of sampling or of the simple search approach used.

ScholarOne Support 1-434/964-4100

For Peer Review

Nevertheless, this stringent search still identified representatives in both Gram-positive and -negative isolates from a variety of habitats, suggesting these systems are important in the evolution of many diverse bacteria. Further analysis will identify whether their disparate distribution represents multiple independent emergences or the horizontal transfer of these loci between highly divergent recipient cells.

The situation in *S. pneumoniae* is somewhat peculiar in that both of the type I phase variable RM systems are part of the core genome (Croucher *et al.* 2014), yet both are absent from many representatives of the *Streptococcus mitis* complex, to which *S. pneumoniae* belongs (Kilian *et al.* 2014). In many other species the presence of phase variable systems is restricted to groups of related strains, as is seen in the case of the clonal complexes in *L. monocytogenes* (Fagerlund *et al.* 2016) or serotypes of *S. suis* (Willemse and Schultsz 2016; Willemse *et al.* 2016). These examples indicate that the phase variable Type I RM systems appear to be associated generally with particular lineages that often do not correspond with species boundaries.

Allele quantification of invertible Type I RM systems

Both quantitative and non-quantitative methods have been used for measuring inversions within phase variable Type I systems. The non-quantitative protocols allow for the rapid detection of systems where inversions are occurring; however, quantitative systems that can determine the proportions of individual S subunits within the population are essential to understand the nature of the phase variation process. When the Hsd system of *M. pulmonis* was first identified, a number of primer pairs were used to detect on-going inversions (Dybvig and Yu 1994; Sitaraman and Dybvig 1997). The primers used were designed to be co-directional within the locus with the intention that they would then only be able to generate a PCR amplicon following an inversion event. This method allowed for the detection of the presence of alternative S subunits, but due to the nature of the PCR methodology it gave no real indication of the proportion of each subunit within the population. SMRT sequencing has made the study of these phase variable methylation systems much more convenient (Clark *et al.* 2012). In the case of species with actively inverting Type I systems it is quite possible for a DNA sample to contain more than one of the possible methylation patterns (Feng *et al.* 2014; Manso *et al.* 2014). If multiple patterns can be detected within a sample, this strongly suggests that any observed variation at a RM locus is having a phenotypic effect on the pattern of DNA methylation. Unfortunately, because the SMRT software requires a minimum number of potentially modified reads to be detected before it can determine whether a nucleotide is methylated, this system cannot currently be used to accurately quantify the abundance of S subunits in a given sample, but they can be directly used for a rough quantification if the read depth is adequate.

ScholarOne Support 1-434/964-4100

For Peer Review

In *S. pneumoniae* a quantitative PCR method was developed to measure the *Spn*III S subunit proportions in a mixed population (Manso *et al.* 2014). In this method, the entire *hsdS* containing region of the *ivr* locus is PCR-amplified using a pair of primers where just the forward primer is FAM labelled. The PCR products generated are then digested using restriction enzymes *Dra*I and *Ple*I resulting in a uniquely sized, FAM-labelled fragment for each active S subunit represented within the sample tested. These fragments are then run on an ABI prism Gene Analyser (Life Technologies) and analysed using the programme Peak Scanner v1.0, which allows the determination of the relative abundance of each S subunit.

Another quantitative method that has been used to measure the proportion of S subunits in both *S. pneumoniae* (Lees *et al.* 2017) and *Lactobacillus salivarius* (Claesson *et al.* 2006) relies on the analysis and quantification of whole genome sequencing reads. In *L. salivarius*, individual sequence reads could be mapped to one of nine different *hsdS* combinations present in the DNA sequenced, and the number of reads for each was used to quantify their relative abundance in the sample. In *S. pneumoniae*, sequence reads having homology to the locus were individually mapped first to TRD1 and then to TRD2. This allowed each read to be assigned to one of the six S subunits and the relative proportions of each in the sequenced genome could be determined (Lees *et al.* 2017).

Recombinase-driven rearrangement of *hsdS* genes

Several of the phase variable Type I RM systems described above have a site specific recombinase associated with the operon. The pneumococcal *Spn*III system, contains a site specific tyrosine recombinase situated between the active and silent *hsdS* genes (Tettelin *et al.* 2001; Manso *et al.* 2014; Li *et al.* 2016); however, it is interesting to note that this recombinase has only been shown to be partially responsible for control of recombination at the locus (Li *et al.* 2016). The *ivr* locus contains three independent pairs of inverted repeats (IRs) of varying sizes (330 bp, 85 bp and 15 bp (Manso *et al.* 2014)); however only recombination at the 15 bp IR sequence is exclusively controlled by the tyrosine recombinase, and in a recombinase knockout strain, recombination on the two larger repeats has been shown to continue (Li *et al.* 2016). There is a conserved 10 bp sequence that is common to all three of the repeats and which may act as the recognition site for a recombinase; however, there is clearly more than one recombination mechanism involved in the inversion of the *spnIII* locus. Both our own work and the work of Li *et al.* (Li *et al.* 2016) has confirmed that recombination on the two larger repeats occurs independently of RecA. Therefore, there is likely to be another facilitator protein or recombinase that still needs to be identified that promotes recombination on these two repeats; this situation is therefore

ScholarOne Support 1-434/964-4100

For Peer Review

similar to the case of the recombination of the flagella subunits in *Salmonella enterica* (Kutsukake *et al.* 2006).

In ST8 strains of *L. monocytogenes* DNA sequence analysis indicates that the phase variable Type I RM system utilizes site-specific recombination to switch between the four possible DNA target recognition site specificities (Fagerlund *et al.* 2016). The presence of two pairs of inverted repeats (5'-AGCTTGGGAACAGCGT-3' and 5'-CTATCGCTCTTCATCAGCGTAAGTTAGAT-3'), which are located in the 5'-end and in the central part of the *hsdS* genes respectively, appear to allow the recombination to occur. Although direct evidence for its role is still missing, an integrase, which contains active site residues similar to tyrosine recombinases is encoded between the two *hsdS* genes and thus is most likely responsible for the reported *hsdS* gene inversions (Fagerlund *et al.* 2016).

Within the Hsd1 locus of *M. pulmonis* there are two pairs of inverted repeats (5'-CAAAGTGCAATA-3' and 5'-TAATTAAGATTATTGAACCT-3'), which allow the generation of four different active S subunits (Sitaraman, Denison and Dybvig 2002). Unlike the systems identified in *S. pneumoniae* and *L. monocytogenes*, inversions in the *hsdS* genes are the result of a single site specific tyrosine recombinase, HvsR, which is not found within the Type I locus (Sitaraman, Denison and Dybvig 2002), but is instead located near the phase variable *vsa* surface protein genes (Shen *et al.* 2000). Furthermore, in what appears to be a unique situation, the recombinase controlling the inversion at the *vsa* locus also has complete control of *hsdS* gene inversions. Using transposon mutagenesis to generate mutants of HvsR, Sitaraman *et al.* were able to prove that, despite the lack of sequence similarity between the two systems, it is indeed HvsR that facilitates recombination at both the *vsa* and *hsdS* loci (Sitaraman, Denison and Dybvig 2002). When the system was reconstructed within an *E. coli* background, it was determined that HvsR alone was capable of facilitating recombination when the longer 20 bp IR sequence was present, whereas in the presence of just the shorter IR sequence it did not appear to be sufficient (Sitaraman, Denison and Dybvig 2002), indicating that another mycoplasma protein may be required.

Recently Willemse *et al.* characterised the genomic differences between porcine and human isolates of the zoonotic pathogen *S. suis* (Willemse *et al.* 2016). They identified two genomic differences that characterise invasive human isolates within clonal complex CC20: a novel prophage remnant that contains a novel phase variable Type I RM system and a pathogenicity island containing virulence genes (Willemse *et al.* 2016). The phase variable RM system was restricted to mainly to serotype 2 isolates, the serotype responsible for human invasive disease. The phase variable system contains two inverted *hsdS* alleles, with the multiple recombination states identified in the sequences of individual isolates (Willemse and Schultsz 2016).

ScholarOne Support 1-434/964-4100

In *B. fragilis*, the specific recombinase associated with the BF1839 Type I system has not yet been determined. The genome contains more than 30 site specific recombinases (Nakayama-Imaohji *et al.* 2009), three of which are in close proximity to the locus (BF1833, 1843, and 1845) (Cerdeño-Tárraga *et al.* 2005); however, there is currently no published evidence showing which, if any, of these recombinases permits inversions at the Type I RM locus.

The *hsdS* inversions investigated in *S. pneumoniae, B. fragilis* and *M. pulmonis* all appear to occur independently of RecA. While this can be reasonably explained in *M. pulmonis* and *B. fragilis* by the fact that the site specific recombinase is in sole control of the loci, in *S. pneumoniae* it is known that this is not the case.

Other phase variable type I RM systems

A number of other recombination mechanisms have been described, across different bacterial species, for being responsible for phase variable Type I RM systems, altering TRD sequences and therefore target specificity and genome methylation status and potentially affecting global gene expression patterns.

In *S. pneumoniae*, in addition to the flipping between inverted repeats previously described, it has been found that phase variation of type I RM loci can occur through DNA translocation between direct repeats (Croucher *et al.* 2014). This process occurs at the translocating variable restriction (*tvr)* locus, which encodes the *Spn*IV RM system (Manso *et al.* 2014). As with other phase variable type I RM loci, the *tvr* locus contains genes encoding the three subunits of a pentameric restriction enzyme and a recombinase gene, *tvrR*, but also has a toxin-antitoxin locus, *tvrAT*; however, unlike the inverting loci, all of the core RM genes of the *tvr* locus are encoded on the same strand. The role of the toxin-antitoxin system in this locus has not yet been studied in detail; however, it is proposed that it would be involved in stabilization of the RM system locus – at a population level – by allowing post-segregational killing of daughter cells in which part of the locus had been lost (Croucher *et al.* 2014). Such a mechanism, not observed in inverting loci, indicates the recombination mechanism facilitating the lateral movement of DNA may involve relatively unstable intermediates that can frequently result in partial deletion of this operon. The activity of this RM system was confirmed through SMRT sequencing of mutants in which functional *tvr* loci had been introduced into backgrounds that previously lacked the entire operon (Croucher *et al.* 2014; Manso *et al.* 2014), identifying methylation at typical bipartite Type I RM motifs. Unlike the *Spn*III system, the *Spn*IV system included notable inter-strain variation in its complement of TRDs. Hence the methylation profile at this locus is determined both by which TRDs are present in the genotype and by the pattern into which they are shuffled by intragenomic recombination.

ScholarOne Support 1-434/964-4100

In *H. pylori*, an alternative arrangement facilitates analogous shuffling of TRDs through a mechanism termed DoMo (domain movement). DoMo is capable of moving TRDs to generate different S subunits; however, this does not occur through DNA inversions of sequences that are found within the same locus, but instead involves the movement of TRDs both within a single *hsdS* gene and between homologous *hsdS* genes distributed at different loci (Furuta, Abe and Kobayashi 2010; Furuta *et al.* 2011). Sequence analysis of multiple *H. pylori* genomes has identified three homology groups of Type I *hsdS* ortholog genes distributed across six loci. In two of these groups, the two TRDs (TRD1 and TRD2) are each flanked by one of 5 different pairs of short (14-53 bp) direct repeat sequences. DNA recombination at the level of these flanking sequences determines the movement of a TRD between genes with similar direct repeats present at different loci. Some TRD sequences were observed to occur in either the TRD1 or TRD2 position in different *hsdS* alleles indicating that more complex recombination events between pairs of the different repeats had occurred. Recombination at the direct repeats can also affect the TRD numbers present in a single specificity gene, either by decreasing two TRDs into one TRD or by increasing them to three (TRD1-TRD2-TRD2). Using single-molecule real-time sequencing technology, DNA methylation sites were determined throughout the *H. pylori* genome for several closely related strains and found to be highly variable (Furuta *et al.* 2014). Each of the DNA methylation sequence motifs found were able to be associated to a specific homology group of the target recognition domains in the specificity-determining genes. These results broadly supported the proposed DoMo mechanism for sequence-specificity changes in DNA methyltransferases. Similar TRD1 and TRD2 movement has been reported for Type I restriction-modification enzyme specificity genes in two more eubacterial species: *Streptococcus pyogenes* and *Mycoplasma agalactiae* (Furuta *et al.* 2011) .

In *Lactococcus lactis* a mechanism called ‘combinational variation' (O’Sullivan *et al.* 2000) has been reported in several studies and appears to represent a general strategy through which bacteria can acquire RM systems with novel specificities. Such diversification of RM loci increases the range of phage against which cells are protected (O’Sullivan *et al.* 2000), of particular economic relevance for lactococci commonly used in food starters and fermentation processes, which can be severely impacted by phage infection. According to this mechanism, DNA recombination occurs between specificity genes encoded on plasmids and chromosomally encoded *hsdS* genes. Schouler *et al.* (Schouler *et al.* 1998a) identified a type I RM system for the first time in *Lactococcus* on the natural plasmid pIL2614 and showed that interaction of different HsdS subunits present in distinct RM systems, located in the genome, generated enhanced phage restriction. Later on, they also demonstrated the acquisition of new RM specificities by transferring natural plasmids containing *hsdS* specificity genes to a lactococcal strain possessing a different *hsd* locus (Schouler *et al.*

ScholarOne Support 1-434/964-4100

For Peer Review

1998b). Other studies have also reported the presence of natural plasmids containing *hsdS* genes, which are able to interact both with plasmid-encoded and with chromosomally encoded HsdR and HsdM subunits generating active RM systems with new specificities (Madsen, Westphal and Josephsen 2000; Seegers, Van Sinderen and Fitzgerald 2000). For example, in *L. lactis subsp. cremoris* strain UC509.9, in addition to the *hsdS* gene encoded in plasmid pCIS3, there are two more *hsdS* genes located on the chromosome and another on a second plasmid, named pCIS1. In another study, the HsdS specificity subunit, S.LlaW12I, was found to be encoded in the naturally occurring 8.0 kb plasmid pAW122 of *L. lactis subsp. cremoris* W12 strain (Madsen, Westphal and Josephsen 2000) .

Acquisition of novel restriction specificities through TRDs shuffling in *L. lactis* can also occur through recombination of *hsdS* subunits encoded on different plasmids (O'Sullivan *et al.* 2000). In this case, the recombination events lead to the formation of a new co-integrate plasmid (pAH90) with two novel hybrid S genes, which were characterized by new target specificities (O'Sullivan *et al.* 2000). Furthermore, in a more recent study, the *L. lactis* subsp. *lactis* IL594 strain has been shown to contain seven plasmids, of which four contained either fully functional or truncated versions of Type I S genes that were proposed to be the source of specificity regions for the other *hsdS* genes (Górecki *et al.* 2011). Of note, only one of these plasmids (pIL6) contained also the *hsdR* and *hsdM* genes, which were located between *orfX* and *hsdS* (Górecki *et al.* 2011).

Phenotypes associated with phase variation in type I RM specificities

The first and most obvious phenotype associated with RM systems is the restriction of phage infection. This is generally measured as a reduction in plaque forming units after infection with heterologous methylated phage when compared to a homologous infection. An obvious advantage offered by phase variation in RM specificity is that a single enzyme can restrict multiple different target sequences across a population of cells, thereby maximising the population's defence against a phage. For instance, in *L. lactis* strain DPC721, O'Sullivan *et al.* (Sullivan *et al*., 2000) showed that the loss of two plasmids, pAH33 and pAH82, together with the formation of the novel co-integration plasmid pAH90, was characterised by novel *hsdS* specificities, which were responsible for the newly acquired bacteriophage insensitivity. In particular, the recombination event that occurred between the HsdS determinants of the two small plasmids led to the activation of a phage adsorption blocking phenotype (Ads) against phage c2 and increased restriction against the small isometric-headed phage 712 (O'Sullivan *et al.* 2000). Using the *Mycoplasma* phage P1, Dybvig *et al.* (Dybvig, Sitaraman and French 1998) determined that populations with different active *hsdS* genes also show differing phage susceptibility. By isolating 147 sub-clones from a single laboratory stock they were able to establish that there were eight individual groups. One of these groups

ScholarOne Support 1-434/964-4100

(consisting of 17 subclones) showed no RM activity at all, suggesting the locus had been inverted such that *hsdR* and *hsdM* were no longer in line with their promoter. This group of strains could be infected by P1 phage derived from any background, i.e. propagated in a population with the same *hsdS* orientation, propagated in a population with a different *hsdS* orientation, or propagated in a population with no RM activity (Dybvig, Sitaraman and French 1998). Six of the remaining groups were susceptible to phage with their own methylation profile, but were capable of restricting phage propagated in any other background. The final group showed varying susceptibility depending upon the strain that was used for phage propagation (Dybvig, Sitaraman and French 1998). The observation that non-restricting cells occurred quite so frequently led the authors to hypothesise that the maintenance of the system might not be driven solely by its utility in phage defence and that the *hsd* loci of *M. pulmonis* might have an essential function associated with pathogenesis. It is also possible that *M. pulmonis* cells with the non-restricting status are a necessary intermediary when switching between two different restricting states to prevent a newly encoded HsdRMS holoenzyme cleaving genomic DNA methylated at the old “wrong” sites resulting in suicide of a newly switched cell.

Now that phase variable type I RMs systems have been described in multiple species, and as more bacteriophage tools become available, it should be relatively straight-forward to model the potential advantage for any bacterial population in having a phase variable phage control mechanism. It has already been observed that “clustered regularly interspaced short palindromic repeats” (CRISPR) systems which provide an adaptive form of immunity (Garneau *et al.* 2010) have a substantial advantage over classical RM systems, which are quite easy for phage to circumvent. Indeed, with a classic RM system once a single phage has been able to initiate a lytic cycle (Figure 3B) and produce methylated progeny, then the whole of the bacterial population will be susceptible to further infection. In contrast, though bacterial populations with a phase variable system will also always produce low amounts of methylated phage, it would be expected that these would then be out-titrated by the resistant bacteria in the population, which carry the alternate phase variable restriction enzymes (Figure 3A). Our unpublished data in *S. pneumoniae* show that phage plaques can only be recognised in a lawn of bacterial cells if the population is composed of >70% of bacteria expressing the same HsdS variant. In populations that are more diverse, plaques are not seen indicating that a significant portion of the population is protected by the action of the phase variable RM system. This type of population variation-based protection seems to have an advantage over cells with the CRISPR system that when encountering a previously unknown phage has no such protection and rely upon “single” survivors to generate a new population which is resistant to re-infection (Samson *et al.* 2013).

For Peer Review

ScholarOne Support 1-434/964-4100

It is now becoming apparent that the biological significance of Type I RM systems can extend well beyond a merely defensive function; those systems encoding phase variable methylation will define multiple epigenetic states across individual bacterial cells which could alter the transcriptome potentially leading to global phenotypic changes. So far, three phenotypes have been associated with changes in methylation caused by phase variable Type I RM systems. In *S. pneumoniae* "phase-locked" mutants, which are incapable of *hsdS* recombination and therefore can only express a single S subunit, have been constructed and used to conduct experiments in mice (Manso *et al.* 2014). These locked mutants showed differences in virulence in the experimental infection models, with the A-variant being less able to colonise the epithelium in a carriage model of infection and the B-variant being less virulent in a bacteraemia model (Manso *et al.* 2014) (for variant nomenclature see Fig. 1). The decrease in virulence in the locked B mutant strain correlated with a lower expression of the capsule operon, although no apparent mechanism for this potential epigenetic effect could be identified (Manso *et al.* 2014). When monitoring an invasive infection with a wild-type strain (containing a high proportion of the E-variant), a shift to the A-variant could be observed over time, suggesting host selection for or against certain variants (Manso *et al.* 2014). In contrast after mining the genome sequences of over 600 paired meningeal and sepsis isolates of *S. pneumoniae* for the prevalence of *hsdS*-variants expressed by each isolate, no evidence could be found for an association of any SpnIII methylation-variant with any clinical parameter (Lees *et al.* 2017). The genome work in *S. suis* has associated both phase variable Type I RM system and a pathogenicity island to serotype 2 isolates responsible for human invasive disease (Willemse and Schultsz 2016; Willemse *et al.* 2016) which led the authors to speculate that virulence in *S. suis* could be controlled by the phase variable RM system.

Two independent reports associate variation in SpnIII methylation with phenotypic differences in colony opacity of *S. pneumoniae* strains (Manso *et al.* 2014)(Li *et al.* 2016). This phenotype is well known to be associated with colonisation (transparent) and invasive disease (opaque), though a detailed molecular mechanism for the trait is still missing (Weiser *et al.* 1994). Manso *et al.* showed that a strain locked into an *S.spnIIIA* could be classified as opaque, while those locked into *S.spnIIIB* were >90% transparent and the other 4 variants showed mostly opaque colonies. More recently, Li *et al.* have also generated "phase-locked" mutants (Li *et al.* 2016) in a variety of different strain backgrounds, including some non-encapsulated strains. Data comparison between the two studies was consistent for the majority of variants, confirming that variations in SpnIII methylation do appear to play an important role in the generation of opaque and transparent colonies. An important difference between the two studies was that one reports an unequivocal association of opaque colony morphology to a single methylation variant (Li *et al.* 2016), while the other

ScholarOne Support 1-434/964-4100

For Peer Review

reports that some of the locked variants can give rise to both opaque and transparent colonies (Manso *et al.* 2014). Furthermore, as the frequency of these opaque and transparent colonies varied depending upon which SpnIII variant was expressed by the strain Manso *et al*. concluded that additional loci, possibly including a second phase variable system, must be involved in regulating colony opacity in the pneumococcus (Manso *et al.* 2014). Despite the multiple observations that change in methylation pattern results in differences in virulence phenotypes, the exact details of the epigenetic mechanism(s) behind this process has not yet been reported.

In vivo work with the murine pathogen *M. pulmonis* has been conducted by Gumulak-Smith *et al* (Gumulak-Smith *et al.* 2001). Rats were intra-nasally challenged with *M. pulmonis* and bacteria recovered from the lungs, trachea and nose were then analysed for changes in *hsdS* (and *vsa*) orientation and expression. PCR analysis showed that, overall, the *M. pulmonis* populations recovered from the trachea were more variable at both their *hsd* and *vsa* loci, when compared to those isolated from the nose. In addition to the alterations in *hsdS* recombination shown within the recovered populations, there were cells that showed no detectable restriction or methylation activity that were also detected (Gumulak-Smith *et al.* 2001). Due to the structure of the *hsd* locus it is possible for the *hsdR* and *hsdM* genes to also be inverted, which would explain the lack of RM activity in these small sub-populations of cells. This finding implies that, in addition to the obvious advantage of being able to express multiple different S subunits, there may also be conditions under which a lack of expression of the RM system would be advantageous for *M. pulmonis*.

In both *Bacteroides* and *Mycoplasma* the recombinases involved in the switching of the *hsdS* alleles also have an impact on recombination in the capsule and surface lipoprotein genes respectively (Sitaraman and Dybvig 1997; Sitaraman, Denison and Dybvig 2002; Cerdeño-Tárraga *et al.* 2005; Nakayama-Imaohji *et al.* 2009). However, there is no evidence for a direct relationship between the variation in methylation by recombination in the RM systems and the changes in virulence.

Concluding remarks

As more genomes are sequenced and as the technological advances that allow the detection and quantification of DNA methylation become more widely available it is likely that the number of studies reporting the existence of phase variable Type I restriction modification systems among different bacterial species will increase substantially over the next few years. It seems very likely that these systems have evolved as a defensive response to selective pressure from bacteriophage; however, it has now been clearly demonstrated, in a number of recent publications, that this ability to reversibly change the sites of global DNA methylation can also have significant physiological effects upon the

ScholarOne Support 1-434/964-4100

For Peer Review

bacteria. So why would multiple diverse bacterial species retain such systems? One possibility is that these alternate epigenetic modes may, as has been proposed for other phase variable systems (Moxon, Bayliss and Hood 2006), act as contingency mechanisms for the adaptation of bacterial populations to changing environments, e.g. during the switch from asymptomatic colonisation to invasive disease which occurs during pneumococcal disease. Indeed such a hypothesis is already supported by existing data that show that pneumococcal strains which are differently methylated by SpnIII exhibit differences in fitness between distinct host niches during experimental infections (Manso *et al.* 2014). The ability to switch rapidly between multiple different epigenetic states could also be beneficial as it would provide bacterial populations with the potential to quickly regain phenotypic diversity after selective or non-selective bottlenecks which typically occur during infection of a host.

There do however remain a number of challenging questions which have yet to be answered even in the relatively well described systems, such as the *Spn*III system of *Streptococcus pneumoniae.* It remains unclear how and when the *hsdS* recombination events occur: if, as it now appears, these are not all simple inversion events governed by the site specific recombinase in the locus, then what other bacterial proteins are involved, and is there perhaps a requirement for genome replication to provide a second copy of the locus to aid in such recombination events? There is also the dilemma that occurs immediately after the active *hsdS* allele has changed and where now any newly translated HsdRMS restriction enzyme will target a DNA sequence different from the one that was recognised by the previous protective methyltransferase: how exactly does the cell avoid cutting up its own DNA before it can completely re-methylate itself? Previous research work had already clearly shown that epigenetic modifications had the ability to affect gene expression and alter important and complex bacterial phenotypes, such as replication or virulence; however, the discovery that global DNA methylation can also be subject to phase variation has offered an exciting new area of investigation elucidating how many bacterial species have exploited this opportunity.

For Peer Review

ScholarOne Support 1-434/964-4100

Funding

This work was supported in part by the Medical Research Council [Grant Number MR/M003078/1] and by the Biotechnology and Biological Sciences Research Council [Grant Number BB/N002903/1]. JR was funded by a BBSRC KTN CASE studentship [Grant Number BB/P504737/1]. NJC was funded by a Sir Henry Dale fellowship, jointly funded by the Wellcome Trust and Royal Society [Grant Number 104169/Z/14/Z].

For Peer Review

Figure legends

Figure 1. Type I RM protein complexes and schematic maps of two phase-variable type I RM loci with inverted *hsdS* genes. The Type I enzyme subunits can form two different complexes: the RTase and the MTase. The RTase is made up by a pentameric complex (panel A) that mediates endonucleolytic cleavage (restriction) and is composed of two restriction subunits (R, light grey), two methylation subunits (M, dark grey) and a specificity subunit (S, green). The black arrows indicate DNA helicase activity, whereas the two irregular parallel lines represent the DNA filament. The MTase is composed by a trimeric complex (panel B) that is responsible for methylation and comprises only two methylation subunits (M, dark grey) and a specificity subunit (S, green). The *ivr* locus encoding the SpnD39III system of *S. pneumoniae* D39 (Manso *et al.* 2014) is shown in panel C and the phase variable locus of a *L. monocytogenes* clonal complex 8 strain (Fagerlund *et al.* 2016) in panel D. The pneumococcal locus includes *hsdR* (SPD_0455), *hsdM* (SPD_0454) and *hsdS* (SPD_0453), a CreX recombinase (phage-type integrase also named IvrR) (SPD_0452), a truncated *hsdS2* gene (SPD_0450) with only one variable target recognition domain (TRD) and a further truncated *hsdS3* gene (SPD_0451) with two variable TRDs. There are three series of inverted repeats that allow allele switching and formation of six different *hsdS* alleles (named from A to F and shown below the locus map). Inverted repeats are of 85 bp (diagonal lines), 333 bp (checked) and 15 bp (horizontal lines), respectively. According to their sequence, TRDs are coloured in grey, white, red, black and blue. Black and dotted lines indicate the possibility of recombination occurring at the level of the inverted repeats. The insert represents an illustration of *hsdS* allele switching and the formation of the six different locus arrangements. The *L. monocytogenes* locus (B) of the strain R479a (Fagerlund *et al.* 2016), as representative of CC8 strains, includes *hsdR* (LMR479A_0528), *hsdM* (LMR479A_0529) and *hsdS* (LMR479A_0530), a phage-type integrase (LMR479A_0531), a truncated *hsdS2* gene (LMR479A_0532) with two variable TRDs. There are two series of inverted repeats that allow allele switching and formation of four different *hsdS* alleles (named from A to D and shown below the locus map). The four possible *hsdS* A-D alleles are represented with their two TRDs coloured in either grey, white, red or black, depending on their amino acid sequence. Inverted repeats are of 24 bp and 155 bp (diagonal and vertical lines, respectively). Black and dotted lines indicate the possibility of recombination occurring at the level of the inverted repeats.

Figure 2. Schematic maps of the phase-variable *hsd* loci in distinct strains of different bacterial species. The strains have been divided according to the presence of either

ScholarOne Support 1-434/964-4100

For Peer Review

inverted (A) or direct *hsdS* repeat (B) sequences. Consistent with the nomenclature of REBASE, in all panels the actively transcribed *hsdS* gene is reported as *hsdS1*, while the untranscribed *hsdS* genes avavilable for recombination are marked *hsdS2* and *hsdS3*. The name of each strain is reported on top of each locus illustration. Green arrows show the *hsdS* genes, dark grey and white arrows correspond to *hsdM* and *hsdR* genes respectively, whereas light grey indicates the *recombinase* genes (REC) or genes encoding for hypothetical proteins. Listed from the top to the bottom the strains and genes reported in each phase variable locus are: **A)** *Mycoplasma pulmonis* UAB CTIP (Genbank accession NC_002771), *hsd1* (MYPU_RS02160), *hsdR* (MYPU_RS02165) *hsdM* (MYPU_RS02170), *hsdS2* (MYPU_RS02175); *Bacteroides fragilis* NCTC 9343 (GenBank GCA_000025985.1), *hsdR* (BF9343_RS08540-), hypothetical protein (BF9343_RS08545), *hsdM* (BF9343_RS08550), *hsd1* (BF9343_RS21605), *hsd2* (BF9343_RS08560), *hsd3* (BF9343_RS08565), *hsd4* (BF9343_RS08570); *Streptococcus pneumoniae* D39 (Genbank NC_008533), *hsdS''* (SPD_0451), *hsdS'* (SPD_0450), a *Cre* recombinase (SPD_0452), *hsdS* (SPD_0453), *hsdM* (SPD_0454) and *hsdR* (locus_tag SPD_0455); *Enterococcus faecalis* ATCC 19433 strain (Genbank ASDA01000009), *hsdS2* (WMC_02595), *integrase* (WMC_02596), hsdS1 (WMC_02597), *hsdM* (WMC_02598) and *hsdR* (WMC_02599); *Listeria monocytogenes* R479a (Genbank NZ_HG813247), *hsdR* (LMR479A_0528), *hsdM* (LMR479A_0529), *hsdS1* (LMR479A_0530), *integrase* (LMR479A_0531), *hsdS2* (LMR479A_0532); *Streptococcus suis* P1/7 (Genbank NC_012925), *hsdS1* (SSU_RS06425), *hsdS2* (SSU_RS06430), *hsdM* (SSU_RS06435), *hsdR* (SSU_RS06440); *Lactobacillus plantarum* WCFS1 (Genbank NC_004567), *hsdR* (lp_0938), *hsdM* (lp_0939), *hsd1* (lp_0940), *integrase* (lp_0941), *hsd2* (lp_0942), *hsd3* (lp_0943). B) The schematic maps of two systems showing direct repeat *hsdS* loci (tvr) are for *Streptococcus pneumoniae* TIGR4 (Genbank NC_003028), *hsdM* (SP_0886), *hsdS1* (SP_0887), hypothetical proteins (SP_0888, SP_0889), *integrase* (SP_0890), *hsd2* (SP_0891), *hsdR* (SP_0892) and *Clostridium botulinum* B Eklund 17B (Genbank NC_010674), *hsdS1* (CLL_A2080), *hsdM* (CLL_A2081), *hsdS2* (CLL_A2082), *hsdR* (CLL_A2083).

Figure 3. Schematic model of the effects of Restriction Modification (RM) systems upon phage infection. This figure depicts bacterial cells with different RM systems being infected by phage. Each DNA methylation state is represented by a different colour. The black irregular line represents the bacterial genome and the circled M indicates its methylation status. The phage is represented by the hexagonally shaped figure, with its tail anchoring to the bacterial cell wall to penetrate it and cause infection. When an appropriately methylated phage infects a bacterial cell, it is able to replicate effectively and to produce progeny with the same methylation status. **A)** Inhibition of phage infection by stable RM

ScholarOne Support 1-434/964-4100

systems. Depending upon the strength of restriction, a minority of the phage population may be able to avoid restriction and infect a bacterial cell which has a different methylation pattern to the source of the virus (one out of three “red” cells failing to restrict “blue” phage in this example). In this case the progeny phage, now appropriately methylated, will be able to efficiently infect neighbouring bacterial cells and will perpetuate the infection (productive infection). **B)** Inhibition of phage infection by phase-variable RM systems. In a bacterial population with an RM system capable of producing several distinctly different methylation patterns phage will efficiently infect neighbouring cells which have the same methylation pattern as themselves, and will be restricted by the differently methylated cells. Infection may quickly reach a dead end when progeny phages are differently methylated to all of the remaining surrounding cells which can therefore restrict phage spread.

Figure 4. Taxonomy cladogram showing the presence and the divergence of the IvrR recombinase in distinct bacterial families and species. This NCBI taxonomy cladogram includes one leaf node for each species in which a protein aligned with the *S. pneumoniae* TIGR4 IvrR recombinase of the *Spn*III system (orthologous with SPD_0452 in Fig 1) with an E value of 0.001 or less, and is based upon a BLASTP search of the non-redundant sequence database. This is the background upon which those species containing an alignment matching IvrR with an E value of 10^{-100} or less are highlighted as being likely candidates for containing phase variable type I RM loci. Grey clades contain no such highly significant hits. Clades containing significant hits are coloured and labelled with the appropriate taxon name; within these are shown representative annotated species which contain significant hits.

For Peer Review

ScholarOne Support 1-434/964-4100

References

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

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

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

Atack JM. A biphasic epigenetic switch controls immunoevasion, virulence and niche adaptation in non-typeable Haemophilus influenza. *Nat Commun* 2015;**6**:7828.

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

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

Blow MJ, Clark TA, Daum CG *et al.* The Epigenomic Landscape of Prokaryotes. *PLoS Genet* 2016, DOI: 10.1371/journal.pgen.1005854.

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

Cerdeño-Tárraga AM, Patrick S, Crossman LC *et al.* Extensive DNA inversions in the B. fragilis genome control variable gene expression. *Science* 2005;**307**:1463–5.

Chambaud I, Heilig R, Ferris S *et al.* The complete genome sequence of the murine respiratory pathogen Mycoplasma pulmonis. *Nucleic Acids Res* 2001;**29**:2145–53.

Claesson MJ, Li Y, Leahy S *et al.* Multireplicon genome architecture of Lactobacillus salivarius. *Proc Natl Acad Sci U S A* 2006;**103**:6718–23.

Clark TA, Murray IA, Morgan RD *et al.* Characterization of DNA methyltransferase specificities using single-molecule, real-time DNA sequencing. *Nucleic Acids Res* 2012;**40**, DOI: 10.1093/nar/gkr1146.

Croucher NJ, Coupland PG, Stevenson AE *et al.* Diversification of bacterial genome content through distinct mechansims over different timescales. *Nat Commun* 2014;**5**:1–12.

Croucher NJ, Harris SR, Fraser C *et al.* Rapid pneumococcal evolution in response to clinical interventions. 2011;**331**:430–4.

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

ScholarOne Support 1-434/964-4100

For Peer Review

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

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

Fagerlund A, Langsrud S, Schirmer BCT *et al.* Genome analysis of Listeria monocytogenes sequence type 8 strains persisting in salmon and poultry processing environments and comparison with related strains. *PLoS One* 2016;**11**:1–22.

Feng Z, Li J, Zhang JR *et al.* Qdnamod: A statistical model-based tool to reveal intercellular heterogeneity of DNA modification from SMRT sequencing data. *Nucleic Acids Res* 2014;**42**:13488–99.

Furuta Y, Abe K, Kobayashi I. Genome comparison and context analysis reveals putative mobile forms of restriction-modification systems and related rearrangements. *Nucleic Acids Res* 2010;**38**:2428–43.

Furuta Y, Kawai M, Uchiyama I *et al.* Domain movement within a gene: A novel evolutionary mechanism for protein diversification. *PLoS One* 2011;**6**, DOI: 10.1371/journal.pone.0018819.

Furuta Y, Namba-Fukuyo H, Shibata TF *et al.* Methylome Diversification through Changes in DNA Methyltransferase Sequence Specificity. *PLoS Genet* 2014;**10**, DOI: 10.1371/journal.pgen.1004272.

Garneau JE, Dupuis ME, Villion M *et al.* The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* 2010;**468**:67–71.

Górecki RK, Koryszewska-Bagińska A, Gołebiewski M *et al.* Adaptative potential of the lactococcus lactis IL594 strain encoded in its 7 plasmids. *PLoS One* 2011;**6**, DOI: 10.1371/journal.pone.0022238.

Gumulak-Smith J, Teachman A, Tu AHT *et al.* Variations in the surface proteins and restriction enzyme systems of Mycoplasma pulmonis in the respiratory tract of infected rats. *Mol Microbiol* 2001;**40**:1037–44.

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

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

Kan N, Lautenberger J, Edgell M *et al.* The nucleotide sequence recognized by the Escherichia coli K12 restriction and modification enzymes. *J Mol Biol* 1979;**130**:191–209.

Kilian M, Riley DR, Jensen A *et al.* Parallel evolution of Streptococcus pneumoniae and

ScholarOne Support 1-434/964-4100

For Peer Review

Streptococcus mitis to pathogenic and mutualistic lifestyles. *MBio* 2014;**5**:e01490–e01414.

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

Lees J, Kremer PHC, Manso AS *et al.* Large scale genomic analysis shows no evidence for repeated pathogen adaptation during the invasive phase of bacterial meningitis in humans. *Microb Genomics* 2017;**3**, DOI: doi 10.1099/mgen.0.000103.

Li J, Li JW, Feng Z *et al.* Epigenetic Switch Driven by DNA Inversions Dictates Phase Variation in Streptococcus pneumoniae. *PLoS Pathog* 2016;**12**:1–36.

Loenen WAM, Dryden DTF, Raleigh EA. *et al.* Type I restriction enzymes and their relatives. *Nucleic Acids Res* 2014a;**42**:20–44.

Loenen WAM, Dryden DTF, Raleigh EA. *et al.* Highlights of the DNA cutters: A short history of the restriction enzymes. *Nucleic Acids Res* 2014b;**42**:3–19.

Loenen WAM, Raleigh EA. The other face of restriction: Modification-dependent enzymes. *Nucleic Acids Res* 2014;**42**:56–69.

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

Manso AS, Chai MH, Atack JM *et al.* A random six-phase switch regulates pneumococcal virulence via global epigenetic changes. *Nat Commun* 2014;**5**:5055.

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

Morgan RD, Luyten YA, Johnson SA *et al.* Novel m4C modification in type I restriction-modification systems. *Nucleic Acids Res* 2016;**44**:9413–25.

Mostowy R, Croucher NJ, Hanage WP *et al.* Heterogeneity in the Frequency and Characteristics of Homologous Recombination in Pneumococcal Evolution. *PLoS Genet* 2014;**10**:1–15.

Moxon R, Bayliss C, & Hood D. Bacterial contingency loci: the role of simple sequence DNA repeats in bacterial adaptation. *Annu. Rev. Genet.* 2006;40:307–33.

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

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

O'Sullivan D, Twomey DP, Coffey A *et al.* Novel type I restriction specificities through domain shuffling of HsdS subunits in Lactococcus lactis. *Mol Microbiol* 2000;**36**:866–75.

ScholarOne Support 1-434/964-4100

Pingoud A, Jeltsch A. Structure and function of type II restriction endonucleases. *Nucleic Acids Res* 2001;**29**:3705–27.

Roberts RJ, Vincze T, Posfai J *et al.* REBASE—a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Res* 2015;**43**, DOI: 10.1093/nar/gku1046.

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

Samson JE, Magadán AH, Mourad S *et al.* Revenge of the phages: defeating bacterial defences. *Nat Rev Microbiol* 2013;**11**:675–87.

Schouler C, Clier F, Lerayer AL *et al.* A type IC restriction-modification system in Lactococcus lactis. *J Bacteriol* 1998a;**180**:407–11.

Schouler C, Gautier M, Ehrlich SD *et al.* Combinational variation of restriction modification specificities in Lactococcus lactis. *Mol Microbiol* 1998b;**28**:169–78.

Seegers JFML, Van Sinderen D, Fitzgerald GF. Molecular characterization of the lactococcal plasmid pCIS3: Natural stacking of specificity subunits of a type I restriction/modification system in a single lactococcal strain. *Microbiology* 2000;**146**:435–43.

Seib KL, Jen FEC, Tan A *et al.* Specificity of the ModA11, ModA12 and ModD1 epigenetic regulator N^6-adenine DNA methyltransferases of Neisseria meningitidis. *Nucleic Acids Res* 2015;**43**:4150–62.

Shen X, Gumulak J, Yu H *et al.* Gene rearrangements in the vsa locus of Mycoplasma pulmonis. *J Bacteriol* 2000;**182**:2900–8.

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

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

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

Srikhanta YN, Gorrell RJ, Steen JA. *et al.* Phasevarion mediated epigenetic gene regulation in Helicobacter pylori. *PLoS One* 2011;**6**:1–9.

Tettelin H, Nelson KE, Paulsen IT *et al.* Complete Genome Sequence of a Virulent Isolate of Streptococcus pneumoniae. *Science* 2001;**293**:498-506.

Vasu K, Nagaraja V. Diverse Functions of Restriction-Modification Systems in Addition to Cellular Defense. *Microbiol Mol Biol Rev* 2013;**77**:53–72.

Weiser JN, Austrian R, Sreenivasan PK *et al.* Phase variation in pneumococcal opacity:

ScholarOne Support 1-434/964-4100

For Peer Review

Relationship between colonial morphology and nasopharyngeal colonization. *Infect Immun* 1994;**62**:2582–9.

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

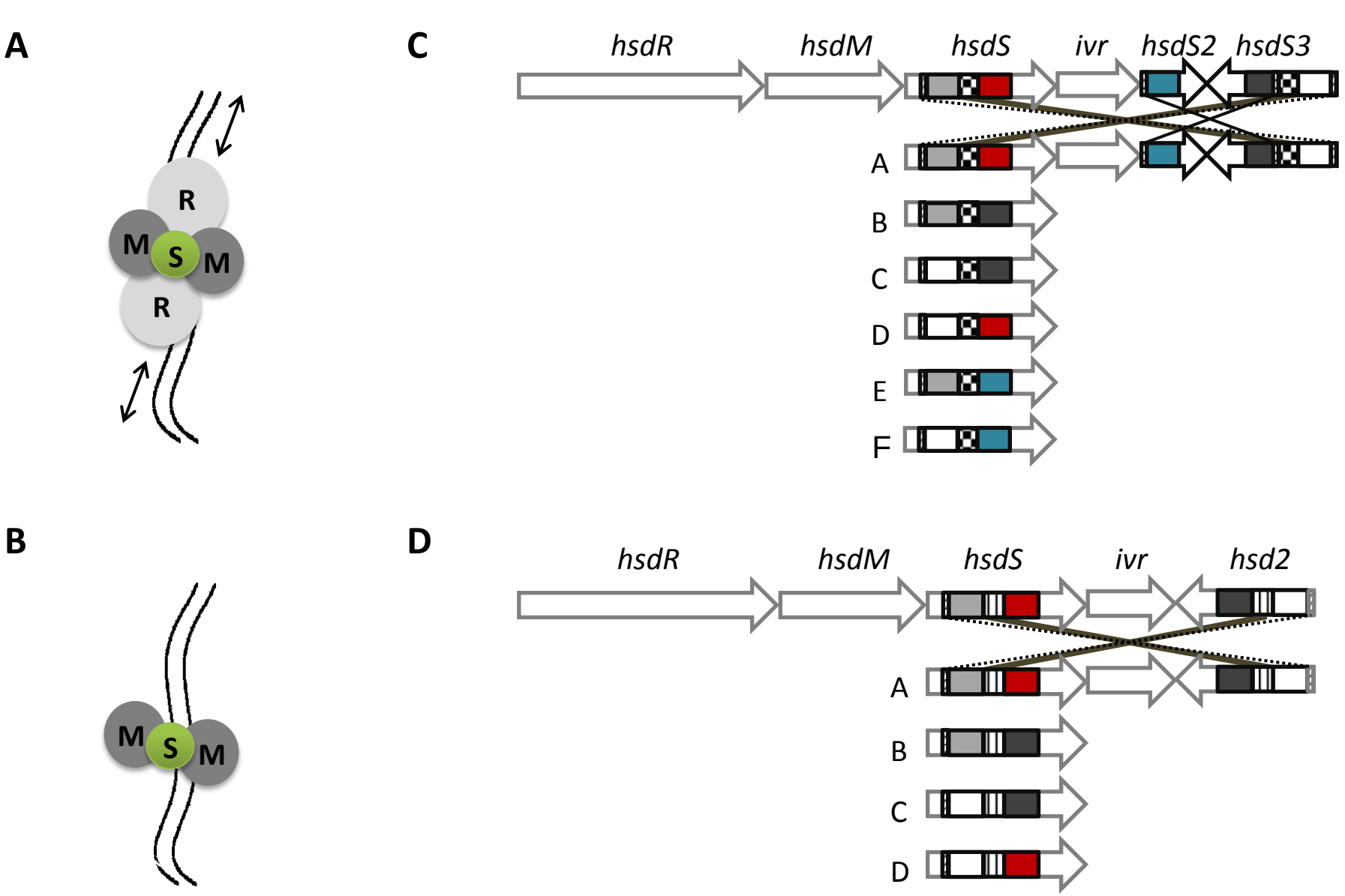
Willemse N, Howell KJ, Weinert LA *et al.* An emerging zoonotic clone in the Netherlands provides clues to virulence and zoonotic potential of Streptococcus suis. *Sci Rep* 2016;**6**:28984.

Willemse N, Schultsz C. Distribution of Type I Restriction–Modification Systems in Streptococcus suis: An Outlook. *Pathogens* 2016;**5**:62.

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

For Peer Review

ScholarOne Support 1-434/964-4100


A
R
M
S
M
R
B
M
S
M
C
hsdR
hsdM
hsdS
ivr
hsdS2
hsdS3
A
B
C
D
E
F
D
hsdR
hsdM
hsdS
ivr
hsd2
A
B
C
D

A

***Mycoplasma pulmonis* UAB CITP**



***Bacteroides fragilis* NCTC 9343**



***Streptococcus pneumoniae D39 Spn*III**



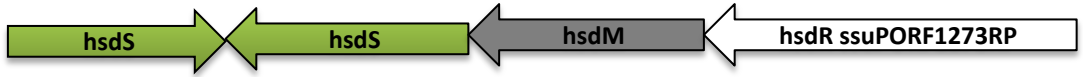
***Enterococcus faecalis* ATCC19433**



***Listeria monocytogenes* L312**



***Streptococcus suis* P1/7**

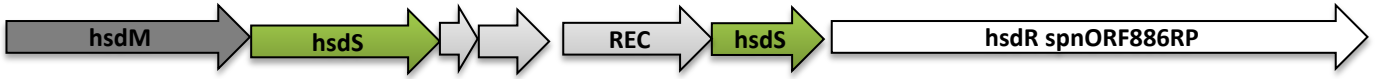


***Lactobacillus plantarum* WCFS1**

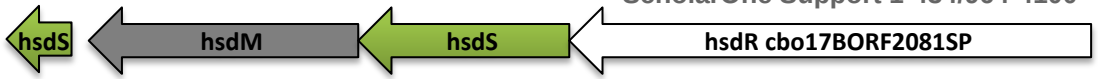


B

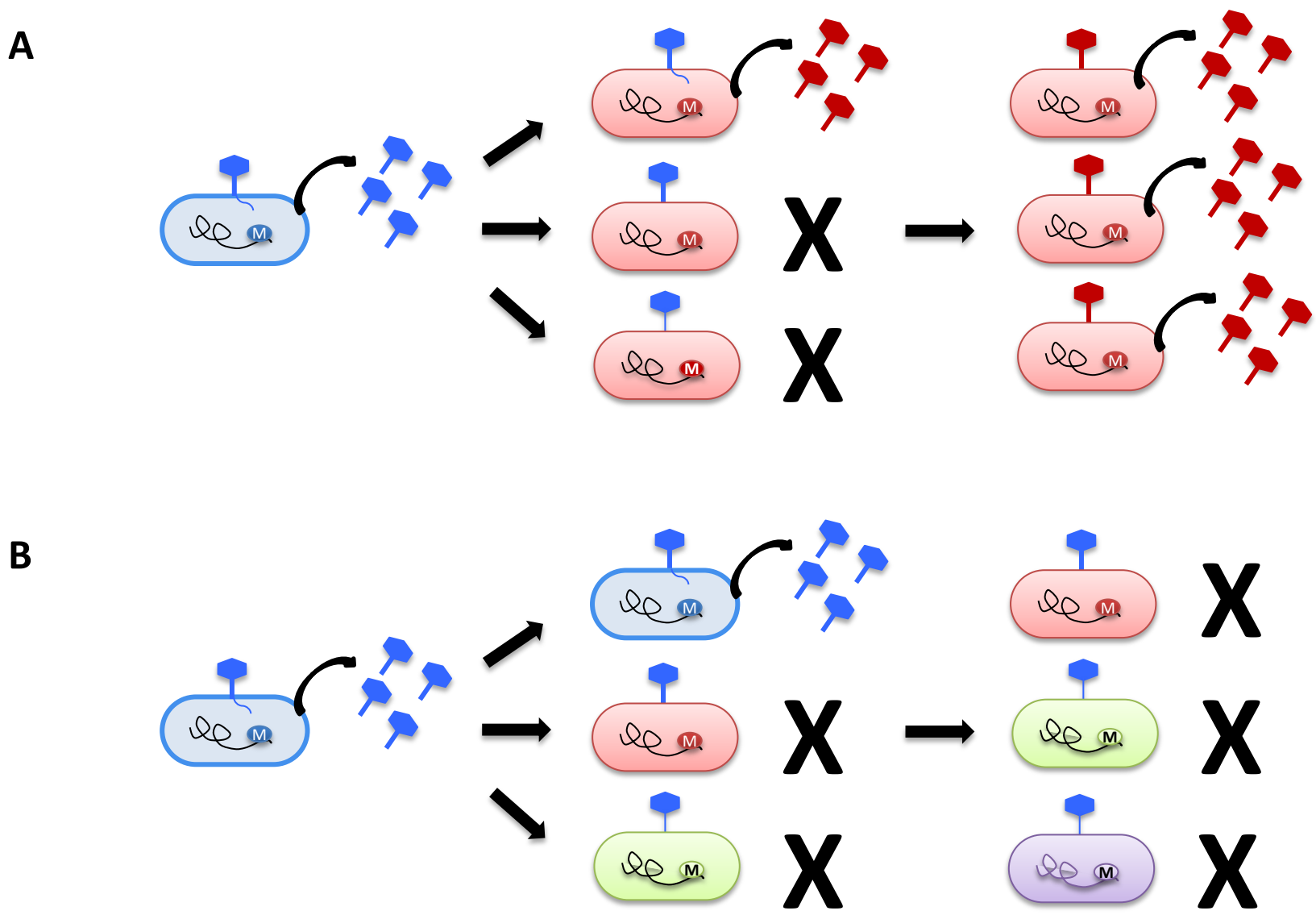
***Streptococcus pneumoniae* TIGR4 *Spn*IV**

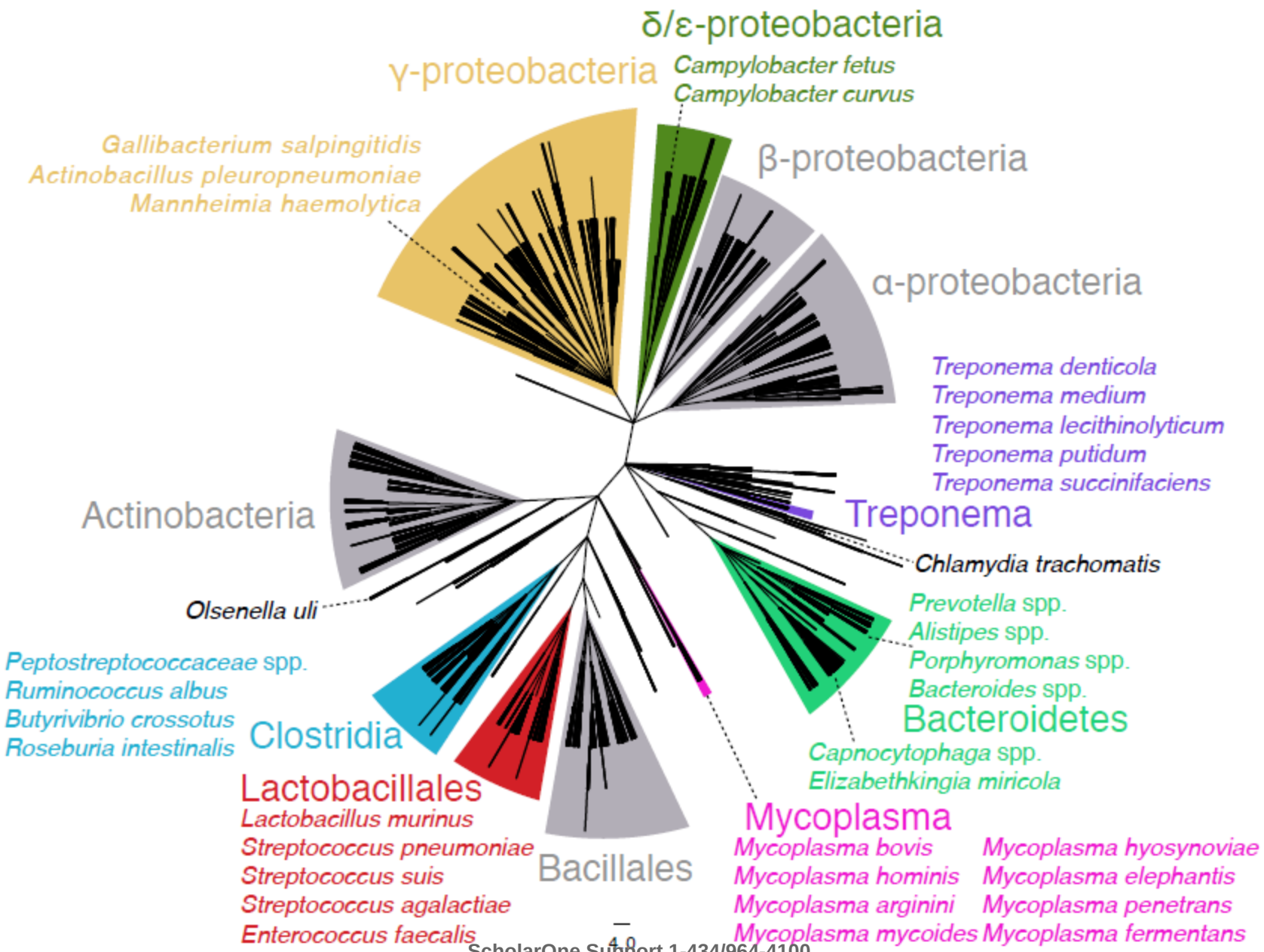


***Clostridium botulinum* B Eklund 17B**



ScholarOne Support 1-434/964-4100


A
M
X
X
B
M
X
X
X


δ/ε-proteobacteria
γ-proteobacteria
Campylobacter fetus
Campylobacter curvus
Gallibacterium salpingitidis
Actinobacillus pleuropneumoniae
Mannheimia haemolytica
β-proteobacteria
α-proteobacteria
Treponema denticola
Treponema medium
Treponema lecithinolyticum
Treponema putidum
Treponema succinifaciens
Treponema
Actinobacteria
Chlamydia trachomatis
Olsenella uli
Prevotella spp.
Alistipes spp.
Porphyromonas spp.
Bacteroides spp.
Bacteroidetes
Peptostreptococcaceae spp.
Ruminococcus albus
Butyrivibrio crossotus
Roseburia intestinalis
Clostridia
Capnocytophaga spp.
Elizabethkingia miricola
Lactobacillales
Lactobacillus murinus
Streptococcus pneumoniae
Streptococcus suis
Streptococcus agalactiae
Enterococcus faecalis
Mycoplasma
Mycoplasma bovis
Mycoplasma hominis
Mycoplasma arginini
Mycoplasma mycoides
Mycoplasma hyosynoviae
Mycoplasma elephantis
Mycoplasma penetrans
Mycoplasma fermentans
Bacillales
1.0