Mechanisms of DNA Modification-Dependent Regulation in Gram-Positive Bacteria

by

Taylor M. Nye

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Molecular, Cellular, and Developmental Biology)
in The University of Michigan
2020

Doctoral Committee:

       Professor Lyle A. Simmons, Chair
       Professor Robert A. Bender
       Professor Matthew R. Chapman
       Assistant Professor Peter L. Freddolino
       Professor Michele S. Swanson

Taylor M. Nye

tnye@umich.edu

ORCID iD:0000-0003-4695-0485

DEDICATION

To Lola and Kevin,

I love you 3000. Always.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE

LIST OF TABLES

ABSTRACT

The presence of DNA modifications is pervasive among both prokaryotic and eukaryotic species. In bacteria, the study of DNA methylation has largely been in the context of restriction-modification systems, where DNA methylation serves to safeguard the chromosome against restriction endonucleases that are intended to cleave invading foreign DNA. There has been a growing recognition that the methyltransferase component of restriction-modification systems can also function in the regulation of gene expression. Outside of restriction modification systems, DNA methylation from orphan methyltransferases, which lack cognate restriction endonucleases, have been shown to regulate critical cellular processes. The majority of research articles focuses on the epigenetic regulatory roles of bacterial DNA methylation in the context of Gram-negative bacteria, with particular bias towards *Escherichia coli*, *Caulobacter crescentus*, and related Proteobacteria. Despite the critical functions of DNA methylation in Gram-negative bacteria, far less is known about how DNA methylation contributes to epigenetic regulation of gene expression in Gram-positive bacteria. In this thesis I investigated the effects of DNA modifications in Gram-positive bacteria. I showed that DNA methylation from an active Type I restriction-modification system in *Streptococcus pyogenes* also functions in the epigenetic regulation of a small subset of virulence genes, all of which are significantly down regulated in the absence of DNA methylation. Moreover, I showed that the methylation-dependent decrease in gene expression results in attenuated virulence of an *S. pyogenes* clinical isolate*,* implicating DNA methylation as an important contributor to *S. pyogenes* pathogenesis. I also characterized the methylomes for two strains of the Gram-positive Firmicute *Bacillus subtilis* and demonstrated that DNA methylation regulates the expression of a small subset of genes involved in chromosome structure and maintenance. I further identified a methylation-sensitive transcriptional regulator, providing some of the first insight into the mechanisms of methylation-dependent gene regulation in Gram-positive bacteria. Finally, I identified a previously uncharacterized gene, *rnhP*, which is a plasmid encoded

RNase HI. I found that RnhP contributes to genome maintenance in *B. subtilis* NCIB 3610 by removing RNA-DNA hybrids with four or more ribonucleotides embedded in DNA. I showed that RnhP does not contribute to plasmid maintenance or hyper-replication. Importantly, I showed that RnhP contributes to genome maintenance by allowing DNA replication forks to progress through the terminus region. Together, my work highlights the importance of DNA modifications and noncanonical nucleotides in Gram-positive bacteria and provides a framework for future studies of epigenetic regulation by RM systems in bacterial pathogenesis and development.

CHAPTER I

A Positive Perspective on DNA Methylation: Regulatory Functions of DNA Methylation Outside of Host Defense in Gram-positive Bacteria

Abstract

The presence of post-replicative DNA methylation is pervasive among both prokaryotic and eukaryotic species. In bacteria, the study of DNA methylation has largely been in the context of restriction-modification systems, where DNA methylation serves to safeguard the chromosome against restriction endonuclease cleavage of invading foreign DNA. There has been a growing recognition that the methyltransferase component of restriction-modification systems can also regulate gene expression, with important contributions to virulence factor gene expression in bacterial pathogens. Outside of restriction-modification systems, DNA methylation from orphan methyltransferases, which lack cognate restriction endonucleases, has been shown to regulate critical processes, including origin sequestration, DNA mismatch repair, and the regulation of gene expression. The majority of research and review articles focuses on the epigenetic regulatory roles of bacterial DNA methylation in the context of Gram-negative bacteria, with emphasis towards *Escherichia coli*, *Caulobacter crescentus*, and related Proteobacteria. Here we summarize the epigenetic functions of DNA methylation outside of host defense in Gram-positive bacteria, with a focus on the regulatory effects of both phase variable methyltransferases and traditional restriction-modification system DNA methyltransferases.

Introduction

The occurrence of genomic DNA methylation is ubiquitous across all three domains of life, where modification events function in diverse and critical cellular processes. In eukaryotes, the predominant type of DNA methylation is 5-methylcytosine (m5C) and the presence of these modifications is necessary for the regulation of gene expression and development (16; 39). In humans, aberrant DNA methylation events are implicated in numerous disease states, including cancer (39; 81; 99). In addition to m5C, the genomes of bacteria are known to include N4-methycytosine (m4C) and N6-methyladenine (m6A) modifications [(10) and references there in]. A recent survey of prokaryotic genomes demonstrates the widespread occurrence of m5C, m4C, and m6A, where at least one type of modification was detected in 93% of the ~230 genomes analyzed (10). For all of the prokaryotes included in the study, DNA methylation was detected using Pacific Biosciences (PacBio) Single-Molecule Real-Time (SMRT) sequencing platform (25). PacBio SMRT sequencing uses inferences from DNA polymerase kinetics during sequencing reactions to detect the presence of DNA base modifications without *a priori* knowledge of the presence of genomic methylation or the sequence contexts in which modifications occur (25). In the survey, 75% of the modifications detected were m6A, which is likely an overrepresentation of m6A relative to cytosine methylation because PacBio SMRT sequencing is more robust for detection of m6A and m4C modifications but is not well suited for the detection of m5C modifications (10; 25). In addition to the Blow *et al.* study, New England Biolabs (NEB) maintains a free database, REBASE, that serves as a repository for bacterial genome methylomics results as well as information about predicted MTases, REases, and their recognition sites (http://rebase.neb.com). This resource is available to scientists interested in understanding if DNA methylation is detected or predicted in a genome of interest.

The importance of DNA methylation in bacterial genomes can also be highlighted by the diverse processes in which they function, including protection from the invasion of foreign DNA (49; 50), phase variation (4; 33), the regulation of DNA replication (31; 64), strand discrimination during DNA mismatch repair (5), and the regulation of gene

expression (13). The majority of the methylation-dependent processes listed above have been extensively studied and reviewed for Gram-negative bacteria (1; 55; 59; 76). This bias in study towards Gram-negative bacteria is reflected in the organisms included in the survey of prokaryotic DNA methylation, where 57% of the prokaryotes included were Gram-negative organisms, 33% were Gram-positive, and 10% were undefined or belonged to the domain Archaea (10) **(Fig 1.1)**. Gram-positive bacteria include members of the high GC content phylum Actinobacteria and the low GC content Firmicutes, accounting for 6.6% and 26.3% of surveyed genomes, respectively (10) **(Fig 1.1)**. Actinobacteria include the genus *Streptomyces*, which are responsible for the production of two thirds of clinically relevant antibiotics, while Firmicutes include several important human pathogens from the genera *Staphylococcus, Streptococcus, Enterococcus,* and *Clostridia.* Despite the importance of Gram-positive bacteria to human health and industry, the functions of DNA methylation outside of host defense have been understudied **(Fig 1.2)**. Here we summarize the current knowledge of the presence and known biological functions of DNA methylation in Gram-positive bacteria with the goal of opening new and important areas of study within this important field.

DNA Methyltransferases: Origins in orphan methyltransferases and host defense systems. Enzymes called DNA methyltransferases (MTases) catalyze post-replicative modifications in DNA by transferring a methyl group from the donor S-adenosylmethionine (SAM) to adenine or cytosine bases in DNA (40). DNA MTases can function as part of a host defense system, such as the well-studied restriction-modification (RM) systems and the newly discovered bacterial exclusion (BREX) systems, or as stand-alone "orphan" MTases **(Fig 1.3)**. RM systems are minimally comprised of an MTase component and a restriction endonuclease (REase) partner. RM systems are hypothesized to predominately function as bacterial defense systems against the invasion of foreign DNA, however they have also been shown to function in phase variation and the regulation of gene expression(23). Similar to RM systems, BREX systems also function as bacterial defense systems and use DNA methylation to distinguish between self and foreign DNA (8; 28). However, as opposed to the cleavage of foreign DNA observed in RM systems, BREX systems function by blocking replication

of phage DNA (8; 28). Orphan MTases, as the name suggests, only have MTase activity and orphan MTases contribute to a variety of DNA processes, including DNA mismatch repair, origin sequestration, and the regulation of gene expression with the majority of orphan MTase characterization occurring in Gram-negative bacteria (1; 55; 59; 76).

Regulatory functions of methylation from orphan MTases. Orphan Mtases are hypothesized to be the products of RM systems that have lost their REase component (79) **(Fig 1.3)**. The most well studied orphan MTases are Dam and CcrM from Gram-negative *Escherichia coli* and *Caulobacter crescentus*, respectively. Dam methylates G**A**TC sites throughout the *E. coli* genome and functions in origin sequestration, strand discrimination during DNA mismatch repair, and the regulation of gene expression (1; 54; 76). CcrM methylates G**A**NTC sites and regulates cell cycle progression in *C. crescentus* (53; 59). While CcrM homologs are only conserved through α-Proteobacteria, Dam homologs are conserved throughout Proteobacteria and even occur in several strains of Gram-positive bacteria (54; 59). Notably, in many Gram-positive systems the Dam homolog is typically paired with a cognate endonuclease as part of an active Type II RM system as in *Streptococcus mutans*, a dental pathogen (6).

The Blow *et al.* survey of DNA methylation in prokaryotic genomes identified 165 candidate orphan MTases, a subset of which were identified in the Gram-positive genera *Clostridia*, *Nocardia,* and *Arthrobacter*. In agreement with previous studies, the authors found that orphan MTases tend to be far more conserved than MTases that belong to an RM system, with 57% and 9% conservation at the genus level, respectively (10; 79). A candidate orphan MTase from two *Arthrobacter* species, which are Gram-positive bacteria belonging to the Actinobacteria phylum, was also conserved in 93% (39/42) of the available *Arthrobacter* genome sequences for which PacBio SMRT sequencing data does not exist. The strong conservation of this orphan MTase in Arthrobacter highlights the potential biological significance to this genus (10). For both *Arthrobacter* and *Nocardia* species, another Gram-positive Actinobacteria, the recognition site for the candidate orphan MTase was enriched in the putative origin of

replication (10). Recognition sites for Dam MTase are also enriched in the *E. coli* origin, where they function in origin sequestration during DNA replication initiation, suggesting that the orphan MTases from *Arthrobacter* and *Nocardia* species may also contribute to the regulation of origin firing (31; 64).

It is worth noting that both *Arthrobacter* and *Nocardia* species also have conserved unmethylated recognition sites upstream of putative transcriptional regulators. In *Nocardia*, unmethylated recognition motifs from the orphan MTase are enriched up to 20-fold in regions upstream of transcriptional regulators (10). In *E. coli*, although the majority (99.9%) of Dam recognition sites are fully methylated, there is a small subset of unmethylated sites on both strands of DNA that have important functions in gene regulation (10; 33; 94). The presence and conservation of unmethylated motifs suggests that the orphan MTases from *Arthrobacter* and *Nocardia* may also function in the regulation of gene expression.

A Type II RM system MTase lacking a cognate endonuclease has also been identified across 36 clinical isolates of the Gram-positive pathogen *Clostridioides difficile*. Oliveira *et al.* identified the CamA MTase, which methylates CAAAA**A** motifs at an average of 7,721 sites across *Clostridioides difficile* genomes (67). Unlike the enrichment of recognition motifs for the putative orphan MTase observed in the origin of *Arthrobacter* species, CamA recognition motifs were not enriched in the origin but were present upstream of genes involved in transcriptional regulation, cell wall protein production, membrane transport, and sporulation (67). Consistent with a regulatory role for CamA-dependent methylation, deletion of *camA* resulted in global transcriptome changes and defects in both sporulation and in animal models for colonization and infection (67). It is worth noting that, unlike the conservation of the putative orphan MTases across the genera *Arthrobacter* and *Nocardia*, CamA is not well conserved across *Clostridiales* and is instead fairly unique to *C. difficile* (67). As a direct role in host defense has not been tested we cannot exclude the possibility that CamA functions both as part of a host defense system and in the regulation of gene expression. Given the important roles of orphan MTases in Gram-negative bacteria, and the conservation of orphan MTases in

Gram-positive bacteria, more studies are needed to understand the contributions of orphan MTases to the regulation of gene expression and chromosome dynamics in Gram-positive bacteria.

DNA methylation from BREX defense systems. Relative to the study of orphan MTases and RM systems, the discovery of the BREX family of defense systems is new. The term BREX (bacterial exclusion) was coined in a 2015 paper characterizing the system from *Bacillus cereus*, a Gram-positive Firmicute (8; 28)*.* BREX systems were identified based on conservation of a putative alkaline phosphatase gene, *plgZ*, which is commonly found on genomic defense islands surrounded by 4-8 conserved BREX systems genes (28). The majority of putative systems identified contain six genes, which include *pglZ*, the putative alkaline phosphatase, *plgX*, which contains a methyltransferase domain, a gene encoding a Lon-like protease domain, a putative RNA binding protein, a gene of unknown function, and a gene containing an ATP binding motif (28). A previous study in the Gram-positive Actinobacteria *Streptomyces coelicolor* showed that the *pgl* gene, along with three surrounding genes, conferred resistance to phage infection following an initial round of infection (17; 87). In the Goldfarb *et al.* study researchers found that the six-gene BREX system from *B. cereus* was sufficient to provide protection from both temperate and virulent phages when expressed in *B. subtilis* (28). The PglX protein, containing the MTase domain, was found to catalyze the formation of m6A at TAGG**A**G sites throughout the host chromosome **(Fig 1.3)**. While the MTase activity is necessary to confer protection against the invasion of foreign DNA, in the *B. cereus* system there is no decrease in cell viability in the absence of the MTase or observable cleavage of foreign DNA, suggesting that BREX systems do not achieve protection through the cleavage mechanism of a canonical RM system (28). Further, although the mechanism(s) of protection remain unclear, it is evident that BREX systems allow for adsorption of phage but not replication of phage DNA. Of the 1,500 bacterial genomes surveyed in Goldfarb *et al.*, 10% contained a putative BREX system across both Gram-positive and Gram-negative bacteria (28). More work will be necessary to understand the mechanism(s) of BREX defense systems and to determine if DNA methylation from BREX MTases has

additional regulatory roles outside of conferring protection to the host by blocking phage replication **(Fig 1.2)**.

DNA methylation from RM systems. While MTases from RM systems methylate the bacterial chromosome subsequent to replication, invading double-stranded foreign DNA from phages often enters the cell unmethylated at these recognition sites, which allows for cleavage of the foreign DNA by the cognate REase activity. There are several different types of RM systems that vary in subunit composition, cofactor requirement, recognition site, and cleavage pattern that are reviewed extensively elsewhere (23; 74; 96). Types I-III all have MTase and REase activities and are reviewed briefly here while Type IV systems, which lack MTase activity and instead cleave methylated DNA, are not discussed further and are reviewed elsewhere (50).

Type I RM systems consist of *hsdM*, *hsdS*, and *hsdR* genes which encode the MTase, specificity, and REase subunits, respectively (23; 61). The specificity subunit is composed of two target recognition domains that recognize specific bipartite recognition sites in the DNA **(Fig 1.3)** (26; 61; 62). The bipartite recognition sites, which are characteristic of Type I RM systems, consist of conserved DNA sequences at the 5' and 3' ends with 6-8 base pairs of degenerate sequence in the middle (61). Methylation is achieved at hemi-methylated bipartite motifs through the complex of two MTase subunits and one specificity subunit, resulting in methylation of both DNA strands (88; 89). Restriction activity requires complex formation of two MTase subunits, two REase subunits, and one specificity subunit. The REase complex recognizes fully unmethylated bipartite recognition sequences and collision of the complex with a DNA binding protein is required for cleavage events, which can occur several kilobases away from the original recognition site (22).

Type II RM systems are most commonly used for biotechnology applications and typically consist of stand-alone MTase and REase genes. A notable exception is the Type IIG family, which consists of a single polypeptide with both MTase and REase activities (70; 74). Type II REase enzymes, which bind to and cleave unmethylated DNA

independent of the MTase, are incredibly diverse and exhibit very low sequence identity (70). The Type II systems generally have 4-8 base pair palindromic recognition motifs, methylate both DNA strands, and cleave unmethyated sites within or near the recognition site **(Fig 1.3)(70)**. The defined cleavage within the recognition sites from REases of Type II RM systems as well as the independent activities of the MTase and REase proteins make them well-suited for applications in biotechnology (70).

Type III systems are comprised of *mod* and *res* genes that encode components for the MTase and REase activities (73). The complex of two Mod subunits is necessary to bind and methylate one strand of DNA at 5-6 base pair non-palindromic motifs **(Fig 1.3)**(11; 73). Restriction activity requires the complex of one or two Res subunits with two Mod subunits, because the DNA binding activity is intrinsic to the Mod subunits and not the Res subunit (36). Cleavage by the REase complex requires two recognition motifs oriented in opposite directions that results in cleavage 25-27 base pairs downstream of the recognition site (30; 57; 69; 73).

Type I-III RM systems are present across Gram-positive bacteria as a means of protection against the invasion of foreign DNA. Oftentimes, RM systems act as a barrier for horizontal gene transfer among closely related bacteria, resulting in clade separation among important pathogens (35; 93). Some Gram-positive species have overcome the restriction barrier to allow for the acquisition of pathogenicity islands in similar strains while maintaining the RM system for protection from phage predation (37; 38). In addition to DNA restriction, these systems also provide underappreciated roles in the regulation of gene expression and virulence potential of Gram-positive pathogens (48; 51; 65). Below, we review the functions of RM system methylation outside of protection from phage predation in Gram-positive bacteria.

Balancing host protection and the benefits of genetic transformation. In addition to host defense, Type I RM systems have been shown to regulate strain separation in Gram-positive bacteria. *Enterococcus faecium* isolates are separated into clades, where clade A consists of multi-drug resistant isolates and clade B consists of drug susceptible

fecal commensals (44). Clade A is further separated into subclades A1 and A2. Subclade A1 isolates are associated with hospital acquired infections and have a larger genome size and higher mutation rate relative to subclade A2 (44). Hou *et al.* identified multiple putative Type I RM systems across clades A and B and showed that the MTase and REase components of a Type I RM system shared greater than 90% sequence identity between these subunits in subclade A1 and clade B strains (35). However, subclades A1 and B showed high variability in their S subunits, which are required for DNA recognition and binding (35). The S subunits were highly conserved between strains from subclade A1 but appeared to be strain-specific across clade B. The authors speculate that the divergence in S subunits and subsequent methylation patterns between the subclades act as a barrier to horizontal gene transfer between members of different clades (35). Type I systems in the human pathogen *Staphylococcus aureus* also mediate horizontal gene transfer by restricting exchange from strains possessing variable S subunits (93).

While the *E. faecium* and *S. aureus* RM systems function to prevent horizontal gene transfer from between clades, other Gram-positive RM systems restrict phage DNA while maintaining mechanisms for acquisition of pathogenicity islands from related strains. Strains of the Gram-positive pathogen *Streptococcus pneumoniae* typically encode one of two Type II RM systems, DpnI or DpnII, which cleave at palindromic GATC sites throughout the genome (42). DpnI represents an atypical system because it cleaves fully methylated sites while DpnII cleaves at fully unmethylated sites. Strains with DpnII encode two upstream DNA MTases, a Dam homolog, DpnM, and a single-stranded DNA MTase, DpnA (14).

The occurrence of both RM systems across strains serves a mixed *S. pneumoniae* population in two ways. First, the occurrence of both systems protects against a broad range of phage predation, allowing for degradation of DNA independent of the methylation status at GATC sites. Second, the mixed population promotes preferential acquisition of DNA from kin. DpnI cells can acquire methylated genomic DNA from DpnII cells because the newly acquired DNA will exist in a hemi-methylated state that

9

DpnI cannot cleave (37; 38). Conversely, uptake of DpnI DNA in DpnII cells would also result in hemi-methylated DNA. If the newly acquired hemi-methylated DNA is not methylated prior to replication, the DNA will exist in a complete unmethylated state and can be cleaved by DpnII. Cleavage of unmethylated DNA in DpnII cells is prevented via methylation of the new DNA from the unique single-stranded DNA MTase DpnA. DpnA is only expressed during genetic competence ensuring that the DpnII RM system remains active against incoming phage DNA but allows for the acquisition of beneficial pathogenicity islands from related DpnI strains (37; 38).

Therefore, in addition to protecting against phage predation, RM systems function as barriers to horizontal gene transfer to maintain strain separation in Gram-positive bacteria such as *E. faecium* and *S. aureus.* Conversely, other Gram-positive species have adapted special mechanisms that use DNA methylation to acquire beneficial DNA (e.g. pathogenicity islands) while maintaining restriction activity to protect against phage predation. In the next sections we will review how RM system methylation functions in epigenetic regulation in bacteria.

Phasevarions: Epigenetic regulation by RM system MTases. Bacteria must have the ability to adapt to rapidly changing environmental conditions in order to survive. One mechanism bacteria use to cope with rapidly changing conditions is through phase variation. Phase variation occurs when certain genes, often those that encode cell surface proteins, undergo random differential expression in a reversible fashion among bacterial subpopulations (32; 68). This variation can be achieved through the presence of simple sequence repeats within genes (e.g. tandem repeats or homopolymer runs), where DNA polymerase is prone to errors that can result in non-functional or non-expressed proteins, subsequently resulting in ON/OFF expression of the gene product within a subpopulation of cells (60; 68; 92). The variation in expression can also occur as a result of genetic exchange of differentially expressed loci through homologous recombination, which typically occurs at inverted repeats within the exchanged loci (68).

Phasevarions (phase variable regulons) consist of multiple genes that are differentially regulated within various subpopulations based on epigenetic control from phase-variable MTases (86). In Gram-positive organisms, MTases from both Type I and Type III RM systems have been shown or predicted to be regulators of phasevarions (for review (4; 20). In Type I systems, homologous recombination occurs at inverted repeats within the genes for multiple specificity subunits to generate unique methylation patterns throughout the genome **(Fig 1.4A)** (19; 24; 48; 51). The subspecies specific methylation patterns act as an epigenetic signal that gives rise to differential gene expression and subsequent phenotypic differences between the subpopulations (48; 51). In Type I and Type III RM systems, variation in simple sequence repeats can result in DNA polymerase errors that give rise to subpopulations with active and inactive MTases, resulting in loss of methylation and subsequent differential gene expression **(Fig 1.4B)** (for review (83) and (2; 3). This mechanism allows for gene expression heterogeneity within a population of cells.

Regulation from S subunit variation in Type I RM systems. In the Gram-positive pathogen *Streptococcus pneumoniae* Type I phasevarions have been shown to regulate virulence via global epigenetic changes (48; 51). In one system, three separate specificity subunit genes containing inverted repeats allow for six possible specificity subunit variants **(Fig 1.5)** (51). Manso *et al.* "locked" the strains into one epigenetic state by expressing only one of the six specificity subunits and then used PacBio SMRT sequencing to show that each variant methylated different motifs, with the frequency of the various motifs differing within the genome **(Fig 1.5)** (51). The locked strains showed differential gene expression relative to one another that resulted in phenotypic consequences. Most notably, the different subtypes varied in colony opacity, which is a reversible morphological change between opaque and transparent colonies (Weiser, Infect Immun, 1994). While some variants were 100% opaque others were as low as 7% opaque colonies **(Fig 1.5)**. The colony opacity phenotypes correlated with invasive disease and carriage phenotypes, where a variant with 100% opaque colonies had poor colonization ability but was highly virulent and the variant with the majority of transparent colonies was greatly attenuated for virulence but not colonization (51; 95).

11

Moreover, the authors showed variant switching with the "unlocked" wild type strain during the course of invasive disease infection, where the cells had predominately switched to the highly virulent state with reduced colonization as early as 4 hours post-challenge (51).

A similar Type I RM system encoding two specificity subunits with inverted repeats has been shown to produce four specificity subunit variants in *S. suis*, a major veterinary pathogen, though no differential expression has been associated with the variants to date (3). In fact, an analysis of 393 *S. suis* genomes identified that 262 strains contained Type I RM systems with multiple *hsdS* specificity subunits containing inverted repeats, suggesting that the occurrence of phase variable Type I RM systems may be pervasive across this species (3). Additionally, the presence of phase variable Type I RM systems have been predicted or identified in strains of *Enterococcus faecalis, Listeria monocytogenes, Clostridium botulinum, and Lactobacillus salivarus* (19; 20; 24). More work needs to be completed to understand how phase variable Type I RM systems affect virulence gene expression across Gram-positive pathogens.

Regulation from bi-phasic MTases in Type III RM systems. In various Gram-negative pathogens, including species of *Haemophilus*, *Neisseria*, *Kingella, Helicobacter,* and *Moraxella*, phase variable Type III *mod* alleles, encoding the Mod protein responsible for MTase activity, have been implicated in the regulation of gene expression (for review (83) and (9; 82; 84-86). The Mod proteins from these Type III systems exhibit ON/OFF expression within a population due to the presence of simple sequence repeats (SSRs) within the *mod* gene, which can cause DNA polymerase slippage at the SSRs (60; 68; 92). While no studies, to our knowledge, have demonstrated a phase variable Type III RM system regulating gene expression in Gram-positive bacteria, the presence of candidate phase variable Type III systems have been identified in *S. thermophiles*, *S. galactiae*, *S. mitis*, and *L. saerimneri* strains (4). These candidate phase variable Type III systems were identified based on the presence of SSRs within the *mod* allele (4). Putative epigenetic regulation by these novel systems remains an area of continued investigation.

In addition to the examples of the Type I and Type III systems discussed above, both SSRs and inverted repeats have been observed in the PglX MTase of BREX systems, resulting in phase variation for expression of the system (28). Phase variable MTases represent an important mechanism of epigenetic regulation in Gram-positive bacteria, allowing for differential methylation patterns and subsequently differential gene expression within various bacterial subpopulations (68). Few studies have investigated the regulatory effects of DNA methylation from active and inactive RM systems outside of Type I RM systems with multiple specificity subunits or Type III RM systems containing short sequence repeats within the *mod* allele. Below we will discuss our current understanding of the important regulatory functions of DNA methylation from non-phase variable RM systems across bacteria from the two Gram-positive phyla, Actinobacteria and Firmicutes.

DNA methylation-dependent mechanisms for the regulation of gene expression in Actinobacteria. The Actinobacteria comprise one of the largest and most diverse bacteria phyla, including Gram-positive filamentous bacteria with high GC content genomes (for review (7; 45)). Actinobacteria can be found in aquatic and terrestrial environments where they are important contributors to diverse ecosystems (7; 29). The impact of DNA methylation outside of RM systems on the cell physiology of Actinobacteria remains largely unexplored, with the first studies focusing on *Streptomyces* and *Mycobacterium*. The soil dwelling *Streptomyces* have been well studied for their multicellular behaviors and complex lifestyles (7; 98). *Streptomyces* are also of tremendous importance to biotechnology and human health as they are responsible for the production of 2/3 of clinically relevant antibiotics (45; 63; 72). *Mycobacterium* species are well known for causing a broad range of human diseases, particularly in immunocompromised individuals, and represent significant burdens on healthcare systems across the world (21; 41; 90). Given the importance of *Streptomyces* and *Mycobacterium* on human health, as well as the impact of other Actinobacteria genera on terrestrial and aquatic ecosystems, the initial studies

suggesting an important regulatory role for DNA methylation in the adaptive lifestyles of these bacteria is of particular importance for on-going and future research.

Mycobacterium tuberculosis is a Gram-positive pathogen that represents a significant worldwide public health burden, causing more than 1.5 million deaths in 2018 ([WHO] (21). The antibiotics rifampin and isoniazid, among others, have been used to cure tuberculosis infections, however multi-drug resistant tuberculosis (MDR-TB) strains, which are resistant to both rifampin and isoniazid, are emerging (41; 90). Among the mechanisms for emerging antibiotic resistance, a study by Chen *et al.* suggests that the extent of methylation differs between rifampicin and isoniazid treated *M. tuberculosis* H37Rv strains compared to the untreated wild type strain (15). A separate study of para-aminosalicylic acid (PAS) resistant Mycobacterium suggests differential methylation in PAS resistant H37Rv, with 1,161 hyper-methylated and 227 hypo-methylated genes relative to the susceptible parent strain (47). These data suggest that DNA methylation contributes to antibiotic resistance of *Mycobacterium tuberculosis* with the strong potential to contribute to formation of persister cells.

Another study suggests that DNA methylation may play an important role in *M. tuberculosis* survival under hypoxic conditions (80). Latent infections with *M. tuberculosis* can last decades, requiring the bacteria to survive, persist, and adapt to a range of environmental conditions within the human host (27). Shell *et al.* discovered a Type II MTase, MamA, present in a subset of *M. tuberculosis* strains that catalyzes m6A at CTGG**A**G sites throughout the genome (80). MamA is also conserved in other Mycobacterium species including *M. smegmatis*, *M. bovis, M. avium,* and *M. leprae.* Upon loss of *mamA* in *M. tuberculosis*, a small but significant decrease in the expression a subset of genes was observed where the MamA recognition site overlapped with putative sigma factor -10 binding boxes. Moreover, the researchers found that the *mamA* deficient cells had decreased viability in hypoxic conditions relative to wild type cells. These hypoxic conditions were used to simulate those of hypoxic granulomas formed in the human host (91). A separate study of nineteen *Mycobacterium tuberculosis* complex strains found that MamA had 13 binding sites that

overlapped with SigA and that strains with inactive MamA variants showed decreased expression of the downstream genes relative to strains with active MamA (18). The same study showed that while methylation from a separate Type I RM system in *M. tuberculosis* strains did not directly influence the expression of genes through overlap with known sigma factor binding sites, loss of methylation indirectly affected expression of a small subset of genes in the absence of a recognition site near the affected genes (18). Therefore these results suggest both direct and indirect mechanisms for DNA methylation in the regulation of gene expression **(Fig 1.6)** highlighting the importance of DNA methylation beyond restriction-modification systems in clinically important Actinobacteria.

In addition to m6A-dependent regulation, m5C modifications have been shown to function in the regulation of antibiotic production and development in Actinobacteria. Streptomycetes are Gram-positive soil-dwelling bacteria that produce two thirds of all clinically relevant secondary metabolites (63; 72). In addition to antibiotic production, *Streptomyces* species are known for their complex life cycles, which include differentiation and programmed cell death (PCD) (for review (7; 98). Briefly, subsequent to uninucleoid spore germination, hyphae growth gives rise to a first/vegetative mycelium (MI) (52). Upon nutrient depletion, PCD occurs as the multinucleated second/differentiated mycelium (MII) develops, which consists of multiple cell types including the aerial mycelium and sporulating mycelium (52). The sporulating mycelium undergoes PCD to form the uninucleoid spore (52). A recent study showed that both antimicrobial production in *Streptomyces* and development are affected by m5C methylation (71). DNA extracted from strains of *S. coelicolor, S. avermitilis, S. griseus,* and *S. lividans* showed less m5C in the MII stages compared to MI in all four species (71). Moreover, the researchers used a gene interruption in the putative MTase *SCO1731* (*SCO1731::Tn5062*) and found significant reduction in the genomic m5C signal in the *S. coelicolor* genome in MI but only a slight reduction in signal in MII (71). Phenotypically, the *SCO1731::Tn5062* strain displayed a substantial delay in differentiation on solid media, with aerial mycelium formation occurring at 96 hours relative to formation at 48 hours in wild type cells. The mutant was also severely

15

impaired for production of the antibiotic actinorhodin (71). *S. coelicolor* encodes 37 putative DNA MTases in addition to *SCO1731*, a subset of which are differentially expressed in MI and MII stages of development (71; 98). Further studies are necessary to determine the extent to which various methylation events regulate development and the expression of clinically relevant secondary metabolites across Streptomyces. Nevertheless, it appears that further studies will reveal an important regulatory contribution for DNA methylation in the complex life cycles of *Streptomyces*, raising broadly conserved biological parallels with the developmental regulatory functions of DNA methylation in eukaryotes.

DNA methylation-dependent mechanisms for the regulation of gene expression in Firmicutes. The Firmicutes phylum includes Gram-positive bacteria with low GC content genomes. In addition to being one of the dominating phyla in the human gut microbiome, members of the Firmicutes also encompass several important human pathogens, including *Stapylococcus*, *Streptococcus, Enterococcus*, *Clostridium*, and *Listeria* species (46). Despite the very limited research available outside of regulation by phase variable MTases, RM system MTases have been shown to regulate gene expression in Firmicutes outside of host defense, prompting important possibilities for the functions of DNA methylation across Firmicutes.

Epigenetic regulation of virulence factors from a Type I RM system has been shown for the important human pathogen *Streptococcus pyogenes*. Loss of m6A from an active Type I RM system resulted in substantial down regulation of 20 genes that clustered into six distinct loci in a clinical isolate of *S. pyogenes* (Chapter II (65)). Many of the differentially expressed genes were part of the core regulon for the stand-alone transcriptional regulator, Mga (Chapter II (65)). The Mga core regulon consists of genes that encode cell surface proteins, including the M-protein, C5a peptidase, which cleaves host complement, and the Mga regulator itself, which are important for adhesion, internalization, and immune evasion phenotypes (34; 56). The m6A-dependent decrease in expression of the Mga regulon resulted in decreased adhesion of *S. pyogenes* cells to host epithelial cells, a decreased ability of the bacteria to survive

16

within host neutrophils, and a decreased ability to evade the host immune response. Interestingly, the *S. pyogenes* genome contains another putative Type I specificity unit (AWM59_04585), which is not surrounded by *hsdM* or *hsdR* genes. However, AWM59_04585 is located 691kb from the S subunit (AWM59_07900) of the active Type I RM system and REBASE annotates AWM59_04585 as unlikely to be a genuine S subunit (http://rebase.neb.com). Thus, more work is necessary to determine if S-subunit switching occurs in *S. pyogenes* as it does it in *S. pneumonae* or if the epigenetic regulation described in Nye *et al.* represents a phase variation independent mechanism of regulation by a Type I RM system. Either biological mechanism would impart regulation of *S. pyogenes* virulence.

Gene regulation in *Streptococcus* is also governed by the presence of a Type II RM system. As previously discussed, in Gram-negative *E. coli* and related Proteobacteria, Dam MTase occurs as a stand-alone orphan MTase that functions in many important cellular processes, including origin sequestration (31; 64), DNA mismatch repair (5; 43), and the regulation of gene expression (13). Homologs of Dam MTase occur in a subset of Gram-positive bacteria, however they often exist as part of an active RM system, such as the DpnM-DpnA-DpnII system from *S. pneumoniae* discussed above (6; 37). Homologs of the DpnM-DpnA-DpnII system occur in a subset of strains from other Gram-positive bacteria such as *Streptococcus mutans, Lactococcus lactis, Streptococcus sanguinis,* and *Streptococcus suis* (6; 58; 78; 97). In *S. mutans*, it was shown that deletion of the DpnM homolog, DamA, resulted in the differential expression of over 100 genes, of which 70 were up regulated and 30 were down regulated at least two fold in the *damA* mutant relative to wild type (6). The differentially expressed genes included virulence factors, bacteriocins, and genes involved in sugar metabolism, which would contribute to the formation of dental caries and tooth decay (6). Importantly, this study showed that the differences in gene expression had effects at the phenotypic level. The up regulation of the cell surface glucan receptor, GpbC, in the *damA* mutant resulted in increased clumping in dextran-dependent aggregation assays and the increases in bacteriocin gene expression resulted in larger zones of clearing in the *damA* mutant against *Streptococcus godonii* and *Lactococcus lactis* strains (6). Thus, in

addition to functioning as part of a restriction-modification system, the *S. mutans* DNA MTase DpnM also functions in the regulation of gene expression. It remains unknown if the DpnM homologs in other *Streptococcus* species have regulatory functions beyond host restriction.

Another example of DNA methylation regulating gene expression in Firmicutes was demonstrated in a recent study of the *Bacillus subtilis* MTase, DnmA. In Nye *et al.* researchers characterized the methylomes of the lab and ancestral strains of *B. subtilis* PY79 and NCIB 3610, respectively (Chapter III (66)). They found that the DnmA MTase from a Type I-like RM system catalyzed the formation of m6A at non-palindromic GACG**A**G sites throughout the chromosome. The absence of DnmA did not affect natural transformation efficiency, suggesting that DnmA either does not have activity as a canonical Type I RM-like system or the activity cannot be measured during natural transformation (Chapter III (66)). Moreover, deletion of *dnmA* resulted in small but significant decreases in expression for a subset of genes that are important for chromosome structure and maintenance. DnmA recognition sites were proximal to the -35 box for sigma factor SigA binding in the promoters of the differentially expressed genes. Further, this study found that the transition state transcriptional repressor ScoC, preferentially bound an unmethylated promoter, providing mechanistic insight into the MTase-dependent regulation of gene expression in Gram-positive bacteria (Chapter III (66)). These data show that ScoC binding to a reporter promoter region is stronger for unmethylated relative to methylated DNA demonstrating that ScoC repressor binding serves to reduce gene expression when methylation is absent (Chapter III (12; 66)).

Conclusions and Future Perspectives

Methylation of genomic DNA is pervasive across bacterial genomes, where it has been most extensively studied as a self-recognition mechanism in host defense. The majority of the pioneering studies exploring the function of DNA methylation outside of host defense have been done in Gram-negative bacteria (1; 53; 54; 59; 75). However, outside of the CamA MTase conserved only in specific species of *Clostridiales*, much less is known about the functions of orphan MTases in Gram-positive bacteria (67). A

critical area of future investigation is understanding the biological contribution for enrichment of orphan MTase recognition sites in the putative origin of replication region for *Arthrobacter* species, which are used for commercial production of glutamic acid, and *Norcardia* species, a subset of which can cause opportunistic infections in susceptible populations (10). The over-representation of MTase sites in their predicted origin region suggests that orphan MTase methylation regulates origin firing in a subset of Gram-positive species. Additionally, unmethylated recognition sites from Gram-positive orphan MTases can be also be found in promoter regions for transcriptional regulators, suggesting an additional contribution in regulated gene expression (10). Given the conservation of putative orphan MTases in Gram-positive bacteria it is tempting to speculate that MTase function is conserved across distantly related species. In our opinion experiments are necessary to determine the function of orphan MTase methylation in Gram-positive bacteria and how methylation regulates cell proliferation and gene expression.

In addition to orphan MTases, the regulatory functions of methylation from RM systems has also focused on Gram-negative bacteria. While phase variable Type I RM MTases have been found to be important for *Streptococcus* virulence (48; 51), as discussed here, most other studies of Type I and Type III phase variable RM systems have been completed in Gram-negative bacteria. Outside of epigenetic regulation from phase variable RM systems, few studies have explored the regulatory consequences of DNA methylation from non-phase variable RM systems in both Gram-positive and Gram-negative bacteria. Here we have discussed epigenetic regulation from non-phase variable RM systems in *Mycobacterium, Streptomyces, Streptococcus,* and *Bacillus* species. In some systems, such as MamA and DnmA from *M. tuberculosis* and *B. subtilis*, respectively, the mechanism of methylation-dependent regulation appears to be direct, where m6A modifications overlap with transcription factor binding sites in differentially expressed genes **(Fig 1.6)** (80) (Chapter III (66)). In *B. subtilis* researchers identified an m6A sensitive transcriptional regulator, ScoC, which bound near the sigma factor binding site, providing some of the first insight into the mechanism of m6A-dependent regulation in Gram-positive bacteria (Chapter III (66)). It remains to be

determined if m6A regulation of ScoC binding is a common mechanism for ScoC regulated genes or specific to particular loci. In other systems, such as the Type I RM systems in *M. tuberculosis* and *S. pyogenes*, the mechanism of methylation-dependent regulation of gene expression appears to be indirect, with modified recognition motifs occurring distal to the differentially expressed genes **(Fig 1.6)** (Chapter II, (18; 65)). Both direct and indirect mechanisms of regulation from non-phase variable RM systems appear to have important consequences for cell physiology, where they affect virulence potential, adaptability to environmental conditions, and bacterial development. Given the widespread occurrence of DNA methylation in Gram-positive bacteria and the importance of Actinobacteria and Firmicutes to human health, industry, and the environment, further study of DNA methylation in Gram-positive bacteria is important for understanding regulatory and phenotypic variations among bacteria within populations.
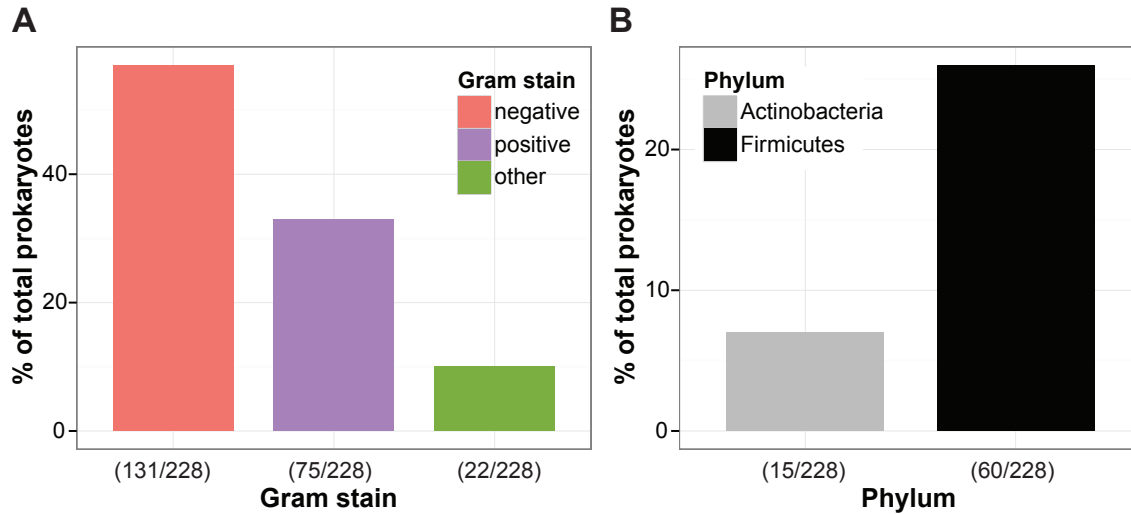
Figure 1.1. DNA methylation has been most intensely studied in Gram-negative bacteria. (A) Gram stains of bacteria included in the Blow *et al.* survey of prokaryotic genome methylation (10). Bacteria were grouped based on Gram-stain. The percent of Gram-negative (pink), Gram-positive (purple), and Other (green) species is indicated on the y-axis. The number of species in each category out of the total surveyed is indicated as a fraction underneath each bar. The 'Other' category consisted of Archaea and bacterial species from *Chloroflexi*, *Plantomycetes,* and *Deinococcus-Thermus,* which exhibit atypical Gram stains based on cell wall structure. **(B)** The percent of representative bacteria from the major Gram-positive phyla in the Blow *et al.* survey (10). The percent of Actinobacteria (gray) and Firmicutes (black) species is indicated on the y-axis with the number of species included out of the total surveyed indicated as a fraction underneath each bar.
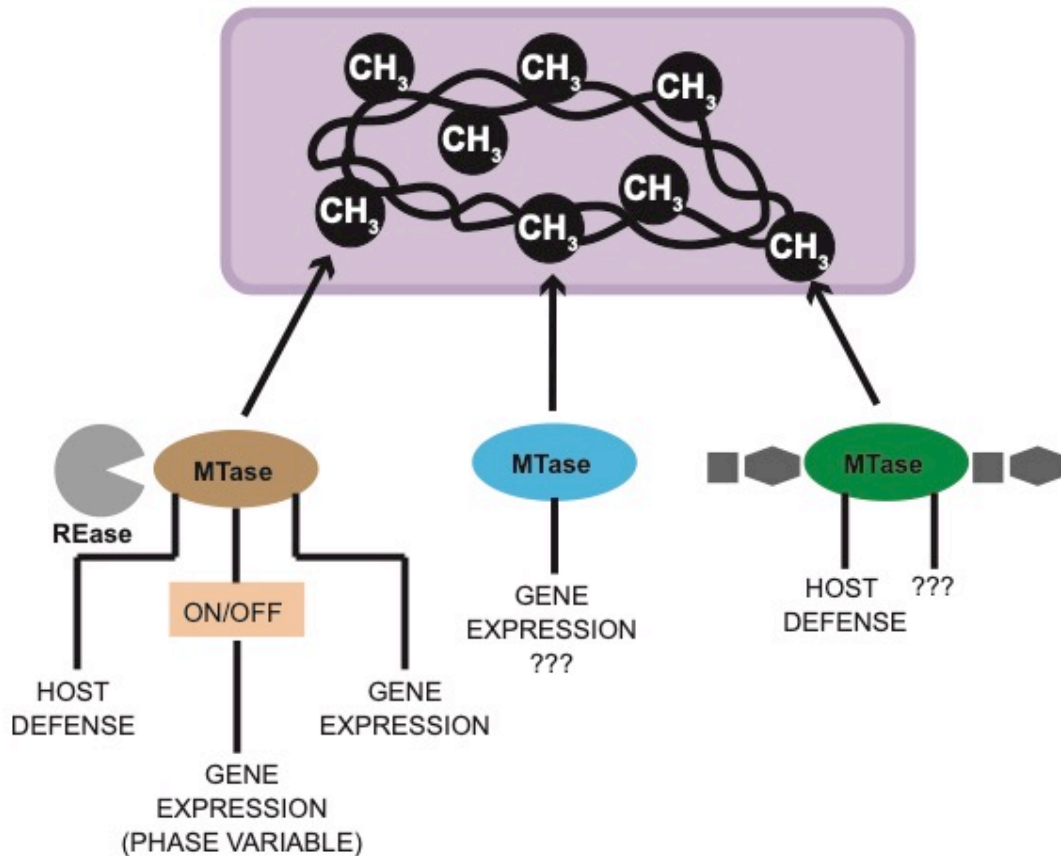
Figure 1.2. The functions of DNA methylation in Gram-positive bacteria. Genomic DNA methylation in Gram-positive bacteria occurs from the activity of RM system MTases (brown), orphan MTases (blue), or BREX MTases (green). Methylation from both BREX and RM MTases has been shown to function in host defense. Both phase variable and non-phase variable MTases from RM systems have been shown to regulate gene expression in Gram-positive bacteria as well. To date, a regulatory function for DNA methylation from BREX system MTases has not been experimentally demonstrated.
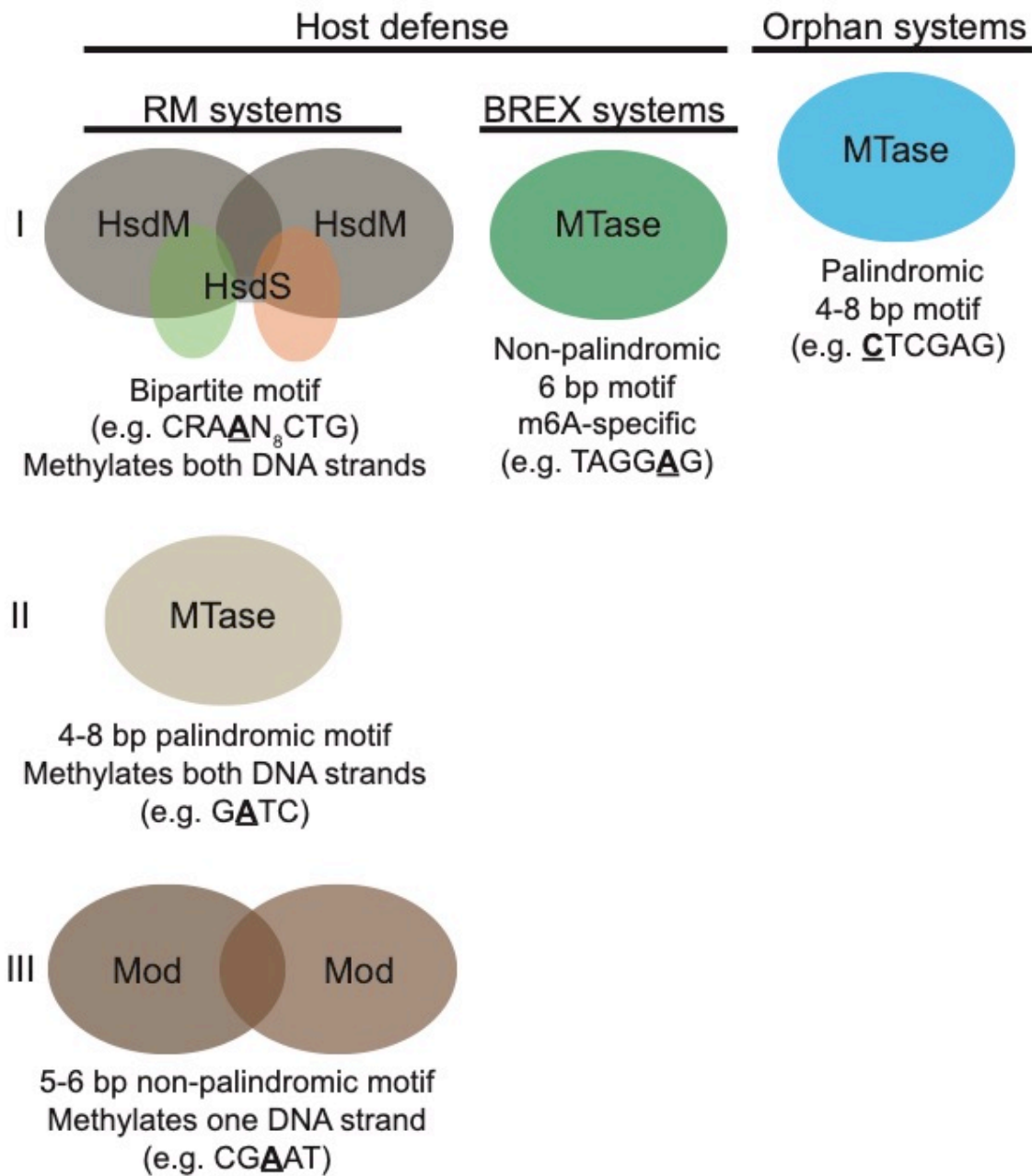
Figure 1.3. DNA MTases in Gram-positive bacteria. DNA methylation in Gram-positive bacteria comes from DNA MTases that exist as part of RM systems (brown), BREX (green), and orphan MTases (blue). The composition of the MTase component from Types I-III RM systems is indicated as well as the typical recognition motifs and methylation patterns. The typical recognition motif and methylation pattern from BREX systems and orphan MTases is also included (8; 28).
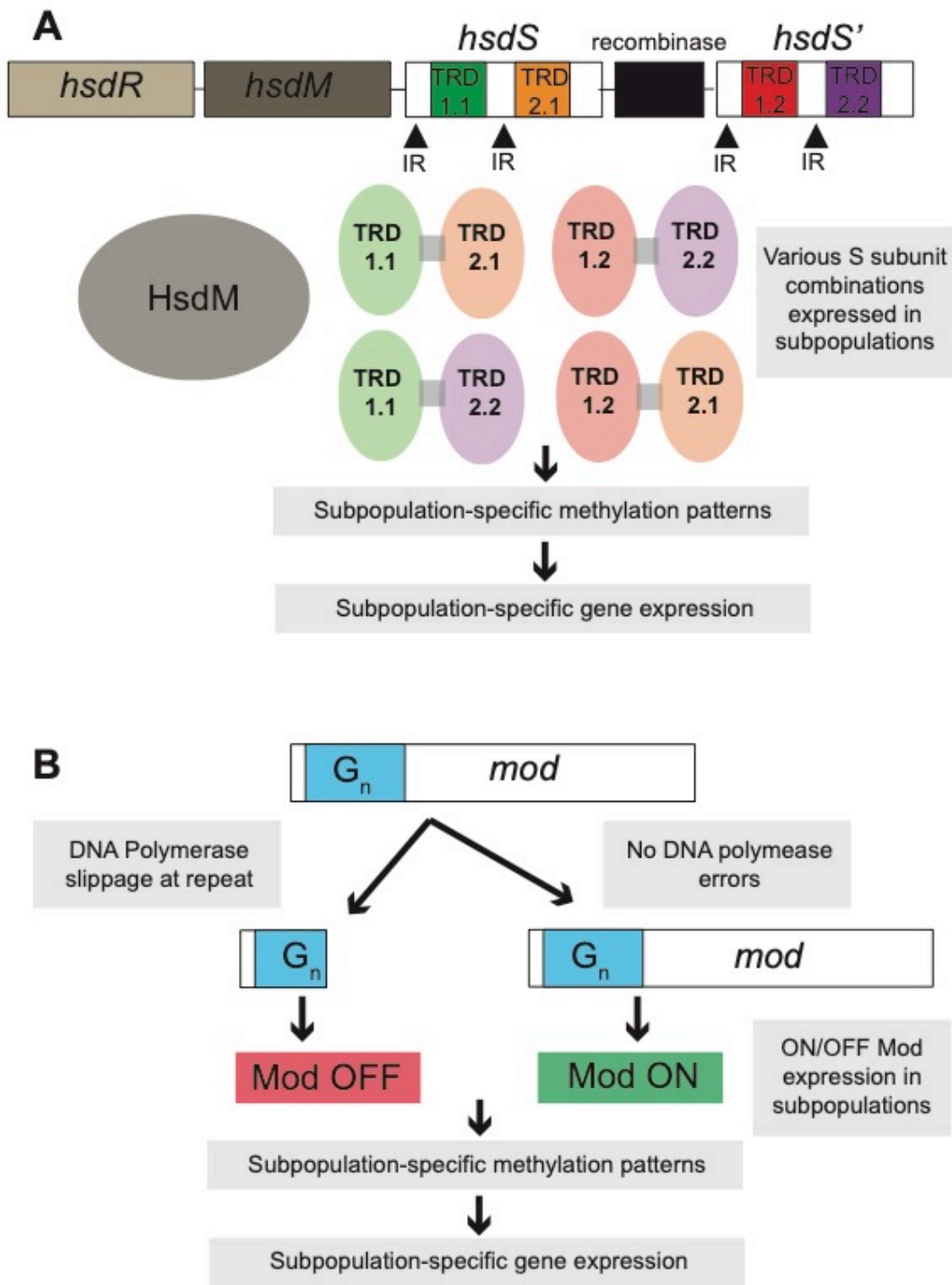
Figure 1.4. Phase variable MTases from Type I and III RM systems. (A) Phase variable MTases from Type I RM systems occur through S-subunit switching. Random recombination of the TRDs from *hsdS* and *hsdS'* occurs at inverted repeats within the genes by the proximally encoded recombinase. The recombination events produce multiple S-subunits with different combinations of TRDs that target the MTase,

comprised of HsdM and HsdS subunits, to different recognition sites throughout the genome resulting in bacterial subpopulations with various methylation patterns. The subpopulation specific methylation patterns can result in differential gene expression between subpopulations. **(B)** Phase variable MTases from Type III RM systems occur through DNA polymerase slippage at SSRs. Random DNA polymerase slippage at a homopolymer track in the coding region of the *mod* allele results in subpopulations with truncated and full length Mod proteins. The subpopulations with the truncated Mod protein lack the DNA methylation present in the population with the functional full length Mod-protein, resulting in subpopulation specific DNA methylation patterns that can result in differential gene expression between the populations (77).
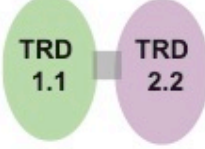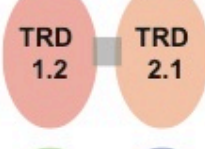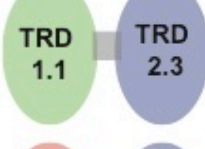
Figure 1.5. Phase variable MTase in *S. pneumoniae* regulates virulence in distinct subpopulations. Shown are the six different S-subunits produced from recombination of the TRDs from three *hsdS* genes to produce systems A-F as described in Manso *et al*. The distinct recognition site for each system is listed according to the color-coded TRDs in the S-subunit. The percent of colonies displaying the opaque phenotype for each subpopulation is also indicated (51). This figure is based on the following reference (51).
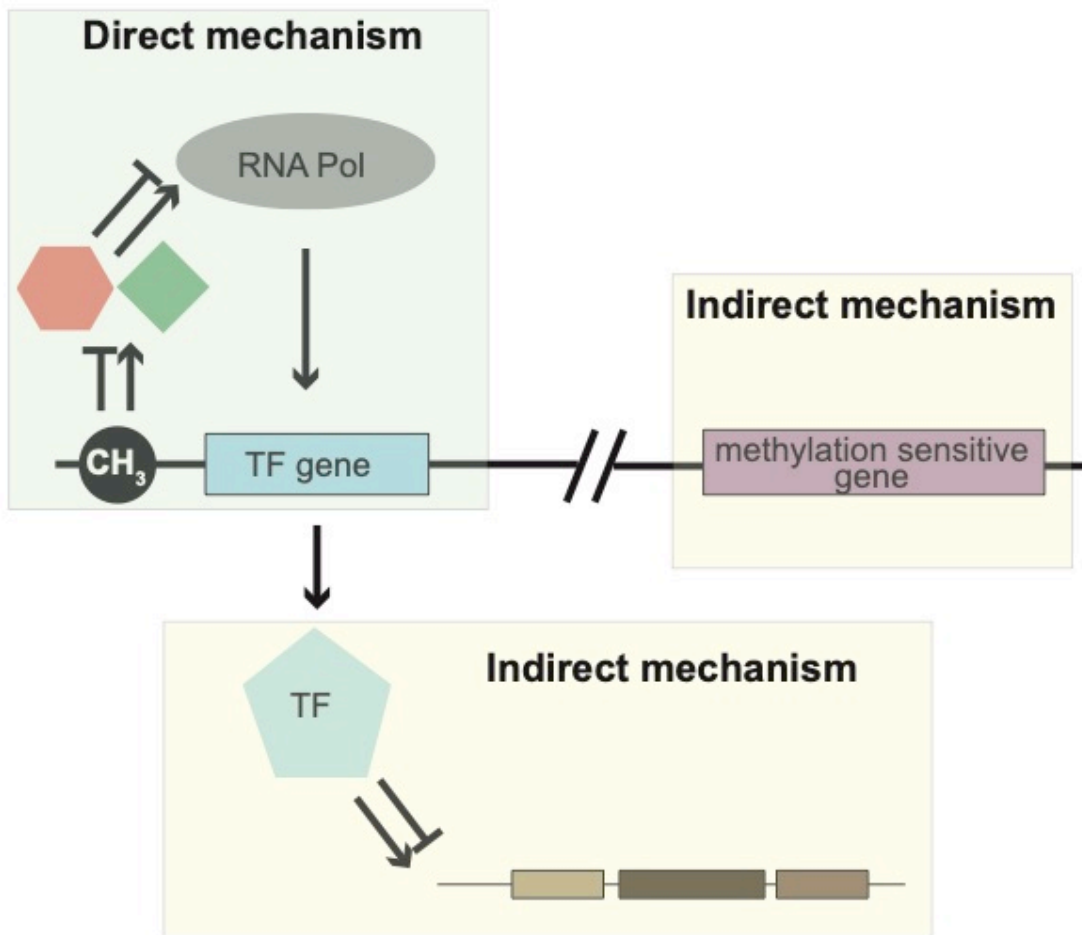
Figure 1.6. Mechanisms of DNA methylation-dependent regulation of gene expression in Gram-positive bacteria. Direct regulatory mechanisms result from the occurrence of methylation within a promoter region of a gene that affects binding of transcriptional regulators that influence RNA polymerase activity, subsequently affecting gene expression. Indirect regulation can occur through differential expression of a gene that is directly regulated by DNA methylation, such as transcription factors (TF). The methylation-dependent differential expression of the TF can result in downstream differential expression of many genes within the TF regulon. Indirect regulation can also occur at genes that are differentially expressed upon loss of DNA methylation but are not proximal to any methylated sites. Such indirect mechanisms are poorly understood but occur in a number of bacteria.

References

1. Adhikari S, Curtis PD. 2016. DNA methyltransferases and epigenetic regulation in bacteria. *FEMS Microbiol Rev* 40:575-91
2. Atack JM, Guo C, Yang L, Zhou Y, Jennings MP. 2020. DNA sequence repeats identify numerous Type I restriction-modification systems that are potential epigenetic regulators controlling phase-variable regulons; phasevarions. *FASEB J* 34:1038-51
3. Atack JM, Weinert LA, Tucker AW, Husna AU, Wileman TM, et al. 2018. Streptococcus suis contains multiple phase-variable methyltransferases that show a discrete lineage distribution. *Nucleic Acids Res* 46:11466-76
4. Atack JM, Yang Y, Seib KL, Zhou Y, Jennings MP. 2018. A survey of Type III restriction-modification systems reveals numerous, novel epigenetic regulators controlling phase-variable regulons; phasevarions. *Nucleic Acids Res* 46:3532-42
5. Bale A, d'Alarcao M, Marinus MG. 1979. Characterization of DNA adenine methylation mutants of Escherichia coli K12. *Mutat Res* 59:157-65
6. Banas JA, Biswas S, Zhu M. 2011. Effects of DNA methylation on expression of virulence genes in Streptococcus mutans. *Appl Environ Microbiol* 77:7236-42
7. Barka EA, Vatsa P, Sanchez L, Gaveau-Vaillant N, Jacquard C, et al. 2016. Taxonomy, Physiology, and Natural Products of Actinobacteria. *Microbiol Mol Biol Rev* 80:1-43
8. Barrangou R, van der Oost J. 2015. Bacteriophage exclusion, a new defense system. *EMBO J* 34:134-5
9. Blakeway LV, Power PM, Jen FE, Worboys SR, Boitano M, et al. 2014. ModM DNA methyltransferase methylome analysis reveals a potential role for Moraxella catarrhalis phasevarions in otitis media. *FASEB J* 28:5197-207
10. Blow MJ, Clark TA, Daum CG, Deutschbauer AM, Fomenkov A, et al. 2016. The Epigenomic Landscape of Prokaryotes. *PLoS Genet* 12:e1005854
11. Brockes JP. 1973. The deoxyribonucleic acid-modification enzyme of bacteriophage P1. Subunit structure. *Biochem J* 133:629-33
12. Caldwell R, Sapolsky R, Weyler W, Maile RR, Causey SC, Ferrari E. 2001. Correlation between Bacillus subtilis scoC phenotype and gene expression determined using microarrays for transcriptome analysis. *J Bacteriol* 183:7329-40
13. Casadesus J, Low DA. 2013. Programmed heterogeneity: epigenetic mechanisms in bacteria. *J Biol Chem* 288:13929-35
14. Cerritelli S, Springhorn SS, Lacks SA. 1989. DpnA, a methylase for single-strand DNA in the Dpn II restriction system, and its biological function. *Proc Natl Acad Sci U S A* 86:9223-7
15. Chen L, Li H, Chen T, Yu L, Guo H, et al. 2018. Genome-wide DNA methylation and transcriptome changes in Mycobacterium tuberculosis with rifampicin and isoniazid resistance. *Int J Clin Exp Pathol* 11:3036-45
16. Chen Z, Zhang Y. 2019. Role of Mammalian DNA Methyltransferases in Development. *Annu Rev Biochem*
17. Chinenova TA, Mkrtumian NM, Lomovskaia ND. 1982. [Genetic characteristics of a new phage resistance trait in Streptomyces coelicolor A3(2)]. *Genetika* 18:1945-52

18. Chiner-Oms A, Berney M, Boinett C, Gonzalez-Candelas F, Young DB, et al. 2019. Genome-wide mutational biases fuel transcriptional diversity in the Mycobacterium tuberculosis complex. *Nat Commun* 10:3994
19. Claesson MJ, Li Y, Leahy S, Canchaya C, van Pijkeren JP, et al. 2006. Multireplicon genome architecture of Lactobacillus salivarius. *Proc Natl Acad Sci U S A* 103:6718-23
20. De Ste Croix M, Vacca I, Kwun MJ, Ralph JD, Bentley SD, et al. 2017. Phase-variable methylation and epigenetic regulation by type I restriction-modification systems. *FEMS Microbiol Rev* 41:S3-S15
21. Dorman SE, Chaisson RE. 2007. From magic bullets back to the magic mountain: the rise of extensively drug-resistant tuberculosis. *Nat Med* 13:295-8
22. Dryden DT, Cooper LP, Thorpe PH, Byron O. 1997. The in vitro assembly of the EcoKI type I DNA restriction/modification enzyme and its in vivo implications. *Biochemistry* 36:1065-76
23. Ershova AS, Rusinov IS, Spirin SA, Karyagina AS, Alexeevski AV. 2015. Role of Restriction-Modification Systems in Prokaryotic Evolution and Ecology. *Biochemistry (Mosc)* 80:1373-86
24. Fagerlund A, Langsrud S, Schirmer BC, Moretro T, Heir E. 2016. Genome Analysis of Listeria monocytogenes Sequence Type 8 Strains Persisting in Salmon and Poultry Processing Environments and Comparison with Related Strains. *PLoS One* 11:e0151117
25. Flusberg BA, Webster DR, Lee JH, Travers KJ, Olivares EC, et al. 2010. Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat Methods* 7:461-5
26. Fuller-Pace FV, Bullas LR, Delius H, Murray NE. 1984. Genetic recombination can generate altered restriction specificity. *Proc Natl Acad Sci U S A* 81:6095-9
27. Getahun H, Matteelli A, Chaisson RE, Raviglione M. 2015. Latent Mycobacterium tuberculosis infection. *N Engl J Med* 372:2127-35
28. Goldfarb T, Sberro H, Weinstock E, Cohen O, Doron S, et al. 2015. BREX is a novel phage resistance system widespread in microbial genomes. *EMBO J* 34:169-83
29. Goodfellow M, Williams ST. 1983. Ecology of actinomycetes. *Annu Rev Microbiol* 37:189-216
30. Hadi SM, Bachi B, Shepherd JC, Yuan R, Ineichen K, Bickle TA. 1979. DNA recognition and cleavage by the EcoP15 restriction endonuclease. *J Mol Biol* 134:655-66
31. Han JS, Kang S, Kim SH, Ko MJ, Hwang DS. 2004. Binding of SeqA protein to hemi-methylated GATC sequences enhances their interaction and aggregation properties. *J Biol Chem* 279:30236-43
32. Henderson IR, Owen P, Nataro JP. 1999. Molecular switches--the ON and OFF of bacterial phase variation. *Mol Microbiol* 33:919-32
33. Hernday AD, Braaten BA, Low DA. 2003. The mechanism by which DNA adenine methylase and PapI activate the pap epigenetic switch. *Mol Cell* 12:947-57
34. Hondorp ER, McIver KS. 2007. The Mga virulence regulon: infection where the grass is greener. *Mol Microbiol* 66:1056-65

35. Huo W, Adams HM, Trejo C, Badia R, Palmer KL. 2019. A Type I Restriction-Modification System Associated with Enterococcus faecium Subspecies Separation. *Appl Environ Microbiol* 85

36. Janscak P, Sandmeier U, Szczelkun MD, Bickle TA. 2001. Subunit assembly and mode of DNA cleavage of the type III restriction endonucleases EcoP1I and EcoP15I. *J Mol Biol* 306:417-31

37. Johnston C, Caymaris S, Zomer A, Bootsma HJ, Prudhomme M, et al. 2013. Natural genetic transformation generates a population of merodiploids in Streptococcus pneumoniae. *PLoS Genet* 9:e1003819

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

39. Jones PA. 2012. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet* 13:484-92

40. Jurkowska RZ, Jeltsch A. 2016. Enzymology of Mammalian DNA Methyltransferases. *Adv Exp Med Biol* 945:87-122

41. Kim DH, Kim HJ, Park SK, Kong SJ, Kim YS, et al. 2008. Treatment outcomes and long-term survival in patients with extensively drug-resistant tuberculosis. *Am J Respir Crit Care Med* 178:1075-82

42. Lacks S, Greenberg B. 1975. A deoxyribonuclease of Diplococcus pneumoniae specific for methylated DNA. *J Biol Chem* 250:4060-66

43. Lahue RS, Au KG, Modrich P. 1989. DNA mismatch correction in a defined system. *Science* 245:160-4

44. Lebreton F, van Schaik W, McGuire AM, Godfrey P, Griggs A, et al. 2013. Emergence of epidemic multidrug-resistant Enterococcus faecium from animal and commensal strains. *mBio* 4

45. Lewin GR, Carlos C, Chevrette MG, Horn HA, McDonald BR, et al. 2016. Evolution and Ecology of Actinobacteria and Their Bioenergy Applications. *Annu Rev Microbiol* 70:235-54

46. Ley RE, Peterson DA, Gordon JI. 2006. Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell* 124:837-48

47. Li HC, Chen T, Yu L, Guo HX, Chen L, et al. 2020. Genome-wide DNA methylation and transcriptome and proteome changes in Mycobacterium tuberculosis with para-aminosalicylic acid resistance. *Chem Biol Drug Des* 95:104-12

48. Li J, Li JW, Feng Z, Wang J, An H, et al. 2016. Epigenetic Switch Driven by DNA Inversions Dictates Phase Variation in Streptococcus pneumoniae. *PLoS Pathog* 12:e1005762

49. Loenen WA, Dryden DT, Raleigh EA, Wilson GG. 2014. Type I restriction enzymes and their relatives. *Nucleic Acids Res* 42:20-44

50. Loenen WA, Dryden DT, Raleigh EA, Wilson GG, Murray NE. 2014. Highlights of the DNA cutters: a short history of the restriction enzymes. *Nucleic Acids Res* 42:3-19

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

52. Manteca A, Fernandez M, Sanchez J. 2005. A death round affecting a young compartmentalized mycelium precedes aerial mycelium dismantling in confluent surface cultures of Streptomyces antibioticus. *Microbiology* 151:3689-97
53. Marczynski GT, Shapiro L. 2002. Control of chromosome replication in caulobacter crescentus. *Annu Rev Microbiol* 56:625-56
54. Marinus MG, Casadesus J. 2009. Roles of DNA adenine methylation in host-pathogen interactions: mismatch repair, transcriptional regulation, and more. *FEMS Microbiol Rev* 33:488-503
55. Marinus MG, Lobner-Olesen A. 2014. DNA Methylation. *EcoSal Plus* 6:doi: 10.1128/ecosalplus.ESP-0003-2013.
56. McIver KS, Scott JR. 1997. Role of mga in growth phase regulation of virulence genes of the group A streptococcus. *J Bacteriol* 179:5178-87
57. Meisel A, Bickle TA, Kruger DH, Schroeder C. 1992. Type III restriction enzymes need two inversely oriented recognition sites for DNA cleavage. *Nature* 355:467-9
58. Moineau S, Walker SA, Holler BJ, Vedamuthu ER, Vandenbergh PA. 1995. Expression of a Lactococcus lactis Phage Resistance Mechanism by Streptococcus thermophilus. *Appl Environ Microbiol* 61:2461-6
59. Mouammine A, Collier J. 2018. The impact of DNA methylation in Alphaproteobacteria. *Mol Microbiol* 110:1-10
60. Moxon R, Bayliss C, Hood D. 2006. Bacterial contingency loci: the role of simple sequence DNA repeats in bacterial adaptation. *Annu Rev Genet* 40:307-33
61. Murray NE. 2000. Type I restriction systems: sophisticated molecular machines (a legacy of Bertani and Weigle). *Microbiol Mol Biol Rev* 64:412-34
62. Nagaraja V, Shepherd JC, Bickle TA. 1985. A hybrid recognition sequence in a recombinant restriction enzyme and the evolution of DNA sequence specificity. *Nature* 316:371-2
63. Newman DJ, Cragg GM. 2007. Natural products as sources of new drugs over the last 25 years. *J Nat Prod* 70:461-77
64. Nievera C, Torgue JJ, Grimwade JE, Leonard AC. 2006. SeqA blocking of DnaA-oriC interactions ensures staged assembly of the E. coli pre-RC. *Mol Cell* 24:581-92
65. Nye TM, Jacob KM, Holley EK, Nevarez JM, Dawid S, et al. 2019. DNA methylation from a Type I restriction modification system influences gene expression and virulence in Streptococcus pyogenes. *PLoS Pathog* 15:e1007841
66. Nye TM, van Gijtenbeek LA, Stevens AG, Schroeder JW, Randall JR, et al. 2020. Methyltransferase DnmA is responsible for genome-wide N6-methyladenosine modifications at non-palindromic recognition sites in Bacillus subtilis. *Nucleic Acids Res*
67. Oliveira PH, Ribis JW, Garrett EM, Trzilova D, Kim A, et al. 2020. Epigenomic characterization of Clostridioides difficile finds a conserved DNA methyltransferase that mediates sporulation and pathogenesis. *Nat Microbiol* 5:166-80
68. Phillips ZN, Tram G, Seib KL, Atack JM. 2019. Phase-variable bacterial loci: how bacteria gamble to maximise fitness in changing environments. *Biochem Soc Trans* 47:1131-41

69. Piekarowicz A, Brzezinski R. 1980. Cleavage and methylation of DNA by the restriction endonuclease HinfIII isolated from Haemophilus influenzae Rf. *J Mol Biol* 144:415-29

70. Pingoud A, Wilson GG, Wende W. 2014. Type II restriction endonucleases--a historical perspective and more. *Nucleic Acids Res* 42:7489-527

71. Pisciotta A, Manteca A, Alduina R. 2018. The SCO1731 methyltransferase modulates actinorhodin production and morphological differentiation of Streptomyces coelicolor A3(2). *Sci Rep* 8:13686

72. Procopio RE, Silva IR, Martins MK, Azevedo JL, Araujo JM. 2012. Antibiotics produced by Streptomyces. *Braz J Infect Dis* 16:466-71

73. Rao DN, Dryden DT, Bheemanaik S. 2014. Type III restriction-modification enzymes: a historical perspective. *Nucleic Acids Res* 42:45-55

74. Roberts RJ, Belfort M, Bestor T, Bhagwat AS, Bickle TA, et al. 2003. A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Res* 31:1805-12

75. Sanchez-Romero MA, Casadesus J. 2020. The bacterial epigenome. *Nat Rev Microbiol* 18:7-20

76. Sanchez-Romero MA, Cota I, Casadesus J. 2015. DNA methylation in bacteria: from the methyl group to the methylome. *Curr Opin Microbiol* 25:9-16

77. Seib KL, Jen FE, Tan A, Scott AL, Kumar R, et al. 2015. Specificity of the ModA11, ModA12 and ModD1 epigenetic regulator N(6)-adenine DNA methyltransferases of Neisseria meningitidis. *Nucleic Acids Res* 43:4150-62

78. Sekizaki T, Osaki M, Takamatsu D, Shimoji Y. 2001. Distribution of the SsuDAT1I restriction-modification system among different serotypes of Streptococcus suis. *J Bacteriol* 183:5436-40

79. Seshasayee AS, Singh P, Krishna S. 2012. Context-dependent conservation of DNA methyltransferases in bacteria. *Nucleic Acids Res* 40:7066-73

80. Shell SS, Prestwich EG, Baek SH, Shah RR, Sassetti CM, et al. 2013. DNA methylation impacts gene expression and ensures hypoxic survival of Mycobacterium tuberculosis. *PLoS Pathog* 9:e1003419

81. Smith ZD, Shi J, Gu H, Donaghey J, Clement K, et al. 2017. Epigenetic restriction of extraembryonic lineages mirrors the somatic transition to cancer. *Nature* 549:543-7

82. Srikhanta YN, Dowideit SJ, Edwards JL, Falsetta ML, Wu HJ, et al. 2009. Phasevarions mediate random switching of gene expression in pathogenic Neisseria. *PLoS Pathog* 5:e1000400

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

84. Srikhanta YN, Fung KY, Pollock GL, Bennett-Wood V, Howden BP, Hartland EL. 2017. Phasevarion-Regulated Virulence in the Emerging Pediatric Pathogen Kingella kingae. *Infect Immun* 85

85. Srikhanta YN, Gorrell RJ, Steen JA, Gawthorne JA, Kwok T, et al. 2011. Phasevarion mediated epigenetic gene regulation in Helicobacter pylori. *PLoS One* 6:e27569

86.   Srikhanta YN, Maguire TL, Stacey KJ, Grimmond SM, Jennings MP. 2005. The phasevarion: a genetic system controlling coordinated, random switching of expression of multiple genes. *Proc Natl Acad Sci U S A* 102:5547-51

87.   Sumby P, Smith MC. 2002. Genetics of the phage growth limitation (Pgl) system of Streptomyces coelicolor A3(2). *Mol Microbiol* 44:489-500

88.   Suri B, Bickle TA. 1985. EcoA: the first member of a new family of type I restriction modification systems. Gene organization and enzymatic activities. *J Mol Biol* 186:77-85

89.   Taylor I, Patel J, Firman K, Kneale G. 1992. Purification and biochemical characterisation of the EcoR124 type I modification methylase. *Nucleic Acids Res* 20:179-86

90.   Tornheim JA, Dooley KE. 2019. The Global Landscape of Tuberculosis Therapeutics. *Annu Rev Med* 70:105-20

91.   Tsai MC, Chakravarty S, Zhu G, Xu J, Tanaka K, et al. 2006. Characterization of the tuberculous granuloma in murine and human lungs: cellular composition and relative tissue oxygen tension. *Cell Microbiol* 8:218-32

92.   van Belkum A, Scherer S, van Alphen L, Verbrugh H. 1998. Short-sequence DNA repeats in prokaryotic genomes. *Microbiol Mol Biol Rev* 62:275-93

93.   Waldron DE, Lindsay JA. 2006. Sau1: a novel lineage-specific type I restriction-modification system that blocks horizontal gene transfer into Staphylococcus aureus and between S. aureus isolates of different lineages. *J Bacteriol* 188:5578-85

94.   Wallecha A, Munster V, Correnti J, Chan T, van der Woude M. 2002. Dam- and OxyR-dependent phase variation of agn43: essential elements and evidence for a new role of DNA methylation. *J Bacteriol* 184:3338-47

95.   Weiser JN, Austrian R, Sreenivasan PK, Masure HR. 1994. Phase variation in pneumococcal opacity: relationship between colonial morphology and nasopharyngeal colonization. *Infect Immun* 62:2582-9

96.   Wilson GG, Murray NE. 1991. Restriction and modification systems. *Annu Rev Genet* 25:585-627

97.   Xu P, Alves JM, Kitten T, Brown A, Chen Z, et al. 2007. Genome of the opportunistic pathogen Streptococcus sanguinis. *J Bacteriol* 189:3166-75

98.   Yague P, Lopez-Garcia MT, Rioseras B, Sanchez J, Manteca A. 2013. Pre-sporulation stages of Streptomyces differentiation: state-of-the-art and future perspectives. *FEMS Microbiol Lett* 342:79-88

99.   Zhou W, Dinh HQ, Ramjan Z, Weisenberger DJ, Nicolet CM, et al. 2018. DNA methylation loss in late-replicating domains is linked to mitotic cell division. *Nat Genet* 50:591-602

CHAPTER II

DNA Methylation from a Type I Restriction Modification System Influences Gene Expression and Virulence in *Streptococcus pyogenes*

Abstract

DNA methylation is pervasive across all domains of life. In bacteria, the presence of N6-methyladenosine (m6A) has been detected among diverse species, yet the contribution of m6A to the regulation of gene expression is unclear in many organisms. Here we investigated the impact of DNA methylation on gene expression and virulence within the human pathogen *Streptococcus pyogenes,* or Group A Streptococcus. Single Molecule Real-Time sequencing and subsequent methylation analysis identified 412 putative m6A sites throughout the 1.8 Mb genome. Deletion of the Restriction, Specificity, and Methylation gene subunits (ΔRSM strain) of a putative Type I restriction modification system lost all detectable m6A at the recognition sites and failed to prevent transformation with foreign-methylated DNA. RNA-sequencing identified 20 genes out of 1,895 predicted coding regions with significantly different gene expression. All of the differentially expressed genes were down regulated in the ΔRSM strain relative to the parent strain. Importantly, we found that the presence of m6A DNA modifications affected expression of Mga, a master transcriptional regulator for multiple virulence genes, surface adhesins, and immune-evasion factors in *S. pyogenes*. Using a murine subcutaneous infection model, mice infected with the ΔRSM strain exhibited an enhanced host immune response with larger skin lesions and increased levels of pro-

inflammatory cytokines compared to mice infected with the parent or complementedmutant strains, suggesting alterations in m6A methylation influence virulence. Further, we found that the ΔRSM strain showed poor survival within human neutrophils and reduced adherence to human epithelial cells. These results demonstrate that, in addition to restriction of foreign DNA, gram-positive bacteria also use restriction modification systems to regulate the expression of gene networks important for virulence.

Author Summary

DNA methylation is common among many bacterial species, yet the contribution of DNA methylation to the regulation of gene expression is unclear outside of a limited number of gram-negative species. We characterized sites of DNA methylation throughout the genome of the gram-positive pathogen *Streptococcus pyogenes* or Group A Streptococcus. We determined that the gene products of a functional restriction modification system are responsible for genome-wide m6A. The mutant strain lacking DNA methylation showed altered gene expression compared to the parent strain, with several genes important for causing human disease down regulated. Furthermore, we showed that the mutant strain lacking DNA methylation exhibited altered virulence properties compared to the parent strain using various models of pathogenesis. The mutant strain was attenuated for both survival within human neutrophils and adherence to human epithelial cells, and was unable to suppress the host immune response in a murine subcutaneous infection model. Together, these results show that bacterial m6A contributes to differential gene expression and influences the ability of Group A Streptococcus to cause disease. DNA methylation is a conserved feature among bacteria and may represent a potential target for intervention in effort to interfere with the ability of bacteria to cause human disease.

Introduction

DNA methylation has been shown to regulate diverse pathways across all domains of life [1]. In eukaryotes, cytosine methylation regulates developmental gene expression

and aberrant DNA methylation patterns have been implicated in many disease states, including cancer [2, 3]. Although studied in a limited number of prokaryotic organisms, DNA methylation has been implicated in a myriad of cellular processes, including protection from the invasion of foreign DNA, cell cycle regulation, DNA mismatch repair, and the regulation of gene expression [4]. It was recently shown that within the genomes of over 200 prokaryotes surveyed greater than 90% contained N6-methyladenosine (m6A), N4-methylcytosine (m4C), or 5-methylcytosine modifications (m5C) [5]. These results demonstrate that DNA methylation among prokaryotes is more pervasive than originally anticipated. What remains uncertain is if DNA methylation imparts any regulatory controls influencing virulence properties or other phenotypes amongst the array of diverse prokaryotic species.

DNA methylation in bacteria has been well characterized in the context of restriction modification (RM) systems [4, 5]. RM systems are a mechanism of bacterial host defense to prevent the invasion of foreign DNA. RM systems are generally comprised of a site-specific restriction endonuclease (REase), methyltransferase (MTase), and, in some cases, a specificity subunit that together form a protein complex that cleaves foreign DNA after it enters the cell. Methylation of the host DNA at the same recognition site serves to safeguard the host chromosome from cleavage. In addition to RM systems, DNA can also be methylated by orphan MTases. Orphan MTases methylate DNA in site-specific sequences and lack an active cognate endonuclease [5, 6]. In bacteria, the two most well studied orphan MTases are *Escherichia coli* DNA adenosine methyltransferase (Dam) and *Caulobacter crescentus* cell cycle regulated methyltransferase (CcrM) [5, 6]. Site-specific DNA methylation by Dam and CcrM has been shown to regulate DNA mismatch repair, cell cycle progression, origin sequestration, and gene expression, demonstrating that DNA methylation imparts critical regulatory functions [6].

Despite the importance of RM systems and orphan MTases, the lack of genome-wide detection tools has hindered the identification of DNA base modifications and characterization of the physiological consequences resulting from MTase inactivation in bacteria. The use of methylation-sensitive restriction endonucleases to identify sites of DNA base modifications is limited by the sequence specificity of the recognition site,

potentially missing many base modifications that could occur outside of a particular sequence context ([5] and references therein). While bisulfite sequencing allows for genome-wide detection of m5C in sequence specific-contexts, no such genome-wide detection tool has been available for the detection of m6A or m4C until the recent advent of Pacific Biosciences (PacBio) Single Molecule Real-Time (SMRT) sequencing platform [7-11]. SMRT sequencing relies on differences in DNA polymerase kinetics to detect base modifications in the template strand in a sequence-context specific manner without *a priori* knowledge of the modification.

Our group previously used the PacBio SMRT sequencing platform to complete whole genome sequencing and reference genome assembly of two strains of the bacterial human pathogen *Streptococcus pyogenes*, or Group A Streptococcus (GAS) [12, 13]. *S. pyogenes* causes a wide variety of human infections, ranging from the relatively common streptococcal pharyngitis and cellulitis to the relatively uncommon, but severe, streptococcal toxic shock syndrome and necrotizing fasciitis, which have high morbidity and mortality rates [14-16]. *S. pyogenes* is a model bacterial pathogen, not only for the infections it produces, but also for the great diversity of toxins and virulence factors expressed by the organism and the highly complex nature of regulatory mechanisms employed to control virulence factor expression [14, 16-18]. Indeed, *S. pyogenes* utilizes over 30 recognized transcriptional regulatory proteins and 13 two-component regulatory systems to coordinate virulence factor expression in response to varying environmental signals (e.g., carbohydrate availability, temperature, pH, oxygen tension, salt concentrations, osmolality, etc.), growth phase, intracellular metabolite concentrations, and signaling pheromones involved in quorum sensing [17, 18]. DNA methylation has not been previously investigated as a significant mechanism influencing virulence factor expression within *S. pyogenes*, and DNA methylation may represent an unrecognized target for therapeutic intervention to help prevent or treat severe streptococcal disease.

In this study, we show that in *S. pyogenes* strain MEW123, a representative derivative of a serotype M28 clinical pharyngitis isolate, the active Type I RM system SpyMEW123I is responsible for the bipartite m6A motif identified throughout the genome. We show that deletion of the RM system and subsequent loss of m6A from *S.*

pyogenes results in the down regulation of a distinct set of operons involved in streptococcal virulence. Importantly, our study shows that methylation by a Type I RM system correlates with differential expression of Mga, a major transcriptional regulator of multiple virulence factors, surface adhesins, and immune evasion factors in *S. pyogenes*. The results presented here demonstrate that RM systems can integrate their methylation signal to influence the expression of gene networks important for bacterial virulence.

Results

SMRT sequencing and methylation analysis identifies m6A modifications in a bipartite recognition sequence in the *S. pyogenes* genome. Previously we completed whole genome assembly using PacBio SMRT sequencing with *S. pyogenes* strain MEW123, a representative serotype M28 isolate used by our group to investigate streptococcal mucosal colonization [12] (for strain list refer to **Table 2.1**). To begin our investigation, we performed methylation analysis of the SMRT sequencing data. We identified m6A DNA base modifications in the MEW123 genome at the consensus sequence 5' GCANNNNNTTYG and its corresponding partner motif 5' CRAANNNNNNTGC, consistent with m6A modification motifs previously reported by Blow *et al.* (**Table 2.2**) [5]. Within the MEW123 genome, 412 occurrences of each m6A site within the bipartite recognition motif were identified; the majority occurred in predicted coding (92%) and intergenic (6%) regions of the MEW123 genome. The bipartite recognition motif is characteristic of Type I RM systems, which are typically comprised of three separate subunits, including a restriction endonuclease, a specificity subunit, and a methyltransferase subunit, that act together as a single protein complex and typically act at large distances from the methylation site. The RM system annotation pipeline used in Blow *et al.* identified the putative Type I restriction modification system, annotated as SpyMEW123I, consisting of a three-gene cluster with separate restriction endonuclease (*hsdR*), specificity (*hsdS*), and methyltransferase (*hsdM*) genes, as a predicted match for modification of the identified m6A motif in *S. pyogenes* [5, 19] (**Fig 2.1A** and **Fig 2.1B**). This three-gene cluster exhibits high amino acid sequence homology to the Type I RM system identified in *S. pyogenes* SF370 at Spy_1904

(*hsdR*), Spy_1905 (*hsdS*), and Spy_1906 (*hsdM*), with 99%, 87%, and 99% identity, respectively [20]. This Type I RM system is present in virtually all sequenced *S. pyogenes* strains to date, with rare exception reported in some *emm1* strains from Japan with spontaneous deletion of a two-component regulatory system and the adjacent Type I RM system [21]. Notably, we did not detect the 5mC modifications at C^mCNGG reported by Euler *et al.* in our PacBio SMRT sequencing results, which is not surprising given the MTase, M.SpyI, is absent from the *S. pyogenes* M28 serotype [22]. The REase and MTase activities of SpyMEW123I are annotated as R.SpyMEW123I and M.SpyMEW123I, respectively.

M.SpyMEW123I is responsible for m6A modifications in the *S. pyogenes* genome. To determine if the SpyMEW123I RM system was responsible for the observed m6A modifications in strain MEW123, an in-frame deletion mutation was constructed using a plasmid vector designed for allelic replacement (pGCP213) as previously described [26] (**Table 2.1** and **Fig 2.1A**). Approximately 95% of the three-gene sequence encoding the *hsdR*, *hsdS*, and *hsdM* genes was deleted producing strain MEW513 (referred to as ΔRSM); the in-frame deletion was confirmed by PCR amplification and Sanger DNA sequencing (**Table 2.1**). Growth of the MEW123 parent strain, referred to as wild-type (WT) and the ΔRSM mutant were not significantly different in rate or final growth density when measured in either the nutrient rich Todd-Hewitt medium with 0.2% yeast extract (THY broth) or the low-carbohydrate C-medium (**Fig 2.8**). To confirm a reduction in m6A base modifications and to determine the sequence context lacking m6A base modifications in the ΔRSM strain, genomic DNA was isolated and sequenced via PacBio SMRT sequencing. Modification analysis showed loss of detectable m6A base modifications at 5' GCANNNNNTTYG and 5' CRAANNNNNTGC sites, demonstrating that streptococci with a SpyMEW123I deletion no longer have m6A DNA base modifications at the consensus sequence identified in the WT strain (**Fig 2.1B, Table 2.3**). A number of additional methylation events were identified in MEW513; however, these occurred at far lower frequencies compared to the modifications at the consensus sequences in the parent strain and the quality of the read scores (Mod QV) were low compared to the RSM-dependent modifications. Based on these low quality read

scores, we feel it is unlikely that these additional modifications reflect compensatory methylation events. Furthermore, SMRT sequencing of the MEW513 genome did not identify any unforeseen mutations outside of the in-frame deletion within *hsdRSM* that we anticipated.

To further confirm that the MTase component of the RSM gene cluster, M.SpyMEW123I, was indeed responsible for producing m6A DNA modifications, genomic DNA was harvested from the WT and the ΔRSM strain and spotted onto a nitrocellulose membrane for immunodetection using an α-m6A antibody. We found that the α-m6A signal was substantially reduced in genomic DNA blots from the ΔRSM strain compared to the WT parent, suggesting a significant and near complete reduction in m6A base modifications in the ΔRSM strain (**Fig 2.1C**). Complementation *in trans* of the ΔRSM mutant with a plasmid encoded copy of the three gene cluster (*hsdRSM*) produced strain MEW552 (referred to as ΔRSM/pRSM) and successfully restored detection of the α-m6A signal to levels comparable to the WT strain (**Fig 2.1C**). These results demonstrate that the MTase activity of SpyMEW123I is responsible for base modifications at 5' GCANNNNNTTYG and 5' CRAANNNNNTGC sites *in vivo*.

The SpyMEW123I RM system influences *S. pyogenes* transformation efficiency demonstrating functional restriction of foreign DNA acceptance. Deletion of the three-gene cluster, *hsdRSM*, containing the predicted endonuclease, specificity, and methylation gene subunits abolished m6A base modifications in the ΔRSM mutant strain. In Type I RM systems, DNA cleavage is dependent on the MTase and specificity subunits, in addition to the REase subunits which are often independently regulated by a separate promoter [29]. Fully unmethylated recognition motifs induce REase activity that results in DNA cleavage typically between two fully unmethylated motifs at sites distant from the recognition sequence; this distance may range from 40 base pairs to several kilobases away from the RM site. Type I MTases can function to add m6A *de novo* on fully unmethylated DNA or act as maintenance MTases at hemi-methylated recognition sites [29-31]. Additional mechanisms also protect DNA from restriction, including proteolysis of the REase subunits or protection by DNA binding proteins that

can protect unmethylated sites from cleavage in the host chromosome [32]. To establish the function of the REase component of SpyMEW123I, a transformation efficiency assay was performed using pJoy3 plasmid DNA methylated in an *E. coli* host (**Table 2.1**). This 6.3 kb plasmid contains eight predicted Dam MTase RM sites (5' GATC) and is delivered in its native double-stranded circular form via electroporation into electrocompetent *S. pyogenes* where the plasmid is maintained and replicates extrachromosomally [27]. In addition to testing the effect of deleting the entire *hsdRSM* gene cluster in the ΔRSM mutant strain, we constructed an additional strain derivative of MEW123 with a spectinomycin-resistance cassette disrupting the *hsdR* REase gene subunit alone producing strain MEW489 (referred to as ΩRE, **Table 2.1**). If the SpyMEW123I RM system has true restriction enzyme activity to foreign-modified DNA, then we would expect that inactivating the *hsdR* gene subunit, either individually or within the entire RSM gene cluster, would enhance the transformation efficiency of the plasmid. Indeed, we found that the rates of transformation with foreign-methylated plasmid DNA increased significantly for both the ΔRSM mutant and the ΩRE mutant strains compared to the WT parent strain, providing evidence that the restriction endonuclease component of SpyMEW123I is active and functional (**Fig 2.2A**). We were unable to compare our complementation strain ΔRSM/pRSM for transformation efficiency as this strain already carries the pJoy3 plasmid encoding the *hsdRSM* gene cluster. As a control, we undertook transformation of a MEW123 mutant in the gene encoding the C5a peptidase, *scpA* (strain 489 or Ω*scpA*), as mutants in this gene would not be expected to show enhanced transformation efficiency; as expected, the transformation efficiency of Ω*scpA* was not significantly different than the WT (**Fig 2.2A**). Interestingly, inactivation of the endonuclease subunit *hsdR* alone in the ΩRE mutant strain conferred significantly greater transformation efficiency than that observed in the ΔRSM mutant (**Fig 2.2A**). In many Type I RM systems the restriction subunit is generally under control of a separate promoter from the specificity and methylation subunits in the RSM gene cluster [29]. We found that the α-m6A signal generated by dot blot of genomic DNA from the ΩRE strain was intermediate in intensity between the WT and ΔRSM strains (**Fig 2.1C**). This result suggests that the methyltransferase subunit was still functional in the ΩRE strain, but that there may have been some degree of

41

polar effect from the spectinomycin-resistance cassette used to inactivate *hsdR* that was reducing transcription of the *hsdS* and *hsdM* gene products compared to WT levels. We speculate that the residual functional activities of the specificity and methyltransferase subunits in the ΩRE mutant strain, even though less than WT levels, may have conferred additional stability to the incoming foreign-methylated plasmid DNA, possibly offering protection from other minor endonucleases, thereby enhancing overall transformation efficiency.

Variation in m6A base modification occurrence at RM recognition sites identified by SMRT sequencing in the *S. pyogenes* genome. As discussed above, the MTase activity of Type I RM systems may function to maintain the state of hemi-methylated or fully methylated DNA, whereas REase cleavage only occurs on fully unmethylated DNA. Thus, RM sites can exist in the genome in a hemi-methylated state while still conferring protection from digestion [29]. Having shown that SpyMEW123I functions as an active RM system, we sought to establish the fraction of reads that were called as methylated at each recognition site. Our analysis of sequencing data from WT *S. pyogenes* MEW123 found substantial variation in the fraction of sequencing reads modified at RM sites (**Fig 2.2B**). The fraction of reads called as modified at a given RM site did not appear to be dependent on orientation or genome position. Of the m6A modifications called at RM sites, 4.9% of sites were called as m6A modified in less than 50% of sequencing reads, 23.7% of sites were called as modified in between 50-75% of sequencing reads, and finally 71.4% of sites were called as modified at greater than 75% of aligned reads. Previous studies have also reported heterogeneity in the frequency of SMRT sequencing reads with base modifications; it has been hypothesized that these differences are due to timing in DNA replication and subsequent methylation [33-35]. Whether there is a temporal component accounting for the heterogeneity in m6A DNA modifications, and whether this impacts other functions of m6A modifications, such as in influencing gene transcription in *S. pyogenes*, is unknown. Given the heterogeneity observed in the fraction of reads called as m6A methylated, we hypothesized that m6A modifications produced by the SpyMEW123I RM

system might have additional functions outside of host protection from foreign DNA prompting the experiments below.

RNA-sequencing shows that deletion of the SpyMEW123I RM system results in the down regulation of several transcripts involved in streptococcal immune evasion and adherence. In addition to functioning in RM systems, m6A base modifications from orphan MTases have been shown to function in cell cycle regulation, DNA mismatch repair, and the regulation of transcription [4]. In the pathogenic *Escherichia coli* serotype O104:H4 strain C227-11 associated with hemolytic uremic syndrome, deletion of the φStx104 RM system results in the differential expression of over 38% of the genes, including genes involved in motility, cell projection, and cation transport [33]. Mismatch repair is not coupled to methylation in *S. pyogenes* or most other gram-positive bacteria [36]. Therefore, we asked if m6A originating from the SpyMEW123I RM system in *S. pyogenes* might have additional functions outside of host defense from foreign DNA. We isolated RNA from streptococcal cells during mid-exponential growth phase in C media broth culture from WT and ΔRSM strains followed by RNA-sequencing. The results of the differential expression analysis showed that 20 genes were differentially expressed in the ΔRSM strain compared to WT (adjusted p. val < 0.05, $\log_2$ fold change >1, data set available at NCBI repository). Interestingly, all 20 genes were down regulated in ΔRSM relative to WT suggesting a common regulatory mechanism (**Fig 2.3A** and **2.3B**, **Table 2.4**). The three genes (*hsdRSM*) of the SpyMEW123I RM gene cluster showed the greatest $\log_2$ fold change in expression of -10.8, -10.7, and -11.7, respectively, which was expected because these genes were deleted in the ΔRSM strain.

The majority of the differentially expressed genes are located in approximately 6 separate operons or gene clusters as indicated in Table 2.4. Interestingly, several of these gene groups are transcriptionally regulated, at least in large part, by activity of the multiple gene regulator protein, Mga [37-39]. During mid-exponential growth phase, Mga acts as a transcriptional activator to regulate a core set of virulence factors at the *mga* locus [37]. The *mga* locus consists of several components: a) the M protein (*emm* gene) a major surface protein involved in resistance to phagocytosis and intracellular

killing by neutrophils and used to distinguish *S. pyogenes* isolates, b) a fibronectin-binding protein that binds host complement regulator factors, c) an *emm*-like protein that binds IgG and fibrinogen, d) the C5a peptidase (ScpA) which cleaves C5a chemotaxin, e) the *enn* protein that binds IgA, and f) the *mga* gene itself. All genes at the *mga* locus displayed $\log_2$ fold changes ranging from -1.2 to -8.7 in the ΔRSM strain relative to WT (**Fig 2.3A**). To confirm this differential expression, we again isolated total RNA from strains during mid-exponential growth phase in C media broth culture and performed quantitative RT-PCR for detection of transcripts *mga*, *emm28*, and *scpA* (**Fig 2.3B**). Consistent with the RNA-seq results, the qRT-PCR results showed that these genes were significantly down regulated in the ΔRSM strain, with approximately 5-fold to over 300-fold decreased expression in the ΔRSM strain relative to WT (**Fig 2.3B**). Complementation *in trans* in the ΔRSM/pRSM strain restored transcript expression patterns similar to WT values. Deletion of the *mga* gene produced qRT-PCR results in a similar trend to the ΔRSM strain for the examined transcripts, with significantly decreased detection of *emm28* and *scpA* transcripts; *mga* transcript was not detected in the Δ*mga* strain (**Fig 2.3B**). Examination of these transcripts in the *hsdR* insertional inactivation mutant ΩRE showed transcript detection of *mga* and *emm28* comparable to WT levels, with detection of *scpA* transcript approximately four to five-fold of WT levels. This transcript pattern was very different than those of the ΔRSM and Δ*mga* mutant strains and more similar to the WT pattern. Even though the spectinomycin resistance cassette insertion into *hsdR* may have produced some polar effect with slightly decreased methyltransferase activity as noted on the anti-m6A dot blot (**Fig 2.1C**), it seems sufficient residual m6A base modifications persisted to not significantly disrupt gene expression (**Fig 2.3B**). Taken together, these results from RNA sequencing and qRT-PCR provide evidence that m6A base modifications correlate with patterns of differential gene expression in *S. pyogenes*, including those of several recognized virulence factors and major regulators of virulence gene expression.

Disruption of m6A DNA modifications enhances the host inflammatory response to streptococcal infection in a murine subcutaneous ulcer model. Given that the genes in the Mga regulon were significantly down regulated in the ΔRSM strain relative

to WT, we were interested in determining the impact of disrupting m6A DNA modifications on *S. pyogenes* virulence using a murine subcutaneous infection model [40, 41]. C57BL/6J mice were inoculated at the shaved flank with 1 x 10^7 CFUs of either MEW123 (WT) or the ΔRSM mutant strain and resulting skin ulcers were photographed daily for sizing the skin ulcer area. As shown in Fig 2.4A, there was no significant difference in skin lesion size at day two post-infection in comparison of the mice infected with either the WT or the ΔRSM strains. However, by three to four days post-infection, and for the remainder of the experiment, the skin lesions of mice infected with the ΔRSM strain were significantly larger than those of mice infected by the WT strain (**Fig 2.4A**). No strain caused a lethal infection among any of the mice with the 1 x 10^7 CFU inoculum. Representative images of skin lesions for mice infected with the WT, the ΔRSM strain, and the complemented ΔRSM/pRSM strain over time are shown in Fig 2.4B, with skin lesions of mice infected with the ΔRSM strain notably larger on average at 4 and 6 days compared to those of mice infected with the WT or complemented strain. Complementation of the ΔRSM mutation *in trans* by strain MEW552 (ΔRSM/pRSM) produced murine skin lesions smaller than the ΔRSM mutant but not significantly different than the WT strain throughout the duration of the experiment (**Fig 2.4C**).

Skin lesion sizes reached a mean peak size at four to six days post infection. To determine if the difference in skin lesion size correlated with the concentration of viable streptococci at the site of infection, the skin lesions of mice were dissected and homogenized at day four post-infection to obtain viable CFU counts. Upon dissection, we made the observation that skin lesions from mice infected with the ΔRSM strain were grossly more purulent than lesions of mice infected by the WT and complemented ΔRSM/pRSM strains. The skin lesions contained on average CFU counts of approximately 1 x 10^6 to 1 x 10^7 CFUs; while there was a slight trend to higher CFU counts on day four post-infection for the ΔRSM streptococci compared to the WT and complemented strain CFUs, there were no statistically significant differences in CFU counts between these groups (**Fig 2.9**). We noted that skin lesions of mice infected with the WT and complemented strain ΔRSM/pRSM strains seemed to heal more quickly than those of mice infected with the ΔRSM strain (**Fig 2.4C**).

With the subcutaneous ulcer model, skin lesion size tends to correlate closely with the degree of the host immune response, with particular regards to the neutrophil influx [40, 41]. To compare the inflammatory response in skin lesions of mice infected with the WT and the ΔRSM strain, we performed skin biopsies for cytokine analysis and histologic examination at six-days post-infection; this time point was chosen as it was the time point with the greatest difference in skin lesion size between the experimental groups. Measurements of interleukin-1 beta (IL-1β), interleukin-6 (IL-6), interleukin-17A (IL-17A), and tumor necrosis factor alpha (TNFα), were obtained as evidence of pro-inflammatory cytokine activity. Cytokine concentrations for all four cytokines measured were significantly greater from mice infected with WT streptococci than mice mock-infected with sterile phosphate-buffered saline (PBS) (**Fig 2.5A**). Cytokine concentrations from mice infected with the ΔRSM strain were significantly greater than mock-infected or mice infected with the WT strain (**Fig 2.5A**). Furthermore, histologic analysis of skin lesions shows predominantly increased neutrophil influx, but also a modest increase in the number of macrophages in the subcutaneous tissue of mice infected with the ΔRSM strain compared with WT (**Fig 2.5B**). Infiltration of T lymphocytes was not appreciably different between skin lesions of mice infected with WT or the ΔRSM strain (**Fig 2.5B**). Cytokines IL-6 and IL-17A, in particular, are important for coordinating neutrophil trafficking to areas of infection [42-44]. Our results in mice infected with the ΔRSM strain showing enhanced pro-inflammatory cytokine detection, increased neutrophil infiltration, and larger skin lesions, suggests an effect of altered gene transcription patterns in the ΔRSM strain and a more robust host inflammatory response compared to mice infected with the WT parent strain. Given the known association of several of the streptococcal gene transcripts down regulated in the ΔRSM strain, including *mga*, *emm28*, and *scpA*, with immune evasion properties, we hypothesized that m6A DNA modifications and proper regulation of gene expression are important contributors to immune evasion strategies and/or disruption of host immune responses by *S. pyogenes*.

To determine if the loss of specific virulence factors recapitulates the phenotype of the ΔRSM strain in the murine subcutaneous ulcer model, we infected mice with derivatives of strain MEW123 with in-frame deletions of *mga* (strain MEW480, ΔMga),

and spectinomycin-resistance cassette disruption mutations of *emm28* (strain 409, Ω*emm28*) and *scpA* (strain 380, Ω*scpA*). As shown in Fig 2.4C, infection of mice by the ∆Mga strain produced skin lesions significantly larger than the WT strain and comparable to the ∆RSM strain in size throughout the experiment. Infection by the Ω*emm28* strain was not statistically different than the WT strain at day 2 and day 4 post-infection; however, by day 6 and day 8 post-infection, the Ω*emm28* strain produced lesions that were statistically significantly larger than the WT (**Fig 2.4C**). Infection of mice by the Ω*scpA* strain produced the widest range of murine skin lesion sizes, with some mice having very large lesions following infection (**Fig 2.4C**); however, at no time point was the average size of the lesions produced by the Ω*scpA* strain statistically different than WT. Overall, these results suggest that the presence of m6A DNA base modifications produced by M.SpyMEW123 activity correlate with differential transcriptional expression of several *S. pyogenes* virulence factors, especially those within the Mga operon, and that these seem to influence host-pathogen interactions at the site of infection.

Disruption of m6A DNA modifications inhibits streptococcal survival within human neutrophils. A major function of the *S. pyogenes* M protein is to promote streptococcal survival, resisting killing by human leukocytes by interfering with bactericidal activity within neutrophils following phagocytosis [45, 46]. Staali *et al.* found that *S. pyogenes* strains with or without M protein underwent phagocytosis by neutrophils to similar levels, but only strains expressing M protein survived intracellularly whereas strains lacking M protein expression were rapidly killed [45]. Given our findings that elimination of m6A DNA modifications was associated with decreased transcript expression for *mga* and *emm28*, we wished to compare survival within human neutrophils. Purified human neutrophils were incubated with WT or ∆RSM *S. pyogenes* strains using a neutrophil bactericidal assay similar to a previous report [45]. Briefly, streptococci and neutrophils were mixed together allowing the neutrophils to internalize *S. pyogenes* strains followed by elimination of extracellular bacteria with penicillin and gentamicin. It was previously determined that there was no significant difference in susceptibility to penicillin and gentamicin at the high concentrations used in these

experiments between the WT or ΔRSM strains (**Fig 2.8**). Streptococcus surviving within neutrophils were liberated by treatment with the detergent saponin and plated for viable CFUs. As shown in Figure 2.6, we utilized serotype M14 HSC5 and a derivative strain with disruption in the M14 *emm* gene (Ω*emm14*) as positive and negative controls, respectively. As expected, the Ω*emm14* mutant was significantly attenuated for intracellular survival within neutrophils compared to the M14 parent strain (**Fig 2.6**). Similarly, we compared survival of the MEW123 parent strain (M28) and its cognate strain with disruption of the M28 *emm* gene (Ω*emm28)* or the ΔRSM mutant. We found that both the Ω*emm28* and the ΔRSM mutant were significantly attenuated for intracellular survival compared to the M28 parent strain, further confirming the role of M protein in promoting intracellular neutrophil survival by the serotype M28 MEW123 strain, in addition to demonstrating correlation of m6A DNA base modifications with differential expression of M protein (**Fig 2.6**). These results provide further support for m6A DNA base modifications in *S. pyogenes* as important for promoting streptococcal virulence, possibly by influencing virulence factor expression.

Disruption of m6A DNA modifications inhibits adherence to human vaginal epithelial cells *in vitro*, but does not appear to impair carriage *in vivo* in a murine vaginal colonization model. From the RNA-seq results we found that the ΔRSM strain had significantly decreased transcript expression of several recognized and known adhesin proteins, including M28, M-like protein, collagen-binding protein, and fibronectin-binding proteins, as well as several hypothetical surface proteins [38, 47-49]. As a group, serotype M28 *S. pyogenes* are overrepresented in cases of human infection within the female urogenital tract, including vulvovaginitis and puerperal sepsis (a.k.a. "childbed fever") [50-53]. Serotype M28 *S. pyogenes* have a particular predilection for cervical and vaginal epithelium due to surface proteins, including protein R28 among others, which may explain the overrepresentation of this serotype with infections in this niche [15, 54]. Therefore, we asked if m6A DNA modifications influenced adherence of the serotype M28 MEW123 strain to human vaginal epithelial cells.

As shown in Figure 2.7A, disruption of m6A DNA modifications in the ΔRSM strain was indeed associated with significantly decreased adherence to human vaginal epithelial cells *in vitro* compared to the WT parent strain. The attenuation in vaginal epithelial cell adherence by the ΔRSM strain was comparable to a strain lacking expression of the M protein (Ω*emm28*), suggesting that decreased expression of M protein, among other adhesins, by the ΔRSM strain was at least partly responsible for decreased adherence (**Fig 2.7A**). To determine if impaired adherence to human vaginal cells *in vitro* translated to impaired vaginal mucosal colonization *in vivo*, we utilized a murine vaginal model and compared streptococcal carriage burdens over time [40, 55]. In contrast to the results of the *in vitro* adherence assay, using the murine vaginal carriage model we found no significant difference in vaginal streptococcal burdens in comparison of mice inoculated with either the WT or the ΔRSM strains over the course of the 28-day experiment (**Fig 2.7B**). Given that human cells are the natural hosts of *S. pyogenes*, this may be an example of the human-restricted nature of *S. pyogenes* in which a murine model cannot adequately replicate the natural human environment in which this pathogen evolved to survive. Nevertheless, our overall results showed several key differences in virulence phenotypes correlating with alterations in gene transcription associated with streptococcal m6A DNA methylation.

Discussion

In this report, we provide evidence that m6A DNA base modifications influence gene transcription patterns and overall virulence properties in a major gram-positive bacterial pathogen of humans, *S. pyogenes*. The *S. pyogenes* RM system, SpyMEW123I, is a Type I RM system and is responsible for the majority of m6A base modifications distributed throughout the *S. pyogenes* genome. The target consensus sequences identified by our study, 5' GCANNNNNTTYG and its corresponding partner motif 5' CRAANNNNNNTGC, were consistent with m6A motifs identified in *S. pyogenes* previously reported by Blow *et al* [5]. We found approximately 412 occurrences of each m6A site with the majority found within coding regions. Interestingly, we found that not all m6A sequence motifs were consistently modified to the same extent; only about 70% of consensus sites were modified in at least 75% of sequencing reads, suggesting that

m6A modifications may be intermittently present with additional functions beyond simple protection from restriction, including influencing gene expression patterns based on timing of hemi- or full-methylation status. It is not known at this time whether all of the m6A sites, or only the sites within the intergenic regions, would participate in influencing transcriptional expression, but methylation events modifying access of transcriptional regulators to intergenic promoter regions would be a potential mechanism.

With the introduction of SMRT sequencing, groups have now identified m6A DNA modifications within a diversity of prokaryotes, including *E. coli*, *Campylobacter jejuni*, *Salmonella enterica* serovar Typhimurium, *Vibrio breoganii*, *Geobacter metallireducens*, *Chromhalobacter salexigens*, *Bacillus cereus*, and *Borrelia burgdorferi* [33-35, 56, 57]. Additional evidence of 5-methylcytosine (m5C) DNA modifications influencing transcriptional expression of multiple genes with an impact on several phenotypic traits has recently been described in *Helicobacter pylori*, further expanding the recognized influence of prokaryotic methylation modifications [58]. Some of the DNA modifications described have been linked to orphan MTases without an associated endonuclease, such as DNA Adenine Methyltransferase (Dam) of *S. enterica*, *E. coli*, and *Haemophilus influenzae* [56, 59, 60]. Uncoupling DNA methylation from restriction endonuclease protection is conceptually easier to envision with an orphan MTase, freeing the orphan MTase to have roles in DNA mismatch repair and influencing gene expression of potential virulence factors [6]. Indeed, Dam-dependent DNA modifications in *S. enterica* have been linked to alterations of gene expression and virulence [56]. However, two examples have recently been reported in *C. jejuni* and *B. burgdorferi* of intact RM systems also influencing gene expression patterns [34, 35]. Both of the RM systems in these organisms are representatives of Type IIG RM systems, which differ significantly from the Type I RM system described here for *S. pyogenes* in that they consist of a single polypeptide with both REase and MTase activity [34, 35, 61]. The effects on gene expression conferred by these systems in *C. jejuni* and *B. burgforferi* were noted by Casselli *et al*. to be more modest in terms of numbers of genes influenced by m6A base modifications when compared to the larger number of transcriptional changes found from the standalone activity of Dam MTase in Salmonella [34, 35, 56]. It would seem that with an intact RM system the conditions involved in determining gene expression is

more stringent and regulates a fewer number of genes than orphan MTases. DNA methylation from Type I RM systems has also been well-established in phase variation in a number of Gram-positive pathogens, including *Streptococcus pneumoniae*, *Streptococcus suis*, *Listeria monocytogenes*, and *Mycoplasma pulmonis*, which can have downstream effects on gene expression [62-65]. In phase variation, switching of specificity subunits of Type I RM systems results in cells with different sites of methylation within the population, which can create heterogeneity in gene expression. The role of methylation in phase variation differs from our findings here as we show that loss of methylation at a single site (i.e. not switching of specificity subunits to create methylation at diverse sites) results in the down regulation of a very defined subset of genes.

The M.SpyMEW123I MTase activity we describe here modifies 412 sites in the MEW123 genome, whereas Dam-modified recognition sites approximate 19,000 per chromosome [56]. Perhaps the context of the m6A recognition motif in a particular intergenic promoter region, combined with specific transcription factors sensitive to the presence or absence of m6A modifications, determines the specificity of which genes an intact RM system will influence. Our results reported here demonstrate that the *S. pyogenes* Type I RM system is functional as a protective mechanism with restricting uptake of foreign DNA (**Fig 2.2A**). Similar results were found by Okada *et al*. in a series of *emm1 S. pyogenes* isolates from Japan with spontaneous deletions in their Type I RM systems; isolates lacking the Type I RM system had significantly increased rates of transformation with foreign plasmid DNA [21]. While their study did not specifically address virulence properties of isolates lacking the RM system, the authors speculated that enhanced rates of DNA uptake and transformation exhibited by strains lacking REase activity may be beneficial by allowing uptake of potentially advantageous genes from the environment contributing to overall fitness.

Inactivation of the SpyMEW123I RM system was associated with significant dysregulation of gene transcript expression in broth culture, with 20 genes from at least six separate gene clusters/operons significantly down regulated (**Table 2.4**). Notable among the down regulated genes were the trans-acting regulator Mga, the M-like protein, M28 protein, C5a peptidase (ScpA), a cell surface protein, a collagen-like

surface protein (SclA), the Serum Opacity Factor (SOF), and a fibronectin-binding protein (SfbX). Most of these genes are regulated by the Mga transcriptional regulator in serotypes that have been investigated. Mga is a ubiquitous stand-alone regulator primarily active during exponential growth phase and is responsible for influencing expression of over 10% of the *S. pyogenes* genome, primarily genes involved in metabolism, but also many virulence factors including adhesins and surface proteins involved in immune evasion [37, 38]. Mga binds to upstream promoter regions to activate high-level transcription of genes in the Mga core regulon [66]. The majority of Mga-regulated promoters, including most of the genes in the core Mga regulon, contain a single Mga binding site centered around position -54 and overlapping the -35 region of the gene promoter, likely interacting with the α-subunit of RNA polymerase [67]. In theory, m6A base modifications at or around this site could potentially influence Mga and RNA polymerase binding to the promoter region, perhaps by stabilizing or localizing Mga to the proper site, promoting activation of gene transcription. Consistent with this hypothesis, examination of the genome sequences upstream of the Mga open reading frame for *S. pyogenes* strains MEW123, MEW427, and SF370, all reveal the existence of the m6A consensus motifs approximately 800 bp upstream of the *mga* start codon [12, 13, 20]. It is unclear if, or how, this m6A motif site located upstream of the predicted Mga promoter region activates Mga expression. The mechanism of m6A-dependent regulation of the *mga* locus is the subject of active investigation by our group.

Regulation of virulence factor expression in response to different environmental cues and stresses is critical to the success of *S. pyogenes* survival and pathogenesis. Over 30 recognized transcriptional regulatory proteins and 13 two-component regulatory systems must function to coordinate virulence factor expression properly [17, 18]. We found that loss of m6A DNA modifications in our ΔRSM mutant correlated with significant changes in virulence properties of *S. pyogenes*. In a murine model of subcutaneous ulcer formation, we noted that mice infected with the ΔRSM mutant displayed enhanced inflammatory responses compared to mice infected with the WT strain, with comparatively larger skin lesions, increased detection of pro-inflammatory cytokine levels, and enhanced neutrophil infiltrates on histologic examination (**Fig 2.4 and 2.5**). Disruption of m6A DNA modifications and an associated dysregulation of gene

transcript expression may result in failed activation of multiple important adhesins and streptococcal proteins involved in evading host immunity (**Fig 2.3** and **Table 2.4**). For example, neutrophilic infiltration in response to bacterial infections is enhanced by activity of host chemotaxins, chiefly complement protein C5a. A major virulence determinant of *S. pyogenes* aiding immune evasion is to degrade complement C5a through activity of ScpA, a surface-expressed, serine-protease specifically degrading host C5a and interfering with neutrophil recruitment [68]. We found that the ΔRSM mutant exhibited significantly decreased transcript expression for ScpA which may partly explain a more exaggerated neutrophil response to infection with the ΔRSM mutant strain, resulting in more inflammation and larger skin lesions (**Fig 2.4C** and **2.5**). Previous investigation into the contribution of ScpA to host immune responses was performed using a murine air sac model of subcutaneous infection performed by Ji *et al*. [69]; air sacs infected with *S. pyogenes* lacking ScpA expression exhibited a significantly enhanced host inflammatory response compared to the WT parent, with a neutrophil predominance analogous to our results. Another report found similar to slightly larger skin lesions in mice infected subcutaneously with *S. pyogenes* lacking ScpA compared to WT [70]. The effect of *S. pyogenes* virulence factors in murine models is not always similar to activity in the human environment; it is known that ScpA does cleave murine C5a, but at slower rates compared to human C5a, and these differences may impact our ability to detect phenotypes in these non-human systems [71]. Similar to our own results with the Ω*scpA* strain infections, the results reported by Li *et al*. were not statistically significant suggesting that the individual contribution of ScpA in this murine model may be modest, but when the expression of multiple virulence factors is disrupted the effects may be more apparent. Indeed, our experiments in the skin lesion model with the ΔRSM and the Δ*mga* strains showed significant differences in lesion size and inflammatory response compared to the WT and complemented mutant strains. Both of these mutant strains would be expected to have similar patterns of differential gene expression and as result they phenocopy each other in this model. Decreased expression of several adhesins and other factors may have contributed to enhanced spread of the infection together with an exaggerated host

inflammatory response resulting in larger areas of inflammation and larger skin lesion formation.

Decreased M protein expression, among other adhesins, also explains the decreased *in vitro* adherence of the ΔRSM mutant to human vaginal epithelial cells. Interestingly, the decreased adherence to human vaginal epithelial cells *in vitro* did not correlate with disrupted carriage in the murine vaginal mucosa colonization model. This suggests that there are additional adhesins not influenced by m6A DNA modifications that are important for promoting and maintaining carriage *in vivo*. One example would be the R28 adhesin of serotype M28 *S. pyogenes* strains, which is a major streptococcal adhesin to human cervical epithelial cells [54]. Our RNA-sequencing experiments did not find significant differences in the transcription of the MEW123 R28 gene (AWM59_02815) between WT and the ΔRSM mutant (full data set available in NCBI repository). With only 20 genes significantly downregulated in the ΔRSM mutant clearly not all major *S. pyogenes* adhesins and virulence factors are impacted by m6A DNA modifications. Our data show that only a few gene operons, or regulons as in the case of Mga, are differentially expressed in the absence of m6A base modifications in *S. pyogenes* and that down regulation of these genes impacts virulence.

In this study, we have demonstrated that the SpyMEW123I RM system and m6A DNA modifications in *S. pyogenes* significantly influence DNA restriction activity, in addition to correlating with differential gene transcription and virulence properties of this important human pathogen. Disruption of the SpyMEW123I Type I RM in *S. pyogenes* altered the transcriptional profile of the mutant strain resulting in attenuated virulence and impaired evasion of the host immune response in both *in vitro* and *in vivo* models. Similar to our results, disruption of Type IIG RM systems in *C. jejuni* and *B. burgdorferi* also interfered with genetic regulation of virulence factors of those pathogens [34, 35]. Together, these findings demonstrate that intact RM systems in these bacterial pathogens, and likely many other prokaryotes, can exert multiple functions, including restriction-mediated protection from foreign DNA in addition to influencing gene expression. Understanding how m6A DNA modifications influence virulence properties in these organisms could potentially identify targets for therapeutic intervention, potentially changing patterns of virulence factor expression resulting in strain

54

attenuation helping to prevent human disease. Further investigation is necessary to fully comprehend the many functions of DNA methylation and the complex nature of bacterial physiology and pathogenesis.

Materials and Methods

Ethics statement. Experimental protocols involving the use of mice were reviewed and approved by the Institutional Animal Care and Use Committee (IACUC) of the University of Michigan Medical School (Ann Arbor, MI, USA). The University of Michigan IACUC complies with the policies and standards as outlined in the Animal Welfare Act and the "Guide for the Care and Use of Laboratory Animals," [72]. The protocol numbers approved by the University of Michigan IACUC are as follows: Skin and Soft Tissue Infection Model of *Streptococcus pyogenes* Virulence (PRO00007495), and Murine Vaginal Colonization Model for *Streptococcus pyogenes* (PRO00007218). For consistency, all experiments utilized female C57BL/6J mice at approximately 6 weeks of age at the time of use. Mice were purchased from The Jackson Laboratories (catalog #000664) (Bar Harbor, ME, USA), and maintained in a University of Michigan animal facility with biohazard containment properties. Following arrival, mice were allowed to acclimatize in the facility for one week prior to beginning experiments. When manipulated, mice were briefly sedated by inhalation of isoflurane via drop jar dosing. Animals were inspected at least once daily for evidence of suffering, manifested by significantly diminished or no activity, decreased appetite, poor grooming, increased respiratory rate, or weight loss greater than 15% of body weight; if evidence of suffering was identified, then the mouse was euthanized. Euthanasia was primarily through carbon dioxide asphyxiation with a subsequent secondary method of euthanasia, including induction of bilateral pneumothorax, decapitation, and/or removal of a vital organ.

Bacterial strains, media, and growth conditions. The principal strain used in this study was *S. pyogenes* MEW123, a streptomycin-resistant ($rpsL_{K56T}$), serotype M28 pharyngeal isolate [55]. Other strains used are listed in Table 2.1. Growth rates and yields of MEW123 and associated mutant strains were measured using a Synergy HTX

plate reader (BioTek, Winooski, VT, USA) in 96 well plates (Greiner Bio-One, Monroe, NC, USA). Briefly, 4µl of overnight culture grown in THY broth was inoculated into 200µl of the described fresh media, with identical strains and conditions measured in at least triplicate. Growth was at 37°C, room air, in static conditions for 12 hours and OD_{620nm} was measured every 3 seconds. Unless otherwise noted, all *S. pyogenes* strains had equivalent growth rates and yields under all *in vitro* conditions tested (S1 Fig). Routine culture of *S. pyogenes* was performed in Todd-Hewitt medium (Becton Dickinson, Franklin Lakes, NJ, USA) supplemented with 0.2% yeast extract (Difco Laboratories, Detroit, MI, USA) (THY media). Where required, Bacto agar (Difco) was added to a final concentration of 1.4% (w/v) to produce solid media. Gene expression experiments used C-Medium, a lower-glucose, higher-protein media that more closely resembles *in vivo* conditions [73]. Incubation was performed at 37°C under anaerobic conditions (GasPack™, Becton Dickinson) for solid media, or in sealed tubes without agitation for broth media. Aerobic culture was conducted as described [74]. For inoculation of mice, *S. pyogenes* was harvested from culture in THY broth at early logarithmic-phase (OD_{600} 0.2), washed once in PBS, briefly sonicated on ice to break up long streptococcal chains, and resuspended in PBS to 10^8 CFU/mL. Molecular cloning used *Escherichia coli* strain DH5a (Invitrogen, Grand Island, NY, USA) cultured in LB broth. When appropriate, antibiotics were added at the following concentrations: erythromycin, 500 µg/mL for *E. coli* and 1 µg/mL for *S. pyogenes*; chloramphenicol, 20 µg/mL for *E. coli* and 3 µg/mL for *S. pyogenes*; spectinomycin,100 µg/mL for both *E. coli* and *S. pyogenes*; and streptomycin, 1000 µg/mL for *S. pyogenes*. In some experiments, growth was monitored in THY broth supplemented with either penicillin, gentamicin, or erythromycin at concentrations ranging from 0.05 µg/mL to 100 µg/mL. All antibiotics were obtained from Sigma Chemical Co., St. Louis, MO, USA.

Gene cloning and mutant construction. *Streptococcus pyogenes* MEW123 was used as a source strain for DNA, Genbank CP014139.1 [12]. Bacterial strains and plasmid vectors are listed in Table 2.1. The primers used for PCR amplification and cloning are listed in Table 2.5. For cloning and routine DNA Sanger sequencing, the Phusion High-Fidelity DNA Polymerase (New England Biolabs, Inc., Ipswich, MA, USA) was used. For

routine endpoint PCR amplification standard Taq DNA Polymerase was used (New England Biolabs, Inc.). Polymerase chain reaction products were digested with indicated restriction enzymes and ligated to pJRS233 or pGCP213 for in-frame deletions, pSPC18 for insertional mutations, or pJoy3 as a plasmid vector for transformation efficiency assays. In-frame deletion mutants and insertional mutants were constructed essentially as described [25, 26], [28], and [27], respectively.

a) **Construction of an in-frame deletion of *SpyMEW123I* gene cluster.** The ΔRSM in-frame deletion allele was cloned by splice overlap extension (SOE) PCR [75]. Corresponding GenBank accession numbers for the MEW123 restriction endonuclease gene *hsdR*, specificity subunit *hsdS*, and the methyltransferase subunit *hsdM*, are AWM59_07895, AWM59_07900, and AWM59_07905, respectively. The upstream region of the gene cluster was PCR amplified using primers MEW123 Del-RSM F1 and MEW123 Del-RSM R2, producing a 1.02 kb amplicon. The downstream region of the gene cluster was PCR amplified using primers MEW123 Del-RSM F3 and MEW123 Del-RSM R4, producing a 1.02 kb amplicon. These two amplicons contain complementary ends that anneal together and essentially will produce an in-frame deletion of the three-gene restriction endonuclease, specificity subunit, and DNA methyltransferase open reading frames. The two amplicons were mixed together as template and further amplified using primers MEW123 Del-RSM F1 and MEW123 Del-RSM R4, the resulting amplicon was approximately 2.04 kb and contained a unique EcoRI site at the 5' end and a unique HindIII site at the 3' end. The resulting amplicon was digested with EcoRI and HindIII, and inserted within same restriction sites of the *E. coli* to *S. pyogenes* temperature-sensitive vector for allelic replacements, pGCP213 [26], producing plasmid pKJ24. The pKJ24 plasmid was confirmed by Sanger DNA sequencing using primers MEW M13 F and MEW M13 R, which bind just outside and flank the multiple cloning site region within pGCP213. Electrocompetent cells of MEW123 were prepared and transformation was performed essentially as previously described [76]. The pKJ24 plasmid carrying the RSM in-frame deletion was transformed into electrocompetent *S. pyogenes* MEW123 through electroporation with conditions as described above. Erythromycin-resistant transformants were handled according to the temperature-sensitive selection protocol as previously described [26]. Final clones of *S.*

pyogenes that had successfully replaced the full-length genomic RSM gene cluster with the in-frame deletion allele were screened by endpoint PCR and confirmed by Sanger DNA sequencing. The resulting strain containing the in-frame deletion allele (∆RSM) was identified as MEW513.

b) Construction of ∆RSM strain complemented *in trans* with plasmid-encoded RSM operon. GenBank accession numbers for the MEW123 restriction endonuclease gene *hsdR*, specificity subunit gene *hsdS*, and the methyltransferase subunit gene *hsdM*, are AWM59_07895, AWM59_07900, and AWM59_07905, respectively. The operon was cloned by PCR using primers pJoy3_123_RSM_F and pJoy3 123 RSM R, producing an amplicon of approximately 6 kb. This fragment was inserted into plasmid pJoy3 linearized by digestion with EcoRI and SphI using the NEBuilder® HiFi DNA Assembly kit (New England Biolabs, Inc.), producing plasmid pEH01. This plasmid was transformed into electrocompetent *S. pyogenes* MEW513 through electroporation with conditions as described above. Chloramphenicol-resistant clones were selected and screened by endpoint PCR, with restoration of m6A methylation activity confirmed by dot blot. The resulting strain containing the plasmid encoded RSM operon for complementation (∆RSM/pRSM) was identified as strain MEW552.

c) Construction of spectinomycin-cassette disruption mutant of restriction endonuclease gene, *hsdR*. The GenBank accession number for the restriction-endonuclease subunit gene, *hsdR*, is AWM59_07895. A fragment of the endonuclease open reading frame was cloned by PCR using primers 123_7895 F and 123_7895 R, producing an amplicon of approximately 950 bp. This fragment was inserted into plasmid pSpc18 linearized by digestion with HindIII and BamHI using the NEBuilder® HiFi DNA Assembly kit (New England Biolabs, Inc.), producing plasmid pKJ19. This plasmid was transformed into electrocompetent *S. pyogenes* MEW123 through electroporation with conditions as described above. Spectinomycin-resistant clones were selected and screened by endpoint PCR, with final confirmation by Sanger DNA sequencing. The resulting strain containing the spectinomycin-resistance cassette insertion disrupting the restriction endonuclease gene *hsdR* (ΩRE) was identified as strain MEW489.

d) Construction of in-frame deletion of *mga*. The GenBank accession number for the MEW123 Mga protein, gene *mga*, is AWM59_08335. An in-frame deletion allele of *mga* was cloned by splice-overlap extension (SOE) PCR [75]. The upstream region of the *mga* gene was cloned using primers <u>M28 Mga 5' SalI</u> and <u>M28 Mga 5' SOE R</u>, producing an amplicon of approximately 420 bp. The downstream region of the mga gene was cloned using primers <u>M28 Mga 3' BamHI</u> and <u>M28 Mga 3' SOE F</u>, producing an amplicon of approximately 410 bp. The two amplicons are mixed together as template and amplified using the outside primers <u>M28 MGA 5' SalI</u> and <u>M28 Mga 3' BamHI</u>, producing an amplicon of approximately 830 bp. This amplicon was subsequently digested with BamHI and SalI and ligated into the *E. coli* to *S. pyogenes* temperature-sensitive vector for allelic replacements, plasmid pJRS233 [25], cut similarly with BamHI and SalI. The resulting plasmid of was named pIL01, with confirmation by Sanger DNA sequencing and PCR verification. Electrocompetent cells of MEW123 were prepared and transformation with plasmid pIL01 was performed essentially as previously described [76]. Erythromycin-resistant transformants were handled according to the temperature-sensitive selection protocol as previously described [26]. Final clones of *S. pyogenes* that had successfully replaced the full-length genomic *mga* allele with the in-frame deletion allele were screened by endpoint PCR and confirmed by Sanger DNA sequencing. The resulting strain containing the in-frame deletion allele (Δ*mga*) was identified as MEW480.

e) Construction of spectinomycin-cassette disruption mutant of strain MEW123 *scpA* gene (ScpA protein). The GenBank accession number for the MEW123 *scpA* gene is AWM59_08315. A fragment of the *scpA* open reading frame was cloned by PCR using primers <u>M28 ScpA SalI F</u> and <u>M28 ScpA SacI R</u>, producing an amplicon of approximately 1.1 kb. The amplicon was digested with SalI and SacI and ligated into plasmid pSpc18 linearized with SalI and SacI, producing plasmid pIL09. This plasmid was transformed into electrocompetent *S. pyogenes* MEW123 through electroporation with conditions as described above. Spectinomycin-resistant clones were selected and screened by endpoint PCR, with final confirmation by Sanger DNA sequencing. The resulting strain containing the spectinomycin-resistance cassette insertion disrupting the *scpA* gene (Ω*scpA*) was identified as strain MEW380.

f) Construction of spectinomycin-cassette disruption mutant of strain MEW123 *emm28* gene (M28 protein). The GenBank accession number for the M28 protein, gene *emm28*, is AWM59_08325. A fragment of the *emm28* open reading frame was cloned by PCR using primers M28 Emm HindIII F and M28 Emm BamHI R, producing an amplicon of approximately 1.1 kb. This amplicon incorporated unique sites for HindIII and BamHI, and the amplicon was accordingly restriction digested and ligated into plasmid pSpc18 opened with HindIII and BamHI, producing plasmid pIL03. This plasmid was transformed into electrocompetent *S. pyogenes* MEW123 through electroporation with conditions as described above. Spectinomycin-resistant clones were selected and screened by endpoint PCR, with final confirmation by Sanger DNA sequencing. The resulting strain containing the spectinomycin-resistance cassette insertion disrupting the *emm28* gene (Ω*emm28*) was identified as strain MEW409.

Transformation efficiency assay. Transformation efficiency was assessed by electroporation of electrocompetent *S. pyogenes* strains with 0.5 µg plasmid pJoy3 conferring chloramphenicol resistance isolated from *E. coli* DH5α. Electroporation was performed using a Gene Pulser II system (BioRad, Hercules, CA, USA) under the following settings; Volts at 1.75 kV, capacitance at 400Ω, and resistance at 25 µf. Transformants were plated onto THY agar supplemented with chloramphenicol. In addition, a separate aliquot of the sample was plated onto THY agar with no antibiotics to determine the total viable cell count. Transformation efficiency was determined as the number of chloramphenicol resistant cells per total viable cell count.

SMRT sequencing. Genomic DNA was purified from *S. pyogenes* strains MEW123 (WT) and MEW513 (ΔRSM) using the Wizard Genomic DNA Purification Kit (Promega, Madison, WI). Genomic DNA preparation, library preparation, and sequencing of MEW123 was performed as previously described [12]. For MEW513, one Single Molecule Real-Time (SMRT) cell was used to sequence the library prepared with 5 kb mean insert size on the Pacific Biosciences RSII sequencer by the University of Michigan Sequencing Core (https://brcf.medicine.umich.edu/cores/dna-sequencing).

Modification and motif analysis was performed using RS_Modification_and_Motif_Analysis.1 version 2.3.0 using the published MEW123 reference genome with an average reference coverage of 501 and 539 for MEW123 and MEW513, respectively. Data generated in this analysis have been deposited in NCBI's Gene Expression Omnibus [77] and are accessible through GEO Series accession number GSE130428 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE130428).

Dot blot assay for m6A modification. Genomic DNA was isolated from *S. pyogenes* strains MEW123 (WT) and MEW513 (∆RSM), as described above. DNA was treated with RNAse during purification to remove any contaminating mRNA or rRNA potentially containing m6A base modifications. DNA was denatured by heating at 98°C for 10 min and then placed immediately on ice for 5 minutes. Denatured DNA or unmodified oligonucleotides as a negative control was then spotted at 500 ng per spot onto nitrocellulose membranes and allowed to air dry. Membranes were then placed onto Whatman paper soaked with PBS containing 0.5% Tween 20 (PBST), and DNA was cross-linked to the membranes using a Bio-Rad GS Genelinker using two 125 mJ delivery cycles. Membranes were blocked in 5% milk protein in PBS for 1 h at room temperature and then incubated with a dilution of anti-m6A primary rabbit antibody (2 μg/mL) (EMD Millipore ABE572 Anti-N6-methyladenosine (m6A) Antibody) in 5% milk PBS overnight at 4°C. Primary antibody was removed and the membrane was washed three times with PBST for 5 minutes each wash. The membrane was then incubated with a 1:5,000 dilution of horseradish peroxidase-conjugated anti-rabbit secondary antibody in 5% milk PBS at room temperature for 1 hour. The secondary antibody was removed, and the membrane washed with PBST three times for 5 minutes each wash. Chemiluminescent substrate (Pierce SuperSignal West Femto HRP Substrate, ThermoFisher Scientific, Waltham, MA) was applied and the membrane was visualized.

RNA-sequencing. Streptococcus from fresh overnight growth on THY agar plates was inoculated into 40 mL of C-media broth and grown statically to mid-log phase OD_{600nm} of

0.6. RNA was then purified using the RiboPure RNA Purification Kit (Life Technologies), for bacteria according to the manufacturer's recommendations. The University of Michigan Sequencing Core performed ribosomal rRNA depletion using the Ribo-Zero Magnetic Kit, bacteria and subsequent library preparation. Fifty-base single end reads were sequenced on the Illumina HiSeq 4000. Sequence alignment was performed using the Burrows-Wheeler Aligner (BWA) version 0.7.8-r455 to the MEW123 reference genome [12]. Subsequent differential expression analysis was performed using the limma package in R [78]. Differentially expressed genes were called as those that had a Benjamini-Hochberg adjusted p-value less than 0.05 and a $\log_2$ fold change greater than 1. $\log_2$ CPM values were computed using edgeR and were subsequently used to construct the heatmap using the *aheatmap* function as part of the NMF package in R [79, 80]. Data generated in this analysis have been deposited in NCBI's Gene Expression Omnibus [77] and are accessible through GEO Series accession number GSE130427 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE130427).

Real-time PCR for comparison of transcript expression. Based on results of the most significantly differentially expressed genes between WT (MEW123) and ΔRSM (MEW513), three genes were selected for independent reverse-transcription cDNA preparation and real-time PCR amplification for relative comparison of transcript expression; *mga*, *emm*28, and *scpA. RNA* was isolated as described above from strains grown in C-media broth to mid-log phase OD_{600nm} of 0.6. Synthesis of cDNA was performed using the iScript™ cDNA Synthesis Kit (BioRad). Real time amplification of select genes was performed using an iCycler Thermocycler (BioRad) and iQ SYBR Green Supermix (BioRad). Sequences for RT-PCR primers are as shown in Table 2.5. Relative transcript levels were determined using the *recA* transcript as reference by the $2^{(-\Delta\Delta Ct)}$ method [81]. All RNA was stored at -80ºC. All cDNA was stored at -20ºC or utilized directly for comparative RT-PCR analysis. For each experiment, three biological replicates were analyzed in duplicate. Statistical significance was examined using the paired t-test in Prism 6 (GraphPad).

Murine subcutaneous infection model. Inflammatory infection of murine subcutaneous tissue was conducted as described in detail [41]. On the day of infection, mice sedated by inhalation of isoflurane received a subcutaneous injection of 100 µl PBS containing 1 x 10^7 *S. pyogenes* into the shaved flank. Following infection, the resulting ulcers were photographed over several days and the areas of the irregular lesions were calculated using ImageJ software as described in detail elsewhere [82, 83]. Skin biopsies were obtained from euthanized mice and homogenized in 1 mL ice cold PBS using a FastPrep-24 homogenizer (MP Biomedicals, LLC., Santa Ana, CA); tissue was homogenized in 2 mL conical screw top vials with 3.2 mm stainless steel beads (Fisher Scientific, Pittsburg, PA) with two FastPrep cycles of speed 6.0 for 45 sec, with a 5 min ice incubation between pulses to prevent overheating.

Murine tissue cytokine analysis. Samples of mouse skin and subcutaneous tissue homogenates were harvested at six-days post-infection. Cytokine protein concentrations were determined by a multiplex murine ELISA assay (EMD Millipore, Billerica, MA, USA) according to the manufacturer's protocol.

Murine tissue histologic analysis. Murine skin biopsies were obtained at six-days post-infection and were fixed in 4% formalin and dehydrated up to 70% ethanol prior to paraffin embedding through the University of Michigan Pathology Core for Animal Research (PCAR). H&E staining and immunohistochemistry services were performed by the PCAR using commercially available anti-CD3 (T lymphocytes), and anti-F4/80 (macrophages) antibodies (Abcam, Cambridge, MA, USA). Digital images were obtained with an EC3 digital imaging system (Leica Microsystems, Buffalo Grove, IL, USA) using Leica Acquisition Software (Leica Microsystems). Adjustments to contrast in digital images were applied equally to all experimental and control images.

Human vaginal epithelial cell *in vitro* assays. Adherence of *S. pyogenes* strains was assessed to an established human vaginal epithelial cell line, VK2/E6E7, using methods

63

similar to those previously described [84-86]. The human vaginal epithelial cell line VK2/E6E7 was purchased from the American Type Culture Collection (ATCC, Manassas, Virginia), and cells were grown and maintained in media and conditions as recommended by ATCC. Human cells were grown to confluence in 12-well tissue culture-treated plates and washed with sterile PBS prior to inoculation with bacteria. *S. pyogenes* strains were grown in THY broth to early stationary phase (OD_{600nm} 0.6), washed twice in sterile PBS, and adjusted to give an inoculum of ~5 x 10^6 CFU in 1 mL per well, for a multiplicity of infection (MOI) of ~5. Bacteria and human cells were incubated at 37°C in 5% carbon dioxide for 60 min, after which time the supernatants were removed and cells were washed four times with 2 mL sterile PBS to remove non-adherent organisms. To recover *S. pyogenes* from the epithelial cells, each well was treated with 0.2 mL 0.25% Trypsin-EDTA (Invitrogen) and incubated at 37°C for 5 min, and then lysed by addition of 0.8 mL sterile water at pH 11. Lysis in water at pH 11 was shown to result in a more complete eukaryotic cellular breakdown with maximal recovery of bacteria from the surface in addition to intracellular reservoirs [87]. This method recovers all cell-associated streptococci, predominantly extracellular adherent cells with a relatively smaller amount of intracellular cells. The cell suspension was serially diluted in PBS and plated onto THY agar for determination of viable CFU count. The total cell-associated CFU percentage was calculated as (total CFU recovered from the well/CFU of the original input inoculum) x 100%.

Murine vaginal colonization model. Experiments were performed as previously described [55]. To synchronize estral cycles, sedated mice were estrogen supplemented by intra-peritoneal injection with 0.5 mg β-estradiol 17-valerate (Sigma) dissolved in 0.1 mL sterile sesame oil (Sigma) 2 days prior to streptococcal inoculation and again on the day of inoculation (considered day #0). On day #0, sedated mice were inoculated with ~1 x 10^6 colony forming units (CFUs) instilled into the vaginal vault using a P20 micropipetter (Gilson, Inc., Middleton, WI) in a total volume of 20 μL PBS. At successive intervals over a 1-month period post-inoculation, the vaginal vaults of sedated mice were gently washed with 50 μL sterile PBS and serial dilutions in sterile

64

PBS were plated onto THY agar plates supplemented with 1000 μg/mL streptomycin to determine viable CFUs. This concentration of streptomycin suppressed growth of normal mouse vaginal flora but had no effect on the plating efficiency of the streptomycin-resistant *S. pyogenes* strains. For colonization experiments, between 5 to 20 mice were tested per *S. pyogenes* strain, as indicated in the relevant figure legends.

Human neutrophil bactericidal activity assay. Human neutrophils were purchased from a commercial supplier (Astarte Biologics, Bothell, WA, USA) and prepared according to supplier recommendations to a concentration of 5 x 10^6 cells/mL in room temperature Hank's Balanced Salt Solution (HBSS). Neutrophil bactericidal assay was performed similar to that reported by Staali *et al*. [45]. Briefly, streptococcal strains were grown in fresh C-media to mid-log phase (OD$_{600nm}$ of 0.6) and were washed twice in HBSS with calcium and magnesium, but without Phenol Red (Sigma, St. Louis, MO, USA). Streptococci were counted using a hemocytometer and adjusted to a concentration of 5 x 10^7 CFU/mL in room temperature HBSS. Neutrophils and streptococci were mixed in a 1:10 ratio of neutrophils to bacteria, and were incubated together for 10 minutes at 37°C. Next, extracellular streptococci were eliminated by addition of gentamicin (100 μg/mL) and penicillin (5 μg/mL) in HBSS for 20 minutes at 37°C. Next, cells were diluted in 1 mL of HBSS, centrifuged at 400g x 5 min, and washed with 1 mL fresh HBSS. The wash was repeated a second time in HBSS and the final cell pellet was resuspended in 50 μL of 2% saponin in distilled water at pH 11 and allowed to remain at room temperature for 20 minutes to lyse neutrophils and release viable intracellular streptococci. The cells were diluted in distilled pH 11 water and aliquots plated onto fresh THY agar media for CFU counts. Three biological replicate experiments for each strain were performed.

Statistical analyses. Comparison of nonparametric data sets was performed using the Mann-Whitney U-test to determine significant differences. Differences between groups for recovery of CFU in vaginal washes were tested using a repeated measures analysis of variance. Differences in relative transcript levels were tested for significance with a

two-tailed paired t-test. Differences in VK cell adherence and in neutrophil bactericidal survival assays were compared using a non-paired t-test. For all tests, the null hypothesis was rejected for $P < 0.05$. Computation utilized the resources available in GraphPad Prism™ (GraphPad Software, Inc., San Diego, CA).
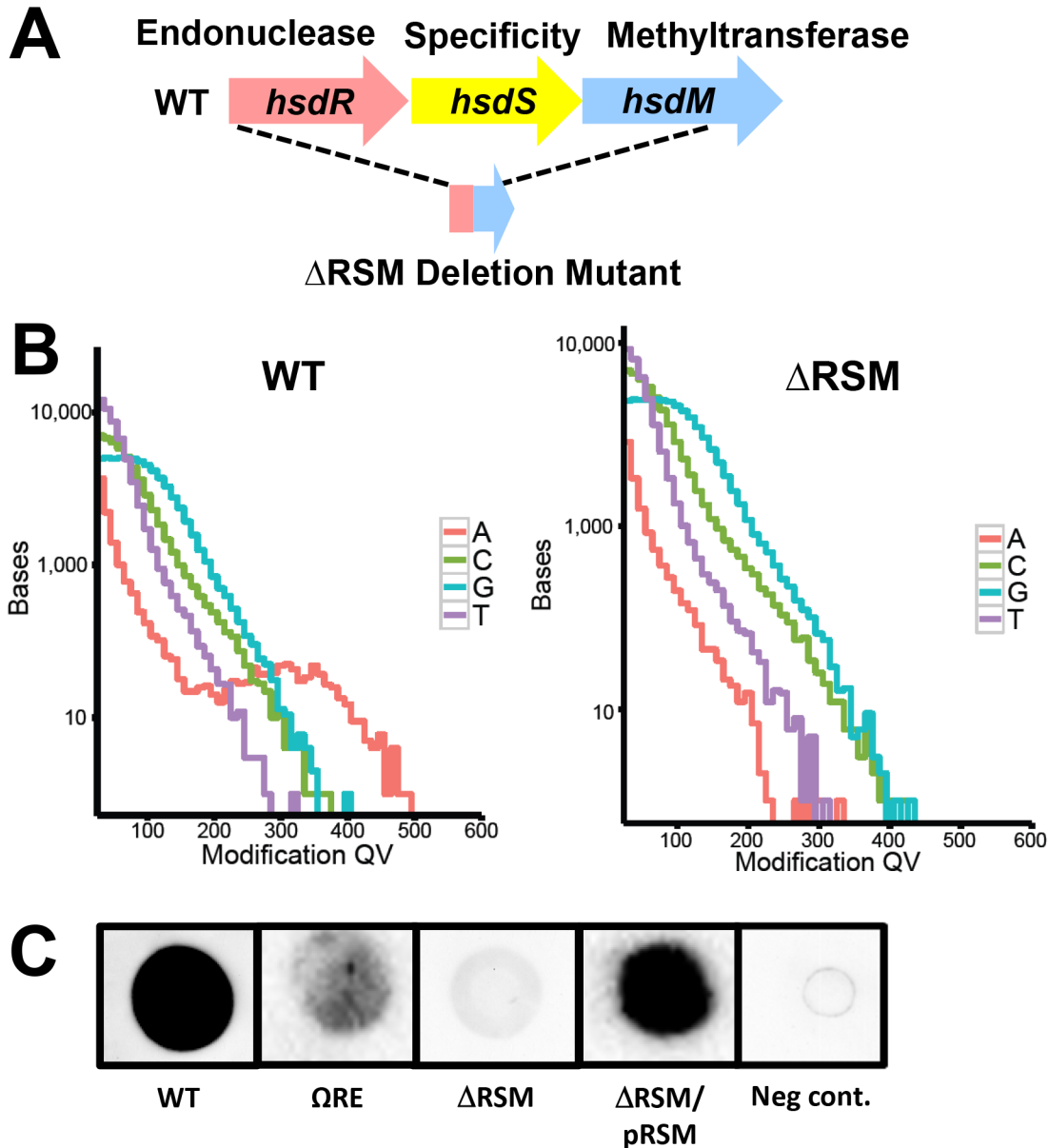
Fig. 2.1. **M.SpyMEW123I dependent m6A modifications in the *S. pyogenes* genome.** A) Genomic organization of the SpyMEW123I Type I RM system gene cluster, *hsdRSM*, in strain MEW123 (WT) and the MEW513 in-frame deletion mutant (ΔRSM). B) Detection of genomic m6A base modifications (red line) in the MEW123 genome or the ΔRSM genome via PacBio SMRT sequencing. Modification quality values (modQVs) are indicated on the x-axis and the number of bases is indicated on the y-axis. ModQVs indicate if the polymerase kinetics at a position differs from the expected background, where a modQV of 30 corresponds to a p-value of 0.001. C) Dot blot with α-m6A antibody on genomic DNA isolated from the following strains: MEW123 (WT), restriction endonuclease *hsdR* antibiotic cassette-disruption mutant MEW489 (ΩRE), in-frame deletion of the *hsdRSM* gene cluster (ΔRSM), the ΔRSM strain complemented with plasmid-encoded *hsdRSM* (ΔRSM/pRSM), and unmodified DNA oligonucleotides serving as a negative control (Neg cont.) (500 ng DNA per spot).
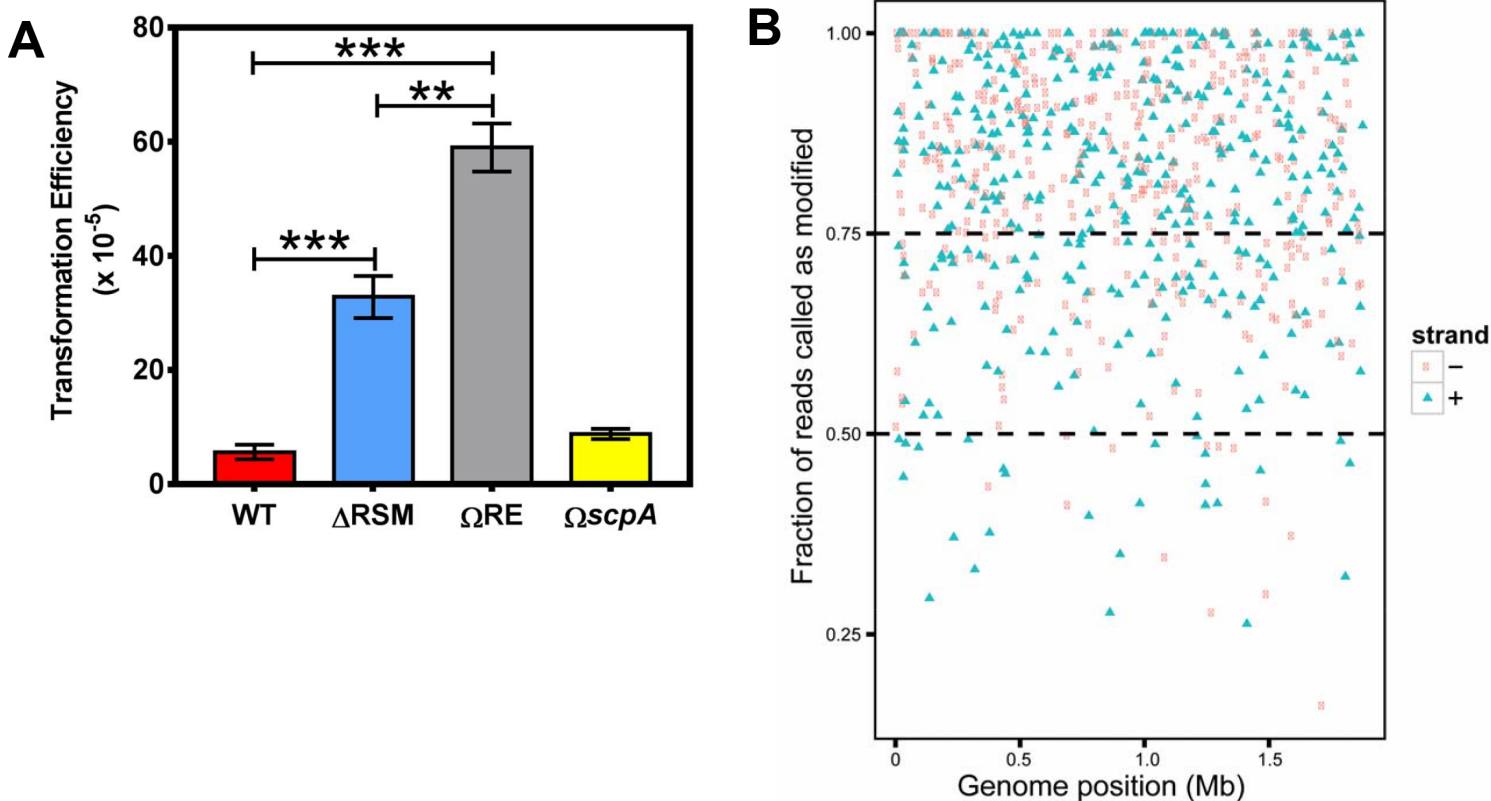
Fig. 2.2. The Type I RM system, SpyMEW123I, is functional for endonuclease activity and influences transformation efficiency. A) Transformation efficiencies of WT (MEW123), the *hsdRSM* in-frame deletion (ΔRSM, MEW513), the *hsdR* single-gene disrupted mutant (ΩRE, MEW489), and the *scpA* insertional inactivation mutant (Ω*scpA*, MEW380). A Mann-Whitney U test was used to identify significant differences between groups for transformation efficiency; ∗∗ $P < 0.01$, ∗∗∗ $P < 0.001$, n = 3-7 replicates per point. B) Each SpyMEW123I RM site is represented on the plot where the position that the site occurs in the genome (from 0 Mb -1.8 Mb) is represented on the x-axis. The fraction of reads called as methylated from PacBio SMRT sequencing for each RM site is represented on the y-axis. RM sites that occur on the + and – strand of the DNA duplex are represented by green triangles and red dots, respectively.
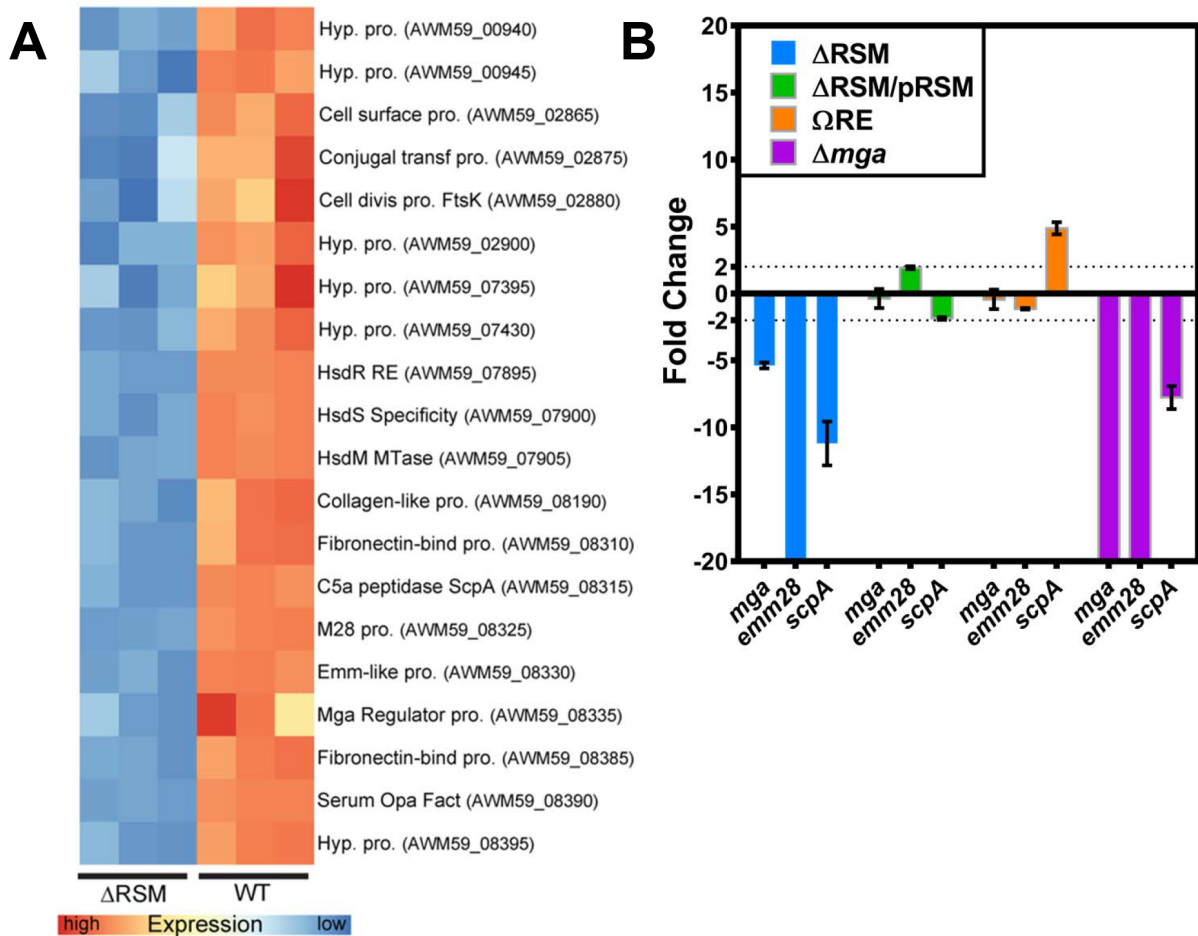
Fig. 2.3. Gene transcripts involved in immune evasion and adherence are down regulated in the ΔRSM strain compared to WT. A) Heatmap of differentially expressed genes between WT and ΔRSM strains. Replicates in triplicate are represented on the x-axis and published/putative gene functions are on the y-axis with gene references to the MEW123 genome. Relative expression is compared row-wise with more highly expressed replicates in red. B) Verification of RNA-Seq data using qRT-PCR with individual primer sets shown on the x-axis for the ΔRSM mutant (blue), complemented mutant strain ΔRSM/pRSM (green), *hsdR* single gene mutant ΩRE (orange), and the Δ*mga* mutant (purple). The y-axis indicates Relative Transcript Levels for individual transcripts compared to *recA* reference transcript. Each gene transcript was analyzed in triplicate. Shown is fold change compared to WT expression, with genes showing greater than two-fold change as significant. Genes extending lower than the x-axis were down-regulated several hundred-fold.
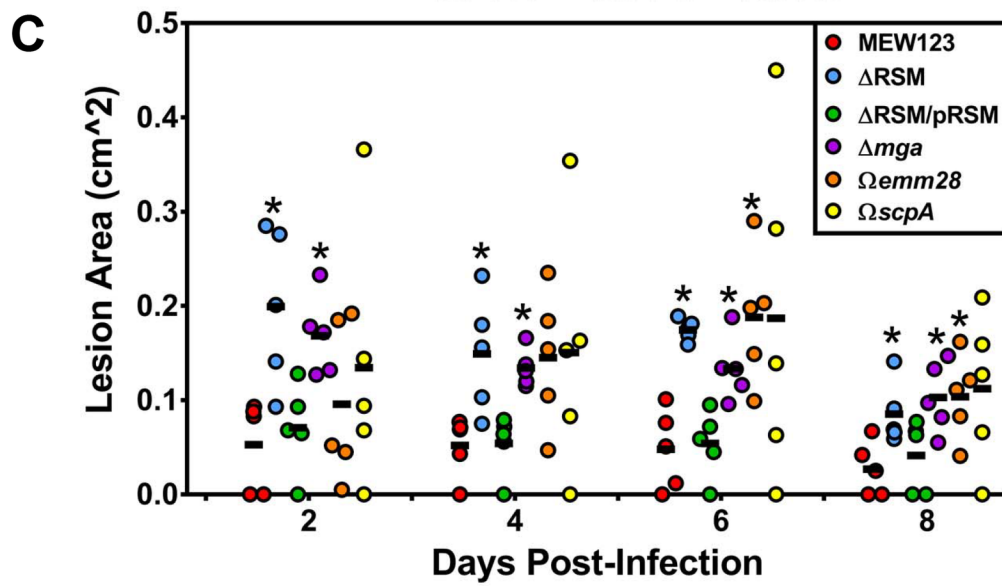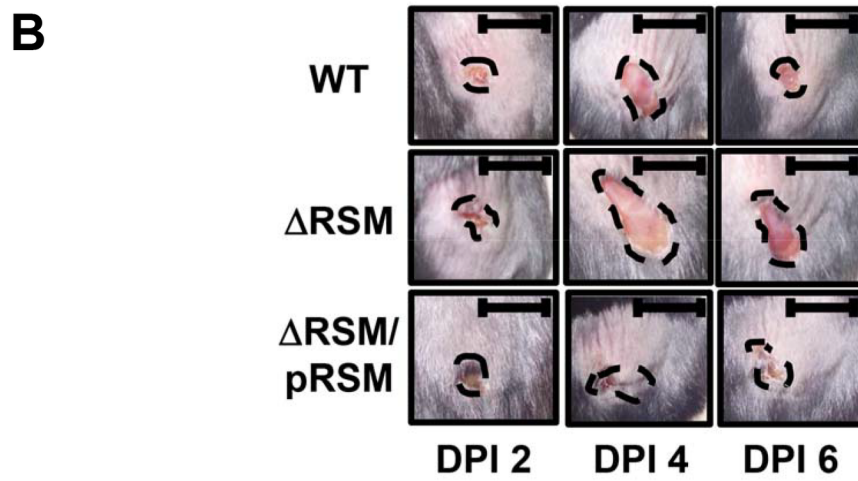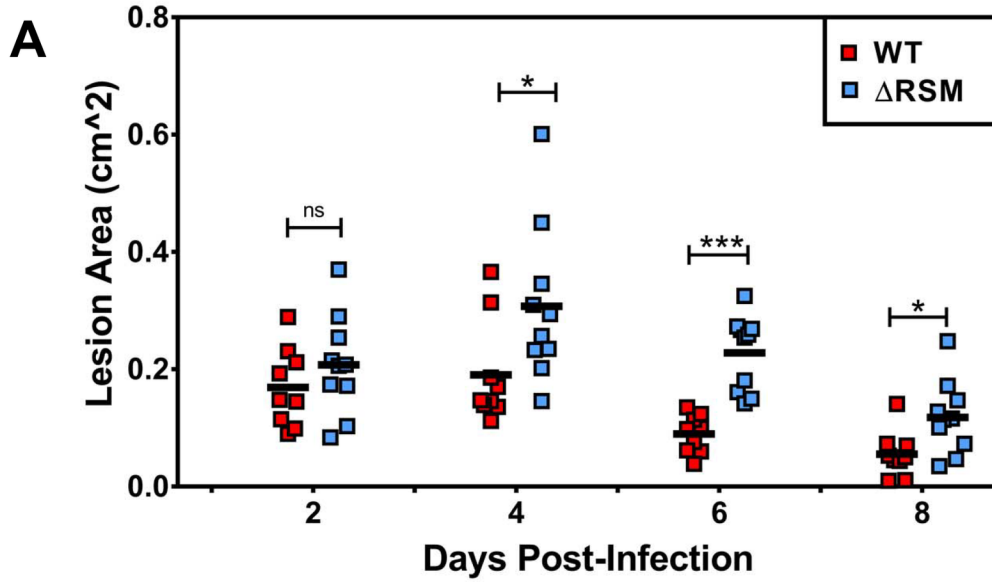
Fig. 2.4. Deletion of the MEW123 RSM gene cluster is associated with larger skin lesion formation in a murine subcutaneous infection model. C57BL/6J mice were inoculated subcutaneously at the shaved flank with 1×10^7 CFUs MEW123 (WT, red squares) or the ΔRSM mutant (blue squares) on Day 0. Lesions were photographed daily and lesion area was calculated using ImageJ software. A) Shown are individual mouse lesion area sizes with mean values (black bars) over Days Post-Infection (DPI). A Mann-Whitney U test was used to identify significant differences in lesion size between strains at each time point; ns, not significant, $*$ $P < 0.05$, $***$ $P < 0.001$. B) Representative images of mice skin lesions over time at days 2-, 4-, and 6-days post-infection. Shown are skin lesions following infection with MEW123 (WT), MEW513 (ΔRSM), and the complemented strain MEW552 (ΔRSM/pRSM). Black bars are 1 cm for reference. Black dashed lines highlight the area of tissue injury. C) Shown are individual mouse lesion area sizes from a representative experiment with mean values (black bars) over Days Post-Infection. Shown are lesions from mice infected with MEW123 (WT), MEW513 (ΔRSM), the complemented strain MEW552 (ΔRSM/pRSM), the in-frame deletion of Mga, MEW480 (Δ*mga*), the cassette-insertion disruption mutant of *emm28*, MEW409 (Ω*emm28*), and the cassette-insertion disruption mutant of *scpA*, MEW380 (Ω*scpA*). A Mann-Whitney U test was used to identify significant differences in lesion size between strains at each time point; $*$ $P < 0.05$.
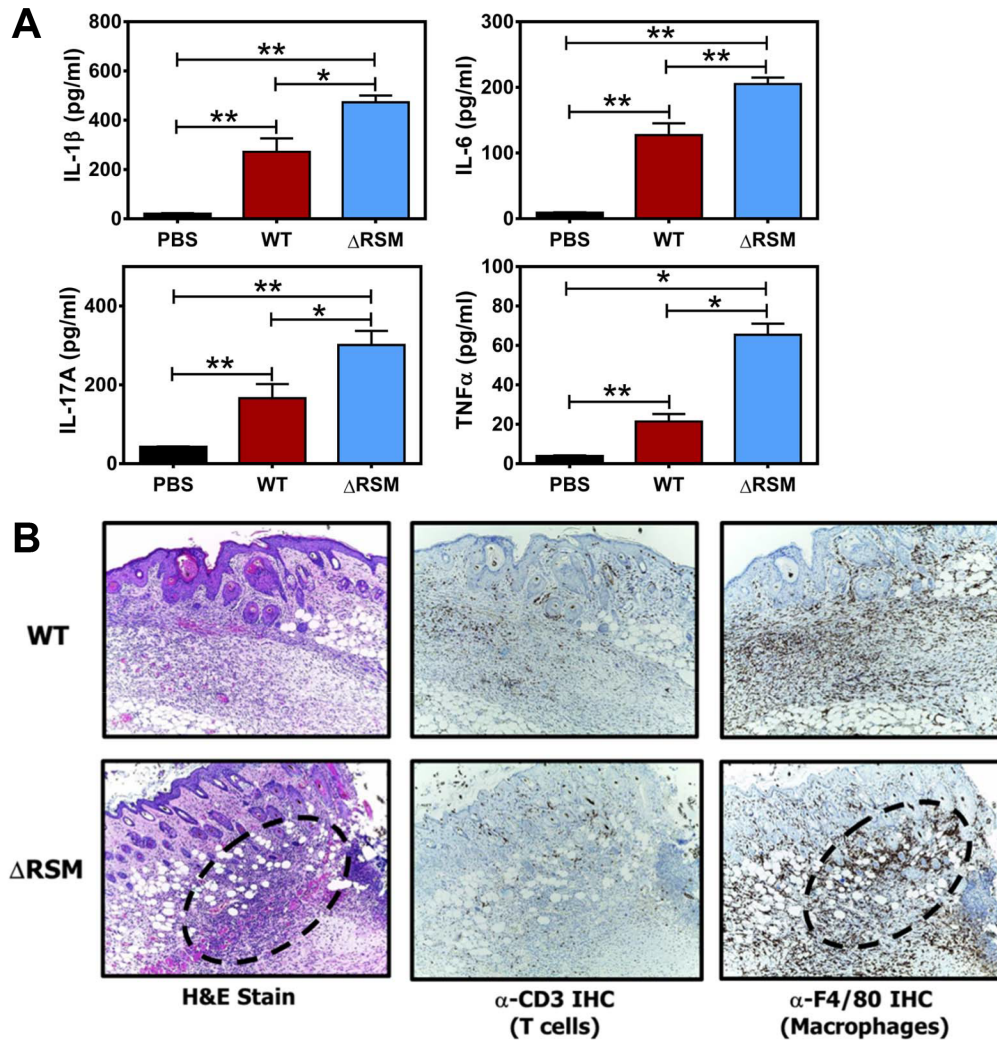
Fig. 2.5. **Deletion of the MEW123 RSM gene cluster increases the host inflammatory cytokine response in the murine subcutaneous infection model.** C57BL/6J mice were inoculated subcutaneously at the shaved flank and resulting lesions were dissected on day 6 post-infection for cytokine analysis (panel A) and histology (panel B). A) Shown are ELISA results of homogenized murine skin biopsy specimens from mice previously inoculated with sterile phosphate buffered saline (PBS, black bars), 1×10^7 CFUs MEW123 (WT, red bars), or 1×10^7 CFUs of the ΔRSM mutant (ΔRSM, blue bars). Results are pooled from biopsies of 3 mice per group with mean and SEM cytokine concentrations. A Mann-Whitney U test was used to identify significant differences in cytokine concentration; ns, not significant, $\ast$ $P < 0.05$, $\ast\ast$ $P < 0.01$. B) Representative images at 10X magnification of skin biopsies from mice inoculated with MEW123 (WT, upper row), or the ΔRSM mutant (ΔRSM, bottom row). Slides were stained with hematoxylin and eosin (H&E) for general neutrophil and overall inflammatory response, or specifically by immunohistochemistry for T cells (α-CD3) or macrophages (α-F4/80). Focal areas of intense inflammation were outlined with dashed lines for comparison.
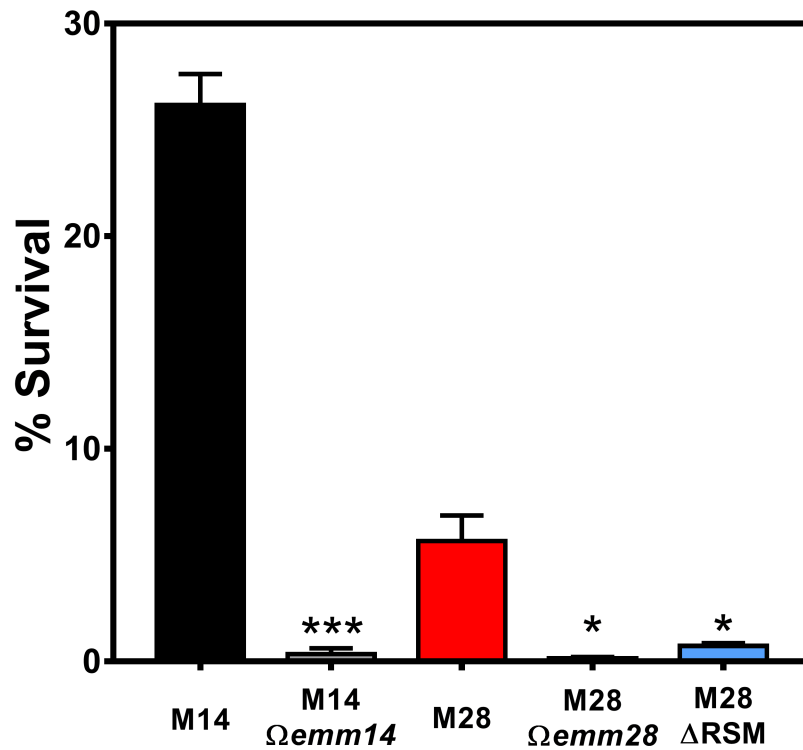
Fig. 2.6. Deletion of MEW123 RSM gene cluster impairs resistance to human neutrophil bactericidal activity. Streptococcal resistance to human neutrophil bactericidal activity was examined using an *in vitro* assay. Human neutrophils were incubated with *S. pyogenes* strains in a 1:10 ratio for 10 min. Extracellular streptococci were eliminated with gentamicin and penicillin for 20 min. Neutrophil lysis and release of viable intracellular streptococci was performed with 2% saponin in pH 11 water for 20 min. Surviving CFUs were plated onto THY agar. Shown are streptococcal strains HSC5 (M14) as a positive control, the HSC5 Ω*emm* (M14 Ω*emm*) as a negative control, MEW123 (M28), MEW123 Ω*emm* (M28 Ω*emm*), and the MEW123 RSM deletion mutant (M28 ΔRSM). Shown are mean ± SEM CFU counts performed in triplicate from a representative experiment. An non-paired t-test was used for statistical significance, comparing mutant to parent strain; ∗ *P* < 0.05, ∗∗∗ *P* < 0.001.
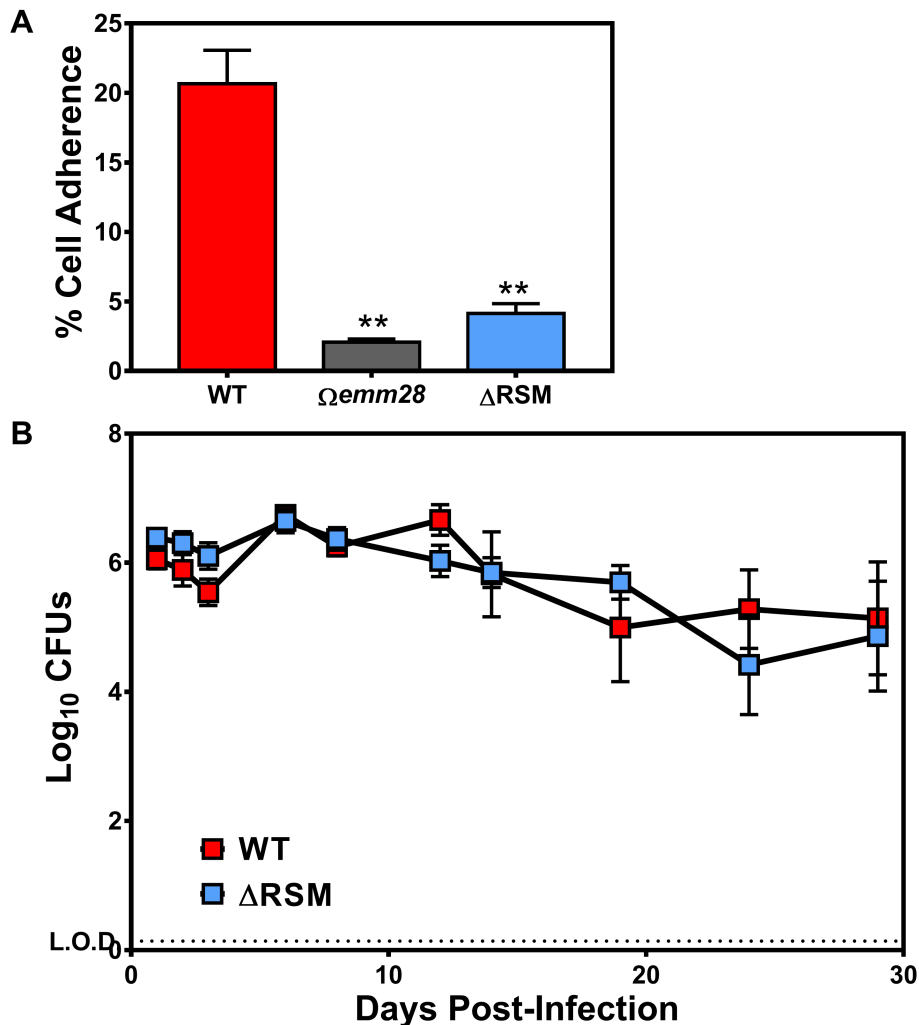
Fig 2.7. Deletion of MEW123 RSM gene cluster impairs streptococcal adherence to human vaginal epithelial cells, but does not impact carriage duration in a murine vaginal colonization model. A) Confluent wells of VK E6E7 human vaginal epithelial cells were inoculated with $5x10^6$ CFUs of *S. pyogenes* MEW123 (WT, red), the *emm28* gene-disrupted mutant (Ω*emm28*, gray), or the RSM deletion mutant (ΔRSM, blue) (multiplicity of infection 5-10:1). Following 1h incubation, non-adherent streptococci were washed away with sterile PBS followed by lysis of the epithelial cells with pH 11 water, serial dilution, and determination of viable streptococci remaining. The % Cell Associated (mostly adherent plus few intracellular) is the percentage of the inoculum CFUs remaining detectable at 1h. Significant differences between groups were calculated by non-paired t-test (** $P < 0.01$, n = 6 replicates per point). B) Estrogenized female C57BL/6J mice were intravaginally inoculated with $1x10^6$ CFU of *S. pyogenes* MEW123 (WT, red), or the RSM deletion mutant (ΔRSM, blue). Mice were

cultured by intravaginal washes with sterile PBS and plated onto selective media (THY supplemented with streptomycin 1000 µg/mL) for quantification. Shown are mean ± SEM Log_{10} CFU counts over time post-infection of 10 mice per group pooled from two separate experiments. A repeated-measures ANOVA was used for statistical significance; at no time point tested over the 28-day experiment were the two groups statistically significantly different.
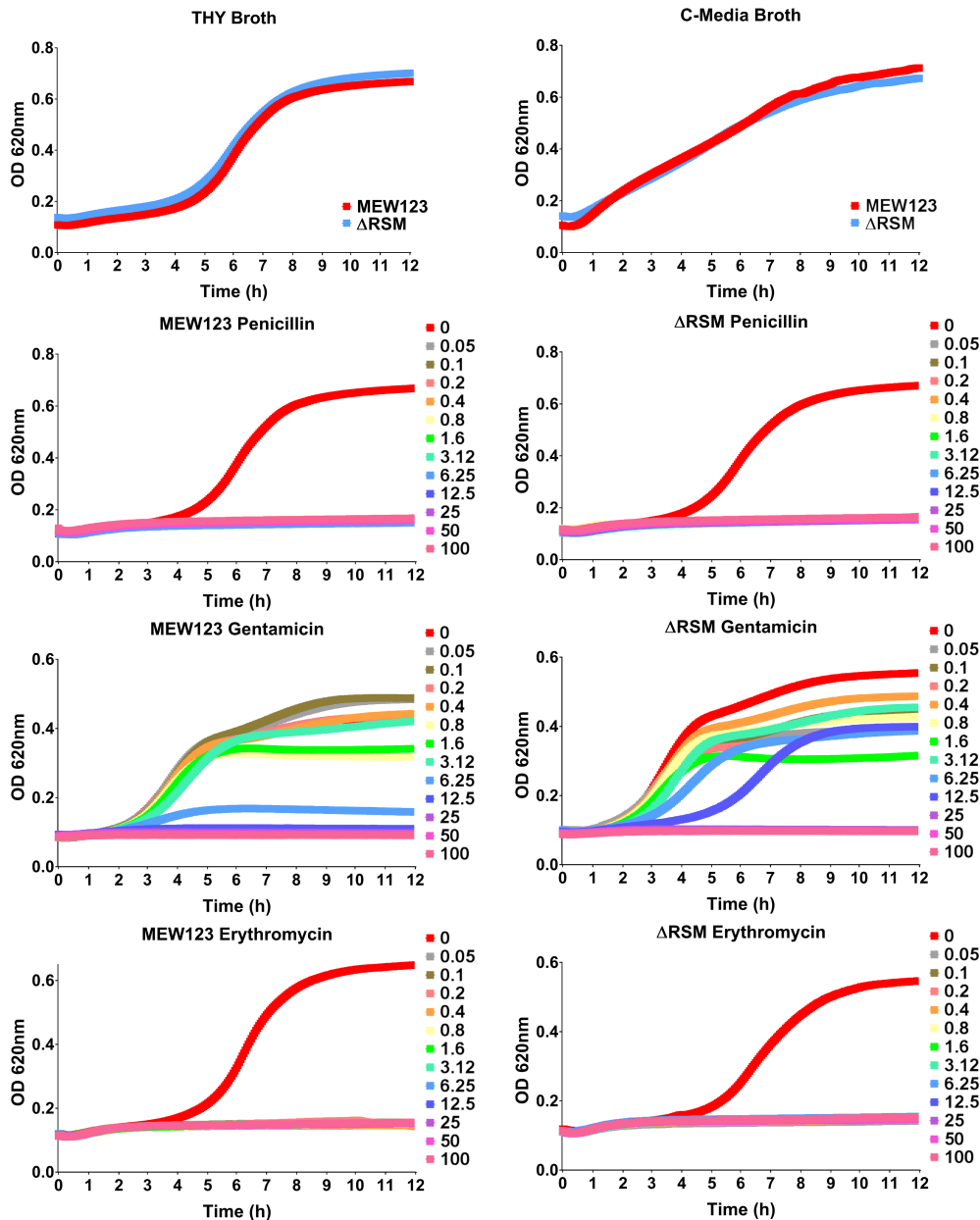
Fig. 2.8. Growth curves of WT and ΔRSM mutant are similar in THY broth and C-media, and in THY-broth containing penicillin, gentamicin, or erythromycin. Growth was monitored using a Synergy HTX plate reader (BioTek) in 96 well plates (Greiner Bio-One). Briefly, 4µl of overnight culture grown in THY broth was inoculated into 200µl of the described fresh media, with identical strains and conditions measured in at least triplicate. Growth was at 37°C, room air, in static conditions for 12 hours (time on X-axis) and OD_{620nm} was measured every 3 seconds (Y-axis). Where indicated, antibiotic was added to THY broth at concentrations ranging from 0.05 µg/mL to 100 µg/mL, with concentrations shown in the color-coded key on the right [µg/mL].
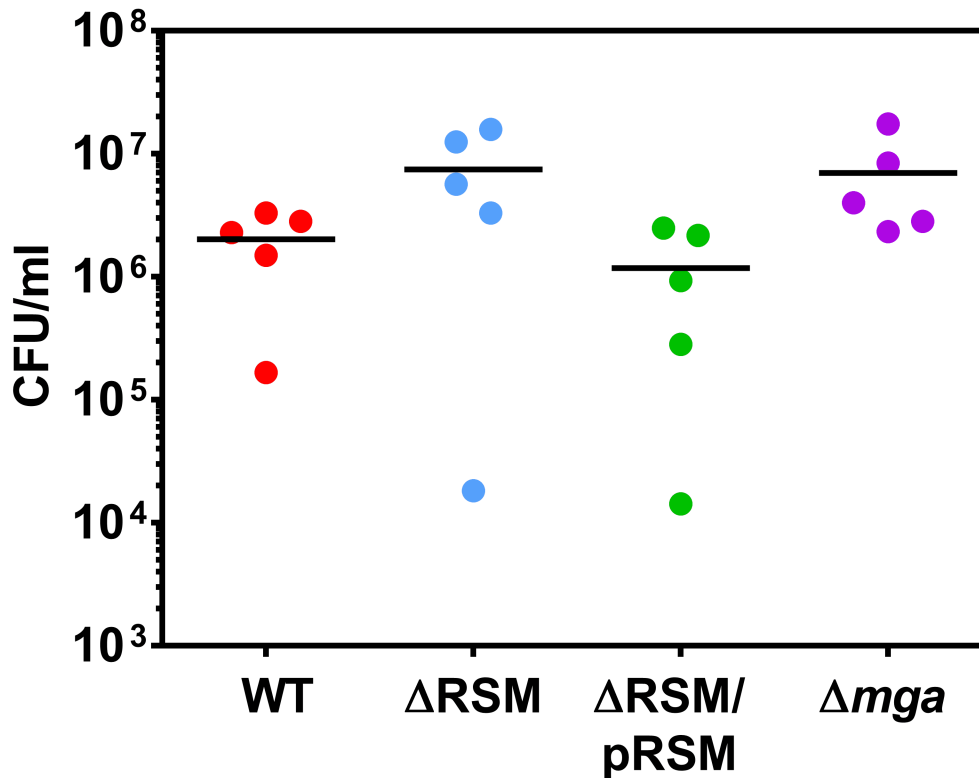
Fig. 2.9. Streptococcal CFU counts in skin lesions at day 4 post-infection. Skin lesions from infected mice were dissected on day#4 post-infection and homogenized in 1 mL sterile PBS. The homogenate was serially diluted in sterile PBS and plated onto THY agar plates containing streptomycin (1000 µg/mL) to select for *S. pyogenes* strain MEW123 and its mutants. Each point represents the CFU counts from one mouse lesion, with black bars indicating mean CFU values for that group. Groups are MEW123 (WT), MEW513 (ΔRSM), MEW552 (ΔRSM/pRSM), and MEW480 (Δ*mga*). The Mann-Whitney U-test was used to test for statistical significance, though none were statistically different from the WT strain.

Table 2.1. Strains used in this study.

Strain/Plasmids	Description	Source
Strains		
Escherichia coli		
DH5α	Standard cloning vector, *recA1 endA1 hsdR17*	Invitrogen
Streptococcus pyogenes		
MEW123	Streptomycin-resistant clone of pediatric throat isolate, serotype M28	[12]
MEW380	MEW123 transformed by plasmid pIL09 to insertionally-inactivate scpA gene by spectinomycin resistance marker (Ω*scpA*)	This study
MEW409	MEW123 transformed with plasmid pIL03 to insertionally-inactivate M protein, *emm* gene, by spectinomycin resistance marker (Ω*emm28*)	This study
MEW480	MEW123 with in-frame deletion of *mga* (Δ*mga*) by allelic exchange after transformation with plasmid pIL01	This study
MEW489	MEW123 transformed with plasmid pKJ19 to insertionally-inactivate restriction endonuclease with spectinomycin resistance marker (ΩRE)	This study
MEW513	MEW123 with restriction-modification system in-frame deletion (ΔRSM) by allelic exchange after transformation with pKJ24	This study
MEW552	MEW513 with RSM operon complemented *in trans* on plasmid pEH01 (ΔRSM/pRSM)	This study
HSC5	Serotype M14 reference strain	[23]
HSC5 Ωemm	HSC5 mutant with the M protein, *emm14* gene, disrupted by spectinomycin resistance marker (Ω*emm14*)	[24]
Plasmids		
pJRS233	Low-copy *E. coli* to *S. pyogenes* temperature-sensitive vector for allelic replacement (erythromycin-resistant, Erm^R)	[25]
pGCP213	High-copy *E. coli* to *S. pyogenes* temperature-sensitive vector for allelic replacement (Erm^R)	[26]
pJoy3	*E. coli* to *S. pyogenes* shuttle vector (chloramphenicol-resistant, Chlor^R)	[27]
pSpc18	integration vector containing aad9 (spectinomycin resistance gene from *Enterococcus faecalis*) (Spc^R)	[28]
pIL01	pJRS233 with in-frame deletion of *mga* (Erm^R)	This study
pIL03	pSpc18 with *emm28* gene fragment disruption (Spc^R)	This study
pIL09	pSpc18 with *scpA* gene fragment disruption (Spc^R)	This study
pKJ19	pSpc18 with restriction endonuclease gene fragment disruption (Spc^R)	This study
pKJ24	pGCP213 with in-frame deletion of restriction-modification gene cluster (Erm^R)	This study
pEH01	pJoy3 with intact RSM operon cloned for complementation (Chlor^R)	This study

Table 2.2. Motif analysis of modified bases from PacBio SMRT sequencing in wild type *S. pyogenes* strain MEW123.

Motif*	Type	Motifs in genome	Mean modQV	Mean coverage
GC**A**NNNNNNNTTYG	m6A	412	332.64	245.02
CRA**A**NNNNNNNTGC	m6A	412	282.18	237.32
THTWGAAGA	unknown	410	44.74	240.57
ANDYVGCAD	m6A	3502	86.17	241.02
TNRRDDDG	unknown	34390	44.88	235.68
TNNNDNNH	unknown	746710	47.81	235.15
THRGCNTWNH	unknown	3107	43.39	237.84
AGNNAVNW	m6A	32122	78.13	239.29
TNNNCRV	unknown	77468	45.11	235.03
V**A**HNBAVYW	m6A	27467	76.34	242.45
THNNDVNG	unknown	96451	43.43	237.89

*modified base is bolded and underlined

Table 2.3. Motif analysis of modified bases from PacBio SMRT sequencing in ΔRSM strain.

Motif*	Type	Motifs in genome	Mean modQV	Mean coverage
TYTWGARGR	unknown	701	46.56	263.00
D**A**GKBANYW	m6A	5826	88.94	256.78
ANNYRGYA	m6A	8371	85.54	259.83
G**A**HBBAACA	m6A	499	125.47	268.90
TNNNDNNH	unknown	746710	49.02	251.20
TNRRDDDG	unknown	34390	45.12	251.97
THRGCNTH	unknown	5464	43.69	253.61
TNNNCRV	unknown	77468	45.81	251.94
D**T**NRVCBNHNH	unknown	29343	44.13	251.93
AHSBAMYW	m6A	9198	75.01	265.11
THNNDVNG	unknown	96451	44.13	256.11

*modified base is bolded and underlined

Table 2.4. Differentially expressed genes in ΔRSM strain compared to WT.

Cluster	MEW123 Locus	Fold-Change (Log2)	Ave Expr	adj.P.Val	Prob	Product Annotation
1	AWM59_00940	-3.88	1.98	0.00	1.00	hypothetical protein
1	AWM59_00945	-3.16	1.30	0.00	0.98	hypothetical protein
2	AWM59_02865	-1.14	6.80	0.01	0.89	cell surface protein
2	AWM59_02875	-1.12	3.04	0.05	0.45	conjugal transfer protein
2	AWM59_02880	-1.29	4.30	0.04	0.46	cell division protein (FtsK)
2	AWM59_02900	-1.14	2.87	0.01	0.87	hypothetical protein

3	AWM59_07395	-1.23	6.09	0.03	0.55	hypothetical protein
3	AWM59_07430	-4.12	5.14	0.00	1.00	hypothetical protein
4	AWM59_07895	-10.73	4.10	0.00	1.00	Endonuclease *hsdR*
4	AWM59_07900	-11.69	3.07	0.00	1.00	Specificity *hsdS*
4	AWM59_07905	-10.79	3.98	0.00	1.00	Methyltransferase *hsdM*
5	AWM59_08190	-5.60	5.26	0.00	0.99	Collagen-like surface protein A (SclA)
5	AWM59_08310	-2.38	6.45	0.00	0.99	LPXTG anchor domain surface protein
5	AWM59_08315	-4.32	5.53	0.00	1.00	peptidase C5 (ScpA)
5	AWM59_08325	-8.65	5.31	0.00	1.00	M28 protein (M28)
5	AWM59_08330	-8.01	6.54	0.00	1.00	emm-like protein
5	AWM59_08335	-1.24	6.39	0.04	0.45	Trans-Acting Positive Regulator (Mga)
6	AWM59_08385	-4.57	6.93	0.00	1.00	fibronectin-binding protein (SfbX)
6	AWM59_08390	-5.11	7.31	0.00	1.00	Serum Opacity Factor (SOF)
6	AWM59_08395	-6.47	3.57	0.00	1.00	hypothetical protein

Table 2.5. Primers used in this study.

Primer	Sequence 5'-3' (restriction sites underlined)
MEW123 Del-RSM F1	ATAT<u>GAATTC</u>GGTTTTTTTGGTAAAAAACTTTTTTAGCA
MEW123 Del-RSM R2	TTTTTTGGTCTTTTTTAATCCCCATTCGACATGATA
MEW123 Del-RSM F3	TGGGGATTAAAAAAGACCAAAAAACACCACAGTAGA
MEW123 Del-RSM R4	ATAT<u>AAGCTT</u>TTAATTTAACAAATATTTCTAAAGAAAATGGATTGG
pJoy3 123 RSM F	ACCATTATTGTGAGGAACTGCGTTACCGATCCCTTAAAAG
pJoy3 123 RSM R	ACCGATAGCACCCGCGCATGGATGAGATGATTCTATTTTGATTTATAG
123_7895 F	TAAAACGACGGCCAGTGCCAATCTCTTAGAAACAGGTGAAAG
123_7895 R	ATTCGAGCTCGGTACCCGGGAGATATCATTTTGCGCATAG
M28 Emm Hindlll F	CCC<u>AAGCTT</u>ATAAACAGTATTCGCTTAGAAAATTAAAAACAGG
M28 Emm BamHI R	CGC<u>GGATCC</u>GTTAGCTGCTTCGCCTGTTGACGGTAACG
M28 ScpA SalI F	CCC<u>GTCGAC</u>CTCAATGCACAATCAGACATTAAAGC
M28 ScpA SacI R	CCC<u>GAGCTC</u>TCAATATCGCCACGTTCAATAAGG
M28 Mga 5' SalI	CCC<u>GTCGAC</u>TGACAATAATGTCACAGAT
M28 Mga 5' SOE R	TTGTTGGCTAGTAAACAATTTACTTACATGC
M28 Mga 3' SOE F	GTTTACTAGCCAACAAGCAACATCATCATAGGATTTCAGACG
M28 Mga 3' BamHI	CGC<u>GGATCC</u>CGCTCTTCGAATACTTTGTT
Emm28 RT-PCR F	CAGACTTAGCAGAAGCAAATAGC
Emm28 RT-PCR R	CAGCTTGTTTAGCCAATTGCTC
Mga RT-PCR F	CTTATCTACCCTCAAACGCCTC
Mga RT-PCR R	CGAATTTGCCTCTCATCTCCTG
ScpA RT-PCR F	CACTGATTTTGATGTGATTGTAGACAA
ScpA RT-PCR R	ATGCAAGTGTCAAACGACGATCT
recA RT-PCR F	ATTGATTGATTCTGGTGCGG
recA RT-PCR R	ATTTACGCATGGCCTGACTC
MEW M13 F	CAGGGTTTTCCCAGTCACGAC
MEW M13 R	GAGCGGATAACAATTTCACACAGG

References

1. Siegfried Z, Simon I. DNA methylation and gene expression. Wiley Interdiscip Rev Syst Biol Med. 2010;2(3):362-71. doi: 10.1002/wsbm.64. PubMed PMID: 20836034.
2. Gray SG, Eriksson T, Ekstrom TJ. Methylation, gene expression and the chromatin connection in cancer (review). Int J Mol Med. 1999;4(4):333-50. Epub 1999/09/24. PubMed PMID: 10493973.
3. Virani S, Colacino JA, Kim JH, Rozek LS. Cancer epigenetics: a brief review. ILAR J. 2012;53(3-4):359-69. Epub 2013/06/08. doi: 10.1093/ilar.53.3-4.359. PubMed PMID: 23744972; PubMed Central PMCID: PMCPMC4021643.
4. Sanchez-Romero MA, Cota I, Casadesus J. DNA methylation in bacteria: from the methyl group to the methylome. Curr Opin Microbiol. 2015;25:9-16. doi: 10.1016/j.mib.2015.03.004. PubMed PMID: 25818841.
5. Blow MJ, Clark TA, Daum CG, Deutschbauer AM, Fomenkov A, Fries R, *et al*. The Epigenomic Landscape of Prokaryotes. PLoS Genet. 2016;12(2):e1005854. doi: 10.1371/journal.pgen.1005854. PubMed PMID: 26870957; PubMed Central PMCID: PMC4752239.
6. Vasu K, Nagaraja V. Diverse functions of restriction-modification systems in addition to cellular defense. Microbiol Mol Biol Rev. 2013;77(1):53-72. Epub 2013/03/09. doi: 10.1128/MMBR.00044-12. PubMed PMID: 23471617; PubMed Central PMCID: PMCPMC3591985.
7. Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, *et al*. Real-time DNA sequencing from single polymerase molecules. Science. 2009;323(5910):133-8. doi: 10.1126/science.1162986. PubMed PMID: 19023044.
8. McCarthy A. Third generation DNA sequencing: pacific biosciences' single molecule real time technology. Chem Biol. 2010;17(7):675-6. Epub 2010/07/28. doi: 10.1016/j.chembiol.2010.07.004. PubMed PMID: 20659677.
9. Kelleher P, Murphy J, Mahony J, van Sinderen D. Identification of DNA Base Modifications by Means of Pacific Biosciences RS Sequencing Technology. Methods Mol Biol. 2018;1681:127-37. Epub 2017/11/15. doi: 10.1007/978-1-4939-7343-9_10. PubMed PMID: 29134592.
10. Flusberg BA, Webster DR, Lee JH, Travers KJ, Olivares EC, Clark TA, *et al*. Direct detection of DNA methylation during single-molecule, real-time sequencing. Nat Methods. 2010;7(6):461-5. Epub 2010/05/11. doi: 10.1038/nmeth.1459. PubMed PMID: 20453866; PubMed Central PMCID: PMCPMC2879396.
11. Clark TA, Murray IA, Morgan RD, Kislyuk AO, Spittle KE, Boitano M, *et al*. Characterization of DNA methyltransferase specificities using single-molecule, real-time DNA sequencing. Nucleic Acids Res. 2012;40(4):e29. Epub 2011/12/14. doi: 10.1093/nar/gkr1146. PubMed PMID: 22156058; PubMed Central PMCID: PMCPMC3287169.
12. Jacob KM, Spilker T, LiPuma JJ, Dawid SR, Watson ME, Jr. Complete Genome Sequence of *emm*28 Type *Streptococcus pyogenes* MEW123, a Streptomycin-Resistant Derivative of a Clinical Throat Isolate Suitable for Investigation of Pathogenesis. Genome Announc. 2016;4(2). doi: 10.1128/genomeA.00136-16. PubMed PMID: 26988051; PubMed Central PMCID: PMC4796130.

13.	Jacob KM, Spilker T, LiPuma JJ, Dawid SR, Watson ME, Jr. Complete Genome Sequence of *emm*4 *Streptococcus pyogenes* MEW427, a Throat Isolate from a Child Meeting Clinical Criteria for Pediatric Autoimmune Neuropsychiatric Disorders Associated with Streptococcus (PANDAS). Genome Announc. 2016;4(2). doi: 10.1128/genomeA.00127-16. PubMed PMID: 26988046; PubMed Central PMCID: PMC4796125.

14.	Cunningham MW. Pathogenesis of group A streptococcal infections. Clin Microbiol Rev. 2000;13(3):470-511. PubMed PMID: 10885988.

15.	Areschoug T, Carlsson F, Stalhammar-Carlemalm M, Lindahl G. Host-pathogen interactions in *Streptococcus pyogenes* infections, with special reference to puerperal fever and a comment on vaccine development. Vaccine. 2004;22 Suppl 1:S9-S14. doi: 10.1016/j.vaccine.2004.08.010. PubMed PMID: 15576204.

16.	Walker MJ, Barnett TC, McArthur JD, Cole JN, Gillen CM, Henningham A, *et al*. Disease manifestations and pathogenic mechanisms of Group A Streptococcus. Clin Microbiol Rev. 2014;27(2):264-301. doi: 10.1128/CMR.00101-13. PubMed PMID: 24696436; PubMed Central PMCID: PMC3993104.

17.	Kreikemeyer B, McIver KS, Podbielski A. Virulence factor regulation and regulatory networks in *Streptococcus pyogenes* and their impact on pathogen-host interactions. Trends Microbiol. 2003;11(5):224-32. PubMed PMID: 12781526.

18.	Vega LA, Malke H, McIver KS. Virulence-Related Transcriptional Regulators of *Streptococcus pyogenes*. In: Ferretti JJ, Stevens DL, Fischetti VA, editors. Streptococcus pyogenes : Basic Biology to Clinical Manifestations. Oklahoma City (OK)2016.

19.	Roberts RJ, Belfort M, Bestor T, Bhagwat AS, Bickle TA, Bitinaite J, *et al*. A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. Nucleic Acids Res. 2003;31(7):1805-12. Epub 2003/03/26. PubMed PMID: 12654995; PubMed Central PMCID: PMCPMC152790.

20.	Ferretti JJ, McShan WM, Ajdic D, Savic DJ, Savic G, Lyon K, *et al*. Complete genome sequence of an M1 strain of *Streptococcus pyogenes*. Proc Natl Acad Sci U S A. 2001;98(8):4658-63. Epub 2001/04/11. doi: 10.1073/pnas.071559398. PubMed PMID: 11296296; PubMed Central PMCID: PMC31890.

21.	Okada R, Matsumoto M, Zhang Y, Isaka M, Tatsuno I, Hasegawa T. Emergence of type I restriction modification system-negative *emm1* type *Streptococcus pyogenes* clinical isolates in Japan. APMIS. 2014;122(10):914-21. Epub 2014/10/31. PubMed PMID: 25356467.

22.	Euler CW, Ryan PA, Martin JM, Fischetti VA. M.SpyI, a DNA methyltransferase encoded on a *mefA* chimeric element, modifies the genome of *Streptococcus pyogenes*. J Bacteriol. 2007;189(3):1044-54. Epub 2006/11/07. doi: 10.1128/JB.01411-06. PubMed PMID: 17085578; PubMed Central PMCID: PMCPMC1797290.

23.	Hanski E, Horwitz PA, Caparon MG. Expression of protein F, the fibronectin-binding protein of *Streptococcus pyogenes* JRS4, in heterologous streptococcal and enterococcal strains promotes their adherence to respiratory epithelial cells.

Infect Immun. 1992;60(12):5119-25. Epub 1992/12/01. PubMed PMID: 1452345; PubMed Central PMCID: PMC258286.

24.   Cho KH, Caparon MG. Patterns of virulence gene expression differ between biofilm and tissue communities of *Streptococcus pyogenes*. Mol Microbiol. 2005;57(6):1545-56. Epub 2005/09/02. doi: 10.1111/j.1365-2958.2005.04786.x. PubMed PMID: 16135223.

25.   Perez-Casal J, Price JA, Maguin E, Scott JR. An M protein with a single C repeat prevents phagocytosis of *Streptococcus pyogenes*: use of a temperature-sensitive shuttle vector to deliver homologous sequences to the chromosome of *S. pyogenes*. Mol Microbiol. 1993;8(5):809-19. Epub 1993/05/01. PubMed PMID: 8355608.

26.   Nielsen HV, Guiton PS, Kline KA, Port GC, Pinkner JS, Neiers F, *et al*. The metal ion-dependent adhesion site motif of the *Enterococcus faecalis* EbpA pilin mediates pilus function in catheter-associated urinary tract infection. MBio. 2012;3(4):e00177-12. doi: 10.1128/mBio.00177-12. PubMed PMID: 22829678; PubMed Central PMCID: PMC3419518.

27.   Ghosh J, Caparon MG. Specificity of *Streptococcus pyogenes* NAD(+) glycohydrolase in cytolysin-mediated translocation. Mol Microbiol. 2006;62(4):1203-14. Epub 2006/10/18. doi: 10.1111/j.1365-2958.2006.05430.x. PubMed PMID: 17042787.

28.   Lyon WR, Madden JC, Levin JC, Stein JL, Caparon MG. Mutation of luxS affects growth and virulence factor expression in *Streptococcus pyogenes*. Mol Microbiol. 2001;42(1):145-57. Epub 2001/10/27. PubMed PMID: 11679074.

29.   Loenen WA, Dryden DT, Raleigh EA, Wilson GG. Type I restriction enzymes and their relatives. Nucleic Acids Res. 2014;42(1):20-44. Epub 2013/09/27. doi: 10.1093/nar/gkt847. PubMed PMID: 24068554; PubMed Central PMCID: PMCPMC3874165.

30.   Dryden DT, Cooper LP, Murray NE. Purification and characterization of the methyltransferase from the type 1 restriction and modification system of *Escherichia coli* K12. J Biol Chem. 1993;268(18):13228-36. Epub 1993/06/25. PubMed PMID: 8514761.

31.   Suri B, Shepherd JC, Bickle TA. The EcoA restriction and modification system of *Escherichia coli* 15T-: enzyme structure and DNA recognition sequence. EMBO J. 1984;3(3):575-9. Epub 1984/03/01. PubMed PMID: 6325176; PubMed Central PMCID: PMCPMC557390.

32.   Makovets S, Titheradge AJ, Murray NE. ClpX and ClpP are essential for the efficient acquisition of genes specifying type IA and IB restriction systems. Mol Microbiol. 1998;28(1):25-35. Epub 1998/05/21. PubMed PMID: 9593294.

33.   Fang G, Munera D, Friedman DI, Mandlik A, Chao MC, Banerjee O, *et al*. Genome-wide mapping of methylated adenine residues in pathogenic *Escherichia coli* using single-molecule real-time sequencing. Nat Biotechnol. 2012;30(12):1232-9. Epub 2012/11/10. doi: 10.1038/nbt.2432. PubMed PMID: 23138224; PubMed Central PMCID: PMCPMC3879109.

34.   Anjum A, Brathwaite KJ, Aidley J, Connerton PL, Cummings NJ, Parkhill J, *et al*. Phase variation of a Type IIG restriction-modification enzyme alters site-specific methylation patterns and gene expression in *Campylobacter jejuni* strain

NCTC11168. Nucleic Acids Res. 2016;44(10):4581-94. Epub 2016/01/21. doi: 10.1093/nar/gkw019. PubMed PMID: 26786317; PubMed Central PMCID: PMCPMC4889913.

35. Casselli T, Tourand Y, Scheidegger A, Arnold WK, Proulx A, Stevenson B, *et al*. DNA Methylation by Restriction Modification Systems Affects the Global Transcriptome Profile in *Borrelia burgdorferi*. J Bacteriol. 2018. Epub 2018/09/27. doi: 10.1128/JB.00395-18. PubMed PMID: 30249703.

36. Lenhart JS, Pillon MC, Guarne A, Biteen JS, Simmons LA. Mismatch repair in Gram-positive bacteria. Res Microbiol. 2016;167(1):4-12. Epub 2015/09/08. doi: 10.1016/j.resmic.2015.08.006. PubMed PMID: 26343983.

37. McIver KS, Scott JR. Role of mga in growth phase regulation of virulence genes of the group A streptococcus. J Bacteriol. 1997;179(16):5178-87. PubMed PMID: 9260962; PubMed Central PMCID: PMCPMC179378.

38. Hondorp ER, McIver KS. The Mga virulence regulon: infection where the grass is greener. Mol Microbiol. 2007;66(5):1056-65. PubMed PMID: 18001346.

39. Vega LA, Malke H, McIver KS. Virulence-Related Transcriptional Regulators of *Streptococcus pyogenes*. In: Ferretti JJ, Stevens DL, Fischetti VA, editors. Streptococcus pyogenes: Basic Biology to Clinical Manifestations. Oklahoma City (OK)2016.

40. Watson ME, Jr., Neely MN, Caparon MG. Animal Models of *Streptococcus pyogenes* Infection. In: Ferretti JJ, Stevens DL, Fischetti VA, editors. Streptococcus pyogenes: Basic Biology to Clinical Manifestations. Oklahoma City (OK)2016.

41. Bunce C, Wheeler L, Reed G, Musser J, Barg N. Murine model of cutaneous infection with gram-positive cocci. Infect Immun. 1992;60(7):2636-40. Epub 1992/07/01. PubMed PMID: 1612733; PubMed Central PMCID: PMC257214.

42. Fielding CA, McLoughlin RM, McLeod L, Colmont CS, Najdovska M, Grail D, *et al*. IL-6 regulates neutrophil trafficking during acute inflammation via STAT3. J Immunol. 2008;181(3):2189-95. Epub 2008/07/22. PubMed PMID: 18641358.

43. Wright HL, Cross AL, Edwards SW, Moots RJ. Effects of IL-6 and IL-6 blockade on neutrophil function in vitro and in vivo. Rheumatology. 2014;53(7):1321-31. Epub 2014/03/13. doi: 10.1093/rheumatology/keu035. PubMed PMID: 24609058.

44. Cho JS, Pietras EM, Garcia NC, Ramos RI, Farzam DM, Monroe HR, *et al*. IL-17 is essential for host defense against cutaneous *Staphylococcus aureus* infection in mice. J Clin Invest. 2010;120(5):1762-73. Epub 2010/04/07. doi: 10.1172/JCI40891. PubMed PMID: 20364087; PubMed Central PMCID: PMC2860944.

45. Staali L, Morgelin M, Bjorck L, Tapper H. *Streptococcus pyogenes* expressing M and M-like surface proteins are phagocytosed but survive inside human neutrophils. Cell Microbiol. 2003;5(4):253-65. PubMed PMID: 12675683.

46. Kihlberg BM, Cooney J, Caparon MG, Olsen A, Bjorck L. Biological properties of a *Streptococcus pyogenes* mutant generated by Tn916 insertion in *mga*. Microb Pathog. 1995;19(5):299-315. PubMed PMID: 8778565.

47. Caparon MG, Scott JR. Identification of a gene that regulates expression of M protein, the major virulence determinant of group A streptococci. Proc Natl Acad

Sci U S A. 1987;84(23):8677-81. Epub 1987/12/01. PubMed PMID: 2446327; PubMed Central PMCID: PMC299609.

48.   Caswell CC, Lukomska E, Seo NS, Hook M, Lukomski S. Scl1-dependent internalization of group A Streptococcus via direct interactions with the alpha2beta(1) integrin enhances pathogen survival and re-emergence. Mol Microbiol. 2007;64(5):1319-31. Epub 2007/06/05. doi: 10.1111/j.1365-2958.2007.05741.x. PubMed PMID: 17542923.

49.   Terao Y, Kawabata S, Kunitomo E, Murakami J, Nakagawa I, Hamada S. Fba, a novel fibronectin-binding protein from *Streptococcus pyogenes*, promotes bacterial entry into epithelial cells, and the fba gene is positively transcribed under the Mga regulator. Mol Microbiol. 2001;42(1):75-86. Epub 2001/10/27. PubMed PMID: 11679068.

50.   Green NM, Beres SB, Graviss EA, Allison JE, McGeer AJ, Vuopio-Varkila J, *et al*. Genetic diversity among type emm28 group A streptococcus strains causing invasive infections and pharyngitis. J Clin Microbiol. 2005;43(8):4083-91. Epub 2005/08/06. doi: 10.1128/JCM.43.8.4083-4091.2005. PubMed PMID: 16081955; PubMed Central PMCID: PMC1233891.

51.   Green NM, Zhang S, Porcella SF, Nagiec MJ, Barbian KD, Beres SB, *et al*. Genome sequence of a serotype M28 strain of group A streptococcus: potential new insights into puerperal sepsis and bacterial disease specificity. J Infect Dis. 2005;192(5):760-70. Epub 2005/08/10. doi: 10.1086/430618. PubMed PMID: 16088825.

52.   Bruins MJ, Damoiseaux RA, Ruijs GJ. Association between group A beta-haemolytic streptococci and vulvovaginitis in adult women: a case-control study. Eur J Clin Microbiol Infect Dis. 2009;28(8):1019-21. PubMed PMID: 19343383.

53.   Anteby EY, Yagel S, Hanoch J, Shapiro M, Moses AE. Puerperal and intrapartum group A streptococcal infection. Infect Dis Obstet Gynecol. 1999;7(6):276-82. Epub 1999/12/22. doi: 10.1155/S1064744999000514. PubMed PMID: 10598916; PubMed Central PMCID: PMC1784762.

54.   Stalhammar-Carlemalm M, Areschoug T, Larsson C, Lindahl G. The R28 protein of *Streptococcus pyogenes* is related to several group B streptococcal surface proteins, confers protective immunity and promotes binding to human epithelial cells. Mol Microbiol. 1999;33(1):208-19. PubMed PMID: 10411737.

55.   Watson ME, Jr., Nielsen HV, Hultgren SJ, Caparon MG. Murine Vaginal Colonization Model for Investigating Asymptomatic Mucosal Carriage of *Streptococcus pyogenes*. Infect Immun. 2013;81(5):1606-17. Epub 2013/03/06. doi: 10.1128/IAI.00021-13. PubMed PMID: 23460515.

56.   Balbontin R, Rowley G, Pucciarelli MG, Lopez-Garrido J, Wormstone Y, Lucchini S, *et al*. DNA adenine methylation regulates virulence gene expression in *Salmonella enterica* serovar Typhimurium. J Bacteriol. 2006;188(23):8160-8. Epub 2006/09/26. doi: 10.1128/JB.00847-06. PubMed PMID: 16997949; PubMed Central PMCID: PMCPMC1698197.

57.   Murray IA, Clark TA, Morgan RD, Boitano M, Anton BP, Luong K, *et al*. The methylomes of six bacteria. Nucleic Acids Res. 2012;40(22):11450-62. Epub 2012/10/05. doi: 10.1093/nar/gks891. PubMed PMID: 23034806; PubMed Central PMCID: PMCPMC3526280.

58. Estibariz I, Overmann A, Ailloud F, Krebes J, Josenhans C, Suerbaum S. The core genome m5C methyltransferase JHP1050 (M.Hpy99III) plays an important role in orchestrating gene expression in *Helicobacter pylori*. Nucleic Acids Res. 2019. Epub 2019/01/10. doi: 10.1093/nar/gky1307. PubMed PMID: 30624738.

59. Watson ME, Jr., Jarisch J, Smith AL. Inactivation of deoxyadenosine methyltransferase (*dam*) attenuates *Haemophilus influenzae* virulence. Mol Microbiol. 2004;53(2):651-64. doi: 10.1111/j.1365-2958.2004.04140.x. PubMed PMID: 15228541.

60. Hu C, Zhao Y, Sun H, Yang Y. Synergism of Dam, MutH, and MutS in methylation-directed mismatch repair in *Escherichia coli*. Mutat Res. 2017;795:31-3. Epub 2017/01/21. doi: 10.1016/j.mrfmmm.2016.12.002. PubMed PMID: 28107644; PubMed Central PMCID: PMCPMC5299848.

61. Morgan RD, Bhatia TK, Lovasco L, Davis TB. MmeI: a minimal Type II restriction-modification system that only modifies one DNA strand for host protection. Nucleic Acids Res. 2008;36(20):6558-70. Epub 2008/10/22. doi: 10.1093/nar/gkn711. PubMed PMID: 18931376; PubMed Central PMCID: PMCPMC2582602.

62. Willemse N, Schultsz C. Distribution of Type I Restriction-Modification Systems in *Streptococcus suis*: An Outlook. Pathogens. 2016;5(4). Epub 2016/11/22. doi: 10.3390/pathogens5040062. PubMed PMID: 27869755; PubMed Central PMCID: PMCPMC5198162.

63. Furuta Y, Namba-Fukuyo H, Shibata TF, Nishiyama T, Shigenobu S, Suzuki Y, *et al*. Methylome diversification through changes in DNA methyltransferase sequence specificity. PLoS Genet. 2014;10(4):e1004272. Epub 2014/04/12. doi: 10.1371/journal.pgen.1004272. PubMed PMID: 24722038; PubMed Central PMCID: PMCPMC3983042.

64. Sitaraman R, Dybvig K. The *hsd* loci of *Mycoplasma pulmonis*: organization, rearrangements and expression of genes. Mol Microbiol. 1997;26(1):109-20. Epub 1998/01/31. PubMed PMID: 9383194.

65. Li J, Li JW, Feng Z, Wang J, An H, Liu Y, *et al*. Epigenetic Switch Driven by DNA Inversions Dictates Phase Variation in *Streptococcus pneumoniae*. PLoS Pathog. 2016;12(7):e1005762. Epub 2016/07/20. doi: 10.1371/journal.ppat.1005762. PubMed PMID: 27427949; PubMed Central PMCID: PMCPMC4948785.

66. McIver KS, Myles RL. Two DNA-binding domains of Mga are required for virulence gene activation in the group A streptococcus. Mol Microbiol. 2002;43(6):1591-601. Epub 2002/04/16. PubMed PMID: 11952907.

67. Almengor AC, McIver KS. Transcriptional activation of sclA by Mga requires a distal binding site in *Streptococcus pyogenes*. J Bacteriol. 2004;186(23):7847-57. Epub 2004/11/18. doi: 10.1128/JB.186.23.7847-7857.2004. PubMed PMID: 15547255; PubMed Central PMCID: PMCPMC529090.

68. Wexler DE, Cleary PP. Purification and characteristics of the streptococcal chemotactic factor inactivator. Infect Immun. 1985;50(3):757-64. Epub 1985/12/01. PubMed PMID: 3905613; PubMed Central PMCID: PMCPMC261145.

69. Ji Y, McLandsborough L, Kondagunta A, Cleary PP. C5a peptidase alters clearance and trafficking of group A streptococci by infected mice. Infect Immun.

1996;64(2):503-10. Epub 1996/02/01. PubMed PMID: 8550199; PubMed Central PMCID: PMC173793.

70. Li J, Zhu H, Feng W, Liu M, Song Y, Zhang X, *et al*. Regulation of inhibition of neutrophil infiltration by the two-component regulatory system CovRS in subcutaneous murine infection with group A streptococcus. Infect Immun. 2013;81(3):974-83. Epub 2013/01/16. doi: 10.1128/IAI.01218-12. PubMed PMID: 23319556; PubMed Central PMCID: PMCPMC3584857.

71. Lynskey NN, Reglinski M, Calay D, Siggins MK, Mason JC, Botto M, *et al*. Multi-functional mechanisms of immune evasion by the streptococcal complement inhibitor C5a peptidase. PLoS Pathog. 2017;13(8):e1006493. Epub 2017/08/15. doi: 10.1371/journal.ppat.1006493. PubMed PMID: 28806402; PubMed Central PMCID: PMCPMC5555575.

72. National Research Council (U.S.). Committee for the Update of the Guide for the Care and Use of Laboratory Animals. Guide for the care and use of laboratory animals. Washington, D.C.: National Academies Press; 2011. Available from: http://lib.myilibrary.com?id=297592.

73. Lyon WR, Gibson CM, Caparon MG. A role for trigger factor and an *rgg*-like regulator in the transcription, secretion and processing of the cysteine proteinase of *Streptococcus pyogenes*. Embo J. 1998;17(21):6263-75. Epub 1998/11/03. doi: 10.1093/emboj/17.21.6263. PubMed PMID: 9799235; PubMed Central PMCID: PMC1170952.

74. Kietzman CC, Caparon MG. CcpA and LacD.1 affect temporal regulation of *Streptococcus pyogenes* virulence genes. Infect Immun. 2010;78(1):241-52. Epub 2009/10/21. doi: 10.1128/IAI.00746-09. PubMed PMID: 19841076; PubMed Central PMCID: PMC2798178.

75. Bryksin AV, Matsumura I. Overlap extension PCR cloning: a simple and reliable way to create recombinant plasmids. Biotechniques. 2010;48(6):463-5. Epub 2010/06/24. doi: 10.2144/000113418. PubMed PMID: 20569222; PubMed Central PMCID: PMC3121328.

76. Caparon MG, Stephens DS, Olsen A, Scott JR. Role of M protein in adherence of group A streptococci. Infect Immun. 1991;59(5):1811-7. Epub 1991/05/01. PubMed PMID: 2019444; PubMed Central PMCID: PMC257920.

77. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. 2002;30(1):207-10. Epub 2001/12/26. PubMed PMID: 11752295; PubMed Central PMCID: PMCPMC99122.

78. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, *et al*. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47. doi: 10.1093/nar/gkv007. PubMed PMID: 25605792; PubMed Central PMCID: PMC4402510.

79. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26(1):139-40. doi: 10.1093/bioinformatics/btp616. PubMed PMID: 19910308; PubMed Central PMCID: PMC2796818.

80. Gaujoux R, Seoighe C. A flexible R package for nonnegative matrix factorization. BMC bioinformatics. 2010;11:367. doi: 10.1186/1471-2105-11-367. PubMed PMID: 20598126; PubMed Central PMCID: PMC2912887.

81. Schmittgen TD, Livak KJ. Analyzing real-time PCR data by the comparative C(T) method. Nat Protoc. 2008;3(6):1101-8. Epub 2008/06/13. PubMed PMID: 18546601.

82. Brenot A, King KY, Janowiak B, Griffith O, Caparon MG. Contribution of glutathione peroxidase to the virulence of *Streptococcus pyogenes.* Infect Immun. 2004;72(1):408-13. Epub 2003/12/23. PubMed PMID: 14688122; PubMed Central PMCID: PMC344014.

83. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. Nat Methods. 2012;9(7):671-5. PubMed PMID: 22930834.

84. Sheen TR, Jimenez A, Wang NY, Banerjee A, van Sorge NM, Doran KS. Serine-rich repeat proteins and pili promote *Streptococcus agalactiae* colonization of the vaginal tract. J Bacteriol. 2011;193(24):6834-42. Epub 2011/10/11. doi: 10.1128/JB.00094-11. PubMed PMID: 21984789; PubMed Central PMCID: PMC3232834.

85. Fichorova RN, Rheinwald JG, Anderson DJ. Generation of papillomavirus-immortalized cell lines from normal human ectocervical, endocervical, and vaginal epithelium that maintain expression of tissue-specific differentiation proteins. Biol Reprod. 1997;57(4):847-55. Epub 1997/10/07. PubMed PMID: 9314589.

86. Okada N, Tatsuno I, Hanski E, Caparon M, Sasakawa C. *Streptococcus pyogenes* protein F promotes invasion of HeLa cells. Microbiology. 1998;144 (Pt 11):3079-86. Epub 1998/12/10. PubMed PMID: 9846743.

87. Decleva E, Menegazzi R, Busetto S, Patriarca P, Dri P. Common methodology is inadequate for studies on the microbicidal activity of neutrophils. J Leukoc Biol. 2006;79(1):87-94. Epub 2005/10/26. doi: 10.1189/jlb.0605338. PubMed PMID: 16244110.

CHAPTER III

Methyltransferase DnmA is Responsible for Genome-wide N6-methyladenosine Modifications at Non-palindromic Recognition Sites in *Bacillus subtilis*

Abstract

The genomes of organisms from all three domains of life harbor endogenous base modifications in the form of DNA methylation. In bacterial genomes, methylation occurs on adenosine and cytidine residues to include N6-methyladenine (m6A), 5-methylcytosine (m5C), and N4-methylcytosine (m4C). Bacterial DNA methylation has been well characterized in the context of restriction-modification (RM) systems, where methylation regulates DNA incision by the cognate restriction endonuclease. Relative to RM systems less is known about how m6A contributes to the epigenetic regulation of cellular functions in Gram-positive bacteria. Here, we characterize site-specific m6A modifications in the non-palindromic sequence GACG^{m}AG within the genomes of *Bacillus subtilis* strains. We demonstrate that the *yeeA* gene is a methyltransferase responsible for the presence of m6A modifications. We show that methylation from YeeA does not function to limit DNA uptake during natural transformation. Instead, we identify a subset of promoters that contain the methylation consensus sequence and show that loss of methylation within promoter regions causes a decrease in reporter expression. Further, we identify a transcriptional repressor that preferentially binds an unmethylated promoter used in the reporter assays. With these results we suggest that m6A modifications in *B. subtilis* function to promote gene expression.

Introduction

DNA methylation is pervasive across all three domains of life. In eukaryotes, 5-methylcytosine (m5C) modifications have been shown to function in development and the regulation of gene expression, with aberrant methylation implicated in human health, including cancer, autoimmune diseases, and metabolic disorders [for review, (1,2)]. m5C in promoter regions has been linked to the repression of downstream gene transcription, whereas gene body methylation has been positively correlated with gene expression [for review (3)]. A lesser-studied modification in the genomes of eukaryotes is N6-methyladenine (m6A). Recent studies have identified m6A in the genomes of *Chlamydomonas*, *Caenorhabditis elegans* and *Drosophila melanogaster* (4-6). In contrast to promoter m5C, m6A modifications appear to function in gene activation in the algae *Chlamydomonas* (4) and promoter m6A is also important in early *Drosophila* development (5). Further, m6A was positively correlated with gene expression in a diverse set of fungi (7). Thus, there is a growing recognition that m6A is critical for the regulation of gene expression in a broad range of eukaryotic organisms.

Bacterial genomes are known to harbor N4-methylcytosine (m4C) in addition to m5C and m6A [(8) and references therein]. All three modifications impart consequences to bacterial cells when methylation is lost (9). The most well understood example of DNA methylation in eubacteria is in the context of restriction-modification (RM) systems [for review (10,11)]. RM systems function as a bacterial host defense mechanism to prevent the invasion of foreign DNA, including phages and other mobile genetic elements (10,11). In organisms with RM systems, unmethylated foreign DNA is targeted for site-specific cleavage by a restriction endonuclease while the host chromosome is protected at the recognition sequence by site-specific DNA methylation (12). Methylation is achieved through the activity of DNA methyltransferases (MTases). MTases catalyze the transfer of a methyl group from the donor S-adenosylmethionine (SAM) to adenosine or cytidine residues in DNA (13,14). MTases that lack a cognate endonuclease and do not function in RM systems are referred to as 'orphan MTases' (15). In a limited set of Gram-negative bacteria, orphan MTases have been shown to function in critical processes including cell cycle control (16), origin sequestration (17,18), DNA mismatch repair (19-21), and the regulation of gene expression [for review

(22)]. DNA methylation from orphan and RM-based MTases has also been shown to establish epigenetic inheritance through phase variation primarily in Gram-negative pathogens (23-25). While much work has been done to characterize RM and orphan MTases from Gram-negative bacteria, much less is known about how m6A contributes to the regulation of the cell cycle or gene expression in Gram-positive bacteria (26).

Until recently, tools for unbiased detection and functional characterization of DNA methylation were limited. Available tools for detection, such as methylation-sensitive restriction endonuclease treatment and bisulfite sequencing, are limited to the sequence context and modification type that can be detected (27). The recent development of the Pacific Biosciences (PacBio) Single Molecule, Real-Time (SMRT) sequencing platform allows for detection of modifications without *a priori* knowledge of their existence (28). SMRT sequencing enables the analysis of real-time DNA polymerase kinetics for inference of DNA base modifications. Base modifications in the template strand result in changes in DNA polymerase kinetics compared to their unmodified counterparts, allowing for reliable, sequence-context specific detection of methylated bases during sequencing reactions (29). While differences in kinetic signatures for 5mC modified cytidine residues are modest, SMRT sequencing is adept for m6A and m4C detection (29).

Using the SMRT sequencing platform, a recent study of 230 diverse prokaryotes detected base modifications in 93% of the genomes surveyed (8). Of the genomes with detected modifications, 75% of the modifications were m6A, which is due in part to the robust signal of m6A modifications in SMRT sequencing relative to other modifications (29). Given the high percentage of prokaryotic genomes with m6A detected and the contribution of m6A to the regulation of eukaryotic gene expression, it seems unlikely that the prevalent m6A modifications in prokaryotes are used exclusively in the context of regulating DNA cleavage by RM systems. As mentioned above, in *Escherichia coli* and *Caulobacter crescentus* m6A from orphan MTases occurs in palindromic recognition sequences and has been shown to mediate protein-DNA interactions (9,30), regulating important cellular processes including gene expression (31-34). Deletion of Dam methyltransferase (*dam*), which is responsible for m6A at GATC sites in *E. coli*, has severe pleiotropic effects (35,36). In *C. crescentus* deletion of the CcrM

91

methyltransferase, which catalyzes the formation of m6A at GA(N)TC sites, is lethal when the CcrM-deficient strain is grown in rich media (16,37).

Much less is known about how m6A regulates cellular functions in Gram-positive bacteria. Recent work in *Streptococcus pyogenes* found that m6A from an active Type I RM system regulates virulence gene expression in a clinical isolate, suggesting that m6A could have important roles for regulating gene expression in Gram-positive systems (26).Therefore, the importance of m6A in *E. coli* and *C. crescentus* and the pervasive occurrence of m6A in prokaryotes (8) highlights the importance of understanding how m6A regulates cellular functions in the numerous and diverse set of bacterial genomes that contain the modification.

Here, we characterize m6A modifications in the Gram-positive bacterium *Bacillus subtilis* strains PY79 and NCIB 3610. Using SMRT sequencing, we show that m6A is present at non-palindromic GACG^m$\underline{A}$G sites throughout the *B. subtilis* chromosome. Further, we characterize the methyltransferase, referred to herein as DnmA, as responsible for detectable m6A modifications in the *B. subtilis* genome of both strains. We found that DnmA does not function as part of an active, canonical Type I or Type II RM system. Moreover, we show that the promoter regions for a subset of genes contain the consensus sequence and that loss of methylation in these *cis* regulatory elements results in a decrease in gene expression. Further, we show that the transcriptional repressor ScoC preferentially binds a promoter region that is unmethylated. Together, our results show that m6A can function as an epigenetic signal in *B. subtilis*.

Results

Characterization of *B. subtilis* PY79 and NCIB 3610 methylomes. It was previously published that *B. subtilis* does not have m6A at the *E. coli* Dam MTase recognition site, GATC, and that ectopic expression of Dam in *B. subtilis* induced the DNA damage response (44,45). However, until recently it remained unknown if *B. subtilis* contains m6A in another sequence context because the detection of m6A without *a priori* knowledge of the sequence context would require a new experimental approach. PacBio SMRT sequencing was used to determine if DNA modifications were present in the genome of several *B. subtilis* strains with the results deposited on the publicly

available web resource REBASE maintained by New England Biolabs. This resource reports m6A occurring in various sequence motif contexts in 19 of 23 *B. subtilis* strains where SMRT sequencing was used. Among the *B. subtilis* strains analyzed, methylation at GACG^{m}AG sites was reported in four of the 23 strains (http://rebase.neb.com). Previously, our group performed PacBio sequencing on the widely used *B. subtilis* laboratory strain PY79 for whole-genome assembly (41). As part of our effort to study DNA methyltransferases, we used PacBio sequencing to characterize the PY79 methylome. We purified genomic DNA from the wild type (WT) *B. subtilis* strain PY79 and analyzed our results using the SMRT sequencing platform, allowing for genome-wide base modification detection in sequence-specific contexts (29).

SMRT sequencing of the *B. subtilis* PY79 chromosome showed that the second adenosine residue within the sequence context 5'-GACG^m<u>A</u>G showed high modification quality values (modQVs), which indicates a statistically significant difference in DNA polymerase kinetics from the expected background at particular loci (**Fig 3.7, Table 3.1**). The interpulse duration (IPD) ratios, which are a comparison of DNA polymerase kinetics at a base within a particular sequence context compared to an unmethylated *in silico* control, were far higher for the second adenosine residue in the GACG^m<u>A</u>G motif compared to any other modified motifs in the *B. subtilis* chromosome (**Table 3.1, Fig 3.7**). Thus, we identify m6A in the sequence context 5'-GACG^m<u>A</u>G in the chromosome of *B. subtilis* PY79, herein referred to as the m6A motif.

We found that 99.7% of m6A motifs (1215/1219) were called as methylated in the PacBio SMRT sequencing analysis at the 3'-adenosine during exponential growth in defined minimal medium. While our sequencing analysis identified other motifs in the *B. subtilis* PY79 chromosome, the average modQVs, IPD ratios, and the percentage of motifs called as modified were far lower compared to m6A identified within the GACG^m<u>A</u>G sequence (**Table 3.6** and **Fig 3.8**). It is likely that most of the other motifs called represent DNA secondary structures that affect DNA polymerase kinetics or sequencing noise instead of genuine nucleic acid modifications (**Table 3.6**). For completeness we chose to report all motifs called during analysis of the SMRT sequencing data (**Table 3.6**).

Of the 1,219 m6A motifs that occur in the *B. subtilis* PY79 genome, 1,118

(91.7%) occur in protein coding regions. Intergenic regions, which compose 11.2% of the genome, contain 7% (85 motifs in 76 regions) of the m6A motifs. With the exception of only a few sites, the majority of m6A sites had greater than 75% of sequencing reads called as methylated independent of genome position or occurrence on the plus or minus strand of the chromosome (**Fig 3.8** and **Table 3.6**).

B. subtilis PY79 is a commonly used laboratory strain, however selection in the lab has caused PY79 to lose many of the robust phenotypes associated with ancestral strains of *B. subtilis* (46). To determine whether m6A is present in the ancestral strain, we purified genomic DNA from *B. subtilis* strain NCIB 3610 (40) for SMRT sequencing and found m6A within the same GACG<u>A</u>G sequence context **(Fig 3.7B** and **Table 3.1)**. In NCIB 3610 94.7% (1208/1275) of m6A sites were called as methylated in the PacBio SMRT sequencing analysis. The chromosome of the ancestral strain is considerably larger than PY79 and harbors an 84-kb plasmid, both of which account for the increased number of m6A motifs (40). The decrease in the percentage of motifs called as modified between PY79 and NCIB 3610 (99.7% → 94.7%) could be the result of biological variation, such as an increase in protein binding or other factors that may occlude methylation of recognition sites. The decrease in motifs called could also be due to technical variation in sequencing reactions. We note that we also detected many additional motifs in the ancestral strain that did not appear in the lab strain PY79, with each motif called listed in supplementary **Table 3.6**. Further, m6A at GACG<u>A</u>G sequences has also been reported for three *B. subtilis* strains other than PY79 and NCIB 3610 on REBASE.

In addition to m6A modifications, SMRT sequencing of the PY79 genome identified cytidine modifications in the sequence ^m<u>C</u>TCGARB (where R represents a purine and B either a cytidine or a guanosine). These results are described in the supplementary results section, where we show using methylation-sensitive restriction digest that m5C formation occurs in the *B. subtilis* PY79 genome through the BsuMI RM system **(Fig 3.9)** previously described for *B. subtilis* Marburg (47).

Distribution of m6A sites across the *B. subtilis* genome shows enrichment on the lagging strand of the left chromosomal arm. To begin to understand the function of

m6A in *B. subtilis*, we used the motif enrichment program DistAMo (42) to determine the location of m6A sites on the *B. subtilis* chromosome. This was done to determine if m6A sites are uniform or showed areas of enrichment and de-enrichment throughout the chromosome **(Fig 3.1)**. We present the analysis using sliding windows of 50 kb to 500 kb over the length of the chromosome by the rings from outside (large) to inside (small) scaling in 50 kb increments. Over (red) and under (blue) enrichment are colored by z-scores in the scale as shown. From the analysis we determine that the locations of m6A sites are certainly not uniform across the chromosome and instead show patterns of enrichment in particular regions. We find that several areas are largely devoid of m6A sites, including the terminus region and the origin of replication **(Fig 3.1)**. Analysis of enrichment shows that locations in the *B. subtilis* chromosome with high z-scores includes the right and left chromosomal arms with the largest enrichment on the lagging strand of the left chromosomal arm **(Fig 3.1C)**. With these results we suggest that m6A is unlikely to function in origin sequestration or DNA mismatch repair as described for Dam methylation in *E. coli* (17,18) due to our finding that the origin does not contain m6A sites and because m6A is non-palindromic and not uniform across the *B. subtilis* chromosome. To be certain, we empirically test if m6A contributes to replication timing, mutagenesis, or recombination in the supplementary results and show no effect (**Fig 3.10, 3.14** and **Table 3.7**).

Methyltransferase YeeA is necessary for m6A formation *in vivo*. DNA methylation is catalyzed by DNA methyltransferases (MTases) (48). To identify putative MTase(s) responsible for the observed m6A modification, we searched all protein coding sequences for the conserved DNA m6A MTase catalytic motif (D/N/S)PPY (48). This search yielded two uncharacterized MTases, coded for by the genes *yabB* and *yeeA* (*dnmA*) (41)*.* We created clean deletions of the Δ*yabB* and Δ*yeeA* (*dnmA*) coding regions as well as a Δ*yabB*Δ*yeeA* double deletion. Each of these strains was viable and none of the deletions conferred a growth defect on *B. subtilis* under the conditions used here **(Fig 3.3A,** described later in the results**)**.

　　To identify the MTase responsible for genomic m6A, DNA was harvested from each strain when cultures reached an OD_{600} of ~0.7 followed by SMRT sequencing.

Subsequent methylation analysis revealed that chromosomal DNA from Δ*yeeA* (*dnmA*) cells lost all detectable methylation at the m6A motif previously identified in WT cells in both PY79 and NCIB 3610 strain backgrounds **(Table 3.2, Fig 3.11,** and **Table 3.8)**. Expression of *yeeA* (*dnmA*) from an ectopic locus in the Δ*yeeA* (*dnmA*) background restored methylation at the m6A site **(Fig 3.11C** and **Table 3.2).** Computational analysis from sequencing data posted on REBASE also predicted YeeA (DnmA) as the MTase responsible for m6A detected in strains of *B. subtilis* with modifications at the m6A motif described here.

Genomic DNA from Δ*yabB* cells retained the methylation at m6A sites **(Fig 3.12, Table 3.9)** whereas detectable modifications at the m6A site were lost in the double deletion strain **(Fig 3.12B, Table 3.9)**. Interestingly, while methylation is maintained at the m6A site in the Δ*yabB* strain, we noticed additional motifs not present in the WT or Δ*yeeA* (*dnmA*) strains that were detected upon loss of *yabB* in the single or double deletion strains (**Table 3.9**). These additional motifs are likely to result from sequencing noise and/or DNA secondary structure given the low IPD ratios (**Table 3.9**). With these results we show that *yeeA* (*dnmA*) is necessary for genomic m6A formation in the sequence context GACG^mAG *in vivo* and we refer to YeeA herein as <u>DNA</u> <u>m</u>ethyltransferase <u>A</u> (DnmA), with the formal name of M.BsuPY79I and M.Bsu3610I for strains PY79 and NCIB 3610, respectively. For simplicity, we will collectively refer to M.BsuPY79I and M.Bsu3610I as DnmA in the work presented below.

DnmA is sufficient for methylation of m6A sites in double stranded (ds)DNA *in vitro*. DNA MTases typically use SAM to catalyze the transfer of a methyl group to a DNA base (9). DnmA (M.BsuPY79I), YabB, and a DnmA catalytically inactive variant (Y465A) were purified **(Fig 3.2A)**. In addition to the predicted ~120-kDa band corresponding to the DnmA monomer, a high molecular weight species was observed in the DnmA purifications. The slower migrating protein was analyzed by mass spectrometry identifying it as multimer of DnmA. We speculate that the DnmA multimer is caused by irreversible disulfide bonding or another crosslink that forms between two purified DnmA monomers during isolation **(Table 3.10)**.

A time course methylation experiment was performed to determine if DnmA is sufficient to catalyze methylation of the m6A motif in DNA **(Fig 3.2B)**. The purified proteins were incubated with tritiated SAM and an oligonucleotide sequence from the *B. subtilis addA* locus containing the m6A (target) motif. Incorporation of the labeled methyl group over time indicates that DnmA is indeed sufficient for methylation at m6A motifs in dsDNA **(Fig 3.2B)**. With the results from the time course methylation experiment we suggest that purified DnmA does not have significant activity on single-strand (ss)DNA. As a control we show that the Y465A catalytically inactive variant was unable to methylate the substrate indicating that the MTase activity we detect is specific to DnmA.

With the *in vitro* methylation assay established, we tested the activity of DnmA and YabB on DNA containing the target sequence and whole cell RNA extracted from a Δ*dnmA*Δ*yabB* double mutant strain by assaying for incorporation of methylation from tritiated SAM. As expected, DnmA showed activity on the dsDNA substrate with the target sequence, but had minimal activity when whole cell RNA was used as a substrate **(Fig 3.2C)**. In support of the *in vivo* results, we show that purified YabB had very little activity on a DNA substrate, whereas YabB did show incorporation when whole cell RNA was used as a substrate. With these results we suggest that YabB may function as an RNA methyltransferase **(Fig 3.2C)**. To test if the m6A motif was necessary for DnmA methylation *in vitro*, the 3'-adenosine residue was substituted with thymidine (non-target sequence) and incubated with DnmA and tritiated SAM. As shown in **Fig 3.2D**, there was no appreciable incorporation of the methyl group by DnmA to the non-target sequence, demonstrating that methylation is specific for the target sequence (m6A motif). We also tested DnmA for methylation of dsRNA, ssDNA, and ssRNA bearing the target sequence. The results show little to no methylation for any of these substrates with the exception of ssDNA, which yielded only weak methylation activity relative to dsDNA **(Fig 3.2D)**. Together, these results provide strong evidence that DnmA is specific for dsDNA containing the m6A motif.

To determine if the lack of methylation at the non-target sequence was caused by an inability of DnmA to bind DNA, an electrophoretic mobility shift assay (EMSA) was performed on 5' end-labeled target (GACGAG), non-target (GACGTG), and a degenerate sequence where the entire target sequence had been removed. Incubation

of DnmA with the target, non-target, and degenerate sequences each resulted in a shift, indicating that the methylation specificity is not due to a loss of DNA binding at other sequences **(Fig 3.13)**. Additionally, the Y465A catalytically inactive variant still bound the target sequence, suggesting that this variant is only dysfunctional for methyltransferase activity **(Fig 3.13)**. We conclude that DnmA is necessary and sufficient to methylate dsDNA that carries the GACG<u>A</u>G sequence *in vivo* and *in vitro* and that Y465 is an important residue for activity.

DnmA does not function as part of an active RM system. We next asked if DnmA functions as part of an active RM system. DnmA shares 38% identity and 57% similarity with the MmeI enzyme, which is a bifunctional protein with a methyltransferase domain and a PD-ExK endonuclease motif in the amino terminal domain. MmeI belongs to a subgroup of Type II RM systems that use DNA hemi-methylation for host chromosome protection (49). DnmA was included in a set of MmeI homologs that lack the endonuclease motif in the amino terminal domain but are flanked by conserved genes similar to *yeeB* and *yeeC*, which are immediately downstream of *dnmA* (49). It was found that under the conditions tested for other MmeI homologs DnmA lacked endonuclease activity, however it is important to note that the downstream *yeeB* and *yeeC* gene products are annotated as a putative helicase and an endonuclease, respectively (49). Deletion of *dnmA* does not result in a growth defect **(Fig 3.3A)**, which would suggest that *yeeB* lacks endonuclease activity associated with typical Type II RM systems, where endonuclease activity is achieved independent of the MTase.

It has been suggested that DnmA, along with YeeB and YeeC, comprise a Type I-like RM system, where restriction endonuclease activity requires the MTase subunit and DNA cleavage would not occur efficiently in the absence of DnmA (49). To test this possibility, we performed a transformation efficiency assay in WT and Δ*dnmA* cells with the plasmid pHP13, which is a 4.7 kb plasmid containing three m6A sites as the donor DNA **(Fig 3.3B)**. Plasmid purified from *E. coli* cells was used to transform competency deficient (Δ*comK*), hyper-competent (Δ*rok*), WT and Δ*dnmA* strains followed by selection for transformants conferring resistance to chloramphenicol. We found that compared to Δ*comK* and Δ*rok* strains, with transformation efficiencies of $< 1 \times 10^{-8}$ and

177 x10^{-5} (SE 13.2 x10^{-5}), respectively, the transformation efficiencies of WT [7.33 x10^{-5} (SE 3.30 x10^{-6})] and Δ*dnmA* [9.44 x10^{-5} (SE 1.25 x10^{-5})] were nearly indistinguishable (**Fig 3.3C**). We show that DnmA, YeeB, and YeeC do not function to restrict DNA update during natural genetic competence. Based on the transformation results and the conservation of these three genes clustering together, we suggest that DnmA, YeeB, and YeeC could be part of an inactive or inefficient Type I-like RM system or perhaps a noncanonical RM system. We also cannot exclude the possibility that restriction activity could be measured under some other circumstance, such as phage predation.

Proximity of m6A sites to -35 boxes of housekeeping sigma factor SigA regulates promoter activity. Due to the enrichment of m6A within particular genomic locations (**Fig 3.1**), we considered a role for m6A in regulating gene expression. Several prior studies have shown that DNA methylation from RM systems can also regulate gene expression (23,25,26). Accordingly, DNA MTase targets that occur within promoter or operator regions have the potential to influence transcription (50). Thus, we hypothesized that DnmA-dependent methylation might exhibit a similar function in *B. subtilis*.

To identify genes that might be affected by DnmA-dependent methylation, we used the list of transcribed regions 5' of *B. subtilis* 168 open reading frames (ORFs) reported previously (51) to prioritize the subset of promoters in *B. subtilis* with m6A sites located on the left chromosomal arm where we observed m6A enrichment. The promoters chosen for analysis included those of non-coding and anti-sense RNAs as well as promoters embedded inside transcriptional units, and we excluded promoters where the target site occurs downstream of the transcriptional start site **(Table 3.11)**. *B. subtilis* PY79 contains 32 transcribed regions 5' of ORFs with the m6A motif in the vicinity of known or predicted sigma factor binding sites **(Table 3.11)**. To examine if m6A in promoter regions influences gene expression in *B. subtilis*, we constructed a series of transcriptional fusions where a *gfp* allele was introduced downstream of the respective m6A motif-containing promoter **(Fig 3.4A)**. All transcriptional fusions were introduced at the ectopic *amyE* locus to separate the promoter from other potential *cis*-acting regulatory elements or chromosome structure contexts that could affect

99

expression **(Fig 3.4B)**. Promoter activity was monitored in WT and Δ*dnmA* strains using fluorescence as a reporter in single cells during mid-exponential growth by flow cytometry (please see Materials and Methods).

We found that loss of m6A in a subset of *B. subtilis* promoters, specifically those that contain an m6A motif in or slightly downstream of the -35 region of the SigA-binding box (*PscpA, Phbs, PrnhC, PyumC, PzapA*), consistently resulted in decreased activity from the unmethylated promoter relative to the methylated counterpart **(Fig 3.4C** and **D)**. The m6A sites in the promoter region for *PscpA, Phbs, PrnhC, PyumC, PzapA* in PY79 are identical to the promoter regions in *B. subtilis* strain NCIB 3610.

We did not observe this trend for the promoter fusions that contained m6A sites away from the -35 box. For example, the activation level of the SigB-inducible *rsbV-rsbW-sigB-rsbX* promoter (PrsbV), with an m6A site directly upstream of the -10 box, was not influenced by the presence of methylation during normal growth or even after stressing the cells with 4% ethanol for 1-hour as described (52) **(Fig 3.4C** and **D)**. Similarly, we did not observe differences in *gfp* expression with the *PcomEA*, *PwprA*, or *PyloA* fusions in the Δ*dnmA* background relative to WT.

The m6A motif was present just upstream and overlapping the -35 region of the SigA binding box for *PzapA* (transcription unit: *zapA-yshB-polX-mutSB-yshE*) and *PyumC*, respectively, and both reporters showed a decrease in activity in Δ*dnmA* cells relative to WT (**Fig 3.4C** and **D**). ZapA is involved in FtsZ ring assembly and YumC is an essential ferrodoxin/flavodoxin reductase (53,54). The m6A site for the remaining three promoter fusions that showed decreased expression upon loss of m6A, *PscpA* (transcription unit: *scpA-scpB-ypuI*), *Phbs* (transcription unit: *S861-hbs*), and *PrnhC*, was located just downstream of the -35 region of the SigA binding box. Interestingly, the gene products for two of the differentially expressed promoter regions, *scpA* and *hbs*, have important roles in chromosome segregation, chromosome structure, and organization (55-60). The changes in *Phbs* activity were mild, which is likely due to the fact that *Phbs* contains two SigA-binding boxes, of which the m6A-positive box is the least dominant of the two promoters (61). Another promoter fusion that exhibited a DnmA-dependent increase in expression was P*rnhC*, which codes for RNase HIII, an

100

enzyme important for cleavage of RNA-DNA hybrids (62,63). One type of RNA-DNA hybrid is an R-loop, which could affect local chromosome structure and transcription (64). Together, decreased expression from *PscpA*, *Phbs*, and *PrnhC* could have impacts on global chromosome structure, altering the expression of other genes.

To further investigate how m6A methylation affects transcription, the m6A site within the P*scpA*-GFP promoter was mutated to GACG**C**G, ensuring loss of methylation at this site in both the WT and Δ*dnmA* backgrounds. The GACG**C**G containing promoter adopted the same activity as observed in the Δ*dnmA* strain, indicating that m6A at the fifth position of the motif stimulates gene expression (**Fig 3.5A** and **B**). Interestingly, an A→T at the fifth position of the m6A site (GACG**T**G) made P*scpA*-GFP behave as if it were m6A (GACG^m**A**G) in both WT and Δ*dnmA* backgrounds **(Fig 3.5A and B middle panel).** The reason for how thymidine in the fifth position of the motif stimulates gene expression to the same extent as m6A is unclear. To further test how integrity of the motif modulates P*scpA* activity, the motif was subsequently changed to GACGA**C** so that the fifth position was unchanged but the DnmA recognition site was lost. The promoter adopted the same activity as quantified in the Δ*dnmA* strain, indicating that m6A or T at the fifth position of the motif stimulates gene expression for the *scpA* promoter **(Fig 3.5A** and **B)**. With these data we suggest that m6A is capable of regulating gene expression when located near the -35 binding site for SigA with methylation promoting gene expression from a subset of promoters in *B. subtilis.*

Transcriptional repressor ScoC binds GACGAG sites. The mechanism for m6A-dependent promotion of gene expression could be explained by an increase in SigA binding at methylated promoter regions or a less direct mechanism, such as competition for SigA binding with a methylation-sensitive transcriptional regulator. To determine if proteins in *B. subtilis* differentially associate with unmethylated DNA, we performed a pull-down in cell extracts using two different oligos. We amplified biotinylated oligos corresponding to the *PscpA* promoter region containing the GACGAG site. We could not obtain complete methylation of the substrate *in vitro* using purified DnmA. Therefore, we amplified the region and introduced a mutation in the m6A motif to GACG**T**G, which behaved like the WT methylated promoter in our reporter assay using the same

promoter region **(Fig 3.5A-B, middle panel)**. We isolated protein lysates from exponentially growing *B. subtilis* cells, incubated the lysates with our biotinylated oligos, performed a streptavidin pull-down, and visualized the proteins from each pull-down experiment via SDS-PAGE. We noted differences in the protein bands for the GACGAG relative to GACGTG oligo in the 20 and 40 kDa molecular weight range. These regions were excised from the gel and the proteins identified using mass spectrometry. Of the top four most abundant proteins across the samples, the transcriptional regulator of the transition state, ScoC (65,66), was the only protein that did not appear in both pull-down experiments **(Fig 3.6A)**. We found that ScoC was only present in the pull-down with the oligo that contained the GACGAG site, representing the unmethylated promoter state. No peptides corresponding to ScoC were identified in the pull-down of the GACG**T**G control site **(Fig 3.6A)**.

To directly test if ScoC binding is affected by the A→T mutation, we purified ScoC **(Fig 3.6B)** and performed electrophoretic mobility shift assays (EMSAs). We used labeled oligos representing the *PscpA* promoter that only differed in the GACGAG and GACGTG sites, which overlap the -35 box but occur just outside of the ScoC consensus binding site **(Fig 3.6C)**. The intensity of the shifted band was quantified and normalized to a no protein control for three independent experiments across a range of protein concentrations and the percent band shifted was compared at 250 nM and 500 nM ScoC. Consistent with the results from our pull-down experiment, we observed a 33.4% (S.E. ±2.6) and 14.7% (S.E. ±1.1) percent band shift at 250 nM ScoC for the GACGAG and GACGTG oligos, respectively **(Fig 3.6D-E)**. We also observed percent band shifts of 70.6% (S.E. ±9.0) and 45.7% (S.E. ±5.1) at 500 nM ScoC for the GACGAG and GACGTG oligos, respectively **(Fig 3.6D-E)**. The increased binding of ScoC to the oligo with the GACGAG site compared to the oligo with the GACGTG site **(Fig 3.6E)** and the decrease in expression we observed from the GACGAG promoter region compared to the GACGTG or GACG^{m}AG promoter **(Fig 3.5)** supports the model that ScoC is a transcriptional repressor (65,66) and that ScoC shows preferential binding to an unmethylated promoter with the m6A site proximal to the ScoC binding site. With these results we suggest that ScoC binds to unmethylated GACGAG sites in promoter regions repressing transcription. When the GACGAG site overlaps or is adjacent to the ScoC

binding site we suggest that methylation or A→T mutation at the fifth position could weaken ScoC binding leading to an increase in gene transcription.

Discussion

We report that DnmA (M.BsuPY79I or M.Bsu3610I) is responsible for endogenous m6A modifications that promote gene expression in *B. subtilis* strain PY79. We have shown that m6A in *B. subtilis* occurs at non-palindromic GACG^{m}AG sites in the chromosome with enrichment on the left chromosomal arm. In *B. subtilis* PY79 there are only 1,219 chromosomal m6A sites in contrast to the ~20,000 and ~4,500 palindromic m6A sites in *E. coli* and *C. crescentus,* respectively (67,68). While non-palindromic sites have been described (8) and have been shown to affect gene expression (25), the palindromic nature of m6A sites in *E. coli* and *C. crescentus* is necessary for function in DNA mismatch repair, origin sequestration, and cell cycle control (67). During these processes, protein binding or activity is dictated by full versus hemi-methylated states of m6A motifs, which determines the downstream regulatory role (67,69). Here, we have shown that loss of m6A at the non-palindromic GACG^{m}AG sites in *B. subtilis* also affects the regulation of gene expression, with loss of methylation resulting in decreased expression of genes, including *scpA* and *hbs*, which code for proteins important for chromosome structure, organization, and maintenance (55-60) (**Fig 3.4C** and **D**). Our data indicate that the presence of m6A promotes the expression of a subset of genes in PY79 that could have important downstream effects on gene expression and chromosome structure.

One mechanism by which m6A regulates gene expression is through dictating transcription factor binding to promoter regions. In prototypical *E. coli* the methylation state of recognition sites for Dam methyltransferase in promoter regions has been shown to affect expression of a subset of genes, including virulence factors (67,69). One such example is the *agn43* promoter, where methylation at the promoter blocks binding of the redox sensitive repressor OxyR, thereby stimulating production of Agn43, which is important for non-fimbrical adhesion (70). Also, uropathogenic *E. coli* use phase variation to evade the host immune system by altering the expression of the pyelonephritis-associated pilus (pap) in a Dam methylation dependent manner (24). In

the Gram-negative pathogen *Neisseria meningitidis* non-palindromic m6A sites from an active Type III RM system also function in phase variation (25). The Gram-negative bacterium *C. crescentus* has a transcriptional activator, GcrA, which associates with RNA polymerase-σ^{70} and recognizes a subset of promoter regions that are methylated at palindromic recognition sites by the CcrM MTase (71).

Here we have demonstrated that m6A regulated promoters in *B. subtilis* PY79 contain the methylation site at or slightly downstream of the -35 region of the housekeeping SigA binding site (72). We have shown that, in the absence of modification at the m6A site, we observe increased binding of the transcriptional repressor ScoC in the promoter region for the gene *scpA* **(Fig 3.6A-E)**. The increased binding of the transcriptional repressor ScoC at the *scpA* promoter containing a GACGAG site relative to the GACGTG site supports our reporter results, showing that the GACGTG site phenocopied the higher expression levels in a wild type strain relative to the Δ*dnmA* strain **(Fig 3.5A-B)**. We speculate that increased binding of the ScoC repressor to unmethylated GACGAG sites is responsible for the decreased gene expression we observe from the *scpA* promoter, representing one mechanism by which m6A could regulate gene expression in *B. subtilis* PY79.

While m6A-mediated binding of ScoC represents one mechanism by which m6A regulates gene expression, we find it likely that many other mechanisms exist. The methylation-responsive promoters identified in the current study do not share an obvious ScoC consensus binding sequence. Future work will be necessary to determine the additional regulatory mechanism(s) that result in increased gene expression at methylated promoter regions in *B. subtilis* PY79 and 3610.

 Each of the promoter fusions tested was ectopically expressed at the *amyE* locus, which allowed us to assay for the effect of promoter methylation status independent of the effects of chromosomal location and local chromosome architecture. This experimental design allows for identification of promoter region activities that were affected by loss of methylation at the m6A site but did not account for other factors. Interestingly, as shown **(Fig 3.4B)**, the genes for many of the downregulated promoter fusions occur toward the terminus (*hbs*, *scpA*, *rnhC*, and *zapA*) and on the left arm of the chromosome, whereas the *amyE* locus is origin proximal and occurs on the right

arm of the chromosome. Thus, we are able to conclude that methylation at the m6A site in *B. subtilis* PY79 promotes gene expression for a subset of genes but we cannot rule out other factors that control gene expression at the endogenous loci or indirect regulatory functions of m6A elsewhere in the chromosome.

In addition to its direct regulatory function at select promoter regions, m6A may have indirect effects on gene expression. It has been shown that m6A can increase the curvature of the DNA that may, in turn, influence protein binding and chromosome architecture (73,74). Alternatively, m6A might directly influence the expression of DNA binding proteins that contribute to chromosome architecture. Consistent with this hypothesis, we observe slight but significant downregulation of the *hbs* gene, which codes for the essential and highly abundant histone-like protein HBsu **(Fig 3.4C)**. A potential decrease in HBsu levels concomitant with the preference of HBsu for highly curved regions of DNA creates the possibility for an m6A-dependent mechanism for changes in overall DNA topology and chromosome architecture. Thus, loss of m6A may affect protein occupancy throughout the chromosome to influence chromosome architecture in such a way that results in more changes to gene expression. It is important to note that both direct and indirect models of m6A-dependent changes are possible and that they are not mutually exclusive

Genomic m6A from orphan and active RM system MTases has been shown to function in the regulation of gene expression [e.g. (23-26)]. Here we demonstrate that loss of MTase DnmA does not affect the natural transformation efficiency of foreign methylated DNA from a plasmid with multiple recognition sites in competent cells. Therefore, we suggest that DnmA is an MTase from an inefficient or inactive RM system. We have also discovered that DnmA-dependent m6A in the promoter regions of a subset of genes promotes gene expression in *B. subtilis* PY79 and we show that transcriptional repressor ScoC binds unmethylated DNA. In addition to influencing ScoC binding, we find it interesting that m6A promotes expression of several genes involved in chromosome structure and maintenance, which could in turn have effects on the expression of other genes. In total, we have shown that DNA methylation from DnmA has an effect on gene expression, prompting further investigation of RM systems and their possible regulatory contribution outside of DNA restriction.

Materials and Methods

General Bacteriology: The antibiotic concentrations used in this study are as follows: 5 µg/mL chloramphenicol, 0.5 µg/mL erythromycin, 100 µg/mL spectinomycin. Unless otherwise indicated, strains were grown in either LB (10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl) or defined $S7_{50}$ minimal media supplemented with 1% glucose (1x $S7_{50}$ salts diluted from 10x $S7_{50}$ salts (104.7 g/L MOPS, 13.2 g/L ammonium sulfate, 6.8 g/L monobasic potassium phosphate, adjusted to pH 7 with potassium hydroxide), 0.1% potassium glutamate, 1% glucose, 40 µg/mL phenylalanine, 40 µg/mL tryptophan, 2 mM $MgCl_2$, 0.7 mM $CaCl_2$, 50 µM $MnCl_2$, 1 µM $ZnCl_2$, 1 µg/mL thiamine-HCl, 20 µM HCl, and 5 µM $FeCl_3$) at 30°C with shaking at 200 rpm.

Strain construction: The strains, plasmids and oligos used in this study can be found in Supplementary Tables 3.3, 3.4, and 3.5. Individual strain and plasmid construction can also be found in the Supplemental Materials and Methods. Deletions were created by ordering *Bacillus subtilis* 168 strains from the Bacillus Genetic Stock Center (http://www.bgsc.org/) where the respective genes were replaced with a *loxP* flanked erythromycin (erm) resistance cassette (BKE strains). Genomic DNA from the BKE strains was purified and used to transform *B. subtilis* strain PY79, and the erm resistance cassette was subsequently removed with Cre recombinase (38). Overexpression strains and all promoter GFP fusions were integrated in the PY79 *amyE* locus via double crossover (39). Three colonies containing the crossover were selected and colony purified on LB plates containing 100 µg/mL spectinomycin. Successful integration of the constructs was verified by PCR, Sanger sequencing, and screening for the ability to utilize starch.

Chromosomal DNA purification: Genomic DNA for Pacific Biosciences SMRT sequencing was purified as follows. Strains were struck out on LB and grown overnight at 30°C. 500 mL LB cultures were inoculated at OD_{600} 0.05 and grown at 37°C. During mid-exponential phase (OD_{600} 0.6-0.8) an equal volume of methanol was added to each culture and centrifuged at 4,000 rpm for 30 minutes. The supernatant was discarded and cells were resuspended in 12.5 mL of 10% sucrose Tris/HCl pH 8 buffer and

transferred to Oakridge tubes. Resuspensions were then treated with 310 µL lysozyme (40 mg/mL in 10% sucrose Tris/HCl pH 8 buffer) for 30 minutes at 37°C and mixed every 5 minutes. 1.25 mL of 0.5 M EDTA was added to each tube and incubated on ice for five minutes followed by addition of 10 mL of freshly prepared lysis solution (0.1% Triton X-100, 62.5 mM EDTA, 50 mM Tris/HCl pH 8). Solutions were centrifuged at 15,000 rpm for 30 minutes and decanted into chilled graduated cylinders. To each lysate 0.95 g/mL of cesium chloride (CsCl) was added and dissolved followed by a 1/10 volume addition of 10 mg/mL ethidium bromide. Solutions were balanced and centrifuged at 44,000 (131,600 x g) rpm for 24 hours. Chromosomal DNA was extracted and subjected to a second round of CsCl purification as described above. Solutions were centrifuged at 44,000 rpm (131,600 x g) for 48 hours. Ethidium bromide was removed by extraction 4x with water-saturated butanol. The aqueous phase was transferred to an Oakridge tube and 1 volume of water and 2 volumes ethanol were then added. The solution was centrifuged at 15,000 rpm for 20 minutes and the supernatant was aspirated. The pellet was washed with 70% ethanol and resuspended in 1 mL TE buffer.

In all other experiments, frozen strains were struck out and grown at 30°C. The plates were washed in $S7_{50}$ minimal media and 25 mL cultures were inoculated at an OD_{600} 0.05 and grown at 37°C with shaking to mid-exponential growth phase (OD_{600} 0.6-0.8). Genomic DNA was purified via phenol chloroform extraction method.

PacBio SMRT sequencing and methylation analysis: Chromosomal DNA was prepared for sequencing as described above. Library preparation and subsequent sequencing was performed as previously described (40,41). Modification and motif analyses were performed using RS_Modification_and_Motif_Analysis.1 version 2.3.0 with the appropriate *B. subtilis* reference genomes. The initial parameters used for modification analysis were performed using 0.75 minimum high quality reads, 50 bps minimum length, and a minimum ModQV call of 30. We also increased minimum high quality reads to >0.85 and minimum length to >1000 bps in subsequent analysis. Modification graphs were generated using functions from BaseModFunctions.v2.1.R

available at: https://github.com/PacificBiosciences/Bioinformatics-Training/tree/master/basemods.

Motif Distribution Analysis: Motif distribution analysis was performed using the DistAMo web based server (42) available at http://computational.bio.uni-giessen.de/distamo searching the GACGAG motif for the *Bacillus subtilis* PY79 genome via accession number NC_022898.1.

Protein Purification (DnmA, DnmA (Y645A), and YabB): Recombinant proteins were purified from *E. coli* BL21$_{DE3}$ cells containing a pE-SUMO vector with the *B. subtilis* gene inserted (*dnmA, dnmA (Y465A)*, or *yabB*). Cultures were grown in 4 L of terrific broth (2.4% yeast extract, 1.2% tryptone, 0.4% glycerol, 250 mM $(NH_4)_2SO_4$, 500 mM KH_2PO_4, 1x metals (1,000x metals: 2.5 mM $FeCl_3$, 1 mM $CaCl_2$, 0.5 mM $ZnCl_2$, 0.1 mM $CoCl_2$, 0.1 mM $CuCl_2$, 0.1 mM $NiCl_2$, 0.1 mM Na_2MoO_4, 0.1 mM Na_2SeO_3, 1 mM H_3BO_3), and 25 µg/mL kanamycin) at 37°C with orbital rotation for 2 hours until reaching an OD_{600} of ~0.7. Overexpression was induced by adding IPTG to 1 mM and the cultures were grown for 3 additional hours 37°C. Cells were then pelleted by centrifugation and frozen in liquid nitrogen to be stored at -80°C. Once thawed, the pellet was resuspended in lysis buffer (50 mM Tris-HCl pH 8, 300 mM NaCl, 10% sucrose, 10 mM imidazole, 1x protease inhibitors (Roche 11873580001)) and cells were sonicated on ice. Cell debris was pelleted by centrifugation. Supernatant was then poured through a 3 mL Ni^{2+}-NTA agarose gravity-flow column. The column was washed with wash buffer (20 mM Tris-HCl pH 8, 10% glycerol, 20 mM imidazole, 2 M NaCl) and eluted with elution buffer (50 mM Tris-HCl pH 8, 150 mM NaCl, 400 mM imidazole). SDS-PAGE was performed to confirm the presence of desired protein. The sample was then dialyzed into anion exchange start buffer (50 mM Tris-HCl pH 8, 25 mM NaCl, 5% glycerol, 1 mM β-mercaptoethanol) and the sample was applied to a Q column (GE: 17115301) using an elution gradient of 50-750 mM NaCl. SDS-PAGE was performed and fractions containing desired protein were pooled and incubated with ULP1 protease at 25°C for 30 minutes. The digestion product was applied to another 3 mL Ni^{2+}-NTA gravity-flow column, washed, and eluted using the same buffers as above. SDS-PAGE

was again performed to confirm the SUMO tag was removed and the protein was concentrated and buffer exchanged into protein storage buffer (50 mM Tris-HCl pH 8, 150 mM NaCl, 50% glycerol), aliquoted, flash frozen in liquid nitrogen, and stored at -80°C.

ScoC purification: Primers oTMN62 and 63 were used to amplify *scoC* from the *B. subtilis* chromosome and were subsequently combined with the pE-SUMO expression vector via Gibson assembly. Recombinant proteins were purified from *E. coli* BL21$_{DE3}$ cells grown in 2 L of LB with 25 µg/ml kanamycin at 37°C with orbital rotation until an OD$_{600}$ of 0.7 was reached. Overexpression was induced by adding 0.5 mM IPTG followed by culture growth for an additional three hours at 37°C with orbital rotation and cultures were subsequently pelleted via centrifugation and stored at -80°C. The pellet was re-suspended in lysis buffer and sonicated on ice as described for DnmA and YabB. Subsequent to centrifugation, the supernatant was applied to a 4 mL Ni^{2+}-NTA agarose gravity-flow column. The column was washed with wash buffer (50 mM Tris-HCl pH 8, 25 mM imidazole, 2 M NaCl, 5% glycerol) and eluted with elution buffer (50 mM Tris-HCl pH 8, 400 mM imidazole, 150 mM NaCl, 5% glycerol). Following elution, 1 mM DTT and SUMO ULP1 protease were added to the elution fraction and incubated for 2 hours at room temperature. The sample was then dialyzed into storage buffer (50 mM Tris-HCl pH 8, 150 mM NaCl, 5% glycerol) overnight at 4°C. The dialyzed sample was then applied to another 4 mL Ni^{2+}-NTA gravity-flow column to separate the recombinant protein from the SUMO tag. SDS-PAGE was performed to confirm the SUMO tag was removed. Glycerol was added to 25% and the protein was aliquoted and flash frozen for storage at -80°C

Methylation Assays: All methylation reactions were performed in a buffer containing 50 mM Tris-HCl pH 8, 50 mM NaCl, and 200 µM MgSO$_4$. The following substrates were annealed in the same buffer at 2.5 µM concentration by heating primers to 100 °C for 30 seconds and then cooling to room temperature on the bench top: dsDNA target (oTNM38, oTMN39); dsDNA non-target (oTMN40, oTMN41); and dsRNA (oJR270, oJR271). The H^3-SAM (Perkin Elmer: NET155H001MC) was used at a concentration of

1 µM in solution. The purified DnmA, YabB, or DnmA (Y465A) was added to a concentration of 1 µM and all substrates were used at 0.25 µM in solution. The proteins were added in excess to determine if there was any off target methylation activity at higher protein concentrations. The total reaction solution came to 10 µL. All reactions were incubated at 37°C for 150 minutes unless otherwise specified. Reactions were stopped using 450 µL of 10% TCA and placed on ice. The samples were filtrated using Glass microfiber filters (GE: 1822-025), washed with cold 70% ethanol, dried, and placed in a scintillation counter to measure mmol incorporation.

Growth Curves: Strains were plated on LB and grown overnight at 30°C. Plates were washed in LB and diluted to an OD_{600} of 0.05 in 10 mL of LB in side-armed flasks. Cultures were grown in shaking water baths at 37°C and optical density was measured using a Klett meter every half hour through late stationary phase. Growth curve experiments were done in triplicate and data was subsequently fit to a Gompertz growth (43) model $\{y = Aexp\{-exp\left[\frac{\mu_m \times e}{A}(\lambda - t) + 1\right]\}$ (where the parameters A, μ_m, and λ represent the time (t) when the growth rate equals zero (asymptote), the maximum growth rate, and the lag time, respectively), to obtain growth rate estimates (μ_m) for each strain.

Transformation efficiency assays: Strains were plated on LB and grown overnight at 30°C. Plates were washed with phosphate buffered saline (PBS) pH 7.4 and the cells were pelleted, the supernatant was aspirated, and a second PBS wash was completed before the cells were resuspended in PBS. The cells were used to inoculate a culture at an OD_{600} of 0.05 into 1 mL of 1x MC media (10x MC media: 615 mM K_2HPO_4, 380 mM KH_2PO_4, 1.11 M dextrose anhydrous, 30 mM sodium citrate dihydrate, 840 µM ferric ammonium citrate, 0.5 g casein hydrolysate, and 125 mM sodium aspartate monohydrate, to 50 mL with ddH_2O and filter sterilize) with 3 µL of 1M $MgSO_4$ and grown at 37°C with aeration for 4 hours. After 4 hours 3 µL of 1M $MgSO_4$ and 300 ng of pHP13 purified from *E. coli* MC1061 cells was added to 300 µL of cells and grown for an additional 1.5 hours at 37°C. 10x serial dilutions were performed into PBS and appropriate dilutions were plated onto LB plates for colony forming unit (CFU) counts

and chloramphenicol plates for transformation forming unit (TFU) counts. Transformation efficiencies were calculated as TFU/CFU and the average transformation efficiency for replicates performed over three separate days was plotted along with the corresponding standard errors.

Flow Cytometry: Cells were grown overnight at 30°C on LB plates containing 100 μg/mL spectinomycin. Exponentially growing colonies were washed from the plates using S7$_{50}$ medium, and washed two more times to remove residual LB agar before diluting the cells in pre-warmed S7$_{50}$ medium to an OD$_{600}$ of 0.05. Cells were grown to an OD$_{600}$ of 0.4 at 30°C after which fluorescence of 200,000 cells was measured using an Attune™ NxT Acoustic Focusing Cytometer (ThermoFisher Scientific) using the following settings: Flow rate, 25 μL/min; FSC voltage, 200; SSC voltage, 250; BL1 voltage, 250.

Streptavidin pull-down: 5' biotinylated primers were used to amplify the 233 bp region of the *scpA* promoter via PCR using genomic DNA from strains LVG087 and LVG102 as a template, which correspond to the GACGAG and GACGTG promoter, respectively. To obtain total cell lysate, 4 L of strain TMN85 (Δ*dnmA*) was grown in S7$_{50}$ medium at 37°C with shaking until the culture reached an OD$_{600}$ of 1.0. After the cells were harvested the pellets were washed with 1x PBS (pH 7.5) and then subsequently washed with Pull- Down Binding Buffer (PDBB; 50 mM Tris-HCl pH 7.5, 0.5 mM EDTA, 100 mM NaCl, 0.01% (v/v) Triton X-100, 25% (v/v) glycerol, and 1 mM DTT) and resuspended in ice-cold 20 mL PDBB supplemented with one tablet of cOmplete™, EDTA-free Protease Inhibitor Cocktail (Roche, Mannheim, Germany). The cell suspensions were sonicated on ice (10s on, 40s off, 70 Hz) until the solutions cleared. Cell debris was removed from the lysate by two subsequent washing steps and the protein content of the supernatant was estimated using a Bradford assay (~20 mg/mL protein). For each pull-down experiment, 100 μL of Dynabeads™ M-270 Streptavidin magnetic bead slurry (ThermoFisher Scientific) was washed three times with 500 μL Pull-Down Wash Buffer (PDWB; 10 mM Tris-HCl pH 7.5, 1 mM EDTA, and 1 M NaCl). The beads were re-suspended in 250 μL PDWB, mixed with 200 pmol biotinylated

probe DNA dissolved in 250 μL nuclease-free water, and incubated for 30 min at 25°C with gentle rotation. The DNA-coated beads were washed three times with PDBB before 100 mg protein and 100 μg salmon sperm DNA (Millipore Sigma) were mixed and added to the DNA-bound beads. After 2 hrs of incubation at room temperature with gentle rotation, the beads were separated and washed once with PDBB, once with PDBB plus 100 μg salmon sperm DNA, and again with PDBB. Bound proteins were eluted using Pull-Down Elution Buffer (PDEB; 50 mM Tris-HCl pH 7.5, 0.5 mM EDTA, 1 M NaCl, 0.01% (v/v) Triton X-100, 25% (v/v) glycerol, and 1 mM DTT). The eluted proteins were desalted and concentrated using TCA precipitation and separated on a 4-20% Mini-PROTEAN TGX precast protein gel (Bio-Rad, Hercules, USA). Bands in the 20 and 40 kDa size range were excised from the gel followed by protein identification using mass spectrometry through the University of Michigan Proteomics Resource Facility, project PRF-2019-L-SIMM-29.

ScoC EMSA: 5' IR dye end-labeled substrates oTN67/oTN68 and oTN70/oTN71, corresponding to the GACGAG and GACGTG oligos, respectively, were annealed at a concentration of 50 nM by heating at 95°C for 1 minute and then snap-cooled on ice. Care was taken to avoid subjecting the IR dye labeled oligos to light. Annealed oligos were mixed at a final concentration of 5 nM with indicated concentrations of purified ScoC in 1x EMSA reaction buffer (5x EMSA reaction buffer: 250 mM Tris-HCl pH 8, 5 mM EDTA, 150 mM KCl, 10 mM $MgCl_2$, 5 mM DTT, 1% Tween 20, 125 μg/mL sheared salmon sperm DNA) to a final volume of 10 μL. Reactions were incubated at 37°C for 15 minutes and subsequently loaded onto and resolved via 6% Native-PAGE, which was performed covered and on ice for 60 minutes at 100V. The samples were visualized with the LI-COR Odyssey imager. The intensity of the shifted band was normalized to the no protein control for each sample to calculate the percent band shifted. Three replicates were completed and quantified across separate days and the average and standard errors for percent band shifted was reported in 3.6.
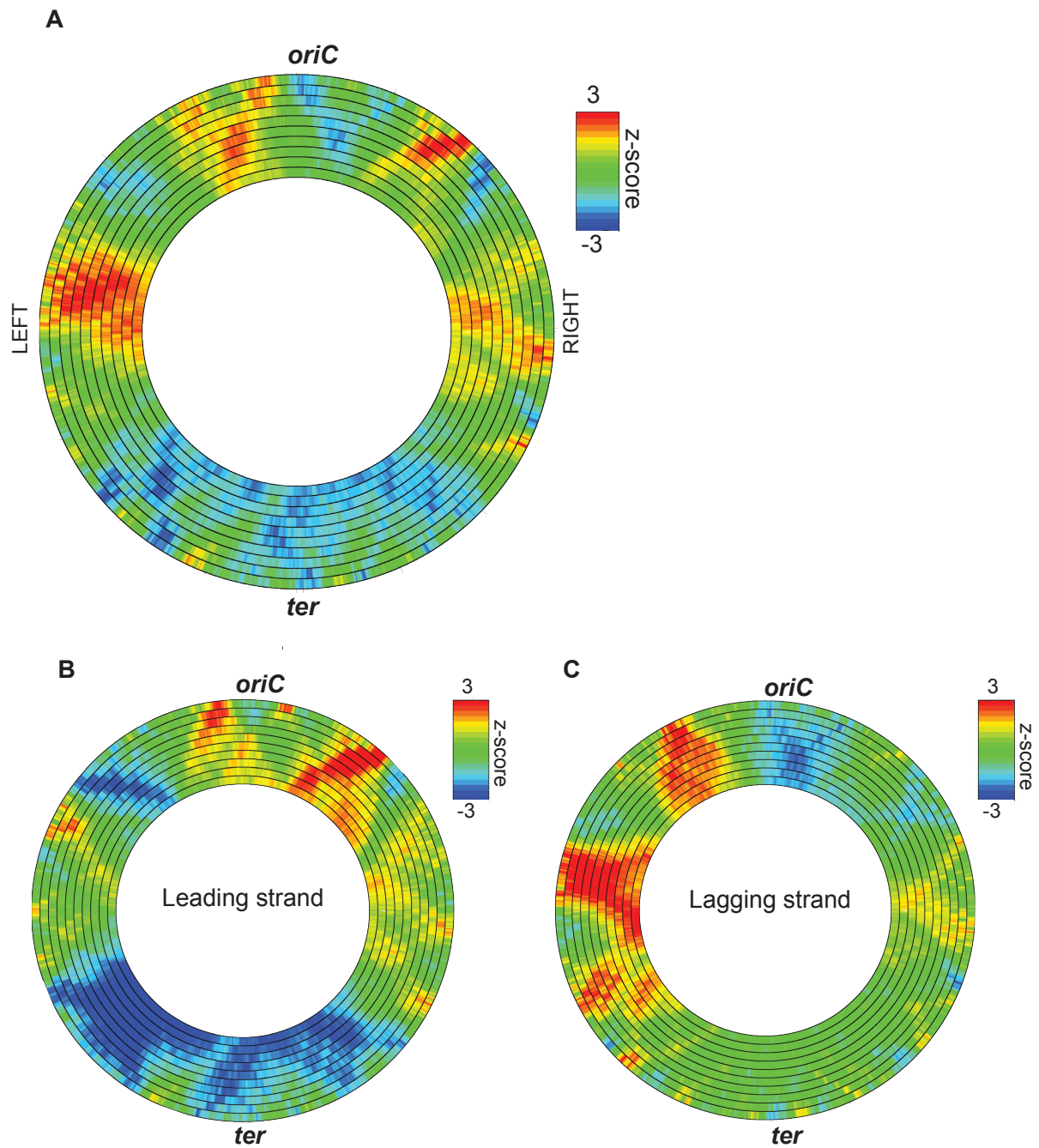
Fig. 3.1. Motif enrichment analysis for m6A sites in the *B. subtilis* PY79 chromosome. Motif enrichment analysis was performed using the DistAMo web based server tool (42). Sliding windows of 50 kb to 500 kb are represented by the rings from outside (large) to inside (small) rings scaling in 50 kb increment increases. Over (red) and under (blue) enrichment are represented by z-scores in the scale indicated. **(A)** m6A motif enrichment for all motifs with *ori* and *ter* regions indicated; **(B)** m6A motif enrichment on the leading strand; **(C)** m6A motif enrichment on the lagging strand.
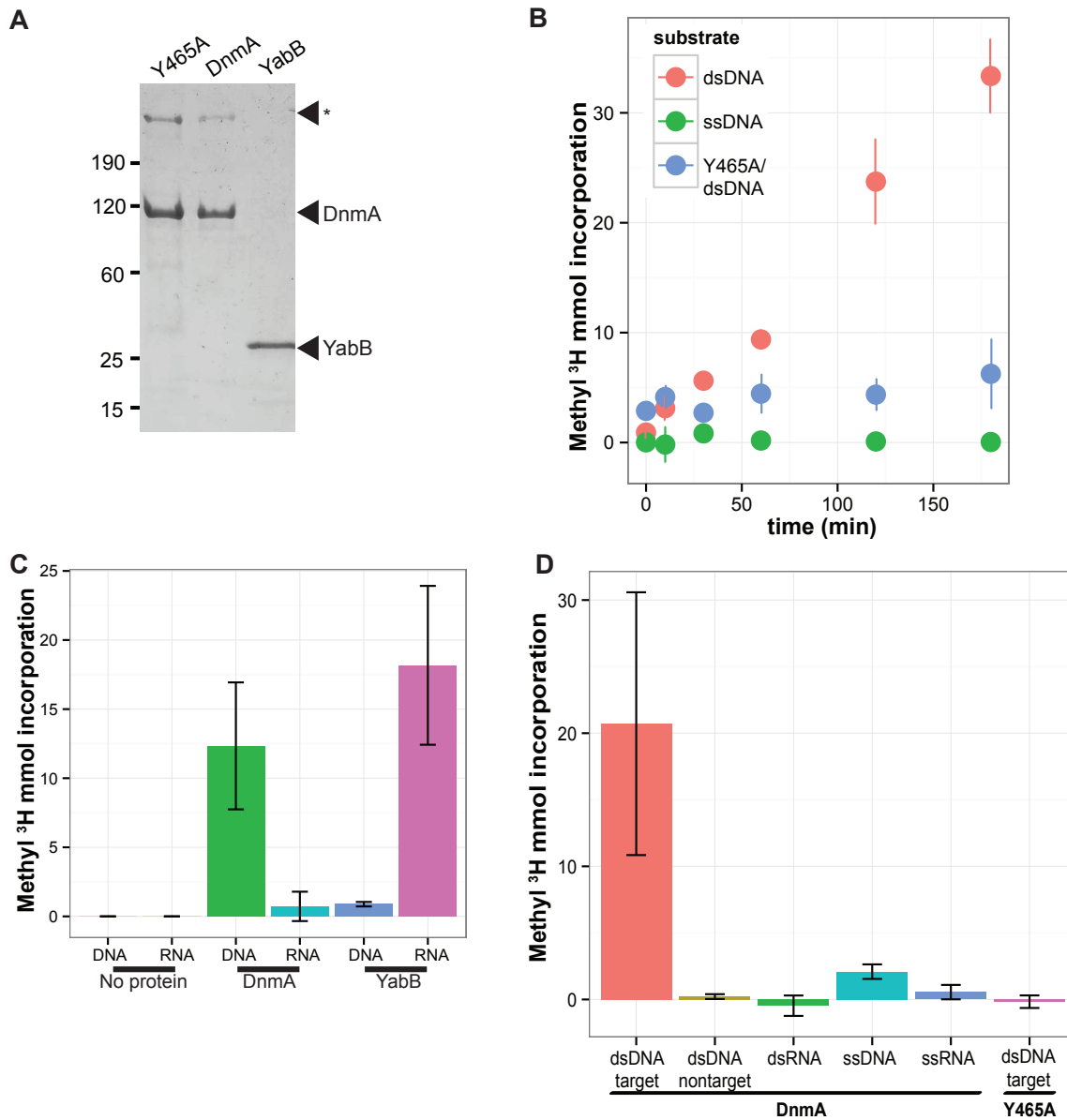
Fig. 3.2. DnmA is sufficient for methylation of dsDNA at 5'GACGAG sites.
(A) SDS-polyacrylamide gel of purified catalytically inactive DnmA variant Y465A, WT
DnmA (M.BsuPY79I), and YabB. (*) indicates DnmA multimer. **(B)** DnmA (M.BsuPY79I)
incorporation of tritiated SAM into dsDNA and ssDNA substrates carrying the GACGAG
sequence over time. Y465A (indicated in blue) is a DnmA (M.BsuPY79I) catalytically
inactive variant. **(C)** Incorporation of tritiated SAM into DNA and RNA substrates by
uncharacterized MTases DnmA (M.BsuPY79I) and YabB. **(D)** DnmA (M.BsuPY79I)
incorporation of tritiated SAM onto indicated substrates. The DnmA catalytically inactive
variant is indicated.

Relevant genotype	Growth rate	95% CI
Wild type	1.52	[1.43, 1.61]
ΔdnmA	1.51	[1.43, 1.59]
ΔyabB	1.51	[1.43-1.58]
ΔdnmAΔyabB	1.46	[1.35-1.57]
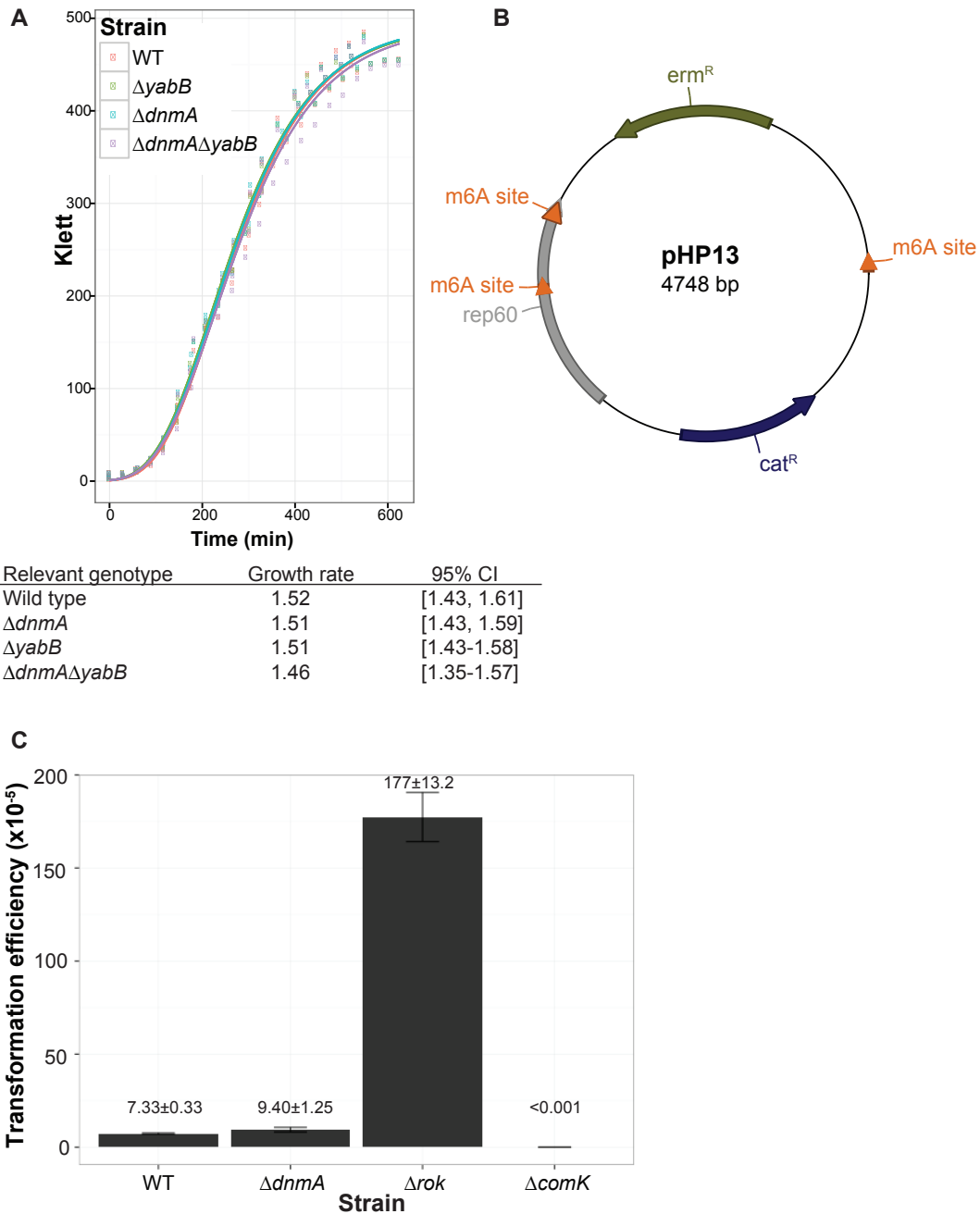
Fig. 3.3. Loss of DnmA does not affect growth rate or transformation efficiency of foreign methylated DNA. (A) Growth curves for WT, Δ*dnmA,* Δ*yabB,* and Δ*dnmA*Δ*yabB* were performed in triplicate and fit to a Gompertz growth model (43) to calculate growth rate. Growth rate and the corresponding 95% confidence interval for each strain are indicated. **(B)** Plasmid map of pHP13 with the location of each m6A site shown. The orange carrots indicate the relative position and strand orientation for each site. **(C)** Transformation efficiency assays were performed using pHP13 plasmid purified from *E. coli* as donor DNA in WT, Δ*dnmA,* Δ*rok,* and Δ*comK* recipient strains. The average transformation efficiency and standard error for each strain is indicated.
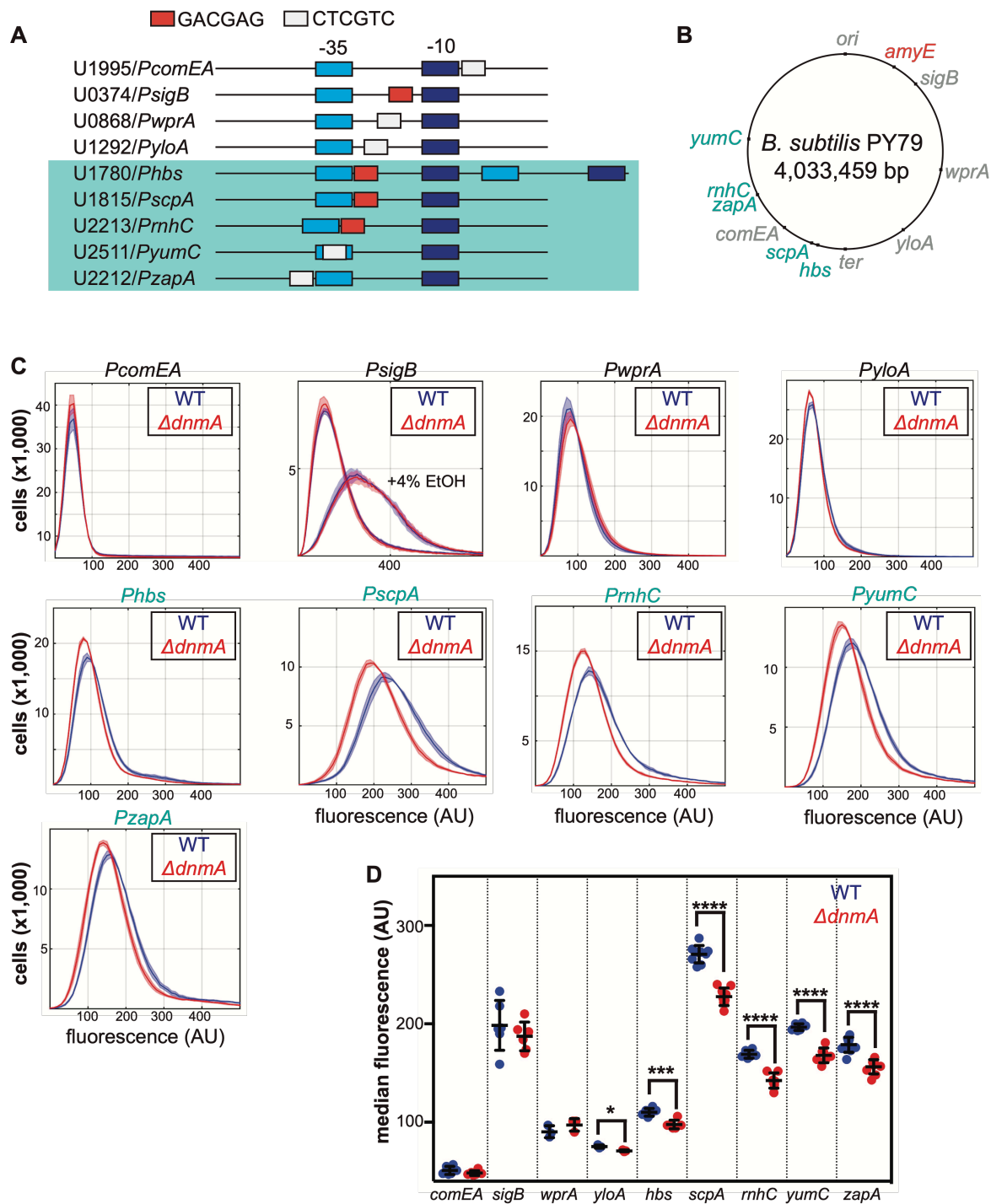
Fig. 3.4. Methylation of DnmA motifs in proximity of -35 boxes affects downstream gene expression. **(A)** Schematic overview of the promoter regions containing DnmA sites that were selected for analysis using transcriptional GFP fusions.

Indicated are the locations of the predicted sigma factor -35 and -10 boxes with respect to the DnmA motifs. U numbers correspond to the transcribed regions 5' of ORFs identified by Nicolas *et al* (51). **(B)** The location of the studied promoters on the PY79 chromosome map with respect to the *amyE* site used for integration and analysis of the promoter-GFP constructs. **(C)** Histograms depicting the GFP fluorescence in 200,000 WT (blue) or Δ*dnmA* (red) cells in three biological replicates that were grown in S7$_{50}$ medium to an OD$_{600}$ of 0.5 at 30°C and measured using flow cytometry. For U0374/P*sigB*, an additional experiment was performed in which the cells were treated with 4% EtOH an hour before analysis with flow cytometry. The standard deviations are represented as shaded areas. Promoter regions that appear methylation sensitive are shown in green. **(D)** Scatter dot plots, with indicated mean and standard deviation, depicting the median fluorescence of each strain taken from the histograms shown in (C) and appended with similar measurements taken on at least one different day. A standard T-test was performed to evaluate differential GFP expression between WT and Δ*dnmA* for each promoter. *p*-values: * = $p < 0.05$, *** = $p < 0.005$, **** = $p < 0.001$.
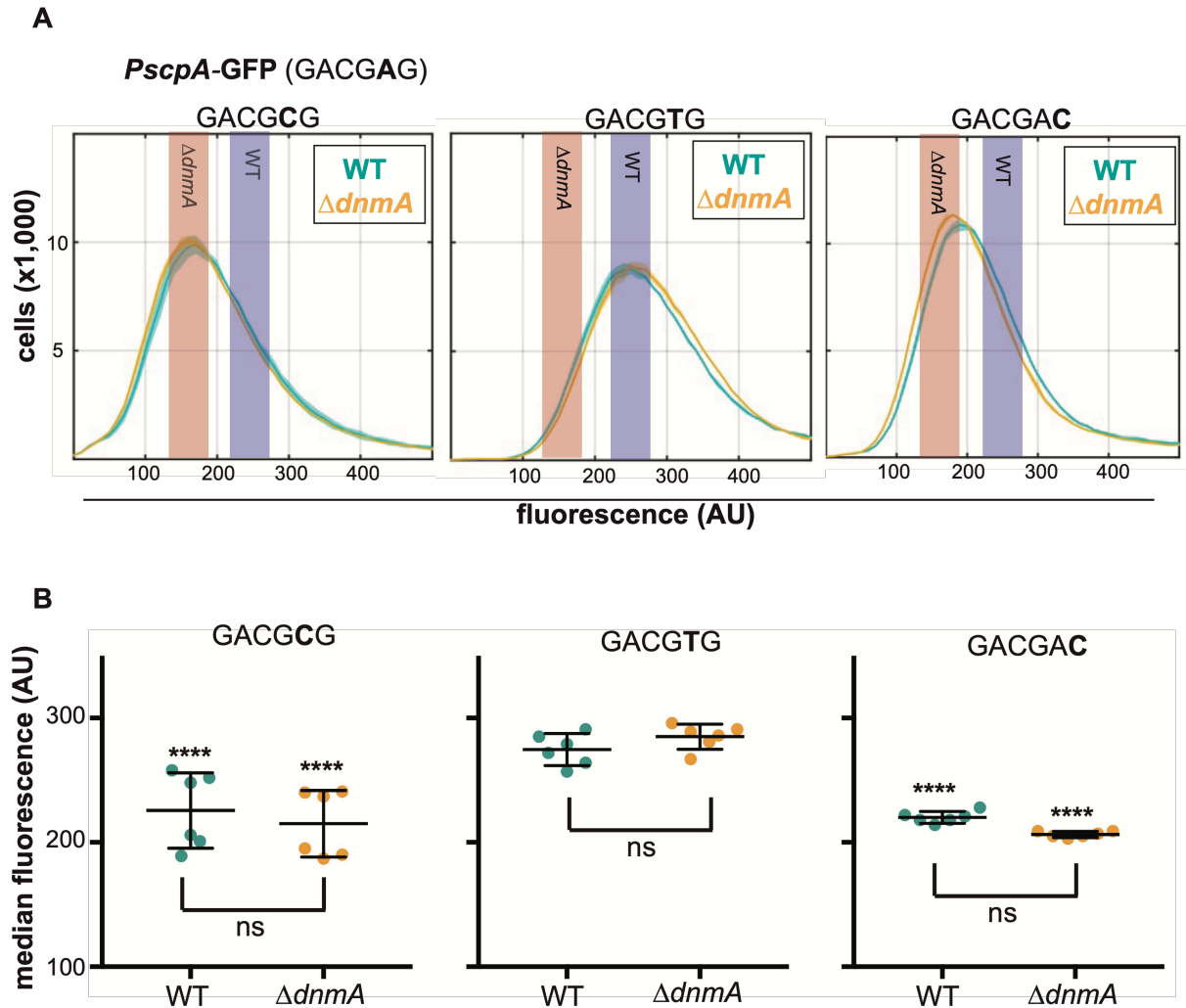
Fig. 3.5. Mutating the DnmA recognition motif is sufficient for differential gene expression in the *PscpA* promoter. (A) Analysis of the effect of mutating WT GACG**A**G to GACG**C**G (first graph), GACG**T**G (second graph), or GACGA**C** (third graph) on the activity of P*scpA*-GFP in WT (teal) and Δ*dnmA* (orange) cells. **(B)** Scatter dot plots, with indicated mean and standard deviation, of the median GFP fluorescence of each strain taken from the histograms shown in (A) and appended with measurements from a similar experiment taken on a separate day. The median values were tested against each other, including the median values from the strain expressing P*scpA*-GFP in WT cells, for differential expression using a one-way ANOVA post-hoc Tukey test. *p*-values: * = $p < 0.05$, *** = $p < 0.005$, **** = $p < 0.001$, ns = not significant.

A

Protein	Function	MW [kDa]	# PSMs GACGAG	# PSMs GACGTG
FabL	Enoyl-acyl carrier protein reductase	27.2	105	110
RpsE	Ribosomal protein	17.6	29	22
RpsG	Ribosomal protein	17.9	23	11
ScoC	Transcriptional regulator of transition state	23.7	16	0

B

C

PscpA:

D

E

Fig. 3.6. Transcription factor ScoC binds the *scpA* promoter with an unmodified GACGAG site. (A) Top protein hits identified in the pull-down of the biotinylated *scpA* promoter regions with GACGAG and GACGTG sites. The #PSMs indicates the total number of peptide spectra identified for each protein using the indicated oligo in the lysate pull-down assay. **(B)** SDS-polyacrylamide gel of ScoC overexpressed and purified from *E. coli* and stained with Coomassie. **(C)** Schematic of the *scpA* promoter region. The ScoC binding consensus sequence is shown in blue, the m6A site is in red, and the -35 box is also indicated. **(D)** ScoC binding to 5' IR dye end-labeled *scpA* promoter region containing a GACGAG or GACGTG site was determined via EMSA. Representative electrophoretic mobility shift assay (EMSA) of ScoC binding to *scpA* promoter regions is shown. The concentration of ScoC is shown with (-) indicating the absence of ScoC from the reaction. Oligos containing the GACGAG or GACGTG site are also indicated at the top of the gel. The DNA substrates used in the reaction are otherwise identical. **(E)** Quantification of the percent band shifted using 250 nM and 500 nM concentrations of ScoC for the GACGAG and GACGTG oligos as indicated on the graph. The percent band shifted was normalized to the no protein control for each substrate. Three replicates were completed with the error bars representing the standard error between reactions.
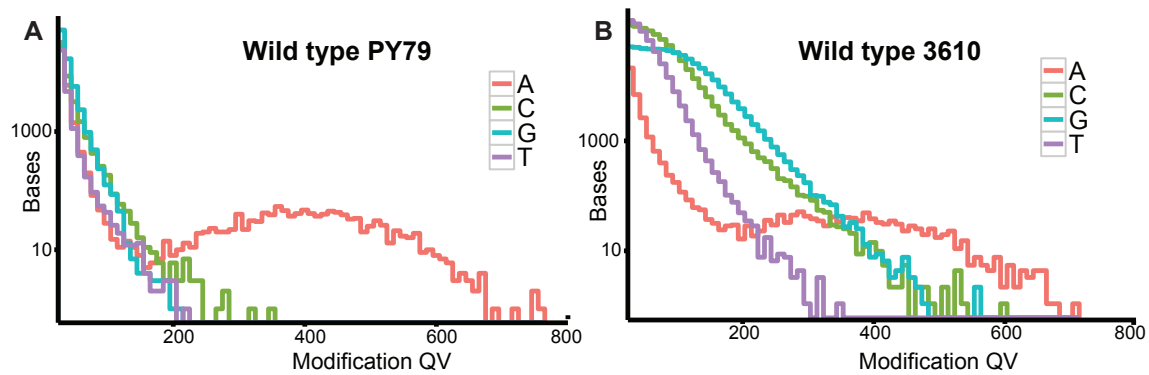
Fig. 3.7. The genome of *B. subtilis* strains contain m6A modifications. (A) PacBio SMRT sequencing of genomic DNA isolated from WT PY79 cells. Modification quality values (modQVs) indicate if the kinetics of the DNA polymerase differs from the expected background at a particular locus, where a modQV of 30 represents a p-value of 0.001. ModQVs are indicated on the x-axis and the number of bases is indicated on the y-axis. Each line represents the modification quality values for a particular nucleotide. **(B)** PacBio SMRT sequencing of genomic DNA isolated from the WT ancestral strain NCIB 3610.
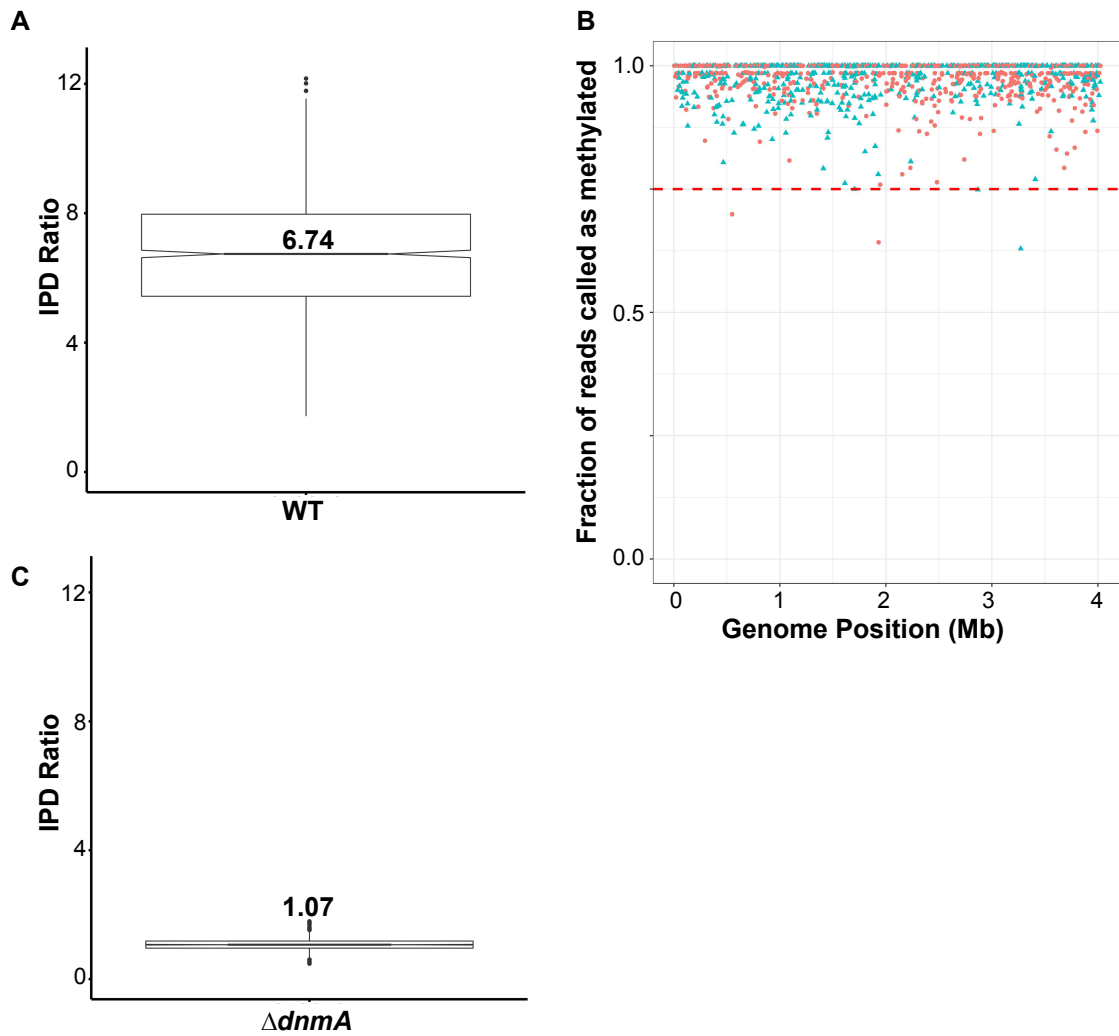
Fig. 3.8. GACGᵐAG sites have high modification scores throughout the *B. subtilis* PY79 genome. (A) Representative boxplot of interpulse duration (IPD) ratio values at GACGAG sites throughout the genome in WT cells. The median IPD ratio value is indicated. **(B)** The genomic location of each GACGᵐ<u>A</u>G site (x-axis) and the corresponding fraction of reads that were called as methylated at that position (y-axis) from PacBio SMRT sequencing is plotted. Sites that appear on the plus strand are indicated as a green triangle and those that appear on the minus strand are indicated as red dots. **(C)** Representative boxplot of the IPD ratio values at GACGAG sites throughout the genome in Δ*dnmA* (M.BsuPY79I) cells. Median IPD ratio value is indicated.
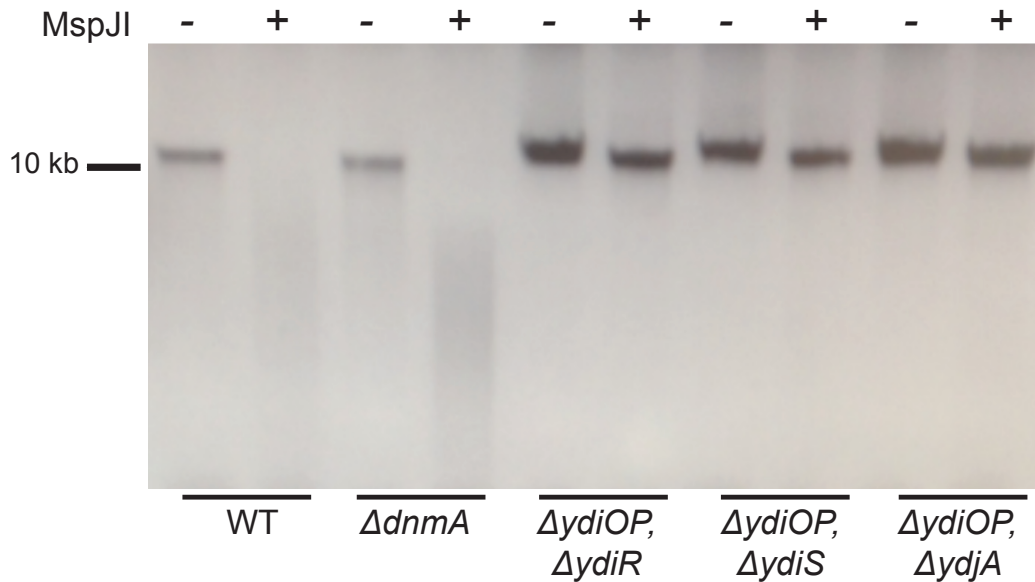
Fig. 3.9. Deletion of the BsuMI RM system eliminates m5C from the *B. subtilis* chromosome. Clean deletions were made for the coding regions of both subunits of the putative methyltransferase (*ydiOP*) in conjunction with separate deletions for each gene in a nearby operon coding for a putative restriction endonuclease (*ydiR, ydiS, ydjA*). DNA purified from these strains was subjected to 6 hours of treatment with a 5-methylcytidine and 5-hydroxymethylcytidine specific endonuclease MspJI. (-) indicates no treatment, (+) indicates treatment with MspJI.
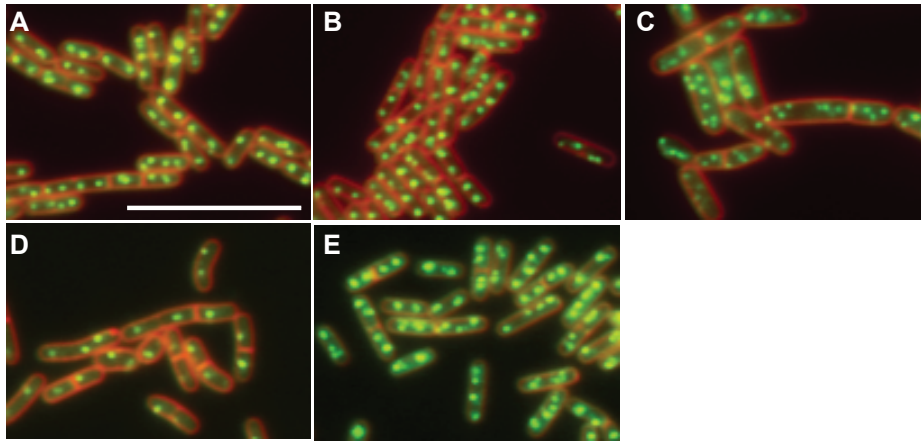
F		Percentage of cells with *n* Spo0J-GFP foci				
Relevant genotype	No. of cells	1	2	3	4	>4
wild type	1079	6	65	15	14	<1
ΔdnmA	1083	7	64	13	15	1
yabA::cat	1059	3	8	9	12	74
DnaAN depletion	1072	47	40	7	3	5
ΔyabB	1029	2	66	15	15	2

Fig. 3.10. Origin firing in *B. subtilis* is not regulated m6A. (A-E) Representative images of fluorescence microscopy for (A) WT, (B) Δ*dnmA*, (C) *yabA::cat*, (D) *dnaAN* depletion, and (E) Δ*yabB* strains expressing *spo0J::spo0J-gfp*, respectively. White bar = 10 μm. **(F)** Quantification of Spo0J-GFP foci for strains A-E.
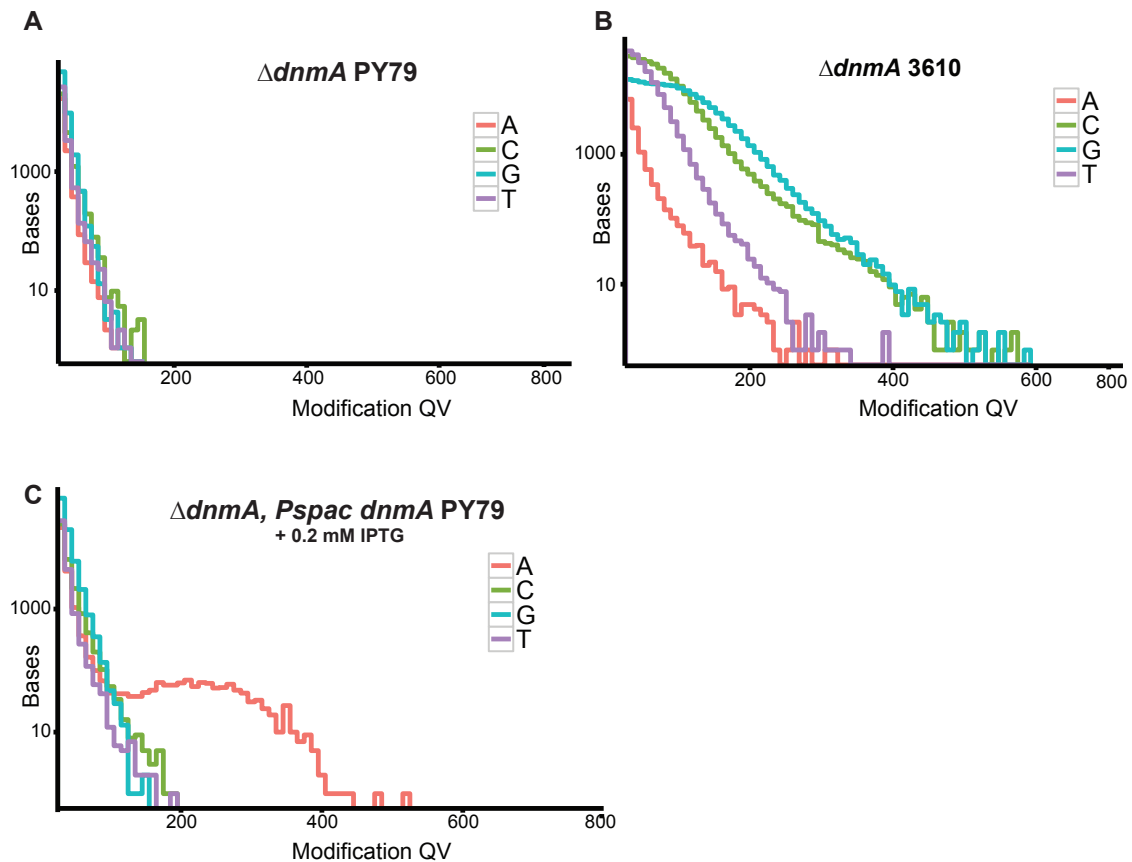
Fig. 3.11. *Bacillus subtilis* **m6A modifications are dependent on methyltransferase DnmA. (A)** PacBio SMRT sequencing of genomic DNA isolated from the Δ*dnmA* PY79 strain. **(B)** PacBio SMRT sequencing of genomic DNA isolated from the ancestral strain NCIB 3610 with a *dnmA* deletion. **(C)** PacBio SMRT sequencing of genomic DNA isolated from Δ*dnmA* cells ectopically expressing *dnmA* from the *amyE* locus with 0.2 mM IPTG.
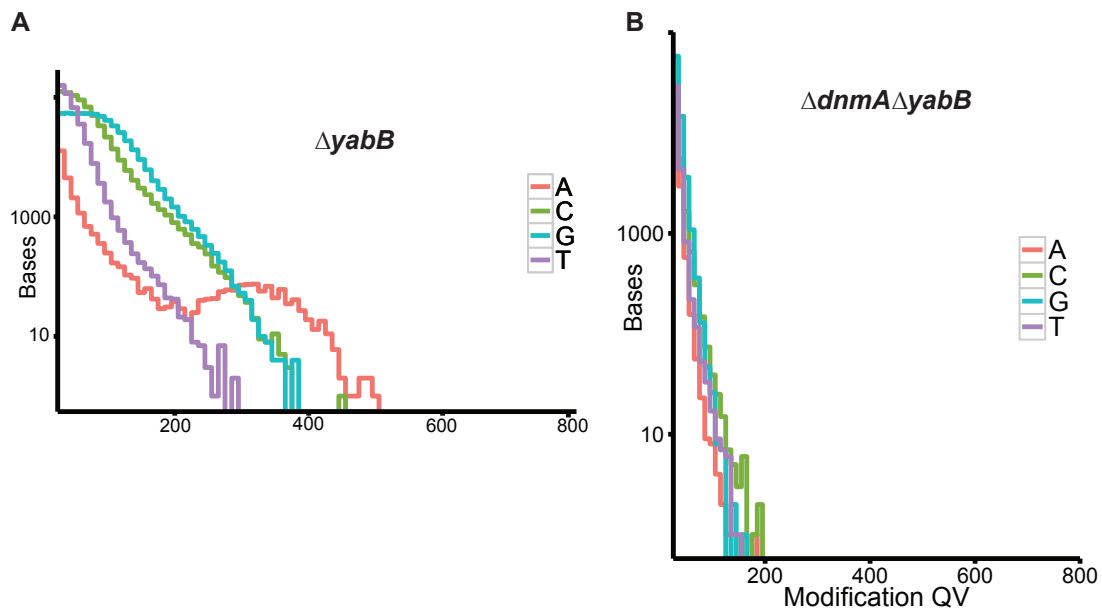
Fig. 3.12. Genomic m6A is present in a *yabB* deletion strain. (A) PacBio SMRT sequencing of genomic DNA isolated fromPY79 Δ*yabB* cells. **(B)** PacBio SMRT sequencing of genomic DNA isolated from PY79 Δ*dnmA*Δ*yabB* cells.
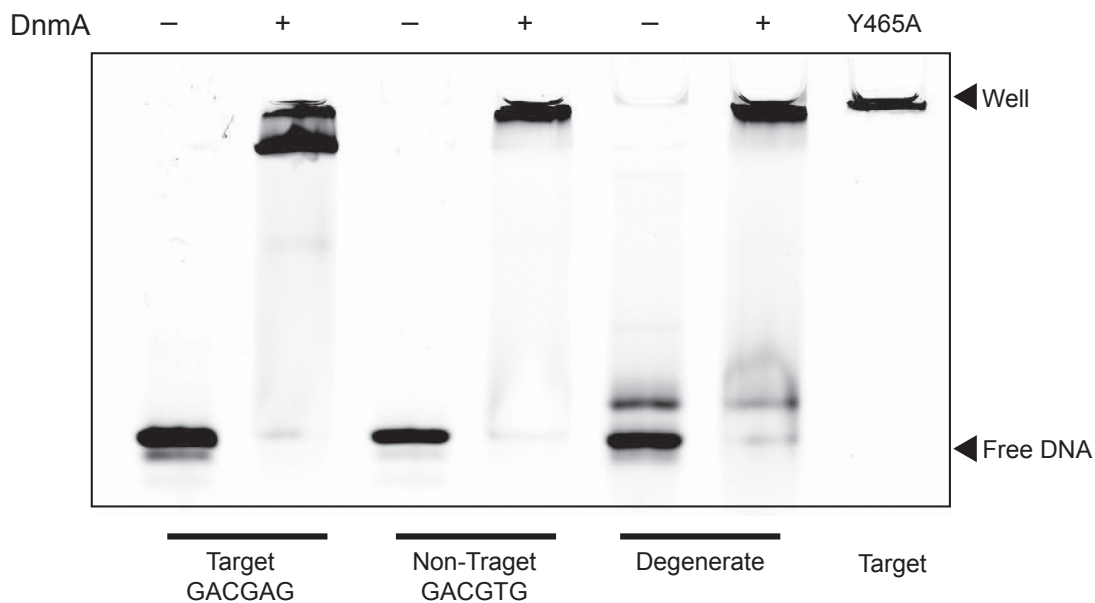
Fig. 3.13. DnmA binds DNA without the m6A motif. DnmA (M.BsuPY79I) substrate binding was determined by electrophoretic mobility shift assay (EMSA) with purified DnmA and varying substrates. 5' IR-labeled substrates include: target substrate (GACGAG), non-target substrate (GACGTG) and a degenerate sequence substrate, which are indicated at the bottom. The (-) indicates the absence of DnmA from the reaction, (+) indicates addition of DnmA to the reaction. As indicated, the final lane includes the DnmA catalytic inactive variant (Y465A) incubated with the target substrate.
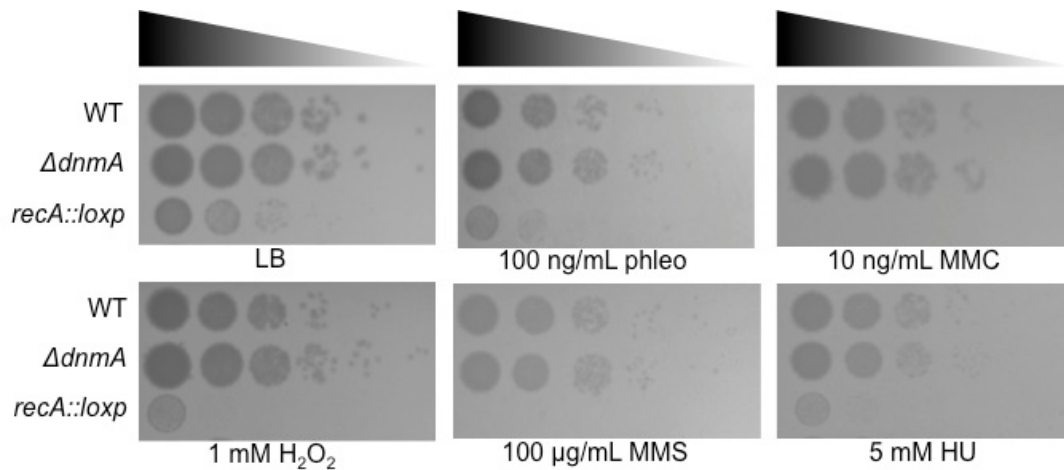
Fig. 3.14. Loss of m6A does not cause an increased susceptibility to genotoxic stress. WT and Δ*dnmA* cells were tested for their sensitivity to several DNA damaging agents and replication fork stress caused by hydroxyurea (HU). Cells were grown to mid-exponential growth phase, serially diluted, and plated on LB agar plates with the following concentrations of DNA damaging agents: 100 ng/mL phleomycin (phleo), 10 ng/mL mitomycin C (MMC), 1 mM hydrogen peroxide (H_2O_2), 100 μg/mL methyl methanesulfonate (MMS), and 5 mM hydroxyurea (HU). Cells with a *recA::loxP* disruption were used as a control.

Table 3.1. Relevant modified motifs detected in *B. subtilis* by PacBio SMRT sequencing.

Motif[a]	Type	%Detected	Mean QV	Mean Cov.	Mean IPD Ratio
WT PY79					
GACGAG	m6A	99.7	388	286	6.72
C**TCGARB**	m5C[b]	70.8	74	270	1.89
WT 3610					
GACGAG	m6A	94.7	362	313	4.84

[a]All motif calls by SMRT sequencing are reported in Table 3.6.
[b]Modification type confirmed via methylation sensitive restriction endonuclease digest as described in the supporting document.

Table 3.2. Relevant modified motifs detected in *B. subtilis* by PacBio SMRT sequencing.

Motif[a]	Type	%Detected	Mean QV	Mean Cov.	Mean IPD Ratio
Δ*dnmA* WT PY79					
C**TCGARB**	m5C[b]	46.7	51	120	2.00
Δ*dnmA* WT 3610					
None[c]				358	
Δ*dnmA, amyE::Pspac dnmA* PY79					
GACGAG	m6A	99.7	213	152	6.32
C**TCGARB**	m5C	52.7	59	149	2.00

[a]All motif calls by SMRT sequencing are reported in Table 3.8.
[b]Modification type confirmed via methylation sensitive restriction endonuclease treatment as described in the supporting document.
[c]GACGAG and CTCGARB were not detected in NCIB 3610 Δ*dnmA*. All other motifs called are reported in Table 3.8. The average coverage is reported for each spurious motif detected.

Table 3.3. Strains used in this study

Strain	Genotype	Source
JWS10	PY79	(90)
JWS261	Δ*ydiOP, ΔydiR*	This work
JWS262	Δ*ydiOP, ΔydiS*	This work
JWS263	Δ*ydiOP, ΔydjA*	This work
TMN1	Δ*yabB*	This work
TMN5	Δ*dnmA* (*M.BsuPY79I*)	This work
TMN7	Δ*dnmA, ΔyabB*	This work
DK1042	NCIB 3610 *comIQ12I*	(91)
TMN47	NCIB 3610*, ΔdnmA* (*M.Bsu3610I*)	This work
TMN16	Δ*dnmA, amyE::*P$_{spac}$*dnmA*	This work
JWS259	*spo0J::spo0J-gfp*	
JWS260	Δ*dnmA, spo0J::spo0J-gfp*	This work
TMN80	Δ*yabB, spo0J::spo0J-gfp*	This work
AK42	*yabA::cat, spo0J::spo0J-gfp*	Lab stock
LAS254	P$_{spac}$*dnaAN::cat, spo0J::spo0J-gfp*	(81)
BTS13	Δ*mutSL::spc*	(92)
LVG066	*amyE::PrbsV-GFP*	This work
LVG067	Δ*dnmA, amyE::PrbsV-GFP*	This work

LVG068	*amyE::PwprA-GFP*	This work
LVG069	*ΔdnmA, amyE::PwprA-GFP*	This work
LVG070	*amyE::PyloA-GFP*	This work
LVG071	*ΔdnmA, amyE::PyloA-GFP*	This work
LVG072	*amyE::PzapA-GFP*	This work
LVG073	*ΔdnmA, amyE::PzapA-GFP*	This work
LVG074	*amyE::PrnhC-GFP*	This work
LVG075	*ΔdnmA, amyE::PrnhC-GFP*	This work
LVG079	*amyE::PcomEA-GFP*	This work
LVG080	*ΔdnmA, amyE::PcomEA-GFP*	This work
LVG081	*amyE::PezrA-GFP*	This work
LVG082	*ΔdnmA, amyE::PezrA-GFP*	This work
LVG087	*amyE::PscpA-GFP*	This work
LVG088	*ΔdnmA, amyE::PscpA-GFP*	This work
LVG102	*amyE::PscpA mut1-GFP*	This work
LVG103	*ΔdnmA, amyE::PscpA mut1-GFP*	This work
LVG105	*amyE::Phbs-GFP*	This work
LVG106	*ΔdnmA, amyE::Phbs-GFP*	This work
LVG108	*ΔdnmA operon, amyE::PscpA-GFP*	This work
LVG109	*ΔdnmA operon, amyE::PscpAmut1-GFP*	This work
LVG118	*amyE::PscpA mut2-GFP*	This work
LVG119	*ΔdnmA operon, amyE::PscpAmut2-GFP*	This work
LVG120	*amyE::PscpA mut3-GFP*	This work
LVG121	*ΔdnmA operon, amyE::PscpAmut3-GFP*	This work

Table 3.4. Plasmids used in this study

Plasmid	Vector	Insert	Source
pJS146	pminiMAD	*ydiOP*	
pTN02	pE-SUMO	*dnmA* (*M.BsuPY79I*)	
pTN03	pDR110	*dnmA* (*M.BsuPY79I*)	
pAS2	pE-SUMO	*dnmA (Y465A)* (*M.BsuPY79I*)	
pTN12	pE-SUMO	*yabB*	
pTN13	pE-SUMO	*scoC*	
pLVG1	pDR111_GFP(Sp)	w/o lacI	(79)
pLVG1-0374	pLVG1	P*rsbV*/U0374	
pLVG1-0868	pLVG1	P*wprA*/U0868	
pLVG1-1292	pLVG1	P*yloA*/U1292	
pLVG1-1995	pLVG1	P*comE*/U1995	
pLVG1-2292	pLVG1	P*ezrA*/U2292	
pHP13		None	BGSC (http://www.bgsc.org)

Table 3.5. Oligonucleotides used in this study

Oligo name	Oligo sequence
oAS1	CCAAGAGCCGGTGGATTACCAAAAACATATACTTCTTC
oAS2	CCACCGGCTCTTGGTTCAAAAAAACAAAACAAAGAACATAAATC
oAS9	/5IRD700/GTGCAGCGTATCCGGACAATGACGAGACGGAAACAGAGCTGCTGCTG ATC
oAS10	/5IRD700/GTGCAGCGTATCCGGACAATGACGTGACGGAAACAGAGCTGCTGCTGATC
oAS11	/5IRD700/ATATAAACATACATACATACATTATTATATAAACATACATACATACATTA
oTMN5	GGCTCACCGCGAACAGATTGGAGGTATGGCGCTCATTGATTTAGAAGA TAAAATTGC
oTMN7	TGGTGGTGGTGGTGGTGCTCGACTACCGTTCTGTCATTTCTTGATACAA TTTAAGCAATAC
oTMN36	CACCGCGAACAGATTGGAGGTATGGTTTCATTACATGATGATGAAAGATTAGATTA
oTMN37	TGGTGGTGGTGGTGCTCGATTATTTGTCTCCATATAAAATGGTCCTGATTTC
oTMN38	GTGCAGCGTATCCGGACAATGACGAGACGGAAACAGAGCTGCTGCTGATC
oTMN39	GATCAGCAGCAGCTCTGTTTCCGTCTCGTCATTGTCCGGATACGCTGCAC
oTMN40	GTGCAGCGTATCCGGACAATGACGTGACGGAAACAGAGCTGCTGCTGATC
oTMN41	GATCAGCAGCAGCTCTGTTTCCGTCACGTCATTGTCCGGATACGCTGCAC
oTMN62	CACCGCGAACAGATTGGAGGTATGAATCGAGTGGAACCGCCCTATG
oTMN63	GGTGGTGGTGGTGGTGCTCGATTAACTGTTTACAGGTTCGAGCTCTTCAG
oTMN67	/5IRD700/CA AAACAGGATATGAAATAGTATTGGACGAGAGCTTTTTGGTGGCTTATACTATAG
oTMN68	/5IRD800/CTATAGTATAAGCCACCAAAAAGCTCTCGTCCAATACTATTTCATATCCTGTTTTG
oTMN70	/5IRD700/CAAAACAGGATATGAAATAGTATTGGACGTGAGCTTTTTGGTGGCTTATACTATAG
oTMN71	/5IRD800/CTATAGTATAAGCCACCAAAAAGCTCACGTCCAATACTATTTCATATCCTGTTTTG
oJR269	TAATGTATGTATGTATGTTTATATAATAATGTATGTATGTATGTTTATAT
oJR270	GUGCAGCGUAUCCGGACAAUGACGAGACGGAAACAGAGCUGCUGCUGAUC
oJR271	GAUCAGCAGCAGCUCUGUUUCCGUCUCGUCAUUGUCCGGAUACGCUGCAC
oLVGLS023A	GCTAGCTGATTAACTAATAAGGAGGACAAAC
oLVGLS023B	GAGAGTCGAATTCCTGCAGC
oLVGLS024A	CCGGGATCCGATGACCTCGTTTCCACCGAATTAGC
oLVGLS024B	CCGGGATCCGCAGGCCATGTCTGCCCGTATTTC
oLVGLS034	CGTATCACGAGGCCCTTTCG
oLVGLS042A	GAAATAGTATTGGACGTGAGCTTTTTGGTGGCTTATAC
oLVGLS042B	GCCACCAAAAAGCTCACGTCCAATACTATTTCATATCCTG
oLVGLS044A	GAAATAGTATTGGACGCGAGCTTTTTGGTGGCTTATAC
oLVGLS044B	GCCACCAAAAAGCTCGCGTCCAATACTATTTCATATCCTG
oLVGLS045A	GAAATAGTATTGGACGACAGCTTTTTGGTGGCTTATAC
oLVGLS045B	GCCACCAAAAAGCTGTCGTCCAATACTATTTCATATCCTG
oKWJ89	TTCTTCGCTTGGCTGAAAAT
oKWJ90	CACCAGGTTTTTGGTTTGCT
oLVG52A	5'Biotin-CTGCAGGAATTCGACTCT
oLVG52B	5'-Biotin-CCTTATTAGTTAATCAGCTAGC
oLVG_U0374F	CGCCATTCGCCAGGGCTGCAGGAATTCGACTCTCCCTGATCTGCAGAAGCTCATTG
oLVG_U0374R	CATGTTTGTCCTCCTTATTAGTTAATCAGCTAGCCTTCAAATCACTAGTTGCTTTATAC

oLVG_U0868F	CGCCATTCGCCAGGGCTGCAGGAATTCGACTCTCGGTCTGCATTTGCCAATTG
oLVG_U0868R	CATGTTTGTCCTCCTTATTAGTTAATCAGCTAGCAAAATAATGAATCTCCTTGAAGG
oLVG_U1292F	CGCCATTCGCCAGGGCTGCAGGAATTCGACTCTCGTTAATCCTTGTTTCATGGACG
oLVG_U1292R	CATGTTTGTCCTCCTTATTAGTTAATCAGCTAGCTCTCATTCTTCCTGCATTCGAT
oLVG_U1995F	CGCCATTCGCCAGGGCTGCAGGAATTCGACTCTCGCGTGACAGCTGATTTTACGG
oLVG_U1995R	CATGTTTGTCCTCCTTATTAGTTAATCAGCTAGCCGCAGTGAAAAAGCAGTTTC
oLVG_U2292F	CGCCATTCGCCAGGGCTGCAGGAATTCGACTCTCCGGAAGTATTGAAGTCGAG
oLVG_U2292R	CATGTTTGTCCTCCTTATTAGTTAATCAGCTAGCCGGAGTATCTATTCTTCCATTG
oLVG_U1780F	CGCCATTCGCCAGGGCTGCAGGAATTCGACTCTCGAATCATAAACGAAGGCTCTGG
oLVG_U1780R	CATGTTTGTCCTCCTTATTAGTTAATCAGCTAGCGTAGAGTAACACATATAAAAAGCCAT
oLVG_U1815F	CGCCATTCGCCAGGGCTGCAGGAATTCGACTCTCCTCTAGTGCTTCTTAGAAAGG
oLVG_U1815R	CATGTTTGTCCTCCTTATTAGTTAATCAGCTAGCTCACTCTCATTGCCGGAAAAAC
oLVG_U2212F	CGCCATTCGCCAGGGCTGCAGGAATTCGACTCTCGTAAGTGAACCGCTGTACG
oLVG_U2212R	CATGTTTGTCCTCCTTATTAGTTAATCAGCTAGCATTCCGCGAGAATCCTAG
oLVG_U2213F	CGCCATTCGCCAGGGCTGCAGGAATTCGACTCTCGTGAAGTGCTGGCCGTAAATG
oLVG_U2213R	CATGTTTGTCCTCCTTATTAGTTAATCAGCTAGCACTTTTCGCTGTATATACCAGTG

Oligo sequences in red represent RNA. /5IRD700/ indicates 5' infrared dye label with excitation at 700 nm.

Table 3.6. Modified motifs detected in *B. subtilis* by PacBio SMRT sequencing.

Motif	Type	%Detected	Mean QV	Mean Cov.	Mean IPD Ratio
WT PY79					
GACGAG	m6A	99.7	388	286	6.72
V**TTCGAG**NR	NA	79.2	75	284	1.90
C**TCGAR**B	m5C*	70.8	74	270	1.89
V**TTVGAG**NBY	NA	40.1	55	283	1.67
G**GN**B	NA	5.6	41	290	1.67
WT 3610					
GACGAG	m6A	94.7	362	313	4.84
R**AWKYAGYA**	m6A	28.7	98	309	1.66
D**TNRADDDG**	NA	22.8	61	305	1.78
D**TWTWGAAG**	NA	21.4	57	327	1.71
A**GCNMAAAWH**	m6A	15.8	107	322	1.53
T**NNNDNNH**	NA	12.6	61	303	1.78
D**TSNVCNTWNH**	NA	11.7	58	304	1.75
T**WGCNNNG**	NA	10.6	58	313	1.75
T**NRGCYNH**	NA	10.1	56	309	1.72
T**NNNCRVH**	NA	9.6	58	304	1.76
T**SNNNNNG**	NA	6.1	57	305	1.75
A**GDNNNNW**	m6A	4.3	104	325	1.71

*Modification type confirmed via methylation sensitive restriction endonuclease digest.
The motifs shown in this table are comprehensive to those presented in Table 3.1.

Table 3.7. Cells with Δ*dnmA* are wild type for mutation rate

Strain	No. of cultures	Mutations per culture	Mutation rate (Mutations per generation 10^{-8} ± [95%CI]	Relative mutation rate
Wild Type	20	0.60	1.7 [0.95-1.68]	1

ΔdnmA	22	0.72	1.8 [0.94-2.2]	1.05
ΔmutSL	22	41.3	120.8 [109-131]	71.1

Mutagenesis assays were done as described using rif[R] as an indicator. Mutation rate and mutations per culture were calculated using the Ma-Sandri-Sarkar Maximum Likelihood Estimator with the web-based tool FALCOR (76).

Table 3.8. Modified motifs detected in *B. subtilis* by PacBio SMRT sequencing.

Motif	Type	%Detected	Mean QV	Mean Cov.	Mean IPD Ratio
ΔdnmA WT PY79					
VT**TCGAGNR**	NA	62.5	53	118	2.00
C**TCGARB**	m5C*	46.7	51	120	2.00
ΔdnmA WT 3610					
VA**TATRGCA**	m6A	54.0	88	363	2.00
R**A**H**KYAGYA**	m6A	31.0	110	357	1.67
DDT**NRGCNTHNH**	NA	20.4	60	356	1.72
DNNDT**GYAADNG**	NA	20.3	65	348	1.81
DT**NRVDDDG**	NA	15.2	61	355	1.73
T**NNNDNNH**	NA	12.3	62	353	1.74
T**NNNCRVH**	NA	9.4	60	359	1.72
A**GNNMRNA**	m6A	9.1	109	359	1.55
T**NNSCBDH**	NA	7.2	58	363	1.68
T**SNNBNNG**	NA	6.4	58	361	1.71
A**GNNDNNW**	m6A	3.4	101	365	1.54
A**NDNNNNH**	m6A	0.8	98	367	1.53
ΔdnmA, amyE::Pspac dnmA PY79					
GACGA**G**	m6A	99.7	213	152	6.32
VT**TCGAGNR**	NA	67.2	59	146	1.98
C**TCGARB**	m5C	52.7	59	149	2.00
MNGACGA**WCC**	NA	47.3	58	152	2.20
VT**TCGAGBB**	NA	38.1	53	157	1.82
WAGACGA**WB**	NA	21.7	53	148	2.19
G**GNNB**	NA	6.6	40	168	1.86

*Modification type confirmed via methylation sensitive restriction endonuclease treatment.
The motifs shown in this table are comprehensive to those presented in Table 3.2.

Table 3.9. Modified motifs detected in *B. subtilis* by PacBio SMRT sequencing.

Motif	Type	%Detected	Mean QV	Mean Cov.	Mean IPD Ratio
ΔyabB					
GACGA**G**	m6A	96.9	304	240	5.15
A**TATRGCA**	m6A	74.0	80	238	2.09
A**DGYACYTV**	m6A	34.7	85	238	2.01
A**DKYASYA**	m6A	29.8	88	238	1.88
A**GCNAAAAWH**	m6A	17.3	95	237	1.58
G**ANNBNRCA**	m6A	13.7	98	241	1.90
T**NNNNNNH**	NA	12.2	46	234	1.69
DT**VVVNNDG**	NA	11.1	44	234	1.67
A**NVBANYW**	m6A	6.4	78	238	1.81
A**GDNVDNW**	m6A	5.4	87	236	1.81
T**BNNDNNG**	NA	5.2	43	235	1.67
A**GBB**	m6A	2.7	99	238	2.02

ΔdnmA ΔyabB					
VT**TCGAG**NR	NA	71.9	60	166	1.92
C**TCGAR**B	NA	57.3	57	168	1.91
VT**TCGAG**BY	NA	33.2	46	181	1.70
G**G**	NA	3.1	39	185	1.77

Table 3.10. Identification of protein species in DnmA protein purification

Identified Protein Accession Number	Molecular Weight	Total Spectrum Count
YEEA_BACSU	101 kDa	190
SMT3_YEAST	12 kDa	42
K2C1_HUMAN	66 kDa	34
HORN_HUMAN	282 kDa	34
TRYP_PIG	24 kDa	18
K1C10_HUMAN	59 kDa	19
K22E_HUMAN	65 kDa	21
K1C9_HUMAN	62 kDa	15
K2C5_HUMAN	62 kDa	4
CYTA_HUMAN	11 kDa	3
K1C14_HUMAN	52 kDa	8
ALBU_HUMAN	69 kDa	2
ARGI1_HUMAN	35 kDa	3
ANXA2_BOVIN (+8)	39 kDa	2
FABP5_HUMAN	15 kDa	2
SBSN_HUMAN	61 kDa	2

Mass spectrometry was completed by the University of Michigan Core on the high molecular weight species, confirming the presence of DnmA (YeeA). The SMT3_Yeast contaminant is likely the result of trace SUMO-tagged DnmA from the protein purification process (see Materials and Methods).

Table 3.11. Promoter upshifts containing the m6A motif

Upshift	Sequence context	Downstream TU
U374	cgatgattttacgttaattgttttgcggagagaaaG<u>GTTTAA</u>cgtctgtcagacga<u>GGGTATA</u>AAGCAACTAGTGatttgaaggaaaatttgaggtgatacga	**rsbV, rsbW, sigB, rsbX**
U868	atattccaaatcatttaaaataaccttaaaattccctgtaagcggtatctcgtcctatgaaatTATGATACCTTCAAGGAGATtcattattttgcaggaggg	**wprA**
U1292	gctctcttatagagatatcactctataagcatgcttaTT<u>CTGA</u>ctcgtcccatttCATGC<u>TATAAT</u>TATCGAATGcaggaagaatgagagggtgtattgca	**yloA**
U1780	aaaaaaggaatattcgttcggtaaatcaccttaaatcC<u>TTGAC</u>gagcaagggattgacg<u>CTTTAAAAT</u>GCTTGATATGGcttttatatgtgttactctac	**S861, hbs**
U1815	ctgtacaaactccttcaaaacaggatatgaaatagTATT<u>TGGAC</u>gagagcttttttgg<u>TGGCTTATACT</u>ATAGGGTAGccagtttttccggcaatgagagtga	**scpA, spcB, ypuI**
U1995	taaaacgatggtttttaaaatgcttttttatgCTT<u>TTGCA</u>gtacagacgaacgTA<u>TGACATA</u><u>CTCGTC</u>TACACatgaaactgcttttttcactgcggaaat	**S963, comEA-EC, S962, holA yshA (zapA), yshB, S1080, polX, mutSB**
U2212	tgtatataccagtgtatcataacagcgggaggctcgtc<u>TTTCCA</u>TTcatttaataaaCG<u>TGTTATGAT</u>AAGAACTAGgattctcgcggaatggaggagaaa	
U2213	gttcttatcataacacgtttattaaatgaatggaaagacgagcctcccgctgt<u>TATGAT</u>ACACTGGTATATACagcgaaaagtgtaaaaaaaggagattat	**rnhC**
U2292	aattttt ggacgagtgtgatgtgaccgactcttt<u>TGGCTT</u>ATAaaacgccgagaGA<u>TGCTACAAT</u>GGAAGAATAgatactccggtaatattgttcatatac	**S1127, S1125, ezrA**
U2511	acggatgggcacgttagatccttacgatttttctgctgac<u>TCGTC</u>AATtgaacggcaaTA<u>TGGTATAAT</u>TAATAACAAtttcatttaggaggcaattt	**yumC**
U624	cactaaagttgatcaaatgacctaagtgcgccaaacgt<u>GTTA</u>CGgggacgagctatctCA<u>TGGTATAAA</u>TGGAATTGTaaacgttatcaggaggtcgtcat	**malA, malR-Q**
U792	atattcagctcagtcctttttgatgcgtctttcccgCACATAactcgtctcattcccGCATATG<u>GTTAAGAAT</u>AAAgaatctatgcaaagggggaggagcc	**yhaL**

134

ID	Sequence	Genes
U1042	agcggtctgaggctttttattagataaaagacctaattgttctaTGAAACTTTCtgacgagatatCCGTATATCATACAGAttgacatcattcacatcaga	S462, S461
U1293	tacaccctctcattcttcctgcattcgataattatAGCATGaaatgggacgagtcaGAATAAGCATGCTTATAGAgtgatatctcataagagagcagggg	yloB, S571, yloC, ylzA, gmk, rpoZ
U1406	atgtgttcataaaaaactaaaaaaaaTATTGAAAatactgacgaggttATATAAGATGAAAATAAGTTagtttgtttaaacaacaaactaataggtgatgt	S634, xylA, S635, xylB, S636
U1861	ctttcccgtcatataaactgctcaaatgaaccgcttttgtcaaacaTTTCTGTATAATAGACGAGAaatcagtttattttttcagtatagaagcatttttat	yqjN
U2116	atcaagggaataaagatcgtttttatgacgagccaaaagggtgaccatAACATAtcaagcataaaaaagACATAGACTGTTAACAGAAccagcaaacaaat	S1041, glnQ, glnH, glnM, glnP
U2235	gtgtattggctttgcggaaaaaagggtggaacCacgattccgttcattcaaCCTCGTCCCTttcataggggggcggggtttttatatgcaaaaaaaggagtg	thrS
U2676	cttactttaaaaagccacgcaacacggttctCGtcacagacgaaggagccgcAaagaagttattaagggatgaataatccctcaatataaatatctctcac	gapA, S1301, S1300, pgk, tpiA, pgm, eno
U3059	gcgaaacacacatgttctccactaaaaaagAGTATAtccggtatagatagaCGAGAAACTGAAAGGGAAAcctcattcgtttacatattggcttcagcgga	aldY
U75	cctgatttccaaatcatttgcggatcgcctgaacTGCttgtcagcaaaaagggcAAGCTATTAGAGACGAGACcgattgcgggcacccgttccagagggaa	secondary internal pabB upshift
U568	aaaagagtatctcgtatagaacataagaagaatgacgagtttttaAGATTAtcatcaattatgtgaGAATAAAATATTATAAGGGaaaatgaaggctgtc	S257 - as-fecF
U797	ctattttttcttttttgcagatgattgGCAGGACATggatgactttaTGTACAAATAAAATAACGAcgagcaaaccgccgaataaagatacatattcatcg	as-yhaJ; secondary yhaI upshift
U809	tttgctgcggaagacatatacagataaagaacagccATAATGacatagacgagaagcGCATACATATAAACAGATTggaaaacaaaataaagaacgagtgt	as-ecsB (ecsB internal)
U895	aagaaggtccgtattaattttttcccaCTCATAAaccttactttacccCACCATACTATTgaagacgactcatctcgtcaaagtatggaagggggcagtctc	S397, yizC
U911	tcctgagaaggttttaacaaatcatgatcttgaaaaaatggtTGAAACTTCTgacgagtggatTCGTACAAGAACAGGAaatagaagaaagaagaatcgcag	secondary internal fabHA upshift
U936	ataaaacattctaaaggcggtgtttccgacggcttcggctcgtcgccaAGCATCtcaaataaaatttctTCATATACATCATATGAGTagctgccaggaac	S416 - as-spxH
U1010	cgggtagccgaccttcggattatcaAACTTGACGaggcagcggcatcACGGCCACACTTGTGATAAAcaaggcgccaagtgatccgatttggcggaacgtcg	as-yjmB
U1122	gttcttaacggttatatgaaccaattcattcgaaaagacaCCTTTACAaacatacgttcgttaAATTATAATAAAACAGACGAgctgccatccagcatccc	S498; independent transcript
U1389	ttgttgttgatattcttgaataaaaaaacccggtttctCGCGAtgaggagccgggtttttttatGAGACGCTCGTCCCCGTCtcggctatgattctaggat	S623 - as-cwlC
U1429	gaatgtaggtctctttgaggctttagcacgaattctctcgtcctccttgttaaattttgttaaattcaCAATATTATAtaccattagcccgggcgctgtt	S653 - secondary internal surA upshift
U1984	aaatattttattcaaagtcagccagtcgacgagTATTTAAacgcactcgaacaGGAAATGATAAAtacaatagcaaaaaccgggcagcctgatctcaaaac	secondary hrcA-grpE-dnaK upshift
U2466	tcatttctttggcgttacgtcattcatcctcgtcatggcctttcgcgcttatcaggaattgaaatcggacgatatggatcaaatgaggggaaatgatcaa	secondary mrpD-G upshift
U1405	ttcatcttatataacctcgtcagtattttcaatatttttTTAGtttttatgaacacATTAGATATAATAAAGGGAAgattcgctatgtactatgttgat	S633, xylR
U1996	ttttacctgacgagtttgtaaaaaatatttttcatattACACctgagaaattaaaggAACGAAATGTAAAAGGAAttattactgacctggataatacgctt	yqeH, aroD, yqeI, nadD, yqeK, yqeL, yqeM
U2030	gagattttggatatagacgaggctcttggttcacGAATTCACCAgatgtgccgtgatTACATAGTGATTATTAGAGGcgatcgaatgcaattaaatcatag	yqzO, yqaN, S982, yqaO, S980
U2255	cctcctctcgtcagcatgtcctattttttatatgTATTCACGCtgcggctgaataTGAATACATTCATCTTAAAGgagggatggcatgtgtttacacaagc	ytwI
U2486	cggaggcaggatgacgagccacagccctctgttGGTTTGAtcgctctcctgagaGGGAAAACTGAAAGAAACGcggtcatccggcagatcgtataccatcc	S1225; as-yukBC
U3058	ttcagtttctcgtctatctataccggatatactctttttTAGTGGAGaacatgtgtgtttCGCCTTATACTGAATATACAgatccttacataagagaggag	yxkF, S1490, msmX
U3138	agacgaggtttcttataagcctttTTCATcctttttcCCTCCTTCCTTGTAAaaaaataggctatcacgcacaactaaatattataatcctctgataattct	yxbC
U3060	agatagacgagaaactgaaagggaaacctcattCGTTTACatattggcttcagcGGAAATAGAAGAAGACATGcaggaccaaaggagggtcatcttatgag	aldY
U118	tttgaaaataaaaaatttaattttcctCTTTACAAacagggggtgacctGTATATAATAACCTTTTGTCAgctcgacgagaacacaacggcccgttggtcaa	trnSL-Glu2, -Val1,-Thr1, -Tyr1, -Gln2, S67
U301	cagcagcgatcgcggcctatgcgaatcaaaacGGATTACttttgctgacagcGGGAATTAACGGTAATATCatccgctttttgacaccgctcgtcatctca	secondary gabD upshift
U902	atcagactcttttgtcacctcactttctgctaaaattggaTTCCCCCTtcgctttttgTATGGTATGATAACTTTTAGaatagaatgagaaggacgaggtg	yjzD, S399
U934	acaacctctatgcttaatgttcatattttttgtcacaaAACATAacgaagtgcattcacTCATATCCTTATAAGGAAAaaggacgaggaggaccgccatgat	yizD
U1428	agcttacttttcataaattcaaaaatgagaagAACAGCGCCCgggctaatggtatATAATATTGTGAATTTAACAaaaatttaacaaggaggacgagagaa	S654, yndL, S657, fosB, S658, S659
U1534	tatagatcagaacaaaagttcgatgtaaatgttggtaataaaatataaaggtcaataatgaTATCCGTAGTATTAATAAAGgagagattcttttcgacgag	yozM
U1794	atttcgagcttttcacttgaactgaaaaaGGTTTGatgcgtcgggtaattGTAAATACTGTTAACGACAttgccttttttcgatgacgagggcatcaatgcc	S869 - sporulation sRNA - as-ypdA
U1806	taaggctcttttttagttgctattcatataatagaaATTTtcaaaaaaaagttgTACGTGTATAATAAAACAAGgtaaagattgaaaggatttgagacgag	aroC
U1860	cagcaattcgtccttcgtctgccaattcattgattccgcccctcctaTGAATAAaaatgcttctataCTGAAAAATAAActgatttctcgtctattataca	yqjM
U2140	tgtaagaaaaaccgattgcatttcacaaagcttttaCGTCTAattcatgggataaggGAATACATTTTTACAAAGAcgagccatcagcatgtctgacggtt	yrzE
U2221	aattgttaaaatgcgtgatatttcatcagtattCCTCGGAGCAatcacggcatcaGGCATAGACTGATACTGAGGcgtcgcatcatatgaatagacgagac	S1083; as-ysfB
U2256	aaacacatgccatccctcctttaagatgaatgtaTTCATAttcagccgcagcgtGAATACATATAAAAAATAGgacatgctgacgagaggaggaccgtttt	ytvI
U2654	tgtttctctttctatattttatcaatcacgctTGCATGCCctccctcgttattTGCGTTATAATAGTGACAGAcgaggtgaaaagtatgaaccaatcagaa	yvaP, S1288, yvaQ
U3003	tacggacaattcagagcatattggctcttcattgcgCCTTTtctttcaatatttgaTGCGTTAAAATGGTAACCGTgtgaaaagatgctagacgaggaaaa	S1473, S1472, S1471, qoxA-D

Subset of transcribed regions 5' of ORFs identified in Nicholas *et al.* (51) that contain the m6A motif (indicated in red). Capitalized letters and underscores indicate predicted sigma factor binding sites. Downstream transcriptional units are listed.

Supporting Text

Supplementary Results

m5C modifications function as part of the BsuMI restriction-modification system.
The analysis of SMRT sequencing detected cytidine methylation in the PY79 genome
(**Table 3.1**). In *B. subtilis* Marburg the BsuMI RM system was first found to recognize 5'
YTCGAR sites and later refined using analysis of transformation efficiency to recognize
5' CTCGAG (47). This work showed that in *B. subtilis* Marburg the *ydiO-ydiP* operon
codes for the methyltransferase (MTase) responsible for m5C modifications of the
BsuMI RM system and that an adjacent operon, *ydiR-ydiS-ydjA,* codes for the cognate
endonuclease (47). Given the sequence similarity between the mC motif detected in the
WT strain PY79, 5' CTCGARB, and the site identified in the Marburg strain, 5'
CTCGAG, we decided to test whether YdiO-YdiP was responsible for cytidine
methylation in PY79. Because PacBio does not robustly detect m5C methylation, we
experimentally determined the modification type by treating DNA with the m5C- and 5-
hydroxymethylcytosine-specific endonuclease, MspJI (**Fig 3.9**). We created PY79
strains with deletions of *ydiO-ydiP* and each subunit of the putative endonuclease, *ydiR,
ydiS,* or *ydjA.* DNA was purified from each of these strains in addition to WT and a strain
with a deletion of N6-methyladenosine methyltransferase Δ*dnmA*, as controls. DNA
from WT, Δ*dnmA*, and strains lacking *ydiO-ydiP* plus the respective restriction
endonuclease subunits were treated with MspJI, recognizes 5- 5hmC and m5C at 5'
ᵐCNNR sites, followed by electrophoresis on an agarose gel. Smearing in WT and
Δ*dnmA* strains indicates the presence of m5C modifications whereas distinct bands in
Δ*ydiOP*Δ*ydiR,* Δ*ydiOP*Δ*ydiS,* Δ*ydiOP*Δ*ydjA* strains indicates loss of m5C modifications,
implicating *ydiOP* as the MTase responsible for all m5C methylation in the *B. subtilis*
genome (**Fig 3.9**). The results we present here confirm the BsuMI RM recognition site
as 5'CTCGARB in *B. subtilis* strain PY79. The m5C motif identified in PY79 was not
detected as modified in NCIB 3610 by PacBio SMRT sequencing (**Table 3.1**).

B. subtilis m6A does not function in replication timing. We sought to determine the
consequence of m6A loss in *B. subtilis* cells. In the Gram-negative bacterium *E. coli*,

GATC-specific m6A functions in origin sequestration (17,18,35), DNA mismatch repair (20,21), and the regulation of gene expression (69). The methylation status of palindromic GATC sites in the *E. coli* origin of replication regulates the binding of SeqA, which inhibits origin firing by sequestering the origin region (17,18,35). Whereas GATC sites are enriched in the *E. coli* origin, as discussed in the main text the m6A motif is not present in the *B. subtilis* replication origin, although a high density of m6A sites flank the origin on the left arm **(Fig 3.1)**. To empirically determine if m6A sites located in the origin proximal region on the left arm influence origin duplication, we assessed the origin proximal copy number in exponentially growing WT and Δ*dnmA* (M.BsuPY79I) cells using Spo0J-GFP (*parB-gfp*) as a marker for origin copy number as done previously (80,81). We show that in WT, Δ*dnmA, ΔyabB* cells, 65%, 64%, 66% of cells showed two Spo0J-GFP foci, respectively. As controls we used a deletion of *yabA*, a negative regulator of origin firing (82), and show that 74% of cells have four or more foci as expected (81). As a hypo-initiation control we used an IPTG regulated promoter (P$_{spac}$*dnaAN*) to deplete the replication initiation protein *dnaA* and the replication sliding clamp *dnaN*. We show a near 8-fold increase in the percentage of cells with a single Spo0J-GFP focus, demonstrating an inhibition of DNA replication initiation (81) **(Fig 3.10)**. With these results we show no difference in origin proximal copy number between WT, Δ*dnmA,* or Δ*yabB* cells as determined by fluorescence microscopy and we conclude that m6A does not contribute to the regulation of DNA replication initiation.

B. subtilis m6A does not function in DNA mismatch repair. In addition to origin sequestration, methylation at GATC sites in *E. coli* also functions in strand discrimination during DNA mismatch repair, thereby ensuring removal of mismatched bases from the nascent strand (20). Both the loss of adenosine methylation at GATC sites and hyper-methylation of the chromosome by overexpression of Dam resulted in an increase in spontaneous mutation rate (83,84). m6A sites are non-palindromic in *B. subtilis* and occur far less frequently (~1,200 sites relative to ~20,000 GATC sites in *E. coli*). The lack of an even distribution on the leading and lagging strands across the genome and the low number of sites does not support a contribution of m6A to strand discrimination during mismatch repair. To be certain, we conducted rifampin resistance

assays as a measure for mutation rate (75,85,86) in WT and Δ*dnmA* strains. No difference in mutation rate between these strains was observed as compared to a mismatch repair deleted control **(Table 3.11)**. These results indicated that the presence or absence of m6A does not influence spontaneous mutagenesis in *B. subtilis* (87,88). Furthermore, because the m6A sites occur multiple times at the *addA* locus and AddA is important for recombinational repair (89), we performed spot titer assays to determine if Δ*dnmA* cells were more sensitized to DNA damaging agents relative to WT cells and found no increase in sensitivity **(Fig 3.14)**.

Supplementary Materials and Methods

Chromosomal DNA digestion by MspJI: Genomic DNA was purified from PY79, Δ*dnmA*, Δ*ydiOP*Δ*ydiR*, Δ*ydiOP*Δ*ydiS*, and Δ*ydiOP*Δ*ydjA* strains as described above and treated for six hours with MspJI according to the manufacturer's recommendations (New England BioLabs). For each control the reaction was set up exactly like the experimental group with an equivalent amount of water added instead of MspJI. Each reaction was then loaded on a 0.7% agarose gel and electrophoresed, stained with ethidium bromide, and visualized by illumination with UV.

DnmA (M.BsuPY79I) Y465A: A PCR reaction was performed using specially designed primers to create two overlapping blocks of DNA coding for *dnmA* with an alanine in the place of the tyrosine usually found in the NPP<u>Y</u> catalytic motif. The 5' block was created by PCR using oTMN5 and oAS1 with *B. subtilis* genomic DNA as the template. The 3' block was created by PCR using oAS2 and oTMN7 with *B. subtilis* genomic DNA as the template. PCR products were gel extracted, purified, and combined with pE-SUMO vector via Gibson assembly to create pAS2. The resulting plasmid was used to transform *E. coli* MC1061 cells and plated on LB agar containing 25 µg/ml kanamycin. Resulting colonies were PCR screened for presence of the *dnmA* gene using oTMN5 and oTMN7 and further verified by Sanger sequencing. BL21$_{DE3}$ cells containing this plasmid were then tested for their ability to overexpress the mutant protein with addition of 200 µM IPTG.

Electrophoretic Mobile Shift Assay (EMSA): EMSAs were performed using 1 µM DnmA and 5' IR dye labeled substrates at 0.62 µM in a buffer containing 100 mM Tris-HCl pH 8, 250 mM NaCl, and 1 mM $MgSO_4$. The substrates were annealed in the same buffer by heating to 100°C for 30 seconds and then allowed to cool back to room temperature on the bench top. Substrates included the target sequence (oAS09, oTMN39), non-target (oAS10, oTMN41), and a degenerate sequence (oAS11, oJR269). A no protein control was used for each substrate and catalytically inactive DnmA (Y465A) was assayed with the target sequence. These assays were performed at 30°C for 15 minutes. Samples were then loaded onto, and resolved via 6% native-PAGE electrophoresed on ice at 100V and visualized with a LI-COR Odyssey imager.

Spot titer assays: The indicated strains were struck from frozen stocks onto LB agar plates and incubated overnight at 30°C. Single colonies were inoculated into 2 mL of LB media and grown in a rolling rack at 37°C to an OD_{600} of 0.6-0.8. Strains were then diluted to an OD_{600} of 0.5 and subsequent 10-fold serial dilutions were performed in 0.85% saline solution. The dilutions (4 µL) were then spotted on LB agar and LB agar plus the indicated concentrations of exogenous DNA damaging agent or HU. Spots were allowed to dry, and the plates were incubated at 30°C overnight.

Mass Spectrometry: Mass spectrometry was performed by The University of Michigan Proteomics & Peptide Synthesis Core, project number MS976/M1516-086. Briefly, the band of interest was excised from SDS-PAGE and placed in 50 µl of distilled water. The band was then digested with trypsin and analyzed using LC/MS/MS on a ThermoFisher Orbitrap mass spectrometer. Resulting data was searched against the NCBI protein database and presented in Supplementary Table 3.10.

Spontaneous mutagenesis assay: Protocol was followed essentially as described (75). Briefly, frozen strains were struck out on LB and grown at 30°C overnight. Single colonies were inoculated into 3 mL of LB media and grow at 37°C to an OD_{600} between 1 and 1.2. At this point, 1.5 mL of culture was pelleted by centrifugation and the supernatant was aspirated. Cells were resuspended in 0.85% saline and two 1,000-fold

serial dilutions were performed in 0.85% saline. 100 μL of the original solution was plated on LB plates containing 100 μg/mL rifampin and grown at 30°C overnight and 100 μL from the 10^{-6} dilution was plated on LB and grown at 30°C overnight. The number of single colonies on each plate was counted the next morning and mutation rate was calculated using the Ma-Sandri-Sarkar Maximum Likelihood Estimator Method through the FALCOR fluctuation analysis calculator (76). All strains were independently grown and plated on at least three different days.

Live cell microscopy: Protocol was followed essentially as described (77). Frozen strains were struck on LB plates and grown overnight at 37°C. Plates were washed with defined S7$_{50}$ minimal media and diluted back to an OD$_{600}$ of 0.05 in 2 mL of defined S7$_{50}$ minimal media and grown at 37°C to mid-exponential growth phase (OD$_{600}$ between 0.6-0.8). 1 mL aliquots were then treated with 1 μL of FM4-64, the vital membrane strain, and spotted onto 1% agarose pads containing 1X Spizizen's salts. Fluorescence microscopy was performed with an Olympus BX61 microscope. The Olympus 100X oil immersion 1.45 numerical aperture (NA) total internal reflection fluorescence microscopy (TIRFM) objective lens was used for all imaging and all strains were independently imaged on at least three different days.

Strain construction

JWS261 (Δ*ydiOP*, Δ*ydiR*): PY79 was transformed with genomic DNA from BKE06090 to make strain JWS245. JWS245 was transformed with pDR224 to make JWS248. JWS248 was transformed with pJS146.

JWS262 (Δ*ydiOP*, Δ*ydiS*): PY79 was transformed with genomic DNA from BKE06100 to make strain JWS246. JWS246 was transformed with pDR224 to make JWS249. JWS249 was transformed with pJS146.

JWS263 (Δ*ydiOP*, Δ*ydjA*): PY79 was transformed with genomic DNA from BKE06110 to make strain JWS247. JWS247 was transformed with pDR224 to make JWS250. JWS250 was transformed with pJS146.

TMN1 and TMN2 (Δ*yabB)*: PY79 was transformed with genomic DNA from BKE00340 to make strain JWS230. JWS230 was transformed with pDR224.

TMN5 and TMN6 (Δ*dnmA*): PY79 was transformed with genomic DNA from BKE06760 to make strain JWS230. JWS230 was transformed with pDR224.

TMN16 (Δ*dnmA, amyE::Pspac dnmA*): TMN5 was transformed with pTN003.

TMN47 (Δ*dnmA* in NCIB 3610*)*: DK1042 was transformed with genomic DNA from JWS230. JWS230 was transformed with pDR224.

JWS260 (Δ*dnmA, spo0J::spo0J-gfp*): TMN5 was transformed with genomic DNA from JWS259.

TMN80 (Δ*dnmA, spo0J::spo0J-gfp)*: TMN2 was transformed with genomic DNA from JWS259.

LVG066 (*amyE*::*PrbsV-GFP*): PY79 was transformed with plasmid pLVG1-374.

LVG067 (Δ*dnmA*, *amyE*::*PrbsV-GFP*): TMN06 was transformed with plasmid pLVG1-0374.

LVG068 (*amyE*::*PwprA-GFP*): PY79 was transformed with plasmid pLVG1-0868.

LVG069 (Δ*dnmA*, *amyE*::*PwprA-GFP*): TMN06 was transformed with plasmid pLVG1-0868.

LVG070 (*amyE*::*PyloA-GFP*): PY79 was transformed with plasmid pLVG1-1292.

LVG071 (Δ*dnmA*, *amyE*::*PyloA-GFP*): TMN06 was transformed with plasmid pLVG1-1292.

LVG072 (*amyE*::*PzapA-GFP*): PY79 was transformed with Gibson assembled fragment fLVG-2213.

LVG073 (Δ*dnmA*, *amyE*::*PzapA-GFP*): TMN06 was transformed with Gibson assembled fragment fLVG-2213.

LVG074 (*amyE*::*PrnhC-GFP*): PY79 was transformed with Gibson assembled fragment fLVG-2212.

LVG075 (Δ*dnmA*, *amyE*::*PrnhC-GFP*): TMN06 was transformed with Gibson assembled fragment fLVG-2212.

LVG079 (*amyE*::*PcomEA-GFP*): PY79 was transformed with plasmid pLVG1-1995.

LVG080 (Δ*dnmA*, *amyE*::*PcomEA-GFP*): TMN06 was transformed with plasmid pLVG1-1995.

LVG081 (*amyE*::*PezrA-GFP*): PY79 was transformed with plasmid pLVG1-2292.

LVG082 (Δ*dnmA*, *amyE*::*PezrA-GFP*): TMN06 was transformed with plasmid pLVG1-2292.

LVG087 (*amyE*::*PscpA-GFP*): PY79 was transformed with Gibson assembled fragment fLVG-1815.

LVG088 (Δ*dnmA*, *amyE*::*PscpA-GFP*): TMN06 was transformed with Gibson assembled fragment fLVG-1815.

LVG108 (Δ*dnmA operon*, *amyE*::*PscpA-GFP*): TMN17 was transformed with Gibson assembled fragment fLVG-1815.

LVG102 (*amyE*::*PscpA^{mut1}-GFP*): PY79 was transformed with Gibson assembled fragment fLVG-1815mut1.

LVG103 (Δ*dnmA*, *amyE*::*PscpA^{mut1}-GFP*): TMN06 was transformed with Gibson assembled fragment fLVG-1815mut1.

LVG109 (Δ*dnmA operon*, *amyE*::*PscpA^{mut1}-GFP*): TMN17 was transformed with Gibson assembled fragment fLVG-1815mut1.

LVG105 (*amyE*::*Phbs-GFP*): PY79 was transformed with Gibson assembled fragment fLVG-1780.

LVG106 (Δ*dnmA*, *amyE*::*Phbs-GFP*): TMN06 was transformed with Gibson assembled fragment fLVG-1780.

LVG118 (*amyE*::*PscpA^{mut2}-GFP*): PY79 was transformed with Gibson assembled fragment fLVG-1815mut2.

LVG119 (Δ*dnmA operon*, *amyE*::*PscpA^{mut2}-GFP*): TMN17 was transformed with Gibson assembled fragment fLVG-1815mut2.

LVG120 (*amyE*::*PscpA^{mut3}-GFP*): PY79 was transformed with Gibson assembled fragment fLVG-1815mut3.

LVG121(Δ*dnmA operon*, *amyE*::*PscpA^{mut3}-GFP*): TMN17 was transformed with Gibson assembled fragment fLVG-1815mut3.

Plasmid construction

General cloning techniques

All pLVG1-derived plasmids and *amyE*-containing linear fragments were assembled using Gibson assembly (78). Enzymatic assembly of overlapping DNA fragments, or overlap extension PCR. Gibson assemblies consisted of 30-80 ng of each PCR product and 1X Gibson assembly mastermix (0.1 M Tris pH 8.0, 5% PEG-8000, 10 mM MgCl2, 10 mM DTT, 0.2 mM dNTPs, 1 mM NAD+, 4 units/mL T5 exonuclease, 25 units/mL Phusion DNA polymerase, 4,000 units/mL Taq DNA ligase) in a total reaction volume of 10-12 μL. The reactions were incubated at 50°C for 90 minutes. Gibson-assembled plasmids were used to transform *E. coli* MC1061. Gibson-assembled linear fragments were purified using spin columns, re-amplified using Phusion polymerase and used to transform PY79 or PY79 derivatives. For overlap extension PCR, 500 ng of each PCR product was mixed and standard PCR cycling was performed using end primers and Q5 polymerase (NEB). PCR fragments were routinely obtained using Phusion polymerase (NEB) or Q5 polymerase (NEB) and gel-purified before Gibson assembly or overlap extension PCR.

Individual plasmid (p) construction

pJS146: The regions 500 base pairs upstream and downstream of the *ydiOP* operon were amplified from PY79 genomic DNA using primers oJS650 and oJS651 (upstream region) and oJS653 and oJS657 (downstream region). The fragments were then combined with the pminiMAD vector using Gibson assembly.

pTN02: The *dnmA* gene was cloned from PY79 genomic DNA using primers oTN3 and oTN8 with overlapping regions to the pDR110 vector. The pDR110 vector and insert were combined using Gibson assembly.

pTN03: The *dnmA* gene was cloned from PY79 genomic DNA using primers oTN5 and oTN7 with overlapping regions to the pE-SUMO vector. The pE-SUMO vector and insert were combined using Gibson assembly.

pAS2: Overlap PCR was used to make the Y→A substitution. The 5' block was created by using oTMN5 and oAS1 with PY79 genomic DNA as a template. The 3' block was created by using oTMN7 and oAS2 with PY79 genomic DNA as a template. PCR products were gel purified and combined with the pE-SUMO vector using Gibson assembly.

pTN12: The *yabB* gene was cloned from PY79 genomic DNA using primers oTN36 and oTN37 with overlapping regions to the pE-SUMO vector. The pE-SUMO vector and insert were combined using Gibson assembly.

pTN13: The *scoC* gene was cloned from PY79 genomic DNA using primers oTN62 and oTN63 with overlapping regions to the pE-SUMO vector. The pE-SUMO vector and insert were combined using Gibson assembly.

pLVG1: To remove the *lacI* gene from pDR111_GFP(Sp) (79), the plasmid was amplified using primers oLVGLS024A and oLVGLS024B, restricted with BamHI and self-ligated with T4 DNA ligase.

pLVG1-0374: The backbone of pLVG1 without P*xyl* was amplified with primers oLVGLS023A and oLVGLS023B and combined with a DNA fragment containing U0374 (P*rsbV*), amplified from PY79 genomic DNA using primers oLVG_U0374F and oLVG_U0374R.

pLVG1-0868: The backbone of pLVG1 without P*xyl* was amplified with primers oLVGLS023A and oLVGLS023B and combined with a DNA fragment containing U0868 (P*wprA*), amplified from PY79 genomic DNA using primers oLVG_U0868F and oLVG_U0868R.

pLVG1-1292: The backbone of pLVG1 without P*xyl* was amplified with primers oLVGLS023A and oLVGLS023B and combined with a DNA fragment containing U1292

(P*yloA*), amplified from PY79 genomic DNA using primers oLVG_U1292F and oLVG_U1292R.

pLVG1-1995: The backbone of pLVG1 without P*xyl* was amplified with primers oLVGLS023A and oLVGLS023B and combined with a DNA fragment containing U1995 (P*comEA*), amplified from PY79 genomic DNA using primers oLVG_U1995F and oLVG_U1995R.

pLVG1-2292: The backbone of pLVG1 without P*xyl* was amplified with primers oLVGLS023A and oLVGLS023B and combined with a DNA fragment containing U2292 (P*ezrA*), amplified from PY79 genomic DNA using primers oLVG_U2292F and oLVG_U2292R.

Individual DNA fragment (f) construction

fLVG-1780: An upstream DNA fragment was amplified from pLVG1 using primers oLVGLS023C and oLVGLS023A. A downstream DNA was amplified from pLVG1 using primers oLVGLS023B and oLVGLS023D. A DNA fragment containing U1780(P*hbs*) was amplified from PY79 genomic DNA using primers oLVG_U1780F and oLVG_U1780R. The three fragments were assembled using Gibson assembly and the correct construct was enriched using end primers oLVGLS034 and oKJW090.

fLVG-1815: An upstream DNA fragment was amplified from pLVG1 using primers oLVGLS023C and oLVGLS023A. A downstream DNA was amplified from pLVG1 using primers oLVGLS023B and oLVGLS023D. A DNA fragment containing U1815 (P*scpA*) was amplified from PY79 genomic DNA using primers oLVG_U1815F and oLVG_U1815R. The three fragments were assembled using Gibson assembly and the correct construct was enriched using end primers oLVGLS034 and oKJW090.

fLVG-1815mut1: To replace 5'-GACGAG with 5'-GACGTG in the *scpA* promoter, an upstream and downstream DNA fragment was amplified from LVG087 genomic DNA

147

using primer pair oLVGLS042A/oKJW89 and oLVGLS042B/oKJW090, respectively. The fragments were assembled by overlap extension PCR.

<u>fLVG-1815mut2</u>: To replace 5'-GACGAG with 5'-GACGCG in the *scpA* promoter, an upstream and downstream DNA fragment was amplified from LVG087 genomic DNA using primer pair oLVGLS044A/oKJW89 and oLVGLS044B/oKJW090, respectively. The fragments were assembled by overlap extension PCR.

<u>fLVG-1815mut3</u>: To replace 5'-GACGAC with 5'-GACGAC in the *scpA* promoter, an upstream and downstream DNA fragment was amplified from LVG087 genomic DNA using primer pair oLVGLS045A/oKJW89 and oLVGLS045B/oKJW090, respectively. The fragments were assembled by overlap extension PCR.

<u>fLVG-2212</u>: An upstream DNA fragment was amplified from pLVG1 using primers oLVGLS023C and oLVGLS023A. A downstream DNA was amplified from pLVG1 using primers oLVGLS023B and oLVGLS023D. A DNA fragment containing U2212 (P*rnhC*) was amplified from PY79 genomic DNA using primers oLVG_U2212F and oLVG_U2212R. The three fragments were assembled using Gibson assembly and the correct construct was enriched using end primers oLVGLS034 and oLVGLS090.

<u>fLVG-2213</u>: An upstream DNA fragment was amplified from pLVG1 using primers oLVGLS023C and oLVGLS023A. A downstream DNA was amplified from pLVG1 using primers oLVGLS023B and oLVGLS023D. A DNA fragment containing U2213 (P*zapA*) was amplified from PY79 genomic DNA using primers oLVG_U2213F and oLVG_U2213R. The three fragments were assembled using Gibson assembly and the correct construct was enriched using end primers oLVGLS034 and oKWJ90.

References

1. Chen, K., Zhao, B.S. and He, C. (2016) Nucleic Acid Modifications in Regulation of Gene Expression. *Cell Chem Biol*, **23**, 74-85.
2. Jin, Z. and Liu, Y. (2018) DNA methylation in human diseases. *Genes Dis*, **5**, 1-8.
3. Jones, P.A. (2012) Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet*, **13**, 484-492.
4. Fu, Y., Luo, G.Z., Chen, K., Deng, X., Yu, M., Han, D., Hao, Z., Liu, J., Lu, X., Dore, L.C. *et al.* (2015) N6-methyldeoxyadenosine marks active transcription start sites in Chlamydomonas. *Cell*, **161**, 879-892.
5. Zhang, G., Huang, H., Liu, D., Cheng, Y., Liu, X., Zhang, W., Yin, R., Zhang, D., Zhang, P., Liu, J. *et al.* (2015) N6-methyladenine DNA modification in Drosophila. *Cell*, **161**, 893-906.
6. Greer, E.L., Blanco, M.A., Gu, L., Sendinc, E., Liu, J., Aristizabal-Corrales, D., Hsu, C.H., Aravind, L., He, C. and Shi, Y. (2015) DNA Methylation on N6-Adenine in C. elegans. *Cell*, **161**, 868-878.
7. Mondo, S.J., Dannebaum, R.O., Kuo, R.C., Louie, K.B., Bewick, A.J., LaButti, K., Haridas, S., Kuo, A., Salamov, A., Ahrendt, S.R. *et al.* (2017) Widespread adenine N6-methylation of active genes in fungi. *Nat Genet*, **49**, 964-968.
8. Blow, M.J., Clark, T.A., Daum, C.G., Deutschbauer, A.M., Fomenkov, A., Fries, R., Froula, J., Kang, D.D., Malmstrom, R.R., Morgan, R.D. *et al.* (2016) The Epigenomic Landscape of Prokaryotes. *PLoS Genet*, **12**, e1005854.
9. Jeltsch, A. (2002) Beyond Watson and Crick: DNA methylation and molecular enzymology of DNA methyltransferases. *Chembiochem*, **3**, 274-293.
10. Loenen, W.A., Dryden, D.T., Raleigh, E.A. and Wilson, G.G. (2014) Type I restriction enzymes and their relatives. *Nucleic Acids Res*, **42**, 20-44.
11. Loenen, W.A., Dryden, D.T., Raleigh, E.A., Wilson, G.G. and Murray, N.E. (2014) Highlights of the DNA cutters: a short history of the restriction enzymes. *Nucleic Acids Res*, **42**, 3-19.
12. Arber, W. and Dussoix, D. (1962) Host specificity of DNA produced by Escherichia coli. I. Host controlled modification of bacteriophage lambda. *J Mol Biol*, **5**, 18-36.
13. Gold, M., Hurwitz, J. and Anders, M. (1963) The Enzymatic Methylation of Rna and DNA, Ii. On the Species Specificity of the Methylation Enzymes. *Proc Natl Acad Sci U S A*, **50**, 164-169.
14. Cheng, X. (1995) DNA modification by methyltransferases. *Curr Opin Struct Biol*, **5**, 4-10.
15. Seshasayee, A.S., Singh, P. and Krishna, S. (2012) Context-dependent conservation of DNA methyltransferases in bacteria. *Nucleic Acids Res*, **40**, 7066-7073.
16. Gonzalez, D., Kozdon, J.B., McAdams, H.H., Shapiro, L. and Collier, J. (2014) The functions of DNA methylation by CcrM in Caulobacter crescentus: a global approach. *Nucleic Acids Res*, **42**, 3720-3735.
17. Han, J.S., Kang, S., Kim, S.H., Ko, M.J. and Hwang, D.S. (2004) Binding of SeqA protein to hemi-methylated GATC sequences enhances their interaction and aggregation properties. *J Biol Chem*, **279**, 30236-30243.

18. Nievera, C., Torgue, J.J., Grimwade, J.E. and Leonard, A.C. (2006) SeqA blocking of DnaA-oriC interactions ensures staged assembly of the E. coli pre-RC. *Mol Cell*, **24**, 581-592.
19. Lahue, R.S., Au, K.G. and Modrich, P. (1989) DNA mismatch correction in a defined system. *Science*, **245**, 160-164.
20. Bale, A., d'Alarcao, M. and Marinus, M.G. (1979) Characterization of DNA adenine methylation mutants of Escherichia coli K12. *Mutat Res*, **59**, 157-165.
21. Pukkila, P.J., Peterson, J., Herman, G., Modrich, P. and Meselson, M. (1983) Effects of high levels of DNA adenine methylation on methyl-directed mismatch repair in Escherichia coli. *Genetics*, **104**, 571-582.
22. Casadesus, J. and Low, D.A. (2013) Programmed heterogeneity: epigenetic mechanisms in bacteria. *J Biol Chem*, **288**, 13929-13935.
23. Atack, J.M., Yang, Y., Seib, K.L., Zhou, Y. and Jennings, M.P. (2018) A survey of Type III restriction-modification systems reveals numerous, novel epigenetic regulators controlling phase-variable regulons; phasevarions. *Nucleic Acids Res*, **46**, 3532-3542.
24. Hernday, A.D., Braaten, B.A. and Low, D.A. (2003) The mechanism by which DNA adenine methylase and PapI activate the pap epigenetic switch. *Mol Cell*, **12**, 947-957.
25. Seib, K.L., Jen, F.E., Tan, A., Scott, A.L., Kumar, R., Power, P.M., Chen, L.T., Wu, H.J., Wang, A.H., Hill, D.M. *et al.* (2015) Specificity of the ModA11, ModA12 and ModD1 epigenetic regulator N(6)-adenine DNA methyltransferases of Neisseria meningitidis. *Nucleic Acids Res*, **43**, 4150-4162.
26. Nye, T.M., Jacob, K.M., Holley, E.K., Nevarez, J.M., Dawid, S., Simmons, L.A. and Watson, M.E., Jr. (2019) DNA methylation from a Type I restriction modification system influences gene expression and virulence in Streptococcus pyogenes. *PLoS Pathog*, **15**, e1007841.
27. Wojciechowski, M., Czapinska, H. and Bochtler, M. (2013) CpG underrepresentation and the bacterial CpG-specific DNA methyltransferase M.MpeI. *Proc Natl Acad Sci U S A*, **110**, 105-110.
28. Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P., Rank, D., Baybayan, P., Bettman, B. *et al.* (2009) Real-time DNA sequencing from single polymerase molecules. *Science*, **323**, 133-138.
29. Flusberg, B.A., Webster, D.R., Lee, J.H., Travers, K.J., Olivares, E.C., Clark, T.A., Korlach, J. and Turner, S.W. (2010) Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat Methods*, **7**, 461-465.
30. Wion, D. and Casadesus, J. (2006) N6-methyl-adenine: an epigenetic signal for DNA-protein interactions. *Nat Rev Microbiol*, **4**, 183-192.
31. Bolker, M. and Kahmann, R. (1989) The Escherichia coli regulatory protein OxyR discriminates between methylated and unmethylated states of the phage Mu mom promoter. *EMBO J*, **8**, 2403-2410.
32. Nou, X., Skinner, B., Braaten, B., Blyn, L., Hirsch, D. and Low, D. (1993) Regulation of pyelonephritis-associated pili phase-variation in Escherichia coli: binding of the PapI and the Lrp regulatory proteins is controlled by DNA methylation. *Mol Microbiol*, **7**, 545-553.

33.   Ogawa, T. and Okazaki, T. (1994) Cell cycle-dependent transcription from the gid and mioC promoters of Escherichia coli. *J Bacteriol*, **176**, 1609-1615.

34.   Lobner-Olesen, A., Marinus, M.G. and Hansen, F.G. (2003) Role of SeqA and Dam in Escherichia coli gene expression: a global/microarray analysis. *Proc Natl Acad Sci U S A*, **100**, 4672-4677.

35.   Lu, M., Campbell, J.L., Boye, E. and Kleckner, N. (1994) SeqA: a negative modulator of replication initiation in E. coli. *Cell*, **77**, 413-426.

36.   Marinus, M.G. and Morris, N.R. (1975) Pleiotropic effects of a DNA adenine methylation mutation (dam-3) in Escherichia coli K12. *Mutat Res*, **28**, 15-26.

37.   Stephens, C., Reisenauer, A., Wright, R. and Shapiro, L. (1996) A cell cycle-regulated bacterial DNA methyltransferase is essential for viability. *Proc Natl Acad Sci U S A*, **93**, 1210-1214.

38.   Koo, B.M., Kritikos, G., Farelli, J.D., Todor, H., Tong, K., Kimsey, H., Wapinski, I., Galardini, M., Cabal, A., Peters, J.M. *et al.* (2017) Construction and Analysis of Two Genome-Scale Deletion Libraries for Bacillus subtilis. *Cell Syst*, **4**, 291-305 e297.

39.   Hardwood, C.R. and Cutting, S.M. (1990) *Molecular Biological Methods for Bacillus*. John Wiley & Sons, Chichester.

40.   Nye, T.M., Schroeder, J.W., Kearns, D.B. and Simmons, L.A. (2017) Complete Genome Sequence of Undomesticated Bacillus subtilis Strain NCIB 3610. *Genome Announc*, **5**, pii: e00364-00317.

41.   Schroeder, J.W. and Simmons, L.A. (2013) Complete Genome Sequence of Bacillus subtilis Strain PY79. *Genome Announcements*, **1**, pii: e01085-01013.

42.   Sobetzko, P., Jelonek, L., Strickert, M., Han, W., Goesmann, A. and Waldminghaus, T. (2016) DistAMo: A Web-Based Tool to Characterize DNA-Motif Distribution on Bacterial Chromosomes. *Front Microbiol*, **7**, 283.

43.   Zwietering, M.H., Jongenburger, I., Rombouts, F.M. and van 't Riet, K. (1990) Modeling of the bacterial growth curve. *Appl Environ Microbiol*, **56**, 1875-1881.

44.   Dreiseikelmann, B. and Wackernagel, W. (1981) Absence in Bacillus subtilis and Staphylococcus aureus of the sequence-specific deoxyribonucleic acid methylation that is conferred in Escherichia coli K-12 by the dam and dcm enzymes. *J Bacteriol*, **147**, 259-261.

45.   Guha, S. and Guschlbauer, W. (1992) Expression of Escherichia coli dam gene in Bacillus subtilis provokes DNA damage response: N6-methyladenine is removed by two repair pathways. *Nucleic Acids Res*, **20**, 3607-3615.

46.   Kearns, D.B., Chu, F., Rudner, R. and Losick, R. (2004) Genes governing swarming in Bacillus subtilis and evidence for a phase variation mechanism controlling surface motility. *Mol Microbiol*, **52**, 357-369.

47.   Ohshima, H., Matsuoka, S., Asai, K. and Sadaie, Y. (2002) Molecular organization of intrinsic restriction and modification genes BsuM of Bacillus subtilis Marburg. *J Bacteriol*, **184**, 381-389.

48.   Schluckebier, G., O'Gara, M., Saenger, W. and Cheng, X. (1995) Universal catalytic domain structure of AdoMet-dependent methyltransferases. *J Mol Biol*, **247**, 16-20.

49. Morgan, R.D., Dwinell, E.A., Bhatia, T.K., Lang, E.M. and Luyten, Y.A. (2009) The MmeI family: type II restriction-modification enzymes that employ single-strand modification for host protection. *Nucleic Acids Res*, **37**, 5208-5221.

50. Sanchez-Romero, M.A., Cota, I. and Casadesus, J. (2015) DNA methylation in bacteria: from the methyl group to the methylome. *Curr Opin Microbiol*, **25**, 9-16.

51. Nicolas, P., Mader, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Bidnenko, E., Marchadier, E., Hoebeke, M., Aymerich, S. *et al.* (2012) Condition-dependent transcriptome reveals high-level regulatory architecture in Bacillus subtilis. *Science*, **335**, 1103-1106.

52. Voelker, U., Luo, T., Smirnova, N. and Haldenwang, W. (1997) Stress activation of Bacillus subtilis sigma B can occur in the absence of the sigma B negative regulator RsbX. *J Bacteriol*, **179**, 1980-1984.

53. Gueiros-Filho, F.J. and Losick, R. (2002) A widely conserved bacterial cell division protein that promotes assembly of the tubulin-like protein FtsZ. *Genes Dev*, **16**, 2544-2556.

54. Seo, D., Kamino, K., Inoue, K. and Sakurai, H. (2004) Purification and characterization of ferredoxin-NADP+ reductase encoded by Bacillus subtilis yumC. *Arch Microbiol*, **182**, 80-89.

55. Mascarenhas, J., Soppa, J., Strunnikov, A.V. and Graumann, P.L. (2002) Cell cycle-dependent localization of two novel prokaryotic chromosome segregation and condensation proteins in Bacillus subtilis that interact with SMC protein. *Embo J*, **21**, 3108-3118.

56. Wang, X., Le, T.B., Lajoie, B.R., Dekker, J., Laub, M.T. and Rudner, D.Z. (2015) Condensin promotes the juxtaposition of DNA flanking its loading site in Bacillus subtilis. *Genes Dev*, **29**, 1661-1675.

57. Wang, X., Montero Llopis, P. and Rudner, D.Z. (2014) Bacillus subtilis chromosome organization oscillates between two distinct patterns. *Proc Natl Acad Sci U S A*, **111**, 12877-12882.

58. Wang, X., Tang, O.W., Riley, E.P. and Rudner, D.Z. (2014) The SMC condensin complex is required for origin segregation in Bacillus subtilis. *Curr Biol*, **24**, 287-292.

59. Carabetta, V.J., Greco, T.M., Cristea, I.M. and Dubnau, D. (2019) YfmK is an N(epsilon)-lysine acetyltransferase that directly acetylates the histone-like protein HBsu in Bacillus subtilis. *Proc Natl Acad Sci U S A*, **116**, 3752-3757.

60. Kohler, P. and Marahiel, M.A. (1998) Mutational analysis of the nucleoid-associated protein HBsu of Bacillus subtilis. *Mol Gen Genet*, **260**, 487-491.

61. Lalanne, J.B., Taggart, J.C., Guo, M.S., Herzel, L., Schieler, A. and Li, G.W. (2018) Evolutionary Convergence of Pathway-Specific Enzyme Expression Stoichiometry. *Cell*, **173**, 749-761 e738.

62. Randall, J.R., Hirst, W.G. and Simmons, L.A. (2017) Substrate specificity for bacterial RNase HII and HIII is influenced by metal availability. *J Bacteriol*, **200**, pii: e00401-00417.

63. Randall, J.R., Nye, T.M., Wozniak, K.J. and Simmons, L.A. (2019) RNase HIII Is Important for Okazaki Fragment Processing in Bacillus subtilis. *J Bacteriol*, **201**, pii: e00686-00618.

64. Stolz, R., Sulthana, S., Hartono, S.R., Malig, M., Benham, C.J. and Chedin, F. (2019) Interplay between DNA sequence and negative superhelicity drives R-loop structures. *Proc Natl Acad Sci U S A*. 116(13):6260-6269.

65. Belitsky, B.R., Barbieri, G., Albertini, A.M., Ferrari, E., Strauch, M.A. and Sonenshein, A.L. (2015) Interactive regulation by the Bacillus subtilis global regulators CodY and ScoC. *Mol Microbiol*, **97**, 698-716.

66. Caldwell, R., Sapolsky, R., Weyler, W., Maile, R.R., Causey, S.C. and Ferrari, E. (2001) Correlation between Bacillus subtilis scoC phenotype and gene expression determined using microarrays for transcriptome analysis. *J Bacteriol*, **183**, 7329-7340.

67. Marinus, M.G. and Lobner-Olesen, A. (2014) DNA Methylation. *EcoSal Plus*, **6**, doi: 10.1128/ecosalplus.ESP-0003-2013.

68. Kozdon, J.B., Melfi, M.D., Luong, K., Clark, T.A., Boitano, M., Wang, S., Zhou, B., Gonzalez, D., Collier, J., Turner, S.W. *et al.* (2013) Global methylation state at base-pair resolution of the Caulobacter genome throughout the cell cycle. *Proc Natl Acad Sci U S A*, **110**, E4658-4667.

69. Casadesus, J. and Low, D. (2006) Epigenetic gene regulation in the bacterial world. *Microbiol Mol Biol Rev*, **70**, 830-856.

70. Waldron, D.E., Owen, P. and Dorman, C.J. (2002) Competitive interaction of the OxyR DNA-binding protein and the Dam methylase at the antigen 43 gene regulatory region in Escherichia coli. *Mol Microbiol*, **44**, 509-520.

71. Haakonsen, D.L., Yuan, A.H. and Laub, M.T. (2015) The bacterial cell cycle regulator GcrA is a sigma70 cofactor that drives gene expression from a subset of methylated promoters. *Genes Dev*, **29**, 2272-2286.

72. Jarmer, H., Larsen, T.S., Krogh, A., Saxild, H.H., Brunak, S. and Knudsen, S. (2001) Sigma A recognition sites in the Bacillus subtilis genome. *Microbiology*, **147**, 2417-2424.

73. Diekmann, S. (1987) DNA methylation can enhance or induce DNA curvature. *EMBO J*, **6**, 4213-4217.

74. Camacho, E.M., Serna, A., Madrid, C., Marques, S., Fernandez, R., de la Cruz, F., Juarez, A. and Casadesus, J. (2005) Regulation of finP transcription by DNA adenine methylation in the virulence plasmid of Salmonella enterica. *J Bacteriol*, **187**, 5691-5699.

75. Bolz, N.J., Lenhart, J.S., Weindorf, S.C. and Simmons, L.A. (2012) Residues in the N-terminal domain of MutL required for mismatch repair in *Bacillus subtilis*. *Journal of Bacteriology*, **194**, 5361-5367.

76. Hall, B.M., Ma, C.X., Liang, P. and Singh, K.K. (2009) Fluctuation analysis CalculatOR: a web tool for the determination of mutation rate using Luria-Delbruck fluctuation analysis. *Bioinformatics*, **25**, 1564-1565.

77. Lenhart, J.S., Brandes, E.R., Schroeder, J.W., Sorenson, R.J., Showalter, H.D. and Simmons, L.A. (2014) RecO and RecR are necessary for RecA loading in response to DNA damage and replication fork stress. *J Bacteriol*, **196**, 2851-2860.

78. Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison, C.A., 3rd and Smith, H.O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods*, **6**, 343-345.

79. Overkamp, W., Beilharz, K., Detert Oude Weme, R., Solopova, A., Karsens, H., Kovacs, A., Kok, J., Kuipers, O.P. and Veening, J.W. (2013) Benchmarking various green fluorescent protein variants in *Bacillus subtilis*, *Streptococcus pneumoniae*, and *Lactococcus lactis* for live cell imaging. *Appl Environ Microbiol*, **79**, 6481-6490.

80. Lee, P.S., Lin, D.C., Moriya, S. and Grossman, A.D. (2003) Effects of the chromosome partitioning protein Spo0J (ParB) on *oriC* positioning and replication initiation in *Bacillus subtilis*. *J Bacteriol*, **185**, 1326-1337.

81. Dupes, N.M., Walsh, B.W., Klocko, A.D., Lenhart, J.S., Peterson, H.L., Gessert, D.A., Pavlick, C.E. and Simmons, L.A. (2010) Mutations in the *Bacillus subtilis* beta clamp that separate its roles in DNA replication from mismatch repair. *J Bacteriol*, **192**, 3452-3463.

82. Noirot-Gros, M.F., Velten, M., Yoshimura, M., McGovern, S., Morimoto, T., Ehrlich, S.D., Ogasawara, N., Polard, P. and Noirot, P. (2006) Functional dissection of YabA, a negative regulator of DNA replication initiation in *Bacillus subtilis*. *Proc Natl Acad Sci U S A*, **103**, 2368-2373.

83. Marinus, M.G. and Morris, N.R. (1974) Biological function for 6-methyladenine residues in the DNA of Escherichia coli K12. *J Mol Biol*, **85**, 309-322.

84. Herman, G.E. and Modrich, P. (1981) *Escherichia coli* K-12 clones that overproduce dam methylase are hypermutable. *J Bacteriol*, **145**, 644-646.

85. Cooper, L.A., Simmons, L.A. and Mobley, H.L. (2012) Involvement of Mismatch Repair in the Reciprocal Control of Motility and Adherence of Uropathogenic *Escherichia coli*. *Infect Immun*, **80**, 1969-1979.

86. Lenhart, J.S., Sharma, A., Hingorani, M.M. and Simmons, L.A. (2013) DnaN clamp zones provide a platform for spatiotemporal coupling of mismatch detection to DNA replication. *Molecular microbiology*, **87**, 553-568.

87. Lenhart, J.S., Pillon, M.C., Guarne, A., Biteen, J.S. and Simmons, L.A. (2015) Mismatch repair in Gram-positive bacteria. *Res Microbiol*. 167(1):4-12.

88. Lenhart, J.S., Schroeder, J.W., Walsh, B.W. and Simmons, L.A. (2012) DNA Repair and Genome Maintenance in *Bacillus subtilis*. *Microbiology and molecular biology reviews : MMBR*, **76**, 530-564.

89. Yeeles, J.T., Gwynn, E.J., Webb, M.R. and Dillingham, M.S. (2011) The AddAB helicase-nuclease catalyses rapid and processive DNA unwinding using a single Superfamily 1A motor domain. *Nucleic Acids Res*, **39**, 2271-2285.

90. Youngman, P., Perkins, J.B. and Losick, R. (1984) Construction of a cloning site near one end of Tn917 into which foreign DNA may be inserted without affecting transposition in *Bacillus subtilis* or expression of the transposon-borne erm gene. *Plasmid*, **12**, 1-9.

91. Konkol, M.A., Blair, K.M. and Kearns, D.B. (2013) Plasmid-encoded ComI inhibits competence in the ancestral 3610 strain of *Bacillus subtilis*. *J Bacteriol*, **195**, 4085-4093.

92. Smith, B.T., Grossman, A.D. and Walker, G.C. (2001) Visualization of mismatch repair in bacterial cells. *Mol. Cell*, **8**, 1197-120

CHAPTER IV

RnhP is a Plasmid-borne RNase HI that Contributes to Genome Maintenance in the Ancestral Strain *Bacillus subtilis* NCIB 3610

Abstract

RNA-DNA hybrids form throughout the chromosome during normal cell growth and under stress conditions. When left unresolved, RNA-DNA hybrids can slow replication fork progression, cause DNA breaks, increase mutagenesis, and reduce gene expression. To remove hybrids, all organisms use ribonuclease H (RNase H) to specifically degrade the RNA portion. Here we show that, in addition to chromosomally encoded RNase HII and RNase HIII, *Bacillus subtilis* NCIB 3610 encodes a previously uncharacterized RNase HI protein, RnhP, on the endogenous plasmid pBS32. Like other RNase HI enzymes, RnhP incises Okazaki fragments, ribopatches, and a complementary RNA-DNA hybrid. We show that while chromosomally encoded RNase HIII is required for pBS32 hyper-replication, RnhP compensates for loss of RNase HIII activity on the chromosome. Consequently, loss of RnhP and RNase HIII impairs bacterial growth. We show that the decreased growth rate can be explained by laggard replication fork progression near the terminus region of the right replichore, resulting in SOS-dependent inhibition of cell division. We conclude that *B. subtilis* NCIB 3610 encodes functional RNase HI, HII, and HIII, and pBS32 encoded RNase HI contributes to replication fork progression and chromosome stability while RNase HIII is important for chromosome stability and plasmid hyper-replication.

Introduction

For all organisms, faithful replication of the chromosome is essential to ensure daughter cells receive an accurate and complete copy of their genetic material. Over the last decade there has been a growing recognition that RNA is often incorporated into genomic DNA, either through hybridization with or covalent linkage to DNA (1-3). These incorporation events can have severe consequences for cell physiology, leading to replication fork stress, genome instability, and adverse effects on transcription (4-8). RNA-DNA hybrids form through a variety of processes throughout each phase of bacterial growth, with each type of hybrid impacting genome integrity in a different way.

In exponentially growing cells a common type of RNA-DNA hybrid occurs in the form of Okazaki fragments (9, 10). During DNA replication, Okazaki fragments on the lagging strand begin with RNA primers generating an RNA-DNA hybrid with a covalent RNA-DNA junction (11, 12). These RNA primers are later removed and replaced with DNA through the activity of several DNA repair proteins. A second type of RNA-DNA hybrid occurs during replication by DNA polymerase error, where an rNTP is used in place of the cognate dNTP, resulting in a sugar error (4, 6, 14). Sugar errors tend to be single replicative DNA polymerase errors and have the potential to occur every few thousand base pairs in exponentially growing cells (6, 14). In states of slow growth, it has been proposed that translesion DNA polymerases could use rNTPs in place of scarce dNTPs in a process termed ribopatch repair (15). Ribopatch repair would generate relatively short polymers of RNA nested in double stranded DNA to provide a temporary solution for sites in need of repair (15). RNA polymers covalently joined to DNA can impact genome integrity because the 2'OH of the ribose sugar can facilitate a nucleophilic attack on the 3' PO_4^-, resulting in a 2', 3' cyclic phosphate at the rNMP and a 5' OH at the adjacent nucleotide (16). The resulting nick in the phosphodiester backbone is refractory to ligation and unable to function in further DNA synthesis (4, 17, 18). Failure to heal the end and repair the nick would then result in a double strand break during the next round of DNA replication.

RNA-DNA hybrids in the form of Okazaki fragments and DNA polymerase incorporation events are similar in that the RNA is covalently linked to DNA through a phosphodiester bond (9, 10). Another prevalent RNA-DNA hybrid forms during

156

transcription when mRNA transcripts are base-paired with the transcribed DNA strand, displacing the coding strand as ssDNA to form an R-loop [for review (1)]. In the case of R-loops, RNA hybridized to DNA lacks a covalent RNA-DNA junction. Persistent R-loops can impair progression of replication forks and DNA synthesis while also decreasing gene expression from the DNA template subsequent to R-loop formation (7, 19, 20). Transcription is required during all growth phases, suggesting that R-loop formation could be prevalent during the entire life cycle of a bacterium. All cells need to resolve each class of RNA-DNA hybrid that occurs *in vivo* to maintain genome integrity and efficient gene expression throughout bacterial growth.

To reduce the detrimental consequences that RNA-DNA hybrids impose on genome integrity and transcription, organisms have enzymes dedicated to hybrid resolution (21, 22). The RNase H family of endoribonucleases binds to and cleaves the RNA component of RNA-DNA hybrids, resolving all classes of hybrids that occur *in vivo* (21, 22). RNase H enzymes are highly conserved, with family members present in bacteria, archaea, eukaryotes and retroviruses, including HIV-1 (23-26). Bacterial RNase H enzymes are grouped into two general types based on amino acid sequence similarity: type I, which includes RNase HI, and type II, which includes RNase HII and HIII (27). RNase HI and RNase HIII enzymes act on both ribopatches (four or more embedded rNMPs) and hybrids lacking a covalent RNA-DNA junction, but are unable to cleave at a single rNMP embedded in DNA (9, 10, 28). Unlike RNase HI and HIII, RNase HII enzymes are adept for cleavage at single embedded rNMPs and ribopatches participating in ribonucleotide excision repair (RER), yet show very poor activity on hybrids that lack a covalent RNA-DNA junction (24, 29). All three enzymes are active on the RNA primer portions of an Okazaki fragment, suggesting that all three bacterial enzymes could have overlapping functions during lagging strand processing and maturation (9, 10, 30). In addition to their important contribution to chromosomal replication, evidence suggests that RNase H enzymes function in the regulation of endogenous plasmid replication (31) and in regulatory aspects of transcription (3, 7).

Comparative sequence analysis of over 300 genomes found that 80% of bacterial genomes contain RNase HI and RNase HII (27). Approximately 17% of bacterial genomes, including the Firmicutes phylum, which includes a group of

important Gram-positive pathogens from Streptococcus, Staphylococcus, and Enterococcus species, lack RNase HI and instead encode RNase HII and RNase HIII (27). Importantly, Firmicutes were the only group with some representatives that appeared to encode all three RNase H genes (9, 27). One Firmicute that seemed to encode all three RNase H enzymes is the soil bacterium *Bacillus subtilis* (9, 27). The RNase HII and HIII enzymes from the lab strain *B. subtilis* PY79 are active and have been characterized *in vitro* and *in vivo* (4, 9, 10, 30). Functional studies of the putative RNase HI-like genes from *B. subtilis* have shown that these genes lack the residues involved in substrate binding and do not possess nuclease activity *in vitro* (9, 27). Furthermore, prior work also showed simultaneous deletion of both RNase HII (*rnhB*) and RNase HIII (*rnhC*) is lethal (32) or results in a mutator phenotype with accumulation of compensatory mutations (10), suggesting that the putative RNase HI-like genes are unable to compensate for loss of both RNase HII and RNase HIII *in vivo.* Of the small subset of bacteria that contain putative RNase HI, HII, and HIII proteins there is no experimental evidence to support the coexistence of *functional* RNase HI and RNase HIII in the same genome (27).

B. subtilis NCIB 3610 (referred herein as 3610) is considered a "wild" ancestral strain that has maintained many of the wild motility and social behaviors associated with *B. subtilis* strains isolated from the soil (33-36). In addition to the 4.2 Mbp chromosome, 3610 contains an endogenous 84 Kbp plasmid, pBS32 (37). Plasmid pBS32 encodes 102 genes, many of which include a large contiguous set of genes that appear to encode for a cryptic prophage (38). Other genes on the plasmid control host cell physiology, such as the inhibitor of biofilm formation RapP, the inhibitor of natural competence for DNA uptake ComI, and the cell death promoting sigma factor SigN (38, 39). The remaining genes on pBS32 are of unknown function, including *zpdC* (*rnhP*), which encodes a putative RNase HI. If *zpdC* encodes a functional RNase HI enzyme this, to the best of our knowledge, would suggest that *B. subtilis* 3610 is the first reported bacterium to encode active RNase HI, HII, and HIII enzymes. Moreover, if ZpdC is an active RNase H it is unknown whether ZpdC activity is important for pBS32 plasmid maintenance, integrity of the *B. subtilis* chromosome, or some other DNA maintenance function.

Here we show that ZpdC (named here as RnhP) is capable of cleaving all substrates typical of RNase HI proteins. Deletion of *rnhP* does not affect pBS32 maintenance, although deletion of *rnhC* results in loss of the pBS32 hyper-replication phenotype. We demonstrate that deletion of *rnhP* and *rnhC* results in a 2-fold increase in doubling time, which is attributed to cell filamentation and induction of the SOS response. Together, our data show that both the plasmid encoded RNase HI (RnhP) and the chromosomally encoded RNase HIII (RnhC) are important for genome maintenance in the ancestral strain of *B. subtilis* NCIB 3610, demonstrating that bacteria can indeed maintain all three RNase H proteins with each enzyme contributing to genome stability.

Results

RnhP is an active RNase HI enzyme. The endogenous 84 Kbp plasmid, pBS32, of the ancestral strain *B. subtilis* NCIB 3610 contains several uncharacterized genes that encode proteins with sequence homology to bacterial DNA replication and repair proteins (38, 47, 48). One such gene, *zpdC* (*rnhP*), shares 38.5% primary structure identity and 50.3% primary structure similarity to the RNase HI protein from *Escherichia coli* **(Fig 4.1A)***. Importantly, all of the catalytic residues involved in metal coordination are conserved between the two sequences as are residues within the α-helix 3 basic protrusion handle, which is involved in substrate binding (49), suggesting that ZpdC might have RNase H activity **(Fig 4.1A)**. As part of our ongoing effort to determine how RNA-DNA hybrids impact genome stability and transcription, we began by purifying ZpdC (RnhP) and a variant with D73N, which has been shown to render *E. coli* RNase HI catalytically inactive (49) **(Fig 4.1B)**.

To assay for RNase H activity, we ordered an oligonucleotide labeled with an IR dye at the 5' end that contained four embedded rNMPs flanked by DNA on either side. This labeled oligonucleotide was annealed to a complementary DNA strand, creating a double stranded RNA-DNA chimeric substrate as previously described (10). We used this substrate because prior work showed that bacterial RNase HI, HII, and HIII are all active on this substrate (9, 10, 30). We incubated ZpdC (RnhP) and the catalytically inactive variant with this substrate for 10 minutes at 37°C in buffers that mimic *in vivo*

relevant metal concentrations (1 mM MgCl$_2$, 10 µM MnCl$_2$,) as described (10). The substrate was also exposed to alkaline hydrolysis in a separate reaction to create a ladder corresponding to the positions of each embedded rNMP. The products of the reaction were separated by electrophoresis on denaturing urea PAG to measure substrate cleavage. Incubation of the substrate with low (4 nM) and high (50 nM) concentrations of protein results in complete cleavage of the substrate, whereas the catalytically inactive variant did not show any cleavage at either concentration **(Fig 4.1C)**.

      To determine if ZpdC (RnhP) has strict RNase H activity, such that it is only capable of cleaving the RNA portion of RNA/DNA hybrids, we tested the ability of ZpdC (RnhP) to cleave double stranded RNA and DNA substrates. We incubated 4 nM and 50 nM ZpdC (RnhP) with an RNA oligo labeled at the 5' end with an IR dye hybridized to a complementary RNA oligo and a DNA oligo labeled at the 3' end with an IR dye hybridized to a complementary DNA oligo under the same buffer and incubation conditions described above for the RNA-DNA chimera substrate. Unlike the RNA-DNA chimera substrate, we did not observe any cleavage of the RNA or DNA substrates when incubated with low or high concentrations of ZpdC **(Fig 4.1D and E)**. From these data we conclude that ZpdC is an active and strict RNase H enzyme. Having established that ZpdC is a plasmid encoded RNase H, we rename *zpdC* to <u>RN</u>ase <u>H</u> from pBS32 (*rnhP*).

RnhP is active with various metals and at varying temperatures. To determine the parameters of activity for RnhP we assayed for cleavage over a range of temperatures and metal concentrations relevant to *B. subtilis* growth (50). RnhP cleaved the RNA-DNA chimeric substrate with four embedded rNMPs when incubated at 25°C, 30°C, and 37°C for 10 minutes with no appreciable difference in activity observed between the three temperatures **(Fig 4.8A).** We also tested the activity of RnhP on the four embedded rNMP substrate with various metals, holding all other reaction and buffer conditions the same between samples (see Materials and Methods). RnhP appeared to be most active when incubated with 10 µM Mn^{2+} as a metal cofactor **(Fig 4.8B)**. RnhP also showed activity when incubated with Mg^{2+}, Zn^{2+}, and Co^{2+}, although RnhP was less

active when compared with activity in the presence of Mn^{2+}. We note a reduction in RnhP activity when incubated with 1 mM $MgCl_2$ and 10 µM $MnCl_2$, compared to the 10 µM $MnCl_2$ alone condition, which could be explained by competition of Mn^{2+} with Mg^{2+} for binding to the RnhP active site **(Fig 4.8B)**. We conclude that RnhP is active both under relevant growth temperatures and with various metal cofactors. We note that, compared to the other cofactors tested here, Mn^{2+} supports the most activity.

RnhP cleaves RNA-DNA covalent chimeras in a different location when compared with *B. subtilis* RNase HII and RNase HIII. Most bacteria encode a functional RNase HII enzyme and either RNase HI or RNase HIII (27). Further, it has been hypothesized that RNase HI and HIII are mutually exclusive (27). RNase HII is classically characterized as having unique activity on a single embedded rNMP within DNA and polymers of rNMPs that are covalently linked to DNA including embedded ribonucleotide polymers (i.e. "ribopatches") and Okazaki fragments (51). RNase HI and HIII recognize substrates that contain four or more embedded rNMPs that are covalently linked to DNA and RNA-DNA hybrids that interact through hydrogen bonding, such as R-loops (9, 10, 51). To empirically determine the cleavage patterns of RNase H enzymes, we purified RnhP, RnhC (RNase HIII), and RnhB (RNase HII) to examine their activities and cleavage patterns on a variety of RNA-DNA hybrid substrates *in vitro*. The purification of RnhC and RnhB has already been described (4, 9, 10).

We began by testing the activities of all three enzymes using a substrate labeled with an IR dye at the 5' end that contained one rNMP flanked by DNA on both sides annealed to a complementary DNA strand (oJR209 and oJR145). Consistent with previously published results (4), we found that only RnhB (RNase HII) had activity on this substrate under the conditions tested here (see Materials and Methods) **(Fig 4.2A)**. We next assayed for activity on the four embedded rNMP substrate. While all three enzymes were capable of incising the substrate at 4 nM and 50 nM protein concentrations, we found that the enzymes differed in their incision patterns. RnhB cleavage yielded a longer product, indicating cleavage between the third and fourth ribonucleotide from the 5' IR-dye end label as expected, whereas RnhC cleavage resulted in a shorter product, with cleavage within the middle of the embedded RNA. In

contrast to functional redundancy with RnhC, RnhP appeared to cleave the four embedded rNMP substrate more similarly to RnhB (RNase HII), resulting in a longer product than RnhC with cleavage between the third/fourth and second/third rNMPs from the 5' IR-dye end label **(Fig 4.2B)**.

Both the single and quadruple embedded rNMP substrates are intended to represent misincorporation events that can occur when DNA polymerases erroneously add rNTPs during replication or ribopatch repair, accounting for as many as 2,000 rNMPs incorporated into the *E. coli* genome per round of replication (14). Significantly more rNMPs (~23,000) are expected to be incorporated into the genome in the form of Okazaki fragments during lagging strand synthesis (1). In *B. subtilis*, these primers are removed and replaced with dNMPs through the combined action of RNase HIII, DNA polymerase I, and YpcP (10). To test how RnhP activity compared to RnhC (RNase HIII) and RnhB (RNase HII) on an Okazaki fragment substrate, we constructed an oligo with rNMPs at the 5' end covalently linked to a stretch of DNA with a 3' IR-dye end label. This oligo was hybridized to another oligo that was complementary to the 5' end of the molecule but was significantly longer to generate a 3' overhang (oJR339 and oJR340). We incubated this substrate, as previously described (9, 10), with RnhB, RnhC, and RnhP to measure activity. Consistent with previous work, we observe substrate cleavage for both RnhB and RnhC (9, 10). Furthermore, we show that RnhP has activity on the Okazaki fragment substrate, with multiple cleavage sites and some sites of incision overlapping with that of RnhC **(Fig 4.2C)**.

RnhP cleaves RNA-DNA hybrids differently than RNase HIII. A defining feature of the RNA-DNA hybrids tested thus far is that each substrate contains a covalent RNA-DNA junction. The single rNMP substrate, the polymer of four embedded rNMPs, and the Okazaki fragment-like substrate each have an RNA-DNA covalent linkage. In contrast, an R-loop represents a different type of RNA-DNA hybrid, which is produced during transcription when RNA hybridizes with complementary DNA in the template strand, displacing the DNA coding strand (3, 52). This substrate differs in that it does not contain a covalent RNA-DNA linkage. To determine if RnhP, like other RNase HI enzymes, is capable of cleaving substrates without a covalent RNA-DNA linkage we

began by testing activity on a substrate labeled at the 5' end with an IR dye that contains an all RNA strand hybridized to a complementary DNA strand (oJR227 and oJR145). Consistent with previously published results, we observed cleavage when the complementary RNA-DNA hybridized substrate was incubated with RNase HIII but not RNase HII (9, 10). We then tested RnhP and demonstrate that RnhP does indeed have activity on an RNA-DNA hybridized substrate lacking an RNA-DNA covalent junction. However, the site of incision differs between RnhP and RNase HIII (RnhC) **(Fig 4.2D)**. This result, along with the results described above, shows that RnhP recognizes the same substrates as canonical RNase HI (51) and RnhP often cleaves at a different location than RNase HIII.

RNase HIII is required for plasmid hyper-replication. We initially hypothesized that RnhP would be important for replication of pBS32 in 3610 simply based on the notion that *rnhP* is plasmid-borne. It was previously reported that overexpressing the plasmid-specific sigma factor, SigN (ZpdN), causes pBS32 to hyper-replicate and promote cell lysis (39). Plasmid copy number was measured by quantitative PCR during normal growth and following hyper-replication by inducing SigN from an IPTG inducible promoter. We found that pBS32 was maintained at a low copy number similar to WT in Δ*rnhC*, Δ*rnhP*, and the double mutant (Δ*rnhC*, Δ*rnhP*) when expression of SigN was not induced **(Fig 4.3A)**. Induction of SigN caused the plasmid to hyper-replicate in the WT and Δ*rnhP* cells, but not the Δ*rnhC* cells **(Fig 4.3A)**. Further, cell viability was assessed in all strains induced with SigN by measuring optical density and counting CFUs every 30 minutes up to four hours post-induction. Induction of SigN in the WT and Δ*rnhP* strains caused a similar loss of cell viability, suggesting that RnhP is not required for the pBS32-mediated cell death phenotype. In contrast, the strain with an *rnhC* deletion showed a slight drop and plateau in OD, while CFU counts were reduced less drastically than the WT or Δ*rnhP* mutant but recovered much slower. In the double mutant, OD reached a plateau while CFU counts dropped and did not recover over a four-hour time course experiment **(Fig 4.3BC)**. Moreover, we found that the double mutant cells displayed a severely filamentous phenotype throughout the course of the experiment **(Fig 4.9)**. With these data we conclude that chromosomally encoded RNase

HIII is important for plasmid hyper-replication and recovery while loss of *rnhP* alone has no effect on pBS32 maintenance or hyper-replication.

Cells lacking both RnhP and RnhC activity have a reduced growth rate. Our results thus far demonstrate that RnhP has activity on RNA-DNA hybrids with four or more ribonucleotides *in vitro* and that RnhP does not contribute to pBS32 maintenance or hyper-replication. Therefore, we asked if RnhP contributes to chromosome maintenance in 3610. If so, it would suggest that 3610 has a fitness advantage when maintaining active RNase HI (RnhP), HII, and HIII enzymes for the purpose of resolving the variety of RNA-DNA hybrids that form on the chromosome. It has been shown that in the absence of RNase HIII, but not RNase HII, there is a decrease in growth rate in *B. subtilis* PY79 (4, 9, 10). To test whether RnhP activity could compensate for the decrease in growth observed in the absence of RNase HIII, we performed growth curves for 3610 in LB media for the WT, Δ*rnhC*, Δ*rnhP*, and Δ*rnhP rnhC::erm* strains. While the doubling times for the Δ*rnhC* and Δ*rnhP* single deletion strains (57.8 and 46.8 min, respectively) appeared to be slower than WT (37.7 min), there was no statistically significant difference in growth rate between WT, Δ*rnhCl,* or Δ*rnhP* strains based on the growth model used here **(Fig 4.4A,** see Materials and Methods**)**. However, upon loss of both *rnhP* and *rnhC* the growth rate was significantly slower than WT (37.7 min) or either of the single deletion strains, with a doubling time over two times greater than WT at 94 minutes **(Fig 4.4A)**. As described in greater detail later in the results, we show that ectopic expression of *rnhP* in a Δ*rnhC* background of the lab strain PY79 rescues Δ*rnhC* growth defects to WT levels **(Fig 4.7B**, described below**)**. With these results we conclude that RnhP can compensate for RNase HIII (*rnhC*) and that these enzymes have overlapping functions in 3610.

Cells lacking *rnhP* and *rnhC* genes filament relative to WT. To test if the differences in growth rate were caused by an inhibition of cell division we assayed cell length in exponentially growing cultures of WT, Δ*rnhC*, Δ*rnhP*, and Δ*rnhP rnhC::erm* strains. Cell membranes were imaged with a lipophilic fluorescent dye and cell length was measured as described (Materials and Methods). Consistent with the slight decrease in growth

rate observed in both Δ*rnhC* and Δ*rnhP* single deletions at 30°C, the average cell lengths of ~4.4 µm and ~4.0 µm for each strain respectively was longer than that of the WT strain measuring at ~3.7 µm **(Fig 4.4C-E)**. A slight tail on the distribution of cell lengths can be observed for each single deletion strain, representing a subpopulation of cells that are slow to complete cell division, resulting in a portion of longer cells **(Fig 4.4G)**. The distribution of cell lengths for the Δ*rnhP rnhC::erm* double mutant has a more pronounced tail and an average cell length greater than WT or the single deletions alone at ~5.6 µm **(Fig 4.4D, G)**. These results support the model that single deletions of *rnhC* or *rnhP* are well tolerated by the cell and we suggest that one gene can compensate for loss of the other. However, the double deletion results in a severe growth defect that is, at least in part, the product of improper cell division, suggesting genome integrity is compromised in the double mutant during normal growth in the absence of exogenous stress.

Cells lacking *rnhP* and *rnhC* activity are induced for the SOS response. During periods of DNA damage cell division is inhibited by the cell division inhibitor, YneA, to allow for the chromosome to be properly replicated before cell division resumes (53-55). The YneA-enforced DNA damage checkpoint ensures that daughter cells receive a complete copy of the chromosome after replication is complete (53-55). Given the defects in growth we observe, and the cell filamentation of the double deletion strain, we asked if cells lacking *rnhP* and *rnhC* are induced for the SOS response. We used the SOS reporter construct *tagC::tagC-gfp*, which like *yneA* is highly up regulated during the SOS response, as a single cell proxy for SOS induction (56). In exponentially growing WT cells (OD_{600} = 0.5-0.7) at 30°C in LB media, we found that ~5.0% of cells expressed the SOS reporter while ~88.0% of WT cells expressed the reporter when the DNA damage response was induced following addition of mitomycin C (MMC) **(Fig 4.5A,C, E-F)**. In contrast to the WT cells, ~71.2% of the double deletion cells expressed the SOS reporter under normal growth conditions, which increased to ~91.3% upon treatment with MMC **(Fig 4.5B,D, E-F)**. Therefore, we show that cells lacking *rnhP* and *rnhC* experience a ~14-fold increase in SOS induction during normal growth conditions, which explains the slow growth and cell elongation results described above. These

results further show that 3610 is able to mitigate the deleterious effects of RNA-DNA hybrids when either RnhP or RNase HIII (*rnhC*) is present. When both genes are nonfunctional, the consequences to genome integrity cause most cells to induce the SOS response delaying cell division and impairing growth.

Cells lacking *rnhP* and *rnhC* exhibit replication stress near the terminus region.
Having established that loss of *rnhP* and *rnhC* results in SOS induction for most cells during normal growth, we investigated the genome-wide replication status of the WT and double deletion strains in exponential phase cultures. We isolated DNA from each strain in triplicate for Illumina DNA-sequencing to determine chromosome and plasmid replication status. The resulting reads were aligned to the NCIB 3610 reference chromosome and plasmid separately and the average coverage was plotted over the length of the reference. There was little to no difference in sequencing coverage between the WT and double deletion strain over the length of pBS32 (**Fig 4.6A**). We found a severe drop in sequencing coverage around 60 Kb for pBS32 in the double deletion strain, which corresponds to the location of the deleted *rnhP* gene (62,030 – 62,497). When visualizing the sequencing coverage map for the chromosome and comparing the WT and double deletion strains, we noticed an abrupt drop in sequencing reads in the terminus region for the right replicore. This result shows that replication fork progression is slowed in this region for the double deletion strain **(Fig 4.6B, 4.10)**.

Replication interference at the terminus region in the double deletion strain could be the result of accumulated R-loops in this region, which cannot be resolved in the double deletion due to the lack of RnhC and RnhP. As previously discussed, R-loops can impair replication fork progression (19), which could explain the observed drop in sequencing coverage. Moreover, the impaired replication forks in the double deletion strain also explain the induction of the SOS response that we observe with the *tagC-gfp* reporter. Taken together, we suggest that in the absence of *rnhP* and *rnhC* replication forks become laggard in the terminus after encountering R-loops that persist in the double mutant causing SOS induction, cell elongation, and a decrease in growth rate.

Cells lacking both *rnhP* and *rnhC* show increased sensitivity to cellular stress.

Having established that defects in both plasmid-encoded *rnhP* and chromosomally encoded *rnhC* genes results in a decreased growth rate and inhibition of cell division, we asked how cells respond to various stressors in the absence of one or both of these RNase H enzymes. In *B. subtilis* PY79, Δ*rnhC* cells are sensitive to a myriad of cellular stresses, including cold shock, osmotic stress, and treatment with genotoxic agents (9, 10, 19). To test how the RNase H genes contribute to genotoxic stress responses, we tested the susceptibility of the single deletions, Δ*rnhB*, Δ*rnhC,* and Δ*rnhP*, as well as pairwise deletion strains, Δ*rnhC rnhB::erm,* Δ*rnhP rnhB::erm,* and Δ*rnhP rnhC::erm*, in 3610 to various stressors. We began by testing sensitivity to cold stress (growth at 25°C), which has been hypothesized to contribute to the stability of Okazaki fragments, and growth on sublethal concentrations of hydroxyurea (HU), which has been hypothesized to inhibit ribonucleotide reductase in *B. subtilis* resulting in increased rNTP:dNTP pools in the cell (10, 57). We found that in contrast to WT *B. subtilis* PY79, deletion of any of the *rnhC* gene resulted in a modest (<10 fold) sensitivity relative to WT to cold shock or growth in the presence of HU. For cold shock and HU treatment conditions, the Δ*rnhP rnhC::erm* double mutant displayed ~100 and ~1,000 fold increases in sensitivity relative to WT cells, respectively **(Fig 4.7A)**.

Given that RnhP can compensate for loss of RnhC activity in the ancestral strain NCIB 3610, we asked if expression of *rnhP* could rescue the cold and HU sensitivities observed in the Δ*rnhC* strain for *B. subtilis* PY79, which lacks pBS32 and *rnhP*. We created a strain that expresses *rnhP* from an IPTG inducible promoter from an ectopic chromosomal locus in the Δ*rnhC* background for PY79 and tested susceptibility to cold and HU stress. In support of our results from 3610, we found that ectopic expression of *rnhP* in a Δ*rnhC* background completely restored cold and HU sensitivities to WT survival, with >100- and >1,000-fold growth relative to the Δ*rnhC* strain for cold stress and HU, respectively **(Fig 4.7B)**. With these results we conclude that RnhP activity can compensate for loss of RNase HIII when challenged with cold stress or HU challenge demonstrating overlapping functions of *rnhC* and *rnhP* genes in *B. subtilis*.

Discussion

RNase H enzymes are biologically universal and required for cleavage of the RNA moiety in an RNA-DNA hybrid (21, 22). RNase H genes are present in the genomes of bacteria, archaea, eukaryotes, and retroviruses (23-26). Eukaryotes show less diversity in the RNase H genes they encode. Almost all eukaryotes contain RNase HI and RNase HII (21, 22). Plants, including Arabidopsis, contain multiple RNase HI homologs because different RNase HIs are targeted to the nucleus, mitochondria, and chloroplast (58). In the genomes of prokaryotes, RNase H enzymes show striking diversity between organisms (21, 22). Phylogenetic studies show that all prokaryotic genomes analyzed contain at least one RNase H with most genomes containing two RNase H genes (27). In general, most bacteria contain RNase HI and HII, while a smaller subset contains RNase HII and RNase HIII (27). As RNase HI and HIII are active on the same class of substrates and because a prokaryotic genome had not been identified to encode functional RNase HI and RNase HIII, it had been proposed that RNase HI and HIII are mutually exclusive (27). We show that *rnhP* (RNase HI) and *rnhC* (RNase HIII) contribute to genome maintenance in 3610 demonstrating that 3610 contains functional RNase HI, RNase HII, and RNase HIII enzymes. RNase HIII is chromosomally encoded while *rnhP* is plasmid encoded. We therefore suggest that *rnhP* was acquired through horizontal gene transfer and has resided on the nonessential plasmid pBS32. Our experiments *in vivo* show that 3610 grows well with an Δ*rnhC* allele, however 3610 grows poorly and experiences constitutive SOS induction when Δ*rnhP* and Δ*rnhC* are deficient, indicating that either RNase HI or RNase HIII are important for normal growth and resolution of RNA-DNA hybrids that form *in vivo*. Therefore, although *rnhP* is plasmid-borne, RNase HI activity from this gene product is important for genome maintenance in 3610 and, to our knowledge, this is the first organism described where functional RNase HI and RNase HIII have been shown to coexist.

As discussed above, we initially hypothesized that RnhP would be required for pBS32 maintenance or hyper-replication. In contrast, we found that RNase HIII (*rnhC*) was required for plasmid hyper-replication while neither *rnhP* nor *rnhC* were important for normal plasmid maintenance. We found that while the *rnhP* deletion alone does not confer a phenotype, the Δ*rnhC* does confer slight growth interference to DNA damage

or from cold stress and hydroxyurea (HU), suggesting that RNase HIII is the more important enzyme *in vivo*. The double deletion of Δ*rnhP rnhC::erm* shows ~100-fold and ~1,000 fold growth interference from cold stress and HU treatment, respectively. If we compare the results of Δ*rnhP rnhC::erm* on HU for 3610 to the phenotype for Δ*rnhC* from *B. subtilis* strain PY79 we find the same extent of growth interference. Therefore, the comparison of phenotypes between 3610 and PY79 shows that Δ*rnhP rnhC::erm* in 3610 largely phenocopies the single Δ*rnhC* deletion for PY79 on HU and for cold sensitivity. Finally, we show that the PY79 Δ*rnhC* phenotype is rescued with ectopic expression of *rnhP*, further demonstrating functional overlap between RnhP and RNase HIII (*rnhC*) in *B. subtilis*.

Biochemical characterization shows that RnhP is an RNase H with specificity for substrates with four or more embedded ribonucleotides. RnhP is not active on a dsDNA or dsRNA substrate. Further, RnhP does not cleave a substrate with a single ribonucleotide nested in duplex DNA. Therefore, biochemical characterization of RnhP shows that it is a strict RNase H with preference for Mn^{2+}. Our prior work characterizing RNase HIII showed that this enzyme was most active with Mg^{2+} on the canonical substrates for RNase HIII (9, 10). One simple explanation for the coexistence of RNase HIII and RnhP is that metals could be scarce for wild Bacillus during growth in the soil. One possibility is that RNase HIII is most active when magnesium concentrations are sufficient to support activity. During conditions when magnesium concentrations are lower and manganese concentrations are sufficient, then RnhP could be more active providing RNase H activity and a fitness advantage for *B. subtilis* cells encoding both *rnhC* and *rnhP*. Given our studies with the double mutant, all experiments point to a model where the growth of 3610 is well supported with either RNase HIII or RnhP. It is the double deletion that grows poorly and is constitutively induced for the DNA damage response. Therefore, we suggest that RNase HIII activity predominates and RnhP activity can be used to supplement RNase HIII during specific growth conditions or when the burden of RNA-DNA hybrid resolution overwhelms the capacity of RNase HIII.

Prior phylogenetic work shows that only a small subset of bacteria in the phylum Firmicutes, including *B. subtilis* and Lactobacillus, contain genes for all three RNase H

proteins (27). Sequence comparisons showed that the predicted RNase HI genes in organisms with RNase HIII lack the catalytic residues and the substrate binding α-helix 3 basic protrusion handle found in active RNase HI enzymes (27). Moreover, prior functional studies of the chromosomally encoded and predicted RNase HI genes from *B. subtilis*, including YpdQ and YpeP, were unable to detect RNase H activity, further supporting the argument that RNase HI and RNase HIII activities do not coexist (9). One possible limitation of prior phylogenetic studies would be if this work only interrogated core genomes. Further, our finding that RnhP has a different metal preference and cleavage site selection relative to RNase HIII could also provide a biochemical difference that allows for these genes to coexist while both contribute to genome maintenance. We speculate that functional RNase HI and RNase HIII are unlikely to coexist as chromosomally encoded genes. We wish to speculate that other bacteria will be identified to have RNase HI and RNase HIII coexist with one gene encoded as part of the accessory genome and the other as part of the core genome. This would allow for acquisition, transfer, and loss of one RNase H gene and maintenance of both when a fitness advantage is conferred. As more genome sequences become available, it will be interesting to learn how many other bacteria encode functional RNase HI and RNase HIII and how these genes contribute to growth and genome integrity.

Materials and Methods

General Bacteriology: Unless otherwise specified, the antibiotic concentrations used in this study are as follows: 0.5 µg/mL erythromycin, 100 µg/mL spectinomycin, 5 mM hydroxyurea, 20 ng/µL mitomycin C, and 100 µg/mL methyl methanesulfonate. Strains were grown in Lysogeny Broth (LB) at 30°C.

RNase H alignments: Global alignments were performed on the GenBank protein sequences for ZpdC (AGQ21310.1) from *Bacillus subtilis* NCIB 3610 and RnhA (NP_414750) from *Escherichia coli* MG1655 using the pairwise sequence alignment tool from Clustal Omega (https://www.ebi.ac.uk/Tools/psa/).

Spot plates: Designated strains were streaked from frozen stocks and grown overnight at 30°C on LB agar plates. Plates were washed in LB liquid media and used to inoculate 2 mL cultures to an OD_{600} of 0.05. Cultures were grown in a 30°C rolling rack to an OD_{600} 0.9-1.5. Cultures were diluted to OD_{600} of 0.5 in phosphate buffered saline (PBS) pH 7.4 followed by 10-fold serial dilutions in 1x PBS. The dilution series was then spotted onto LB plates plus the indicated antibiotic or incubated at the indicated temperatures.

Growth rate analysis: Designated strains were streaked onto LB agar plates from frozen stocks and grown overnight at 30°C. Plates were washed in LB liquid media and inoculated at an OD_{600} of 0.05 into 25 mL of pre-warmed LB liquid media. The cultures were grown in a shaking water bath at 200 RPM at 30°C. The OD_{600} measurement for each culture was recorded every 30 minutes. Biological replicates were performed in triplicate on three separate days for each strain and the average growth measurement with corresponding standard errors were plotted. A modified Gompertz growth model in the form $y = A exp\{-exp\left[\frac{\mu_m \times e}{A}\ (\lambda - t) + 1\right]\}$ was fit to the replicates for each strain to obtain estimated growth rates (40). The parameters A, μ_m, and λ represent the time (t) when the growth rate equals zero (asymptote), the maximum growth rate, and the lag time, respectively (40). The estimated growth rate (μ_m) from the Gompertz model was then used to calculate doubling time estimates as $ln(2)/\mu_m$ for each strain (10, 40).

Genomic DNA purifications: Designated strains were streaked from frozen stocks onto LB agar plates and grown overnight at 30°C. Plates were washed in LB liquid media and used to inoculate 10 mL of LB liquid media at an OD_{600} of 0.05. The strains were grown in triplicate over three separate days prior to harvesting chromosomal DNA. At an OD_{600} of 0.5-0.7, the cells were pelleted via centrifugation, washed in 1 mL of re-suspension buffer (50 mM Tris-HCl pH 8 with 5% glycerol) and mixed in a final volume of 150 μL of re-suspension buffer. For cell lysis, Triton 100 was added to 1% (v/v), 10 μL of 10 mg/mL RNase A, and lysozyme from the MasterPure™ Gram-positive DNA purification kit (Lucigen) were added and used as described. Subsequent lysis and purification steps were performed as described in the MasterPure™ Gram-positive DNA

purification kit (Lucigen) protocol per the manufacturer's instructions with the exception of the RNase treatment step, which was omitted because RNase treatment was performed during cell lysis.

DNA sequencing and chromosome coverage analysis: Library preparation and DNA sequencing was performed by the University of Michigan DNA Sequencing Core. Sequencing reads were aligned using bwa (v 0.7.8-r455) to the NCIB 3610 chromosome reference (CP020102.1) and pBS32 (CP020103.1) reference (37, 41). The bam files were sorted and filtered using samtools (v 0.1.18) for quality values greater than 30 (42) and PCR duplicates were removed using Picard tools (https://github.com/broadinstitute/picard). The filtered bam files were used to calculate the genome coverage at each base using genomeCoverageBed from bedtools (v 2.29.1). The coverage at each base was averaged for the three replicates. The average coverage over 10kb windows was plotted every 1kb throughout the length of the chromosome using the packages ggplot2 and zoo in R (v 3.1.3).

RnhP (D73N). To create a catalytically inactive RnhP variant we mutated the aspartic acid residue (GAT) responsible for metal ion coordination at position 73 to asparagine (AAT) using overlapping PCR. The 5' blocks were created using *B. subtilis* NCIB 3610 genomic DNA as a template with either primer oTN56 or oTN58 and oTN61 for the pDR110$_{rnhPD73N}$ and pE-SUMO$_{rnhPD73N}$ vectors, respectively. Similarly, the 3' blocks were created using either primer oTN57 or oTN59 with oTN60 for the pDR110$_{rnhPD73N}$ and pE-SUMO$_{rnhPD73N}$ vectors, respectively. A PCR cleanup was performed and the purified products were combined using Gibson assembly (43) to create pDR110$_{rnhPD73N}$ and pE-SUMO$_{rnhPD73N}$. pDR110$_{rnhPD73N}$ and pE-SUMO$_{rnhPD73N}$. Each plasmid generated was subsequently used to transform competent MC1061 cells and plated on 100 µg/mL spectinomycin or 25 µg/mL kanamycin. Resulting colonies were screened by PCR using primers oTN56 and oTN57 for the pDR110$_{rnhPD73N}$ vector and oTN58 and oTN59 for the pE-SUMO$_{rnhPD73N}$ vector and the insert sequences were verified as correct using Sanger sequencing through the University of Michigan Core sequencing facility.

172

Protein Purification: Recombinant proteins were purified from *E. coli BL21$_{DE3}$* cells containing pE-SUMO$_{rnhP}$ and pE-SUMO$_{rnhPD73N}$ as described (4, 10). Briefly, Cultures were grown in 2 liters of LB with 25 µg/mL kanamycin at 37°C shaking to an OD of 0.7. Overexpression was induced by adding IPTG to 0.5 mM followed by growth for 3 additional hours. Cells were then pelleted by centrifugation and stored at -80°C. Once thawed, the pellet was resuspended in lysis buffer [50 mM Tris-HCl pH 8, 300 mM NaCl, 10% sucrose, 10 mM imidazole, 1x protease inhibitors (Roche 11873580001)] and cells were sonicated at 10 seconds on 20 seconds off cycles for a total of 5 minutes on ice. Cell debris was cleared and pelleted by centrifugation. Supernatant was then applied to a 4 mL Ni^{2+}-NTA agarose gravity-flow column. The column was washed with wash buffer (50 mM Tris-HCl pH 8, 25 mM imidazole, 2 M NaCl, 5% glycerol) and eluted with elution buffer (50 mM Tris-HCl pH 8, 400 mM imidazole, 150 mM NaCl, 5% glycerol). Following elution, 1 mM DTT and SUMO protease were added to the eluate and incubated for 2 hours at room temperature. The SUMO protease treated sample was dialyzed into storage buffer (50 mM Tris-HCl pH 8, 150 mM NaCl, 5% glycerol) overnight at 4°C. The product was fractionated by application to a 4 mL Ni^{2+}-NTA gravity-flow column to separate the recombinant protein from the SUMO tag. SDS-PAGE was performed to confirm the SUMO tag was removed. The sample was then dialyzed into cation exchange start buffer (20 mM Tris-HCl pH 7.5, 5% glycerol, 1 mM DTT) overnight at 4°C. The dialyzed sample was purified using a HiTrap SP HP column (GE: 17-1152-01) with an elution gradient of 50-500 mM NaCl at a flow rate of 1 mL/min over 90 minutes. SDS-PAGE was performed and fractions containing pure protein were pooled. The RnhP (D73N) protein was concentrated, glycerol was added to 25%, aliquoted, flash frozen in liquid nitrogen, and stored at -80 °C. The RnhP protein eluted slightly earlier from the S-column and required further purification with size exclusion chromatography. The concentrated protein was applied to a HiPrep 16/60 Sephacryl S200 HR column (GE: 17-1166-01) at a flow rate of 0.6 mL/min with sizing column buffer (20 mM Tris-HCl pH 7.5, 200 mM NaCl, 1mM DTT) and eluted in one peak. SDS-PAGE was again performed and fractions containing only pure protein were pooled, concentrated, glycerol was added to 25%, aliquoted, flash frozen, and stored at -80°C.

RNase H activity assays: The end infrared (IR) dye-labeled substrates for the 1-rNMP, 4-rNMP, and all RNA substrates were created by mixing oJR209, oJR210, and oJR227 respectively, with oJR145 in a 1:2 µM ratio diluted in Buffer A (20 mM Tris-HCl pH8, 50 nM NaCl, 1 mM DTT) (9, 10). The Okazaki fragment substrate was assembled by mixing oJR339 with oJR340 in a 1:2 µM ratio in Buffer A. The RNA/RNA and DNA/DNA hybrids were created by mixing oJR227/oJR166 and oJR348/oJR365, respectively, in 1:2 µM ratios in Buffer A. The oligos were annealed by heating at 98°C for 1 min followed by cooling on the bench top to room temperature. Reactions totaling 10 µL in volume included 100 nM substrate and 4 or 50 nM protein as indicated (diluted from stock concentrations in Buffer A) in the *in vivo* metal concentration buffer (20 mM Tris-HCl pH8, 50 nM NaCl, 1 mM $MgCl_2$, 10 µM $MnCl_2$, and 1 mM DTT) (9, 10). For NaOH treated samples, 200 mM of NaOH was added to 500 nM substrate. Reactions were allowed to proceed for 10 minutes at 37°C unless otherwise indicated. For all reactions except the RNA/RNA hybrid, 10 µL of stop buffer (95% formamide, 20 mM EDTA, 0.01% bromophenol blue) was added after 10 minutes and reactions were placed at 98°C for 5 minutes and subsequently snap-cooled on ice. A denaturing 8M urea 20% polyacrylamide gel was prepared by pre-electrophoresing the gel at 250V for 30 minutes in TBE buffer. The gel was subsequently loaded with 4 µL of each reaction and electrophoresed at 250V for 1.5 hours. For the RNA/RNA hybrid, 10 µL of RNA hybrid stop buffer (66% formamide, 9% formaldehyde, 17.5 mM EDTA, and 0.65x MOPS Buffer (10x MOPS buffer: 200 mM MOPS, 50 mM sodium acetate, 10 mM EDTA) was added after 10 minutes and the reactions were placed at 55°C for 15 minutes. A denaturing 8M urea 20% polyacrylamide gel was prepared by pre-electrophoresing the gel at 100V for 30 minutes in 0.5x MOPS buffer. The gel was subsequently loaded with 4 µL of each reaction and electrophoresed at 150V for 2 hours. For all gels, the products were visualized with a LI-COR Odyssey imager.

Plasmid growth analysis following induction: Strains grown overnight in LB at 22°C were subcultured into 50 mL fresh LB to an OD_{600} of 0.1 and cultured at 37°C. OD was measured every 30 minutes until OD reached between 0.07-0.12. IPTG was added to a final concentration of 1 mM, and OD was measured every 30 minutes for a total of 4

hours post-induction. Simultaneously, 100 µL of multiple 10-fold serial dilutions of each sample culture was plated on LB plates containing spectinomycin and incubated overnight at 37°C. The following day, plates containing individual colony forming units (CFUs) were counted to determine CFU/mL.

Plasmid copy number following induction: Strains grown overnight in LB at 22°C were subcultured into 50 mL fresh LB to an OD_{600} of 0.1 and cultured at 37°C until OD reached between 0.07-0.12. IPTG was added to a final concentration of 1 mM, and strains continued to grow for an additional 60 minutes. Four OD units of each sample was pelleted, and genomic and plasmid DNA was isolated from cells by Qiagen DNeasy Blood & Tissue Kit (Cat #69504). The concentration of isolated DNA was quantified by Nanodrop, samples were standardized to a DNA concentration of 10 ng/µl, and diluted 10- and 100-fold to 1 ng/µl and 0.1 ng/µl, respectively. Quantitative PCR was performed with all three dilutions to determine plasmid copy number as previously described (44). Improved determination of plasmid copy number using quantitative real-time PCR for monitoring fermentation processes). Primers 3106/3107 (*sigA*) were used to measure chromosomal DNA, and primers 6527/6528 (*zpdE*) were used to measure pBS32 DNA.

Live cell microscopy: Each strain imaged was streaked from frozen stocks onto LB agar plates and grown at 30°C for 16 hours. Plates were then washed with LB and 25 mL cultures were inoculated to an initial OD_{600} of approximately 0.05. Cultures were then placed in a water bath at 30°C with shaking at 212 RPM until reaching an OD_{600} of 0.6 – 0.9. Once the desired OD_{600} was reached, the cultures were filtered to concentrate the cells, and subsequently washed three times with 1x PBS. Cells were then pelleted via centrifugation and resuspended in 200 µL 1x PBS. The membrane was stained using 0.25 ng/µL of FM4-64. 200 µL of cells were then placed onto a microscope slide with a 1% agarose pad as described (45). After each slide was prepared, the slides were placed under an Olympus BX61microscope equipped with a Hamamatsu camera (46). The microscope was focused under an exposure of 20-30 ms. Then the microscope was switched from DIC to RFP setting in order to observe the membrane

stain of FM4-64 with an exposure of 300 ms. Once the microscope was properly focused, an image was recorded. After a combined total of approximately 900 cells were imaged for each strain, images were adjusted for brightness, contrast and gamma using the CellSense software (Olympus). The length of the cells was measured using the polyline tool of CellSense (Olympus). In order for a cell to be considered scorable, the cell membrane had to clearly imaged from pole to pole. For cells undergoing division incomplete septa were scored as one cell and complete septa were scored as two cells.

A

```
Eco-HI  MLKQVEIFTDGSCLGN---PGPGGYGAILRYRGREKTFSAGYTRTTNNRM    47
Bs-ZpdC -MKKVVIYCDGAARNNGKDNNVGGFGAVLRYGDHVKTIKAGFRNVTNNMM    49

Eco-HI  ELMAAIVALEALK-EHCEVILSTDSQYVRQGITQ-WIHNWKKRGWKTADK    95
Bs-ZpdC EIRAAIEALKQLKTTNIPVEINTDSAYLCNCMNOGWYKKWMNNGWVTAGK    99

Eco-HI  KPVKNVDLWQRLDAALGQHQ-IKWEWVKGHAGHPE--------NERCDEL   136
Bs-ZpdC KPVENRQLWIELIELVEQFPFITENKVKGHSGIPDNELADRLANEAMDEL   149

Eco-HI  ARAAAMNPTLEDTGYQVEV     155
Bs-ZpdC TRGAAV------------     155
```
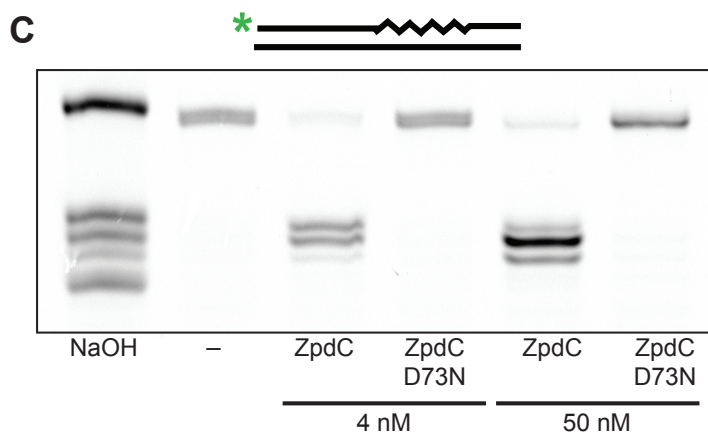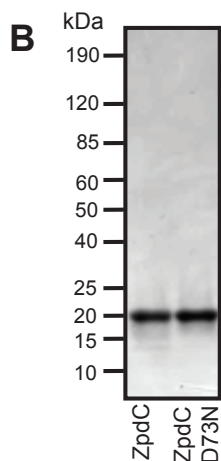
B

C

D

E

Figure 4.1. Plasmid encoded ZpdC is an active RNase HI protein. (A) Sequence alignment of ZpdC with *E. coli* RNase HI. Identical and similar residues are indicated in black and gray, respectively. Red indicates conserved catalytic residues. (*) denotes

177

catalytic residue mutated in catalytically inactive variant (D73N). The α-helix 3 basic protrusion handle is boxed. **(B)** SDS-PAGE stained with Coomassie brilliant Blue of purified ZpdC and catalytically inactive variant D73N. **(C)** ZpdC and D73N were incubated with a ribopatch substrate. The 5' end IR-labeled oligo containing four embedded rNMPs (squiggly lines) within an otherwise DNA oligo (straight lines) was annealed to a complementary DNA oligo (oJR210 and oJR145). A ladder was generated via alkaline hydrolysis of the substrate at the embedded rNMPs (lane one). **(D)** Incubation of ZpdC with an RNA-RNA substrate. A 5' end IR-labeled RNA oligo (squiggly line) was annealed to a complementary RNA oligo (oJR227 and oJR166). **(E)** Incubation of ZpdC with a DNA-DNA substrate. A 3' end IR-labeled DNA oligo (straight line) was annealed to a complementary DNA oligo (oJR348 and oJR365). For C-E, the reactions were assembled as described in "Materials and Methods" and products were separated on a 20% denaturing urea-PAGE and subsequently visualized with a LI-COR Odyssey imager.
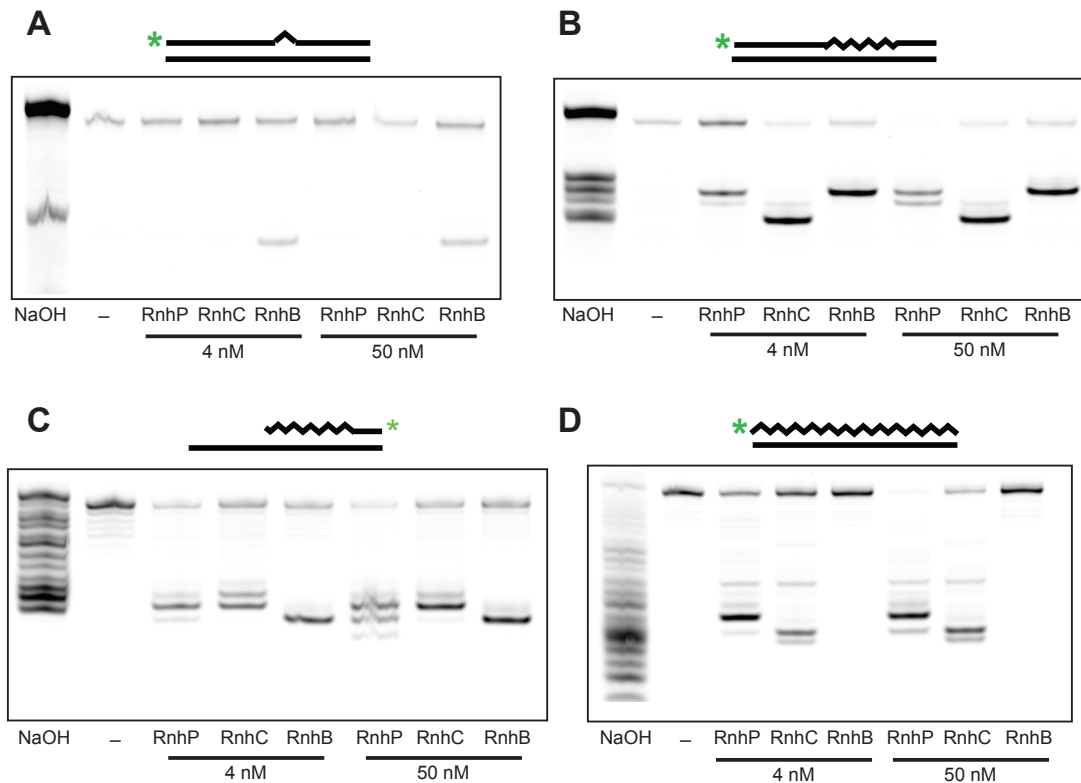
Figure 4.2. RnhP cleaves several different RNA-DNA hybrid substrates. For each reaction, the indicated substrate was incubated separately with RnhP, RnhB, or RnhC in reaction buffer for 10 minutes at 37°C (see Materials and Methods). For each substrate, a ladder was created via alkaline hydrolysis of the substrate at the rNMPs (lane one). The products were separated on a 20% denaturing urea-PAGE and subsequently visualized with a LI-COR Odyssey imager. **(A)** Incubation of RnhP, RnhB, and RnhC with a single rNMP substrate. A 5' end IR-labeled oligo containing one embedded rNMP (triangle) within an otherwise DNA oligo (straight lines) was annealed to a complementary DNA oligo (oJR209 and oJR145). **(B)** Incubation of RnhP, RnhB, and RnhC with a ribopatch substrate. A 5' end IR-labeled oligo containing four embedded rNMPs (squiggly lines) within an otherwise DNA oligo (straight lines) was annealed to a complementary DNA oligo (oJR210 and oJR145). **(C)** Incubation of RnhP, RnhB, and RnhC with an Okazaki fragment-like substrate. A 3' IR-dye end labeled oligo with rNMPs at the 5' end covalently linked to a stretch of DNA was hybridized to an oligo that was complementary at the 5' end of the molecule but was significantly longer to generate a 3' overhang (oJR339 and oJR340). **(D)** Incubation of RnhP, RnhB, and RnhC with a complementary RNA-DNA hybrid substrate. A 5' end IR-labeled RNA oligo (squiggly line) was annealed to a complementary DNA oligo (straight lines) to create an RNA-DNA hybrid (oJR227 and oJR145).
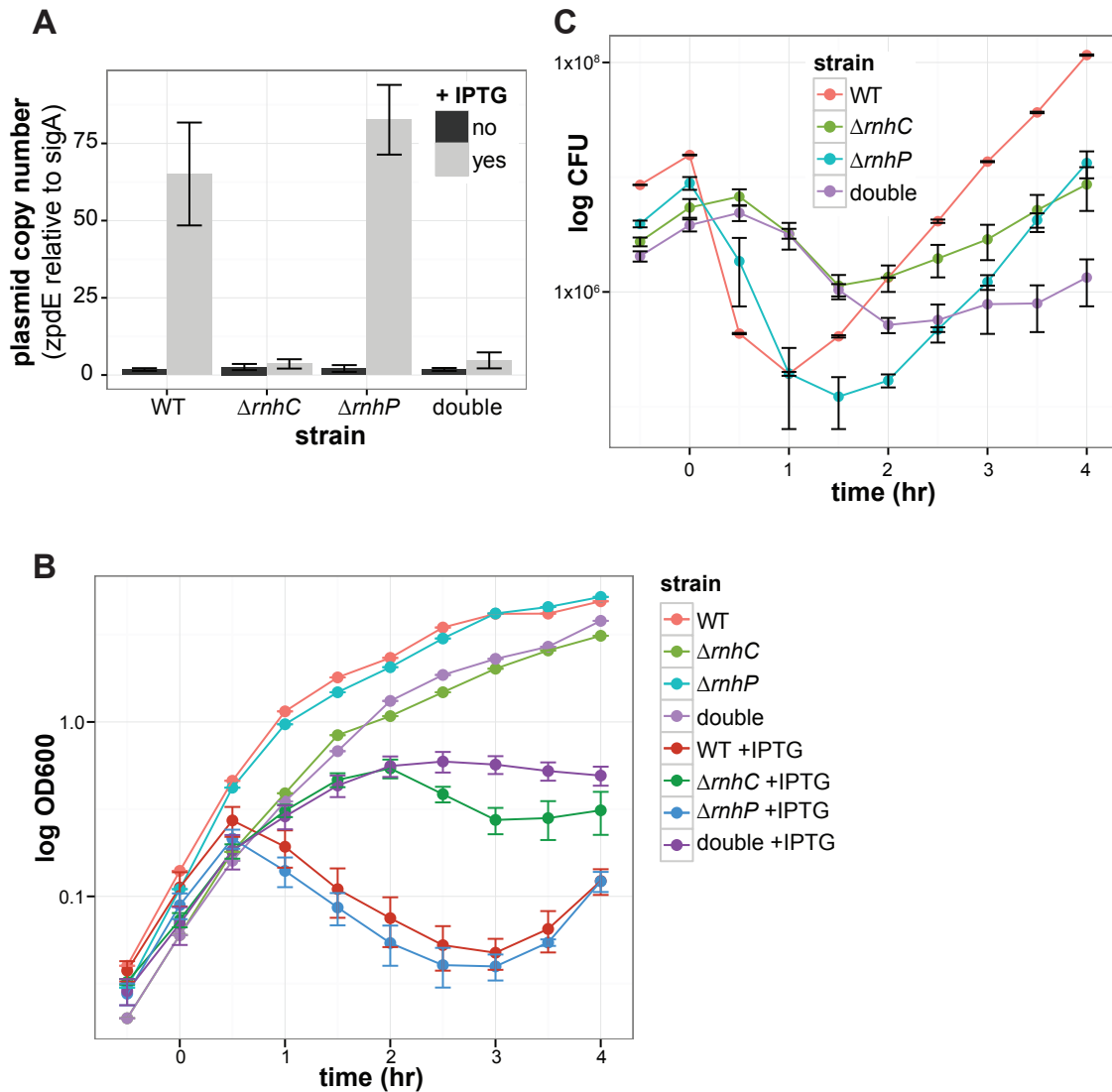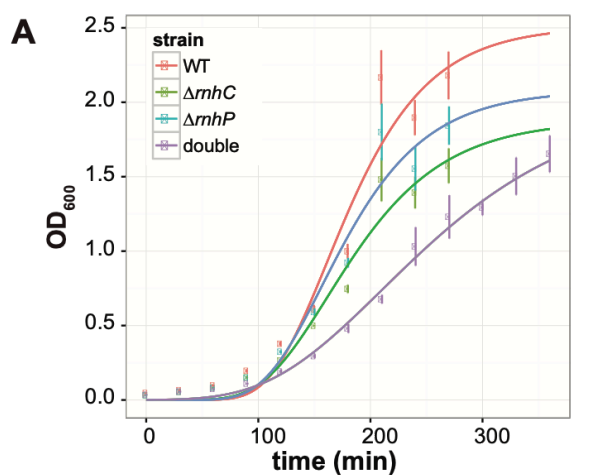
179

Figure 4.3. RnhC, not RnhP, is required for plasmid hyper-replication. (A) Plasmid copy number for IPTG inducible *sigN* strains in WT, Δ*rnhC*, Δ*rnhP*, and Δ*zpdC rnhC::erm* backgrounds with (light gray bars) and without (dark gray bars) IPTG. The plasmid copy number was assessed via qPCR ratio of the plasmid encoded *zpdE* gene to the chromosomally encoded housekeeping sigma factor *sigA*. **(B)** Average OD_{600} (y-axis) of *sigN* inducible strains in WT, Δ*rnhC*, Δ*rnhP*, and Δ*rnhP rnhC::erm* backgrounds over time (x-axis). Representative curves for uninduced strains are indicated. For each IPTG induced strain the average and standard error for three independent replicates is reported. IPTG was added at time 0. **(C)** Average colony forming units (y-axis) for three replicates of *sigN* inducible strains in WT, Δ*rnhC*, Δ*rnhP*, and Δ*rnhP rnhC::erm* backgrounds over time (x-axis). The standard errors are indicated.
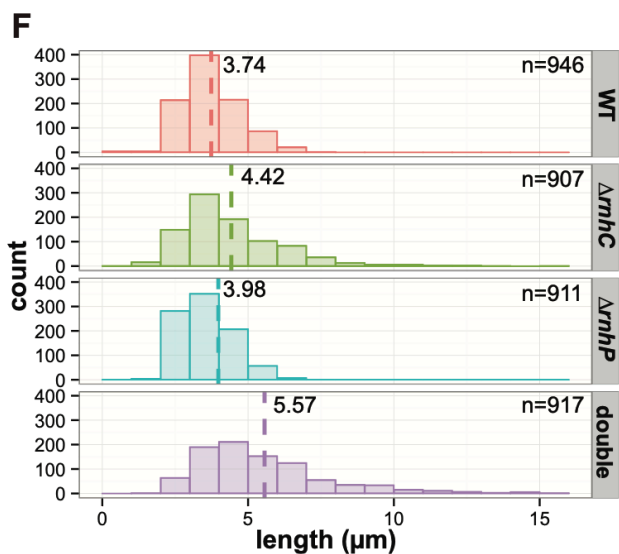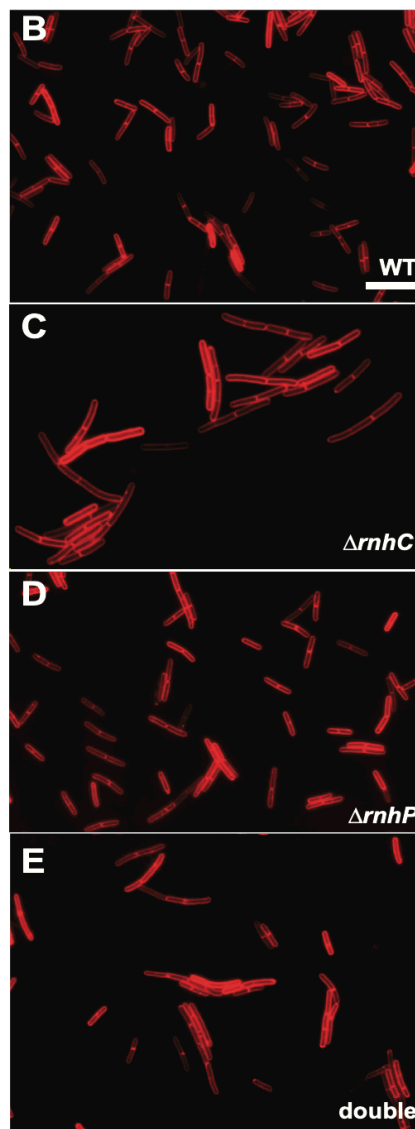
Figure 4.4. Loss of RnhP and RnhC results in decreased cell growth and increased cell length during exponential growth. (A) Growth curves for WT, Δ*rnhC*, Δ*zpdC*, and Δ*rnhP rnhC::erm* in LB media with shaking at 30°C. The growth curves were fit to a Gompertz growth model and the estimated growth rates and corresponding doubling times are indicated with 95% confidence intervals. **(B-E)** Representative images for scoring cell length for WT, Δ*rnhC*, Δ*rnhP*, and Δ*rnhP rnhC::erm*,

respectively. Cells were grown in LB media with shaking at 30°C to mid-exponential growth and treated with a membrane strain for subsequent imaging. **(F)** The distributions of cell lengths plotted for each strain. The dashed line for each strain indicates the average cell length. The number of scored cells is indicated.

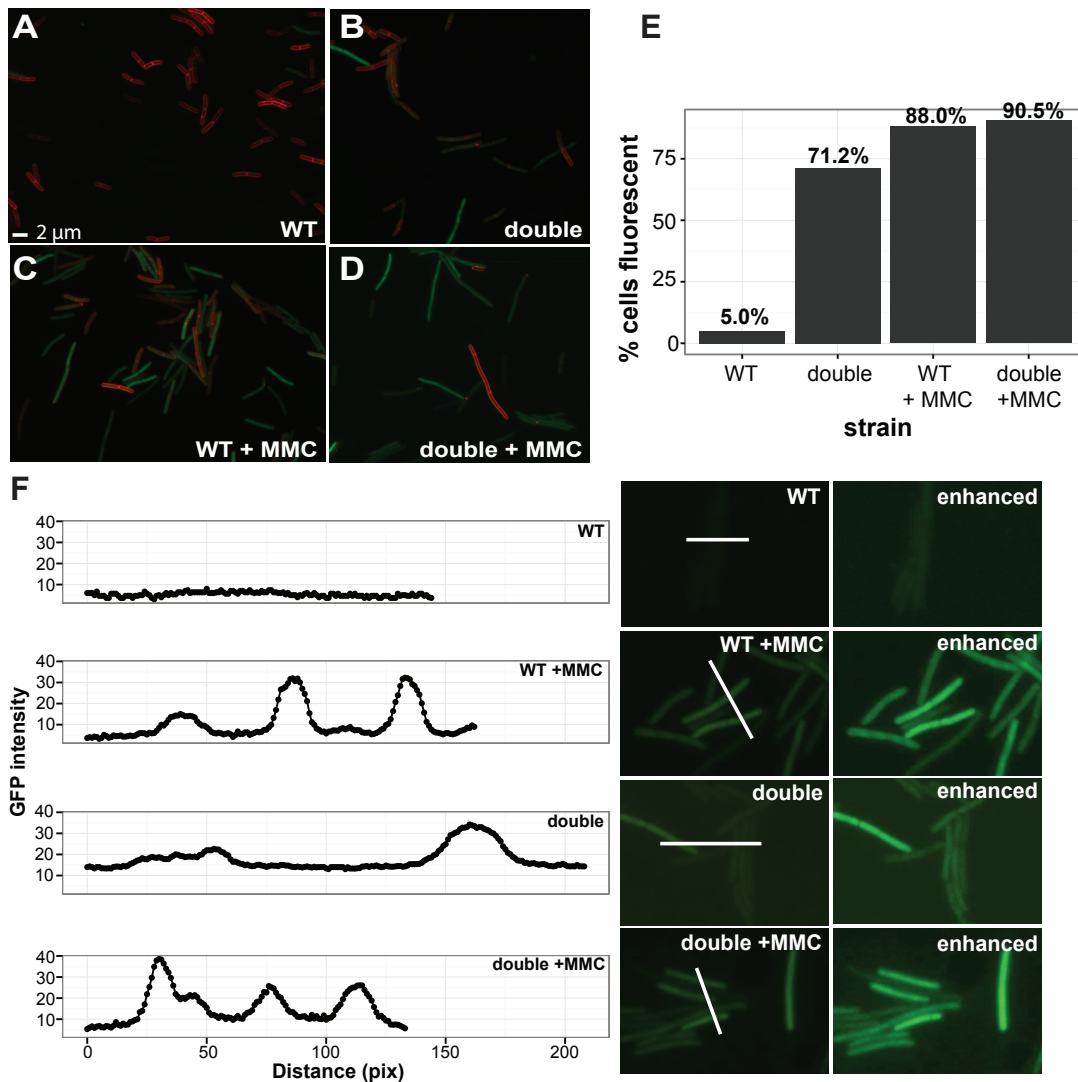**Figure 4.5. Loss of RnhP and RnhC results in induction of the SOS response under normal growth conditions. (A-B)** Representative images for *tagC::tagC-gfp* reporter strains in WT and Δ*rnhP rnhC::erm* backgrounds. **(C-D)** Representative images for *tagC::tagC-gfp* reporter strains in WT and Δ*rnhP rnhC::erm* backgrounds plus treatment with mitomycin C. (**E**) Quantitation of cells expressing the *tagC::tagC-gfp* reporter in WT and Δ*rnhP rnhC::erm* with and without mitomycin C treatment. The strain backgrounds and treatment status are indicated on the x-axis and the percent of cells expressing the reporter is indicated on the y-axis. The percent of fluorescent cells for each strain is indicated above the bar. For WT, WT with MMC treatment, Δ*rnhP rnhC::erm,* and Δ*rnhP rnhC::erm* with MMC treatment reporters 799, 744, 799, and 767 cells were scored per strain, respectively. **(F)** Single image GFP intensities for *tagC::tagC-gfp* reporter strains in WT and Δ*rnhP rnhC::erm* backgrounds plus treatment with mitomycin C. The GFP intensity per pixel was quantified for each strain and plotted. The white line used to quantify pixels for GFP intensity is indicated in each image. An

enhanced image for each strain is also shown. The GFP intensity per pixel was quantified for each strain and plotted to demonstrate background fluorescence in (WT) relative to the fluorescence intensity observed in cells inducing SOS as measure by TagC-GFP fluorescence.
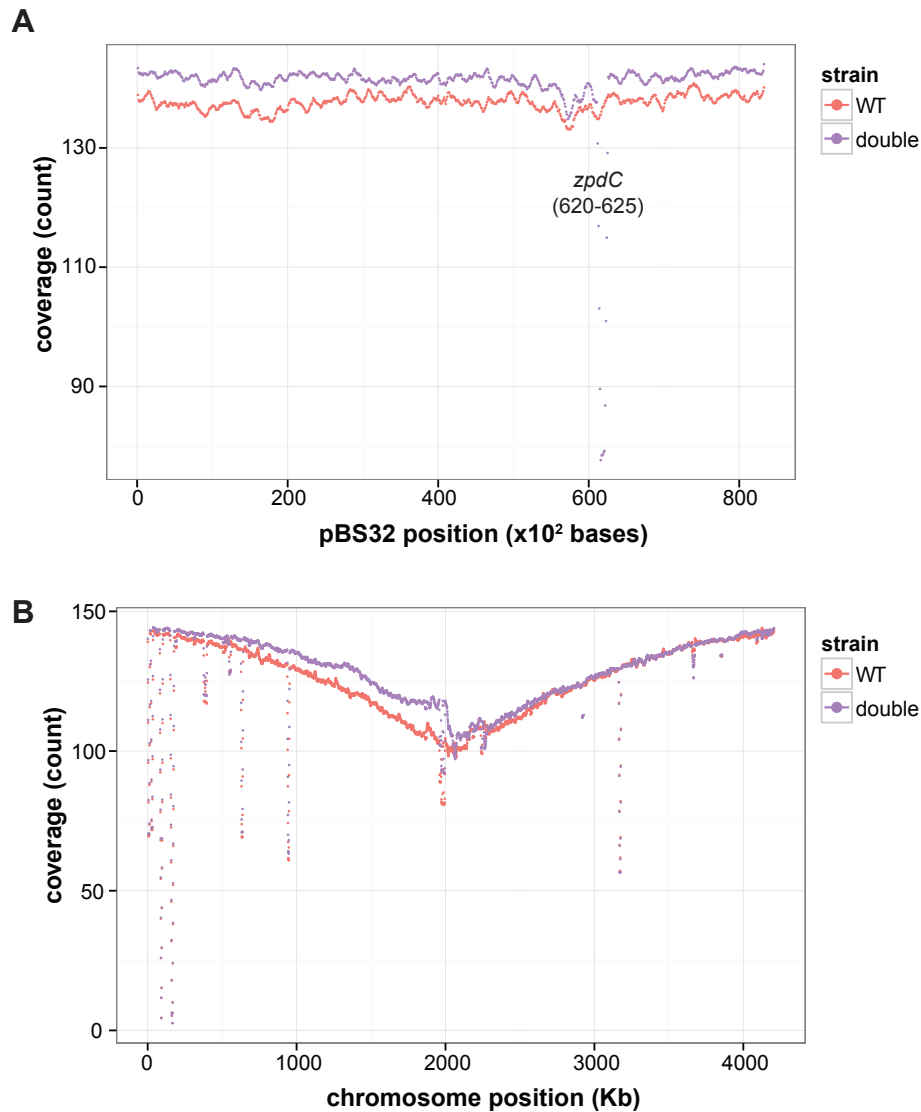
Figure 4.6. Loss of RnhP and RnhC results in replication conflicts around the terminus. (A) Average plasmid coverage of exponentially growing WT and Δ*rnhP rnhC::erm* cells. The average sequencing coverage (y-axis) of three independent replicates for reads aligned to the pBS32 reference over 1kb regions are plotted in 100 bp sliding windows over the length of the plasmid on the x-axis. The plots for the WT and Δ*rnhP rnhC::erm* strains are indicated. **(B)** Average genome coverage of exponentially growing WT and Δ*rnhP rnhC::erm* cells. Average sequencing coverage (y-axis) of three independent replicates for reads aligned to the NCIB 3610 reference (37) chromosome over 10 Kb regions is plotted in 1Kb sliding windows over the length of the chromosome (x-axis). The first origin proximal base in the reference genome represents position 1. The plots for the WT and Δ*rnhP rnhC::erm* are indicated.

185
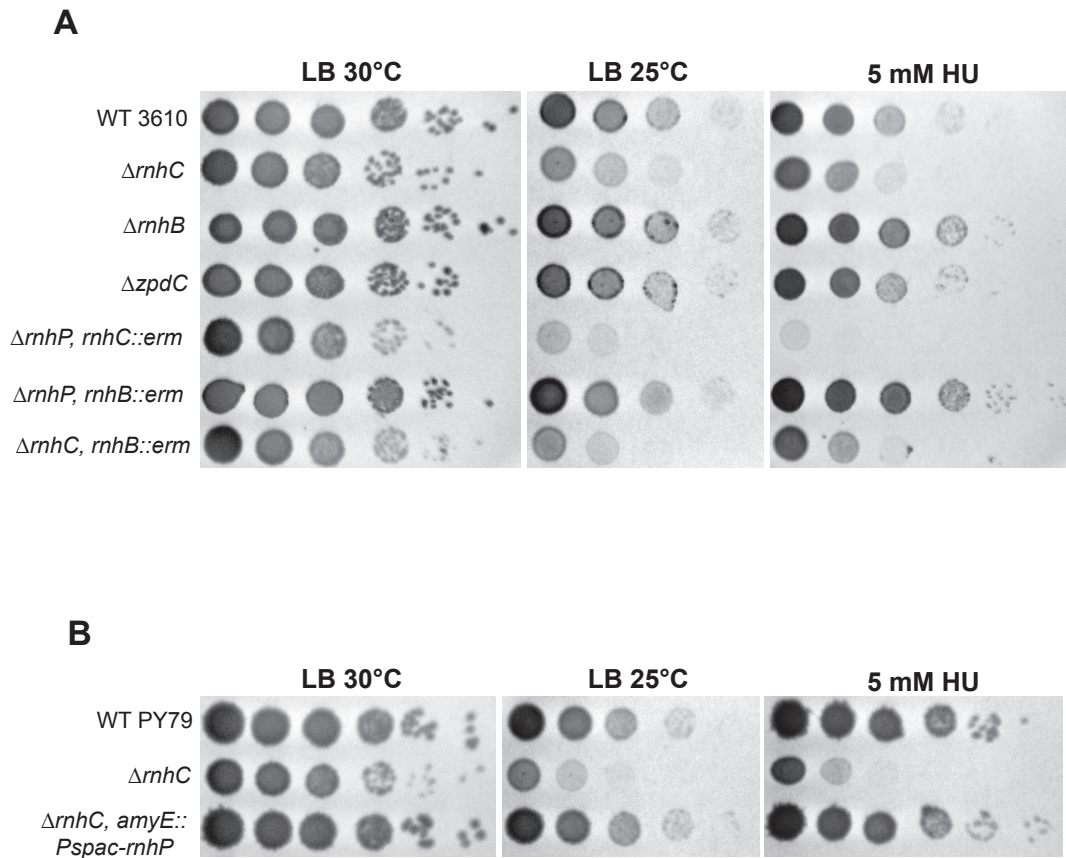
Figure 4.7. RnhP contributes to the mitigation of cell stress caused by DNA damage. (A) Single and pairwise deletion strains in NCIB 3610 were serially diluted 10-fold and spotted onto LB agar media at 30°C, 25°C, and with 5mM hydroxyurea added to the plates. Plates were imaged after overnight incubation at indicated temperatures. **(B)** PY79 strains were serially diluted 10-fold and spotted onto LB agar media at 30°C, 25°C, and with 5 mM hydroxyurea added to the plates. Plates were imaged after overnight incubation at indicated temperatures.
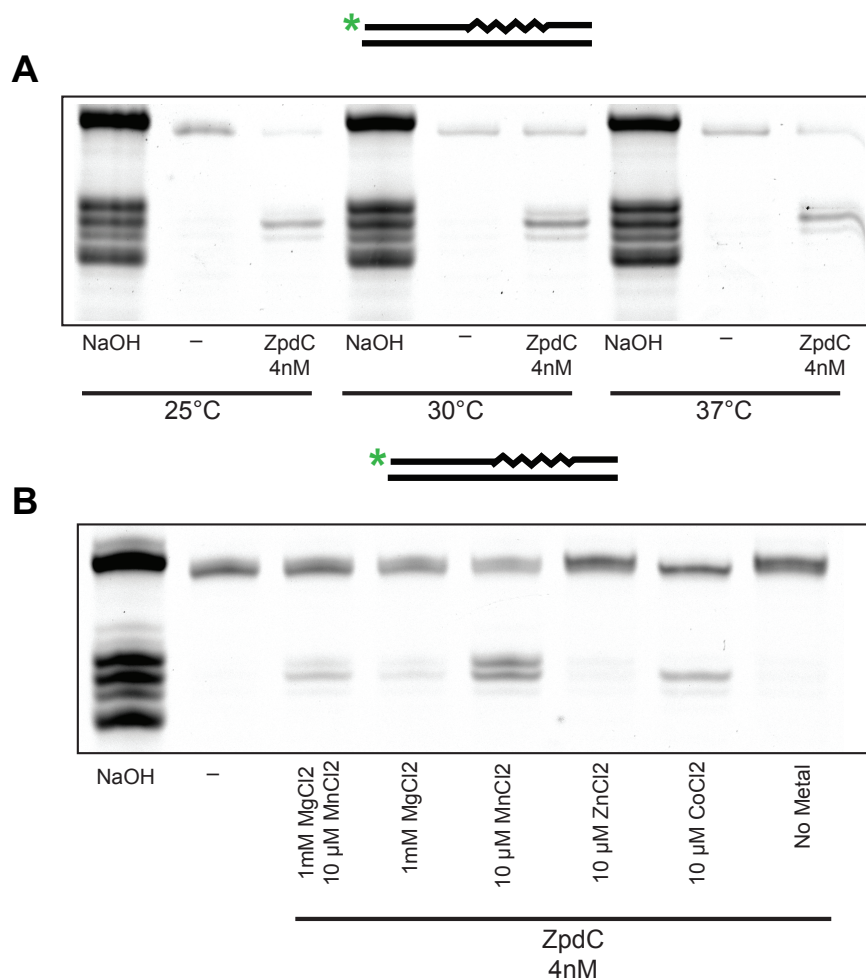
Figure 4.8. ZpdC is active at various temperatures and prefers Mn^{2+} as a metal cofactor. (A) ZpdC was incubated in reaction buffer (see Materials and Methods) with the ribopatch substrate at the indicated temperatures for ten minutes. A 5' end IR-labeled oligo containing 4 embedded rNMPs (squiggly lines) within an otherwise DNA oligo (straight lines) was annealed to a complementary DNA oligo (oJR210 and oJR145) and treated with 4 nM ZpdC at the indicated temperatures. **(B)** ZpdC was incubated with the ribopatch substrate described in (A) in a reaction buffer (20 mM Tris-HCl pH8, 50 nM NaCl, and 1 mM DTT) containing the indicated concentrations of metal ions. For both experiments, a ladder was created via alkaline hydrolysis of the substrate at the embedded rNMPs (lane one). The products were separated on a 20% denaturing urea-PAGE gel and subsequently visualized with a LI-COR Odyssey imager.
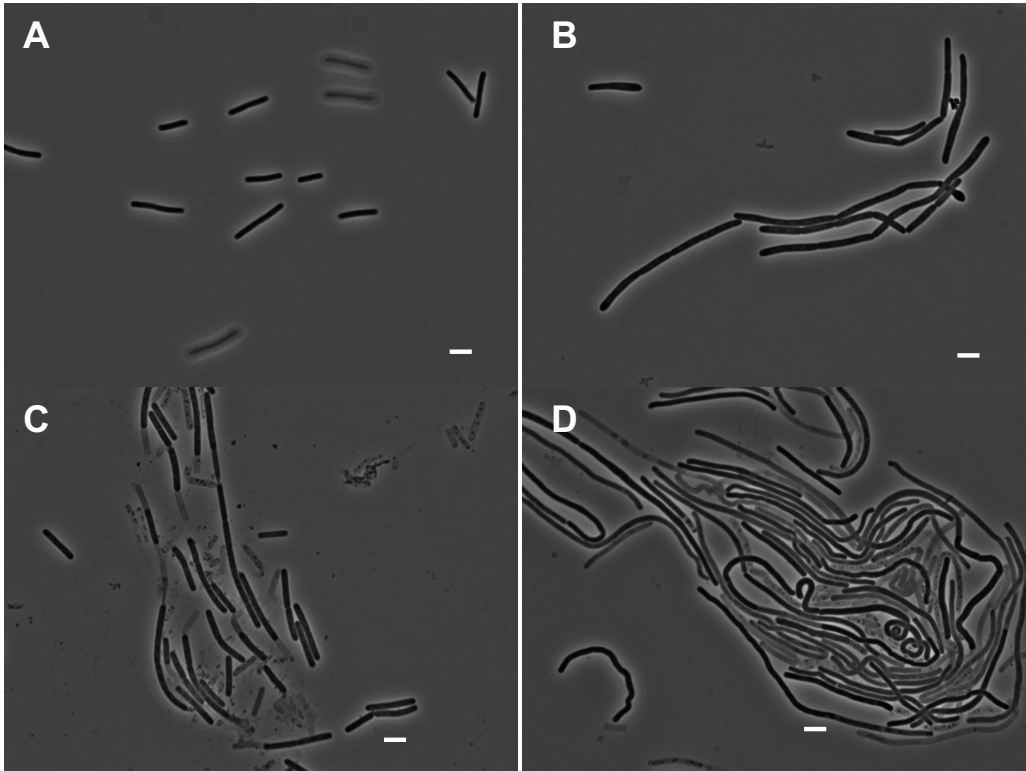
Figure 4.9. Double deletion cells are elongated during induction of plasmid hyper-replication. (A-B) Representative images for IPTG inducible *sigN* strains in WT and Δ*rnhP rnhC::erm* backgrounds pre-induction with IPTG, respectively. **(C-D)** Representative images for IPTG inducible *sigN* strains in WT and Δ*rnhP rnhC::erm* backgrounds 4 hours post induction with IPTG, respectively.
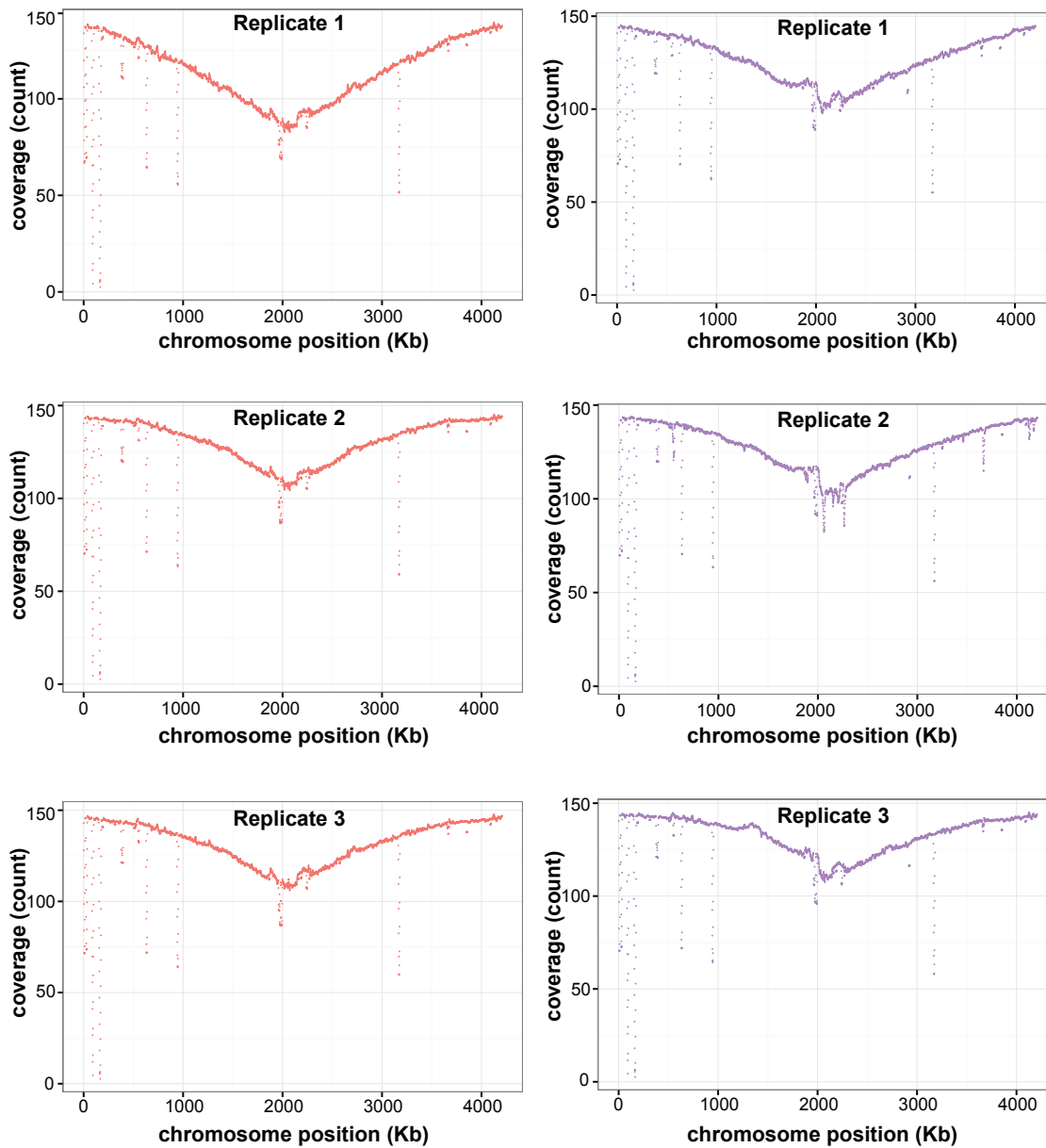
Fig 4.10.

Sequencing coverage of replicates for WT and double deletion strains. Genome coverage for three independent replicates of exponentially growing WT and Δ*rnhP rnhC::erm* cells. Sequencing coverage (y-axis) of reads aligned to the NCIB 3610 reference chromosome over 10 Kb regions is plotted in 1Kb sliding windows over the length of the chromosome (x-axis). The first origin proximal base in the reference genome represents position 1. The independent replicates for each strain are plotted separately.

Table 4.1. Strains used in this study

Strains	Genotype	Source
TMN73	PY79	(59)
DK1042	NCIB 3610 comIQ12I	(38)
JWS207	PY79 Δ*rnhC*	(14)
BKE28620	*rnhC::lox-erm-lox*	BGSC
BKE16060	*rnhB::lox-erm-lox*	BGSC
TMN107	DK1042 Δ*rnhC*	
DK7047	DK1042 Δ*zpdC*	
TMN110	DK1042 Δ*zpdC, rnhC::erm*	
TMN103	PY79 Δ*rnhC, amyE::Pspac-zpdC*	
TMN104	BL21 x pE-SUMO$_{zpdC}$	
TMN112	BL21 x pE-SUMO$_{zpdCD73N}$	
KJW7	PY79 *tagC::tagC-gfp*	(60)
TMN115	DK1042 *tagC::tagC-gfp*	
TMN128	DK1042 Δ*zpdC,rnhC::erm, tagC::tagC-gfp*	
DK1634	Δ*PBSX ΔSPβ ΔcomI amyE::hyspank-zpdN*	(61)
DK7765	Δ*PBSX ΔSPβ ΔcomI ΔzpdC amyE::hyspank-sigN*	
DK7814	Δ*PBSX ΔSPβ ΔcomI ΔzpdC rnhC::erm amyE::hyspank-sigN*	
DK7868	comIQ12L rnhC::erm *amyE::hyspank-sigN*	

Unless otherwise indicated, the 'Source' is this study.

Table 4.2. Plasmids used in this study

Plasmid	Vector	Insert	Source
pTN$_{zpdC1}$	pDR110	*rnhP*	This study
pTN$_{zpdC2}$	pE-SUMO	*rnhP*	This study
pTN$_{zpdC2}$	pE-SUMO	*rnhP* D73N	This study
pAMB32	pMiniMAD2		(62)

Table 4.3. Oligos used in this study

Primer name	Primer sequence
oTN56	TTAGTCGACTAAGGAGGTATACATATGAAAAAGGTTGTAATTTAC
oTN57	TTGCATGCGGCTAGCttaTCATACGGCAG
oTN58	CGCGAACAGATTGGAGGTATGAAAAAGGTTGTAATT
oTN59	GTGGTGGTGCTCGATCATACGGCAGC
oTN60	CCCTGTAGAAATCAATACTAATTCTGCATATCTGTGCAAC
oTN61	GTTGCACAGATATGCAGAATTAGTATTGATTTCTACAGGG
oJR209	/5IRD800CWN/CGATCGTAArGCTAGCTCTGC
oJR210	/5IRD800CWN/CGATCGTArArGrCrUAGCTCTGC
oJR227	/5IRD800CWN/rCrGrArUrCrGrUrArArGrCrUrArGrCrUrCrUrGrC
oJR145	GCAGAGCTAGCTTACGATCG
oJR339	rArGrUrArGrUrGrArArCrCrATGCTTACG/3IR800CWN/
oJR340	CGTAAGCATGGTTCACTACTCGCGCTTGATGC
oJR166	rGrCrArGrArGrArCrUrArGrCrUrUrArCrGrArUrCrG
oJR348	AGTAGTGAACCATGCTTACG/3IRD800CWN/
oJR365	CGTAAGCATGGTTCACTACT
oAB6715	AGGAGGAAGCTTGCCCGAAAATGATGATTATGG
oAB6716	CCTCCTGTCGACGTAAATTACAACCTTTTTCATTAAAG
oAB6717	AGGAGGGTCGACGCCGTATGAATGAATCAGTCTTC
oAB6718	CCTCCTGGTACCGAGCAATAGGATATGCCCGAC

Black and red text represents DNA and RNA sequences, respectively. IRDXXX represents infrared dye with excitation at 700 or 800 nM either at the 5' (5) or 3' (3) end of the oligo. CWN is NHS ester conjugation.

Supplemental text

Supplemental materials and methods

Strain construction: Chromosome deletion strains were created by transforming competent DK1042 cells (38) with genomic DNA purified from *Bacillus subtilis* 168 strains with the gene of interest replaced by an erythromycin resistance cassette flanked by *loxP* sites obtained from the Bacillus Genetic Stock Center (http://www.bgsc.org/). The erythromycin resistance cassette was subsequently removed with Cre recombinase (63).

To generate the Δ*rnhP* in-frame markerless deletion plasmid pAMB32 was constructed. The region 5' to *rnhP* was amplified using the primer pair 6715/6716 and subsequently digested with HindIII and SalI, and the region 3' of *rnhP* was amplified with primer pair 6717/6718 and digested with SalI and KpnI. The two fragments were simultaneously ligated into HindIII/KpnI-digested pMiniMAD2, which contains a temperature-sensitive origin of replication and an erythromycin resistance cassette (62). *Escherichia coli* TG1 was transformed with the resulting product to generate pAMB32. The pAMB32 plasmid was introduced into DK1042 by transformation at the permissive temperature for plasmid replication (22°C) using mls resistance as a selection. The resulting strain (DK7021) was grown on plates containing mls at the restrictive temperature for plasmid replication (37°C) to force integration of the extra-chromosomal plasmid into pBS32. To evict the plasmid, the strain was incubated in 3 mL LB at the permissive temperature for 14h, diluted 30-fold in fresh LB, and incubation continued at the permissive temperature for another 24h. Cells were serially diluted and plated on LB agar at 37°C. Individual colonies were replica patched onto LB plates ad LB plates containing mls to identify mls-sensitive colonies that evicted the plasmid. Colonies that had evicted the plasmid were screened by PCR using primers 6715/6718 to assess which isolates retained the Δ*rnhP* allele.

The *tagC::tagC-gfp* reporter strains were created by transforming genomic DNA purified from *tagC::tagC-gfp* in *B. subtilis* PY79 (KJW7) into the appropriate background and verified via resistance to the selectable marker spectinomycin and microscopy (60). The

inducible *sigN* strains were created by transduction with lysate from DK1634 and subsequently verified

References

1.   Schroeder JW, Randall JR, Matthews LA, Simmons LA. 2015. Ribonucleotides in bacterial DNA. Crit Rev Biochem Mol Biol 50:181-93.
2.   Williams JS, Kunkel TA. 2014. Ribonucleotides in DNA: origins, repair and consequences. DNA Repair (Amst) 19:27-37.
3.   Santos-Pereira JM, Aguilera A. 2015. R loops: new modulators of genome dynamics and function. Nat Rev Genet 16:583-97.
4.   Schroeder JW, Randall JR, Hirst WG, O'Donnell ME, Simmons LA. 2017. Mutagenic cost of ribonucleotides in bacterial DNA. Proc Natl Acad Sci U S A 114:11733-11738.
5.   Nick McElhinny SA, Kumar D, Clark AB, Watt DL, Watts BE, Lundstrom EB, Johansson E, Chabes A, Kunkel TA. 2010. Genome instability due to ribonucleotide incorporation into DNA. Nature chemical biology 6:774-81.
6.   Nick McElhinny SA, Watts BE, Kumar D, Watt DL, Lundstrom EB, Burgers PM, Johansson E, Chabes A, Kunkel TA. 2010. Abundant ribonucleotide incorporation into DNA by yeast replicative polymerases. Proceedings of the National Academy of Sciences of the United States of America 107:4949-54.
7.   Huertas P, Aguilera A. 2003. Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. Mol Cell 12:711-21.
8.   Kouzminova EA, Kadyrov FF, Kuzminov A. 2017. RNase HII Saves rnhA Mutant Escherichia coli from R-Loop-Associated Chromosomal Fragmentation. J Mol Biol 429:2873-2894.
9.   Randall JR, Hirst WG, Simmons LA. 2017. Substrate specificity for bacterial RNase HII and HIII is influenced by metal availability. J Bacteriol 200:pii: e00401-17.
10.   Randall JR, Nye TM, Wozniak KJ, Simmons LA. 2019. RNase HIII Is Important for Okazaki Fragment Processing in Bacillus subtilis. J Bacteriol 201: pii: e00686-18.
11.   Rowen L, Kornberg A. 1978. A ribo-deoxyribonucleotide primer synthesized by primase. J Biol Chem 253:770-4.
12.   Rowen L, Kornberg A. 1978. Primase, the dnaG protein of Escherichia coli. An enzyme which starts DNA chains. J Biol Chem 253:758-64.
13.   Cronan GE, Kouzminova EA, Kuzminov A. 2019. Near-continuously synthesized leading strands in Escherichia coli are broken by ribonucleotide excision. Proc Natl Acad Sci U S A 116:1251-1260.
14.   Yao NY, Schroeder JW, Yurieva O, Simmons LA, O'Donnell ME. 2013. Cost of rNTP/dNTP pool imbalance at the replication fork. Proceedings of the National Academy of Sciences of the United States of America 110:12942-7.
15.   Ordonez H, Uson ML, Shuman S. 2014. Characterization of three mycobacterial DinB (DNA polymerase IV) paralogs highlights DinB2 as naturally adept at ribonucleotide incorporation. Nucleic Acids Res 42:11056-70.
16.   Oivanen M, Kuusela S, Lonnberg H. 1998. Kinetics and Mechanisms for the Cleavage and Isomerization of the Phosphodiester Bonds of RNA by Bronsted Acids and Bases. Chem Rev 98:961-990.

17. Das U, Chauleau M, Ordonez H, Shuman S. 2014. Impact of DNA3'pp5'G capping on repair reactions at DNA 3' ends. Proc Natl Acad Sci U S A 111:11317-22.

18. Das U, Shuman S. 2013. Mechanism of RNA 2',3'-cyclic phosphate end healing by T4 polynucleotide kinase-phosphatase. Nucleic Acids Res 41:355-65.

19. Lang KS, Hall AN, Merrikh CN, Ragheb M, Tabakh H, Pollock AJ, Woodward JJ, Dreifus JE, Merrikh H. 2017. Replication-Transcription Conflicts Generate R-Loops that Orchestrate Bacterial Stress Survival and Pathogenesis. Cell 170:787-799 e18.

20. Prado F, Aguilera A. 2005. Impairment of replication fork progression mediates RNA polII transcription-associated recombination. EMBO J 24:1267-76.

21. Ohtani N, Haruki M, Morikawa M, Kanaya S. 1999. Molecular diversities of RNases H. J Biosci Bioeng 88:12-9.

22. Cerritelli SM, Crouch RJ. 2009. Ribonuclease H: the enzymes in eukaryotes. FEBS J 276:1494-505.

23. Li TK, Barbieri CM, Lin HC, Rabson AB, Yang G, Fan Y, Gaffney BL, Jones RA, Pilch DS. 2004. Drug targeting of HIV-1 RNA.DNA hybrid structures: thermodynamics of recognition and impact on reverse transcriptase-mediated ribonuclease H activity and viral replication. Biochemistry 43:9732-42.

24. Sparks JL, Chon H, Cerritelli SM, Kunkel TA, Johansson E, Crouch RJ, Burgers PM. 2012. RNase H2-initiated ribonucleotide excision repair. Mol Cell 47:980-6.

25. Ohtani N, Haruki M, Muroya A, Morikawa M, Kanaya S. 2000. Characterization of ribonuclease HII from Escherichia coli overproduced in a soluble form. J Biochem 127:895-9.

26. Ohtani N, Yanagawa H, Tomita M, Itaya M. 2004. Cleavage of double-stranded RNA by RNase HI from a thermoacidophilic archaeon, Sulfolobus tokodaii 7. Nucleic Acids Res 32:5809-19.

27. Kochiwa H, Tomita M, Kanai A. 2007. Evolution of ribonuclease H genes in prokaryotes to avoid inheritance of redundant genes. BMC Evol Biol 7:128.

28. Nowotny M, Gaidamakov SA, Ghirlando R, Cerritelli SM, Crouch RJ, Yang W. 2007. Structure of human RNase H1 complexed with an RNA/DNA hybrid: insight into HIV reverse transcription. Mol Cell 28:264-76.

29. Chon H, Sparks JL, Rychlik M, Nowotny M, Burgers PM, Crouch RJ, Cerritelli SM. 2013. RNase H2 roles in genome integrity revealed by unlinking its activities. Nucleic Acids Res 41:3130-43.

30. Ohtani N, Haruki M, Morikawa M, Crouch RJ, Itaya M, Kanaya S. 1999. Identification of the genes encoding Mn2+-dependent RNase HII and Mg2+-dependent RNase HIII from Bacillus subtilis: classification of RNases H into three families. Biochemistry 38:605-18.

31. Itoh T, Tomizawa J. 1980. Formation of an RNA primer for initiation of replication of ColE1 DNA by ribonuclease H. Proc Natl Acad Sci U S A 77:2450-4.

32. Fukushima S, Itaya M, Kato H, Ogasawara N, Yoshikawa H. 2007. Reassessment of the in vivo functions of DNA polymerase I and RNase H in bacterial cell growth. J Bacteriol 189:8575-83.

33. Kearns DB, Chu F, Branda SS, Kolter R, Losick R. 2005. A master regulator for biofilm formation by Bacillus subtilis. Mol Microbiol 55:739-49.

34.   Kearns DB, Chu F, Rudner R, Losick R. 2004. Genes governing swarming in Bacillus subtilis and evidence for a phase variation mechanism controlling surface motility. Mol Microbiol 52:357-69.

35.   Kearns DB, Losick R. 2005. Cell population heterogeneity during growth of Bacillus subtilis. Genes Dev 19:3083-94.

36.   McLoon AL, Guttenplan SB, Kearns DB, Kolter R, Losick R. 2011. Tracing the domestication of a biofilm-forming bacterium. J Bacteriol 193:2027-34.

37.   Nye TM, Schroeder JW, Kearns DB, Simmons LA. 2017. Complete Genome Sequence of Undomesticated Bacillus subtilis Strain NCIB 3610. Genome Announc 5:pii: e00364-17.

38.   Konkol MA, Blair KM, Kearns DB. 2013. Plasmid-encoded ComI inhibits competence in the ancestral 3610 strain of Bacillus subtilis. J Bacteriol 195:4085-93.

39.   Burton AT, DeLoughery A, Li GW, Kearns DB. 2019. Transcriptional Regulation and Mechanism of SigN (ZpdN), a pBS32-Encoded Sigma Factor in Bacillus subtilis. mBio 10.

40.   Zwietering MH, Jongenburger I, Rombouts FM, van 't Riet K. 1990. Modeling of the bacterial growth curve. Appl Environ Microbiol 56:1875-81.

41.   Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754-60.

42.   Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078-9.

43.   Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA, 3rd, Smith HO. 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods 6:343-5.

44.   Skulj M, Okrslar V, Jalen S, Jevsevar S, Slanc P, Strukelj B, Menart V. 2008. Improved determination of plasmid copy number using quantitative real-time PCR for monitoring fermentation processes. Microb Cell Fact 7:6.

45.   Simmons LA, Davies BW, Grossman AD, Walker GC. 2008. Beta clamp directs localization of mismatch repair in Bacillus subtilis. Mol Cell 29:291-301.

46.   Lenhart JS, Sharma A, Hingorani MM, Simmons LA. 2013. DnaN clamp zones provide a platform for spatiotemporal coupling of mismatch detection to DNA replication. Molecular microbiology 87:553-68.

47.   Earl AM, Eppinger M, Fricke WF, Rosovitz MJ, Rasko DA, Daugherty S, Losick R, Kolter R, Ravel J. 2012. Whole-genome sequences of Bacillus subtilis and close relatives. Journal of Bacteriology 194:2378-9.

48.   Earl AM, Losick R, Kolter R. 2007. Bacillus subtilis genome diversity. J Bacteriol 189:1163-70.

49.   Katayanagi K, Miyagawa M, Matsushima M, Ishikawa M, Kanaya S, Ikehara M, Matsuzaki T, Morikawa K. 1990. Three-dimensional structure of ribonuclease H from E. coli. Nature 347:306-9.

50.   Hardwood CR, Cutting SM. 1990. Molecular Biological Methods for Bacillus. John Wiley & Sons, Chichester.

51.   Gupta R, Chatterjee D, Glickman MS, Shuman S. 2017. Division of labor among Mycobacterium smegmatis RNase H enzymes: RNase H1 activity of RnhA or

RnhC is essential for growth whereas RnhB and RnhA guard against killing by hydrogen peroxide in stationary phase. Nucleic Acids Res 45:1-14.

52.   Asai T, Kogoma T. 1994. D-loops and R-loops: Alternative mechanisms for the initiation of chromosome replication in *Escherichia coli*. J Bacteriol 176:1807-1812.

53.   Burby PE, Simmons ZW, Schroeder JW, Simmons LA. 2018. Discovery of a dual protease mechanism that promotes DNA damage checkpoint recovery. PLoS Genet 14:e1007512.

54.   Kawai Y, Moriya S, Ogasawara N. 2003. Identification of a protein, YneA, responsible for cell division suppression during the SOS response in Bacillus subtilis. Mol Microbiol 47:1113-22.

55.   Mo AH, Burkholder WF. 2010. YneA, an SOS-induced inhibitor of cell division in Bacillus subtilis, is regulated posttranslationally and requires the transmembrane region for activity. J Bacteriol 192:3159-73.

56.   Simmons LA, Goranov AI, Kobayashi H, Davies BW, Yuan DS, Grossman AD, Walker GC. 2009. Comparison of responses to double-strand breaks between Escherichia coli and Bacillus subtilis reveals different requirements for SOS induction. J Bacteriol 191:1152-61.

57.   Davies BW, Kohanski MA, Simmons LA, Winkler JA, Collins JJ, Walker GC. 2009. Hydroxyurea induces hydroxyl radical-mediated cell death in Escherichia coli. Molecular Cell 36:845-60.

58.   Yang Z, Hou Q, Cheng L, Xu W, Hong Y, Li S, Sun Q. 2017. RNase H1 Cooperates with DNA Gyrases to Restrict R-Loops and Maintain Genome Integrity in Arabidopsis Chloroplasts. Plant Cell 29:2478-2497.

59.   Youngman P, Perkins JB, Losick R. 1984. Construction of a cloning site near one end of Tn917 into which foreign DNA may be inserted without affecting transposition in Bacillus subtilis or expression of the transposon-borne erm gene. Plasmid 12:1-9.

60.   Britton RA, Kuster-Schock E, Auchtung TA, Grossman AD. 2007. SOS induction in a subpopulation of structural maintenance of chromosome (Smc) mutant cells in Bacillus subtilis. J Bacteriol 189:4359-66.

61.   Myagmarjav BE, Konkol MA, Ramsey J, Mukhopadhyay S, Kearns DB. 2016. ZpdN, a Plasmid-Encoded Sigma Factor Homolog, Induces pBS32-Dependent Cell Death in Bacillus subtilis. J Bacteriol 198:2975-2984.

62.   Patrick JE, Kearns DB. 2008. MinJ (YvjD) is a topological determinant of cell division in Bacillus subtilis. Mol Microbiol 70:1166-79.

63.   Koo, B.M., Kritikos, G., Farelli, J.D., Todor, H., Tong, K., Kimsey, H., Wapinski, I., Galardini, M., Cabal, A., Peters, J.M. *et al.* (2017) Construction and Analysis of Two Genome-Scale Deletion Libraries for Bacillus subtilis. *Cell Syst*, **4**, 291-305 e297.

CHAPTER V

Concluding Remarks and Future Research

Introduction

There has been a growing recognition that post-replicative DNA methylation regulates critical cellular functions across all three domains of life. In bacteria, the study of the regulatory functions of DNA methylation has largely been confined to orphan MTases from Gram-negative *E. coli, C. crescentus*, and related Proteobacteria (for review (15; 20). Studies investigating the regulatory effects of Type III RM system MTases have primarily focused on Gram-negative species as well (for review (32-35)). In this dissertation I describe the important contribution of DNA methylation to gene regulation in Gram-positive bacteria. In Chapter II, I describe how DNA methylation from a Type I RM system promotes the expression of a subset of genes in the Gram-positive pathogen *Streptococcus pyogenes* (Chapter II (22)). I show that among the differentially expressed genes is the Mga core regulon, which includes genes involved in adhesion, internalization, and immune evasion phenotypes (12; 16), and that all of these genes are substantially down regulated in the absence of the RM system MTase. Further, I demonstrate that the m6A-dependent decrease in gene expression results in attenuated adherence and virulence of the RM system mutant relative to the wild type strain (Chapter II (22)). In Chapter III I describe how genomic m6A modifications from the previously uncharacterized MTase, DnmA, functions to promote expression of a small subset of genes involved in chromosome maintenance in the Gram-positive Firmicute *Bacillus subtilis* (Chapter III (23)). I also identify an m6A-sensitive transcription factor, providing some of the first mechanistic insight into m6A-dependent regulation in Gram-positive bacteria (23). Finally, in Chapter IV I describe how RNA incorporated into DNA, representing the most frequent type of noncanonical nucleotide found in DNA (29; 38) can be resolved by a plasmid encoded RNase H protein providing the first example of a

bacterial organism encoding functional RNase HI, HII, and HIII enzymes. My work establishes the importance of DNA modifications to regulation of cell physiology in Firmicutes and opens many new paths of study for future research.

Elucidating direct and indirect mechanisms of m6A-dependent changes in gene expression. The gene expression changes described in Chapters II and III can be the result of both direct and indirect regulation by m6A **(Fig 1.6)**. A direct regulatory mechanism would consist of m6A influencing transcription factor binding directly within the promoter region of a differentially expressed gene, such as the m6A-sensitive ScoC binding described in Chapter III. Indirect mechanisms of regulation can occur downstream of direct regulation, such as differential expression of a transcription factor or DNA binding protein that subsequently results in many genes being differentially expressed, or may be independent of a direct change in gene expression. Here I describe how this thesis directly contributes to future studies of both direct and indirect mechanisms of m6A-dependent changes in gene expression.

Direct mechanisms for m6A-dependent regulation of gene expression in Firmicutes. In Chapter III I described a subset of genes that were down regulated in the absence of m6A in *B. subtilis* (23)*.* The differentially expressed genes contained a DnmA recognition site proximal to the -35 binding box of SigA. To elucidate the mechanism of m6A-dependent gene expression changes, I performed a pull down of *B. subtilis* whole cell lysates using biotinylated oligos for the promoter region of *scpA*, one of the differentially expressed genes upon loss of m6A. I found that ScoC, a transcriptional repressor of genes expressed in the transition state, preferentially bound to an umethylated promoter sequence. I further confirmed the differential binding of ScoC to the *scpA* promoter region using gel shift assays with purified ScoC and labeled promoter probes (23). Previous studies in Gram-negative *E. coli* and *C. crescentus* have identified transcriptional regulators whose binding is dependent on m6A from orphan MTases Dam and CcrM methylation, respectively (2; 10; 11; 18; 28). To my knowledge, ScoC is the first example of m6A-sensitive transcription factor binding in Gram-positive bacteria. Moreover, ScoC provides a rare example of methylation

sensitive transcription factor binding that is not dependent on methylation from an orphan MTase.

ScoC mediated transcriptional repression is complex. ScoC acts to repress many genes that are expressed in the transition from exponential to stationary growth and is regulated by the transcriptional repressor, CodY (1; 4). CodY influences the expression of over 200 genes, most of which are involved in nutrient acquisition, and binds DNA in nutrient rich conditions when amino acids are plentiful (3; 19), and for review (31)). As amino acids become unavailable, CodY-mediated repression is lifted and a subset of genes within the CodY regulon is up regulated, including ScoC (1; 30). Intriguingly, the increased expression of ScoC results in further repression of genes within the CodY regulon that share both ScoC and CodY binding sites, resulting in a redundancy in the repression of these genes (1). Thus, in order to better understand the downstream regulatory affects of m6A-mediated ScoC binding, future studies will have to investigate gene expression effects in amino acid limited media to understand ScoC mediated repression without the inhibitory effects of CodY expression. It is also worth noting that ScoC has not previously been shown to be a regulator of the *scpA* promoter region as demonstrated in Chapter III of this work. A **Ch**romosome **I**mmuno**p**recipitation followed by deep sequencing (ChIP-seq) experiment of ScoC binding throughout the *B. subtilis* genome in amino acid limited media as well as PacBio sequencing of genomic DNA under the same amino acid limiting conditions is necessary to better understand how ScoC binding is affected by the presence of DNA methylation in *B. subtilis.*

The initial pull down experiment of *B. subtilis* lysates with the *scpA* promoter region revealed the potential for several differentially bound proteins between the two promoter sequences. Due to technical limitations at the time of the experiment, I used a promoter region containing a thymine at the m6A position for the methylated promoter, as this mutation showed wild type expression levels in the gene expression experiments. Also, I was only able to identify the differentially bound proteins between 20-40 kDa in size. With the necessary conditions for the pull-down established, this experiment could easily be repeated with an IDT-synthesized methylated promoter region. Additionally,

protein identification via mass spectrometry can be performed directly from the beads used in the pull-down, allowing for differentially bound proteins of all sizes to be identified in subsequent experiments. This protocol could also be adapted for the other differentially expressed promoter regions identified in Chapter III, allowing for further identification of the m6A-senstive transcriptional regulators in *B. subtilis.* These experiments would provide the most comprehensive and mechanistic understanding of m6A-dependent gene regulation in Gram-positive bacteria to date.

Indirect mechanisms of m6A-dependent regulation resulting from direct regulation of gene expression. In addition to the m6A-dependent decrease in gene expression of *scpA* discussed in the previous section, in Chapter III I also identified a subset of genes that were down regulated in the absence of m6A (23). Among the differentially expressed genes was *hbs*, encoding the highly abundant and essential histone-like protein HBsu (17)*.* Although the m6A-dependent decrease of *hbs* expression was small, given the high expression of the *hbs* gene, which is expected to result in 50,000 HBsu monomers per cell (27), a slight decrease in gene expression could have important consequences for protein levels. Altered HBsu binding patterns throughout the *B. subtilis* genome due to decreased protein expression could result in altered chromosome structure (14), which could in turn have downstream effects on gene expression (7). Thus, the direct m6A-dependent decrease in *hbs* levels could result in indirect differential gene expression throughout the *B. subtilis* chromosome.

In order to assess HBsu protein levels and occupancy throughout the *B. subtilis* chromosome, I purified HBsu for antiserum production. Preliminary Western Blots with the HBsu antiserum against whole cell protein lysates and purified HBsu suggest that the antiserum recognizes and is specific for HBsu. However, further validation experiments need to be completed because *hbs* is essential, preventing use of a deletion control for the pull-down experiments. Once these experiments are complete, the antiserum can be used in quantitative Western Blots to determine if the m6A-dependent decrease in *hbs* gene expression also results in decreased levels of HBsu. Furthermore, as a part of ongoing experiments in collaboration with Dr. Peter

Freddolino's Lab, genome-wide protein occupancy can be assessed in the wild type and m6A deficient cells using the *In vivo* **P**rotein **O**ccupancy **D**isplay in **H**igh **R**esolution (IPOD-HR) technique (7). Briefly, the technique involves crosslinking proteins to DNA, degrading all DNA not occupied by protein, reversing the crosslinks, and sequencing the remaining DNA to map all genomic loci bound by protein (8; 37). Moreover, the HBsu antiserum will allow for ChIP-seq to be performed with the IPOD experiments to determine which loci HBsu specifically occupies in the wild type and m6A deficient strains. Subsequent RNA-seq experiments can be performed to determine if differences in protein occupancy between the strains result in differences in gene expression. This work will help to elucidate indirect mechanisms of m6A regulation of gene expression that arise from the direct m6A-dependent promoter regulation described above.

Indirect mechanisms of m6A-dependent regulation of gene expression independent of direct regulatory mechanisms. While my study of m6A-dependent regulation of gene expression in *B. subtilis* revealed possibilities for both direct and indirect mechanisms of regulation, my study of the regulatory functions of m6A in *S. pyogenes* strongly suggest an indirect role for DNA methylation in tuning gene expression (22). In Chapter II I established that loss of m6A from an active Type I RM system resulted in significant down regulation of 20 genes, a subset of which comprised genes in the core Mga regulon. Mga is a stand-alone transcriptional regulator that regulates ~10% of the *S. pyogenes* genome during exponential phase growth, including genes involved in host cell adhesion, internalization, and immune evasion phenotypes (12; 16). The *mga* gene encoding the Mga transcriptional regulator showed a $\log_2$ fold-change of -1.24 in the absence of m6A (22). While the decrease in expression of the transcriptional activator explains the reduced expression of genes within the regulon, the nearest Type I RM system recognition site occurs 800 base pairs upstream of *mga* within the coding region of another gene, making it an unlikely candidate for direct regulation of *mga* expression.

Indirect regulation of gene expression by a Type I RM system would not be unique to *S. pyogenes.* As discussed in the Chapter I, m6A from a Type I RM system in *M.*

tuberculosis also appears to indirectly affect gene expression, with no recognition sites proximal to differentially expressed genes (6). To date, no known mechanism for m6A-dependent indirect regulation of gene expression has been described. In order to better understand how indirect regulation occurs, IPOD-HR could provide genome-wide protein occupancy information in *S. pyogenes* (8)*.* Differences in protein occupancy could result in gene expression changes through differential binding of proteins in promoter or gene body regions that affect transcription initiation and elongation, respectively. Alternatively, more global changes in occupancy could result in varied chromosome conformation, resulting in changes in RNA-polymerase occupancy and gene expression. The relatively small number of differentially expressed genes in *S. pyogenes* suggests that the changes are less likely to be the result of global changes in chromosome conformation. To further elucidate the indirect mechanism of regulation, subsequent to IPOD-HR, locus specific pull-down experiments can be performed to identify differentially bound proteins at specific loci and gross chromosome structure changes can be determined by Chromosome Conformation Capture followed by deep sequencing (Hi-C). Determining the indirect mechanisms of m6A-dependent gene expression changes remains an open and largely unexplored area of research across all bacteria.

RNA modifications within genomic DNA. In Chapter IV I describe a plasmid-borne RNase HI protein that contributes to genome maintenance in the ancestral strain of *B. subtilis* NCIB 3610. RNase H proteins are ubiquitous throughout all three domains of life and function to remove the RNA component of RNA-DNA hybrids within genomic DNA (5; 24). Prior to research presented in Chapter IV, no bacterial species had been identified to carry active RNase HI, HII, and HIII proteins. Most bacteria encode active RNase HI and HII or RNase HII and HIII proteins (13). While genes encoding all three putative RNase H proteins were found in the chromosomes of some species, one of the three types of RNase H proteins was found to be inactive (25). The observation that organisms lack of all three active RNase H proteins has previously been explained by the redundant activity of RNase HI and RNase HIII enzymes, which recognize and cleave the same substrates (13; 21; 25; 26). However, in Chapter IV I showed that,

despite having activity on the same substrates, the RNase HI and HIII proteins in *B. subtilis* cleave the substrates at different locations within the hybrids, which could affect the accessibility and efficiency of downstream repair proteins (26). Thus, RNase HI and HIII enzymes could have varying efficiencies for resolving the overlapping substrates that occur during varied cellular processes. In the next section I describe open areas of research to further investigate if RNase HI and HIII are redundant in function and how these enzymes contribute to genome maintenance in *B. subtilis.*

Contribution of plasmid-borne RNase HI to genome maintenance. The different incision preferences of RNase HI and HIII from *B. subtilis* NCIB 3610 on ribopatch, Okazaki fragment, and hybridized RNA-DNA substrates could affect the accessibility and efficiency of downstream repair proteins that remove the remaining RNA, fill the gaps with DNA, and ligate the break in the DNA backbone (26). Variation of efficiencies in RNase HI and HIII-based repair pathways could in turn explain how both enzymes co-exist in the same genome. To test the efficiency of hybrid resolution, DNA polymerase I (Pol I) extension assays can be completed, as done previously in the Simmons lab, wherein various hybrid substrates are incubated with DNA polymerase I subsequent to incubation with RNase HI or RNase HIII to assay for the efficiency of extension from the RNase H treated hybrid substrates (26). In previous experiments it was found that treatment with RNase HIII stimulated Pol I activity on an Okazaki fragment substrate to a greater extent than treatment with RNase HII, suggesting that differences in incision pattern on various substrates might also affect the efficiency of repair between RNase HI and HIII (26).

In addition to challenging the model that RNase HI and HIII are truly functionally redundant, the work presented in Chapter IV provides important mechanistic insight into the phenotypic effects observed upon loss of RNase HI and HIII activity in *B. subtilis.* Similar to previous reports, I have shown that loss of RNase HI and HIII activities results in a cell elongation phenotype and constitutive expression of the DNA damage response in normally growing cells (9). Furthermore, using whole genome re-sequencing I identify the *pps* operon, encoding genes for production of the antibiotic

plipastatin (36), as a locus with decreased sequencing coverage in the strain lacking both RNase HI and HIII proteins, suggesting DNA replication and transcription conflicts at this region. The *pps* operon consists of five very long genes totaling ~38 kb in length that are all oriented in the head-on direction relative to DNA replication. In unpublished work from the Simmons and Freddolino labs, this region has also been shown to accumulate RNA-DNA hybrids in the lab strain *B. subtilis* PY79. In a separate set of unpublished experiments, we show that deletion of the entire *pps* operon results in loss of the constitutive SOS induction phenotype observed in cells lacking RNase HI and HIII, suggesting that replication conflicts at the *pps* operon are responsible for the replication conflicts that induce the SOS response. Intriguingly, despite the loss of SOS response induction, cells lacking both RNase HI and HIII activity are still elongated in the absence of the *pps* operon, suggesting that the cell elongation phenotype is either independent of the SOS response or that the microscopy-based reporter assay used to determine the status of the SOS response is not sensitive enough to account for the expression changes that result in cell elongation. While much of the work in Chapter IV establishes a foundation for future studies, further work will need to be done to understand the mechanism(s) of cell elongation and SOS induction upon loss of RNase HI and RNase HIII activity in *B. subtilis* NCIB 3610.

To my knowledge, all previous studies have limited the search of functional RNase H proteins to the chromosome (13). As shown in Chapter IV, while the ancestral strain *B. subtilis* NCIB 3160 does not encode a functional RNase HI protein on the chromosome, the plasmid encoded RNase HI is functional. Moreover, while the chromosomally encoded RNase HIII protein appears to be necessary for plasmid hyper-replication, the plasmid encoded RNase HI is not, and appears instead to contribute to genome maintenance and survival of cellular stress. The research presented in Chapter IV challenges the current opinion in the field that RNase HI and HIII are mutually exclusive and encourages researchers to investigate the contributions of plasmid-encoded RNase H proteins to genome maintenance in bacteria.

Conclusion

In this thesis I have investigated the physiological consequences of modifications to genomic DNA in the form of both DNA methylation and RNA misincorporation events. In Chapters II and III I describe how DNA methylation in Gram-positive *S. pyogenes* and *B. subtilis* regulates gene expression in both bacteria (Chapters II and III (22; 23)). This work provided some of the first investigation into the regulatory effects of DNA methylation outside of phase variation in Firmicutes, and demonstrated that DNA methylation could be an important contributor to virulence in other Gram-positive pathogens as well. In Chapter IV I show that the ancestral strain of *B. subtilis* encodes functional RNase HI, HII, and HIII enzymes for the removal of RNA incorporated into DNA, providing the first example of an organism that encodes all three active RNase H enzymes.

As described throughout this thesis, despite the importance and prevalence of DNA methylation across bacteria, few studies have investigated the regulatory effects of DNA methylation in Gram-positive bacteria outside of phase variable MTases in *Streptococcus* species. The Firmicutes phylum includes many important Gram-positive pathogens that represent significant public health burdens. My work strongly suggests that interference of DNA MTase activity would not only make pathogens such as *S. pyogenes* more susceptible to phage predation but can also severely attenuate the virulence potential of these pathogens. I predict that further study of DNA MTases across Firmicutes will reveal important regulatory roles in pathogenicity and development, providing strong candidates for vaccine development and therapeutic interventions in Gram-positive bacteria during a period of emerging antibiotic resistance.

References

1. Belitsky BR, Barbieri G, Albertini AM, Ferrari E, Strauch MA, Sonenshein AL. 2015. Interactive regulation by the Bacillus subtilis global regulators CodY and ScoC. *Mol Microbiol* 97:698-716
2. Bolker M, Kahmann R. 1989. The Escherichia coli regulatory protein OxyR discriminates between methylated and unmethylated states of the phage Mu mom promoter. *EMBO J* 8:2403-10
3. Brinsmade SR, Alexander EL, Livny J, Stettner AI, Segre D, et al. 2014. Hierarchical expression of genes controlled by the Bacillus subtilis global regulatory protein CodY. *Proc Natl Acad Sci U S A* 111:8227-32
4. Caldwell R, Sapolsky R, Weyler W, Maile RR, Causey SC, Ferrari E. 2001. Correlation between Bacillus subtilis scoC phenotype and gene expression determined using microarrays for transcriptome analysis. *J Bacteriol* 183:7329-40
5. Cerritelli SM, Crouch RJ. 2009. Ribonuclease H: the enzymes in eukaryotes. *FEBS J* 276:1494-505
6. Chiner-Oms A, Berney M, Boinett C, Gonzalez-Candelas F, Young DB, et al. 2019. Genome-wide mutational biases fuel transcriptional diversity in the Mycobacterium tuberculosis complex. *Nat Commun* 10:3994
7. Dillon SC, Dorman CJ. 2010. Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. *Nat Rev Microbiol* 8:185-95
8. Freddolino P, Goss T, Amemiya H, Tavazoie S. 2020. Dynamic landscape of protein occupancy across the Escherichia coli chromosome. *bioRxiv*
9. Fukushima S, Itaya M, Kato H, Ogasawara N, Yoshikawa H. 2007. Reassessment of the in vivo functions of DNA polymerase I and RNase H in bacterial cell growth. *J Bacteriol* 189:8575-83
10. Haakonsen DL, Yuan AH, Laub MT. 2015. The bacterial cell cycle regulator GcrA is a sigma70 cofactor that drives gene expression from a subset of methylated promoters. *Genes Dev* 29:2272-86
11. Hernday AD, Braaten BA, Low DA. 2003. The mechanism by which DNA adenine methylase and PapI activate the pap epigenetic switch. *Mol Cell* 12:947-57
12. Hondorp ER, McIver KS. 2007. The Mga virulence regulon: infection where the grass is greener. *Mol Microbiol* 66:1056-65
13. Kochiwa H, Tomita M, Kanai A. 2007. Evolution of ribonuclease H genes in prokaryotes to avoid inheritance of redundant genes. *BMC Evol Biol* 7:128
14. Kohler P, Marahiel MA. 1997. Association of the histone-like protein HBsu with the nucleoid of Bacillus subtilis. *J Bacteriol* 179:2060-4
15. Marinus MG, Casadesus J. 2009. Roles of DNA adenine methylation in host-pathogen interactions: mismatch repair, transcriptional regulation, and more. *FEMS Microbiol Rev* 33:488-503
16. McIver KS, Scott JR. 1997. Role of mga in growth phase regulation of virulence genes of the group A streptococcus. *J Bacteriol* 179:5178-87
17. Micka B, Marahiel MA. 1992. The DNA-binding protein HBsu is essential for normal growth and development in Bacillus subtilis. *Biochimie* 74:641-50

18. Mohapatra SS, Fioravanti A, Biondi EG. 2014. DNA methylation in Caulobacter and other Alphaproteobacteria during cell cycle progression. *Trends Microbiol* 22:528-35

19. Molle V, Fujita M, Jensen ST, Eichenberger P, Gonzalez-Pastor JE, et al. 2003. The Spo0A regulon of Bacillus subtilis. *Mol Microbiol* 50:1683-701

20. Mouammine A, Collier J. 2018. The impact of DNA methylation in Alphaproteobacteria. *Mol Microbiol* 110:1-10

21. Nowotny M, Gaidamakov SA, Ghirlando R, Cerritelli SM, Crouch RJ, Yang W. 2007. Structure of human RNase H1 complexed with an RNA/DNA hybrid: insight into HIV reverse transcription. *Mol Cell* 28:264-76

22. Nye TM, Jacob KM, Holley EK, Nevarez JM, Dawid S, et al. 2019. DNA methylation from a Type I restriction modification system influences gene expression and virulence in Streptococcus pyogenes. *PLoS Pathog* 15:e1007841

23. Nye TM, van Gijtenbeek LA, Stevens AG, Schroeder JW, Randall JR, et al. 2020. Methyltransferase DnmA is responsible for genome-wide N6-methyladenosine modifications at non-palindromic recognition sites in Bacillus subtilis. *Nucleic Acids Res*

24. Ohtani N, Haruki M, Morikawa M, Kanaya S. 1999. Molecular diversities of RNases H. *J Biosci Bioeng* 88:12-9

25. Randall JR, Hirst WG, Simmons LA. 2017. Substrate specificity for bacterial RNase HII and HIII is influenced by metal availability. *J Bacteriol* 200:pii: e00401-17

26. Randall JR, Nye TM, Wozniak KJ, Simmons LA. 2019. RNase HIII Is Important for Okazaki Fragment Processing in Bacillus subtilis. *J Bacteriol* 201: pii: e00686-18.

27. Ross MA, Setlow P. 2000. The Bacillus subtilis HBsu protein modifies the effects of alpha/beta-type, small acid-soluble spore proteins on DNA. *J Bacteriol* 182:1942-8

28. Sanchez-Romero MA, Casadesus J. 2020. The bacterial epigenome. *Nat Rev Microbiol* 18:7-20

29. Schroeder JW, Randall JR, Hirst WG, O'Donnell ME, Simmons LA. 2017. Mutagenic cost of ribonucleotides in bacterial DNA. *Proc Natl Acad Sci U S A* 114:11733-8

30. Shivers RP, Sonenshein AL. 2004. Activation of the Bacillus subtilis global regulator CodY by direct interaction with branched-chain amino acids. *Mol Microbiol* 53:599-611

31. Sonenshein AL. 2005. CodY, a global regulator of stationary phase and virulence in Gram-positive bacteria. *Curr Opin Microbiol* 8:203-7

32. Srikhanta YN, Dowideit SJ, Edwards JL, Falsetta ML, Wu HJ, et al. 2009. Phasevarions mediate random switching of gene expression in pathogenic Neisseria. *PLoS Pathog* 5:e1000400

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

34. Srikhanta YN, Gorrell RJ, Steen JA, Gawthorne JA, Kwok T, et al. 2011. Phasevarion mediated epigenetic gene regulation in Helicobacter pylori. *PLoS One* 6:e27569

35. Srikhanta YN, Maguire TL, Stacey KJ, Grimmond SM, Jennings MP. 2005. The phasevarion: a genetic system controlling coordinated, random switching of expression of multiple genes. *Proc Natl Acad Sci U S A* 102:5547-51

36. Tsuge K, Ano T, Hirai M, Nakamura Y, Shoda M. 1999. The genes degQ, pps, and lpa-8 (sfp) are responsible for conversion of Bacillus subtilis 168 to plipastatin production. *Antimicrob Agents Chemother* 43:2183-92

37. Vora T, Hottes AK, Tavazoie S. 2009. Protein occupancy landscape of a bacterial genome. *Mol Cell* 35:247-53

38. Williams JS, Kunkel TA. 2014. Ribonucleotides in DNA: origins, repair and consequences. *DNA Repair (Amst)* 19:27-37