Mechanisms of Genome Maintenance and Mutagenesis in *Bacillus subtilis*

by

Jeremy W. Schroeder

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
2015

Doctoral Committee:

Associate Professor Lyle A. Simmons, Chair
Associate Professor Matthew R. Chapman
Associate Professor Patrick O'Brien
Assistant Professor Andrzej Wierzbicki

To laugh often and much; to win the respect of intelligent people and the affection of children; to earn the appreciation of honest critics and endure the betrayal of false friends; to appreciate beauty; to find the best in others; to leave the world a bit better, whether by a healthy child, a garden patch or a redeemed social condition; to know even one life has breathed easier because you have lived. This is to have succeeded.

Ralph Waldo Emerson (1803 - 1882)

For my daughters and wife.

ACKNOWLEDGEMENTS

There are few moments in a person's life they can definitively say, without question, were life-changing. One such moment in my life to date was dinner at Palio's following my interview with MCDB, sitting next to Lyle. I applied to MCDB with no intention of studying basic biology in a bacterial model system, but over dinner Lyle had convinced me to rotate in his lab that Summer. I cannot thank him enough for the complete joy of being a part of his lab's research through the past six years. In a conversation I had on the bus with a fellow MCDB graduate student who knew Lyle well, she said of him, "Every time I speak with Lyle I am left feeling," she paused to consider the right word, "valued." I agreed that this was the perfect way to summarize how my interactions with Lyle have played out throughout my thesis work. Thank you, Lyle, for being a dedicated mentor and friend, for inspiring, guiding, and giving me the freedom to ask my own questions and answer them by any means necessary.

To the other members of the Simmons Lab, past and current, thank you for your support and advice. Each and every one of you has contributed to making my time in graduate school the best years of my life. Brian, we have had a very rewarding relationship, both on a personal and a professional level. Together we have been able to have some of the most serious conversations two people can share in the most hilarious way imaginable. That is something I will always cherish. Justin Lenhart, you are a fierce and loyal friend. Your insights on life after graduate school and critical interpretation of data have been invaluable. Taylor and Pete, never stop being philosophers. I believe there is no greater pursuit in life than the pursuit of wisdom,

and of all the people I know, you two are the most likely someday to find it. I am deeply thankful for all our conversations. Justin Randall, I have never seen you be anything but kind to anyone. The world is a dangerous place for you, so be careful out there, buddy. Will, thanks for all the time we spent together doing mutation accumulation lines. Lindsay, thank you for the shared trips to Comet and some great chats. Your friendship is truly irreplaceable.

To my Thesis Committee, I have enjoyed hearing your feedback and receiving your guidance throughout my candidacy. Thank you for the critical role you have played in my development as a scientist.

To my dearest Heather, you have been my rock, unwavering in support and love for me through graduate school, work and the arrival of our beautiful daughters into the world. You have kept me securely grounded in what is most important in life. I cannot imagine what the next eighty years will bring, but I know that with you they will leave me craving eighty more. Thank you for everything. Ever thine, ever mine, ever ours. UU.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

ABSTRACT

Mechanisms of Genome Maintenance and Mutagenesis in *Bacillus subtilis*

by

Jeremy W. Schroeder

Chair: Lyle A. Simmons

Deoxyribonucleic acid is used in all domains of life for the storage, replication and transmission of genetic information. Accurate replication of genomic DNA is essential to maintain gene function, and mechanisms have evolved to limit the mutagenic potential of DNA replication errors and impart stability to the genome. DNA replication and repair are well-studied, but there remain significant unanswered questions. In this dissertation I present my findings regarding the following questions: How does the mismatch repair protein MutS detect a single mismatch among millions of correctly paired nucleotides? Are there genome contexts that predispose loci to mutagenesis? What are the effects of unrepaired ribonucleotides in DNA on genome stability? Using a combination of biochemical, cell biological, genetic and genomic approaches, I have leveraged the power of the *Bacillus subtilis* model organism to show that mismatches produced during DNA replication are detected by MutS near their site of synthesis. This process is dependent on MutS interaction with replisome subunits. Using mutation accumulation lines, I have shown that local sequence contexts, and not global factors such as gene presence, direction of transcription and transcript abundance, affect replication error rate. I also reveal using mutation accumulation lines that

ribonucleotides in bacterial DNA that go unrepaired cause GC→AT transitions. This is likely due to nucleotide excision repair effecting the removal of ribonucleotides in genomic DNA. This work provides important, novel insight to the fields of DNA mismatch repair and mutagenesis, and opens new avenues of exploration into the developing field of ribonucleotide excision repair.

CHAPTER I

Introduction

1.1 Introduction

DNA is a remarkably stable medium for the storage and replication of genetic information, and it carries out this role in all living cells. Base-pairing errors (mismatches) rarely occur during DNA replication, but they carry a high potential for mutagenesis when they are produced. In order to minimize the potential mutagenic consequences of mismatches made by replicating DNA polymerases, DNA mismatch repair has evolved to detect and replace them before they cause mutation. Another type of error made during DNA replication is sugar error, in which a ribonucleotide (rNTP) is inserted by the DNA polymerase rather than a deoxyribonucleotide. Sugar error is $\approx$ 26,000 times more common than mismatches, and because RNA is more prone to hydrolysis of the phosphodiester backbone than DNA, misinserted rNMPs may be a significant source of genome instability if they go unrepaired. DNA replication can be adversely affected by rNMPs in the template strand, rNMPs can increase the likelihood of mutation if they go unrepaired by ribonucleotide excision repair, and any nicks present in DNA due to hydrolysis at rNMPs could also lead to replication stress in

Modified versions of this chapter were published in *Microbiology and Molecular Biology Reviews* by Justin Lenhart, Brian Walsh, Lyle Simmons and I (*Lenhart et al*., 2012), and in *Critical Reviews in Biochemistry and Molecular Biology* by Lindsay Matthews, Justin Randall, Lyle Simmons and I (*Schroeder et al*., 2014).

the form of lethal double-stranded breaks. This dissertation focuses on mechanisms that have evolved to limit the mutagenic effects of mismatches and rNMPs placed in genomic DNA during DNA replication, factors that predispose some genomic loci to mismatch formation, and the effects on genome stability should rNMPs go unrepaired.

1.2 Mismatch Repair

Mismatch repair (MMR) is a process used to correct base-pairing errors made during DNA replication [reviewed in (*Kunkel and Erie*, 2005; *Schofield and Hsieh*, 2003)]. In *Escherichia coli*, MutS binds a mismatch, followed by recruitment of MutL [reviewed in (*Kunkel and Erie*, 2005; *Schofield and Hsieh*, 2003)]. MutL recruits and activates the MutH endonuclease, causing MutH to nick the unmethylated and thus nascent strand bearing the mismatch [reviewed in (*Kunkel and Erie*, 2005; *Schofield and Hsieh*, 2003)]. The helicase UvrD is then loaded at the site of the nick, where it unwinds the DNA strand containing the error [reviewed in (*Kunkel and Erie*, 2005; *Schofield and Hsieh*, 2003)]. The error-containing strand is then degraded by one of many exonucleases, depending on the polarity of the excised strand (*Viswanathan et al.*, 2001). The resulting gap is synthesized by Pol III, and the nick is sealed by DNA ligase [reviewed in (*Kunkel and Erie*, 2005; *Schofield and Hsieh*, 2003)]. The MMR pathway in *E. coli* requires MutS, MutL, MutH, UvrD, and Dam methylase to methylate adenine in the d(GATC) sequence, allowing for identification of the newly replicated, mismatch-bearing strand.

Bacillus subtilis contains the highly conserved MutS and MutL proteins (*Ginetti et al.*, 1996); however, it lacks MutH, Dam, and a clear UvrD ortholog known to function in mismatch repair (*Culligan et al.*, 2000; *Eisen*, 1998; *Eisen and Hanawalt*, 1999). MutS is the sensor that recognizes the mismatch (*Simmons et al.*, 2008), whereas MutL is the linker, which functions to link the remaining proteins in the pathway together to allow for efficient repair of replication errors in *B. subtilis* [reviewed in (*Kunkel and Erie*, 2005; *Schofield and Hsieh*, 2003)]. Because neither *dam* nor *mutH* is present within the *B.*

subtilis genome, it has been hypothesized that *B. subtilis* uses a methylation-independent mismatch repair pathway, unlike the mismatch repair pathway characterized for *E. coli*. Consistent with this idea, d(GATC) sequences in *B. subtilis* and *Staphylococcus aureus* are not methylated, suggesting that a functional analog of Dam is not present in these organisms (*Dreiseikelmann and Wackernagel*, 1981). It is also important to note that the methylation-directed mismatch repair pathway characterized for *E. coli* is absent from most bacteria, and it is hypothesized that most bacteria utilize a methylation-independent mismatch repair system (*Pillon et al.*, 2010), just as in eukaryotic systems it is hypothesized that identifying the newly replicated strand in *B. subtilis* and other organisms that lack a methylation-directed signal relies on strand discontinuities located at or near DNA replication forks *in vivo* [reviewed in (*Larrea et al.*, 2010)].

1.2.1 Mismatch Detection by MutS is Coupled to DNA Synthesis

Several lines of evidence show that mismatch repair proteins assemble into complexes at the site of DNA synthesis (*Kleczkowska et al.*, 2001; *Simmons et al.*, 2008; *Smith et al.*, 2001). Many replication proteins localize in cells as discrete foci marking the site of DNA synthesis (*Berkmen and Grossman*, 2006, 2007; *Dervyn et al.*, 2001; *Lemon and Grossman*, 1998; *Meile et al.*, 2006). Some of the first evidence suggesting that mismatch repair is coupled to DNA synthesis came from visualizing the formation of mismatch repair foci in human cell culture and in live *B. subtilis* cells (*Kleczkowska et al.*, 2001; *Simmons et al.*, 2008; *Smith et al.*, 2001). In *B. subtilis*, MutS-GFP and MutL-GFP fusion alleles expressed from their native promoter were shown to form foci in a small proportion of cells ($\approx$ 5 to 10%) during exponential-phase growth, and this proportion was increased when cells were challenged with the mismatch-forming agent 2-aminopurine (2-AP) (*Simmons et al.*, 2008; *Smith et al.*, 2001). It should be noted that the *mutS-gfp* allele is functional and the *mutL-gfp* allele is nonfunctional with

respect to mutant occurrence; however, the MutL-GFP fusion protein does form foci in response to mismatches, suggesting active recruitment in response to replication errors (*Smith et al.*, 2001). The focus formation response by MutS-GFP and MutL-GFP requires ongoing DNA replication in *B. subtilis* (*Smith et al.*, 2001). When MutS-GFP and MutL-GFP form foci, they preferentially localize to the midcell area, the site in the cell where DNA synthesis occurs (*Smith et al.*, 2001). Moreover, MutS-YFP foci colocalize with the replisome (DnaX-CFP) in $\approx$ 48% of live cells (*Smith et al.*, 2001). These data suggest that mismatch repair proteins are coupled to or function near the site of DNA replication in *B. subtilis*.

In support of a mechanism that couples mismatch repair to DNA synthesis in *B. subtilis*, it was recently shown that MutS alters the subcellular localization of the essential DNA polymerase DnaE in response to replication errors (*Klocko et al.*, 2011). In this work, ectopically expressed DnaE-GFP foci decreased in cells challenged with 2-AP or in cells that bore a proofreading-deficient *polC* allele (mut-1) (*Klocko et al.*, 2011). The decrease in DnaE-GFP foci required MutS, suggesting that the effect takes place at the step of mismatch recognition. Protein far-Western blot experiments demonstrated that both mismatch repair proteins MutS and MutL directly bind DnaE, suggesting that a strong interaction between these proteins may exist *in vivo* (*Klocko et al.*, 2011). Thus, MutS detection of mismatches affects the subcellular localization of an essential DNA polymerase in *B. subtilis*, suggesting that MutS is able to signal to or perturb the replication complex following mismatch identification in live cells (*Klocko et al.*, 2011). One candidate protein for recruiting MutS to active replication forks is the processivity β clamp, discussed below.

Work in *S. cerevisiae* made use of a functional fusion of the MutS homolog MSH6 to an S-phase-specific cyclin to restrict MSH6 protein expression to S phase. In this experiment, the MSH6–S-phase cyclin fusion protein conferred wild-type levels of mismatch repair (*Hombauer et al.*, 2011b). In contrast, when MSH6 expression was

limited to G2/M by fusion to a G2/M-specific cyclin, mismatch repair was nonfunctional (*Hombauer et al.*, 2011b). In addition, live-cell imaging using fully functional fluorescent fusions to MSH6 showed colocalization with the replication machinery in live *S. cerevisiae* cells (*Hombauer et al.*, 2011a). Colocalization of MSH6 to replication forks corresponded to $\approx$ 15% of MMR in *S. cerevisiae*, and interestingly, it was required for functional mismatch repair in the absence of ExoI (*Hombauer et al.*, 2011a). This work shows that mismatch repair is coupled to DNA replication in *S. cerevisiae*, demonstrating that the coupling of mismatch repair to DNA replication is conserved in eukaryotes and is not unique to *Bacillus* (*Hombauer et al.*, 2011a,b).

1.2.2 Involvement of β Clamp in Mismatch Repair

The DNA replication processivity factor β clamp is important for linking mismatch repair foci to the DNA replication status (*Dupes et al.*, 2010; *Simmons et al.*, 2008). *B. subtilis* MutS contains a five-amino-acid motif originally identified through bioinformatic analysis as a putative β clamp-binding site (*Dalrymple et al.*, 2001). Residues 810QLSFF814 provide a mostly hydrophobic plug that fits into a hydrophobic cleft on the β clamp (*Bunting et al.*, 2003; *Dalrymple et al.*, 2001). Replacement of this motif with five alanine residues caused an $\approx$ 40-fold increase in measurements of mutant frequency, and a mutant form of MutS fused to GFP and bearing a replacement of QLSFF with five alanine residues was reduced $\approx$ 3-fold for focus formation in response to 2-AP challenge (*Simmons et al.*, 2008). Biochemical analysis shows that the β clamp binds to peptides bearing the QLSFF residues (*Simmons et al.*, 2008). Therefore, the MutS QLSFF motif is critical for interaction between MutS and the β clamp for efficient mismatch repair in *B. subtilis* (*Simmons et al.*, 2008).

Although mutation of the β clamp-binding motif in MutS increased the frequency of mutants of live cells, deletion of the 58-residue unstructured C-terminal region bearing this motif (resulting in the MutS800 protein) nearly abolished the interaction between

B. subtilis MutS and the β clamp in vitro (*Simmons et al.*, 2008). Importantly, purified MutS800 protein bound to a mismatch at the same level as wild-type MutS in vitro, suggesting that removal of the C-terminal 58 amino acids does not diminish mismatch binding (*Simmons et al.*, 2008). Although it was expressed at wild-type levels, the *mutS800* allele conferred a spontaneous mutant frequency close to that for a strain disrupted in the *mutS* gene. MutS800 fused with GFP failed to localize as a focus in response to 2-AP, and YFP fused to only the C-terminal 58 amino acids of MutS bearing the β clamp-binding motif was sufficient to target YFP for localization when the fusion was overexpressed (*Simmons et al.*, 2008). These data support the hypothesis that the β clamp aids in the formation of MutS-GFP foci in response to mismatches in vivo (*Simmons et al.*, 2008). This work also found that a conditional allele of *dnaN* (coding for the β clamp) which caused a partial defect in mismatch repair had a reduced capacity for supporting MutS-GFP focus formation *in vivo* (*Simmons et al.*, 2008). In addition, intragenic suppressors of the temperature-sensitive phenotype caused by this β clamp allele (*dnaN5*) maintained defects in mismatch repair while rescuing the DNA replication defect conferred by this allele. These data further support a role for the β clamp in mismatch repair in *B. subtilis* (*Dupes et al.*, 2010).

An important role for the β clamp (DnaN) in mismatch repair is strongly supported by studies of *Bacillus anthracis*. Interestingly, *B. anthracis* contains two *dnaN* genes: *dnaN1* and *dnaN2* (*Yang and Miller*, 2008). Both *dnaN*-encoded clamps support growth *in vivo*, yet deletion of *dnaN1* confers a rate of spontaneous mutagenesis that is identical to that of *B. anthracis* cells disrupted for mismatch repair (*Yang and Miller*, 2008). These results show that both *dnaN* genes allow for proper DNA synthesis; however, only *dnaN1* is capable of functioning in mismatch repair (*Yang and Miller*, 2008). These data further establish a role for the β clamp in correction of DNA replication errors in the genus *Bacillus* (*Yang and Miller*, 2008). The data from *B. subtilis* and *B. anthracis* show that MutS interaction with the β clamp is crucial for mismatch

repair in some Gram-positive organisms, and it may be found to be important in other organisms that lack a Dam-directed repair pathway, although this remains to be established. The involvement of the β clamp in mismatch repair in Dam-directed systems such as that of *E. coli* is unclear and requires further study (*López de Saro et al.*, 2003, 2006; *López de Saro and O'Donnell*, 2001; *Pluciennik et al.*, 2009).

1.2.3 *B. subtilis* MutL is a Latent Endonuclease

One of the major differences between *E. coli* and *B. subtilis* is that *B. subtilis* MutL is an endonuclease and *E. coli* MutL is not (*Kadyrov et al.*, 2006b; *Pillon et al.*, 2010). In the *E. coli* methylation-directed mismatch repair system, MutH endonuclease activity is required for mismatch correction (*Ahrends et al.*, 2006; *Hall and Matson*, 1999). The endonuclease active site in *B. subtilis* MutL is identical to the active site in human and *Saccharomyces cerevisiae* MutLα, suggesting a strong conservation in mechanism between *B. subtilis* and eukaryotic organisms (*Kadyrov et al.*, 2006b). Recently, the crystal structure of the endonuclease domain of *B. subtilis* MutL was solved (*Pillon et al.*, 2010). Two critical observations came from this work. First, MutL contains a zinc-binding loop, and mutations that abolish zinc binding inactivate mismatch repair *in vivo* (*Pillon et al.*, 2010). The zinc-binding loop is hypothesized to play a structural role in allowing for the proper positioning of DNA into the active site for subsequent cleavage (*Pillon et al.*, 2010). Second, a β clamp-binding site with the sequence 487QEMIVP492 was identified in the C-terminal domain of MutL (*Pillon et al.*, 2010), and this site was indeed shown to bind the β clamp (*Pillon et al.*, 2011). One favored mode of action is that the β clamp binds to the C terminus of MutL, opening the active site and allowing for DNA cleavage during repair (*Pillon et al.*, 2011). Interaction between MutL and the β clamp may function to both position and activate MutL for cleavage of the mismatch-containing strand. In support of this idea, mutation of the β clamp-binding motif blocks β clamp binding *in vitro* and prevents mismatch repair *in*

vivo (*Pillon et al.*, 2010, 2011). In contrast, mutations to the homologous site in *E. coli* MutL have a less pronounced effect on mismatch repair *in vivo* (*Pillon et al.*, 2011). Thus, it is attractive to consider that the β clamp may help to orient MutL to the nascent strand by directing MutL endonuclease activity. Once the newly synthesized strand is nicked, the nick site could then serve as an entry point for other repair proteins involved in mismatch correction in *B. subtilis*, including the putative 5′ $\rightarrow$3′ exonuclease WalJ (see Chapter III and *Yang et al.* 2013). Although this is an attractive model, these steps have yet to be shown experimentally. We provide a current model for mismatch repair in *B. subtilis* in Figure 1.1.

1.2.4 WalJ is a Putative 5′ $\rightarrow$3′ Mismatch Repair Exonuclease

A transposon-insertion mutagenesis screen using papillation as an experimental readout for mutation rate revealed the gene *walJ* (formerly *yycJ*) to be involved in mutation suppression in *Bacillus anthracis* (*Yang et al.*, 2011a). Disruption of *walJ* yielded an increased mutation rate and a mutation spectrum suggestive of involvement in MMR (*Yang et al.*, 2011a). Subsequent purification and biochemical characterization of WalJ showed the protein to be a Mn^{2+}-dependent 5′ $\rightarrow$3′ DNA exonuclease (*Yang et al.*, 2013). These results provide evidence a role for WalJ as an exonuclease functioning in MMR, but deletion of both *walJ* and *mutS* in *B. anthracis* gave rise to a rifampin-resistant mutant frequency $\approx$ 5-fold higher than deletion of either *mutS* or *walJ* alone (*Yang et al.*, 2013), so MMR and *walJ* are clearly not epistatic in *B. anthracis*. It is likely, therefore, that WalJ may play other roles in mutation suppression in *B. anthracis* than only in MMR. Loss of *walJ* in *B. subtilis* also caused increased papillation, but whether this was due to loss of MMR *per se* was not tested (*Yang et al.*, 2013).

1.3 Ribonucleotide Excision Repair (RER)

Ribonucleotide excision repair (RER) refers to the removal of ribonucleotides (rNMPs) and their replacement with dNMPs. Any process that removes rNMPs from DNA can be referred to as RER; here I will focus on the RNase HII-dependent pathway because this represents the dominant process for removing rNMPs from genomic DNA. The RER pathway for *S. cerevisiae* has been reconstituted with purified proteins to understand the discrete steps (*Sparks et al.*, 2012). During this process, RNase H2 nicks a single ribonucleotide generating a 5′-PO_4-, the processivity clamp PCNA and Pol δ or ϵ catalyze strand displacement synthesis followed by removal of the RNA-containing flap by FEN1 or Exo I. The nick is then sealed by DNA ligase to complete RER (*Sparks et al.*, 2012).

In bacteria, RER is less clear and we expect considerable redundancy in this pathway. Woodgate and co-workers have gained insight into RER in *E. coli* by taking advantage of the propensity of UmuCY11A to readily incorporate rNMPs (*Vaisman et al.*, 2014). In *E. coli* RER, RNase HII nicks at the rNMP generating a 5′-PO_4- rNMP at the nick. DNA pol I, by nick translation, can remove the RNA-containing strand while filling in the gap with dNMPs (*Vaisman et al.*, 2014). The nick is then sealed by DNA ligase to finish repair. In *E. coli*, pol I was shown to have a critical role in RER although pol III can function in the absence of pol I (*Vaisman et al.*, 2014). Furthermore, pol I is most-likely responsible for removing the 5′ flap although redundancy with other 5′ flap endo- and 5′ to 3′ single stranded exonucleases capable of providing the same function has been shown (*Vaisman et al.*, 2014).

1.3.1 Ribonucleotide Incorporation by DNA Polymerases

The ability of DNA polymerases to discriminate between rNTPs and dNTPs has been extensively studied *in vivo* [reviewed in (*Brown and Suo*, 2011; *Joyce*, 1997)]. The

general feature that contributes to sugar selection in the DNA polymerase active site is "exclusion" of the 2′-OH on the ribose sugar moiety by a tyrosine, phenylalanine or glutamic acid residue side chain 1.2 (*Kasiviswanathan and Copeland*, 2011; *Astatke et al.*, 1998; *Beck et al.*, 2002; *Bonnin et al.*, 1999; *Brown et al.*, 2010; *Cases-Gonzalez et al.*, 2000; *DeLucia et al.*, 2003; *Gao et al.*, 1997; *Gardner and Jack*, 1999; *Patel and Loeb*, 2000; *Yang et al.*, 2002). The use of a steric gate residue side chain to limit rNTP utilization by DNA polymerases is common, although some DNA polymerases from the X-family use a segment of the protein backbone to reduce rNTP incorporation (*Garcia-Diaz et al.*, 2005; *Pelletier et al.*, 1994; *Sawaya et al.*, 1997; *Brown et al.*, 2010). In addition, DNA polymerase domains change from an "open" to "closed" conformation during catalysis and the dynamics of this transition can be altered in the presence of rNTPs providing another mechanism for sugar discrimination (*Joyce et al.*, 2008). Interestingly, depending on the type of DNA polymerase and the specific rNTP:dNTP pair, sugar discrimination can range from just over 2-fold discrimination against rNTPs relative to dNTPs to over a million-fold preference for dNTPs [reviewed in (*Brown and Suo*, 2011; *Joyce*, 1997)].

Most polymerases are also limited in the number of rNTPs that can be incorporated consecutively, which prevents long stretches from occuring even when rNTPs are plentiful (*Astatke et al.*, 1998; *Zhu and Shuman*, 2005; *Ruiz et al.*, 2003; *Nick McElhinny and Ramsden*, 2003). This is not related to the steric gate, but instead stems from the developing RNA/DNA hybrid adopting an A-form conformation. This demonstrates an additional mechanism by which ribonucleotide incorporation is minimized (*Astatke et al.*, 1998).

In bacteria the main replicative DNA polymerase is a family C-type DnaE or PolC enzyme [for review (*McHenry*, 2011; *Johnson and O'Donnell*, 2005)], while in eukaryotes the leading strand polymerase, pol ϵ and the lagging strand polymerase, pol δ are B-family replicases (*Braithwaite and Ito*, 1993; *Pursell et al.*, 2007). Studies in

Saccharomyces cerevisiae have demonstrated that rNMPs are incorporated into genomic DNA *in vivo* and *in vitro* by eukaryotic pols α, δ and ϵ (*Nick McElhinny et al.*, 2010; *Nick McElhinny et al.*, 2010; *Sparks et al.*, 2012; *Williams and Kunkel*, 2014). These studies have shown that approximately 10,000 rNMP misincorporations are expected during a single round of DNA replication for *S. cerevisiae* (*Nick McElhinny et al.*, 2010; *Nick McElhinny et al.*, 2010). Extrapolating the rNMP incorporation rates for *S. cerevisiae* and mice to humans allows prediction that $\approx$ 3 million rNMP errors are made per round of replication for a human cell (*Nick McElhinny et al.*, 2010; *Reijns et al.*, 2012). Therefore, in eukaryotes, rNMPs embedded into genomic DNA represent the most abundant nucleotide in need of repair.

Using *E. coli* DNA polymerase III (pol III), as a model for C-family replicases, it was shown that a ribonucleotide incorporation event can be expected every 2,300 bps (*Yao et al.*, 2013). Considering the rate of ribonucleotide incorporation *in vitro*, Yao *et al.* expect nearly 2,000 misincorporation events per round of replication for the *E. coli* genome (*Yao et al.*, 2013). If we consider that *E. coli* pol III makes a dNTP base pairing error only once every $\approx$ 6.6 rounds of replication, it becomes clear that sugar errors are by far the most frequent mistake made by bacterial replicative DNA polymerases (*Lee et al.*, 2012). Of course the number of 2,000 per round of replication for *E. coli* pol III does not take into consideration the possible contributions of DNA pol I, II, IV or V to sugar error during genome replication.

Of the roughly 2,000 errors made by *E. coli* pol III approximately 1,500 of these are expected to be rAMP followed by over 300 misinsertions of rCMP, a little over 100 misinsertions of rGMP and only 6 rUMP (*Yao et al.*, 2013). Therefore, pol III shows excellent exclusion of rUTP during replication. *in vivo* studies of the low GC Gram-positive bacterium *Bacillus subtilis* showed that rNMPs are indeed incorporated into genomic DNA (*Yao et al.*, 2013). *B. subtilis* uses both DnaE and PolC for chromosomal replication and the contributions of each of these replicases to sugar

error are not yet established. Given that *S. cerevisiae*, *E. coli*, mouse, *B. subtilis* and *Chlamydophila pneumoniae* have all shown a significant amount of rNMP misinsertions either *in vitro*, *in vivo* or both, it seems likely that ribonucleotide incorporation by replicative DNA polymerases is a biologically conserved event (*Lu et al.*, 2012b; *Nick McElhinny et al.*, 2010; *Nick McElhinny et al.*, 2010; *Reijns et al.*, 2012; *Yao et al.*, 2013).

1.3.2 Removal of rNMPs from Genomic DNA

RNase H enzymes cleave RNA in an RNA/DNA hybrid. Bacterial RNase H enzymes have established roles in DNA replication, DNA repair and transcription [for review (*Kogoma*, 1997)]. RNases H from bacteria are categorized into type 1 and type 2 enzymes based on conservation of their primary structure [for review (*Tadokoro and Kanaya*, 2009)]. RNase HI is a type 1 RNase H, while type 2 enzymes include RNases HII and HIII. Bacteria usually contain either RNases HI and HII or RNases HII and HIII [for review (*Ohtani et al.*, 1999b; *Tadokoro and Kanaya*, 2009)] (Figure 1.3A). Here we discuss a few examples of RNase H enzymes. It should be noted that prokaryotic RNase H enzymes show impressive diversity in their activity and substrate recognition [for review (*Ohtani et al.*, 1999b; *Tadokoro and Kanaya*, 2009)]. In addition to the RNase H system employed for the removal of RNA from genomic DNA, topo I removes single rNMPs in *S. cerevisiae* and nucleotide excision repair (NER) can remove both single rNMPs and stretches of rNMPs in *E. coli* (*Kim et al.*, 2011; *Vaisman et al.*, 2013; *Cai et al.*, 2014). In this section, we review the processes impacting removal of rNMPs incorporated into genomic DNA.

1.3.3 Bacterial RNase HI

The *E. coli rnhA* gene encodes RNase HI, which cleaves stretches of RNA in an RNA/DNA hybrid, generating a 5′-PO_4- (*Miller et al.*, 1973; *Berkower et al.*, 1973). RNase HI enzymes are divided into two groups: (1) the archaeal-type hybrid binding domain

(HBD) HBD-RNase HI and (2) the *E. coli* type RNase HI (*Tadokoro and Kanaya*, 2009). The HBD recognizes the RNA/DNA hybrid and is important for enzyme activity (*Nowotny et al.*, 2008, 2005) (Figure 1.3B-C). In bacterial systems *E. coli* RNase HI is the best-understood enzyme with respect to its biochemical activity and the physiological consequences that ensue in *rnhA*-deficient cells (*Ogawa and Okazaki*, 1984; *Ogawa et al.*, 1984; *Hong and Kogoma*, 1993). *E. coli* RNase HI cleaves stretches of RNA with a minimum of four consecutive ribonucleotides (*Hogrefe et al.*, 1990). The major *in vivo* functions of RNase HI in *E. coli* are to remove R-loops resulting from transcription and to aid pol I in processing Okazaki fragments during lagging strand synthesis (*Kogoma et al.*, 1993; *Ogawa and Okazaki*, 1984; *Hong et al.*, 1995, 1996; *Kogoma and von Meyenburg*, 1983). In contrast, ribonucleotide incorporation errors by replicative DNA polymerases are unlikely to generate a substrate for RNase HI, although “ribo patches” incorporated by Y-family polymerases could generate a suitable substrate for RNase HI. It should be noted that several bacteria lack RNase HI and instead have RNase HIII. RNase HIII appears to provide an analogous function to RNase HI (see Section 1.3.5).

1.3.4 Bacterial RNase HII

E. coli rnhB (RNase HII) was first identified as a multicopy suppressor of the *rnh*-temperature-sensitive phenotype (Itaya, 1990). Enzymatically, it cleaves both single ribonucleotides and stretches of consecutive ribonucleotides (**?**). Therefore, RNase HII could serve as a backup or in conjunction with RNase HI in the processing of Okazaki fragments, while also providing the role of removing single rNMPs incorporated during DNA replication. Interestingly, most complete archaeal genomes show a single RNase H enzyme, RNase HII, demonstrating the strong conservation of this enzyme among prokaryotes (Ohtani et al., 2004). Deletion of the *rnhB* gene (RNase HII) from prokaryotes or inactivation of RNase H2 from eukaryotes shows the accumulation of ribonucleotides *in vivo*, as judged by alkaline sensitivity of isolated genomic DNA (*Lu*

et al., 2012b; *McDonald et al.*, 2012; *Nick McElhinny et al.*, 2010; *Yao et al.*, 2013; *Reijns et al.*, 2012). In bacterial systems loss of RNase HII does not impart a growth disadvantage or a strong phenotype. In *B. subtilis* loss of RNase HII activity increases overall spontaneous mutagenesis about 2-fold, while in *E. coli* loss of RNase HII does not affect mutation rate (*Yao et al.*, 2013).

The focus of much of this introduction is on bacterial enzymes, however it should be noted that an RNase H2 defect is embryonic lethal in mice, and people with the rare autosomal disorder Aicardi-Goutieres syndrome (AGS) have neurological dysfunction linked to mutations in genes important for nucleic acid metabolism including defects in RNase H2 (*Reijns and Jackson*, 2014; *Figiel et al.*, 2011; *Crow et al.*, 2006b,a). Therefore, the phenotype resulting from RNase H2 inactivation is more striking in mammals, which likely stems from eukaryotes relying more heavily on RNase H2 for RNA removal *in vivo* than prokaryotes. Furthermore, it is worth noting that superimposition of crystal structures of Thermotoga maritima RNase HII and Bacillus stearothermophilus RNase HIII allowed identification of amino acid residues in the catalytic subunit of *S. cerevisiae* RNase H2 responsible for each of its dual roles in hydrolysis of singly-misinserted rNMPs in DNA and digestion of R-loops. Two amino acid substitutions in its catalytic subunit yielded *S. cerevisiae* RNase H2 with greatly reduced efficiency for hydrolysis of covalent junctions between rNMPs and DNA, but maintained the ability to efficiently hydrolyze the RNA strand of an RNA/DNA hybrid (Chon et al., 2013). The fact that these substitutions in the *S. cerevisiae* protein were able to be rationally designed based on structural data from RNases HII and HIII from two distinct bacterial systems highlights the impressive conservation of these enzymes and the importance of continued study of both eukaryotic and prokaryotic ribonucleotide correction.

1.3.5 Bacterial RNase HIII

While the genomes of most bacteria and eukaryotes encode both RNase HI and HII genes, several bacteria, including *B. subtilis*, *B. stearothermophilus*, *Streptococcus pneumoniae*, *Chlamydophila pneumoniae*, and *Aquifex aeolicus*, lack RNase HI and instead contain RNase HIII [for review (*Kanaya*, 2001)]. RNase HIII is considered a type 2 RNase H and is related to RNase HII by sequence and structural comparison (*Ohtani et al.*, 1999a,b) (Figure 1.3B-C). Even though RNase HIII is considered a type 2 enzyme, biochemical characterization indicates it is functionally analogous to *E. coli* RNase HI (*Lu et al.*, 2012b; *Ohtani et al.*, 1999a). RNase HIII enzymes differ from other RNases H in that they contain a TATA-Box binding Protein (TBP)-like N-terminal domain and a distinctive active site motif modification (DEDE) where a glutamic acid replaces the aspartic acid found in type 1 and other type 2 RNases H (DEDD) (*Chon et al.*, 2006).

In vivo, RNase HIII enzymes cleave RNA/DNA substrates containing four or more rNMPs supporting the hypothesis that RNase HIII serves as a functional replacement for bacteria that lack RNase HI (*Itaya et al.*, 1999). For a long time the function of the large N-terminal TBP-like domain, unique to RNase HIII, was unknown. Recently the function of this domain has been determined to be important for substrate binding (*Jongruja et al.*, 2012; *Miyashita et al.*, 2011). Truncation of the N-terminal domain decreases enzymatic efficiency *in vivo* and recent data in *B. stearothermophilus* have shown that six amino acids in this N-terminal domain when mutated individually to alanine show reduced enzymatic activity and/or reduced substrate binding (*Miyashita et al.*, 2011). Purified *Thermovibrio ammonificans* RNase HIII TBP domain alone is sufficient to bind an RNA/DNA hybrid, but does not bind DNA/DNA or RNA/RNA duplex molecules (*Figiel and Nowotny*, 2014; *Miyashita et al.*, 2011). The crystal structure of T. ammonificans RNase HIII bound to its RNA/DNA hybrid substrate reveals two potential "steric gate" residues in the TBP-like domain that interact with the DNA strand of the RNA/DNA substrate and would sterically clash with the 2′-OH of RNA, but this has not

been experimentally tested (*Figiel and Nowotny*, 2014).

RNase HIII enzymes are not known to cleave at single rNMPs, with one interesting exception. *Chlamydophila pneumoniae* RNase HIII (Cpn-HII) can complement an *E. coli rnhB* knockout when cells are grown in the presence of high Mn^{2+} concentrations. Alternatively Cpn-HIII can also complement an *E. coli rnhA* knockout when grown in the presence of Mg^{2+} (*Lu et al*., 2012b). Purification and subsequent *in vivo* experiments demonstrated that Cpn-HIII cleaves at a single rNMP in the presence of Mn^{2+} but not with Mg^{2+}. Therefore Cpn-HIII is the first RNase HIII that we are aware of to be shown to cleave at stretches of four or more rNMPs in RNA-DNA/DNA hybrids and at a single rNMP nested in duplex DNA. This activity has been attributed to Ser 94, which binds the substrate for Cpn-HIII at the RNA-DNA junction, possibly shifting the conformation and allowing a G(R/K)G motif to form hydrogen bonds with the single rNMP. The G(R/K)G motif is responsible for recognizing single ribonucleotides in RNase HII enzymes (*Lu et al*., 2012a) and is not conserved amongst all RNases HIII. Nicking activity at single rNMPs in the presence of Mn^{2+} has yet to be detected in RNase HIII from any other organism (*Lu et al*., 2012a). Interestingly, high levels of Mn^{2+} also inhibit RNase HII in *Chlamydophila pneumoniae*, suggesting a possible adaptation to high levels of manganese (*Lu et al*., 2012b).

1.3.6 Removal of Ribonucleotides by Topo I and NER

In addition to the well-characterized role for the RNases H in removal of ribonucleotides from DNA, other pathways also participate. In yeast, topoisomerase I removes single embedded rNMPs yielding a 2-5 bp deletion, which can be mitigated by helicase Srs2 in collaboration with Exo I (*Kim et al*., 2011; *Sekiguchi and Shuman*, 1997; *Williams et al*., 2013; *Potenski et al*., 2014). Woodgate and co-workers again used the frequent rNMP incorporation by UmuCY11A variant and found that nucleotide excision repair (NER) serves as a backup pathway for removal of rNMPs in cells deficient for *rnhB*

[(*Vaisman et al.*, 2013) for review (*Cai et al.*, 2014)]. Specifically, this work showed that base excision repair and mismatch repair have minimal to no effect on ribonucleotide excision repair *in vivo*. Their work did show that UvrABC incises RNA in duplex DNA containing a single, two, or five consecutive rNMPs in a purified system (*Vaisman et al.*, 2013). The involvement of NER in removal of rNMPs from genomic DNA was an important step forward in identifying other pathways capable of removing ribonucleotide misincorporations. How does NER recognize single rNMPs in DNA? A recent model has been proposed where the 2′-OH in the ribose sugar forms electrostatic and hydrogen bonding interactions with tyrosine residues in the β-hairpin of UvrB resulting in "damage" recognition (*Cai et al.*, 2014). Although this model needs to be tested biochemically, it provides a basis for understanding ribose sugar recognition by NER. It is not yet known if eukaryotic NER also recognizes and incises at rNMPs in genomic DNA. If so, it will provide yet another pathway capable of removing rNMPs from genomic DNA in eukaryotic cells.

1.4 Consequences of rNMPs in Genomic DNA

Phosphodiester bonds 3′ to rNMPs are inherently less stable than those 3′ to dNMPs due to the presence of the 2′-OH in ribose, which can act as a nucleophile to attack the 3′-PO_4, yielding 5′-OH and a cyclic 2′, 3′-PO_4- that is quickly hydrolyzed to 2′- or 3′-PO_4- (*Oivanen et al.*, 1998). Furthermore, rNMPs in the template strand for DNA synthesis can stall or slow the DNA polymerase (*Yao et al.*, 2013). Should ribonucleotides go unrepaired in DNA, this could have severe consequences for genome integrity. However, the ubiquity and great frequency of ribonucleotide incorporation by DNA polymerases in systems from bacteria to mammals suggests that organisms may have evolved mechanisms to benefit from rNMPs embedded in DNA. Indeed, a role for ribonucleotides in DNA during NHEJ has been discovered (*Zhu and Shuman*, 2008), but will not be covered further. A long-standing body of evidence also suggests that DNA

replication is able to initiate from transcripts that anneal to template DNA (R-loops) in a process known as constitutive stable DNA replication (cSDR). Here, we discuss potential consequences of ribonucleotide incorporation into and hybridization with DNA, including genome instability, DNA mismatch repair and DNA replication.

1.4.1 Genome instability

Ribonucleotides that go unrepaired in DNA can hydrolyze to yield 2′- or 3′-PO_4-. Interestingly, an RNA ligase found in all three domains of life, RtcB, may play a role in repair of 3′-PO_4- in DNA (*Englert et al.*, 2012; *Popow et al.*, 2011; *Tanaka et al.*, 2011; *Das et al.*, 2013). Work performed using *E. coli* RtcB revealed that as part of its catalytic mechanism, RtcB becomes covalently linked to GMP, which is then transferred to the 3′-PO_4- to form a 3′-guanylate (DNA-3′-P-GMP) intermediate (Figure 1.4A). When acting on single-stranded DNA or RNA, RtcB then catalyzes attack of the 5′-OH on the DNA-3′-P-GMP such that a new phosphodiester bond is formed, with GMP as a leaving group. Completion of ligation requires RtcB act on a single-stranded substrate, but it is able to guanylate 3′-PO_4- in nicked duplex DNA (*Das et al.*, 2013). The 3′-guanylate cap is also able to protect DNA from exonucleolytic digestion and can be used to prime DNA synthesis by Klenow fragment and *Mycobacterium smegmatis* DinB1, which therefore seals the rGMP into the DNA (*Das et al.*, 2014). The *in vivo* relevance of this process has not been tested, nor has the effect of RtcB on a 3′-PO_4- following an rNMP at a nick in duplexed DNA. Further study of the biochemical pathways processing hydrolyzed DNA 3′ to ribonucleotides will be of great interest in understanding how bacteria avoid death when rNMPs go unrepaired in genomic DNA (Figure 1.4A).

Action of RNase HII on single rNMPs in DNA results in a 5′-rNMP (see Chapter 1.3.4). RER in eukaryotic systems and bacteria has been discussed (Chapter 1.3), but there are a number of enzymes that could act aberrantly on the 5′-PO_4- at a 5′-rNMP generated by RNase HII enzymes. For instance, it was recently determined that ATP-dependent

DNA ligases frequently carry out abortive ligation *in vivo* when presented with a 5′-rNMP at a nick in DNA (*Tumbale et al.*, 2014). Abortive ligation is the process by which a DNA ligase begins to act on a 5′-PO_4- but ultimately fails to complete ligation, resulting in a 5′-adenylated product. In eukaryotes, archaea, and viruses all DNA ligases identified are ATP-dependent (*Wilkinson et al.*, 2001). However, all bacteria encode NAD(+)-dependent DNA ligases (*Wilkinson et al.*, 2001), and the *E. coli* DNA ligase, LigA is able to complete ligation of DNA to an oligonucleotide with a 5′-rNMP end (*Tumbale et al.*, 2014). This is an interesting finding, given that both ATP- and NAD(+)-dependent DNA ligases employ a reaction mechanism in which 5′-adenylation is an intermediate. It is an intriguing possibility that use of NAD(+)-dependent DNA ligases that are apparently able to carry out complete ligation at 5′-rNMPs in bacteria may circumvent the need for aprataxin homologs responsible for reversing 5′-adenylation in eukaryotic systems (*Tumbale et al.*, 2014). Interestingly, aprataxin is also able to resolve a 3′-guanylate cap to a 3′-PO_4- (*Das et al.*, 2014). It is not known whether bacteria which also express ATP-dependent DNA ligases exhibit abortive ligation on 5′ rNMP ends or whether NAD(+)-dependent ligases act on 5′-rNMPs *in vivo*. If they do, then this action would likely be antagonistic to RER and may represent a backup pathway for dealing with 5′-rNMPs in DNA should RER fail after RNase HII nicking.

We have discussed the implications of rNMPs in DNA with respect to their increased rate of hydrolysis, but what happens during DNA replication when rNMPs are in the template strand? On average, *E. coli* pol III takes $\approx$ 3 ms to replicate over a dNMP in template DNA *in vivo* (*Yao et al.*, 2013). When pol III encounters rUMP in the template the pause is increased to $\approx$ 14 ms, and with rGMP the pause is $\approx$ 90 ms (*Yao et al.*, 2013). This represents a 30-fold increase in the time required for pol III to replicate over rGMP. This large effect on DNA polymerase kinetics is not entirely surprising, given that the C3′-endo sugar pucker of an rNMP in DNA will distort the geometry of the local DNA toward the A-form (*Rychlik et al.*, 2010; *DeRose et al.*, 2012). In eukaryotic systems, loss

of RNase H2 results in what is likely a replication-dependent DNA damage response (*Allen-Soltero et al.*, 2014; *Reijns et al.*, 2012). The effect of template rNMPs on DNA polymerase accuracy is unknown, but there is a role for rNMPs and RER in the accuracy of the overall process of DNA replication, specifically through their effect on DNA mismatch repair.

1.4.2 Mismatch Repair

Is rNMP incorporation by DNA polymerases an unfortunate reality of the replication process or do rNMPs have other important biological roles? Again, in the case of bacterial NHEJ a single rNMP is critical for the ligation step during DSB repair providing evidence for a biological role (*Zhu and Shuman*, 2008). In *E. coli* Dam methylation at d(GATC) sequences provides the basis for strand discrimination during mismatch repair (*Iyer et al.*, 2006; *Lahue et al.*, 1989). In most bacteria and all eukaryotic organisms the mechanism used by the mismatch repair machinery to distinguish the nascent strand from the template strand has remained unclear. Several lines of evidence demonstrate that nicks in the DNA can direct mismatch repair to the error containing strand in *Streptococcus pneumoniae* and several eukaryotic organisms (*Holmes et al.*, 1990; *Thomas et al.*, 1991; *Lacks et al.*, 1982). A prominent model is that the DNA ends corresponding to Okazaki fragments transiently formed during discontinuous DNA synthesis provide strand discrimination signals for mismatch repair (*Pavlov et al.*, 2003). However, in eukaryotes and most bacteria the strand discontinuities that are used by the mismatch repair pathway to target correction to the nascent leading strand have remained unclear.

Using variants of the leading strand (pol ϵ) and lagging strand (pol δ) polymerases in *S. cerevisiae* that increase rNMP incorporation it was shown that loss of RER through inactivation of RNase H2 reduced the efficiency of mismatch repair on the leading strand only (*Lujan et al.*, 2013). Furthermore, a single rNMP and RNase H2 provide an

initiation site for mismatch repair targeted to the rNMP-containing strand *in vivo* (*Ghodgaonkar et al.*, 2013). In *B. subtilis*, loss of RNase HII increases mutation rate suggesting a role in mismatch repair, although the mechanism underlying the mutagenesis is unknown (*Yao et al.*, 2013). Considering these studies, evidence is emerging that rNMPs contribute to the efficiency of mismatch repair during replication of the leading strand.

1.5 Conclusion

DNA polymerase errors during DNA replication have the potential to result in high levels of genome instability. Mismatches will become mutations if they go unrepaired by the next round of DNA replication. Mismatches are efficiently detected by MutS and a role for DNA replication proteins in MutS function has been uncovered, but the extent to which MutS is able to bind mismatches independently of the replisome is unclear. The effects of ribonucleotides that go unrepaired are poorly understood. Ribonucleotide incorporation into genomic DNA occurs ubiquitously in living organisms, and despite most polymerases having structural deterrents against incorporation of both single and multiple consecutive rNMPs, both are still incorporated. This can serve many biological roles including providing primers for DNA replication and perhaps providing a means for MMR to identify the nascent strand during mismatch repair. However, regardless of the nature of inclusion, ribonucleotide incorporation creates a high risk for the cell. Consequently, an arsenal of enzymes is employed to continuously cull rNMPs from the genome. By further investigating these systems we can begin to reconcile how cells tolerate the errors produced DNA replication, including the mechanisms by which the errors are made, detected, and repaired. Results of such inquiry will provide insight into the far-reaching consequences of DNA replication errors on bacterial physiology and evolution.

1.6 Dissertation Goals

The genome is constantly buffeted by potential sources of mutation, both intrinsic and environmental. Mechanisms have evolved to limit mutagenesis and allow faithful transmission of genetic information from parent to offspring. *Bacillus subtilis* represents an experimentally tractable model organism with a genome size that is amenable to leveraging the power of high-throughput sequencing technologies to study the processes of genome maintenance and mutagenesis. The goal of this dissertation is to provide novel insight into two methods of maintaining genome stability, DNA mismatch repair and ribonucleotide excision repair, and to determine how mutations occur *in vivo*. In Chapter II I show that MutS binds mismatches in DNA strictly near the site of DNA replication, in Chapter III I consider genomic factors contributing to DNA replication errors, and in Chapter IV I discuss exciting ongoing research into the effects of unrepaired ribonucleotides on genome stability. My thesis contributes vital knowledge to the fields of genome maintenance and mutagenesis, including the mechanism of mismatch searching within the cell, the local sequence contexts that predispose certain genomic loci to DNA polymerase errors, and the contribution of ribonucleotides in genomic DNA to mutagenesis.

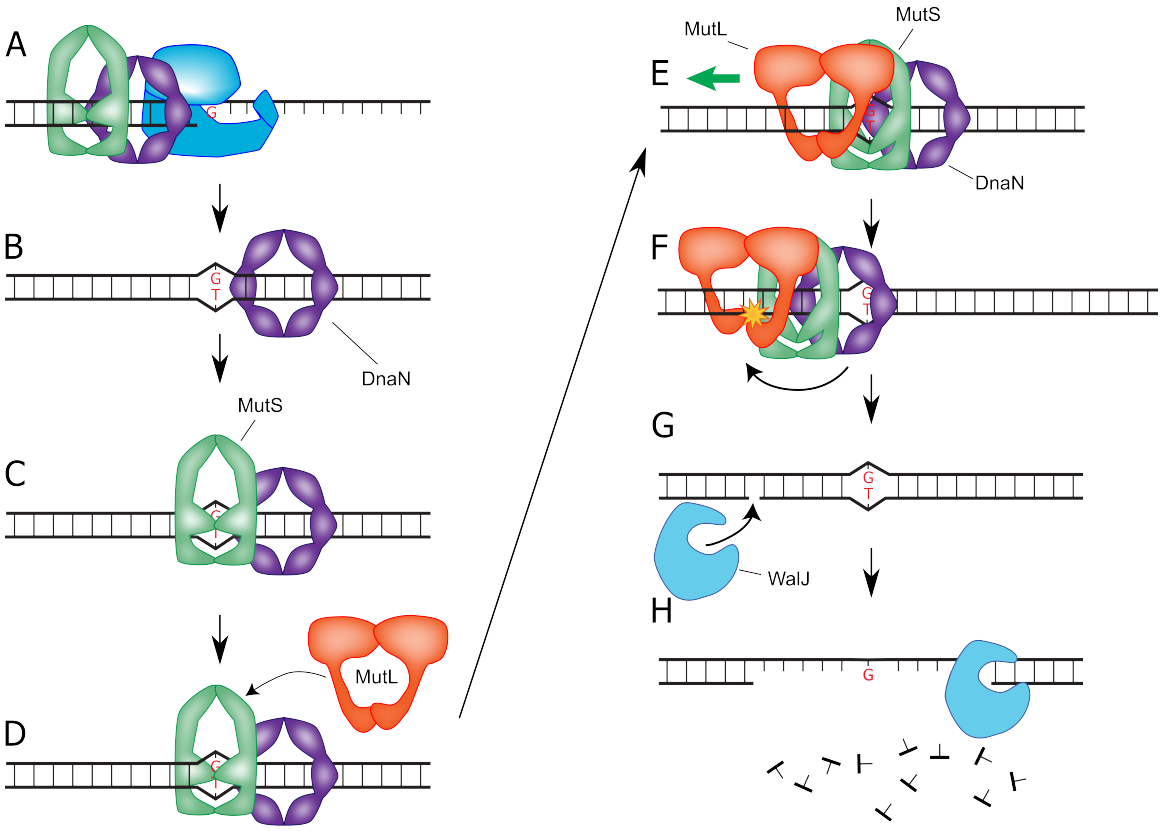


Figure 1.1: Mismatch Repair in *Bacillus subtilis*. Model for mismatch repair in *B. subtilis*. (A - C) A mismatch is produces by the DNA polymerase and β clamp recruits MutS to the site of the mismatch. (D) MutS recruits MutL to the site of the mismatch. (E and F) We speculate that the complex slides along the DNA until the latent endonuclease activity of MutL is stimulated, possibly through interaction with the β clamp, causing MutL to nick the nascent strand. (G) The error-containing strand is then digested, likely by WalJ (see Chapter III). New homoduplex DNA will be synthesized in the gap, and the new strand ligated to complete mismatch correction.

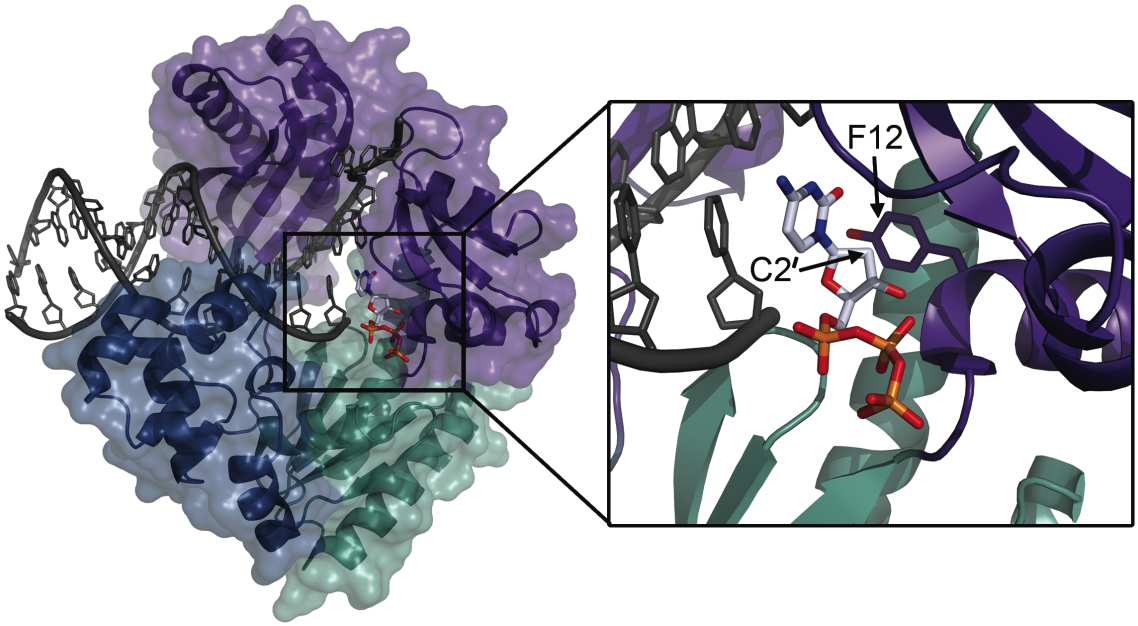
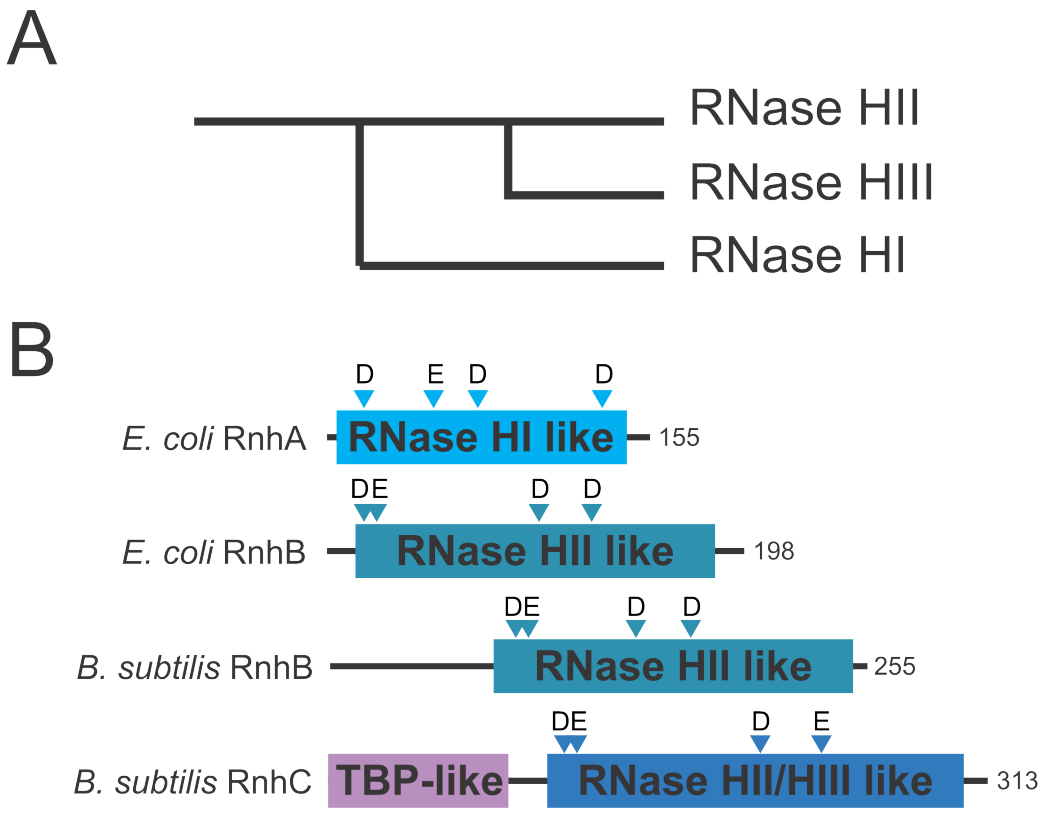


Figure 1.2: Steric Gate Residues Clash with Ribonucleotides. DNA polymerase IV from *Sulfolobus solfataricus* (PDB 4QW8; (*Gaur et al.*, 2014)) is shown as a ribbon diagram bound to template DNA and a nucleotide (dCTP) represented as sticks. The palm (green), thumb (blue) and finger (purple) subdomains have a semitransparent surface. The zoomed view demonstrates the close proximity of the deoxyribose C2′ from dCTP with the phenylalanine steric gate (F12). Oxygen atoms are colored in red, nitrogen in blue, and phosphorous in orange.



C

	R	RRRR	RRRRRRRRRR
RNase HI		✓	✓
RNase HII	✓	✓	✓
RNase HIII		✓	✓
NER	✓	✓	
Pol I	✓	✓	✓

Figure 1.3: Bacterial recognition and removal of rNMPs. A) Schematic evolutionary history of the family of RNase H enzymes with respect to sequence similarity (*Ohtani et al.*, 1999b). B) Domain structures of functional *E. coli* and *B. subtilis* RNases H. Triangles denote the location and identity of the residues stabilizing the metal ions necessary for catalytic activity (*Tadokoro and Kanaya*, 2009). C) The ability of the RNases H, NER and pol I to recognize or engage in repair of different RNA/DNA hybrids (*Tadokoro and Kanaya*, 2009; *Vaisman et al.*, 2013, 2014).

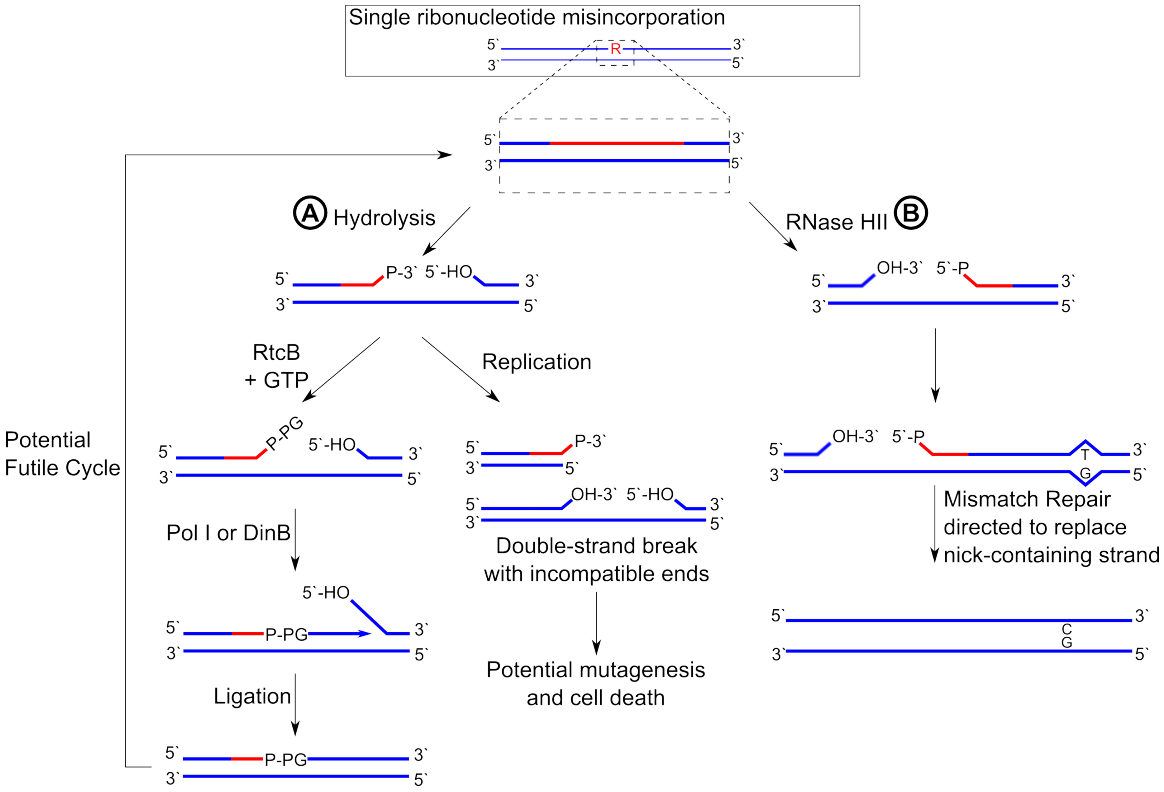


Figure 1.4: Consequences of ribonucleotides in genomic DNA. Shown is a single ribonucleotide embedded in DNA, which may have many consequences to genome integrity. A) If the ribonucleotide remains in the genome it is prone to hydrolysis, producing a 2′- or 3′-PO_4- (*Oivanen et al.*, 1998). The resulting nick could cause a double-stranded break during the next round of DNA replication, or it may be guanylated by an RtcB-like enzyme so that it can be used to prime DNA synthesis by pol I or DinB (*Das et al.*, 2014). This would result in the original rNMP as well as the rGMP added by RtcB to remain in the genome. Thus, such a mechanism for tolerance of rNMPs in DNA is likely to represent a futile cycle and may even serve to exacerbate problems caused by rNMPs. B) RNase HII can generate a nick 5′ to the rNMP, which can then direct DNA mismatch repair to target the rNMP-containing strand for replacement (*Ghodgaonkar et al.*, 2013; *Lujan et al.*, 2013).

CHAPTER II

Single-molecule Motions and Interactions in Live Cells Reveal Target-search Dynamics in Mismatch Repair

2.1 Abstract

MutS is responsible for initiating the correction of DNA replication errors. To understand how MutS searches for and identifies rare base-pair mismatches, we characterized the dynamic movement of MutS in real time using super-resolution microscopy and single-molecule tracking. We found that MutS dynamics are heterogeneous in cells, with one MutS population exploring the nucleoid rapidly while another MutS population moves to and transiently dwells at the replisome region, even without appreciable mismatch formation. Analysis of MutS motion shows that the speed of MutS is correlated with its separation distance from the replisome and that MutS motion slows when it enters the replisome region. We also show that mismatch detection increases MutS speed supporting the model for MutS sliding clamp formation after mismatch recognition. Using variants of MutS and the replication processivity clamp to impair mismatch repair we find that MutS dynamically moves to

A modified version of this chapter was in revision at *Proceedings of the National Academy of Sciences* during the writing of this thesis document. We thank Dr. Daniel Jarosz for comments.

Yi Liao, Lyle Simmons, Julie Biteen and I designed research; I constructed bacterial strains and performed biochemistry, molecular biology and genetic analyses; Yi Liao performed single-molecule imaging and analyzed data; Burke Gao performed DnaN super-resolution imaging; Yi Liao, Lyle Simmons, Julie Biteen and I discussed the results and wrote the paper.

and from the replisome before mismatch binding to scan for errors. Furthermore, a block to DNA synthesis shows that MutS is only capable of binding mismatches near the replisome. It is well established that MutS engages in an ATPase cycle, which is necessary for signaling downstream events. We show that a variant of MutS with an ATP binding defect is no longer capable of dynamic movement to and from the replisome demonstrating that proper nucleotide binding is critical for MutS to locate the replisome *in vivo*. Our results provide mechanistic insight into the trafficking and movement of MutS in live cells as it searches for mismatches.

2.2 Significance

We integrate single-molecule super-resolution imaging with biochemical and genetic approaches to understand how the mismatch repair protein MutS efficiently identifies DNA mismatches during real time in living cells. We found that MutS molecules move fast, exploring the entire nucleoid, but can transition to a slow moving population that is localized at the replisome even before a mismatch is produced. We show that bacterial MutS must initiate mismatch binding in very close proximity to the replisome. We also show that mismatch detection increases MutS speed supporting the model for MutS sliding clamp formation after mismatch recognition. Our results provide fundamental insight into the searching behavior of single MutS molecules during DNA replication.

2.3 Introduction

DNA mismatch repair (MMR) is the highly conserved process responsible for correcting DNA replication errors (*Jiricny*, 2013). Although replication errors occur infrequently in bacteria ($\approx$ 1 error per 31,000,000 base pairs) (*Lee et al.*, 2012), the consequences of MMR failure on human health are severe (*Fishel et al.*, 1993). The first

protein involved in the MMR pathway, MutS, is responsible for detecting rare base-pairing errors. In *Bacillus subtilis*, MutS then recruits MutL, an endonuclease in most bacteria and eukaryotic organisms, to incise the DNA (*Kadyrov et al.*, 2006a; *Pillon et al.*, 2010). After MutL incision, the error-containing strand is removed and the DNA resynthesized by a process not well understood in most organisms other than *Escherichia coli* (*Tran et al.*, 1999).

The mechanism by which MutS homologs locate a single mismatch among millions of correctly paired nucleotides has been studied extensively by bulk biochemistry, by *in vitro* atomic force microscopy and single-molecule imaging, and by visualizing MutS using *in vivo* cell biology approaches (*Kleczkowska et al.*, 2001; *Smith et al.*, 2001; *Wang et al.*, 2003; *Gorman et al.*, 2007; *Simmons et al.*, 2008; *Gorman et al.*, 2010; *Sass et al.*, 2010; *Hombauer et al.*, 2011a; *Jeong et al.*, 2011; *Cho et al.*, 2012; *Gorman et al.*, 2012; *Qiu et al.*, 2012; *Lenhart et al.*, 2013b; *DeRocco et al.*, 2014). Single-molecule studies largely indicate that MutS operates as a searching clamp as it diffuses along DNA in a 1D search (*Gorman et al.*, 2007; *Jeong et al.*, 2011; *Cho et al.*, 2012). In this model, following mismatch recognition, MutS can dwell at the mismatch before exchanging ADP and ATP, converting into a stable ATP-bound sliding clamp, and subsequently diffuse away from the mismatch at a faster rate in search of MutL and possible strand discrimination signals (*Jeong et al.*, 2011; *Cho et al.*, 2012; *Gorman et al.*, 2012). Studying the mechanistic steps of the search process with DNA curtains has provided evidence that MutS may identify errors through a combination of 1D sliding and a 3D pathway (*Gorman et al.*, 2012). These DNA curtain results show that MutS predominantly searches DNA by 1D diffusion, but is also capable of mismatch recognition via a 3D pathway in which MutS binds the mismatch without engaging in a prior 1D search. The 3D pathway for mismatch recognition has been proposed to help MutS circumvent protein barriers that would exist *in vivo*. Further analysis of MutS on DNA curtains suggests that nucleosomes largely prevent 1D diffusion by MutS and indicates that

requiring strict 1D diffusion as the *in vivo* search mechanism would present a significant challenge for the identification of a replication error, though other studies suggest MutS can reposition nucleosomes (*Javaid et al.*, 2009; *Gorman et al.*, 2010). Therefore, it is still not entirely clear how MutS searches for errors within the supercoiled, compacted and protein-laden nucleoid in the crowded 3D environment of a living bacterial cell.

Each of these *in vitro* single molecule studies has provided important insight into target search dynamics of MutS, and these studies and other bulk studies have elucidated many of the steps in the mechanism of mismatch identification on naked DNA *in vitro*. Of course, most of these studies have been limited to analysis of MutS in isolation on protein-free DNA; they do not incorporate replication proteins or other *in vivo* obstacles that are likely to impact the *in vivo* search process. The search dynamics of MutS have not been investigated at single-molecule resolution for any MutS homolog *in vivo*.

The need to understand how the biochemistry of MutS homologs translates into mismatch identification in live cells has led to several studies across many organisms using fluorescent protein fusions and bulk fluorescence imaging approaches. These studies have shown that MutS homologs form foci that colocalize with the DNA replication machinery (replisome), and that interaction of MutS with replication processivity clamps (PCNA or β-clamp) is important for focus formation (*Kleczkowska et al.*, 2001; *Smith et al.*, 2001; *Pluciennik et al.*, 2009; *Hombauer et al.*, 2011a; *Klocko et al.*, 2011; *Lenhart et al.*, 2013b). However, the colocalization between bacterial MutS and the replisome is far from absolute. For example, in the absence of an exogenous mutagen, bulk microscopy detects fluorescent MutS foci in only $\approx$ 9% of *B. subtilis* cells. Furthermore, only about half of these 9% of cells with a MutS focus showed colocalization of MutS to the replisome (*Lenhart et al.*, 2013b). Therefore, only about 4.5% of cells under normal growth conditions in *B. subtilis* show colocalization between MutS and the replisome (*Lenhart et al.*, 2013b). Obviously, the behavior of

MutS in the remaining 91% of cells without a MutS focus has been impossible to determine using ensemble fluorescence techniques and would require single-molecule imaging to reveal the location and behavior of this MutS pool, which represents the vast majority of cellular MutS.

In *Saccharomyces cerevisiae*, fluorescent foci of MutSα (Msh2-Msh6) are observed to colocalize constitutively with replication centers (*Hombauer et al.*, 2011a). In *S. cerevisiae* there exists an Exo1-dependent MMR pathway that appears only to depend weakly on MutSα colocalization with the replisome (subunit of polymerase ϵ) (*Hombauer et al.*, 2011a); however the Exo1-independent pathway is dependent on MutSα interaction with the replication processivity factor PCNA. Mismatch repair and DNA replication have also been temporally correlated in *S. cerevisiae*. These studies showed that MutSα expressed during S-phase corrects errors in reporters replicated during S phase and MutSα expressed during G2/M corrected a region of the genome replicated later (*Hombauer et al.*, 2011b).

Therefore, ensemble fluorescence imaging and other experiments have yielded important information regarding the positioning of MutS foci *in vivo* supporting the model that mismatch repair is coupled to DNA synthesis. In *B. subtilis*, most of the MutS pool cannot be visualized and interestingly, in *S. cerevisiae* only $\approx$ 15% of MMR was correlated with replisome-colocalized MutSα in otherwise wild type cells (*Hombauer et al.*, 2011a,b). Therefore, we have been unable to image and quantify the behavior of MutS outside of large static foci visible during ensemble fluorescence microscopy, although in bacteria, large foci account for only 10% of cellular MutS, that is, $\approx$ 8 dimers out of the 80 dimers present in a cell. To visualize and study MutS *in vivo*, only a more sensitive method with higher spatial and temporal resolutions can unambiguously determine the extent to which MutS is enriched near the replisome, and answer two open questions: Is MutS enrichment at the replisome constitutive in bacteria? Does MutS search for mismatches away from the replisome *in vivo*?

Ensemble fluorescence methods do not allow investigation of the behavior of individual molecules of MutS and its homologs throughout MMR.

In the complex cellular environment, obstacles from molecular crowding, chromosome packing and numerous DNA-binding proteins could easily impede MutS in its search for replication errors. To further understand the process of MMR inside cells and to gain new insight into the movement and location of MutS molecules *in vivo* throughout MMR, we probed the spatial and genomic distribution and dynamics of MutS in live bacteria with super-resolution fluorescence imaging (*Betzig et al.*, 2006; *Hess et al.*, 2006; *Rust et al.*, 2006), single-molecule tracking (*Yildiz et al.*, 2003; *Elmore et al.*, 2005; *Biteen and Moerner*, 2010) and genomic approaches. We investigated the effect of replisome association, active DNA synthesis, mismatch binding, ATP binding and the presence of MutL on the dynamics of MutS movement *in vivo*. This study applies single molecule imaging to a dedicated DNA repair pathway and captures the trafficking behavior of MutS, a protein that is conserved from bacteria to humans. A high degree of spatial and temporal resolution has allowed us to view the dynamic search process and movement of single MutS protomers *in vivo*.

2.4 Results

2.4.1 Localization and dynamics of MutS in live *B. subtilis*.

We constructed *B. subtilis* strains natively expressing MutS fused to the photoactivatable fluorescent protein PAmCherry1 (*Subach et al.*, 2009) as the sole source of MutS (Figure 2.1A). MutS-PAmCherry retains MMR activity and the fusions are stable against proteolytic loss of PAmCherry (Figure 2.2). To investigate the dynamics of each MutS copy with sub-diffraction resolution ($<$ 35 nm), a 405-nm laser was used at low power to produce 1–3 copies of emissive PAmCherry per cell at a time (Figure 2.1B). The MutS-PAmCherry copies in this photoactivated subset were imaged and

tracked in real time until the PAmCherry molecules photobleached. Through 20–30 iterations of this photoactivated localization microscopy (PALM) activation/imaging/photobleaching cycle, super-resolution images were constructed, and multiple single-molecule trajectories were recorded for each cell (*Biteen et al.*, 2008; *Manley et al.*, 2008; *Karunatilaka et al.*, 2014; *Haas et al.*, 2015). In cells expressing MutS-PAmCherry, we observe that MutS explores the entire cell while also experiencing significant enrichment and confinement near the cell center or cell quarter positions (Figure 2.1C), where the replisome was expected to reside based on previous studies (*Lemon and Grossman*, 1998; *Berkmen and Grossman*, 2006).

2.4.2 Localization and dynamics of the DNA replication machinery in live *B. subtilis*.

To visualize the replisome position as a control, and to probe whether the sites of MMR and of DNA replication coincide, we labeled DnaX, part of the processivity clamp loader and a proxy for the replisome, with the yellow fluorescent protein mCitrine (Figure 2.3A). Consistent with earlier studies, we found that DnaX forms clusters either at the mid-cell or at the quarter-cell positions (Figure 2.3B). Because each replisome contains multiple copies (*Reyes-Lamothe et al.*, 2010) of DnaX-mCitrine with overlapping fluorescence signals (Figure 2.3B), and because existing algorithms designed to extract single-molecule positions from images of densely populated fluorophores are not capable of accurately recovering positions if multiple emitters are separated from each other by less than 100 nm (*Holden et al.*, 2011; *Zhu et al.*, 2012), we used a photobleaching-assisted approach (*Burnette et al.*, 2011; *Simonson et al.*, 2011) to achieve single-molecule localization of DnaX (Figure 2.3C): upon identifying mCitrine molecules that photobleach from one imaging frame to the next (*Watkins and Yang*, 2005), the mean of the frames after photobleaching is subtracted from the mean of the frames before photobleaching to produce the point spread function (PSF) of the

photobleached molecule. The position of this PSF is subsequently determined from a fit to a 2D Gaussian function. We found that within a cluster, DnaX molecules are on average separated from each other by 54 nm (early on in the replication cycle) or 119 nm (during the later stage of the replication cycle) (Figure 2.3D) (*Lemon and Grossman*, 1998). In addition, from time-lapse imaging without photobleaching, we also observed that DnaX clusters are neither mobile nor strictly stationary; instead, they engage in subtle motions exploring a small domain of size 84 ± 20 (s.d.) nm as measured by the radius of gyration of the centroid position (Figure 2.3E). Our results support models describing the *B. subtilis* replisome as a confined assembly (*Lemon and Grossman*, 1998; *Imai et al.*, 2000), as opposed to the model of a mobile replisome complex that tracks on chromosomal domains that has been described for *E. coli* (*Reyes-Lamothe et al.*, 2008).

2.4.3 Relative positions and dynamics of DnaX-mCitrine and MutS-PAmCherry.

Imaging cells expressing both DnaX-mCitrine and MutS-PAmCherry, we found that MutS accumulates near the replisome despite being overall more mobile than DnaX (Figure 2.4A). The instantaneous speed from single-molecule MutS tracks (*Lakadamyali et al.*, 2003) also shows dependence on the separation distance between MutS and the replisome: upon entering the replisome region (separation distance $<$ 100 nm), MutS slows down to match the average speed of DnaX clusters (Figure 2.4B), likely as a result of the known direct interaction between MutS and the replication processivity clamp (β-clamp) (*Simmons et al.*, 2008; *Lenhart et al.*, 2013b), or due to MutS engaging in mismatch searching and binding on replisome-proximal DNA (*Simmons et al.*, 2008; *Lenhart et al.*, 2013b), or a combination of both possibilities. Cross-correlation analysis shows that MutS instantaneous speed does indeed positively correlate with the separation distance from the replisome (Figure 2.4C). However, the relatively low correlation amplitude of $\approx$ 0.2 indicates heterogeneity among MutS subpopulations, that is, some MutS do not slow down or slow down only transiently when passing by

the replisome. To quantify how much time a MutS protein spends within the replisome region, we fit the cumulative probability of the dwell time of MutS in the replisome region (P(t) $>$ t) with a two-term exponential decay function (Figure 2.4D) and obtained two dwell time constants of 25 ms (42%) and 188 ms (58%). These dwell time constants represent a lower bound, as we only analyzed single-MutS trajectories that start outside the replisome, remain trackable within the replisome, and end outside the replisome (Figure 2.4D inset). As a result, photobleaching and blinking of the fluorophore do not affect the dwell time analysis, but only MutS trajectories that start and end outside the replisome are taken into account. We conclude that the majority of MutS molecules that pass by the replisome (58%) dwell there for a considerable time (at least 188 ms), likely to engage in a local search for mismatches, while the other population (42%) is freely diffusing within the cell.

2.4.4 MutS accumulates at the replisome regardless of mismatch formation rate.

It is necessary to sample a large number of cells to make meaningful conclusions since there are only approximately 160 copies of MutS-PAmCherry in each cell and PAmCherry photoactivation efficiency is 4-50% (*Durisic et al.*, 2014; *Wang et al.*, 2014). Furthermore, only some small fraction of the imaged cells has mismatches to which MutS is able to respond at all (Figure 2.5E). Therefore, to compare intracellular DnaX-mCitrine and MutS-PAmCherry positions across many cells, we plotted the percentage of DnaX and MutS localizations at corresponding positions inside a normalized cell as probability density maps (Figure 2.6) (*Wang et al.*, 2011). A density map of wild-type (WT) cells treated with the mismatch-forming drug 2-aminopurine (2-AP) (WT+; Figure 2.5E and supplemental text) shows that regions with the highest DnaX densities are also those most populated by MutS (Figure 2.5C). Importantly, the same pattern was also found in WT cells without 2-AP (WT–) (Figure 2.5A and compare Figure 2.5B to 2.5C). Although previous bulk fluorescence studies of *B. subtilis* have

shown that $<$ 10% of WT– cells form MutS foci near the replisome (*Smith et al.*, 2001; *Lenhart et al.*, 2013b), here at the single-molecule level with improved sensitivity to capture transient dwelling behaviors (Figure 2.4B), we reveal that the enrichment of MutS near the replisome is much more prevalent than previously concluded, even in cells with only the very low natural mismatch formation rate (i.e. no 2-AP addition). Furthermore, this mismatch-independent recruitment of MutS in *B. subtilis* resembles the behavior of MutSα in *S. cerevisiae* cells observed using bulk fluorescence (*Hombauer et al.*, 2011a), suggesting that mismatch recognition is an integral component of replisome function and has been conserved from bacteria to eukaryotes. However, we show that a difference between bacteria and yeast is that bacterial MutS is highly dynamic, displaying heterogeneous motion, in addition to its replisomal recruitment.

2.4.5 MutS speed increases after mutagen treatment.

To understand the spatial dependence of MutS motion, we analyzed the average diffusion coefficient of MutS as a function of separation distance from the nearest replisome (Figure 2.5D). The diffusion coefficient, D, is calculated from the mean square displacement, and only data from the first quarter of the time lags for each of over 1000 trajectories longer than 10 frames were analyzed to minimize errors associated with higher time lag values (*Saxton*, 1997). Under both normal and mutagenic growth conditions, the MutS diffusion coefficients are similar to those of DNA ligase and pol I in live *E. coli* (*Uphoff et al.*, 2013) and to those of DNA-bound proteins exhibiting a combination of 1D and 3D motions in other *in vivo* systems (*Elf et al.*, 2007; *Gebhardt et al.*, 2013). However, despite the nearly identical net MutS localization patterns for WT– and WT+ cells, (Figure 2.5B and 2.5C), on the single protein level, we found that MutS exhibits an overall faster motion in WT+ cells compared to WT– (Figure 2.5D). This *in vivo* difference in speed is consistent with *in vitro* observations that MutS switches

from a rotation-coupled sliding configuration to a faster and rotation-free sliding after mismatch binding (*Jeong et al.*, 2011; *Cho et al.*, 2012; *Gorman et al.*, 2012). Regardless of whether 2-AP is present, we observe that MutS slows as it approaches the replisome in both WT+ and WT– cells, suggesting that error searching, and possibly subsequent binding events, are restricted to nascent DNA in the neighborhood of DNA replication sites, and are likely initiated by interactions between MutS and the replisome.

2.4.6 The MutS/replisome interaction is necessary for MutS recruitment and MMR *in vivo*.

To further understand the relationship between DNA replication and the position and dynamics of single MutS molecules, we constructed four strains, each designed to impair one of four MMR steps: 1) MutS binding to β-clamp, 2) mismatch recognition, 3) MutS nucleotide binding, and 4) subsequent MutL recruitment (Figure 2.7A). We tested the effect of replisome interaction perturbation on MutS motion and location using a two-pronged approach: first, we examined MutS800, a MutS variant with the domain that has β-clamp affinity removed. MutS800 is able to bind mismatches (*Simmons et al.*, 2008) and maintains similar ATPase activity to full-length MutS (Figure 2.8). Second, we complemented the MutS800 investigations with studies of the β-clamp allele *dnaN5*, which is compromised for interaction with MutS (*Simmons et al.*, 2008). Relative to the WT– and WT+ density maps, the MutS localization pattern is drastically changed in untreated MutS800 (MutS800–) cells (compare Figure 2.7C to Figure 2.5B-C). While DnaX locations remained largely the same, MutS800 was uniformly distributed throughout the cell, lowering the Pearson correlation between DnaX and MutS800 from 0.83 in WT– cells to 0.50 in MutS800–. The significant decrease in preferential enrichment of MutS near the replisome confirms that the recruitment of MutS to the replisome observed in WT cells depends in part on interactions between MutS and the β-clamp.

The density maps after 2-AP treatment (Figure 2.7D) suggest that MutS800 still

partially responds to mismatch incorporation (the correlation coefficient increases from 0.50 to 0.70 with addition of 2-AP). Although MutS800 is still capable of mismatch binding *in vitro* (*Simmons et al.*, 2008), only 10-12% of MMR activity is retained in MutS800 cells (Figure 2.2) (*Lenhart et al.*, 2013b), further demonstrating the importance of replisomal interactions for efficient mismatch recognition *in vivo*, and we interpret the slightly higher densities near the quarter positions of the MutS800+ density map (Figure 2.7D) as MutS800 responsible for the remaining MMR activity. In addition, the diffusion rates for MutS800 were significantly increased relative to WT– and WT+ (Figure 2.7E). One possible contributing factor to the increase in diffusion rate is that mismatch-stimulated ATPase activity of MutS800 is 2-fold diminished relative to that of the wild type MutS, although the basal ATPase activity between the two proteins in the presence of homoduplex DNA is the same (Figure 2.8). However, our results for the complementary experiment in the DnaN5 mutant strain also show a partially compromised colocalization pattern between MutS and DnaX and a faster diffusion profile for MutS in the DnaN5 strain compared to WT cells (Figure 2.9), demonstrating that the increase in MutS800 diffusion should not be solely attributed to a slight decrease in stimulation of ATPase activity in response to a mismatch. One potential explanation for this change in dynamics is that the MutS/β-clamp interaction facilitates MutS binding to DNA strands and in the absence of such an interaction, MutS cannot efficiently engage in slow 1D searching motion.

2.4.7 MutS recruitment to the replisome occurs independently of mismatch recognition.

As the MutS positioning did not change upon 2-AP treatment in WT cells, we tested if the recruitment of MutS to the replisome is contingent upon mismatch binding. We analyzed the distribution and motion of MutS[F30A], which is unable to recognize mismatches (*Lenhart et al.*, 2013a). Both with and without 2-AP, this mutant preserved

the elevated MutS density around the replisome seen in WT cells (compare Figure 2.7G-H to Figure 2.5B-C). Also, the dependence of diffusion rates on separation distance (Figure 2.7I) for both MutS[F30A]– and MutS[F30A]+ was virtually identical to those of WT–. Thus, as expected, because MutS[F30A] cannot bind mismatches (*Lenhart et al.*, 2013b), all aspects of MutS[F30A] localization and motion remain unresponsive to mismatches caused by 2-AP (Figure 2.7G-I). The highly similar positioning and dynamics of MutS[F30A]– and MutS in WT– cells supports the notion that the recruitment of MutS by the replisome not only precedes mismatch recognition, but also occurs independently of it. This dynamic is consistent with the ability of MutS to efficiently respond to very rare mismatches (*Lee et al.*, 2012). Because MutS is enriched at the replisome, in close proximity to potential mismatches and prior to errors occurring, MutS has access to newly replicated, "naked" DNA strands along which prolonged sliding by MutS would be possible.

2.4.8 Nucleotide binding is necessary for MutS recruitment.

Another MMR step that is poorly understood in bacterial cells is how nucleotide binding affects the search phase, movement and localization of MutS homologs in living cells (Haber1991, Alani1997, Sharma2013). Abundant *in vitro* evidence shows that MutS carries out an ATPase cycle in which it is able to engage in mismatch searching while bound to ADP, and upon mismatch detection, MutS binds ATP inducing a conformational change to form a stable clamp capable of more rapid 1D diffusion along the DNA (*Jeong et al.*, 2011; *Qiu et al.*, 2012; *Sharma et al.*, 2013). MutS engages in an ATPase cycle during MMR wherein MutS searches for mismatches in an ADP-bound state, and upon mismatch detection, MutS binds ATP and switches into a stable sliding clamp on DNA (*Jeong et al.*, 2011; *Qiu et al.*, 2012; *Sharma et al.*, 2013). To better understand how the ATPase cycle affects MutS localization and dynamics *in vivo*, we constructed a strain harboring MutS[K608M]-PAmCherry as its only source of MutS.

MutS[K608M] in *B. subtilis* is the homologous substitution to *E. coli* MutS[K620M], which has a far reduced affinity for nucleotide and therefore little ATPase activity (*Junop et al.*, 2001). The ability of *E. coli* MutS[K620M] to bind mismatched DNA is not clear. One study shows mismatch binding by this MutS variant (*Junop et al.*, 2001) while another does not (*Acharya*, 2008). On the other hand, MutS[K608M] changes the highly conserved Walker A motif in MutS, and purified *B. subtilis* MutS[K608M] has no measurable ATPase activity (Figure 2.8) (*Junop et al.*, 2001). Strikingly, MutS[K608M] displayed highly diffusive behavior both with (MutS[K608M]+) and without (MutS[K608M]-) 2-AP treatment (Figure 2.7J-L). MutS[K608M] motion was also unresponsive to 2-AP treatment, and similar to MutS800, the diffusion of MutS[K608M] did not depend on its separation distance from the replisome (Figure 2.7M). However, in contrast to MutS800 motion, which was significantly faster, this constant MutS[K608M] diffusion rate throughout the cell closely resembled the rate of WT+ MutS at positions $>$ 250 nm from the replisome (Figure 2.7M). These results suggest that processivity clamp interaction alone is not entirely sufficient to recruit MutS to the site of DNA synthesis, and proper nucleotide binding by MutS is also necessary for positioning of MutS within the cell, possibly to return MutS to its mismatch searching state near the replisome. Importantly, whether or not MutS[K608M] is able to recognize mismatches as efficiently as wild type MutS, MutS[F30A], which is completely deficient in mismatch recognition, localizes to the replisome. This leads us to the conclusion that replisome interaction and nucleotide binding, but not mismatch detection, are necessary for MutS positioning to the replisome.

2.4.9 MutS only recognizes mismatches spatially close to the replisome.

Following mismatch detection, MutS binds MutL to form a complex (*Lenhart et al.*, 2013a), which is proposed to then slide away from the mismatch in search of MutL and strand-discrimination signals along the DNA (*Smith et al.*, 2001; *Simmons et al.*, 2008;

Gorman et al., 2012). To further test whether MutS binds mismatches near or distal to the replisome, we probed the effect of MutL binding on MutS dynamics by imaging DnaX and MutS in a Δ*mutL* strain having no MMR activity. Without 2-AP (Δ*mutL*–), the localization of MutS in Δ*mutL* largely resembles that observed in the WT and MutS[F30A] strains (compare Figure 2.7N to Figure 2.5B-C and Figure 2.7G-H), indicating that the pre-loading of MutS at the replisome is unaffected by the absence of MutL. However, it is notable that when mismatches were induced in Δ*mutL* cells by 2-AP treatment (Δ*mutL*+), the density of MutS in replisome-proximal regions diminished (compare Figure 2.7O to Figure 2.5C and Figure 2.7H). We postulate that the decline in MutS enrichment at the replisome is caused by MutS remaining mismatch-bound in the absence of MutL, and thus being carried away from the replisome with newly synthesized mismatch-containing DNA as DNA replication proceeds. We tested this hypothesis by imaging MutS in the Δ*mutL* strain incubated with 2-AP for one hour, followed by treatment with HPUra, which blocks DNA replication. Consistent with our hypothesis, enrichment of MutS at the replisome was restored after DNA replication was arrested (compare Figure 2.7P to Figure 2.7O), demonstrating that, in Δ*mutL*, MutS is a marker of mismatch positions on the newly replicated DNA and in the absence of HPUra, the attenuated MutS accumulation observed in Δ*mutL*+ (Figure 2.7O) is caused by the mismatch-bound MutS being carried away from the replisome during ongoing DNA synthesis. The process of MMR, from mismatch detection by MutS through ultimate replacement of the error-containing strand of DNA must occur quickly, as deletion of MutL was necessary to observe an effect of 2-AP treatment on MutS position (Figure 2.7N and O).

Overall, the dependence of mismatch recognition on replisome coupling, as observed in MutS800 and DnaN5 strains, is further supported by the MutS localization in Δ*mutL*. Here, DNA mismatches will have accumulated throughout the genomic DNA of the cells in Figure 2.7O prior to HPUra treatment due to prolonged 2-AP exposure

without MutL. Therefore, if MutS could bind mismatches away from the replisome, the MutS distribution would resemble the diffuse pattern for Δ*mutL*+ cells (Figure 2.7O). Rather, HPUra causes MutS to resume its normal enrichment at the replisome in Figure 2.7P. This, together with the loss of MMR activity in MutS800 and partial loss of MMR in DnaN5 (*Simmons et al*., 2008) (Figure 2.2), indicates that proximity to the replisome is critical for MutS to efficiently locate mismatches *in vivo*, and also that the proximity of mismatched DNA to the replisome plays a fundamental role in MMR initiation, as mismatches distal to the replisome are not efficiently targeted by MutS. In contrast to WT cells, MutS from the Δ*mutL* strain moves overall more quickly when distant from the replisome and strikingly, it slows down upon 2-AP treatment (Figure 2.7Q), which presumably results from the increased number of mismatch-bound MutS molecules, as the difference in diffusion coefficients between Δ*mutL*– and Δ*mutL*+ is largest outside the replisome region.

In contrast to treatment with 2-AP, which had no effect on replisome localization, blocking DNA replication with HPUra changed the locations of both the replisome and MutS. For pre-divisional cells with two DNA replication sites, the replisome and MutS were both shifted inward from quarter positions toward the cell center (Figure 2.7P), likely due to the fact that, with DNA replication paused, cells continue to grow but fail to partition the replisome to daughter cells. The simultaneously shifted colocalization pattern observed in HPUra-treated cells again highlights the functional correlation between DNA replication and repair.

2.4.10 MutS interacts with newly replicated DNA and essential DNA polymerases both *in vivo* and *in vitro*.

To biochemically test whether MutS physically associates with replicating DNA, we synchronized cells for replication initiation using a temperature-sensitive allele of the replicative helicase loader DnaB (DnaB134), followed by ChIP-sequencing (ChIP-seq) of

MutS and the essential DNA polymerase subunits PolC and DnaE to determine the location of each protein on the chromosome (*Dervyn et al*., 2001; *Sanders et al*., 2010). The *B. subtilis* genome consists of a single circular chromosome with a single origin of replication (*oriC*). DNA replication commences at *oriC* and proceeds bidirectionally toward *terC*. In synchronized culture during pre-initiation, we observed little to no enrichment of MutS, DnaE or PolC on the chromosome (Figure 2.10A). In contrast, 10 minutes after replication has commenced, PolC, DnaE and MutS are co-enriched at the site of replication initiation (*oriC*), each displaying about 2-fold enrichment (Figure 2.10A). The enrichment is quite broad, occupying a region over 200 kbps wide. The amount of enrichment near *oriC* is, therefore, substantial given the broad distribution of MutS. Importantly, there was little enrichment in ChIP-seq of MutS800 at 10 minutes after replication initiation, consistent with the MutS800 variant being deficient in recruitment to the replisome, and ChIP-seq using antibodies directed against MutS in a strain lacking *mutS* yielded no enrichment at all (Figure 2.10A). Given these controls, the less than 80% synchrony within our system (*Burnett and Wake*, 1977), and the very similar patterns of enrichment we observed in our independent ChIP-seq of two essential components of the *B. subtilis* DNA polymerase, we are confident that the broad 2-fold enrichment observed represents the *bona fide* location of the DNA polymerase and MutS on the chromosome in the plurality of the cells in our culture synchronized for replication initiation. We conclude that MutS is physically associated with the site of ongoing DNA synthesis and is rapidly loaded at *oriC* upon chromosomal replication initiation.

Furthermore, we observed from co-immuno¬precipitation (co-IP) that MutS binds both PolC and DnaE *in vivo* before mismatch formation (Figure 2.10B). Using a far western blot, we also detected direct *in vitro* interaction of MutS with PolC and DnaE (Figure 2.11). Because we carried out the co-IP with the reversible membrane-permeable crosslinker DSP, it is possible that the co-IP of PolC and DnaE

with MutS occurs indirectly via MutS interaction with β-clamp. Taken together with our imaging and ChIP-seq results, these data suggest that whether the interaction of MutS with PolC and DnaE *in vivo* is direct or indirect, in the absence of 2-AP, MutS is capable of searching DNA in extremely close proximity (DSP has a spacer arm length of 12.0 Å) to the actively replicating DNA polymerase complex, and with 2-AP in the growth medium, MutS scans newly replicated DNA in clamp zones trailing the replisome.

2.5 Discussion

In the crowded cellular environment, 3D diffusion alone is too slow to allow for efficient detection of base-pairing mistakes (*von Hippel and Berg*, 1989). Previous *in vitro* studies have demonstrated that 1D sliding is another mechanism by which MutS can locate DNA mismatches (*Gorman et al.*, 2012). To further our understanding of the corresponding process from the *in vivo* perspective, and to complement and clarify our existing knowledge of MMR across species, we performed single-molecule super-resolution microscopy in live *B. subtilis*. With nanometer-scale spatial resolution and millisecond-scale temporal resolution, we directly visualized and quantified the behavior of MutS *in vivo* in real time. We find that MutS molecules move to and from the replisome even under normal growth conditions with exceedingly low mismatch formation rates. When MutS reaches the replisome it dwells for $\geq$ 200 ms before releasing and exploring the rest of the nucleoid. *In vitro* single molecule studies suggest that MutS can search 700 bps/sec (*Jeong et al.*, 2011). In our studies, MutS molecules dwell for $\geq$ 200 ms, we therefore suggest that when MutS enters the replisome region it searches on average about 140 bps before diffusing away from the replisome. MutS is also highly dynamic throughout the nucleoid, a behavior that has not been observed before. We found that the diffusion rate of MutS increases with separation distance from the replisome and following mismatch binding. We demonstrate that mismatch detection by MutS must occur near the site of DNA replication and that MutS remains

associated with mismatches until MutL is recruited. We have also revealed a possible MutL-dependent mechanism for ATP-bound MutS hopping between molecules of DNA *in vivo*. Furthermore, we show that MutS is loaded at the origin with the replisome in synchronized cells and broadly associates with newly replicated DNA. Investigation of protein-protein interactions has uncovered a physical interaction between MutS and the two essential DNA polymerase catalytic subunits PolC and DnaE, providing further evidence that MutS interaction with the replisome is not simply restricted to processivity clamps. Our study provides a single-molecule view of MutS during the search for mismatches in live cells revealing a highly dynamic process where MutS molecules constitutively move to and from the replisome rapidly and dwell at the replisome for short intervals.

Some of the MutS diffusion coefficients measured *in vivo* in the current study differ from those observed by *in vitro* Förster resonance energy transfer (FRET) experiments (*Cho et al.*, 2012). We observed MutS diffusion faster than 0.2 $\mu m^2/s$ at sites distant from the replisome region ($\geq$ 200 nm) for WT cells treated with 2-AP (WT+). Similar *in vivo* diffusion coefficients, and similar discrepancies in diffusion coefficients measured *in vivo* and *in vitro*, have also been previously reported (*Elf et al.*, 2007), and it has been suggested that the higher diffusion coefficients measured *in vivo* likely result from the fact that *in vivo* experiments simultaneously track proteins in 1D and 3D. Thus, we propose here that detection of a mismatch *in vivo* is followed by the formation of a stable, quickly sliding ATP-bound MutS clamp as has been observed *in vitro* (*Jeong et al.*, 2011; *Gorman et al.*, 2012). However, in the complex environment of the cell, an ATP-bound MutS clamp may be able to transiently move between molecules of DNA or around obstacles in DNA through a 3D diffusion process. Indeed, studies of purified *E. coli* MutS and MutL have shown that MutL binds ATP-bound MutS, effecting its release from mismatched DNA (*Acharya et al.*, 2003). Although *E. coli* MutL differs significantly from MutL in *B. subtilis*, such a hopping behavior of ATP-bound MutS *in vivo* would

manifest itself in a higher diffusion coefficient than that observed *in vitro* on a 1D DNA substrate, and this diffusion coefficient would depend on mismatch binding.

Our results reveal that the heterogeneous behavior of MutS depends on mismatch binding state as well as intracellular location, MutL expression, and nucleotide binding by MutS. We have demonstrated that nucleotide binding is necessary for the MutS cycle within the cell, and that if MutS cannot engage in nucleotide binding and hydrolysis, it is unable to properly localize to the replisome region *in vivo*. Because MutS[F30A] is recruited to the replisome, we can effectively rule out the possibility that any potential defect in mismatch recognition by MutS[K608M] is responsible for its change in diffusion coefficient or localization. We show that there is a highly dynamic and transient interplay between MutS and the replisome which positions MutS to sites of ongoing DNA replication before mistakes occur such that MutS can constantly monitor the newly synthesized DNA. Such behavior is somewhat similar to observations with MutSα in eukaryotic cells (*Hombauer et al.*, 2011a), suggesting that replisome-association is a highly conserved mechanism for mismatch detection across species. On the other hand, unlike MutSα which may be able to bind mismatches independent of the replisome by engaging in Exo1-dependent MMR (*Hombauer et al.*, 2011a), our results obtained from *B. subtilis* suggest that mismatch binding by MutS must occur at the replisome. One possible explanation for this difference in mechanism between bacteria and *S. cerevisiae* is that bacterial DNA replication occurs continuously, each replication fork moves at a rate of $\approx$ 500 nt/sec, and in rich growth conditions, multiple rounds of replication initiation can occur prior to a single cell division. Conversely, in eukaryotic cells, DNA replication occurs only in S-phase, the replication fork moves 27 nt/sec, and the behavior of biochemical pathways can be regulated in a cell cycle-dependent manner (*Wang et al.*, 2007; *Sekedat et al.*, 2010; *Tomimatsu et al.*, 2014). Therefore, mismatches must be detected very quickly in a rapidly proliferating bacterium like *B. subtilis* or they will become mutations within

mere minutes when the next round of DNA replication duplicates the mismatched DNA. The large volumes of eukaryotic nuclei compared to typical bacterial cell volumes could also contribute to the emergence of this second mechanism to catch replication errors that have escaped initial proofreading at the replisome in a much larger cell volume. In contrast, in small bacterial cells with confined replisomes such as *B. subtilis*, near the replisome is arguably the best place for MutS to scan DNA for mismatches barrier-free. At the replisome, MutS has access to newly synthesized DNA largely free of proteins, and is also able to recognize rare replication errors as they are produced due to the spatial proximity between the two. Therefore, the replisome provides a scaffold that allows MutS to target a single mistake among tens of millions of correctly paired nucleotides in a timely manner, guarding the bacterial genome against mutations that could otherwise have deleterious effects on bacterial growth and fitness.

2.6 Materials and Methods

2.6.1 Sample preparation for single-molecule imaging.

Bacillus subtilis PY79 cells were grown at 30 °C in $S7_{50}$ minimal medium with starting OD600 $\approx$ 0.1. To induce ectopic expression of DnaX-mCitrine, 0.125% xylose was added to the medium. To cause mismatch formation, 2-AP was added at OD $\approx$ 0.35 to a final concentration of 0.6 mg/mL, and cells were harvested during exponential phase when OD reached $\approx$ 0.55 – 0.65. HPUra (final concentration 162 μM) was added to the culture immediately prior to imaging. 2 μL of cell culture was pipetted onto a 1% agarose in $S7_{50}$ pad, which was sandwiched between two coverslips that had been cleaned by oxygen plasma (Plasma Etch PE50) for 20 minutes. The sample was then mounted on the microscope objective for imaging.

2.6.2 Microscopy and imaging parameters.

DnaX-mCitrine was first imaged under 2.1 x 10^{-4} mW/cm^2 488-nm laser illumination (Coherent Sapphire 488-50) and then MutS-PAmCherry1 was photoactivated using a 200-ms 405-nm pulse (Coherent 405-100) and imaged with a 561-nm (Coherent Sapphire 561-50) laser with power densities of 6.6 x 10^{-5} mW/cm^2 and 1.0 x 10^{-4} mW/cm^2, respectively. For photobleaching-assisted DnaX-mCitrine localization, a higher power density of 2.0 x 10^{-3} mW/cm^2 was used. Widefield single-molecule epifluorescence microscopy was performed on an Olympus IX71 inverted microscope. Fluorescence emission was collected by a 1.40-NA 100x oil-immersion phase-contrast objective and detected on a 512x512 pixel Photometrics Evolve EMCCD at a rate of 25 Hz. The photobleaching-assisted DnaX-mCitrine localization experiment was performed using a frame rate of 50 Hz. Appropriate dual-color dichroic and band-pass filters (Semrock) in the emission pathway rejected scattered laser light and maximized the signal-to-noise ratio.

2.6.3 Post-processing of fluorescence images

Raw fluorescence images were band-pass-filtered to produce initial estimates for the locations of molecules. The precise location of each candidate molecule was then determined by fitting single-molecule emission intensity profiles in the raw image to symmetric 2D Gaussian functions using homebuilt MATLAB code. Localization precision was determined by imaging individual PAmCherry1 and mCitrine proteins immobilized on a coverslip with polymethyl methacrylate (PMMA) under the same imaging conditions as in live-cell experiments (*Pavani et al.*, 2009).The full-width-at-half-maximum (FWHM) of the localization distribution gives a localization precision of 35 nm for PAmCherry1 and 41 nm for mCitrine. The localization precision for mCitrine is likely to be worse in photobleaching-assisted microscopy experiments due to the relatively low signal-to-noise ratio caused by high and fluctuating cellular

background fluorescence and the fact that single-molecule PSFs are indirectly obtained by taking the intensity difference between different frames. Localized DnaX-mCitrine signals were grouped into clusters using the k-means algorithm. To quantify the size of each cluster, the radius of gyration was calculated according to:

$$R_g = \left[\frac{1}{N} \sum_{k=1}^{N} (r_k - r_{mean}) \right]^{\frac{1}{2}} \tag{2.1}$$

where N and r respectively denote the total number of localizations belonging the cluster and the position of each localization.

PALM super-resolution reconstruction images were obtained by plotting the localization of each fit convolved with a Gaussian blur with width equal to its localization uncertainty (*Biteen et al.*, 2008). Trajectories were constructed by linking localizations from consecutive frames such that the sum of the distance between each pair was minimized. Single-molecule guessing, fitting and tracking were all performed within cell boundaries to avoid false signal identifications and artificial trajectories crossing to neighboring cells.

2.6.4 Localization probability density maps

Cells shown in phase-contrast images were segmented using the “valley” filter as described by Sliusarenko *et al.* (*Sliusarenko et al.*, 2011). The cell contour is rotated such that its principal axes are aligned with the image frame, and coordinates of single-molecule localizations are normalized with respect to the rotated cell contour. Based on the probability of finding a molecule in a certain region within the cell, two 2D localization probability density maps are constructed for each cell, one for DnaX-mCitrine and the other for MutS-PAmCherry. The final density maps are obtained by averaging localization probabilities over all cells with 2 DnaX clusters and at least 100 fits in both color channels. Because there is no distinction between the left and the

right or between the top and the bottom in *B. subtilis* regarding DNA replication and mismatch repair, the maps are symmetrized with respect to the cell center (Figure 2.6) (*Wang et al.*, 2011).

2.6.5 ChIP-sequencing (ChIP-seq)

ChIP was performed using a temperature-sensitive replicative helicase loader, *dnaB134* (*Burnett and Wake*, 1977), essentially as previously described (*Breier and Grossman*, 2009), with the following protocol modification: cultures were grown at 30 °C in $S7_{50}$ minimal medium supplemented with 1% glucose to OD600 $\approx$ 0.35, at which time they were shifted to the restrictive temperature (46 °C) for replication initiation for 1 hour. To initiate DNA replication synchronously, cultures were quickly shifted to 30 °C by addition of the appropriate amount of 4 °C medium. The experiment was performed twice, once using 0 minute and 10 minute time points, and the second time using 0.5 minute and 10 minute time points. Because the 0 minute and 0.5 minute time points yielded the same outcome, the data were combined during analysis. Therefore, analysis of the pre-initiation and the 10 minutes after initiation time points is based on the combined data from two independent experiments. Formaldehyde crosslinking was performed for ten minutes at room temperature by adding formaldehyde and sodium phosphate buffer, pH 7.4, to final concentrations of 1% and 10 mM, respectively. After crosslinking, cultures were moved to 4 °C for 30 minutes, and cells were pelleted and washed twice with phosphate-buffered saline (PBS), pH 7.4. Cell pellets were resuspended in 500 μL solution A (10 mM Tris, pH 8, 20% sucrose, 50 mM NaCl, 10mM EDTA) containing 1 mg/mL lysozyme and 10 μL protease inhibitor cocktail (Thermo Scientific, product number 78410) and incubated for 30 minutes at 37 °C. This was followed immediately by addition of 500 μL of 2x IP buffer (100mM Tris, pH 7, 300mM NaCl, 2% Triton X-100) to lyse cells. The mixture was incubated 10 minutes at 37 °C.

DNA was sheared by sonication, and insoluble debris was removed by centrifugation. 3.3 μL of the supernatant was saved for input, and the remaining supernatant was split three ways (333 μL each for MutS ChIP, PolC ChIP, and DnaE ChIP). Chromatin immunoprecipitation was performed using affinity-purified antisera directed against PolC, DnaE, or MutS for 1 hour at 25 °C, followed by addition of 20 μL Millipore Magna ChIP protein A+G magnetic bead slurry for 1 hour at 25 °C.

Magnetic beads were collected on a magnetic rack and washed 5 x 5 minutes with 1x IP buffer. The beads were then collected and washed 2 x 5 minutes with 1 mL TE (10 mM Tris, pH 8, 0.1 mM EDTA). Prior to the final TE wash, the beads were moved to a fresh microcentrifuge tube. The washed beads were resuspended in 50 μL of TE, the 1% input sample was brought to 50 μL with TE, and crosslinks were reversed by incubation of both the 1% input and ChIP samples at 65 °C for 15 hours. Beads were collected and supernatant was saved. The beads were washed for 5 minutes at 37 °C in 50 μL TE. Beads were collected and supernatant combined with previously collected supernatant. Proteinase K was added to a final concentration of 0.1 mg/mL and samples (1% input as well as ChIP samples) were incubated at 55 °C for 2 hours. Phenol/chloroform extraction was performed followed by ethanol precipitation to purify the input and ChIP DNA.

ChIP and input DNA was amplified using the SEQXE kit from Sigma-Aldrich, and library preparation and sequencing was performed at the University of Michigan DNA Sequencing Core. 50-base single-end sequencing was carried out using an Illumina HiSeq 2000 instrument.

2.6.6 ChIP-seq analysis

50-base single-end ChIP-seq data were aligned to the *B. subtilis* PY79 reference genome, accession number CP006881.1 (*Schroeder and Simmons*, 2013) using bwa v0.5.9-r16 (*Li and Durbin*, 2009) and alignment files were generated using samtools (*Li*

et al., 2009). Only uniquely aligning reads were kept for analysis and PCR duplicates were removed. The statistical package R was used to determine enrichment of ChIP DNA versus input using the R package SPP (*Kharchenko et al.*, 2008).

2.6.7 Co-Immunoprecipitation (Co-IP)

Cells were grown in $S7_{50}$ minimal medium with 1% glucose to OD600 $\approx$ 0.7 and harvested by centrifugation. Pellets were twice washed with crosslinking buffer (40 mM HEPES, pH 7.4, 500 mM Sucrose, 2 mM $MgCl_2$, 150mM NaCl, 0.02% Tween-20) and resuspended in 0.5 mL crosslinking buffer per 10 OD x mL cells. The membrane-permeable crosslinker dithiobis-succinimidyl propionate (DSP) was added to a final concentration of 0.5 mM and crosslinking was carried out for 30 minutes at 25 °C with constant mixing. Excess DSP was quenched by adding Tris-HCl, pH 8.5 to a final concentration of 20 mM, and quenching was allowed to proceed for 30 minutes at 25 °C. Cells were lysed by sonication, and insoluble debris collected by centrifugation. The supernatant was collected and concentrated to a volume of 0.5 mL and protease inhibitor cocktail and DNAse were added. 5% of the resulting solution was set aside for the input fraction. The remaining 95% was used for immunoprecipitation of MutS using affinity-purified antibodies bound to protein A/G magnetic beads (Thermo Scientific Prod number 88803). Immunoprecipitation was carried out overnight at 4 °C.

Beads were washed 5 x 5 minutes in 0.5 mL crosslinking buffer. Beads were then moved to a fresh microcentrifuge tube, collected, and washed once in 0.5 mL crosslinking buffer. Proteins were eluted by washing 3 x 5 minutes in 0.3 mL antibody stripping buffer (5 mM glycine, pH 2.4, 150 mM NaCl), with each wash collected and combined into a single microcentrifuge tube. 90 μL of 1M Tris-HCl, pH 8.5 was added to the eluate, and DTT was added to a final concentration of 50 mM to both the immunoprecipitation and input fractions to reverse crosslinks. Crosslinks were reversed for 30 minutes at 37 °C.

Proteins in the immunoprecipitation fraction were concentrated via TCA precipitation. The pellet was dissolved in 1x SDS-PAGE sample buffer and sample buffer was added to the input fraction to a 1x final concentration. The input and immunoprecipitation were subject to SDS-PAGE and Western blot to detect MutS, DnaE and PolC using affinity-purified antisera.

2.6.8 Far Western blotting

Far Western blotting was performed as previously described (*Walsh et al.*, 2012). Purified PolC, DnaE and MutL were spotted onto nitrocellulose in the indicated amounts. BSA was used as a negative control. BSA was obtained from New England Biolabs and spotted at amounts equal to those of the purified proteins. MutS retained on the membrane was detected with affinity-purified antisera.

2.6.9 Protein purification

MutL, PolC and DnaE were overexpressed recombinantly in *E. coli* from plasmid pET28a-PB, which encodes an N-terminal hexahistidine tag cleavable by Prescission protease, leaving the tetrapeptide GPGS on the N-terminus of the purified protein. Proteins were overexpressed overnight at 20 °C using 200 μM IPTG and cells pelleted by centrifugation. Pellets were resuspended in buffer HA (50 mM Tris-HCl, pH 7.4, 300 mM NaCl, 1mM -mercaptoethanol, EDTA-free protease inhibitor cocktail (Thermo scientific product number: 78410)) with 10% sucrose, 20 mM spermidine-trihydrochloride and 20 mM imidazole. Lysis was carried out via French press and lysate was cleared immediately by centrifugation followed immediately by running cleared supernatant over a HisTrap HP column (GE product number: 17-5247-01). The column was washed with buffer HA containing 40 mM imidazole and 500 mM NaCl. Bound proteins were eluted with a linear gradient of buffer HA with 40 mM imidazole to buffer HA with 400 mM imidazole over 20 column volumes. Fractions containing the protein of interest

were dialyzed against buffer HA with GST-tagged PreScission protease overnight at 4 °C. The dialysate was run through a HisTrap HP column and flow-through was collected and GST-PreScission protease was removed by binding in batch to glutathione-agarose resin.

PolC and DnaE were desalted into buffer QA (20 mM Tris-HCl, pH 7.4, 50 mM NaCl, 1 mM DTT, 5% (w/v) glycerol). Protein solutions were applied to a HiTrap Q column, washed for 10 column volumes with buffer QA, and eluted with a linear gradient over 20 column volumes to buffer QA with 600 mM NaCl. Fractions with PolC or DnaE were then purified using gel filtration using buffer SE (50 mM HEPES, pH 7.4, 150 mM NaCl, 5% (w/v) glycerol, 1 mM DTT). Once purified, buffer exchange was carried out on a desalting column such that MutL was stored in 50 mM HEPES, pH 7.4, 300mM NaCl, 5% (w/v) glycerol, 1 mM DTT.

MutS and MutS variants were overexpressed recombinantly in *E. coli* as a fusion protein N-terminally tagged a with hexahistidine-SUMO tag. Overexpression was carried out using 200 μM IPTG for 3 hours at 37 °C. Cells were pelleted by centrifugation, resuspended in buffer HA with 20 mM spermidine-trihydrochloride and 20 mM imidazole. Cells were lysed via sonication (10 cycles of 10 second pulse at 85% amplitude with 20 seconds rest on ice). Unclarified lysate was immediately purified using a 5 mL HisTrap FF Crude column (GE product number: 17-5286-01) and the column was washed with buffer HA containing 10% (w/v) glycerol, 40 mM imidazole, and 500 mM NaCl. Protein was eluted from the column with buffer HA containing 10% (w/v) glycerol, 400 mM imidazole. SUMO protease was added to the pooled MutS-containing fractions and the solution was dialyzed overnight at 4 °C against 1 liter of buffer SB (50 mM Tris-HCl, pH 8.0, 300 mM NaCl, 10% (w/v) glycerol, 1 mM -mercaptoethanol). Dialysate was exposed for 3 hours at 4 °C to Ni-NTA agarose beads equilibrated in buffer SB. Mixture was applied to a gravity-flow column and flow-through was collected, concentrated to 1.5 mL using a centrifugal filter unit, and

applied to a size exclusion column using 50 mM Tris-HCl, pH 7.4, 150 mM NaCl, 1 mM DTT. Fractions containing MutS were concentrated using a centrifugal filter unit.

2.6.10 ATPase assay

ATPase assays were carried out essentially as described (*Junop et al.*, 2003), with some minor modifications. DNA oligonucleotides (Table 2.2) were mixed and heated to 98 °C, allowed to cool slowly to room temperature, and protein was added. ATP was added to initiate the reaction. Components were used in the following final concentrations: 1 μM MutS, 10 μM dsDNA, varying ATP concentrations from 12.5 μM up to 200 μM. Reaction buffer consisted of 50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1 mM DTT, 5% (w/v) glycerol, 5 mM MgCl2. 15 μL reactions were allowed to proceed for 10 minutes in the case of MutS[K608M], 4 minutes in the case of MutS800, and 1 minute in the case of MutS, at which time 15 μL of 50 mM EDTA was added. Reactions without protein were analyzed for background subtraction. 2 μL of each reaction was spotted onto a PEI-cellulose TLC plate and 0.75 M K_2HPO_4 was used as the mobile phase. Hydrolysis of ATP was observed using 0.45 μCi (9.7 nM final concentration in each reaction) α-[^{32}P]-ATP (Perkin Elmer product number BLU003H250UC). Images were captured on a phosphorimager and quantified using ImageJ software (*Schneider et al.*, 2012). The Michaelis-Menten equation was fit to the data using the statistical software R.

2.6.11 Peptide array analysis

Peptide array analysis was performed essentially as described previously (*Lenhart et al.*, 2013a) using 100 nM myc-MutS as prey and 1 mM ADP with 2 mM MgCl.

2.6.12 Determination of mutation rate

Determination of the mutation rate resulting in rifampin resistance was performed as described previously (*Walsh et al.*, 2014).

2.7 Supplemental Text

2.7.1 MutS binds the PolC and DnaE replicases *in vitro*.

We tested whether MutS binds the catalytic subunits of the *B. subtilis* replisome with far-Western blot analysis. PolC, DnaE, MutS and MutL were purified, and PolC, DnaE and MutL were applied to a nitrocellulose membrane in serial dilutions such that 0.25-10 pmol of each protein was immobilized on the blot as described (*Klocko et al.*, 2011). BSA was also immobilized on the membrane as a negative control. MutS was incubated with the membrane followed by detection using affinity-purified antiserum against MutS. We were able to detect MutS retention on the membrane by PolC and DnaE down to 0.25 and 0.5 pmol quantities, respectively. We observed retention of MutS to the 2.5 pmol quantity of immobilized MutL, the positive control. The negative control BSA did not retain any MutS (Figure 2.11A). We interpret this result to mean that MutS binds PolC and DnaE more strongly than MutS binds MutL.

To identify potential binding sites important for MutS interaction with PolC, we performed a peptide array analysis (Figure 2.11B). In this experiment epitope-tagged MutS (MutS-myc) was incubated with a peptide array representing the entire PolC amino acid sequence as 475 12-mer oligopeptides offset by 3 amino acids each. We determined that distinct surface-exposed peptides retained MutS-myc onto the array. We found that most of the peptides identified in this analysis mapped to the face of PolC where replicated DNA exits. These data suggest that MutS could associate with PolC to bind newly formed mismatches in nascent DNA.

2.7.2 Localization of the processivity clamp

In our study we marked the replisome positions in cells with mCitrine fusions of the clamp loader protein DnaX. To determine that DnaX-mCitrine localizes in the same way as other replisome subunits, we imaged the processivity clamp protein DnaN-mCitrine

in live *B. subtilis* cells (Figure 2.12). Similar to DnaX-mCitrine, DnaN-mCitrine forms compact clusters at quarter positions as well as at cell centers. These results also demonstrate the highly confined localization for both DnaX and DnaN. Because the DnaN-mCitrine fusion partially compromises the MMR activity *in vivo*, we used DnaX-mCitrine as a proxy to the replisome throughout the study.

2.7.3 Enrichment of mismatches by addition of 2-aminopurine

2-aminopurine is a base analog that crosses the bacterial cell membrane efficiently followed by conversion to its deoxynucleotide triphosphate form (2-APTP) (*Ronen*, 1979). 2-AP is likely to be incorporated into genomic DNA by base pairing with T. In the next round of DNA replication, a C may then be placed across from 2-AP. At a concentration of 7.4 mM 2-AP the probability that an AT base pair will be mutated to GC is a 1.5 x 10^{-5} (*Goodman et al*., 1977; *Ronen*, 1979). Under our experimental conditions, with 600 μg/mL (4.44 mM) 2-AP, we expect a 9 x 10^{-6} probability that a given AT base pair will be converted to GC with 2-AP treatment. The probability of observing a cell which accumulates a given number of mismatches in its DNA during a specified window of time is a Poisson process:

$$P\left[N(t) = k\right] = \frac{(Rt)^k * e^{-Rt}}{k!} \tag{2.2}$$

where k is the number of mutations observed, R is the mutation rate in terms of time, and t is the duration of time the cell is observed.

To determine the probability that a single cell has at least one mismatch during time, t, we calculate the complement of P[N(t) = 0]:

$$P\left[N(t) = 0\right] = e^{-Rt} \tag{2.3}$$

$$P\left[N(t) > 0\right] = 1 - e^{-Rt} \tag{2.4}$$

In *E. coli*, the base pair substitution (BPS) rate in the absence of MMR is 273 x 10^{-10} BPS/base replicated (*Lee et al.*, 2012). We assume this approximates the mismatch incorporation rate of the replicating DNA polymerase, and applying this mismatch rate in the absence of 2-AP to *B. subtilis*, and using an overall replication rate of 1000 nt/s (combined rate, i.e., two forks) (*Wang et al.*, 2007), we calculate a mismatch rate of 2.73 x 10^{-11} mismatches per second. This calculation ignores insertion/deletion rate, which will not significantly affect the resulting probability of observing a mismatch in a cell during a 210 second observation. We determine that during a 210 second window of time, which is the approximate time each cell is observed in our PALM experiments, we will observe a cell with a mismatch with a probability of 5.7 x 10^{-9}. This indicates that in untreated conditions, we effectively never observe a cell in which a mismatch arose during our observation period.

Under conditions of 2-AP treatment during the time of our observation the probability of an observed cell obtaining a mismatch is quite high. To calculate R under conditions of 2-AP treatment, we consider that with 2265514 AT pairs in the *B. subtilis* PY79 genome, 2-AP treatment causes approximately 20.4 (2265514 x 9 x 10^{-6}) AT$\rightarrow$GC transitions per round of replication (*Schroeder and Simmons*, 2013). With 4033459 total base pairs in the genome, there are about 197719 (4033459/20.4) base pairs on average between mismatches caused by 2-AP. Two replication forks replicate the genome, each polymerizing DNA at a rate of 500 nucleotides per second, for a combined rate of 1000 nt/s (*Wang et al.*, 2007). This means there is one 2-AP insertion resulting in an AT$\rightarrow$GC transition every 197.7 seconds, or 5.05 x 10^{-3} mutations caused by 2-AP per second. We have excluded the natural rate of mismatch formation from this calculation, because it is so low that it is negligible. Using equation 2.4, we find there is a probability of 0.65 that a 2-AP-treated cell will have at least one mismatch during our 210 second

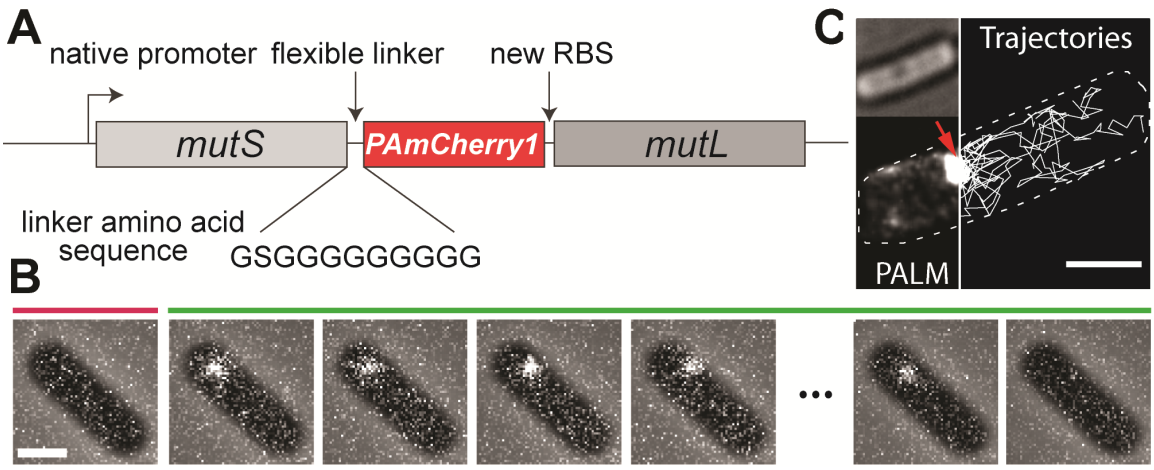
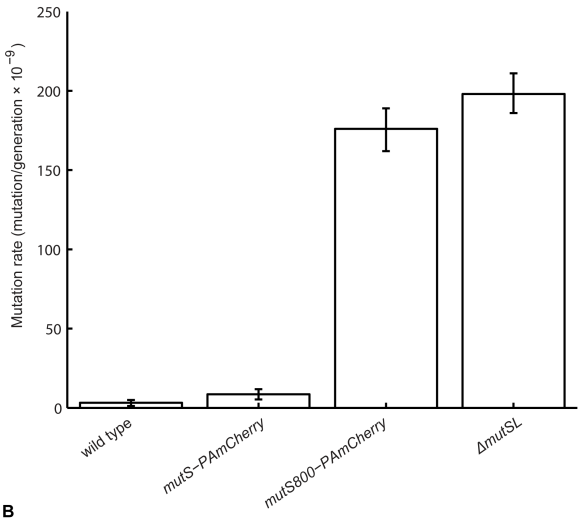


Figure 2.1: Location and dynamics of MutS in live *B. subtilis*. (A) Labeling scheme for MutS-PAmCherry. RBS: ribosome binding site. (B) Representative frames showing the photo-activation of a single copy of MutS-PAmCherry in a cell. Purple and green lines above the frames correspond to the photo-activation laser pulse and the imaging laser. (C) PALM reconstruction (lower left) and single-molecule trajectories (right) of MutS-PAmCherry in a live *B. subtilis* cell (upper left). Each sub-diffraction-limited coordinate of MutS-PAmCherry is plotted in the PALM image as a Gaussian blur with width equal to its localization uncertainty. Red arrow: region of MutS accumulation. White dashed lines: computer-detected cell boundary. Scale bars: 1 μm.

observation.

2.8 Figures and Tables

A



B

Strain	Mutation rate	95% CI	% MMR activity
wild type	3.29	1.16 – 4.94	100
mutS-PAmCherry	8.60	5.31 – 11.8	97
mutS800-PAmCherry	176	162 – 189	12
ΔmutSL	198	186 – 211	0

C

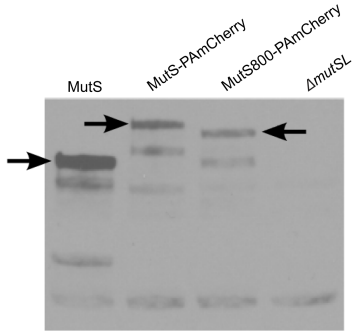


Figure 2.2: MutS-PAmCherry is functional in mismatch repair and is stably expressed *in vivo*. (A) Mutation rates of the indicated strains. Rates were determined by maximum likelihood estimation implemented in the online tool FALCOR (*Hall et al.*, 2009). (B) Mutation rate and 95% confidence interval (expressed as mutations per generation × 10^9), and percent mismatch repair of the indicated strains. (C) Western blot using antiserum directed against MutS. Full-length proteins are indicated by arrows. Truncated protein fragments are due to nonspecific proteolysis, and not due to cleavage of the PAmCherry tag from MutS, as they appear in the wild-type MutS lane as well as the MutS-PAmCherry and MutS800-PAmCherry lanes.

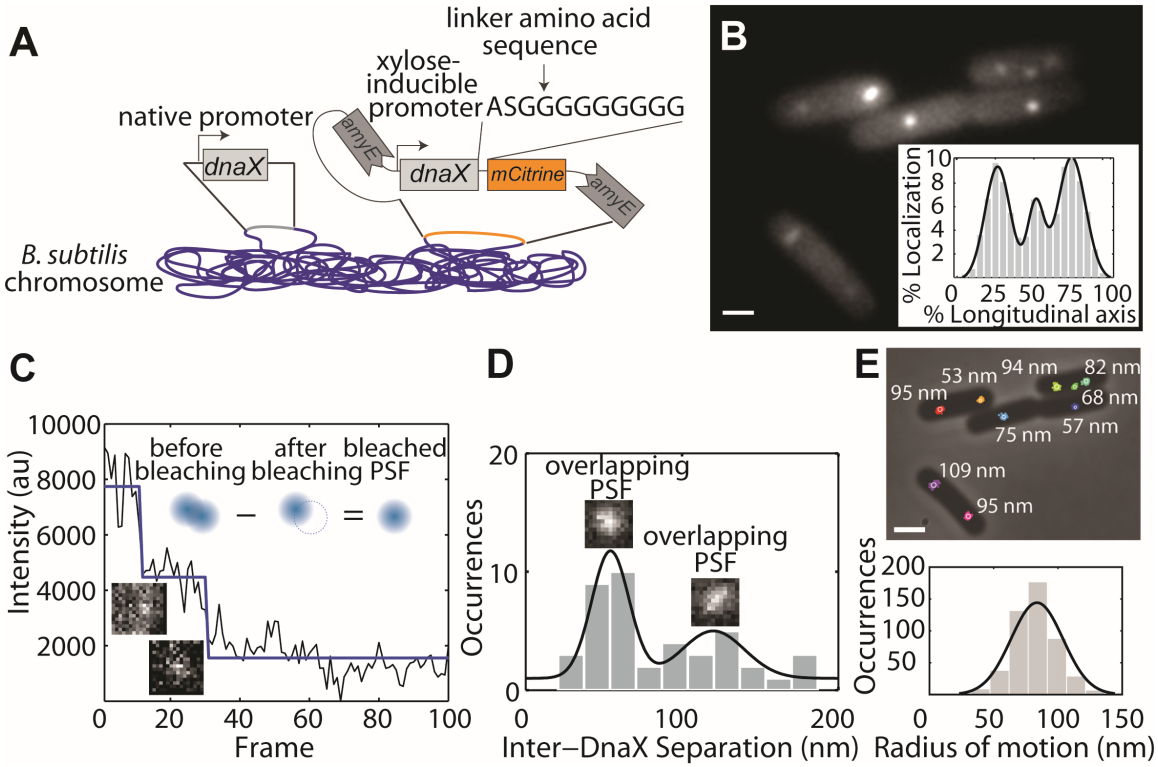


Figure 2.3: Positioning of the replisome. (A) Labeling scheme for DnaX-mCitrine. (B) Sample fluorescence image of DnaX-mCitrine in five cells and localization probability of DnaX in 161 cells along the longitudinal cell axis (inset). The replisome appears most frequently at the quarter positions in log-phase cells. (C) Photobleaching-assisted localization of single DnaX-mCitrine molecules within a cluster. The intensity of a cluster is plotted against time (20 ms/frame) and photobleaching events are identified by maximum likelihood estimation (*Watkins and Yang*, 2005). The PSFs of photobleached single molecules can then be obtained by subtracting the average intensity of the frames following the photobleaching from that of preceding frames (schematic representation shown in inset), allowing the precise location of the photobleached molecule to be determined through 2D Gaussian fitting. Representative PSFs for two photobleached mCitrine molecules are shown above the cluster intensity trajectory. (D) Distribution of separation distance between DnaX-mCitrine within a replisome as determined from photobleaching-assisted localization, with two sample overlapping PSFs shown. (E) The radii explored by DnaX-mCitrine as calculated by tracking the motion of each cluster centroid using low-power time-lapse imaging (upper) and the size distribution of the domain explored by each cluster (lower), illustrating the subtle replisome motion that explores small domains of $\approx$ 84 nm in radius on average. Scale bar: 1 μm.

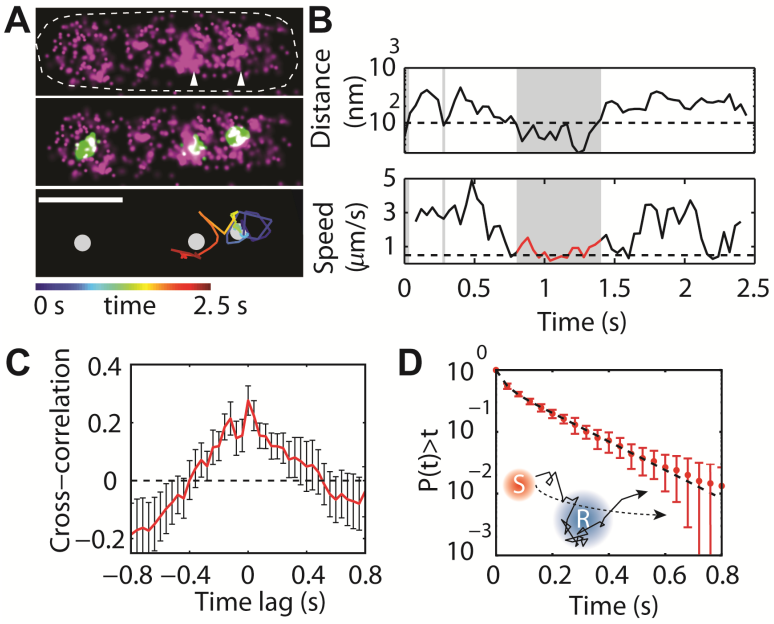


Figure 2.4: Two color imaging results from single cells expressing both MutS-PAmCherry and DnaX-mCitrine. (A) PALM reconstruction of MutS-PAmCherry (magenta) in a cell with MutS-enriched regions indicated with white arrowheads (upper), and overlaid with DnaX-mCitrine clusters (green) (middle). A representative time-coded trajectory showing MutS entering, dwelling at, and leaving one of the replisome regions is shown in the lower panel. Scale bars: 1 μm. (B) Separation distance from the replisome (upper) and instantaneous speed (lower) as a function of time for the MutS trajectory shown in (A). Grey windows: time spent in the replisome region. Red curve: a prolonged period of decreased MutS speed. Black dashed lines: 100-nm MutS-DnaX separation distance (upper) and average DnaX speed (lower; 0.5 μm/s, as measured by tracking cluster centroids). (C) Cross-correlation between the separation between MutS and the center of DnaX cluster and the instantaneous speed of MutS from 11 cells, normalized from −1 to 1 (*Lewis*, 1995). Error bars represent the standard error of the mean. (D) Cumulative probability distribution of time period MutS (red) spends within the same replisome region (blue), fit to a two-term exponential decay function (dashed line) $P = A_1 e^{(-t/\tau \times 1)} + A_2 e^{(-t/\tau \times 2)}$, where A_1 = 0.42, A_2 = 0.58, τ_1 = 25 ms and τ_2 = 188 ms. The error bars are standard deviations from 7 rounds of bootstrapping. The dwell time distribution is constructed using trajectories at least 10 frames long for molecules that can be tracked from the time they enter the replisome region until they leave the replisome (inset).

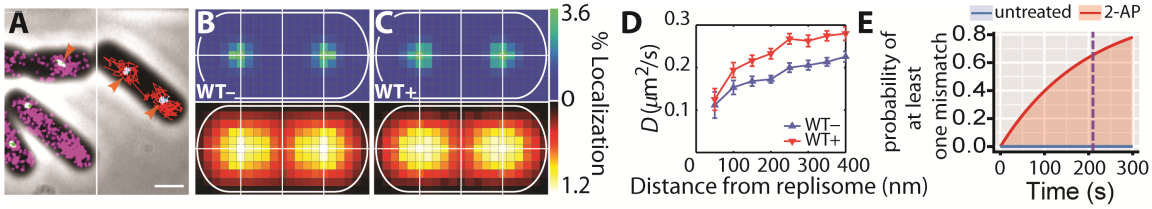


Figure 2.5: MutS localization and dynamics in WT cells. (A) PALM reconstruction (magenta) and single-molecule trajectories (red) of MutS-PAmCherry, overlaid with DnaX-mCitrine (green and blue) and phase-contrast cell images. Overlapping signals are colored in white. Orange arrows: replisome regions at which preferential MutS enrichment or dwelling is observed. Scale bar: 1 μm; (B and C) Localization probability density maps of DnaX-mCitrine (top; blue-green) and MutS-PAmCherry (bottom; red-yellow) within a normalized cell. White lines designate the $\frac{1}{4}$, $\frac{1}{2}$ and $\frac{3}{4}$ positions along the cell long axis and the $\frac{1}{2}$ position in the transverse direction. 108 WT– cells (B) and 91 WT+ cells (C) with two replisome clusters were used to generate the corresponding density maps. To allow for quantitative comparison of colocalization between different cases, the Pearson correlation coefficients between each pair of DnaX/MutS density maps are calculated. The correlation coefficients for WT– and WT+ cells are 0.83 and 0.81 respectively. Grid pixel size: 100$\approx$ 200 nm; (D) Diffusion coefficients of MutS-PAmCherry as a function of separation distance from the nearest replisome. Error bars: 95% confidence interval. (E) Distribution of the probability of a mismatch occurring in an observed cell under normal growth condition (blue) and under 2-AP treatment (red) over time. The vertical purple dashed line indicates the average duration (210 s) of observation for each cell in PALM experiments.

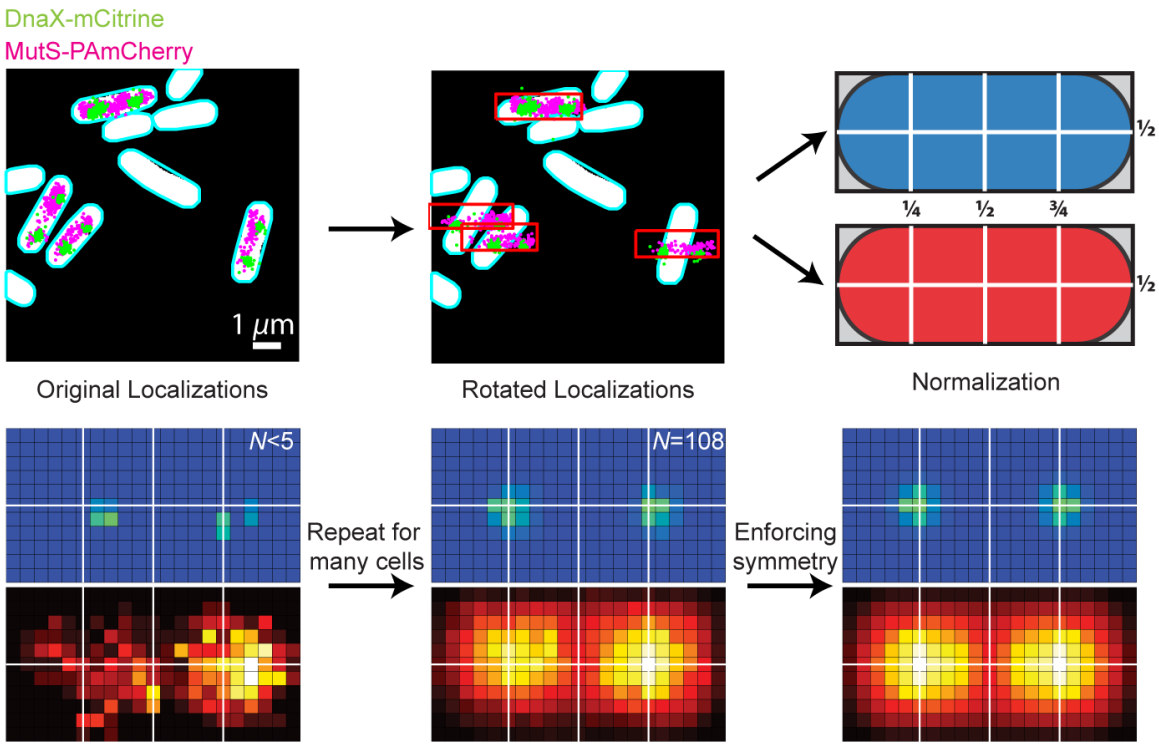


Figure 2.6: Localization probability density maps. Six steps in the procedure for generating localization probability density maps, using data from WT− cells (Figure 2.5B) as an example. All cells with two DnaX-mCitrine clusters and having $>$ 100 localizations in both color channels are analyzed.

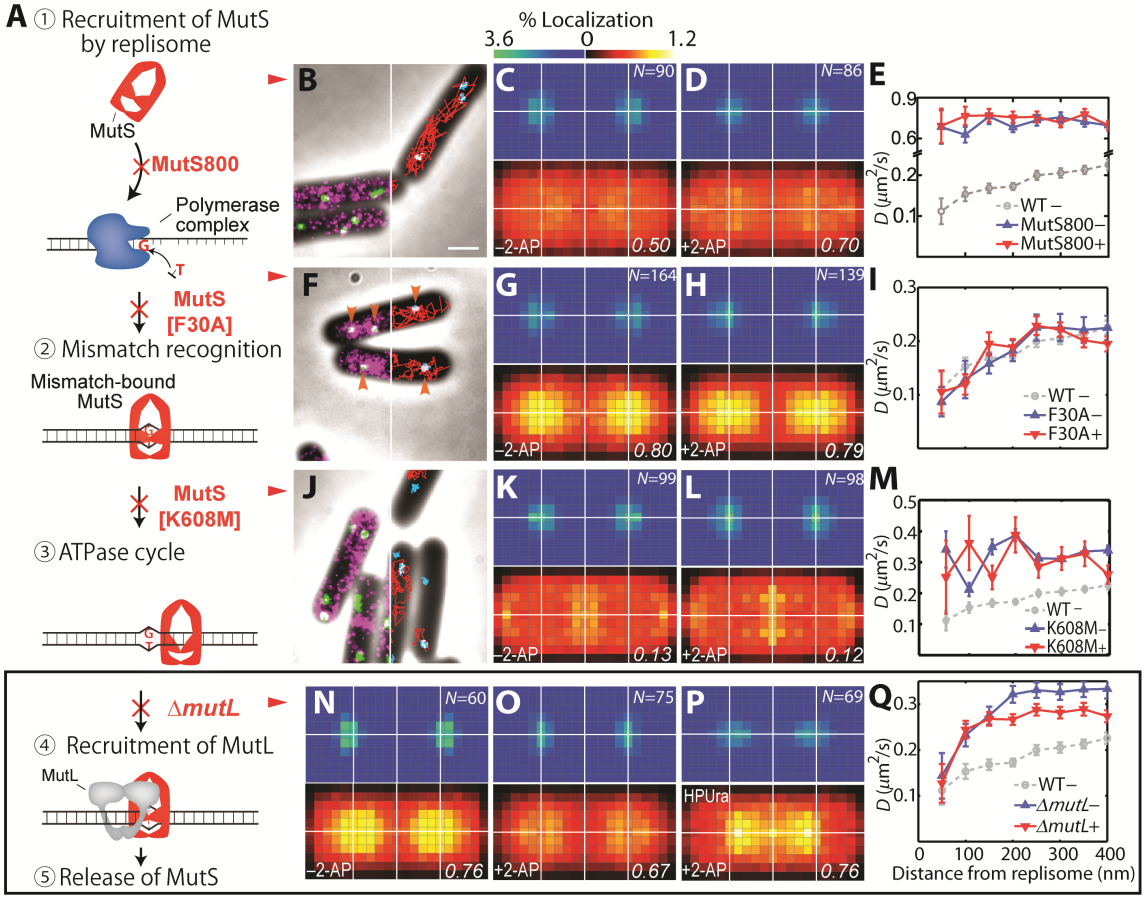
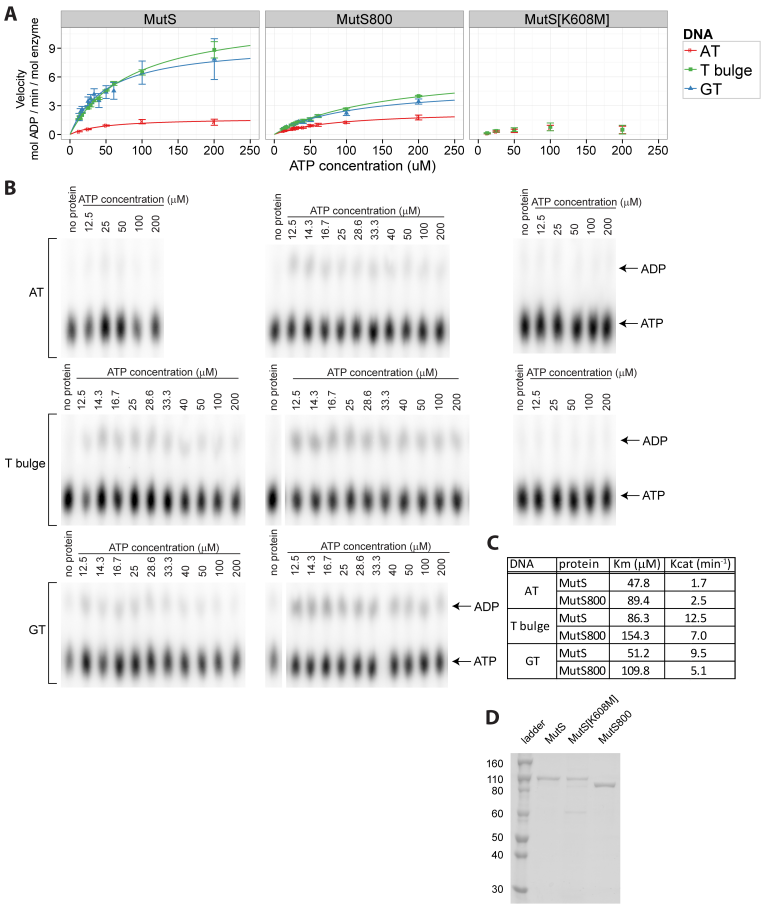


Figure 2.7: Response of MutS to sequential blocking of mismatch repair steps. (A) Schematic diagrams showing the first four steps of MMR, including replisome binding, mismatch recognition, ATPase activity and MutL recruitment, each of which is blocked in one of four mutant strains; (B, F, J) Two-color images of representative cells from MutS800, MutS[F30A] and MutS[K608M] strains. (C, G, K, N) Localization probability density maps of untreated cells for each strain, generated from N cells with two replisome clusters. Pearson correlation coefficients between DnaX and MutS densities are listed at the lower right corner of corresponding MutS density maps. Note that MutS density maps may exhibit similar intensity levels but have different correlation coefficients due to differences in the positioning of corresponding DnaX density maps. (D, H, L, O) Density maps of 2-AP treated cells. (P) Density maps for 2-AP/HPUra double treated cells from the Δ*mutL*+ strain. HPUra restores MutS enrichment around the replisome and also causes simultaneous shifting of DnaX and MutS localizations towards cell center. (E, I, M, Q) Diffusion coefficients of MutS-PAmCherry variants as a function of separation distance from the nearest replisome. Error bars: 95% confidence interval.



DNA	protein	Km (μM)	Kcat (min^{-1})
AT	MutS	47.8	1.7
	MutS800	89.4	2.5
T bulge	MutS	86.3	12.5
	MutS800	154.3	7.0
GT	MutS	51.2	9.5
	MutS800	109.8	5.1

Figure 2.8: ATPase activity of MutS, MutS800 and MutS[K608M]. ATPase assays were carried out using MutS or MutS800 in the presence of either homoduplex (AT) DNA, DNA with a GT mismatch, or DNA with a T bulge. In the case of MutS[K608M], ATPase assays were carried out with either AT DNA, or T bulge DNA. (A) Plots of ATPase activity for the indicated protein and DNA. Points and error bars represent the mean $\pm$ SEM of three independent experiments. Lines represent the Michaelis-Menten equation fit to the data. (B) Representative phosphorimaging results of TLC plates used for each protein/DNA combination. The positions of ADP and ATP on the plates are indicated on the right. (C) Parameters derived from fitting the Michaelis-Menten equation to the ATPase assay data for each protein/DNA combination. Km and Kcat were not determined for MutS[K608M]. MutS and MutS800 hydrolyze ATP similarly in the presence of homoduplex DNA, and the ATPase activity of both proteins are stimulated by the presence of mismatched DNA, but wild type MutS ATPase activity responds more strongly to mismatched DNA. (D) A coomassie-stained gel of purified MutS, MutS[K608M], and MutS800. 0.5 μg of each protein was subject to 7.5% SDS-PAGE.

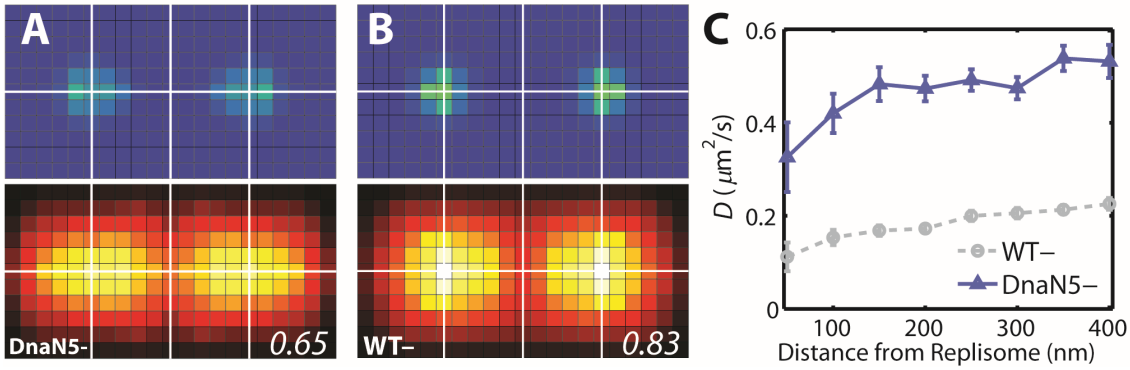


Figure 2.9: Localization and diffusion of MutS in the DnaN5 mutant strain compared to MutS from WT– cells. DnaN5 is a strain that contains a substitution that renders the β-clamp compromised for binding MutS. Here we use this strain to probe whether MutS/β-clamp interaction affects the recruitment of MutS to the replisome from a perspective complementary to MutS800. We observed a noticeable decrease in the correlation between MutS and DnaX from the DnaN5 strain (A) compared to the WT– case (B), although the attenuation in colocalization is not as severe as that observed in the MutS800– cells, presumably because DnaN5 retains some residual MutS binding. The diffusion rate of MutS is also increased in the DnaN5 strain (C).

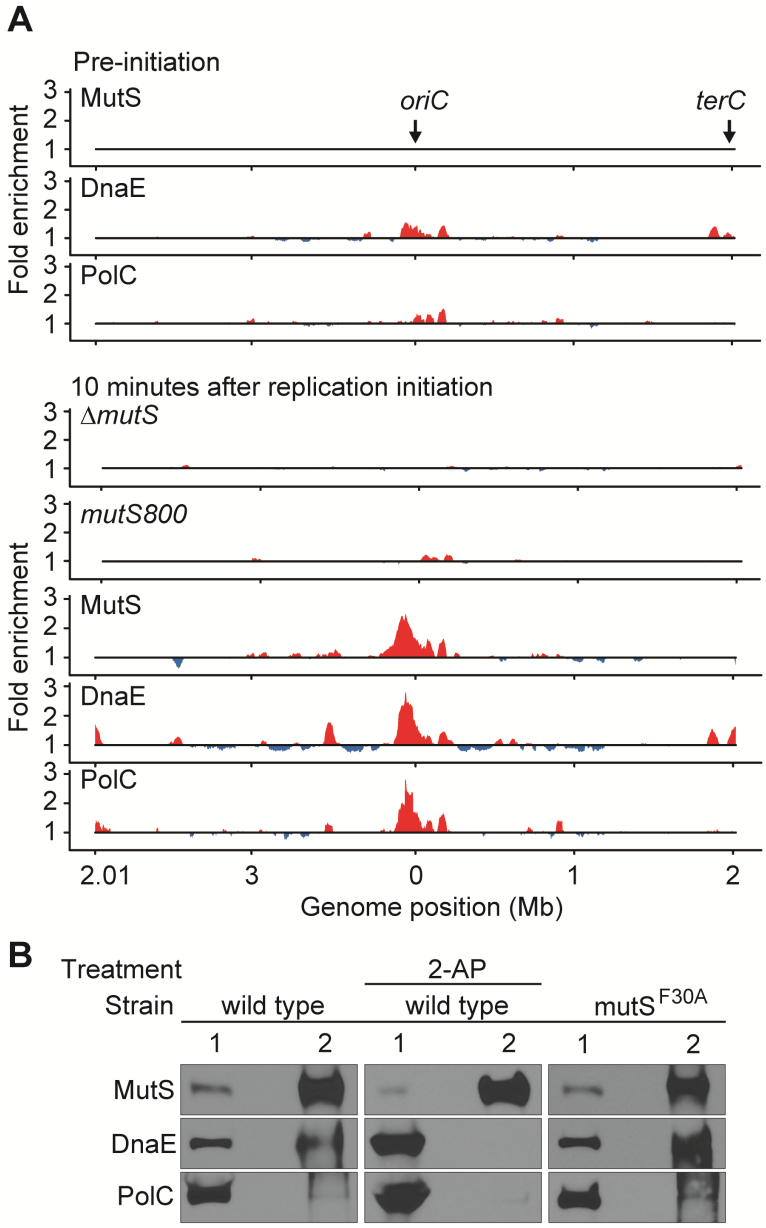


Figure 2.10: MutS recruitment to DNA upon replication initiation and interaction with DNA polymerase subunits. (A) Analysis of pooled ChIP-seq data from two independent experiments showing the enrichment levels of MutS and the polymerases DnaE and PolC along the chromosome prior to (upper) and 10 minutes after (lower) DNA replication initiation. (B) Co-IP of DnaE and PolC with MutS using affinity-purified antiserum directed against MutS. Lane 1: 5% input. Lane 2: anti-MutS immunoprecipitation.

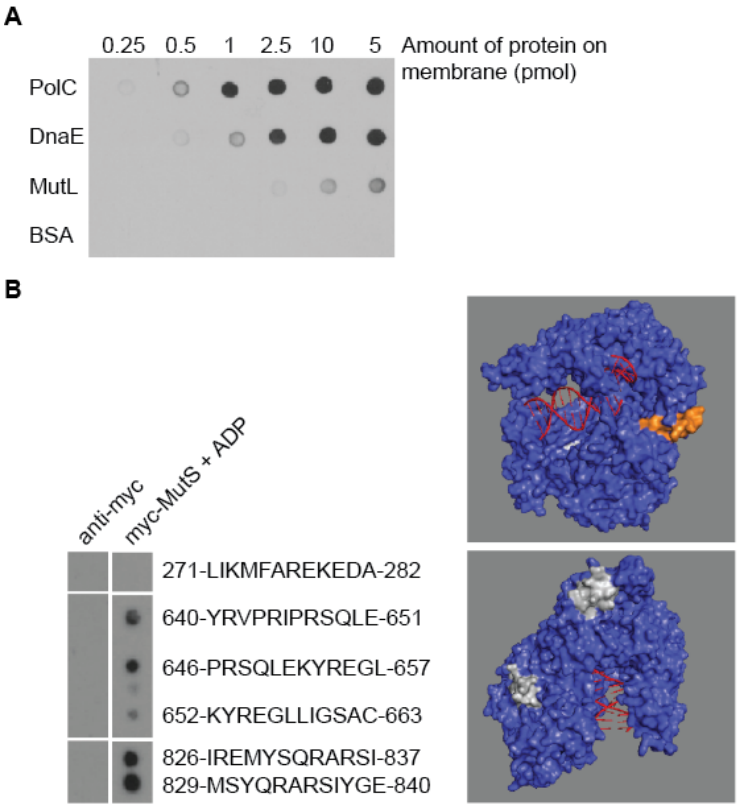


Figure 2.11: Interactions between MutS and PolC, DnaE and MutL *in vitro* (A) Far-western blot with the indicated proteins bound to the membrane. BSA serves as a negative control. MutS bound to proteins on the membrane was detected using affinity-purified antisera. (B) Peptide array analysis of purified myc-MutS retained on a PolC peptide array. The peptides retaining myc-MutS are mapped to a model of *B. subtilis* PolC based on the structure of *Geobacillus kaustophilus* PolC (PMID: 3F2B) (*Evans et al.*, 2008) using pymol. Regions of PolC corresponding to the peptides that retained myc-MutS are shaded gray (amino acids 640-663 and 826-840) and orange (amino acids 271-282). Myc-MutS retained by peptides was detected with monoclonal anti-myc antibodies.

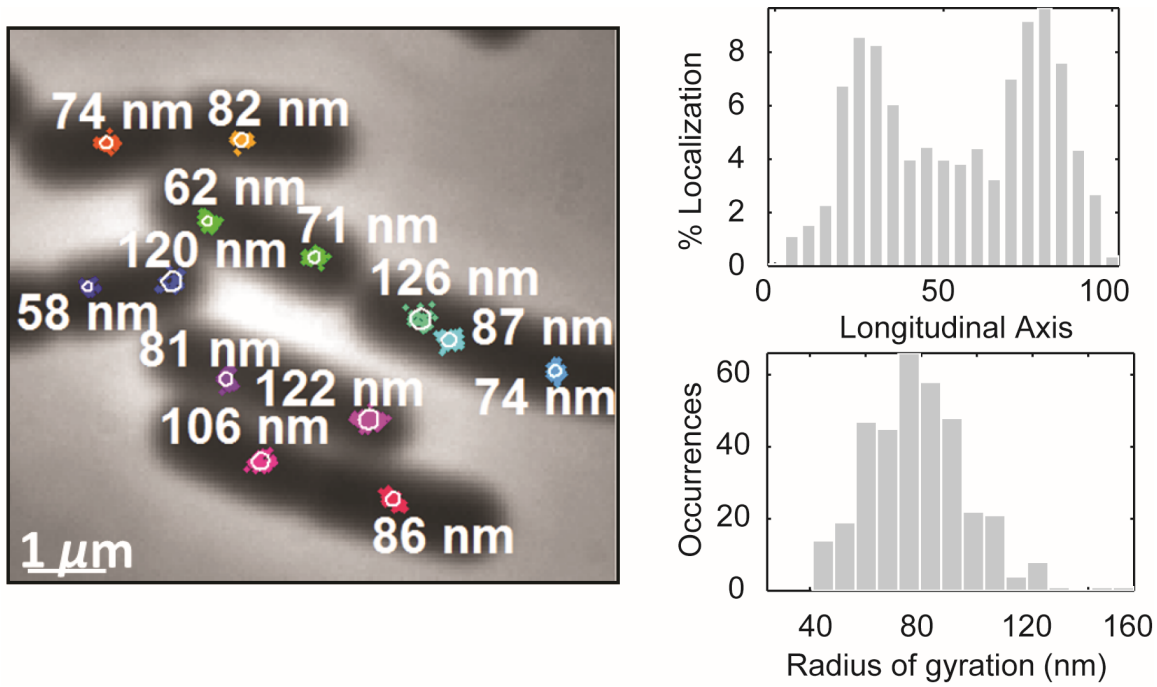


Figure 2.12: Localization of the processivity clamp DnaN-mCitrine in over 100 *B. subtilis* cells during exponential phase. To determine that DnaX-mCitrine localizes in the same way as other replisome subunits, we imaged the processivity clamp protein DnaN-mCitrine in live *B. subtilis* cells. Similar to DnaX-mCitrine, DnaN-mCitrine forms compact clusters with an average radius of motion $\approx 79 \pm 18$ (s.d.) nm, and these clusters are most frequently found at quarter positions or, to a lesser degree, cell centers. Because the DnaN-mCitrine fusion partially compromises the MMR activity *in vivo*, we used DnaX-mCitrine as a proxy to the replisome throughout the study.

Table 2.1: Strains used in this study.

Strain	Relevant genotype	Reference
PY79	Wild type prototroph, SPβ°	*Youngman et al.* 1984
JWS108	Δ*mutSL*	*Yao et al.* 2013
JWS121	*mutS-PAmCherry-RBS-mutL*	This study
JWS134	*amyEPxyl:dnaX-mCitrine*	This study
JWS154	*amyE::Pxyl:dnaN-mCitrine*	This study
JWS161	*mutS[F30A]-PAmCherry1-RBS-mutL*	This study
JWS162	*dnaB134(ts) zhb83::Tn917 (tet)* x PY79	*Burnett and Wake* 1977
JWS170	Δ*mutS*, *dnaB134(ts) zhb83::Tn917 (tet)*	This study
JWS176	*mutS[F30A]-PAmCherry1-RBS-mutL*, *amyE::Pxyl:dnaX-mCitrine*	This study
JWS185	*mutS[K608M]-PAmCherry-RBS-mutL*	This study
JWS194	*mutS[K608M]-PAmCherry-RBS-mutL*, *amyEPxyl:dnaX-mCitrine*	This study
JWS220	*mutS-PAmCherry-RBS-mutL*, *dnaN5*, *spoIIIJkan*	This study
JWS221	*mutS800-PAmCherry-RBS-mutL*	This study
JWS222	*mutS-PAmCherry-RBS*, Δ*mutL*	This study
JWS225	*mutS-PAmCherry-RBS*, Δ*mutL*, *amyEPxyl:dnaX-mCitrine*	This study
JWS243	*mutS800-PAmCherry*, *amyE::Pxyl:dnaX-mCitrine*	This study
JWS258	*mutS::mutS800*, *amyE::mutL*, *dnaB134(ts) zhb83::Tn917 (tet)*	This study
JWS297	*dnaN5*, *spoIIIJ::kan*; *amyE::Pxyl:dnaX-mCitrine*; *mutS-PAmCherry-RBS-mutL*	This study

Table 2.2: Oligonucleotides used in this study.

oJS740	GCGACGCTAGCG**T**GCGGCTCGTCC	Used in all ATPase assays with MutS. Mismatched or bulged base in bold	*Obmolova et al.* 2000
oJS741	GGACGAGCCGCCGCTAGCGTCGC	T-bulge substrate when annealed to oJS740	*Obmolova et al.* 2000
oJS742	GGACGAGCCGCGCGCTAGCGTCGC	GT mismatch substrate when annealed to oJS740	*Obmolova et al.* 2000
oJS743	GGACGAGCCGCACGCTAGCGTCGC	Homoduplex substrate when annealed to oJS740	This study

CHAPTER III

Determination of Genome-wide Factors Affecting DNA Polymerase Error Rate

3.1 Introduction

All living organisms must replicate their genetic information accurately to ensure its faithful transmission to new generations. Replicative polymerase errors provide an important source of genetic variation that can drive evolution if a mutation provides a fitness advantage. Understanding the origins of genetic variation will inform our understanding of evolution and the development of genetic diseases. A number of factors have been proposed to affect mutation occurrence *in vivo* (*Fijalkowska et al.*, 1998; *Lind and Andersson*, 2008; *Ma et al.*, 2012; *Lee et al.*, 2012; *Schaibley et al.*, 2013; *Paul et al.*, 2013; *Foster et al.*, 2013; *Lujan et al.*, 2014; *Sung et al.*, 2015; *Million-Weaver et al.*, 2015). Many studies have relied on selectable mutation reporters, which bias analyses to very specific mutations in static sequence contexts, and are therefore not ideally suited to understanding mutation occurrence genome-wide. Of the studies

A modified version of this chapter was in submission to *Nature* during the writing of this thesis document. A table containing information on all 7247 variants detected in this study is available upon request. I wish to thank Heather Schroeder for her help and advice in statistics. We thank Dr. Lindsay Matthews and Peter Burby for comments and feedback on the manuscript. Gabriella Szewczyk performed rifampin resistance assays and western blots. William Hirst and I performed mutation accumulation line propagation and DNA purifications. I analyzed the data, and Lyle Simmons and I discussed the results and wrote the paper.

using mutation accumulation lines to observe mutation occurrence genome-wide, none have performed a detailed analysis of the effect of gene expression and orientation relative to replication on mutation occurrence. Here, we used mutation accumulation lines, whole-genome sequencing and whole-transcriptome analysis in *Bacillus subtilis* to observe the locations and rate at which mutations arise in bacteria with and without MMR with as little selection bias as possible (*Barrick and Lenski*, 2013). Our analysis of replication errors generated in over 180 sequenced lines that underwent a total of more than 370,000 generations has provided new insights into how DNA polymerase errors sculpt genetic variation and drive evolution. Strikingly and in contrast to recently proposed models (*Paul et al.*, 2013; *Million-Weaver et al.*, 2015), the direction and transcript level of a coding sequence has no significant effect on its mutation rate, nor its ratio of nonsynonymous to synonymous base pair substitutions. Homonucleotide runs, which cause insertion and deletion mutations, are enriched outside of coding regions, a pattern that has likely been selected for so that insertions and deletions, which result in frameshifts, rarely occur within coding regions.

DNA is the storage medium for genetic information throughout biology. Although it is known that DNA polymerase errors provide genetic variation and drive evolution, many prior studies have imposed a selection or stress prior to analysis of mutational reporters (*Fijalkowska et al.*, 1998; *Paul et al.*, 2013; *Million-Weaver et al.*, 2015), which biases observation to a small number of possible mutations in a single sequence context. In order to determine factors affecting mutation occurrence *in vivo*, we observed mutations that occur throughout the genome of the model Gram-positive bacterium *Bacillus subtilis* with as little selection bias as possible by using mutation accumulation (MA) lines (*Barrick and Lenski*, 2013). We inactivated DNA mismatch repair (MMR) to determine where and in what context base pair substitutions (BPSs) and insertions and deletions (indels) are produced by DNA polymerase. Integrating our MA line analysis with RNA-seq data placed us in the unique position of being able to

determine whether gene expression and context affect DNA polymerase accuracy.

3.2 Results

The result of our MA line experiments are briefly summarized in Table 3.1. In general, the distribution of the number of mutations observed in the MA lines follows the expected Poisson distribution (Figure 3.1). Loss of *mutSL* gave rise to a 60-fold increase in overall mutation rate (Table 3.1), which is similar to what we observed using rifampin resistance as a reporter (Figure 3.2 and Figure 3.3). In the absence of *mutSL* an average of 15.5 generations pass before a single mutation occurs, and in wild type *B. subtilis* DNA replication is astoundingly accurate, with an average of 909 generations between mutations (Table 3.1). This is in close agreement with *E. coli*, which undergoes $\approx$ 1000 generations between mutations (*Lee et al.*, 2012). Additionally, comparison of the mutation rates and spectra of Δ*mutSL* and Δ*walJ* revealed WalJ, a 5′ $\rightarrow$3′ exonuclease in *Bacillus anthracis* (*Yang et al.*, 2013), to be necessary for 74% of MMR (Table 3.1 and Figure 3.2). Because deletion of *walJ* and *mutSL* and elimination of *mutL* endonuclease activity all cause defects in MMR, for the remainder of the study, we binned data from the Δ*walJ*, Δ*mutSL* and *mutL[E468K]* MA lines such that we have 6952 mutations in the combined MMR- MA lines and 295 mutations in the MMR+ wild type MA lines. The high number of observed mutations in MMR- MA lines affords us high statistical power to understand where mutations occur in the genome.

Indels and BPSs were approximately evenly distributed throughout the genome (Figure 3.4), but when transitions were grouped into their four possible categories we found that regardless of whether MMR was intact, complementary transitions, i.e. C $\rightarrow$ T and G $\rightarrow$ A, or T $\rightarrow$ C and A $\rightarrow$ G, accumulated symmetrically between the two replichores (Figure 3.5b and c). This distribution cannot be explained by bias in the distribution of nucleotides between the replichores, as our calculation of mutation rate in Figure 3.5c was conditioned on the number of each nucleotide found in each

replichore. A very similar complementary symmetry was observed in undomesticated *B. subtilis*, *Mesoplasma florum* and *E. coli* MA lines (*Lee et al.*, 2012; *Sung et al.*, 2015), and in *S. cerevisiae* MA line experiments it was found that mutations accumulate with complementary symmetry between origins of replication (*Lujan et al.*, 2014). We attribute the complementary symmetry of transition accumulation between the replichores to differences in fidelity of leading and lagging strand replication.

Sequence context has been shown to affect DNA polymerase error rate *in vitro* and mutation rate *in vivo* in a variety of experimental systems (*Kunkel et al.*, 1981; *Petruska and Goodman*, 1985; *Sinha*, 1987; *Zhu et al.*, 2014; *Lujan et al.*, 2014; *Sung et al.*, 2015). We therefore tested whether a similar phenomenon existed in our MA line data. In our analysis we present sequence context in terms of the leading strand for each replichore. The effect of neighboring nucleotides on transition rate in MMR- MA lines is in very close agreement with prior work (*Sung et al.*, 2015) (Figure 3.6). Transitions in our wild type lines show a similar effect, although we were unable to analyze the two replichores independently due to loss of resolution from having only 198 transitions in the wild type data (Figure 3.7). We analyzed context dependence for transversions as well, but as they are rare (2.7% of all BPSs in this study) in comparison to transitions, we were unable determine which contexts, if any, promote transversion occurrence (Figure 3.8). We conclude that transition rate is highly influenced by neighboring nucleotide context, and, because the effect of nucleotide context on transition occurrence is similar between MMR- and MMR+ data, the neighboring nucleotide context primarily affects polymerase accuracy rather than MMR efficiency.

Indels are enriched immediately outside of coding sequences (CDSs) (Figure 3.9a). Given that long homopolymer runs are also enriched intergenically (Figure 3.10), and indel mutation rate in long homopolymer runs is very high (*Lee et al.*, 2012) (Figure 3.11), we tested whether intergenic enrichment of indels was caused by concomitant enrichment of homopolymer runs. Most indels observed immediately

outside CDSs were in homopolymer runs ranging in length from six to eight (Figure 3.9b). Correcting for the bias in homopolymer run distribution (see Supplementary Text), we found that enrichment of indels near CDSs can be entirely explained by coenrichment of homopolymer runs (Figure 3.9). We propose that CDSs containing long homopolymer runs have been selected against due to the high propensity of long homopolymers to obtain indels, which will cause frameshifts and loss of CDS function.

Transcription of genes encoded on the lagging strand causes RNA polymerase to collide head-on with DNA polymerase, and it has recently been suggested that this leads to an increased mutation rate and nonsynonymous to synonymous BPS ratio (dN/dS) in head-on oriented genes (*Paul et al.*, 2013; *Million-Weaver et al.*, 2015). We therefore tested whether expression and orientation of CDSs relative to replication had an effect on CDS mutation rate by performing RNA-seq and integrating transcript abundance data with our MA line data for BPSs. We performed multiple linear regression using the number of BPSs in each CDS as the dependent variable against CDS orientation, length, steady-state transcript abundance (RPKM), the interaction between orientation and length, and the interaction between orientation and RPKM (see Supplementary Equation 3.5 and accompanying description in Supplementary Methods for details). We found that orientation of a CDS was a significant predictor of the number of BPSs it accumulated only if outliers were included in the regression (Figure 3.12). Performing a separate regression using only those genes within three standard deviations of the mean length or RPKM (11 head-on and 90 codirectional of the 4163 CDSs were determined to be outliers) yielded gene length as the only significant predictor of the number of BPSs in a CDS (Figure 3.12 and Supplementary Table 3.4). Furthermore, we found that dN/dS did not differ significantly between codirectional and head-on CDSs, was not significantly different than the expected value derived by Monte Carlo simulation (Figure 3.13). Therefore, in the absence of

strong selection for specific BPSs, transcript abundance and CDS orientation have no discernible influence on BPS rate and effect.

3.3 Discussion

As the drivers of evolution and genetic disease, it is difficult to overstate the importance of understanding the processes by which mutations are generated. Mutations have classically been viewed as occurring somewhat randomly, but recent models propose that genes encoded on the lagging strand evolve more quickly due to a higher intrinsic mutation rate of those genes caused by replication/transcription conflicts (*Paul et al.*, 2013; *Million-Weaver et al.*, 2015). These studies have relied heavily on reporters requiring specific gain of function mutations in a static sequence context. Our work highlights the paramount importance of either using reporters that are sensitive to a wide variety of mutations in a multitude of sequence contexts or observing mutation accumulation genome-wide with as little selection bias as possible in order to make broad conclusions about factors affecting mutagenesis *in vivo*. In doing so, we have shown that local sequence context is the main determinant of mutation occurrence driving genome variation in *Bacillus subtilis*.

3.4 Materials and Methods

3.4.1 Bacteriological methods and strain construction

Strains were constructed using pMiniMAD2 following the protocol as described (*Patrick and Kearns*, 2008). For Δ*walJ* and Δ*mutSL*, only the open reading frame of *walJ* was removed from the genome and *mutS*, *mutL* and 15 nucleotides of intergenic space were removed from the genome, respectively. For the *mutL[E468K]* background, a single nucleotide substitution was introduced, changing the guanosine at nucleotide position 1402 in the *mutL* gene to adenosine. The Δ*walJ* Δ*mutSL* double mutant was constructed by cleanly deleting the *walJ* open reading frame from the Δ*mutSL* strain background using pMiniMAD2. Construction of *walJ[H60A]* was performed by cleanly substituting nucleotides 178 and 179 (CA) with GC. The appropriate strain was verified by PCR and Sanger sequencing of the affected locus.

3.4.2 Estimation of mutation rates via rifampin resistance and fluctuation tests

Rifampin resistance assays and calculation of mutation rate were performed as described (*Walsh et al.*, 2014).

3.4.3 Mutation accumulation line protocol

A founder colony for each strain was selected after two rounds of colony purification from a frozen stock. This founder colony was once more colony purified to achieve hundreds of single colonies while also being used to inoculate 3 mL LB to save a frozen stock and for genomic DNA purification. The next day, from the plate with hundreds of single colonies, each of which came from the single founder colony, we randomly chose between 40-85 colonies to initiate our MA lines. Colony purification was performed by restreaking on LB agar plates every 24 hours, incubating plates at 30° C. Thoughout the MA line experiment, the same six lines shared a plate. In order to reduce

bias originating from colony choice for colony purification, we selected the single, well-isolated colony that was closest to the boundary clockwise on the plate to the next MA line.

Upon achieving the desired number of colony purifications, a colony from each MA line was used to inoculate 3 mL of liquid LB at 37°C. Cells were harvested from 1 mL of culture by centrifugation during mid-exponential phase of growth ($OD_{600} \approx 0.6$) and genomic DNA was purified using the MasterPure Gram Positive DNA Purification Kit (Epicentre catalog number MGP04100) according to the manufacturer's instructions. 1 mL of culture was also used to save a frozen stock of each line.

3.4.4 Determination of generations per colony

Throughout the course of the MA line experiments, we frequently harvested a single colony from each strain background, resuspended it in saline and plated serial dilutions to count the cfu in the colony. A log_2 transformation of the cfu gave the number of generations (times the cells in the colony doubled), and taking the mean of the number of generations for each strain background, we determined the number of generations per 24 hour growth period on LB agar at 30° C was 25.42 and was unaffected by our genetic manipulations.

3.4.5 RNA-seq

Wild type *B. subtilis* strain PY79 was grown to mid-exponential phase at 30°C in LB. Following addition of one volume of ice cold methanol cells were harvested by centrifugation. RNA was purified and rRNA was depleted from total RNA using the RiboPure RNA Purification Kit (Life Technologies) and the Ribo-Zero Magnetic Kit (Bacteria), respectively, according to the manufacturers' recommendations. cDNA, sequencing library preparation, and sequencing were performed by the University of Michigan Sequencing Core. Fifty-base single end reads were sequenced on the Illumina

HiSeq 2000 platform. Sequence alignment to the *B. subtilis* PY79 reference genome (Accession number CP006881.1 (*Schroeder and Simmons*, 2013)) was performed using bwa, version 0.7.8-r455. Subsequent analysis of transcript abundance was performed using the statistical software R, using the package limma (*Law et al.*, 2014). RNA-seq was performed with three biological replicates.

3.4.6 Genome Sequencing

Library preparation and 100-base paired end sequencing on the Illumina Hi-Seq 2000 or 2500 platforms were performed by the University of Michigan DNA Sequencing Core.

3.4.7 Alignment and variant calling

100-base paired end reads were aligned to the *B. subtilis* PY79 reference genome (Accession number CP006881.1 (*Schroeder and Simmons*, 2013)). Aligments were performed using the software package bwa (v0.5.9, revision 16) with the default parameters, except when running the bwa samse program, when we set the "-n" parameter to one to allow each read to align only once to the reference (*Li and Durbin*, 2009). Samtools was used to generate and sort alignment files (*Li et al.*, 2009). Prior to variant calling, duplicate reads were marked using Picard tools (http://broadinstitute.github.io/picard). Variant calling was performed using the software package freebayes with standard filters and setting ploidy to one (*Garrison and Marth*, 2012).

Variants called in each founder colony were subtracted from those in MA lines originating from that founder colony with the software package vcftools, using the command vcf-isec with the "-f" option (*Danecek et al.*, 2011).

Data analysis subsequent to variant calling was performed using a combination of the programming languages Python and R. A variety of R packages were used,

including VariantAnnotation, GenomicRanges and ggplot2 (*Obenchain et al.*, 2014; *Lawrence et al.*, 2013; *Wickham*, 2009).

3.4.8 Data filtering

For lines of each genetic background, any variants within 50 nucleotides of the genomic deletion were removed from further analysis. For example, in lines of the Δ*mutSL* background, variants within 50 nucleotides 5′ to *mutS* and 50 nucleotides 3′ to *mutL* were discarded. If any two MA lines that shared a plate had one or more identical variants called, all data belonging to one of those MA lines was removed at random from further analysis. In all, 75 wild type, 38 Δ*walJ*, 44 *mutL[E468K]*, and 36 Δ*mutSL* MA lines were included in the final analysis. Additionally, we binned the alignments belonging to our wild type founder colony (JWS208) into 100 base pair wide bins, offset by 50 and identified the bins in which the median mapping quality was less than 20 (Supplementary Table 3.3). Any variant found within these low mapping quality regions was removed from further analysis, and in all calculations of mutation rate per generation per nucleotide we define the number of nucleotides in the genome as the number with mapping quality $\geq$ 20 (Supplementary Table 3.3).

3.4.9 Calculation of conditional mutation rate

Calculation of conditional mutation rate in homopolymer repeats was performed using the following equation, as described (*Lee et al.*, 2012):

$$\frac{M_l}{R_l \times GN} \tag{3.1}$$

where M_l is the number of mutations observed in homopolymer repeats of length l, R_l is the total number of nucleotides in homopolymer repeats of length l in the reference genome and GN is the total number of generations the lines went through

during the MA line colony purifications (272700 for wild type, 105020 for MMR-).

Conditional mutation rate in triplet sequence contexts was calculated in a similar manner, as described (*Sung et al.*, 2015):

$$\frac{M_t}{R_t \times GN} \tag{3.2}$$

where M_t is the number of transitions at the focal (middle) nucleotide position observed in a given triplet context, R_t is the number of times the given triplet occurs in the genome, and GN is the total number of generations the lines went through during the colony purifications.

Mutation rate for each transition type genome-wide was calculated conditionally based on the number of nucleotides with each base in the reference sequence of each replichore using the following equation:

$$\frac{M_{b,r}}{R_{b,r} \times GN} \tag{3.3}$$

where $M_{b,r}$ is the number of transitions in nucleotides with base b observed in replichore r, $R_{b,r}$ is the number of nucleotides with base b occur in the replichore r, and GN is the total number of generations the lines went through during the colony purifications.

3.4.10 Determining homopolymer-run-length-corrected indel counts

Indels within one kb of CDS start or end sites were counted in 50 nucleotide wide bins offset by 25. Therefore, any single indel will be counted at least twice in this analysis, as adjacent bins overlap by half, and the indel may exist within one kb of multiple CDS start and end sites. We also counted homopolymer repeats of each length, from 1 to 10, near CDS start/end sites in 50 nucleotide wide bins offset by 25. Conditional mutation rates for indels in homopolymer runs of each length across all

bins were calculated for the new dataset as described above, as was the mutation rate for indels across all homopolymer runs and bins.

The expected number of indels (U_i) in bin i after correcting for homopolymer run length bias was calculated using the following equation, as described (*Long et al.*, 2015), with minor modifications to adapt the equation for use with homopolymer runs rather than GC content:

$$U_i = \sum_{l=1}^{10} \left(\frac{u_{i,l}}{\overline{\mu_l}} \right) \times \mu_{indel} \tag{3.4}$$

where $u_{i,l}$ is the indel count in homopolymers of length l in bin i, $\overline{\mu_l}$ is the conditional mutation rate for indels in homopolymers of length l across all bins, and μ_{indel} refers to the overall mutation rate for indels in all homopolymer run lengths across all bins. Therefore, in the above equation, the factor $\sum_{l=1}^{10} \left(\frac{u_{i,l}}{\overline{\mu_l}} \right)$ represents the expected number of nucleotides in bin i and by multiplying this factor by the overall mutation rate for indels, μ_{indel}, we arrive at the expected number of indels in bin i if nucleotides in all homopolymer run lengths had the same indel mutation rate.

3.4.11 Multiple linear regression

Expression and direction of a gene have been proposed to affect mutation rate (*Paul et al.*, 2013; *Million-Weaver et al.*, 2015). To test this, we performed multiple linear regression to determine the extent to which CDS length (in kb), expression (in RPKM), orientation, interaction between length and orientation, and interaction between expression and orientation are associated with the number of BPSs in CDSs. We counted the number of BPSs in all 4163 CDSs, and our regression model follows:

$$\begin{aligned} BPS\_count = & \beta_0 + \beta_1 Head\_on + \beta_2 Length + \beta_3 RPKM \\ & + \beta_4 Head\_on^* Length + \beta_5 Head\_on^* RPKM \end{aligned} \tag{3.5}$$

The estimated βs refer to the associated variables' effects on average BPS count in CDSs. For instance, if $\beta_2 = 2.0$, this would denote that if all other variables are held constant, for every one kb increase in CDS length, the average BPS count would increase by 2. Head_on*Length and Head_on*RPKM are the interaction terms for CDS orientation and length, and orientation and expression, respectively. If the interaction term β_5 were significanly associated with mean BPS_count, this would indicate that the effect of length differs depending on the orientation of the CDS relative to DNA replication.

3.4.12 Monte Carlo simulation of expected nonsynonymous to synonymous BPS ratio

The nonsynonymous to synonymous BPS ratio (dN/dS) to be expected by random chance was determined using Monte Carlo simulation in the following manner: We grouped each position in the high-mapping-quality reference genome (Supplementary Table 3.3) according to its leading strand neighboring 5′ and 3′ sequence context. For each round of simulation we then selected randomly without replacement from each group the sites of the reference to serve as transitions according to the probability of each site in that context resulting in transitions in our observed data. We simulated transversions in the same manner. For contexts in which no transitions or transversions were observed, the probability of mutation in that context was set to ten-fold lower than the lowest observed probability. The variant bases for transversions were selected at random from the two possibilities, i.e., if the reference base was a purine, either pyrimidine had an equal probability of becoming the variant base in our simulation. This simulation was carried out 10000 times, each time predicting the resulting dN/dS for all CDSs, CDSs transcribed head-on with replication, and CDSs transcribed codirectionally with replication.

3.4.13 Statistical analysis

Statistical analysis, data manipulation, multiple linear regression and plotting were performed using the statistical computing software R. Loess curve fitting was performed with the span argument set to 0.15.

3.5 Supplementary Results and Discussion

3.5.1 WalJ is involved in DNA mismatch repair.

WalJ in *B. anthracis* is a 5′ →3′ exonuclease suggested to function in MMR (*Yang et al.*, 2011b, 2013). To test whether *walJ* is involved in MMR in *B. subtilis*, we performed epistasis analysis to determine the mutation rates of Δ*mutSL*, Δ*walJ* and a Δ*mutSL* Δ*walJ* double mutant strain via rifampin resistance (Rif^R) (Figure 3.2a). We found that the Rif^R mutation rate in Δ*walJ*, Δ*mutSL* and Δ*mutSL* Δ*walJ* double mutants were all elevated compared to wild type, and their mutation rates were all within error, indicating an epistatic interaction between *walJ* and *mutSL* (Figure 3.2a). WalJ is a metallo-beta-lactamase (*Yang et al.*, 2013), with a conserved HXHXDH motif. We substituted the central histidine of the metal-ion-coordinating HXHXDH motif for an alanine to construct *walJ[H60A]*. WalJ[H60A] yielded a mutation rate within error of Δ*walJ*, Δ*mutSL* and Δ*mutSL* Δ*walJ*. Western blotting showed that MutSL levels were unchanged in Δ*walJ* (Figure 3.2b). Using MA lines, which are less biased in the mutations that can be observed (*Barrick and Lenski*, 2013), we found that deletion of *walJ* conferred a mutation rate corresponding to 74% of Δ*mutSL* (Figure 3.2c). Mutation spectra can be diagnostic for a defect in a specific DNA repair pathway, and comparing the genome-wide mutation spectrum in Δ*walJ* MA lines to that of wild type and Δ*mutSL* MA lines, we observed that Δ*mutSL* MA lines harbor a nearly identical mutation spectrum to Δ*walJ* (Figure 3.2d). We therefore conclude that WalJ is an exonuclease necessary for 74% of MMR in *B. subtilis*.

3.5.2 Mismatch repair is most efficient at preventing AT → GC transitions.

Of the possible BPSs, MMR is best equipped to prevent AT→GC transitions, as Δ*mutSL* resulted in an 85-fold increase in the rate of AT→GC transitions and a 67-fold increase in the rate of GC→AT transition accumulation in Δ*mutSL* (Figure 3.3). In agreement with prior published work, we find that MMR has evolved to prevent transitions, which are the most common mismatches produced by replicative DNA polymerases, and it has relatively little effect on transversion rate, the former mutation rate increasing 73-fold in Δ*mutSL*, and the latter increasing 5.5-fold in Δ*mutSL* (Figure 3.3). However, loss of *mutSL* did increase the mutation rate of GC→CG transversions appreciably.

3.5.3 CDS orientation does not affect dN/dS ratio

In order to test whether dN/dS ratio was affected by gene orientation relative to replication, we calculated the observed dN/dS ratio for all CDSs, those in the head-on orientation, and codirectional CDSs. Within MMR- and MMR+ MA lines, we found that the dN/dS ratios were within error of each other (Figure 3.13a). The dN/dS ratios were lower overall for MMR- than for MMR+, but this agrees well with expectations, given the increase in transition abundance in MMR- MA lines and the fact that transitions are less likely to be nonsynonymous. Indeed, comparison of the observed dN/dS values with the distribution of expected values obtained by Monte Carlo simulation showed the observed values either to agree well with expectation, or to be slightly higher than expected (Figure 3.13). A slight overabundance of nonsynonymous substitutions compared to the expected distribution indicates that selection was effectively minimized by our MA line procedure.

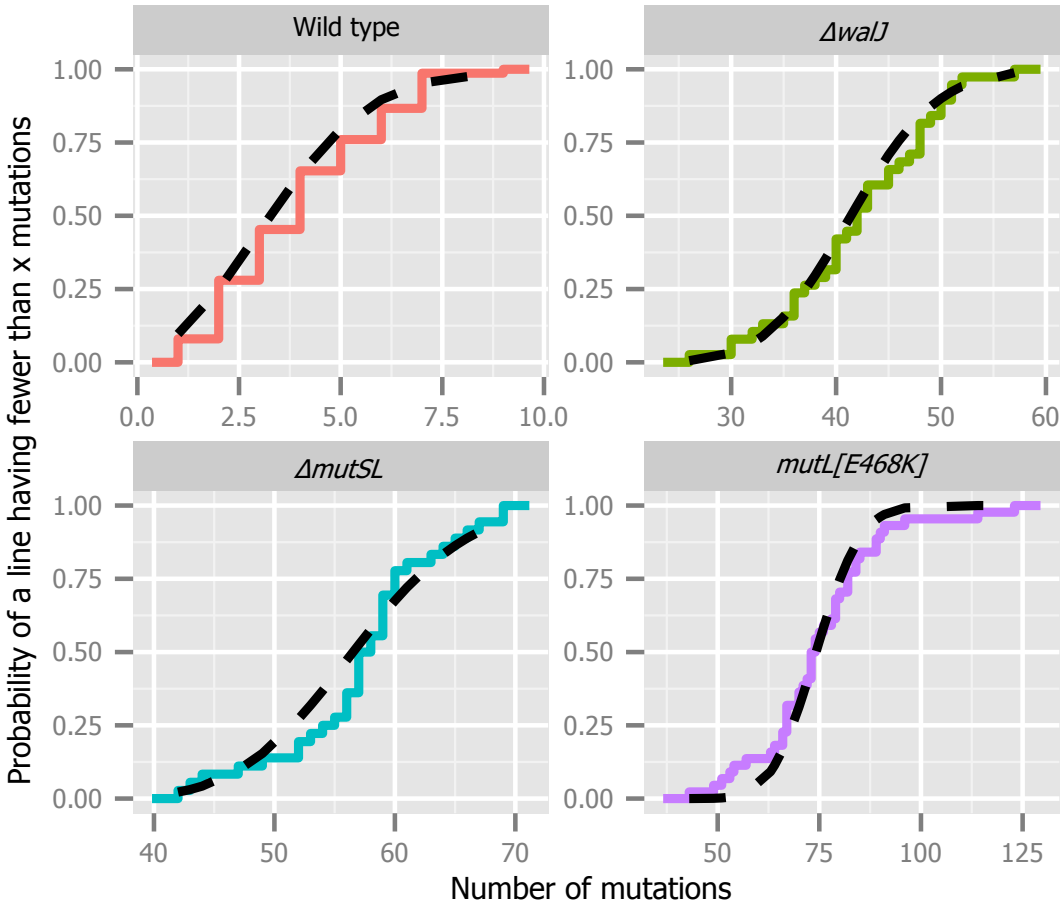


Figure 3.1: Cumulative distributions of mutation counts. The empirical cumulative distribution function for the number of mutations observed in MA lines of the indicated genotype is plotted. The black dashed line is the Poisson distribution fit to the data.

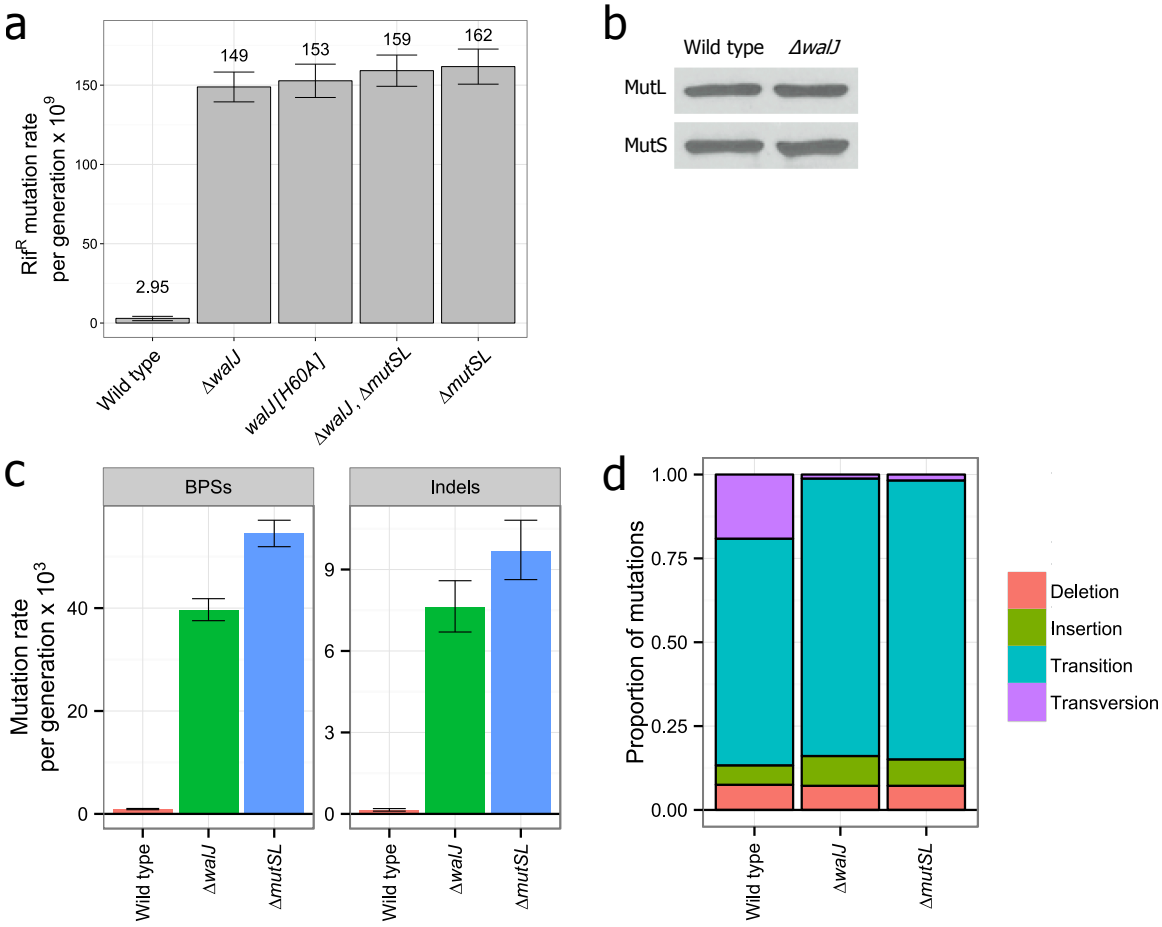


Figure 3.2: WalJ is involved in DNA mismatch repair. **a**, A barplot of the mutation rates as determined by rifampin resistance for the indicated strains shows epistasis between *walJ* and *mutSL*. The y-axis tick labels and mutation rates above the bars represent the mutation rate resulting in rifampin resistance per generation multiplied by 10^9. WalJ[H60A] is a variant of WalJ that is catalytically inactive (see Supplementary Results and Discussion). **b**, Western blot of MutL and MutS proteins in wild type and Δ*walJ* strains indicates that MutSL protein levels are unchanged upon deletion of *walJ*. **c**, Mutation rates derived by mutation accumulation lines are presented as barplots for BPSs and indels. **d**, Mutation spectra for wild type, Δ*walJ*, and Δ*mutSL* strains are shown. Δ*walJ* and Δ*mutSL* result in nearly identical proportions of each type of mutation.

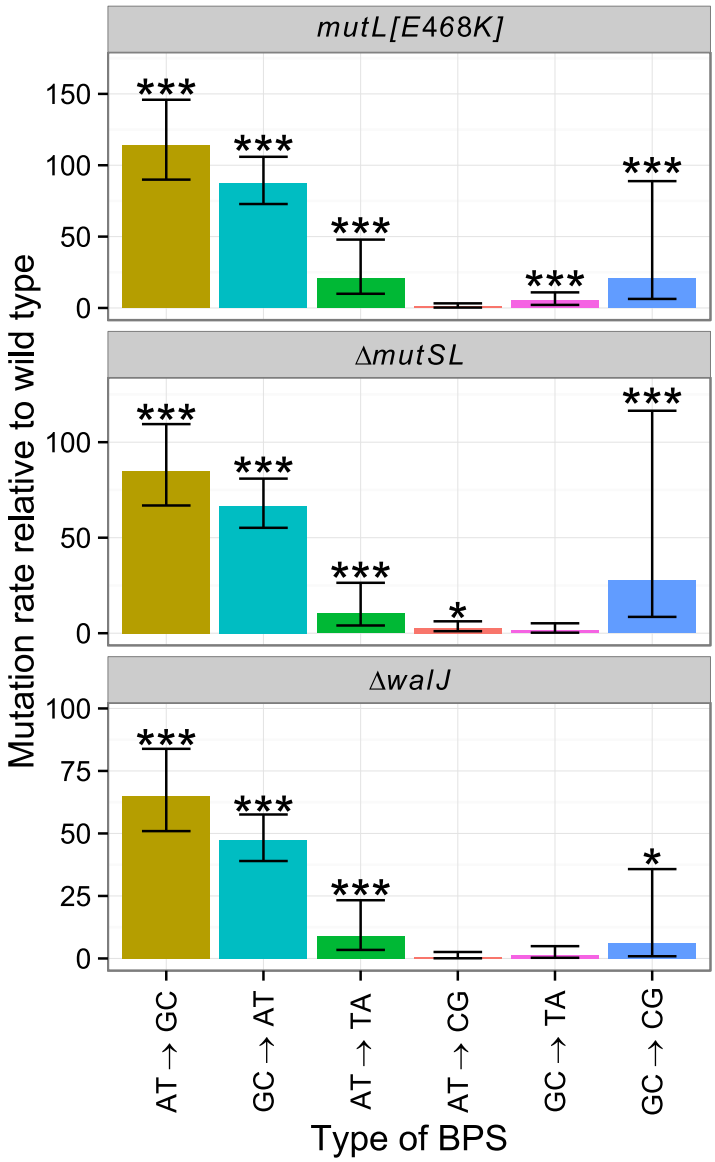


Figure 3.3: MMR efficiently repairs transitions. Bar plots of fold change in mutation rates for the indicated BPSs in the *mutL[E468K]*, Δ*mutSL*, or Δ*walJ* MA lines relative to wild type. Error bars represent exact Poisson 95% confidence intervals for the relative rate. *** denotes $p \leq 0.001$, ** denotes $p \leq 0.01$, and * denotes $p \leq 0.05$.

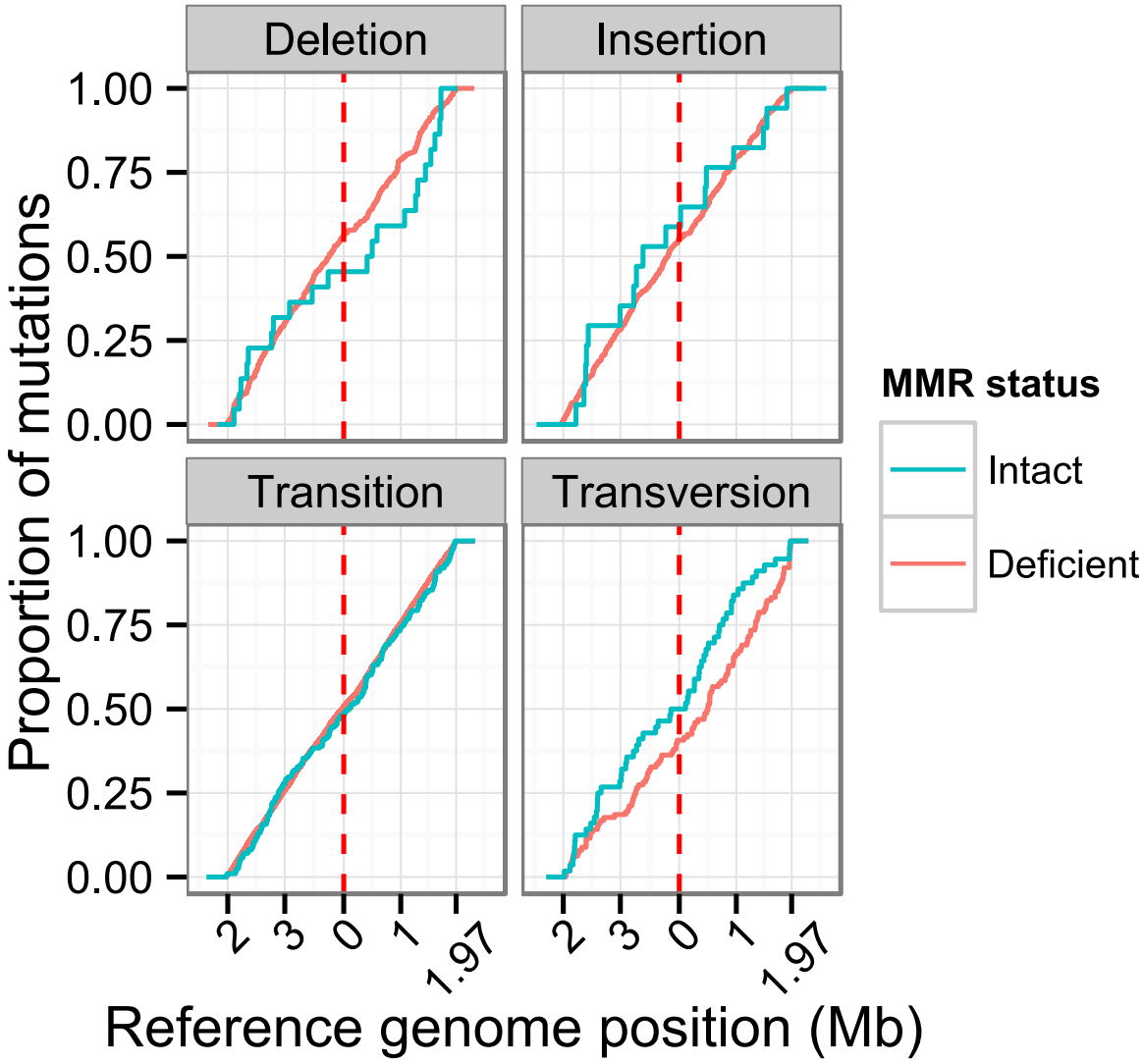


Figure 3.4: Cumulative distributions of mutations by type. Cumulative distributions of the indicated types of mutations along the genome. The origin of replication is indicated by the vertical dashed red line and the terminus is at both ends. MMR intact refers to wild type MA line data and MMR deficient refers to the pooled data for Δ*mutSL*, Δ*walJ*, and *mutL[E468K]*.

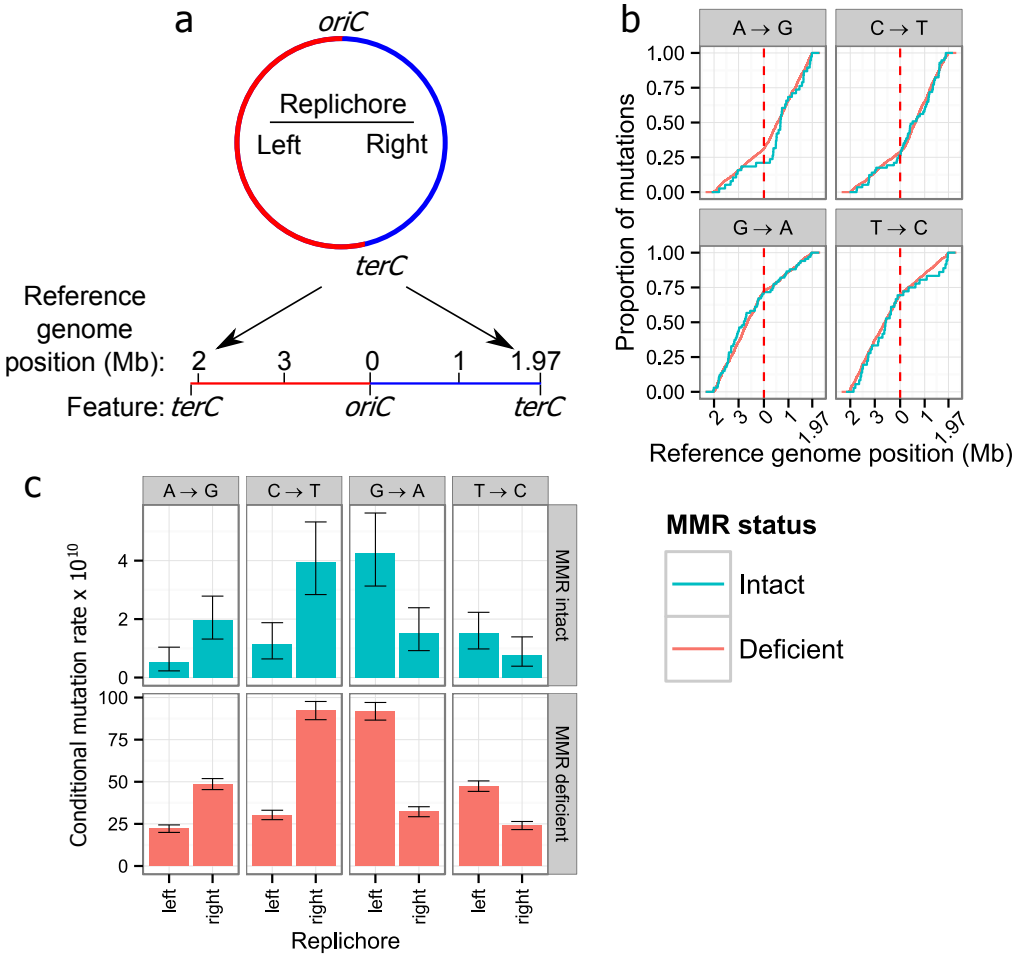


Figure 3.5: Transitions display complementary symmetry between the *B. subtilis* replichores. **a**, Schematic representation of the *B. subtilis* chromosome. DNA replication initiates at *oriC*, which is at position zero in the reference genome, and proceeds bidirectionally toward *terC*, which is at 1.97 Mb in the reference genome. The left and right replichores of the chromosome are in red and blue, respectively. **b**, Cumulative distributions of the indicated types of transitions along the genome. The origin of replication is indicated by the vertical dashed red line and the terminus is at both ends. **c**, A barplot displaying the mutation rate for the indicated types of transitions binned by replichore. The mutation rate is conditioned on the number of reference bases in each replichore as described in Materials and Methods. Error bars represent 95% confidence intervals determined by bootstrapping. MMR intact refers to wild type MA line data and MMR deficient refers to the pooled data for Δ*mutSL*, Δ*walJ*, and *mutL[E468K]*.

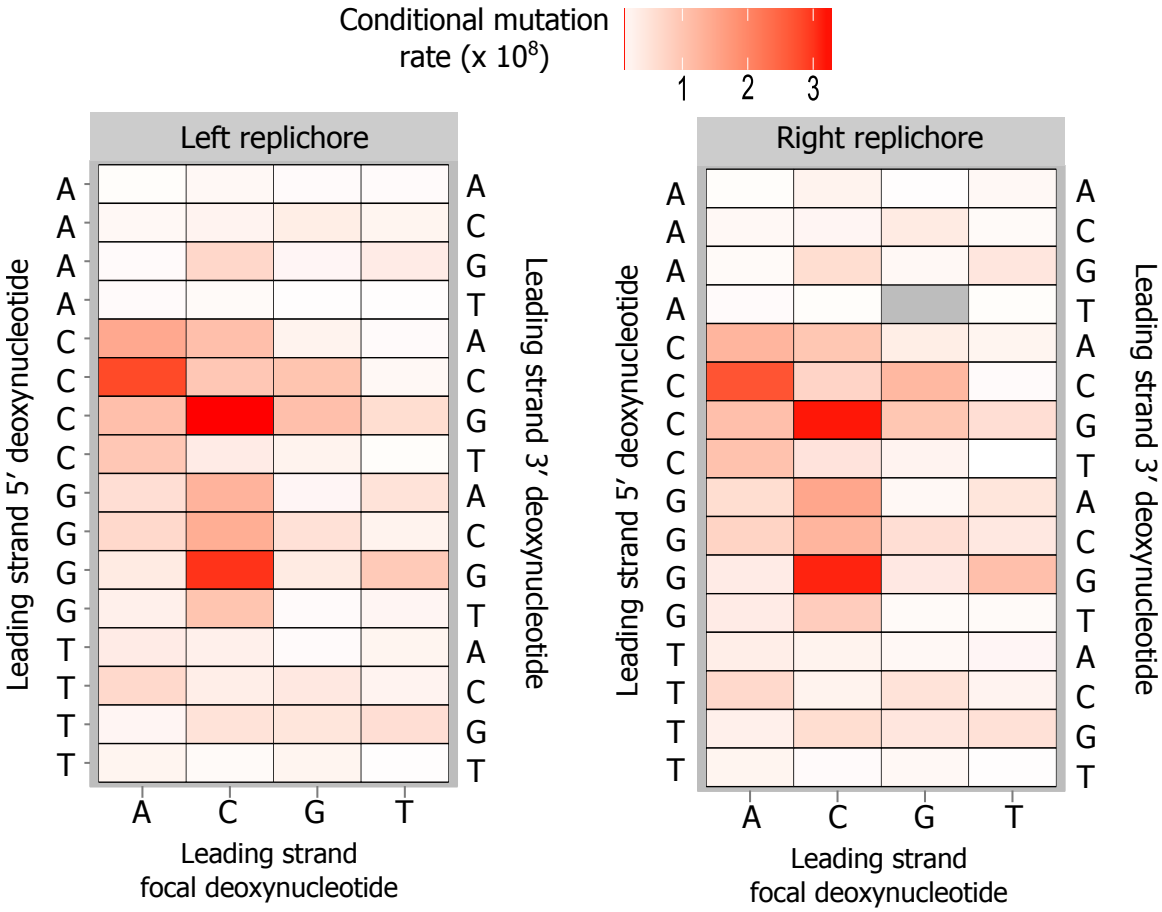


Figure 3.6: Leading strand neighboring nucleotide context affects transition rate symmetrically between the replichores. Heatmaps of transition rate for all 64 possible triplet nucleotide sequence contexts is plotted for each replichore in MMR- MA lines. The indicated sequences are with respect to the leading strand. The gray box indicates no mutations were observed for the given context.

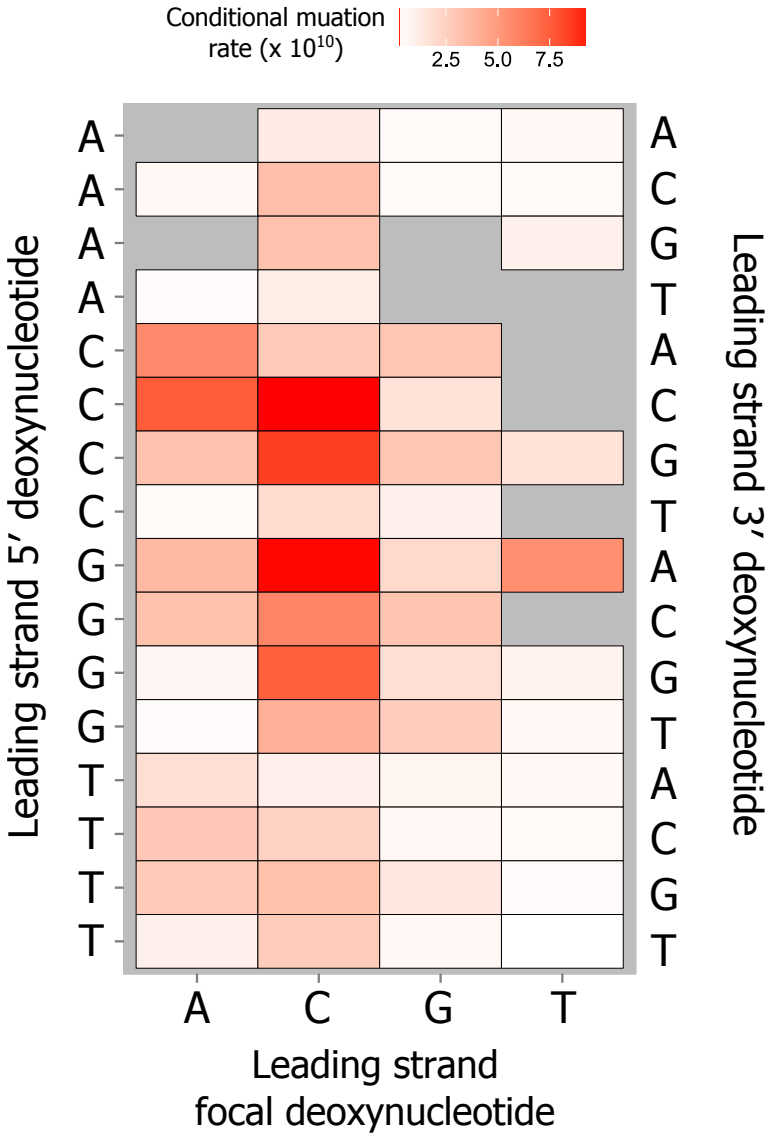


Figure 3.7: Wild type MA lines display context-dependent transition rates. Transition rates for each of the 64 possible triplet nucleotide sequence contexts in MMR+ MA lines are plotted in a heatmap. The indicated sequences are with respect to the leading strand. Gray boxes indicate no mutations were observed for the given context.

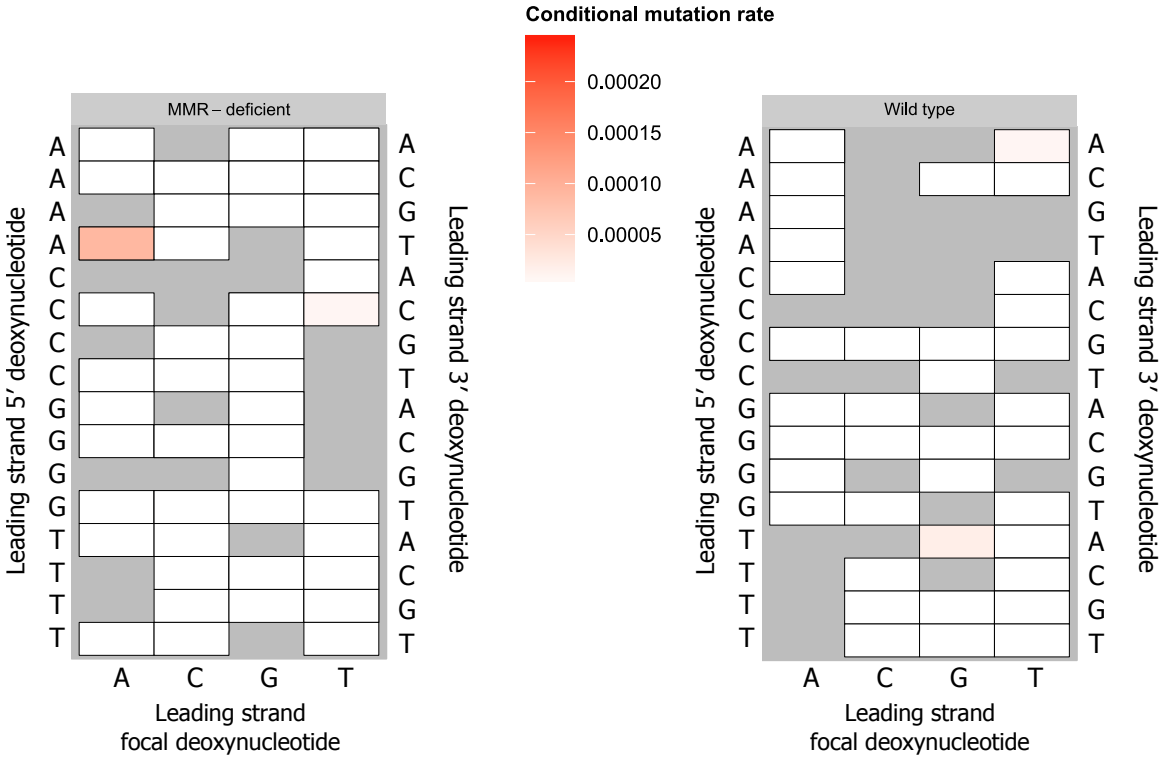


Figure 3.8: Paucity of transversions disallows detailed analysis of context dependence. Transversion rates for each of the 64 possible triplet nucleotide sequence contexts in MMR-deficient MA lines (left) and wild type MA lines (right) are plotted in heatmaps. The indicated sequences are with respect to the leading strand. Gray boxes indicate no mutations were observed for the given context.

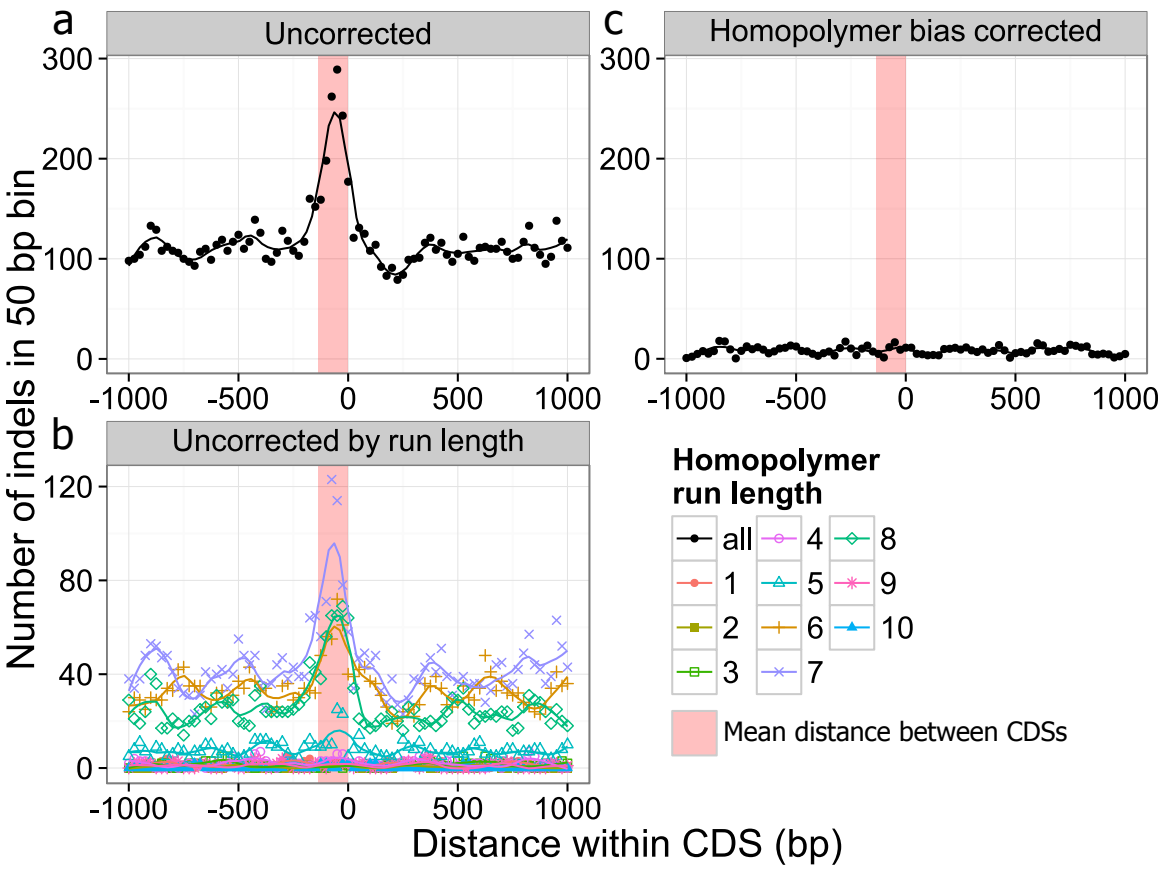


Figure 3.9: Indels in homopolymer runs are enriched outside of coding regions. All starts and ends of CDSs were aligned at relative position zero. Indels were counted in 50 bp bins, offset by 25. Negative distances indicate the indel was 5′ to a CDS start site. The lines in each plot represent the locally weighted polynomial regression (loess) fit to the data. **a**, The number of indels found in each bin without correcting for homopolymer run bias. **b**, Uncorrected indel counts separated by the homopolymer run length in which the indel was produced. **c**, Expected indel count in each bin after applying a correction for homopolymer run bias (see Supplementary Methods).

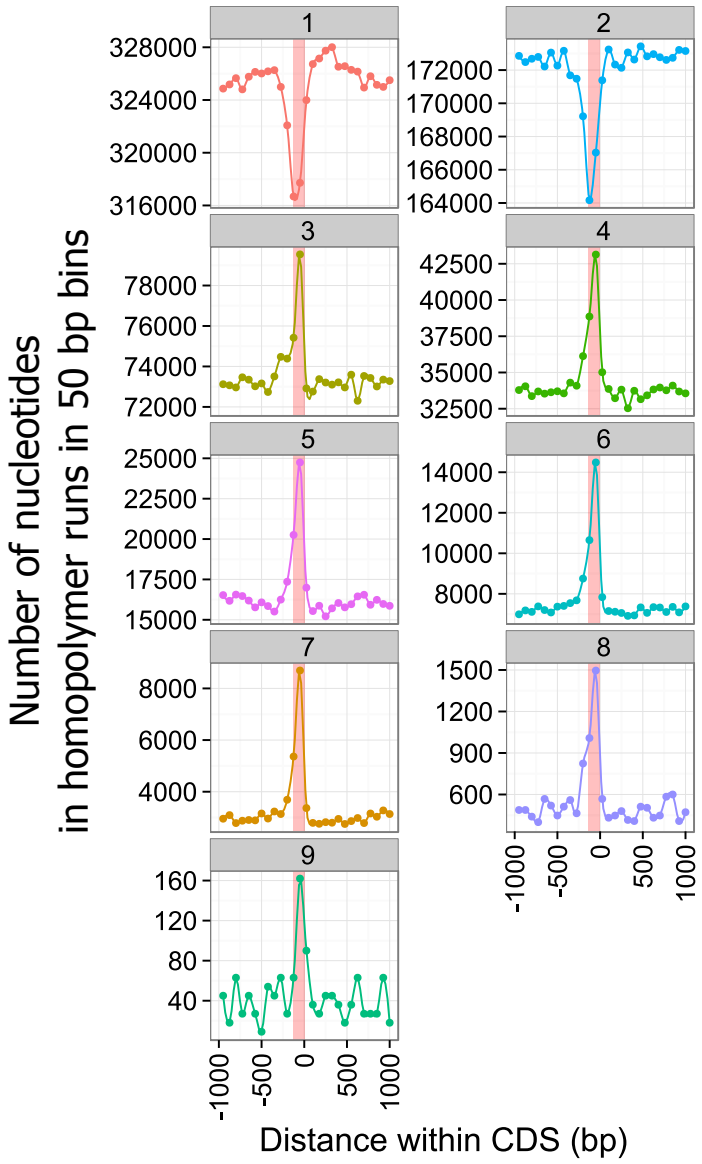


Figure 3.10: Homopolymer runs longer than two nucleotides are enriched outside of coding sequences. The number of nucleotides in homopolymers in 50 nucleotide wide bins offset by 25 is plotted, one plot for each homopolymer run length, 1 to 9. Because the bins overlap and any given homopolymer run may be within 1000 bps of multiple CDS start or end sites, a single homopolymer run may be counted multiple times. Negative distances indicate the nucleotide was either 5′ to a CDS start site or 3′ to a CDS end site. The counts for nucleotides in homopolymers of length 10 are not shown because they are extremely noisy due to there only being two such homopolymers in the *B. subtilis* genome. The vertical pink strip in each plot denotes the mean distance between CDSs.

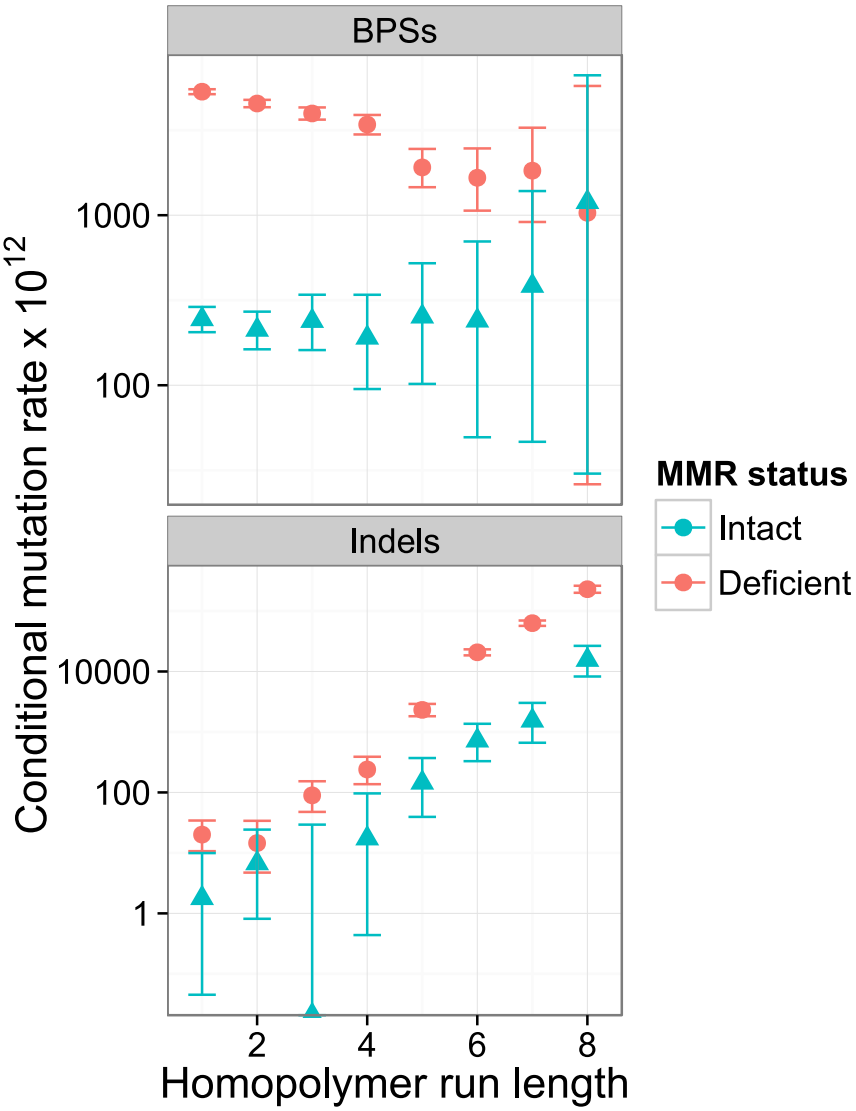
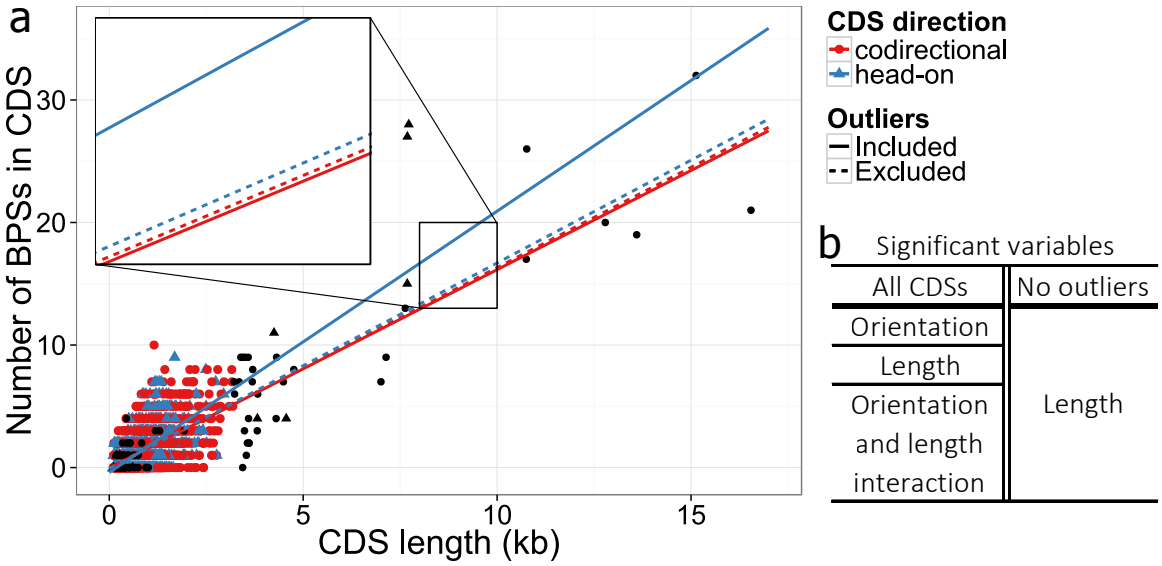


Figure 3.11: Indel mutation rate increases exponentially as homopolymer run length increases. Trends in conditional mutation rate are plotted versus homopolymer run length. Conditional mutation rate in homopolymer runs was calculated as discussed in Supplementary Methods. Error bars represent exact Poisson 95% confidence intervals.



b Significant variables

All CDSs	No outliers
Orientation	Length
Length	
Orientation and length interaction	

Figure 3.12: Coding sequence orientation and transcript abundance do not affect mutation occurrence. **a**, A graphical representation of linear regression analysis. Points in black are outliers in either CDS expression or length. **b**, A table listing the variables determined to be significantly associated with the average number of BPSs found in CDSs either with or without outliers. See Supplemental Table 3.4 for detailed results and Equation 3.5 for the model.

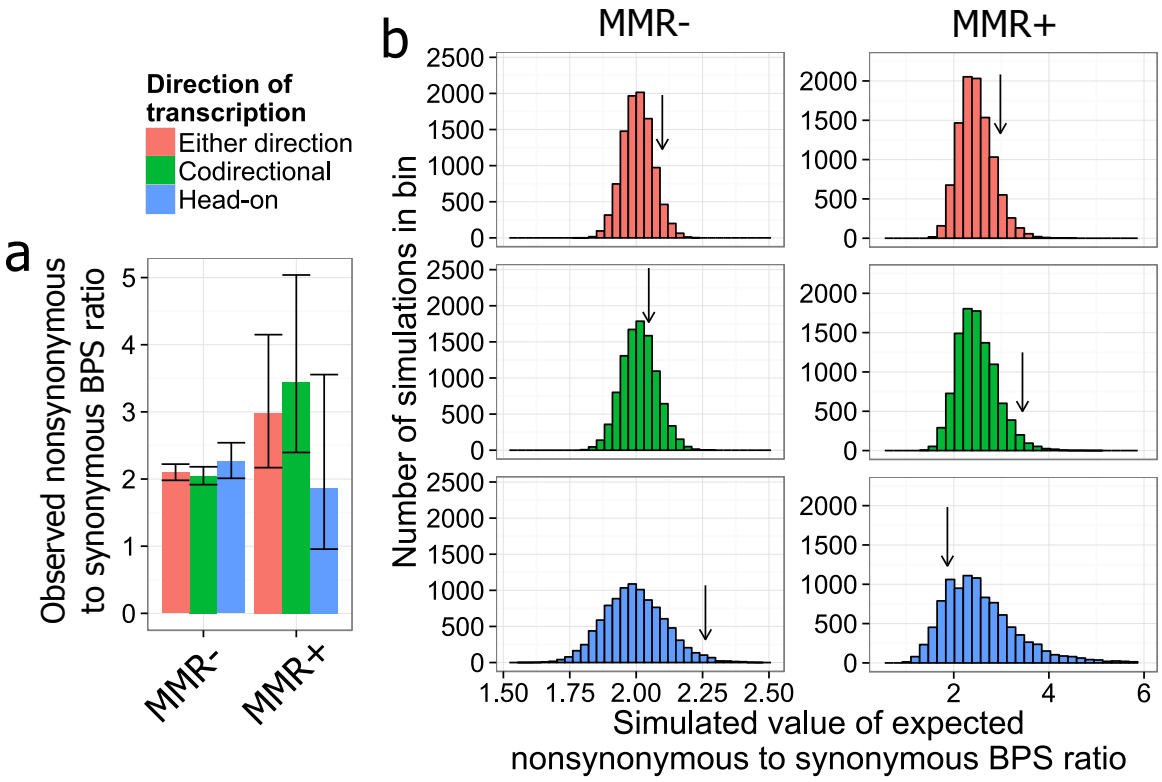


Figure 3.13: The ratio of nonsynonymous to synonymous BPSs is independent of CDS direction. a, A barplot displaying the dN/dS ratio observed in our MA lines for either all CDSs, those transcribed codirectionally with replication, or those head-on with replication. Error bars represent 95% confidence intervals determined by bootstrapping. **b**, Histograms showing the distributions of expected dN/dS ratios from 10000 Monte Carlo simulations (see Supplementary Methods). Black arrows indicate the observed dN/dS value in plot.

Table 3.1: Summary of mutation accumulation lines.

Strain	Number of lines	Generations per line	Total generations	BPSs per line	Indels per line	Mutation rate per genome replication ($x10^{3}$) [95% CI][a]	Mutation rate per genome replication per nucleotide ($x10^{10}$) [95% CI][a]	Generations between mutations
Wild type	75	3636	272700	3.4	0.5	1.1 [1.0-1.2]	2.7 [2.4-3.0]	909
Δ*walJ*	38	890	33820	35.3	6.8	47.3 [45.0-49.7]	118.5 [112.8-124.5]	21.1
Δ*mutSL*	36	890	32040	48.5	8.6	64.3 [61.5-67.1]	161.1 [154.2-168.2]	15.6

We observed 7247 mutations in total, 3294 of which were in *mutL[E468K]* MA lines, where MutL[E468K] is an endonuclease-deficient variant of MutL in *B. subtilis* (*Pillon et al*., 2010), and are not included in the table due to the high similarity of *mutL[E468K]* results to those of Δ*mutSL*.
a, 95% confidence intervals represent exact Poisson confidence intervals.

Table 3.2: Strains used in this study.

Strain	Relevant genotype	Reference
PY79	Prototrophic wild type, SPβ°	*Youngman et al.* 1984
JWS108	Δ*mutSL*	*Yao et al.* 2013
JWS112	Δ*walJ*	This work
JWS113	Δ*walJ* Δ*mutSL*	This work
JWS129	*walJ[H60A]*	This work
JWS208	founder colony for wild type MA lines	This work
JWS210	founder colony for Δ*mutSL* MA lines	This work
JWS211	founder colony for Δ*walJ* MA lines	This work
JWS224	*mutL[E468K]*	This work
JWS251	founder colony for *mutL[E468K]* MA lines	This work

Table 3.3: Regions of the genome with low (< 20) median mapping quality.

Start	End
10000	14600
30400	35000
90800	95200
96400	101000
161000	165600
166400	171000
171600	176200
598600	603200
909800	914200
2992000	2994800

Total number of nucleotides with low median mapping quality: 43810

Total number of nucleotides with high mapping quality: 3989649

Table 3.4: Effect of CDS length, expression and orientation on BPS count: regression estimates

	All CDSs	**No Outliers**
Intercept	-0.03	-0.03
	(0.03)	(0.04)
Head_on orientation	-0.37^{***}	-0.05
	(0.07)	(0.08)
Length (kb)	1.62^{***}	1.64^{***}
	(0.03)	(0.04)
RPKM	-4.1×10^{-6}	-9.0×10^{-5}
	(1.2×10^{-5})	(4.9×10^{-5})
Head_on orientation*Length	0.51^{***}	0.04
	(0.06)	(0.08)
Head_on orientation*RPKM	7.4×10^{-6}	3.2×10^{-5}
	(3.8×10^{-5})	(1.3×10^{-4})
Number of CDSs	4163	4062
R^2	0.54	0.35
Resid. sd	1.24	1.18

Standard errors in parentheses

*** $p \leq 0.001$, ** $p \leq 0.01$, and * $p \leq 0.05$

CHAPTER IV

Concluding Remarks and Future Directions

4.1 Introduction

Throughout the course of this dissertation, mechanisms of ribonucleotide misincorporation, excision and repair have been discussed in considerable detail (Chapter I), investigations of the target-search dynamics carried out by single MutS molecules *in vivo* have been discussed (Chapter II), and a detailed analysis of genome-wide factors affecting mismatches produced during DNA replication was performed (Chapter III). This chapter is centered on a brief discussion of questions that remain open within the fields of mismatch repair (MMR), mutagenesis and ribonucleotide excision repair (RER). Future directions for research in these topics will be considered, emphasizing ongoing work in RER.

4.2 Mismatch repair

There are still many open questions in the field of mismatch repair (MMR), arguably the greatest of which is the origin of strand-discrimination signals in organisms lacking methyl-directed MMR systems. For any system in which MMR has been reconstituted *in vitro*, a single nick in a strand of DNA is able to direct MMR to replace the nick-containing strand (*Holmes et al.*, 1990; *Thomas et al.*, 1991; *Lacks et al.*, 1982). In the

lagging strand, nicks between Okazaki fragments are likely to define the nascent strand for MMR, but in the leading strand the source of the nicks that direct MMR is unclear. As discussed in Chapter I, *Saccharomyces cerevisiae* MMR has been reported by Lujan and colleagues to use misincorporated ribonucleotides (rNMPs) to gain a slight increase in MMR efficiency on the leading strand (*Lujan et al.*, 2013), though this has been challenged by others (*Allen-Soltero et al.*, 2014). Nicks generated at rNMPs *in vitro* are able to direct strand choice, but whether this is the case *in vivo* remains unclear (*Ghodgaonkar et al.*, 2013). Future and ongoing work aims to identify whether rNMPs can serve as strand discrimination signals *in vivo* for *Bacillus subtilis* (see Chapter 4.4). Biochemical characterization of the role WalJ plays in MMR in *B. subtilis* would also prove very interesting, but WalJ is insoluble under the numerous conditions in which its overexpression and purification have been attempted (Lindsay Matthews, personal communication; data not shown).

4.3 Mutagenesis

What is the source of mutation? There is no single suitable answer to this question, as there are many potential sources of mutation and they occur under varying circumstances and in variable contexts. During times of stress, bacteria undergo stress-induced mutagenesis, which was at one point thought to target the genes that must, by necessity, be mutated to overcome the stressor (*Cairns et al.*, 1988). It is now appreciated that although mutagenesis is stimulated under stressful conditions, the mutations that accumulate during stress-induced mutagenesis are random (*Foster*, 2007). Stress-induced mutagenesis is, itself, an umbrella term for myriad sources of mutation that come into play during various stress conditions (*Foster*, 2007). Of course errors made during DNA replication under relatively stress-free growth conditions are a source of mutation in bacteria as well, and evolution occurs due to natural selection acting on genomes which are dynamic due to multiple potential sources of mutation.

Understanding mutation, whether due to stress or normal cellular processes, will yield a greater understanding of evolution.

The issue of whether mutations are driven by stress or normal cellular processes - one could, of course, argue many stress responses to be part of normal life for bacteria - is touched on in Chapter III, where I conclude that when natural selection's influence on mutations that are observable in a population is nearly eliminated, it is clear that local sequence context is the only identifiable factor associated with the likelihood of mutation occurring. Thus, my model for mutagenesis states that under relatively stress-free conditions the major contribution to mutation is replicating DNA polymerase errors produced more often in specific sequence contexts, and this is unaffected by gene presence or direction. In *S. cerevisiae* it has been noted that mutations are more likely to occur near Okazaki junctions and locations of the genome where proteins are commonly bound (*Lujan et al.*, 2014; *Reijns et al.*, 2015). The error-prone DNA polymerase pol α was found to be responsible for the increased mutagenesis near Okazaki junctions in *S. cerevisiae*, and perhaps the analogous DNA polymerase subunit in *B. subtilis*, DnaE, may be responsible for the complementary symmetry of transitions between the replichores of the genome (see Chapter 3.2 and Figure 3.5), but we currently have no experimental evidence that this is the case. Furthermore, the mutation accumulation (MA) line study described in Chapter III is in some ways not entirely representative of nature, where bacteria must be under stressful conditions more often than during MA lines such as those described in this dissertation.

Published work suggesting a higher mutation rate for genes oriented such that transcription and replication will engage in head-on collisions may be an accurate depiction of mutagenesis when specific conditions are met (*Paul et al.*, 2013; *Million-Weaver et al.*, 2015). However, in order to rigorously test whether mutations arise more frequently in genes transcribed head-on with DNA replication in a naturally

evolved genome rather than in a reporter gene placed in a foreign context, a genome-wide view of the mutations that occur under stressful conditions will be necessary. Future work making use of MA lines under stressful conditions will be very illuminating in determining genome-wide factors affecting stress-induced mutagenesis without the use of mutational reporters that greatly limit the mutations and sequence contexts that can be observed. MA lines using variants of PolC and DnaE with defined mutation spectra will also be fruitful in determining each polymerase subunit's contribution to overall DNA replication and mutagenesis, but care must be taken in the interpretation of data resulting from mutator DNA polymerase variants (*Strauss et al.*, 2000; *Johnson et al.*, 2015).

4.4 Ribonucleotide excision repair

As it is a relatively new field, there are innumerable questions of great importance for RER. The single greatest question currently in bacterial RER is the following: what are the effects of unrepaired ribonucleotides (rNMPs) in DNA on bacterial genome stability? The consequences for *S. cerevisiae* are beginning to be understood (*Williams and Kunkel*, 2014), but in bacteria, very little is understood about this important question. A number of possibilities, some experimentally tested and others untested, were discussed in Chapter 1.4. One such possibility is that nicks generated at rNMPs in genomic DNA, either by uncatalyzed hydrolysis or during ribonucleotide excision repair (RER), may signal strand discrimination for MMR in organisms lacking a methyl-directed MMR system. I show in the Appendix that loss of either *rnhB*, encoding RNase HII (see Chapter 1.3.4), or *rnhC*, encoding RNase HIII (see Chapter 1.3.5), in *B. subtilis* leads to a modest increase in overall mutation rate, but with differing mechanisms for the increased mutagenesis. A model for increased mutagenesis in Δ*rnhB* and future directions are discussed. Importantly, the model does not invoke an effect of rNMPs on MMR.

More experiments evaluating to what extent rNMPs are linked to MMR *in vivo* are necessary. MA lines have proven extremely useful in determining the effect of local sequence context on mutagenesis caused by rNMPs *in vivo*, revealing that from the leading strand (or lagging strand template) perspective, the G in the motif 5′-GCC(T/C)T-3′ is about 4 times more likely to undergo transition in the absence of *rnhB* (Figure A.7). A working model suggests that the motif exerts its influence on rNMP-dependent transitions when it resides in the lagging strand template. This is arrived at by the logic that rA is the most common type of sugar error (*Yao et al.*, 2013). Therefore, the effect of thymidine four nucleotides 3′ to G is likely due to misinsertion of rA in the lagging strand across from the thymidine in question (Figure A.8). In the absence of canonical, RNase HII-dependent RER, NER is able to recognize the strand-distorting rA to excise the DNA between eight phosphodiester bonds 5′ and five phosphodiester bonds 3′ to the rA (*Vaisman et al.*, 2013). The repair DNA polymerase Pol I then fills the gap generated during NER, and due to the high GC content at the 3′ end of the gap, is inaccurate at that location, thus inserting a T across from the G that is to undergo transition (Figure A.8). MMR will not efficiently recognize the resulting G/T mismatch, as the replisome has likely moved too far from the site of repair to recruit MutS to the mismatch (see Chapter II), and during the subsequent round of DNA replication A will be placed across from the misinserted T and the GC$\rightarrow$AT transition will become fixed in the genome as a mutation (Figure A.8).

4.5 Discussion of ongoing work

New frontiers in MMR, mutagenesis and RER await exploration. In particular, investigation of the effects of rNMPs in genomic DNA on bacterial genome stability and physiology will prove fruitful in advancing our understanding of the ubiquitously conserved process of rNMP misincorporation (Chapter 1.3.1). Of course the proposed model (Figure A.8) requires testing *in vitro*, including biochemical evaluation of DNA

polymerase accuracy with respect to sugar selection and the frequency of mismatch formation near rNMPs and given differing sequence contexts, specifically the accuracy of Pol I in the contexts described in Figure A.8. Indeed, the assertion that NER is able to excise rA-containing DNA in *B. subtilis* is an extrapolation from what is known to be the case in *E. coli*, and this notion must be rigorously tested *in vitro* using purified *B. subtilis* proteins. Future MA lines making use of *B. subtilis* lacking NER, RER and MMR will also be instrumental in testing the model *in vivo*.

APPENDICES

APPENDIX A

Genome-wide Consequences of Unrepaired Ribonucleotides in Bacterial DNA

A.1 Ongoing work in ribonucleotide excision repair

In this section, ongoing analyses of MA lines performed using Δ*rnhB*, Δ*rnhC* and Δ*rnhB*, Δ*mutSL* strains of *B. subtilis* will be discussed and these new data will be incorporated with Δ*mutSL* and PY79 (wild type) MA line data from Chapter III. These ongoing experiments and analyses are aimed at finding an answer to the question, what are the effects of unrepaired rNMPs in DNA on bacterial genome stability?

A.1.1 Base pair substitution rate increases upon loss of *rnhB* or *rnhC*

Mutation accumulation lines were carried out as described in Chapter 3.4, and an overall summary of the MA lines which are pertinent to this chapter is provided in Table A.1. In agreement with prior work (*Yao et al.*, 2013), loss of *rnhB* or *rnhC* yielded an increase in overall mutation rate to $\approx$1.5-fold that of wild type *B. subtilis* (Table A.1).

William Hirst and I carried out the mutation accumulation line procedure, and Heather Schroeder was very helpful in providing advice on statistics.

Specifically, base pair substitutions (BPSs) are increased in both strain backgrounds (Figure A.1), with GC→AT transitions occurring ≈ 2-fold more frequently in Δ*rnhB*, and transversions occurring more frequently in Δ*rnhC* (Figure A.2). An increased proportion of mutations that were transversions in Δ*rnhC* suggested that SOS may be induced in the absence of RNase HIII, which we confirmed using RNA-seq (Chapter A.1.2).

A.1.2 Loss of RNase HIII causes replication stress

We carried out RNA-seq to assess whether a transcriptional response including SOS induction was initiated upon loss of RNase HIII. Upon initial investigation of the RNA-seq results, it seemed the SOS response may have been initiated, so we additionally performed RNA-seq in *lexA::erm* and *rnhC::erm*, *lexA[G92D]* strains. The *lexA::erm* strain has SOS response constitutively activated, and LexA[G92D] is a variant of LexA that is unable to initiate the SOS response. Clustering of the $\log_2$(fold change) in gene expression for the genes determined to be differentially expressed in all three strains revealed that genes in the SOS regulon clustered together and were regulated in a manner consistent with SOS induction upon loss of *rnhC*, i.e., they were highly expressed in *lexA::erm* and Δ*rnhC*, but their expression was close to wild type levels in *rnhC::erm*, *lexA[G92D]* (Figure A.3). Induction of the SOS response implied replication stress may result from loss of RNase HIII.

Marker frequency analysis by genome sequencing can be used to determine the DNA replication status along the genome (*Srivatsan et al.*, 2010; *Rudolph et al.*, 2013; *Walsh et al.*, 2014; *Maduike et al.*, 2014; *Ivanova et al.*, 2015). In order to determine whether replication stress was evident in Δ*rnhC*, and to what extent it was alleviated by SOS induction, we carried out marker frequency analysis by sequencing genomic DNA from wild type, Δ*rnhC*, *lexA::erm*, *lexA[G92D]* strains as well as a *rnhC::erm*, *lexA[G92D]* strain. Deletion of *rnhC* resulted in striking changes in the distribution of genome sequencing coverage. By analyzing marker frequency in terms of the $\log_2$(fold

enrichment) in sequencing coverage of Δ*rnhC* compared to wild type, we found that the left half of the reference genome in Δ*rnhC* displayed enrichment of reads, and the right half exhibited a lower number of reads than expected (Figure A.4). There was also a conspicuous peak in sequencing coverage enrichment at genomic cooridinate ≈ 1.3 Mb in the Δ*rnhC* strain (Figure A.4). This is the location of the defective prophage PBSX, which is known to be induced for gene expression upon SOS induction in *B. subtilis* (*Goranov et al.*, 2006). However, this peak is absent from the *lexA::erm* enrichment profile (Figure A.4), suggesting that perhaps R-loops that go unresolved at this region in Δ*rnhC* are priming replication to commence from the PBSX locus. This would also explain the higher sequencing coverage in Δ*rnhC* on the left half of the reference genome, as replication from both *oriC* and PBSX would result in about twice the coverage of the half of the chromosome on which PBSX resides. Furthermore, it is clear that SOS is important for alleviating replication stress in Δ*rnhC* cells, as genome-wide sequencing coverage is severely altered in a strain lacking both *rnhC* and SOS (*rnhC::erm*, *lexA[G92D]*) (Figure A.4). Work is currently ongoing to test whether replication is able to commence from PBSX in Δ*rnhC* cells.

A.1.3 Transcriptional response to loss of RNase HII

Few genes exhibited differential expression upon deletion of *rnhB*. The top eleven are summarized in Table A.3. Of particular interest is the gene *yjqB*, as it encodes a putative cyclic phosphatase. It is an intriguing possibility that *yjqB* could encode a protein analogous to RtcB (see Chapter 1.4.1) of *Escherichia coli* (*Das et al.*, 2013). In the absence of RNase HII, hydrolysis of the sugar-phosphate backbone of DNA would result in 2′,3′-cyclic phosphate ends which may be partially resolved by action of the protein encoded by *yjqB*. We are currently deleting *yjqB* in cells also lacking *rnhB* to determine whether there is a synergistic effect of the two deletions.

A.1.4 Some mismatches are corrected independently of mismatch repair in Δ*rnhB*.

In order to test whether rNMPs in genomic DNA are involved in strand discrimination for DNA mismatch repair (MMR) we carried out MA lines in a strain lacking MMR and RNase HII (Δ*rnhB*, Δ*mutSL*). If nicks generated at singly misinserted rNMPs are able to direct strand discrimination for MMR, RER and MMR will be epistatic with respect to mutation rate. Strikingly, Δ*rnhB*, Δ*mutSL* exhibited a BPS rate significantly lower than that of Δ*mutSL* alone (Figure A.5). If nucleotide excision repair (NER) is able to effect the removal of rNMPs in *B. subtilis* DNA as is the case in *E. coli* (*Vaisman et al*., 2013), NER's action on rNMPs may also carry out the removal of mismatches that are near rNMPs. This would require that mismatches be very close to rNMPs, as NER in *E. coli* excises a patch of DNA 12-13 nucleotides in length, comprising the DNA from the eighth phosphodiester bond 5′ to, and the fourth to fifth phosphodiester bond 3′ to the rNMP recognized by NER. This model for an indirect effect of rNMPs on MMR requires base pairing errors to be more likely within four to five nucleotides 3′ to rNMP misincorporations in order for NER to remove both the rNMP and the mismatch. We assert this is likely to be the case, as rNMPs distort the helical geometry of DNA (*Rychlik et al*., 2010; *DeRose et al*., 2012), which may decrease the accuracy of nucleotide selection or the efficiency of proofreading by the replicating DNA polymerase. If rNMPs affect MMR *in vivo*, they likely do so to a small extent and under very specific circumstances, i.e., when canonical RER is nonfuntional. Ongoing work testing whether rNMPs affect MMR in *B. subtilis* includes MA lines in strains lacking MMR, RER and NER as well as biochemical assessment of DNA polymerase accuracy near rNMPs in the primer strand.

A.1.5 GC→AT transitions caused by ribonucleotides are strongly strand- and sequence-context dependent

Observing the accumulation of transitions along the genome as cumulative distributions, it is evident that the increase in GC→AT transitions in Δ*rnhB* shows a strong strand-dependence (Figure A.6). Keeping in mind that the right replichore's reference sequence represents the leading strand and the left replichore's reference sequence represents the lagging strand, GC→AT transitions are increased due to ribonucleotides at genomic loci where G is in either the leading strand or the lagging strand template. We tested whether transitions due to misinserted rNMPs were more likely in certain sequence contexts by calculating conditional mutation rates for transitions in the sixteen possible dinucleotide sequence contexts. All sequence contexts were considered from the perspective of the leading strand, and the analysis was performed separately for 5′ and 3′ contexts. We did not detect any significant effects of 5′ neighboring nucleotides on rNMP-induced transition rate, but 3′ neighboring context had a very strong effect (Figure A.7A). Specifically, if a G is present in the leading strand followed 3′ by a C, the G is ≈ 4 times more likely to undergo a transition in Δ*rnhB* than in wild type (Figure A.7A).

Upon finding a strong dependence for GC→AT transition on neighboring sequence context in Δ*rnhB*, logistic regression was performed genome-wide to determine whether sequence context at distances up to 5 nucleotides 3′ to a G in the leading strand could influence mutation rate at the G (see Equation A.1 and associated Methods). Strikingly, nucleotide identity up to 5 nucleotides 3′ to a G in the leading strand were associated with GC→AT transition rate in Δ*rnhB* (Figure A.7B). Using the bioniformatic tool MEME (*Bailey and Elkan*, 1994), the sequence motif determined to be present at GC→AT transitions in which G was in the leading strand sequence in Δ*rnhB* was 5′-G̲CC(T/C)T-3′, where the underlined G indicates the position to become a transition (Figure A.7). A working model for ribonucleotides' strand- and sequence

context-dependent increase in GC→AT transitions is discussed in Chapter 4.5 and presented in Figure A.8.

A.2 Materials and Methods

A.2.1 Mutation accumulation line protocol

MA lines were carried out as described in Chapter 3.4.3

A.2.2 Calculation of conditional mutation rate

Mutation rates were conditioned on the number of opportunities for mutations in given sequence contexts in a manner similar to that used in Equation 3.2 in Chapter 3.4.9.

A.2.3 Logistic regression of sequence context effect on mutation

To determine the impact of local sequence context on G to A transitions in the leading strand (or lagging strand template), the logit of the probability of obtaining a G to A transition against nucleotide base identity up to five nucleotide positions in either direction of a G in the leading strand. This revealed significant effects from one position in the minus (leading strand 5′) direction and through five positions in the plus (leading strand 3′) direction. Therefore, the regression was repeated, including only positions minus one through plus 5 in the model:

$$\begin{aligned} ln\left(\frac{p}{1-p}\right) = \alpha &+ \beta Base_{-1} + \beta Base_{+1} \\ &+ \beta Base_{+2} + \beta Base_{+3} \\ &+ \beta Base_{+4} + \beta Base_{+5} \end{aligned} \tag{A.1}$$

where p is the probability of a given G in the leading strand having a transition occur and $Base_i$ represents either A, C, T or G at position i relative to the G in question. The log-odds ratios were calculated relative to A as the baseline.

A.2.4 Motif identification at leading strand G to A transitions

A mutifasta file containing the local sequence from all 162 sites where a G in the leading strand was transitioned to an A in Δ*rnhB* was prepared along with a multifasta file containing local sequence contexts at 1000 randomly selected G in the leading strand for a negative control. Specifically, the sequence contexts for motif detection extended one position 5′ through six positions 3′ to the transition site, from the leading strand perspective. MEME was run in discriminative mode using the file containing the 1000 negative control sequences with the minimum possible motif width set to two (*Bailey and Elkan*, 1994). All other parameters were default values. The resulting MEME logo was trimmed to include only positions one through four nucleotides 3′ to the transition site.

A.2.5 RNA-seq

RNA-seq and subsequent data analysis was performed as described in Chapter 3.4.5.

A.2.6 Statistical analysis

Statistical analysis, data manipulation, logistic regression and plotting were performed using the statistical computing software R. Throughout the Appendix, *** denotes $p \leq 0.001$, ** denotes $p \leq 0.01$, and * denotes $p \leq 0.05$.

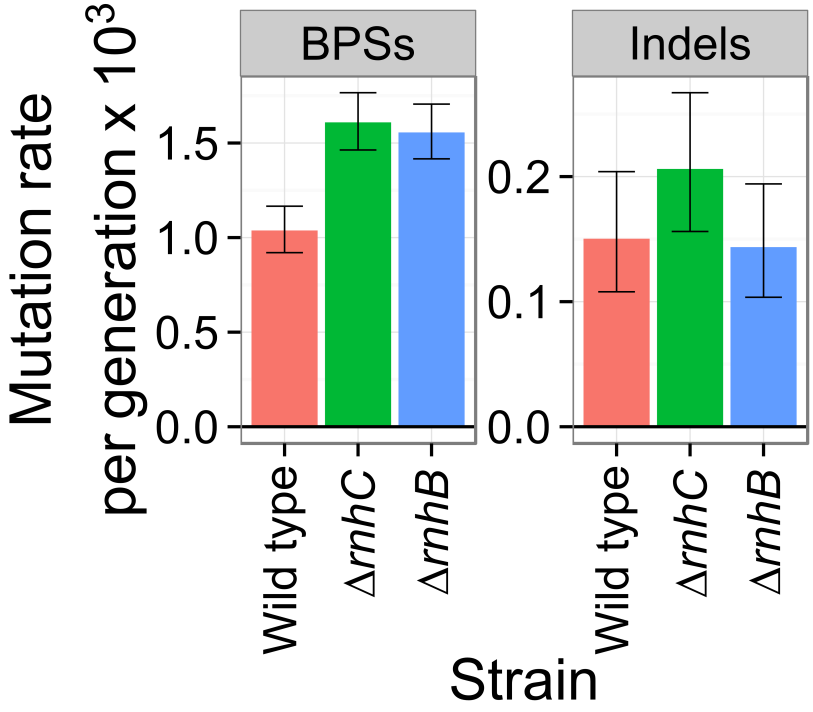


Figure A.1: Unrepaired ribonucleotides increase base pair substitution rate. Barplots displaying the mutation rate for base pair substitutions (BPSs) and insertions/deletions (Indels) are presented. Error bars represent 95% Poisson confidence interval.

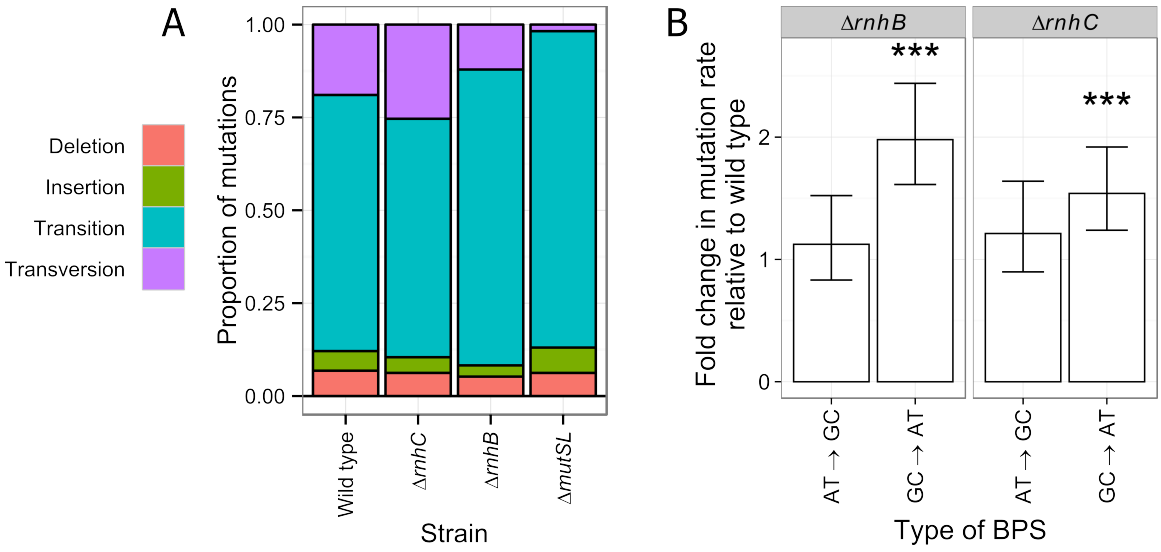


Figure A.2: Transitions are increased in Δ*rnhB* and transversions are increased in Δ*rnhC*. (A) Mutation spectra for wild type, Δ*rnhB*, Δ*rnhC* and Δ*mutSL* MA lines are shown as barplots. (B) Bar plots showing the rate of AT→GC and GC→AT transition subtypes relative the their rates for wild type MA lines. Error bars represent 95% Poisson confidence interval.

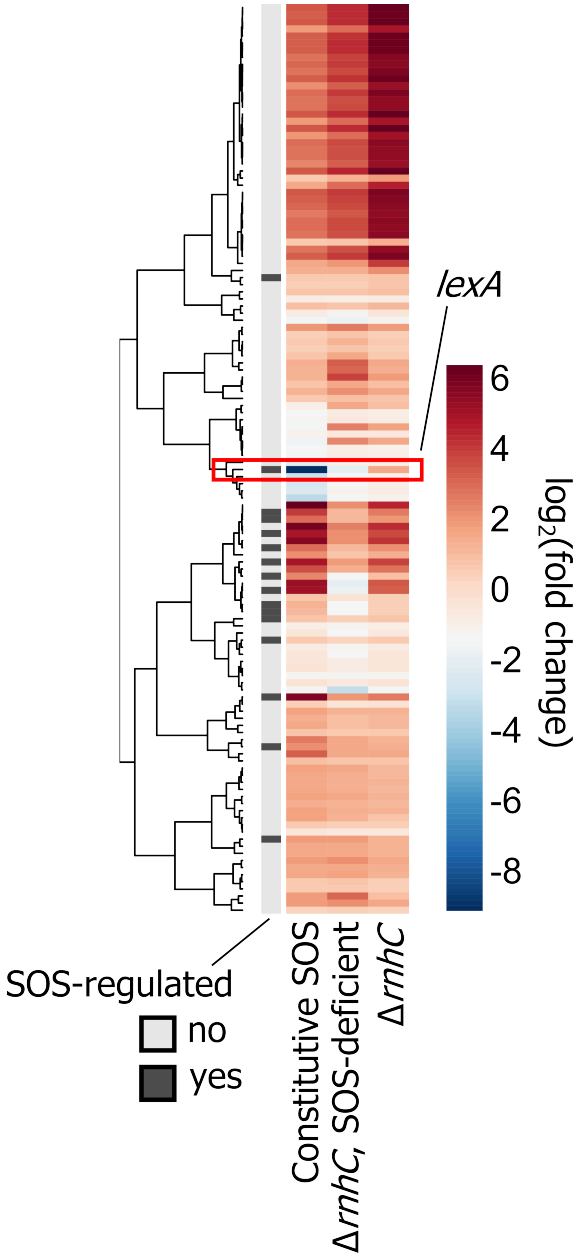


Figure A.3: Heatmap of Δ*rnhC* RNA-seq data. Genes with differential expression in Δ*rnhC*, *lexA::erm*, and the *rnhC::erm*, *lexA[G92D]* strain were clustered according to their Pearson correlation distance using complete linkage. The row corresponding to the *lexA* gene is outlined with a red box. The heatmap's scale indicates the $\log_2$(fold change) in gene expression for the given gene and strain compared to wild type.

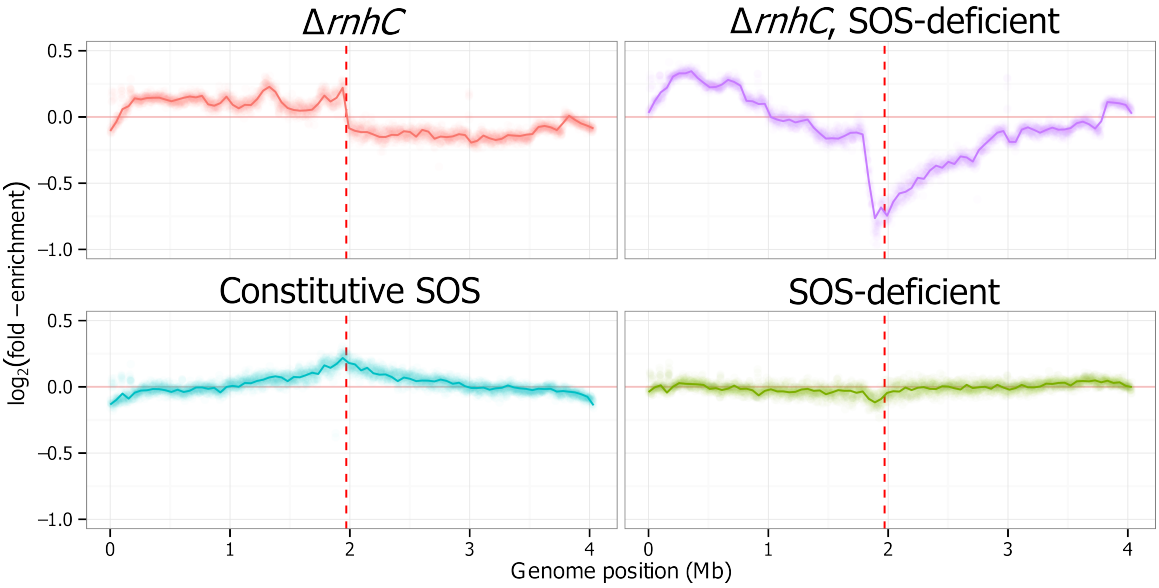


Figure A.4: Replication stress in Δ*rnhC*. Log_2(fold enrichment) of sequencing coverage along the genomes of the indicated strains compared to wild type coverage from mid-exponential phase cultures is plotted versus genome position. The vertical dashed red line indicates the location of *terC*, i.e., the terminus of replication, and *oriC* is located on each end.

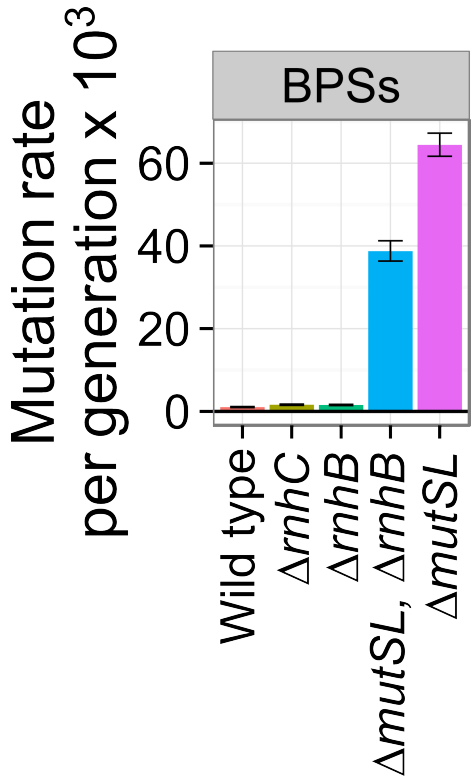


Figure A.5: Some mismatches are corrected independently of mismatch repair in Δ*rnhB*. A barplot depicts the BPS rates for the indicated strains. Error bars represent 95% Poisson confidence interval.

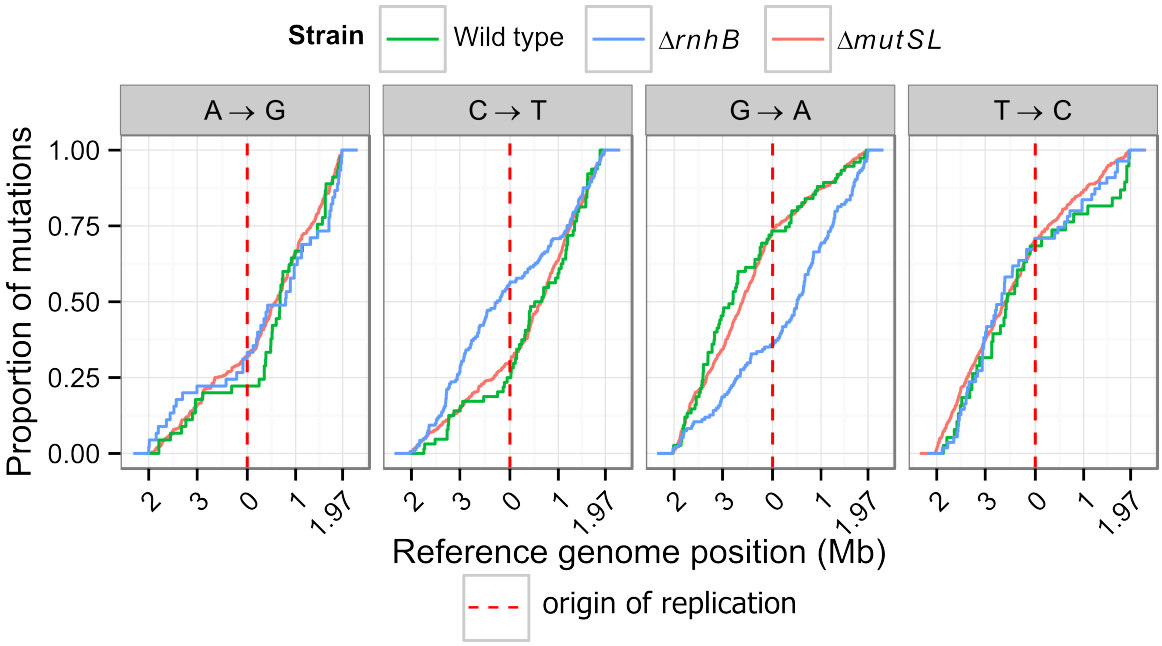


Figure A.6: Increase in GC→AT transition in Δ*rnhB* is strand-biased. The genome-wide distributions of the four possible transition subtypes, i.e., A→G, G→A, T→C and T→C, are shown for wild type, Δ*mutSL* and Δ*rnhB* as cumulative distributions. The origin of replication is denoted by a vertical red dashed line in each plot. Note that a steeper slope in the cumulative distribution is indicative of a higher rate of accumulation in that region of the genome.

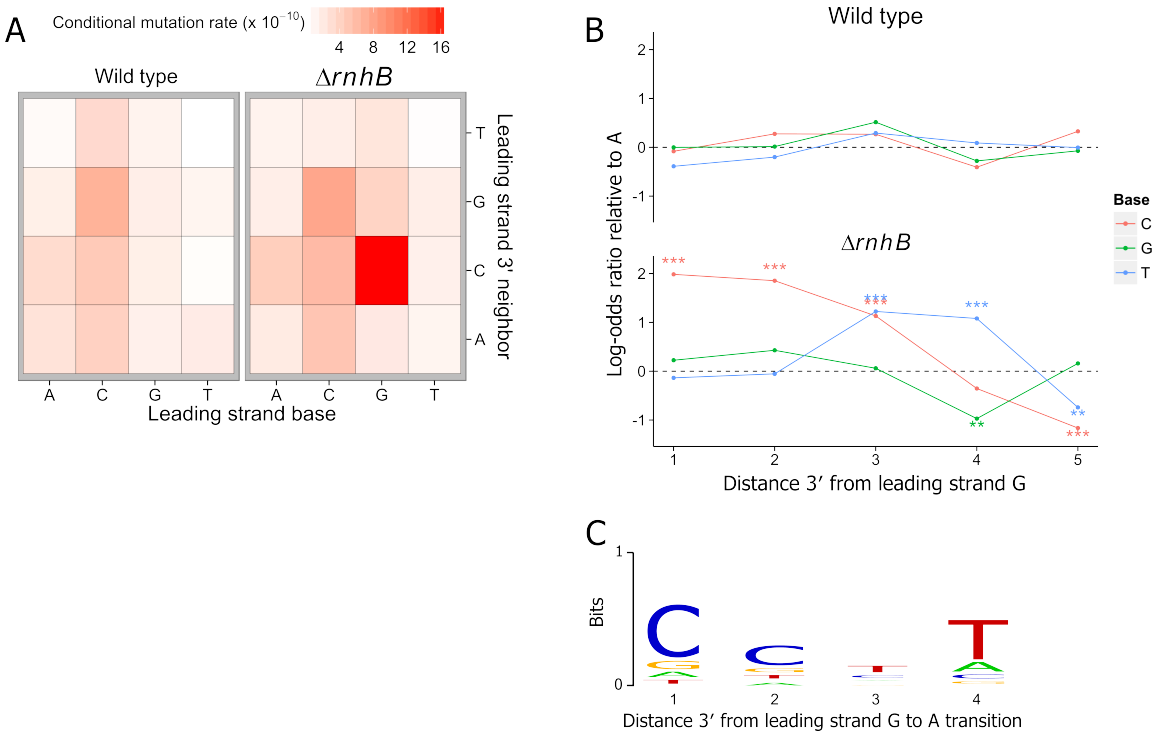


Figure A.7: Mutagenesis due to ribonucleotides is sequence context-dependent. (A) A heatmap of the mutation rate in wild type or Δ*rnhB* for transitions in the sixteen possible dinucleotide sequence contexts. The contexts are considered with respect to the leading strand sequence, regardless of the chromosome arm on which they are found. (B) The log-odds ratios for the effect of nucleotides with the given base on transition rate up to 5 nucleotides 3′ to G in the leading strand sequence. Note that because log-odds ratios are presented, the effects of cytosine, guanosine and thymidine are shown relative to the effect of adenosine, the effect of which is zero by definition. (C) A sequence logo representing the motif immediately 3′ to guanosines in the leading strand of Δ*rnhB* that underwent transitions.

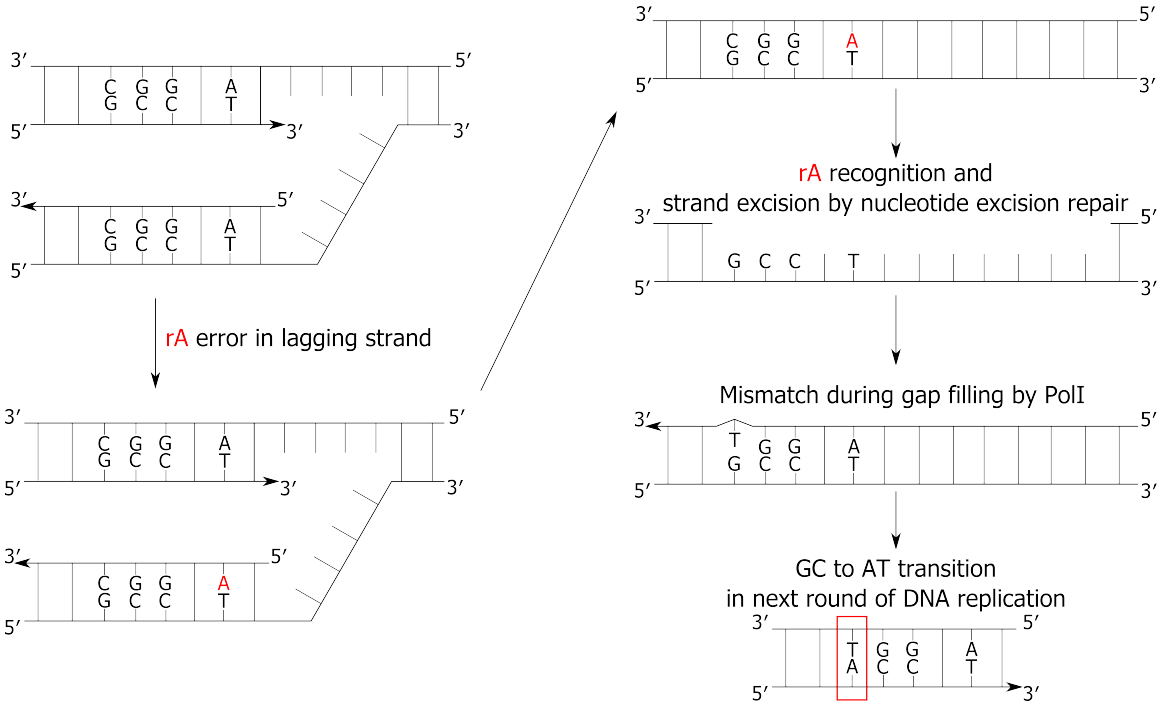


Figure A.8: Model for mutagenesis due to unrepaired ribonucleotides. A model is proposed by which a G in the lagging strand template followed 3′ by the motif CC(T/C)T undergoes a transition to A in a manner that is dependent on unrepaired ribonucleotides incorporated during DNA replication.

Table A.1: Summary of mutation accumulation lines discussed in this appendix.

Strain	**Number of lines**	**Total generations**	**Mutation rate per genome replication ($x10^3$) [95% CI][a]**	**Mutation rate per genome replication per nucleotide ($x10^{10}$) [95% CI][a]**
PY79 (wild type)	75	272700	1.2 [1.1-1.3]	3.0 [2.7-3.3]
Δ*rnhB*	81	292410	1.7 [1.6-1.9]	4.3 [3.9-4.7]
Δ*rnhC*	75	276525	1.8 [1.7-2.0]	4.6 [4.2-5.0]
Δ*rnhB*, Δ*mutSL*	30	25170	52.8 [50.0-55.7]	132.3 [125.3-139.7]
Δ*mutSL*	36	32040	74.3 [71.3-77.3]	186.1 [178.7-193.7]

a, 95% confidence intervals represent exact Poisson confidence intervals.

Table A.2: Strains used in this study.

Strain	Relevant genotype	Reference
PY79	Prototrophic wild type, SPβ°	*Youngman et al.* 1984
JWS105	Δ*rnhB*	*Yao et al.* 2013
JWS108	Δ*mutSL*	*Yao et al.* 2013
JWS207	Δ*rnhC*	*Yao et al.* 2013
JWS208	founder colony for wild type MA lines	Chapter III
JWS209	founder colony for Δ*rnhB* MA lines	This work
JWS210	founder colony for Δ*mutSL* MA lines	Chapter III
JWS224	founder colony for Δ*rnhC* MA lines	*Yao et al.* 2013
JWS266	*lexA[G92D]*	This work
JWS267	*lexA::erm*	This work
JWS268	*rnhC::erm*, *lexA[G92D]*	This work
JWS288	Δ*rnhB*, Δ*mutSL*	This work
JWS292	founder colony for Δ*rnhB*, Δ*mutSL* MA lines	This work

Table A.3: RNA-seq results for $\triangle$*rnhB*.

Gene name	$\log_2$(fold change)	Probability of differential expression	Product
bcd	2.0	1.0	Leucine dehydrogenase
ptb	1.6	0.99	Probable phosphate butyryltransferase
rnhB	-6.4	0.99	Ribonuclease HII
ycgT	1.7	0.98	Ferredoxin--NADP reductase 1
walK	0.81	0.96	Sensor histidine kinase WalK
*r*dgB	0.85	0.95	Non-canonical purine NTP pyrophosphatase
yxkD	-0.98	0.93	UPF0750 membrane protein YxkD
yjqB	1.2	0.92	Putative cyclic phosphatase
motB	-1.0	0.92	Motility protein B
hemAT	-1.3	0.91	Heme-based aerotactic transducer HemAT
buk	1.0	0.90	Probable butyrate kinase

BIBLIOGRAPHY

BIBLIOGRAPHY

Acharya, S. (2008), Mutations in the signature motif in MutS affect ATP-induced clamp formation and mismatch repair, *Mol Microbiol*, *69*(6), 1544--59, doi:10.1111/j.1365-2958.2008.06386.x.

Acharya, S., P. L. Foster, P. Brooks, and R. Fishel (2003), The coordinated functions of the *E. coli* MutS and MutL proteins in mismatch repair, *Mol Cell*, *12*(1), 233--46.

Ahrends, R., J. Kosinski, D. Kirsch, L. Manelyte, L. Giron-Monzon, L. Hummerich, O. Schulz, B. Spengler, and P. Friedhoff (2006), Identifying an interaction site between MutH and the C-terminal domain of MutL by crosslinking, affinity purification, chemical coding and mass spectrometry., *Nucleic Acids Res*, *34*(10), 3169--3180, doi:10.1093/nar/gkl407.

Allen-Soltero, S., S. L. Martinez, C. D. Putnam, and R. D. Kolodner (2014), A *Saccharomyces cerevisiae* RNase H2 interaction network functions to suppress genome instability, *Mol Cell Biol*, *34*(8), 1521--34, doi:10.1128/MCB.00960-13.

Astatke, M., K. Ng, N. D. Grindley, and C. M. Joyce (1998), A single side chain prevents *Escherichia coli* DNA polymerase I (klenow fragment) from incorporating ribonucleotides, *Proc Natl Acad Sci U S A*, *95*(7), 3402--7.

Bailey, T. L., and C. Elkan (1994), Fitting a mixture model by expectation maximization to discover motifs in biopolymers., *Proc Int Conf Intell Syst Mol Biol*, *2*, 28--36.

Barrick, J. E., and R. E. Lenski (2013), Genome dynamics during experimental evolution., *Nat Rev Genet*, *14*(12), 827--839, doi:10.1038/nrg3564.

Beck, J., M. Vogel, and M. Nassal (2002), dNTP versus NTP discrimination by phenylalanine 451 in duck hepatitis B virus P protein indicates a common structure of the dNTP-binding pocket with other reverse transcriptases, *Nucleic Acids Res*, *30*(7), 1679--87.

Berkmen, M. B., and A. D. Grossman (2006), Spatial and temporal organization of the *Bacillus subtilis* replication cycle, *Mol Microbiol*, *62*(1), 57--71, doi:10.1111/j.1365-2958.2006.05356.x.

Berkmen, M. B., and A. D. Grossman (2007), Subcellular positioning of the origin region of the *Bacillus subtilis* chromosome is independent of sequences within *oriC*, the site of replication initiation, and the replication initiator DnaA., *Mol Microbiol*, *63*(1), 150--165, doi:10.1111/j.1365-2958.2006.05505.x.

Berkower, I., J. Leis, and J. Hurwitz (1973), Isolation and characterization of an endonuclease from *Escherichia coli* specific for ribonucleic acid in ribonucleic acid-deoxyribonucleic acid hybrid structures, *J Biol Chem*, *248*(17), 5914--21.

Betzig, E., G. H. Patterson, R. Sougrat, O. W. Lindwasser, S. Olenych, J. S. Bonifacino, M. W. Davidson, J. Lippincott-Schwartz, and H. F. Hess (2006), Imaging intracellular fluorescent proteins at nanometer resolution, *Science*, *313*(5793), 1642--1645.

Biteen, J. S., and W. E. Moerner (2010), Single-molecule and superresolution imaging in live bacteria cells, *Cold Spring Harb Perspect Biol*, *2*(3), a000,448, doi:10.1101/cshperspect.a000448.

Biteen, J. S., M. A. Thompson, N. K. Tselentis, G. R. Bowman, L. Shapiro, and W. Moerner (2008), Super-resolution imaging in live *Caulobacter crescentus* cells using photoswitchable eyfp, *Nature methods*, *5*(11), 947--949.

Bonnin, A., J. M. Lazaro, L. Blanco, and M. Salas (1999), A single tyrosine prevents insertion of ribonucleotides in the eukaryotic-type phi29 DNA polymerase, *Journal of Molecular Biology*, *290*(1), 241--51, doi:10.1006/jmbi.1999.2900.

Braithwaite, D. K., and J. Ito (1993), Compilation, alignment, and phylogenetic relationships of DNA polymerases, *Nucleic Acids Res*, *21*(4), 787--802.

Breier, A. M., and A. D. Grossman (2009), Dynamic association of the replication initiator and transcription factor DnaA with the *Bacillus subtilis* chromosome during replication stress., *J Bacteriol*, *191*(2), 486--493, doi:10.1128/JB.01294-08.

Brown, J. A., and Z. Suo (2011), Unlocking the sugar "steric gate" of DNA polymerases, *Biochemistry*, *50*(7), 1135--42, doi:10.1021/bi101915z.

Brown, J. A., K. A. Fiala, J. D. Fowler, S. M. Sherrer, S. A. Newmister, W. W. Duym, and Z. Suo (2010), A novel mechanism of sugar selection utilized by a human X-family DNA polymerase, *J Mol Biol*, *395*(2), 282--90, doi:10.1016/j.jmb.2009.11.003.

Bunting, K. A., S. M. Roe, and L. H. Pearl (2003), Structural basis for recruitment of translesion DNA polymerase Pol IV/DinB to the beta-clamp., *EMBO J*, *22*(21), 5883--5892, doi:10.1093/emboj/cdg568.

Burnett, L., and R. G. Wake (1977), Initiation and termination of chromosome replication at 45 degree C in a temperature-sensitive deoxyribonucleic acid initiation mutant of *Bacillus subtilis* 168, TsB134, *J Bacteriol*, *130*(1), 538--9.

Burnette, D. T., P. Sengupta, Y. Dai, J. Lippincott-Schwartz, and B. Kachar (2011), Bleaching/blinking assisted localization microscopy for superresolution imaging using standard fluorescent molecules, *Proc Natl Acad Sci U S A*, *108*(52), 21,081--6, doi:10.1073/pnas.1117430109.

Cai, Y., N. E. Geacintov, and S. Broyde (2014), Ribonucleotides as nucleotide excision repair substrates, *DNA Repair (Amst)*, *13*, 55--60, doi:10.1016/j.dnarep.2013.10.010.

Cairns, J., J. Overbaugh, and S. Miller (1988), The origin of mutants., *Nature*, *335*(6186), 142--145, doi:10.1038/335142a0.

Cases-Gonzalez, C. E., M. Gutierrez-Rivas, and L. Menendez-Arias (2000), Coupling ribose selection to fidelity of DNA synthesis. the role of tyr-115 of human immunodeficiency virus type 1 reverse transcriptase, *The Journal of Biological Chemistry*, *275*(26), 19,759--67, doi:10.1074/jbc.M910361199.

Cho, W.-K., C. Jeong, D. Kim, M. Chang, K.-M. Song, J. Hanne, C. Ban, R. Fishel, and J.-B. Lee (2012), ATP alters the diffusion mechanics of MutS on mismatched DNA., *Structure*, *20*(7), 1264--1274, doi:10.1016/j.str.2012.04.017.

Chon, H., H. Matsumura, Y. Koga, K. Takano, and S. Kanaya (2006), Crystal structure and structure-based mutational analyses of RNase HIII from *Bacillus stearothermophilus*: a new type 2 RNase H with TBP-like substrate-binding domain at the N terminus, *J Mol Biol*, *356*(1), 165--78, doi:10.1016/j.jmb.2005.11.017.

Crow, Y. J., et al. (2006a), Mutations in genes encoding ribonuclease h2 subunits cause aicardi-goutieres syndrome and mimic congenital viral brain infection, *Nature genetics*, *38*(8), 910--6, doi:10.1038/ng1842.

Crow, Y. J., et al. (2006b), Mutations in the gene encoding the 3'-5' DNA exonuclease trex1 cause aicardi-goutieres syndrome at the ags1 locus, *Nat Genet*, *38*(8), 917--20, doi:10.1038/ng1845.

Culligan, K. M., G. Meyer-Gauen, J. Lyons-Weiler, and J. B. Hays (2000), Evolutionary origin, diversification and specialization of eukaryotic MutS homolog mismatch repair proteins., *Nucleic Acids Res*, *28*(2), 463--471.

Dalrymple, B. P., K. Kongsuwan, G. Wijffels, N. E. Dixon, and P. A. Jennings (2001), A universal protein-protein interaction motif in the eubacterial DNA replication and repair systems., *Proc Natl Acad Sci U S A*, *98*(20), 11,627--11,632, doi:10.1073/pnas.191384398.

Danecek, P., et al. (2011), The variant call format and vcftools, *Bioinformatics*, *27*(15), 2156--2158.

Das, U., A. K. Chakravarty, B. S. Remus, and S. Shuman (2013), Rewriting the rules for end joining via enzymatic splicing of DNA 3'-po4 and 5'-oh ends, *Proc Natl Acad Sci U S A*, *110*(51), 20,437--42, doi:10.1073/pnas.1314289110.

Das, U., M. Chauleau, H. Ordonez, and S. Shuman (2014), Impact of DNA3'pp5'g capping on repair reactions at DNA 3' ends, *Proc Natl Acad Sci U S A*, *111*(31), 11,317--22, doi:10.1073/pnas.1409203111.

DeLucia, A. M., N. D. Grindley, and C. M. Joyce (2003), An error-prone family y DNA polymerase (dinb homolog from sulfolobus solfataricus) uses a 'steric gate' residue for discrimination against ribonucleotides, *Nucleic Acids Research*, *31*(14), 4129--37.

DeRocco, V. C., L. E. Sass, R. Qiu, K. R. Weninger, and D. A. Erie (2014), Dynamics of MutS-mismatched DNA complexes are predictive of their repair phenotypes, *Biochemistry*, *53*(12), 2043--52, doi:10.1021/bi401429b.

DeRose, E. F., L. Perera, M. S. Murray, T. A. Kunkel, and R. E. London (2012), Solution structure of the dickerson DNA dodecamer containing a single ribonucleotide, *Biochemistry*, *51*(12), 2407--16, doi:10.1021/bi201710q.

Dervyn, E., C. Suski, R. Daniel, C. Bruand, J. Chapuis, J. Errington, L. Janniere, and S. D. Ehrlich (2001), Two essential DNA polymerases at the bacterial replication fork, *Science*, *294*(5547), 1716--9, doi:10.1126/science.1066351.

Dreiseikelmann, B., and W. Wackernagel (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*(1), 259--261.

Dupes, N. M., B. W. Walsh, A. D. Klocko, J. S. Lenhart, H. L. Peterson, D. A. Gessert, C. E. Pavlick, and L. A. Simmons (2010), Mutations in the *Bacillus subtilis* beta clamp that separate its roles in DNA replication from mismatch repair., *J Bacteriol*, *192*(13), 3452--3463, doi:10.1128/JB.01435-09.

Durisic, N., L. Laparra-Cuervo, A. Sandoval-Alvarez, J. S. Borbely, and M. Lakadamyali (2014), Single-molecule evaluation of fluorescent protein photoactivation efficiency using an in vivo nanotemplate, *Nat Methods*, *11*(2), 156--62, doi:10.1038/nmeth.2784.

Eisen, J. A. (1998), A phylogenomic study of the MutS family of proteins., *Nucleic Acids Res*, *26*(18), 4291--4300.

Eisen, J. A., and P. C. Hanawalt (1999), A phylogenomic study of DNA repair genes, proteins, and processes., *Mutat Res*, *435*(3), 171--213.

Elf, J., G. W. Li, and X. S. Xie (2007), Probing transcription factor dynamics at the single-molecule level in a living cell, *Science*, *316*(5828), 1191--4, doi:10.1126/science.1141967.

Elmore, S., M. Muller, N. Vischer, T. Odijk, and C. L. Woldringh (2005), Single-particle tracking of oric-gfp fluorescent spots during chromosome segregation in *Escherichia coli*, *J Struct Biol*, *151*(3), 275--87, doi:10.1016/j.jsb.2005.06.004.

Englert, M., et al. (2012), Structural and mechanistic insights into guanylylation of RNA-splicing ligase rtcb joining RNA between 3'-terminal phosphate and 5'-oh, *Proc Natl Acad Sci U S A*, *109*(38), 15,235--40, doi:10.1073/pnas.1213795109.

Evans, R. J., et al. (2008), Structure of PolC reveals unique DNA binding and fidelity determinants., *Proc Natl Acad Sci U S A*, *105*(52), 20,695--20,700, doi:10.1073/pnas.0809989106.

Figiel, M., and M. Nowotny (2014), Crystal structure of RNase H3-substrate complex reveals parallel evolution of RNA/DNA hybrid recognition, *Nucleic Acids Res*, *42*(14), 9285--94, doi:10.1093/nar/gku615.

Figiel, M., H. Chon, S. M. Cerritelli, M. Cybulska, R. J. Crouch, and M. Nowotny (2011), The structural and biochemical characterization of human RNase H2 complex reveals the molecular basis for substrate recognition and aicardi-goutieres syndrome defects, *J Biol Chem*, *286*(12), 10,540--50, doi:10.1074/jbc.M110.181974.

Fijalkowska, I. J., P. Jonczyk, M. M. Tkaczyk, M. Bialoskorska, and R. M. Schaaper (1998), Unequal fidelity of leading strand and lagging strand DNA replication on the *Escherichia coli* chromosome., *Proc Natl Acad Sci U S A*, *95*(17), 10,020--10,025.

Fishel, R., M. K. Lescoe, M. R. Rao, N. G. Copeland, N. A. Jenkins, J. Garber, M. Kane, and R. Kolodner (1993), The human mutator gene homolog msh2 and its association with hereditary nonpolyposis colon cancer, *Cell*, *75*(5), 1027--38.

Foster, P. L. (2007), Stress-induced mutagenesis in bacteria., *Crit Rev Biochem Mol Biol*, *42*(5), 373--397, doi:10.1080/10409230701648494.

Foster, P. L., A. J. Hanson, H. Lee, E. M. Popodi, and H. Tang (2013), On the mutational topology of the bacterial genome., *G3 (Bethesda)*, *3*(3), 399--407, doi:10.1534/g3.112.005355.

Gao, G., M. Orlova, M. M. Georgiadis, W. A. Hendrickson, and S. P. Goff (1997), Conferring RNA polymerase activity to a DNA polymerase: a single residue in reverse transcriptase controls substrate selection, *Proc Natl Acad Sci U S A*, *94*(2), 407--11.

Garcia-Diaz, M., K. Bebenek, J. M. Krahn, T. A. Kunkel, and L. C. Pedersen (2005), A closed conformation for the pol lambda catalytic cycle, *Nat Struct Mol Biol*, *12*(1), 97--8, doi:10.1038/nsmb876.

Gardner, A. F., and W. E. Jack (1999), Determinants of nucleotide sugar recognition in an archaeon DNA polymerase, *Nucleic Acids Res*, *27*(12), 2545--53.

Garrison, E., and G. Marth (2012), Haplotype-based variant detection from short-read sequencing, *arXiv:1207.3907*.

Gaur, V., R. Vyas, J. D. Fowler, G. Efthimiopoulos, J. Y. Feng, and Z. Suo (2014), Structural and kinetic insights into binding and incorporation of l-nucleotide analogs by a y-family DNA polymerase, *Nucleic Acids Res*, *42*(15), 9984--95, doi:10.1093/nar/gku709.

Gebhardt, J. C., D. M. Suter, R. Roy, Z. W. Zhao, A. R. Chapman, S. Basu, T. Maniatis, and X. S. Xie (2013), Single-molecule imaging of transcription factor binding to DNA in live mammalian cells, *Nat Methods*, *10*(5), 421--6, doi:10.1038/nmeth.2411.

Ghodgaonkar, M. M., et al. (2013), Ribonucleotides misincorporated into DNA act as strand-discrimination signals in eukaryotic mismatch repair., *Mol Cell*, *50*(3), 323--332, doi:10.1016/j.molcel.2013.03.019.

Ginetti, F., M. Perego, A. M. Albertini, and A. Galizzi (1996), *Bacillus subtilis* mutS mutL operon: identification, nucleotide sequence and mutagenesis., *Microbiology*, *142 (Pt 8)*, 2021--2029.

Goodman, M. F., R. Hopkins, and W. C. Gore (1977), 2-aminopurine-induced mutagenesis in t4 bacteriophage: a model relating mutation frequency to 2-aminopurine incorporation in DNA, *Proc Natl Acad Sci U S A*, *74*(11), 4806--10.

Goranov, A. I., E. Kuester-Schoeck, J. D. Wang, and A. D. Grossman (2006), Characterization of the global transcriptional responses to different types of DNA damage and disruption of replication in *Bacillus subtilis*., *J Bacteriol*, *188*(15), 5595--5605, doi:10.1128/JB.00342-06.

Gorman, J., A. Chowdhury, J. A. Surtees, J. Shimada, D. R. Reichman, E. Alani, and E. C. Greene (2007), Dynamic basis for one-dimensional DNA scanning by the mismatch repair complex msh2-msh6, *Mol Cell*, *28*(3), 359--70, doi:10.1016/j.molcel.2007.09.008.

Gorman, J., A. J. Plys, M. L. Visnapuu, E. Alani, and E. C. Greene (2010), Visualizing one-dimensional diffusion of eukaryotic DNA repair factors along a chromatin lattice, *Nat Struct Mol Biol*, *17*(8), 932--8, doi:10.1038/nsmb.1858.

Gorman, J., F. Wang, S. Redding, A. J. Plys, T. Fazio, S. Wind, E. E. Alani, and E. C. Greene (2012), Single-molecule imaging reveals target-search mechanisms during DNA mismatch repair, *Proc Natl Acad Sci U S A*, *109*(45), E3074--83, doi:10.1073/pnas.1211364109.

Haas, B. L., J. S. Matson, V. J. DiRita, and J. S. Biteen (2015), Single-molecule tracking in live vibrio cholerae reveals that toxr recruits the membrane-bound virulence regulator tcpp to the toxt promoter, *Mol Microbiol*, *96*(1), 4--13, doi:10.1111/mmi.12834.

Hall, B. M., C.-X. Ma, P. Liang, and K. K. Singh (2009), Fluctuation analysis calculator: a web tool for the determination of mutation rate using luria-delbruck fluctuation analysis., *Bioinformatics*, *25*(12), 1564--1565, doi:10.1093/bioinformatics/btp253.

Hall, M. C., and S. W. Matson (1999), The *Escherichia coli* MutL protein physically interacts with MutH and stimulates the MutH-associated endonuclease activity., *J Biol Chem*, *274*(3), 1306--1312.

Hess, S. T., T. P. K. Girirajan, and M. D. Mason (2006), Ultra-high resolution imaging by fluorescence photoactivation localization microscopy., *Biophys J*, *91*(11), 4258--4272, doi:10.1529/biophysj.106.091116.

Hogrefe, H. H., R. I. Hogrefe, R. Y. Walder, and J. A. Walder (1990), Kinetic analysis of *Escherichia coli* RNase H using DNA-RNA-DNA/DNA substrates, *J Biol Chem*, *265*(10), 5561--6.

Holden, S. J., S. Uphoff, and A. N. Kapanidis (2011), DaoSTORM: an algorithm for high-density super-resolution microscopy, *Nat Methods*, *8*(4), 279--80, doi:10.1038/nmeth0411-279.

Holmes, J., J., S. Clark, and P. Modrich (1990), Strand-specific mismatch correction in nuclear extracts of human and *Drosophila melanogaster* cell lines, *Proc Natl Acad Sci U S A*, *87*(15), 5837--41.

Hombauer, H., C. S. Campbell, C. E. Smith, A. Desai, and R. D. Kolodner (2011a), Visualization of eukaryotic DNA mismatch repair reveals distinct recognition and repair intermediates, *Cell*, *147*(5), 1040--53, doi:10.1016/j.cell.2011.10.025.

Hombauer, H., A. Srivatsan, C. D. Putnam, and R. D. Kolodner (2011b), Mismatch repair, but not heteroduplex rejection, is temporally coupled to DNA replication, *Science*, *334*(6063), 1713--6, doi:10.1126/science.1210770.

Hong, X., and T. Kogoma (1993), Absence of a direct role for RNase HI in initiation of DNA replication at the oric site on the *Escherichia coli* chromosome, *J Bacteriol*, *175*(20), 6731--4.

Hong, X., G. W. Cadwell, and T. Kogoma (1995), *Escherichia coli* recg and reca proteins in r-loop formation, *EMBO J*, *14*(10), 2385--92.

Hong, X., G. W. Cadwell, and T. Kogoma (1996), Activation of stable DNA replication in rapidly growing *Escherichia coli* at the time of entry to stationary phase, *Mol Microbiol*, *21*(5), 953--61.

Imai, Y., N. Ogasawara, D. Ishigo-Oka, R. Kadoya, T. Daito, and S. Moriya (2000), Subcellular localization of dna-initiation proteins of *Bacillus subtilis*: evidence that chromosome replication begins at either edge of the nucleoids, *Mol Microbiol*, *36*(5), 1037--48.

Itaya, M., A. Omori, S. Kanaya, R. J. Crouch, T. Tanaka, and K. Kondo (1999), Isolation of RNase H genes that are essential for growth of *Bacillus subtilis* 168, *Journal of Bacteriology*, *181*(7), 2118--23.

Ivanova, D., et al. (2015), Shaping the landscape of the escherichia coli chromosome: replication-transcription encounters in cells with an ectopic replication origin., *Nucleic Acids Res*, doi:10.1093/nar/gkv704.

Iyer, R. R., A. Pluciennik, V. Burdett, and P. L. Modrich (2006), DNA mismatch repair: functions and mechanisms, *Chem Rev*, *106*(2), 302--23.

Javaid, S., M. Manohar, N. Punja, A. Mooney, J. J. Ottesen, M. G. Poirier, and R. Fishel (2009), Nucleosome remodeling by hmsh2-hmsh6, *Mol Cell*, *36*(6), 1086--94, doi:10.1016/j.molcel.2009.12.010.

Jeong, C., W.-K. Cho, K.-M. Song, C. Cook, T.-Y. Yoon, C. Ban, R. Fishel, and J.-B. Lee (2011), MutS switches between two fundamentally distinct clamps during mismatch repair., *Nat Struct Mol Biol*, *18*(3), 379--385, doi:10.1038/nsmb.2009.

Jiricny, J. (2013), Postreplicative mismatch repair., *Cold Spring Harb Perspect Biol*, *5*(4), a012,633, doi:10.1101/cshperspect.a012633.

Johnson, A., and M. O'Donnell (2005), Cellular DNA replicases: components and dynamics at the replication fork, *Annu Rev Biochem*, *74*, 283--315.

Johnson, R. E., R. Klassen, L. Prakash, and S. Prakash (2015), A major role of DNA polymerase δ in replication of both the leading and lagging DNA strands., *Mol Cell*, *59*(2), 163--175, doi:10.1016/j.molcel.2015.05.038.

Jongruja, N., D. J. You, C. Angkawidjaja, E. Kanaya, Y. Koga, and S. Kanaya (2012), Structure and characterization of RNase H3 from aquifex aeolicus, *FEBS J*, *279*(15), 2737--53, doi:10.1111/j.1742-4658.2012.08657.x.

Joyce, C. M. (1997), Choosing the right sugar: how polymerases select a nucleotide substrate, *Proc Natl Acad Sci U S A*, *94*(5), 1619--22.

Joyce, C. M., O. Potapova, A. M. Delucia, X. Huang, V. P. Basu, and N. D. Grindley (2008), Fingers-closing and other rapid conformational changes in DNA polymerase I (klenow fragment) and their role in nucleotide selectivity, *Biochemistry*, *47*(23), 6103--16, doi:10.1021/bi7021848.

Junop, M. S., G. Obmolova, K. Rausch, P. Hsieh, and W. Yang (2001), Composite active site of an abc ATPase: MutS uses ATP to verify mismatch recognition and authorize DNA repair, *Mol Cell*, *7*(1), 1--12.

Junop, M. S., W. Yang, P. Funchain, W. Clendenin, and J. H. Miller (2003), In vitro and in vivo studies of MutS, MutL and muth mutants: correlation of mismatch repair and DNA recombination, *DNA Repair (Amst)*, *2*(4), 387--405.

Kadyrov, F. A., L. Dzantiev, N. Constantin, and P. Modrich (2006a), Endonucleolytic function of MutLalpha in human mismatch repair, *Cell*, *126*(2), 297--308, doi:10.1016/j.cell.2006.05.039.

Kadyrov, F. A., L. Dzantiev, N. Constantin, and P. Modrich (2006b), Endonucleolytic function of MutL alpha in human mismatch repair., *Cell*, *126*(2), 297--308, doi:10.1016/j.cell.2006.05.039.

Kanaya, S. (2001), Prokaryotic type 2 RNases H, *Methods Enzymol*, *341*, 377--94.

Karunatilaka, K. S., E. A. Cameron, E. C. Martens, N. M. Koropatkin, and J. S. Biteen (2014), Superresolution imaging captures carbohydrate utilization dynamics in human gut symbionts, *MBio*, *5*(6), doi:10.1128/mBio.02172-14.

Kasiviswanathan, R., and W. C. Copeland (2011), Ribonucleotide discrimination and reverse transcription by the human mitochondrial DNA polymerase, *J Biol Chem*, *286*(36), 31,490--500, doi:10.1074/jbc.M111.252460.

Kharchenko, P. V., M. Y. Tolstorukov, and P. J. Park (2008), Design and analysis of chip-seq experiments for DNA-binding proteins., *Nat Biotechnol*, *26*(12), 1351--1359, doi:10.1038/nbt.1508.

Kim, N., S. N. Huang, J. S. Williams, Y. C. Li, A. B. Clark, J. E. Cho, T. A. Kunkel, Y. Pommier, and S. Jinks-Robertson (2011), Mutagenic processing of ribonucleotides in DNA by yeast topoisomerase i, *Science*, *332*(6037), 1561--4, doi:10.1126/science.1205016.

Kleczkowska, H. E., G. Marra, T. Lettieri, and J. Jiricny (2001), hMSH3 and hMSH6 interact with PCNA and colocalize with it to replication foci., *Genes Dev*, *15*(6), 724--736, doi:10.1101/gad.191201.

Klocko, A. D., J. W. Schroeder, B. W. Walsh, J. S. Lenhart, M. L. Evans, and L. A. Simmons (2011), Mismatch repair causes the dynamic release of an essential DNA polymerase from the replication fork., *Mol Microbiol*, *82*(3), 648--663, doi:10.1111/j.1365-2958.2011.07841.x.

Kogoma, T. (1997), Stable DNA replication: interplay between DNA replication, homologous recombination, and transcription, *Microbiol. Mol. Biol. Rev.*, *61*(2), 212--238.

Kogoma, T., and K. von Meyenburg (1983), The origin of replication, oric, and the dnaa protein are dispensable in stable DNA replication (sdra) mutants of *Escherichia coli* k-12, *EMBO J*, *2*(3), 463--8.

Kogoma, T., X. Hong, G. W. Cadwell, K. G. Barnard, and T. Asai (1993), Requirement of homologous recombination functions for viability of the *Escherichia coli* cell that lacks RNase HI and exonuclease v activities, *Biochimie*, *75*(1-2), 89--99.

Kunkel, T. A., and D. A. Erie (2005), DNA mismatch repair., *Annu Rev Biochem*, *74*, 681--710, doi:10.1146/annurev.biochem.74.082803.133243.

Kunkel, T. A., R. M. Schaaper, R. A. Beckman, and L. A. Loeb (1981), On the fidelity of DNA replication. effect of the next nucleotide on proofreading., *J Biol Chem*, *256*(19), 9883--9889.

Lacks, S. A., J. J. Dunn, and B. Greenberg (1982), Identification of base mismatches recognized by the heteroduplex-DNA-repair system of streptococcus pneumoniae, *Cell*, *31*(2 Pt 1), 327--36.

Lahue, R. S., K. G. Au, and P. Modrich (1989), DNA mismatch correction in a defined system, *Science*, *245*(4914), 160--4.

Lakadamyali, M., M. J. Rust, H. P. Babcock, and X. Zhuang (2003), Visualizing infection of individual influenza viruses., *Proc Natl Acad Sci U S A*, *100*(16), 9280--9285, doi:10.1073/pnas.0832269100.

Larrea, A. A., S. A. Lujan, and T. A. Kunkel (2010), SnapShot: DNA mismatch repair., *Cell*, *141*(4), 730.e1, doi:10.1016/j.cell.2010.05.002.

Law, C. W., Y. Chen, W. Shi, and G. K. Smyth (2014), voom: Precision weights unlock linear model analysis tools for RNA-seq read counts, *Genome Biol*, *15*(2), R29, doi:10.1186/gb-2014-15-2-r29.

Lawrence, M., W. Huber, H. Pagès, P. Aboyoun, M. Carlson, R. Gentleman, M. T. Morgan, and V. J. Carey (2013), Software for computing and annotating genomic ranges., *PLoS Comput Biol*, *9*(8), e1003,118, doi:10.1371/journal.pcbi.1003118.

Lee, H., E. Popodi, H. Tang, and P. L. Foster (2012), Rate and molecular spectrum of spontaneous mutations in the bacterium *Escherichia coli* as determined by whole-genome sequencing., *Proc Natl Acad Sci U S A*, *109*(41), E2774--E2783, doi:10.1073/pnas.1210309109.

Lemon, K. P., and A. D. Grossman (1998), Localization of bacterial DNA polymerase: evidence for a factory model of replication., *Science*, *282*(5393), 1516--1519.

Lenhart, J. S., J. W. Schroeder, B. W. Walsh, and L. A. Simmons (2012), DNA repair and genome maintenance in *Bacillus subtilis*., *Microbiol Mol Biol Rev*, *76*(3), 530--564, doi:10.1128/MMBR.05020-11.

Lenhart, J. S., M. C. Pillon, A. Guarné, and L. A. Simmons (2013a), Trapping and visualizing intermediate steps in the mismatch repair pathway in vivo., *Mol Microbiol*, *90*(4), 680--698, doi:10.1111/mmi.12389.

Lenhart, J. S., A. Sharma, M. M. Hingorani, and L. A. Simmons (2013b), Dnan clamp zones provide a platform for spatiotemporal coupling of mismatch detection to DNA replication., *Mol Microbiol*, *87*(3), 553--568, doi:10.1111/mmi.12115.

Lewis, J. (1995), Fast normalized cross-correlation, *Vision interface*, *10*(1), 120--123.

Li, H., and R. Durbin (2009), Fast and accurate short read alignment with burrows-wheeler transform., *Bioinformatics*, *25*(14), 1754--1760, doi:10.1093/bioinformatics/btp324.

Li, H., et al. (2009), The sequence alignment/map format and samtools., *Bioinformatics*, *25*(16), 2078--2079.

Lind, P. A., and D. I. Andersson (2008), Whole-genome mutational biases in bacteria., *Proc Natl Acad Sci U S A*, *105*(46), 17,878--17,883, doi:10.1073/pnas.0804445105.

Long, H., W. Sung, S. F. Miller, M. S. Ackerman, T. G. Doak, and M. Lynch (2015), Mutation rate, spectrum, topology, and context-dependency in the DNA mismatch repair-deficient *Pseudomonas fluorescens* ATCC948., *Genome Biol Evol*, *7*(1), 262--271, doi:10.1093/gbe/evu284.

López de Saro, F. J., and M. O'Donnell (2001), Interaction of the beta sliding clamp with MutS, ligase, and DNA polymerase I., *Proc Natl Acad Sci U S A*, *98*(15), 8376--8380, doi:10.1073/pnas.121009498.

López de Saro, F. J., R. E. Georgescu, and M. O'Donnell (2003), A peptide switch regulates DNA polymerase processivity., *Proc Natl Acad Sci U S A*, *100*(25), 14,689--14,694, doi:10.1073/pnas.2435454100.

López de Saro, F. J., M. G. Marinus, P. Modrich, and M. O'Donnell (2006), The beta sliding clamp binds to multiple sites within MutL and MutS., *J Biol Chem*, *281*(20), 14,340--14,349, doi:10.1074/jbc.M601264200.

Lu, Z., J. Hou, Y. Wang, and J. Liu (2012a), Involvement of ser94 in RNase HIII from chlamydophila pneumoniae in the recognition of a single ribonucleotide misincorporated into double-stranded DNA, *Biochim Biophys Acta*, *1824*(7), 859--65, doi:10.1016/j.bbapap.2012.04.003.

Lu, Z., R. Liang, X. Liu, J. Hou, and J. Liu (2012b), RNase HIII from chlamydophila pneumoniae can efficiently cleave double-stranded DNA carrying a chimeric ribonucleotide in the presence of manganese, *Mol Microbiol*, *83*(5), 1080--93, doi:10.1111/j.1365-2958.2012.07990.x.

Lujan, S. A., J. S. Williams, A. R. Clausen, A. B. Clark, and T. A. Kunkel (2013), Ribonucleotides are signals for mismatch repair of leading-strand replication errors., *Mol Cell*, *50*(3), 437--443, doi:10.1016/j.molcel.2013.03.017.

Lujan, S. A., et al. (2014), Heterogeneous polymerase fidelity and mismatch repair bias genome variation and composition., *Genome Res*, *24*(11), 1751--1764, doi:10.1101/gr.178335.114.

Ma, X., M. V. Rogacheva, K. T. Nishant, S. Zanders, C. D. Bustamante, and E. Alani (2012), Mutation hot spots in yeast caused by long-range clustering of homopolymeric sequences., *Cell Rep*, *1*(1), 36--42, doi:10.1016/j.celrep.2011.10.003.

Maduike, N. Z., A. K. Tehranchi, J. D. Wang, and K. N. Kreuzer (2014), Replication of the *Escherichia coli* chromosome in RNase HI-deficient cells: multiple initiation regions and fork dynamics, *Mol Microbiol*, *91*(1), 39--56, doi:10.1111/mmi.12440.

Manley, S., J. M. Gillette, G. H. Patterson, H. Shroff, H. F. Hess, E. Betzig, and J. Lippincott-Schwartz (2008), High-density mapping of single-molecule trajectories with photoactivated localization microscopy, *Nat Methods*, *5*(2), 155--7, doi:10.1038/nmeth.1176.

McDonald, J. P., A. Vaisman, W. Kuban, M. F. Goodman, and R. Woodgate (2012), Mechanisms employed by *Escherichia coli* to prevent ribonucleotide incorporation into genomic DNA by pol v, *PLoS Genet*, *8*(11), e1003,030, doi:10.1371/journal.pgen.1003030.

McHenry, C. S. (2011), Breaking the rules: bacteria that use several DNA polymerase iiis, *EMBO Rep*, *12*(5), 408--14.

Meile, J.-C., L. J. Wu, S. D. Ehrlich, J. Errington, and P. Noirot (2006), Systematic localisation of proteins fused to the green fluorescent protein in *Bacillus subtilis*: identification of new proteins at the DNA replication factory., *Proteomics*, *6*(7), 2135--2146, doi:10.1002/pmic.200500512.

Miller, H. I., A. D. Riggs, and G. N. Gill (1973), Ribonuclease h (hybrid) in *Escherichia coli*. identification and characterization, *J Biol Chem*, *248*(7), 2621--4.

Million-Weaver, S., et al. (2015), An underlying mechanism for the increased mutagenesis of lagging-strand genes in *Bacillus subtilis*., *Proc Natl Acad Sci U S A*, *112*(10), E1096--E1105, doi:10.1073/pnas.1416651112.

Miyashita, S., T. Tadokoro, C. Angkawidjaja, D. J. You, Y. Koga, K. Takano, and S. Kanaya (2011), Identification of the substrate binding site in the n-terminal tbp-like domain of RNase H3, *FEBS Lett*, *585*(14), 2313--7, doi:10.1016/j.febslet.2011.05.064.

Nick McElhinny, S. A., and D. A. Ramsden (2003), Polymerase mu is a DNA-directed DNA/RNA polymerase, *Mol Cell Biol*, *23*(7), 2309--15.

Nick McElhinny, S. A., D. Kumar, A. B. Clark, D. L. Watt, B. E. Watts, E. B. Lundstrom, E. Johansson, A. Chabes, and T. A. Kunkel (2010), Genome instability due to ribonucleotide incorporation into DNA, *Nature chemical biology*, *6*(10), 774--81, doi:10.1038/nchembio.424.

Nick McElhinny, S. A., B. E. Watts, D. Kumar, D. L. Watt, E.-B. Lundström, P. M. J. Burgers, E. Johansson, A. Chabes, and T. A. Kunkel (2010), Abundant ribonucleotide incorporation into DNA by yeast replicative polymerases., *Proc Natl Acad Sci U S A*, *107*(11), 4949--4954, doi:10.1073/pnas.0914857107.

Nowotny, M., S. A. Gaidamakov, R. J. Crouch, and W. Yang (2005), Crystal structures of RNase H bound to an RNA/DNA hybrid: substrate specificity and metal-dependent catalysis, *Cell*, *121*(7), 1005--16, doi:10.1016/j.cell.2005.04.024.

Nowotny, M., S. M. Cerritelli, R. Ghirlando, S. A. Gaidamakov, R. J. Crouch, and W. Yang (2008), Specific recognition of RNA/DNA hybrid and enhancement of human RNase H1 activity by hbd, *EMBO J*, *27*(7), 1172--81, doi:10.1038/emboj.2008.44.

Obenchain, V., M. Lawrence, V. Carey, S. Gogarten, P. Shannon, and M. Morgan (2014), VariantAnnotation: a Bioconductor package for exploration and annotation of genetic variants., *Bioinformatics*, *30*(14), 2076--2078, doi:10.1093/bioinformatics/btu168.

Obmolova, G., C. Ban, P. Hsieh, and W. Yang (2000), Crystal structures of mismatch repair protein MutS and its complex with a substrate DNA, *Nature*, *407*(6805), 703--10, doi:10.1038/35037509.

Ogawa, T., and T. Okazaki (1984), Function of RNase H in DNA replication revealed by RNase H defective mutants of *Escherichia coli*, *Mol Gen Genet*, *193*(2), 231--7.

Ogawa, T., G. G. Pickett, T. Kogoma, and A. Kornberg (1984), RNase H confers specificity in the dnaa-dependent initiation of replication at the unique origin of the *Escherichia coli* chromosome in vivo and in vitro, *Proc Natl Acad Sci U S A*, *81*(4), 1040--4.

Ohtani, N., M. Haruki, M. Morikawa, R. J. Crouch, M. Itaya, and S. Kanaya (1999a), 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*(2), 605--18, doi:10.1021/bi982207z.

Ohtani, N., M. Haruki, M. Morikawa, and S. Kanaya (1999b), Molecular diversities of rnases h, *J Biosci Bioeng*, *88*(1), 12--9.

Oivanen, M., S. Kuusela, and H. Lonnberg (1998), Kinetics and mechanisms for the cleavage and isomerization of the phosphodiester bonds of RNA by bronsted acids and bases, *Chem Rev*, *98*(3), 961--990.

Patel, P. H., and L. A. Loeb (2000), Multiple amino acid substitutions allow DNA polymerases to synthesize RNA, *J Biol Chem*, *275*(51), 40,266--72, doi:10.1074/jbc.M005757200.

Patrick, J. E., and D. B. Kearns (2008), MinJ (YvjD) is a topological determinant of cell division in *Bacillus subtilis*., *Mol Microbiol*, *70*(5), 1166--1179, doi:10.1111/j.1365-2958.2008.06469.x.

Paul, S., S. Million-Weaver, S. Chattopadhyay, E. Sokurenko, and H. Merrikh (2013), Accelerated gene evolution through replication-transcription conflicts., *Nature*, *495*(7442), 512--515, doi:10.1038/nature11989.

Pavani, S. R., M. A. Thompson, J. S. Biteen, S. J. Lord, N. Liu, R. J. Twieg, R. Piestun, and W. E. Moerner (2009), Three-dimensional, single-molecule fluorescence imaging beyond the diffraction limit by using a double-helix point spread function, *Proc Natl Acad Sci U S A*, *106*(9), 2995--9, doi:10.1073/pnas.0900245106.

Pavlov, Y. I., I. M. Mian, and T. A. Kunkel (2003), Evidence for preferential mismatch repair of lagging strand DNA replication errors in yeast, *Curr Biol*, *13*(9), 744--8.

Pelletier, H., M. R. Sawaya, A. Kumar, S. H. Wilson, and J. Kraut (1994), Structures of ternary complexes of rat DNA polymerase beta, a DNA template-primer, and ddctp, *Science*, *264*(5167), 1891--903.

Petruska, J., and M. F. Goodman (1985), Influence of neighboring bases on DNA polymerase insertion and proofreading fidelity., *J Biol Chem*, *260*(12), 7533--7539.

Pillon, M. C., J. H. Miller, and A. Guarné (2011), The endonuclease domain of MutL interacts with the β sliding clamp., *DNA Repair (Amst)*, *10*(1), 87--93, doi:10.1016/j.dnarep.2010.10.003.

Pillon, M. C., et al. (2010), Structure of the endonuclease domain of MutL: unlicensed to cut., *Mol Cell*, *39*(1), 145--151, doi:10.1016/j.molcel.2010.06.027.

Pluciennik, A., V. Burdett, O. Lukianova, M. O'Donnell, and P. Modrich (2009), Involvement of the beta clamp in methyl-directed mismatch repair in vitro., *J Biol Chem*, *284*(47), 32,782--32,791, doi:10.1074/jbc.M109.054528.

Popow, J., et al. (2011), Hspc117 is the essential subunit of a human tRNA splicing ligase complex, *Science*, *331*(6018), 760--4, doi:10.1126/science.1197847.

Potenski, C. J., H. Niu, P. Sung, and H. L. Klein (2014), Avoidance of ribonucleotide-induced mutations by RNase H2 and srs2-exo1 mechanisms, *Nature*, *511*(7508), 251--4, doi:10.1038/nature13292.

Pursell, Z. F., I. Isoz, E. B. Lundstrom, E. Johansson, and T. A. Kunkel (2007), Yeast DNA polymerase epsilon participates in leading-strand DNA replication, *Science*, *317*(5834), 127--30, doi:10.1126/science.1144067.

Qiu, R., V. C. DeRocco, C. Harris, A. Sharma, M. M. Hingorani, D. A. Erie, and K. R. Weninger (2012), Large conformational changes in MutS during DNA scanning, mismatch recognition and repair signalling, *EMBO J*, *31*(11), 2528--40, doi:10.1038/emboj.2012.95.

Reijns, M. A., and A. P. Jackson (2014), Ribonuclease h2 in health and disease, *Biochem Soc Trans*, *42*(4), 717--25, doi:10.1042/BST20140079.

Reijns, M. A. M., H. Kemp, J. Ding, S. M. de Procé, A. P. Jackson, and M. S. Taylor (2015), Lagging-strand replication shapes the mutational landscape of the genome., *Nature*, *518*(7540), 502--506, doi:10.1038/nature14183.

Reijns, M. A. M., et al. (2012), Enzymatic removal of ribonucleotides from DNA is essential for mammalian genome integrity and development., *Cell*, *149*(5), 1008--1022, doi:10.1016/j.cell.2012.04.011.

Reyes-Lamothe, R., C. Possoz, O. Danilova, and D. J. Sherratt (2008), Independent positioning and action of *Escherichia coli* replisomes in live cells., *Cell*, *133*(1), 90--102, doi:10.1016/j.cell.2008.01.044.

Reyes-Lamothe, R., D. J. Sherratt, and M. C. Leake (2010), Stoichiometry and architecture of active DNA replication machinery in *Escherichia coli*, *Science*, *328*(5977), 498--501, doi:10.1126/science.1185757.

Ronen, A. (1979), 2-aminopurine, *Mutat. Res.*, *75*, 1--47.

Rudolph, C. J., A. L. Upton, A. Stockum, C. A. Nieduszynski, and R. G. Lloyd (2013), Avoiding chromosome pathology when replication forks collide, *Nature*, *500*(7464), 608--11, doi:10.1038/nature12312.

Ruiz, J. F., R. Juarez, M. Garcia-Diaz, G. Terrados, A. J. Picher, S. Gonzalez-Barrera, A. R. Fernandez de Henestrosa, and L. Blanco (2003), Lack of sugar discrimination by human pol mu requires a single glycine residue, *Nucleic Acids Res*, *31*(15), 4441--9.

Rust, M. J., M. Bates, and X. Zhuang (2006), Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM)., *Nat Methods*, *3*(10), 793--795, doi:10.1038/nmeth929.

Rychlik, M. P., H. Chon, S. M. Cerritelli, P. Klimek, R. J. Crouch, and M. Nowotny (2010), Crystal structures of RNase H2 in complex with nucleic acid reveal the mechanism of RNA-DNA junction recognition and cleavage, *Mol Cell*, *40*(4), 658--70, doi:10.1016/j.molcel.2010.11.001.

Sanders, G. M., H. G. Dallmann, and C. S. McHenry (2010), Reconstitution of the *B. subtilis* replisome with 13 proteins including two distinct replicases, *Mol Cell*, *37*(2), 273--81.

Sass, L. E., C. Lanyi, K. Weninger, and D. A. Erie (2010), Single-molecule FRET TACKLE reveals highly dynamic mismatched DNA-MutS complexes, *Biochemistry*, *49*(14), 3174--90, doi:10.1021/bi901871u.

Sawaya, M. R., R. Prasad, S. H. Wilson, J. Kraut, and H. Pelletier (1997), Crystal structures of human DNA polymerase beta complexed with gapped and nicked DNA: evidence for an induced fit mechanism, *Biochemistry*, *36*(37), 11,205--15.

Saxton, M. J. (1997), Single-particle tracking: the distribution of diffusion coefficients, *Biophys J*, *72*(4), 1744--53, doi:10.1016/S0006-3495(97)78820-9.

Schaibley, V. M., et al. (2013), The influence of genomic context on mutation patterns in the human genome inferred from rare variants., *Genome Res*, *23*(12), 1974--1984, doi:10.1101/gr.154971.113.

Schneider, C. A., W. S. Rasband, and K. W. Eliceiri (2012), Nih image to ImageJ: 25 years of image analysis, *Nat Methods*, *9*(7), 671--5.

Schofield, M. J., and P. Hsieh (2003), DNA mismatch repair: molecular mechanisms and biological function., *Annu Rev Microbiol*, *57*, 579--608, doi:10.1146/annurev.micro.57.030502.090847.

Schroeder, J. W., and L. A. Simmons (2013), Complete genome sequence of *Bacillus subtilis* strain PY79., *Genome Announc*, *1*(6), doi:10.1128/genomeA.01085-13.

Schroeder, J. W., J. R. Randall, L. A. Matthews, and L. A. Simmons (2014), Ribonucleotides in bacterial DNA., *Crit Rev Biochem Mol Biol*, pp. 1--13, doi:10.3109/10409238.2014.981647.

Sekedat, M. D., D. Fenyo, R. S. Rogers, A. J. Tackett, J. D. Aitchison, and B. T. Chait (2010), Gins motion reveals replication fork progression is remarkably uniform throughout the yeast genome, *Mol Syst Biol*, *6*, 353, doi:10.1038/msb.2010.8.

Sekiguchi, J., and S. Shuman (1997), Site-specific ribonuclease activity of eukaryotic DNA topoisomerase i, *Mol Cell*, *1*(1), 89--97.

Sharma, A., C. Doucette, F. N. Biro, and M. M. Hingorani (2013), Slow conformational changes in MutS and DNA direct ordered transitions between mismatch search, recognition and signaling of DNA repair, *J Mol Biol*, *425*(22), 4192--205, doi:10.1016/j.jmb.2013.08.011.

Simmons, L. A., B. W. Davies, A. D. Grossman, and G. C. Walker (2008), Beta clamp directs localization of mismatch repair in *Bacillus subtilis*., *Mol Cell*, *29*(3), 291--301, doi:10.1016/j.molcel.2007.10.036.

Simonson, P. D., E. Rothenberg, and P. R. Selvin (2011), Single-molecule-based super-resolution images in the presence of multiple fluorophores, *Nano Lett*, *11*(11), 5090--6, doi:10.1021/nl203560r.

Sinha, N. K. (1987), Specificity and efficiency of editing of mismatches involved in the formation of base-substitution mutations by the 3'-5' exonuclease activity of phage T4 DNA polymerase., *Proc Natl Acad Sci U S A*, *84*(4), 915--919.

Sliusarenko, O., J. Heinritz, T. Emonet, and C. Jacobs-Wagner (2011), High-throughput, subpixel precision analysis of bacterial morphogenesis and intracellular spatio-temporal dynamics., *Mol Microbiol*, *80*(3), 612--627, doi:10.1111/j.1365-2958.2011.07579.x.

Smith, B. T., A. D. Grossman, and G. C. Walker (2001), Visualization of mismatch repair in bacterial cells., *Mol Cell*, *8*(6), 1197--1206.

Sparks, J. L., H. Chon, S. M. Cerritelli, T. A. Kunkel, E. Johansson, R. J. Crouch, and P. M. Burgers (2012), RNase H2-initiated ribonucleotide excision repair., *Mol Cell*, *47*(6), 980--986, doi:10.1016/j.molcel.2012.06.035.

Srivatsan, A., A. Tehranchi, D. M. MacAlpine, and J. D. Wang (2010), Co-orientation of replication and transcription preserves genome integrity., *PLoS Genet*, *6*(1), e1000,810, doi:10.1371/journal.pgen.1000810.

Strauss, B. S., R. Roberts, L. Francis, and P. Pouryazdanparast (2000), Role of the dinB gene product in spontaneous mutation in *Escherichia coli* with an impaired replicative polymerase., *J Bacteriol*, *182*(23), 6742--6750.

Subach, F. V., G. H. Patterson, S. Manley, J. M. Gillette, J. Lippincott-Schwartz, and V. V. Verkhusha (2009), Photoactivatable mcherry for high-resolution two-color fluorescence microscopy., *Nat Methods*, *6*(2), 153--159, doi:10.1038/nmeth.1298.

Sung, W., M. S. Ackerman, J.-F. Gout, S. F. Miller, E. Williams, P. L. Foster, and M. Lynch (2015), Asymmetric context-dependent mutation patterns revealed through mutation-accumulation experiments., *Mol Biol Evol*, doi:10.1093/molbev/msv055.

Tadokoro, T., and S. Kanaya (2009), Ribonuclease h: molecular diversities, substrate binding domains, and catalytic mechanism of the prokaryotic enzymes, *FEBS J*, *276*(6), 1482--93, doi:10.1111/j.1742-4658.2009.06907.x.

Tanaka, N., B. Meineke, and S. Shuman (2011), Rtcb, a novel RNA ligase, can catalyze tRNA splicing and hac1 mRNA splicing in vivo, *J Biol Chem*, *286*(35), 30,253--7, doi:10.1074/jbc.C111.274597.

Thomas, D. C., J. D. Roberts, and T. A. Kunkel (1991), Heteroduplex repair in extracts of human hela cells, *J Biol Chem*, *266*(6), 3744--51.

Tomimatsu, N., et al. (2014), Phosphorylation of exo1 by cdks 1 and 2 regulates DNA end resection and repair pathway choice, *Nat Commun*, *5*, 3561, doi:10.1038/ncomms4561.

Tran, H. T., D. A. Gordenin, and M. A. Resnick (1999), The 3'-->5' exonucleases of DNA polymerases delta and epsilon and the 5'-->3' exonuclease exo1 have major roles in postreplication mutation avoidance in *Saccharomyces cerevisiae*, *Mol Cell Biol*, *19*(3), 2000--7.

Tumbale, P., J. S. Williams, M. J. Schellenberg, T. A. Kunkel, and R. S. Williams (2014), Aprataxin resolves adenylated RNA-DNA junctions to maintain genome integrity, *Nature*, *506*(7486), 111--5, doi:10.1038/nature12824.

Uphoff, S., R. Reyes-Lamothe, F. Garza de Leon, D. J. Sherratt, and A. N. Kapanidis (2013), Single-molecule DNA repair in live bacteria, *Proc Natl Acad Sci U S A*, *110*(20), 8063--8, doi:10.1073/pnas.1301804110.

Vaisman, A., J. P. McDonald, D. Huston, W. Kuban, L. Liu, B. Van Houten, and R. Woodgate (2013), Removal of misincorporated ribonucleotides from prokaryotic genomes: an unexpected role for nucleotide excision repair, *PLoS Genet*, *9*(11), e1003,878, doi:10.1371/journal.pgen.1003878.

Vaisman, A., J. P. McDonald, S. Noll, D. Huston, G. Loeb, M. F. Goodman, and R. Woodgate (2014), Investigating the mechanisms of ribonucleotide excision repair in *Escherichia coli*, *Mutat Res Fundam Mol Mech Mutagen*, *761*, 21--33, doi:10.1016/j.mrfmmm.2014.01.005.

Viswanathan, M., V. Burdett, C. Baitinger, P. Modrich, and S. T. Lovett (2001), Redundant exonuclease involvement in *Escherichia coli* methyl-directed mismatch repair., *J Biol Chem*, *276*(33), 31,053--31,058, doi:10.1074/jbc.M105481200.

von Hippel, P. H., and O. G. Berg (1989), Facilitated target location in biological systems, *J Biol Chem*, *264*(2), 675--8.

Walsh, B. W., J. S. Lenhart, J. W. Schroeder, and L. A. Simmons (2012), Far western blotting as a rapid and efficient method for detecting interactions between DNA replication and DNA repair proteins., *Methods Mol Biol*, *922*, 161--168, doi:10.1007/978-1-62703-032-8_11.

Walsh, B. W., S. A. Bolz, S. R. Wessel, J. W. Schroeder, J. L. Keck, and L. A. Simmons (2014), Recd2 helicase limits replication fork stress in *Bacillus subtilis*., *J Bacteriol*, *196*(7), 1359--1368, doi:10.1128/JB.01475-13.

Wang, H., et al. (2003), DNA bending and unbending by MutS govern mismatch recognition and specificity, *Proc Natl Acad Sci U S A*, *100*(25), 14,822--7, doi:10.1073/pnas.2433654100.

Wang, J. D., G. M. Sanders, and A. D. Grossman (2007), Nutritional control of elongation of DNA replication by (p)ppgpp, *Cell*, *128*(5), 865--75.

Wang, S., J. R. Moffitt, G. T. Dempsey, X. S. Xie, and X. Zhuang (2014), Characterization and development of photoactivatable fluorescent proteins for single-molecule-based superresolution imaging, *Proc Natl Acad Sci U S A*, *111*(23), 8452--7, doi:10.1073/pnas.1406593111.

Wang, W., G.-W. Li, C. Chen, X. S. Xie, and X. Zhuang (2011), Chromosome organization by a nucleoid-associated protein in live bacteria., *Science*, *333*(6048), 1445--1449, doi:10.1126/science.1204697.

Watkins, L. P., and H. Yang (2005), Detection of intensity change points in time-resolved single-molecule measurements, *J Phys Chem B*, *109*(1), 617--28, doi:10.1021/jp0467548.

Wickham, H. (2009), *ggplot2: elegant graphics for data analysis*, Springer New York, 978-0-387-98140-6.

Wilkinson, A., J. Day, and R. Bowater (2001), Bacterial DNA ligases, *Mol Microbiol*, *40*(6), 1241--8.

Williams, J. S., and T. A. Kunkel (2014), Ribonucleotides in DNA: origins, repair and consequences, *DNA Repair (Amst)*, *19*, 27--37, doi:10.1016/j.dnarep.2014.03.029.

Williams, J. S., D. J. Smith, L. Marjavaara, S. A. Lujan, A. Chabes, and T. A. Kunkel (2013), Topoisomerase 1-mediated removal of ribonucleotides from nascent leading-strand DNA., *Mol Cell*, *49*(5), 1010--1015, doi:10.1016/j.molcel.2012.12.021.

Yang, G., M. Franklin, J. Li, T. C. Lin, and W. Konigsberg (2002), A conserved tyr residue is required for sugar selectivity in a pol alpha DNA polymerase, *Biochemistry*, *41*(32), 10,256--61.

Yang, H., and J. H. Miller (2008), Deletion of dnaN1 generates a mutator phenotype in *Bacillus anthracis*., *DNA Repair (Amst)*, *7*(3), 507--514, doi:10.1016/j.dnarep.2007.10.003.

Yang, H., C. Sikavi, K. Tran, S. M. McGillivray, V. Nizet, M. Yung, A. Chang, and J. H. Miller (2011a), Papillation in *Bacillus anthracis* colonies: a tool for finding new mutators., *Mol Microbiol*, *79*(5), 1276--1293, doi:10.1111/j.1365-2958.2011.07519.x.

Yang, H., M. Yung, C. Sikavi, and J. H. Miller (2011b), The role of *Bacillus anthracis* RecD2 helicase in DNA mismatch repair., *DNA Repair (Amst)*, *10*(11), 1121--1130, doi:10.1016/j.dnarep.2011.08.009.

Yang, H., M. Yung, L. Li, J. A. Hoch, C. M. Ryan, U. K. Kar, P. Souda, J. P. Whitelegge, and J. H. Miller (2013), Evidence that YycJ is a novel 5'-3' double-stranded DNA exonuclease acting in *Bacillus anthracis* mismatch repair., *DNA Repair (Amst)*, *12*(5), 334--346, doi:10.1016/j.dnarep.2013.02.002.

Yao, N. Y., J. W. Schroeder, O. Yurieva, L. A. Simmons, and M. E. O'Donnell (2013), Cost of rntp/dNTP pool imbalance at the replication fork., *Proc Natl Acad Sci U S A*, *110*(32), 12,942--12,947, doi:10.1073/pnas.1309506110.

Yildiz, A., J. N. Forkey, S. A. McKinney, T. Ha, Y. E. Goldman, and P. R. Selvin (2003), Myosin v walks hand-over-hand: single fluorophore imaging with 1.5-nm localization, *Science*, *300*(5628), 2061--5, doi:10.1126/science.1084398.

Youngman, P., J. B. Perkins, and R. Losick (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), 1--9.

Zhu, H., and S. Shuman (2005), A primer-dependent polymerase function of pseudomonas aeruginosa ATP-dependent DNA ligase (ligd), *J Biol Chem*, *280*(1), 418--27, doi:10.1074/jbc.M410110200.

Zhu, H., and S. Shuman (2008), Bacterial nonhomologous end joining ligases preferentially seal breaks with a 3'-oh monoribonucleotide, *J Biol Chem*, *283*(13), 8331--9, doi:10.1074/jbc.M705476200.

Zhu, L., W. Zhang, D. Elnatan, and B. Huang (2012), Faster STORM using compressed sensing, *Nat Methods*, *9*(7), 721--3, doi:10.1038/nmeth.1978.

Zhu, Y. O., M. L. Siegal, D. W. Hall, and D. A. Petrov (2014), Precise estimates of mutation rate and spectrum in yeast., *Proc Natl Acad Sci U S A*, *111*(22), E2310--E2318, doi:10.1073/pnas.1323011111.