The consequences of head-on replication-transcription conflicts on replication restart and genomic instability in *B. subtilis*

Samuel Lewis Million-Weaver

A dissertation submitted in partial fulfillment of requirements for the degree of

Doctor of Philosophy

University of Washington
2015

Reading Committee:
Houra Merrikh, Chair
Nina Salama
Joshua Woodward

Program authorized to offer the degree:
Department of Microbiology

University of Washington

Abstract

The consequences of head-on replication-transcription conflicts on replication restart and genomic instability in *B. subtilis*

Samuel Lewis Million-Weaver

Chair of the Supervisory Committee:
Assistant Professor Houra Merrikh
Department of Microbiology

Concurrent bacterial replication and transcription lead to conflicts between the two machineries. These encounters, which impede replication and destabilize genomes, are especially detrimental when the replisome and RNA polymerase encounter each other head-on, on the lagging strand. Despite the negative consequences of conflicts, 26% of genes remain in the lagging strand orientation in *Bacillus subtilis*. My thesis research aimed to uncover the coping mechanisms activated and consequences experienced by cells in the event of head-on collisions between replication and transcription. The work herein first describes a mechanism required for replication restart in the immediate aftermath of collisions with transcription in *B. subtilis*. Subsequent chapters investigate the role of conflicts in shaping the *B. subtilis* genome over evolutionary time. We found that lagging strand genes experienced increased mutation rates compared to those on the leading strand, and experimentally demonstrated that transcription asymmetrically increases mutation rates for head-on genes. I then identified a cellular factor, the Y-family polymerase PolY1, that is required for asymmetric mutagenesis at conflict regions. PolY1 acts to promote mutagenesis at head-on genes through participation in transcription-coupled nucleotide excision repair, suggesting that conflicts may locally increase the susceptibility of the DNA to bulky lesions. Consistent with this model, I determined that transcription asymmetrically promotes pyrimidine dimer formation in head-on oriented genes. Overall the work presented in this dissertation provides new insight into mechanisms of genome maintenance and stability in *B. subtilis*, as well as the diverse effects of lagging strand transcription on DNA replication.

Acknowledgements

This dissertation would have been impossible without the unending support, encouragement, and guidance from so many people. I am eternally grateful to my thesis advisor, Dr. Houra Merrikh for teaching me to think critically and creatively, to work tenaciously, and to tirelessly persevere in pursuit of answers. I cannot express my gratitude for the unending support, inspiration, and perspective provided by my parents and grandparents throughout this Ph.D. Finally, I thank my many colleagues in the microbiology department and the Merrikh lab as perpetual sources for scientific inspiration, motivation, or commiseration…sometimes all at the same time. In particular I wish to extend well-wishes to my dear friends and classmates Catherine Armbruster, Sam Carpentier, Paul Munson, and Katie Semmens—I hope that our tight-knit cadre remains in contact as we each go on to make our marks on the world.

"Curiosity, especially intellectual inquisitiveness is what separates the truly alive from those who are merely going through the motions"

-Tom Robbins

Table of Contents

List of figures and tables

Chapter 1

Chapter 2

Chapter3

Chapter 4

Dissertation overview

Concurrent replication and transcription lead to conflicts between the two machineries(1-5). Conflicts occur in one of two orientations: co-directionally for leading strand genes, or head-on when a gene is expressed from the lagging strand (1, 6-9). Conflicts destabilize genomes across all domains of life, necessitating resolution factors to preserve genomic integrity (10-12). The chromosomes of eukaryotic organisms contain large, non-coding intergenic regions and maintain a degree of temporal segregation between replication and transcription, minimizing the probability of collisions. Bacterial genomes, by contrast, are densely packed with actively expressed genes, which inevitably causes conflicts at a high frequency (1, 13-15).

The unfavorable effects of head-on collisions between replication and transcription are reflected in the organization of bacterial genomes. All sequenced bacterial genomes are biased for the majority of genes to be encoded on the leading strand, though the degree varies between clades(16). *Bacillus subtilis*, the model Gram-positive firmicute, encodes 74% of its genome on the leading strand of replication(17). Additionally, lagging strand genes in *B. subtilis* mutate at a higher rate than genes on the leading strand(17, 18). Head-on collisions between replication and transcription potentially cause genomic instability as well as fitness defects across a variety of organisms (1, 5, 19). However, despite the potential detrimental effects of the head-on orientation, genes remain on the lagging strand, raising the question: **how do cells manage the consequences of head-on replication-transcription conflicts?**

The work described in this thesis characterizes two mechanisms employed by bacteria to cope with the effects of lagging strand gene expression. The first chapter reveals a requirement for homologous recombination to restart replication after colliding with transcription. The second chapter uncovers an underlying mechanism for the increased mutagenesis of lagging strand genes in *B. subtilis*, namely, that the action of an error-prone, Y-family polymerase, PolY1, in transcription-coupled nucleotide excision repair increases genetic diversity as a consequence of expression from the lagging strand.

Chapter 1. Introduction

After providing an overview of the architecture of the prokaryotic replisome and the two types of replication-transcription conflicts that may occur, I outline results of seminal, though at times contradictory, experiments *in vitro* and *in vivo* to understand the consequences of collisions between replication and transcription. I then describe characterized mechanisms to minimize the negative effects of replication-transcription collisions. I conclude with a discussion of the potential role played by conflicts in shaping prokaryotic genomes over evolutionary time.

1.1 The architecture of the prokaryotic replication fork

The prototypical prokaryotic replisome is a multi-protein machine consisting of 11 distinct subunits in *Escherichia coli* (20-22), and 13 subunits in *B. subtilis* (23). The replisome completes chromosomal duplication within 40 minutes *in vivo* (24), suggesting that the complex synthesizes at least 1,000 base pairs of DNA per second. In line with this estimation, *in vitro* experiments typically measure replication rates on the order of 500-750 nucleotides per second (25, 26). The protein components of the replisome and their associated functions in chromosomal duplication are listed in **table 1.1**.

The 5'-3' polarity of DNA synthesis gives rise to a continuously synthesized leading strand of replication and a discontinuously replicated lagging strand (27). The lagging strand in bacteria is copied in 1-2 kb fragments (28), called Okazaki pieces. These fragments are joined together by DNA liagase (29). The molecule responsible for the bulk of chromosomal duplication is DNA Polymerase III (Pol III) (30). Recent studies of replisome stoichiometry in *E. coli* determined that three Pol III molecules are present at the replication fork (21): one presumably engaged in processive leading strand polymerization and two performing alternate cycles of Okazaki fragment synthesis on the lagging strand.

Polymerase III is encoded by a unique *dnaE* gene in *E. coli* (30, 31). The genomes of *B. subtilis,* and other Firmicutes, by contrast, harbor two distinct Polymerase III paralogues: *dnaE*, and *polC* (32, 33). The DnaE product lacks proofreading capabilities (ε) and is more active on the lagging strand of

replication (33). Genetic and *in vitro* evidence indicates that DnaE initiates replication from RNA primers and synthesizes roughly 1000 base pairs before disengaging from the DNA, where PolC subsequently takes over for processive elongation (23, 34).

The high processivity of the replicative DNA polymerases arises from engagement with a sliding clamp, which prevents premature dissociation from the template (30). Clamps are assembled onto DNA by the hetero-heptameric loader complex, made up of Ψ, x, δ, δ', γ and three τ subunits (20-22, 35, 36). The clamp-loader complex connects the synthetic core of the replisome to the replicative helicase (21, 23, 35, 36). The helicase, DnaC (DnaB in *E. coli*), is a homohexamer that translocates ahead of the fork on the lagging strand, unwinding the parental duplex using energy derived from ATP hydrolysis (37-39).

1.2 The two types of replication-transcription conflicts

Because the replisome and RNA polymerase synthesize and move 5' to 3' along the DNA, the template strand used for transcription determines the orientation of a gene with respect to replication. When a gene is encoded on the leading strand of replication, the machineries move co-directionally. When a gene is expressed from the lagging strand, the replicative helicase and RNA polymerase encounter each other moving in opposite directions on the same DNA strand.

Co-directional conflicts in bacteria modestly impede replication(8), likely due to the slow rate of transcription (50-100 nucleotides per second) (40, 41) as compared to replication (upwards of 1,000 base pairs per second) (20, 24). In Eukaryotic organisms, by contrast, replication is slower and proceeds at roughly the same rate as transcription (42), alleviating the severity of co-directional encounters (5). Head-on conflicts present significant obstacles to replication in all domains of life, causing high levels of replication stress and genomic instability (9, 43-45).

1.3 The consequences of collisions between replication and transcription

A) *in vitro*

Early *in vitro* studies reconstituted replication coming into conflict with either actively transcribing or arrested RNA polymerases (RNAP) using *E. coli* phage T4 and *B. subtilis* phage ϕ29

replisomes (46, 47). The two systems yielded conflicting results about the effects of conflicts in each orientation, and were further confounded by the fact that phage replisomes employ distinct helicases from the bacterial machinery, such as T4 Dda (48), which may themselves act as conflict-resolution factors.

Studies in the T4 system suggested that the replication machinery directly bypasses elongating RNAP in the co-directional orientation without interruption and that head-on encounters temporarily pause replisome progression (46, 49, 50). Experiments using ϕ29 demonstrated that head-on collisions with elongating RNAP have minimal effects, whereas co-directional encounters reduce the rate of replication by 50% (47, 51).

Later *in vitro* work using replication proteins from *E. coli* showed that head-on collisions with elongating RNAP arrest the replisome, although the fork resumes synthesis immediately after removal of the obstacle (52). Another group demonstrated that co-directional collisions with permanently backtracked RNAP cause DNA breaks (53). Overall, *in vitro* studies have provided conflicting results with regards to the effects of conflicts in either orientation. The apparent contradictions may arise because partial complements of replication proteins polymerizing on short DNA substrates cannot recapitulate the dynamic multi-component, topologically complex nature of replication forks *in vivo* (54).

B) *in vivo*

The first indications that replication-transcription conflicts carry detrimental consequences to cells arose from observations about the organization of bacterial genomes. Upon mapping the location of the rDNA operons in the *E. coli* chromosome, Nomura noted that their co-directional orientation avoided head-on replication-transcription conflicts (55). Bonita Brewer was the first to formally propose that the entire *E. coli* genome is arranged to minimize head-on collisions by maintaining highly expressed operons on the leading strand (56).

One of the historical challenges to studying replication-transcription conflicts *in vivo* is that relatively few lagging strand genes are highly expressed during laboratory growth conditions, making natural head-on conflicts difficult to detect (57-59). The first *in vivo* demonstration that head-on

conflicts with transcription impede replication placed an inducible origin of replication either directly upstream or downstream of the highly transcribed rRNA operons in *E. coli* cells, then monitored replication fork progression through the region using electron microscopy (60). This elegant experiment indicated that forks pass co-directionally through the rRNA genes seemingly without interruption, whereas head-on transcription significantly impairs replication progression.

Subsequent studies have again relied on head-on transcription from rRNA genes to investigate the effects of conflicts with replication. Strains of *B. subtilis* and *E. coli* harboring large chromosomal inversions to flip rRNA genes to the lagging strand of replication displayed replication stress, significant fitness defects, and DNA breaks (43, 44). Experiments in *E. coli* strains with inverted rRNA operons led to the identification of conflict resolution factors, such as accessory helicases (10) and RNA polymerase modulators (11) that promote replisome progression and cellular survival in response to this obstacle.

Although inverted rRNA operons create severe head-on conflicts due to their high expression levels these loci are highly dissimilar to protein-coding genes found on the lagging strand in natural bacterial genomes. The rRNA genes consist of highly repetitive operons averaging up to 5.5 kb in length (61), roughly 10-fold longer than the average coding lagging-strand gene (17). Additionally, transcription of rRNA genes is distinct from protein-coding gene expression; anti-termination and elongation factors such as NusG and ribosomal protein S4 associate with RNA polymerase at rRNA promoters, making the machinery more processive through these regions (62-64). Finally, historical experiments never detected any negative consequences of rRNA gene transcription in the native orientation.

Subsequent studies demonstrated that even co-directional rRNA transcription impedes replication, leading to preferential association of the replicative helicase and restart proteins in these regions (8). The improved sensitivity afforded by next-generation sequencing has identified additional endogenous, highly transcribed co-directional loci that inhibit replisome progression in the absence of conflict resolution factors (65). Additionally, reporters consisting of identical, inducible protein-coding genes in either orientation have demonstrated that head-on collisions outside the rDNA cause replication stress, mutagenesis, and a need for recombination proteins (17-19). Our understanding of

the strategies that cells use to cope with conflicts and their physiological consequences continues to blossom as the tools and detection methodologies improve.

1. 4 Conflict-mitigation strategies: genome organization

Bacterial genomes are organized in a manner that minimize head-on collisions (56, 59, 66). All sequenced bacterial phyla display a leading-strand bias, although the degree varies by clade (67). *E. coli,* encodes 55% of essential genes on the leading strand of replication, the *B. subtilis* genome is 76% co-oriented with respect to replication, and *Mycobacterium smegmatis* displays a 95% strand bias (16, 66).

The replication-related organization of bacterial genomes remains an ongoing topic of debate, with numerous hypotheses as to the most important contributing factors including gene essentiality, gene-dosage, selection against the production of truncated peptides, and conflicts with transcription (14, 43, 67-70). Regardless of the underlying mechanism, all sequenced bacterial genomes display some degree of leading strand bias.

Multiple models for the co-orientation bias have been proposed. Observations suggesting that the orientation of genes correlates more strongly with essentiality than expression levels argued against the detrimental effects of head-on collisions on replication as major forces shaping bacterial genomes (68, 69). Instead, these models suggested that the head-on orientation of genes inhibits transcription, potentially causing the production of truncated RNA messages (71). Therefore, fitness defects associated with interruptions to essential gene expression select against the lagging strand orientation. However, regardless of whether the underlying cause preserves transcriptional integrity or processive replication, interplay between the two processes clearly shapes bacterial genomes across evolutionary time (59, 66)

1.5 Conflict-mitigation strategies: resolution factors

Upon collisions that disrupt the replisome, cells must clear the obstacle from the DNA template (72), then re-establish replication forks in the region (73, 74). Classically, conflict resolution

factors fall into one of two categories: helicases or RNA polymerase modulators that remove the impediment, and recombination proteins that rearrange the DNA at collapsed forks.

Studies in *E. coli* demonstrated that the accessory helicases UvrD and Rep are required for efficient replication progression through inverted rRNA operons (10). *In vitro* investigations identified a role for the RNA polymerase modulator Mfd in preventing double-stranded breaks in case of co-directional encounters between replication and transcription (53). Work in our lab demonstrated that the *B. subtilis* paralogue of UvrD and Rep, PcrA, is a conflict resolution mechanism that promotes efficient replication fork progression through the most highly transcribed head-on and co-directional genes genome-wide. Additionally, our group showed that Mfd acts in a transcription-coupled nucleotide excision repair pathway, active at head-on conflict regions (18).

In order for replication to proceed after removing an obstacle, oftentimes the DNA must be remodeled to be refractory for replisome assembly, often through homologous recombination (75-77). Although *in vitro* studies indicated that forks may restart directly after colliding with a head-on RNA polymerase (52), genetic studies demonstrated that *E. coli* strains harboring inverted rRNA operons required RecBCD for survival (44). The *in vivo* results are consistent with the well-established requirement for recombination to restart replication after encounters that stall the replisome, but do not break the DNA (such as at bulky DNA lesions) (78-83).

Intriguingly, conflict resolution factors themselves may play conflicting roles at forks stalled due to encounters with transcription. UvrD, an accessory helicase that promotes replisome progression through head-on conflicts in *E. coli*, also removes RecA from the DNA, inhibiting recombination (and therefore, potentially, restart) (84). The apparent contradiction may arise because, RecA did not contribute to efficient survival of rRNA operon inversions in *E.coli* (10, 44), suggesting that recombination proteins processed the DNA through a fork-reversal mechanism distinct from classical recombinational break repair at these regions (78, 79). Whether these findings are truly applicable to the requirements for restart across all bacterial phyla, and at replication-transcription collisions occurring outside of inverted rDNA, remains controversial.

1.6 Increased mutagenesis in lagging strand genes as a consequence of conflicts

Bioinformatic analyses have suggested that lagging strand genes appear to mutate at a higher rate than their leading strand counterparts (17, 18, 85, 86). Initial proposals that the increased variability arises from the inherent asymmetry of the replication fork (87), were largely cast aside with the discovery of multiple pathways actively causing differential mutagenesis between the strands (88, 89). The underlying reason why a repair pathway should be more active on the lagging strand remained unclear until experiments measuring mutation rates of reporter genes in both orientations at high and low levels of expression showed that head-on transcription asymmetrically increases mutagenesis (17, 19, 90). Further work from our lab revealed that the increased mutagenesis arose from the activity of a Y-family polymerase, PolY1, in transcription-coupled nucleotide excision repair at expressed lagging strand genes (18).

The maintenance of genes in the head-on orientation is an ongoing biological paradox. Head-on gene expression carries potential fitness defects through impediments to replication and increased mutagenesis. While mutations, in general, have the potential to be deleterious by disrupting the function of an essential gene, without mutations bacteria could never generate the necessary diversity to adapt to changing environments (91-95). The increased mutation rate observed on the lagging strand could reflect a simple lack of purifying selection in these genes (69, 96). Alternatively, the lagging strand orientation could serve to promote mutagenesis for specific genes, while maintaining globally low mutation rates for the majority of the co-directionally oriented genome (93, 97, 98).

Bacteria employ mechanisms to increase mutation rates in response to environmental conditions such as the induction of mutagenic DNA polymerases during the SOS (99-102) and stringent responses (103-105). However, these mechanisms increase mutation rates genome-wide. Head-on transcription represents a mechanism to increase mutation rates in specific genes, thereby enabling certain loci to explore novel sequence space while keeping global mutation rates low.

Strikingly, the genes on the lagging strand are over-represented for membrane pumps, transporters, adhesins, and other proteins that the cell uses to respond to stress (**figure 1.4)**. Additionally, bioinformatics analyses revealed evidence for convergent mutations and positive selection

occurring in lagging strand genes. Although the model is provocative, maintaining certain genes on the lagging strand may be a mechanism to target accelerated evolution to specific genes (17).

The existence of a conserved factor (PolY1) that specifically increases mutation rates in expressed lagging-strand loci (18) supports the notion that cells may harness conflicts with transcription for the purposes of targeted genetic variation. Though the strategy is innately risky, the potential benefits associated with adaptive variation may outweigh any fitness defects (92, 93, 106). My thesis research uncovered repair mechanisms that minimize the detrimental effects of head-on collisions and the factors responsible for generating mutations when encounters occur.

B. subtilis core replisome	Architecture	Domain	Enzymatic activity	Function	Gene name
PolC Polymerase III	Monomer	NTD	Interacts with τ	Polymerase traffic	*polC*
		PHP	Polymerase and histidinol phosphatase	Initiation of DNA synthesis	
		Exo	3'=>5' Exonuclease	Proofreading	
		Pol3	Polymerase core	DNA synthesis	
		OB	Oligonucleotide binding	DNA binding	
		HhH	Helix-hairpin-helix	Beta-clamp interaction	
DnaE Polymerase III	Monomer	PHP	Polymerase and histidinol phosphatase	Initiation of DNA synthesis	*dnaE*
		Pol3	Polymerase core	DNA synthesis	
		OB	Oligonucleotide binding	DNA binding	
		HhH	Helix-hairpin-helix	Beta-clamp interaction	
		CTD	Interacts with τ	Polymerase traffic	
Sliding clamp	ββ homodimer	β	dsDNA binding	Processivity clamp	*dnaN*
Clamp loader	τ3δ1δ1 heteropentamer	τ	ATPase	Clamp closer, motor, links core polymerase to clamp loader & helicase	*dnaX*
		δ	Interacts with β	Clamp-opener, wrench	*holA*
		δ'	Interacts with β & ϒ	Stationary platform for δ & ϒ motion	*holB*

Table 1.1 Core components of the *B. subtilis* replisome. The core replicase of *B. subtilis* consists of two distinct Polymerase III molecules, each engaged with a processivity clamp and clamp loader complex. The DnaE polymerase lacks proofreading and extends from RNA primers. The PolC polymerase synthesizes DNA 5'➔3' from DNA primers. The clamp loader complex connects the core synthesis machinery to the replicative helicase.

E. coli core replisome	Architecture	Subunit	Enzymatic activity	Function	Gene name
Polymerase III	αεθ heterotrimer	α	5'=>3' DNA polymerase	DNA synthesis	*dnaE*
		ε	3'=>5' Exonuclease	Proofreading	*dnaQ*
		θ	Stimulates ε activity	Unknown, proposed fidelity factor	*holE*
Sliding clamp	ββ homodimer	β	dsDNA binding	Processivity clamp	*dnaN*
Clamp loader	$\gamma_3\delta_1\delta_{1'}\Psi_1$ heteropentamer	γ	ATPase	Clamp-closer, motor	*dnaX* (truncated product)
		δ	Interacts with β	Clamp-opener, wrench	*holA*
		δ'	Interacts with β & γ	Stationary platform for δ & γ motion	*holB*
		χ	Interacts with SSB and primase	Polymerase cycling	*holC*
		Ψ	Interacts with χ	Strengthens complex	*holD*
τ/γ Complex	$\tau_2\gamma_1$	τ	Interacts with α, β & γ	Links core polymerase to clamp loader & helicase	*dnaX* (full-length product)

Table 1.2 Core components of the *E. coli* replisome. *E. coli*, distinct from *B. subtilis*, encodes one Polymerase III, with three distinct subunits. Other differences include the presence of the *holC* & *holD* gene products, and the Y isoform of the *dnaX* gene within the core replication machinery.

Other essential replisome components	Activity	Function	*B. subtilis* gene	*E. coli* gene
Helicase	5'=>3' Helicase/ATPase	Unwinds duplex DNA	*dnaC*	*dnaB*
Primase	DNA-directed RNA polymerase	RNA primer synthesis	*dnaG*	*dnaG*
Single-strand binding protein	Single-stranded DNA binding	Coats and protects ssDNA	*ssbA*	*ssbA*
Primosomal factor Y	3'=>5' Helicase/ATPase	Replication initiation/restart	*priA*	*priA*
Primosome assembly protein	Interacts with PriA and DnaB	Primosome assembly	*dnaD*	*dnaT*
Helicase loader	Interacts with PriA, DnaD, & DnaC	Helicase loader	*dnaI*	*dnaC*
Helicase loader	Interacts with DnaI, DnaD, & DnaC	Helicase loader	*dnaB*	No *E. coli* homologue
Ligase	5'-3' phosphodiester catalysis	Seals Okazaki fragmens	*ligA*	*ligA*
Topoisomerase IV	Relaxes positive supercoils	Separates linked sister chromosomes	*parC* & *parE*	*parC* & *parE*
Gyrase	Introduces negative supercoils	Relieves positive supercoils	*gyrA* & *gyrB*	*gyrA* & *gyrB*

Table 1.3 Other essential replisome components. Essential proteins for DNA unwinding, primer synthesis, supercoil relaxation, helicase loading, and DNA ligation are listed. *B. subtilis* and *E. coli* employ distinct primosome assembly pathways (DnaDIB- vs DnaTC-mediated helicase loading).

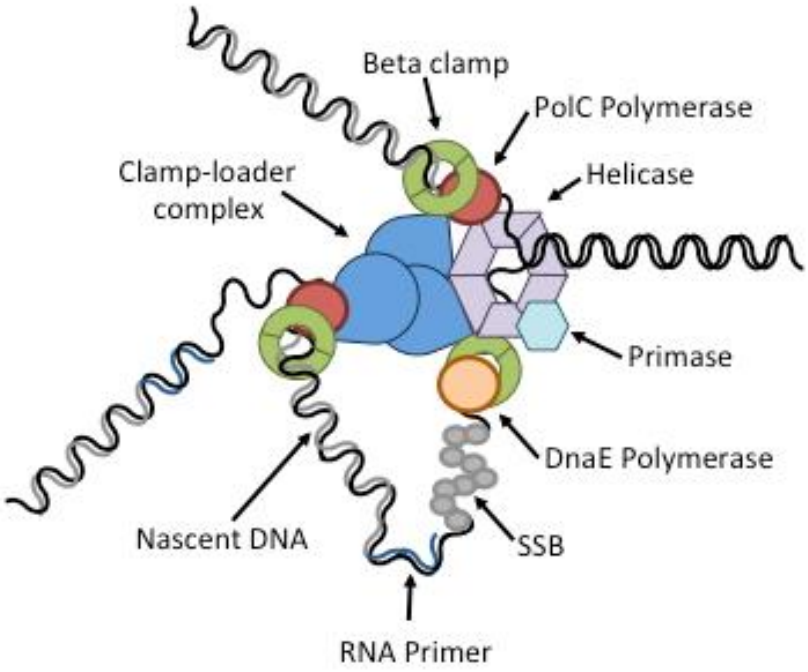


Figure 1.1 The architecture of the *B. subtilis* replisome. Cartoon representation of the *Bacillus subtilis* replisome with three active polymerases engaged in simultaneous leading and lagging strand synthesis. Newly synthesized DNA is indicated in light gray, the parental duplex is colored black. A complete list of replication proteins, as well as their associated functions is given in **table 1.1**.

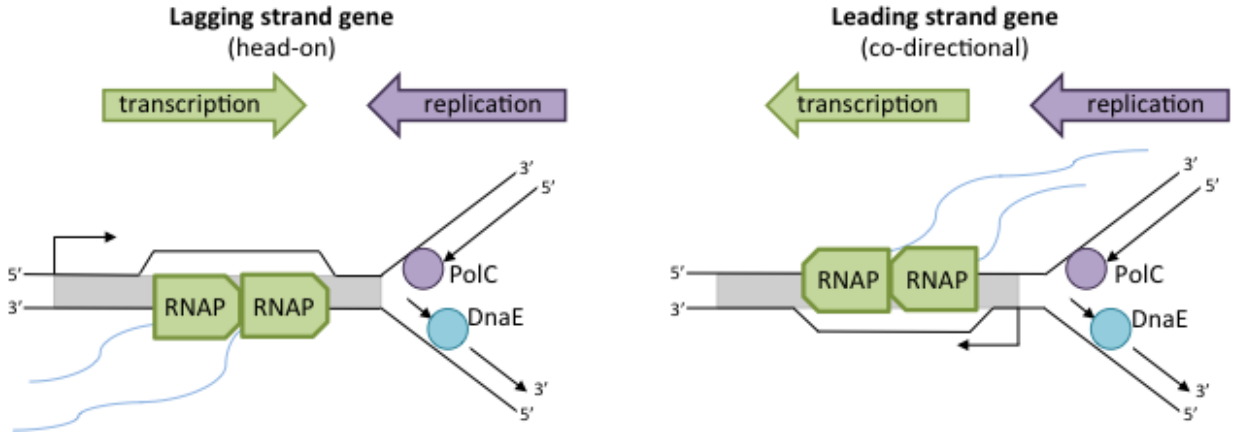


Figure 1.2 Two possibilities for replication-transcription conflicts. RNA and DNA Polymerases proceed in a 3' to 5' direction on the template strand while synthesizing 5' to 3'. For a lagging strand-encoded gene, the two machineries move in opposite directions, and the template strand is replicated discontinuously. Leading strand encoded genes are transcribed in the same direction as the continuously replicated template strand.

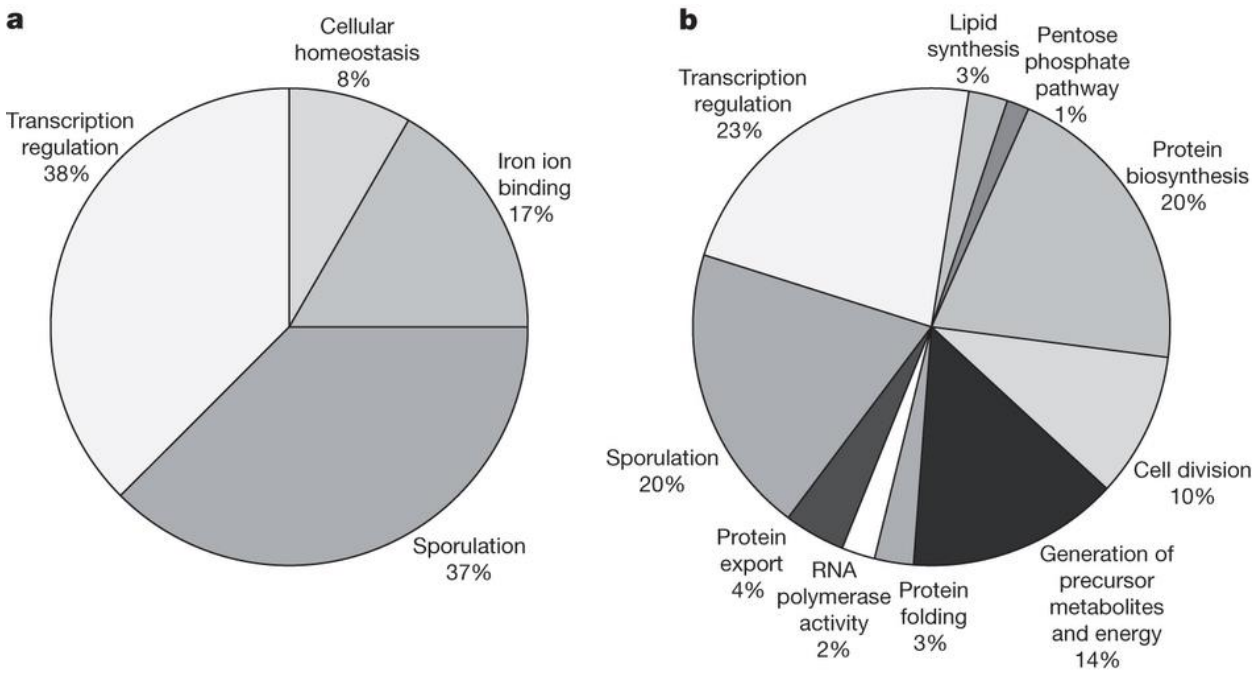


Figure 1.3 General characteristics of lagging-strand genes. Pie charts presenting significantly ($p < 0.05$) overrepresented functional categories (as determined by medium stringency using DAVID) of protein products from lagging (A) and leading (B) strand of replication for *B. subtilis* strain168. The genes categorized for the lagging strand are: *abrB, acuA, acuB, ahpC, ahpF, antE, cotM, cspD, cueR, def, fer, gerPF, hpr, ispG, katA, lexA, mrgA, nasE, sacY, sda, sigM, sigV, sinI, sinR, spoIISB, sspB, sspD, sspI, sspK, sspL, sspM, sspO, tnrA, ycnC, ydbP, ydgJ, yhdK, yjbI, ynzD, ypoP, ytzE, yutI, yuxN, ywoH* (17).

Chapter 2.

Replication restart after collisions with transcription requires RecA in *B. subtilis*

Originally published as an article in *Journal of Bacteriology*

Milion-Weaver, S*, Samadpour, A.*, Merrikh, H. Replication restart after replication-transcription conflicts requires RecA in B. subtilis. *J Bacteriol*. 2015 May 4. pii: JB.00237-15

Abstract

Efficient duplication of genomes depends on reactivation of replication forks outside of the origin. Replication restart can be facilitated by recombination proteins, especially if single or double stranded breaks form in the DNA. Each type of DNA break is processed by a distinct pathway, though both depend on the RecA protein. One common obstacle that can stall forks, potentially leading to breaks in the DNA, is transcription. Though replication stalling by transcription is prevalent, the nature of DNA breaks and the prerequisites for replication restart in response to these encounters remain unknown. Here, we use an engineered site-specific replication-transcription conflict to identify and dissect the pathways required for the resolution and restart of replication forks stalled by transcription in *Bacillus subtilis*. We find that RecA, its loader proteins, RecO and AddAB, and the Holliday junction resolvase, RecU, are required for efficient survival and replication restart after conflicts with transcription. Genetic analyses show that RecO and AddAB act in parallel to facilitate RecA loading at the site of the conflict, but that they can each partially compensate for the other's absence. Finally we find that RecA, and either RecO or AddAB, is required for the replication restart and helicase loader protein DnaD to associate with the engineered conflict region. These results suggest that conflicts can lead to both single stranded gaps and double stranded breaks in the DNA, and that RecA loading and Holliday junction resolution are required for replication restart at regions of replication-transcription conflicts.

2.1 Introduction

Bacterial chromosomal replication initiates from a single origin of replication and proceeds bi-directionally until completion at the terminus. Replication forks regularly encounter obstacles and need to be restarted as they traverse the chromosome (107, 108). The essentiality of restart proteins supports the notion that every single fork initiated from *oriC* will be disrupted at least once, if not more often, before reaching the terminus (82, 109). Transcription frequently impedes replication in bacteria, necessitating numerous factors to resolve conflicts between the two machineries (1). Head-on collisions between replication and transcription, which occur when a gene is encoded on the lagging strand, cause mutagenesis, genomic instability, single-stranded DNA accumulation, and potentially double-stranded DNA breaks (3-5, 18).

Previous studies investigating the consequences and coping strategies of cells with transcription from the lagging strand have relied on the use of strains harboring large chromosomal inversions that oriented the highly transcribed, repetitive rDNA operons head-on to replication (10, 43, 44). In *E. coli*, recombination-mediator proteins were required for efficient survival of rDNA inversions, though the repair protein RecA itself did not contribute to viability in these experiments (10, 44). Inversion of the rDNA leads to RecA-GFP focus formation in *B. subtilis* (43). Subsequent studies in *B. subtilis* demonstrated that RecA localizes to loci containing highly expressed head-on protein-coding genes (18). DNA processing by recombination proteins and/or break repair is likely important for replication-transcription conflict resolution. However, the type of DNA damage occurring upon collisions between replication and transcription, and the role, if any, that RecA plays at these regions remains elusive.

RecA catalyzes strand invasion as the first step in DNA break repair by homologous recombination (110). Depending on the type of DNA break, RecA is loaded onto the DNA (107) by either the RecFOR or AddAB pathway (111-114). RecFOR recognizes single-stranded gaps in the DNA (115), likely through an interaction with SSB (116). AddAB, by contrast, processes double-stranded ends to generate 3' overhangs (117, 118) in a manner analogous to RecBCD in *E. coli* (119). Whether AddAB itself loads RecA onto DNA in *B. subtilis* (an activity mediated by RecB in *E. coli*(120)), is an ongoing topic of debate.

RecA loading and strand invasion lead to the formation of four-way Holliday junction DNA intermediates (121, 122). Cells rely on resolvases, such as the RuvC protein in *E. coli*, to cleave four-way Holliday junction DNA, and separate linked sister chromosomes (123-125). RecU, the homologue of *E. coli* RuvC, carries out this function in *B. subtilis* (126-128). When recombination occurs at stalled replication forks, restart proteins associate with resolved recombination intermediates for growth to continue (76, 80, 81).

Restart in *B. subtilis* involves an ordered association of the essential primosomal proteins, PriA and DnaD with the DNA, followed by recruitment of the helicase loader proteins, DnaB and DnaI. DnaB and DnaI then load the helicase, DnaC, onto the DNA, allowing replication to proceed (74, 129, 130). Purified PriA protein can bind both stalled fork structures and D-Loops *in vitro*, though it has much higher affinity for the latter (75, 80, 131). Consistent with this, genetic analyses indicated that recombination proteins remodel stalled forks for PriA binding to occur (132). Reconstituted restart reactions *in vitro* demonstrated that both RecBCD and RecOR facilitate restart (133). However, the requirements for PriA binding to DNA *in vivo* remain poorly studied.

We set out to determine if RecA played a role in resolving conflicts between replication and transcription. To address this question, we inserted a single protein coding gene into the chromosome of *B. subtilis*, oriented either head-on or co-directionally to replication. We find that when the engineered construct is transcribed in the head-on, but not the co-directional, orientation the presence of at least one RecA loading mediator, RecO or AddAB, and the Holliday junction resolvase, RecU, are required for efficient survival. The transcription- and orientation-dependence of these phenotypes suggests that replication-transcription conflicts are the underlying cause of the observed survival defects. Epistasis analyses indicate that RecO, AddAB and RecU act in the same pathway as RecA to promote survival in response to, specifically, head-on transcription. These data suggested that recombination is required for resolution of replication-transcription conflicts. Consistent with this, we find that both AddAB and RecO load RecA onto DNA at the conflict region in a transcription-dependent manner, as indicated by chromatin immunoprecipitations (ChIPs) of RecA. Additionally, we found that RecA is required for the restart protein, DnaD, to associate with the engineered head-on conflict region. Our findings demonstrate that resumption of replication upon conflicts with transcription

primarily depends on a classical recombination-mediated and RecA-dependent replication restart pathway in *B. subtilis*.

2.2 Results

RecA, along with RecO and AddAB, promotes survival of cells experiencing severe head-on replication-transcription conflicts

Previous studies in *E. coli* indicated that head-on encounters between replication and transcription from inverted rDNA operons caused RecA-independent replication fork reversal (10, 44). We wondered if RecA played a role in overcoming conflicts between replication and transcription of a protein-coding gene in *B. subtilis*. There is precedence for this hypothesis, given that RecA-GFP localization has been observed in in *B. subtilis* strains harboring rDNA inversions(43), and that RecA association with regions of head-on transcription has been detected by ChIP (18).

To address if and how RecA contributes to survival of replication-transcription collisions, we utilized a previously described site-specific, inducible conflict (8, 18). Strains harboring this reporter carry a copy of the *lacZ* gene under control of the ICE*Bs*1 promoter, P_{xis}, in the head-on orientation with respect to replication at the *thrC* locus of the *B. subtilis* chromosome. In strain backgrounds cured of the ICEBs1 element the P_{xis} promoter is highly expressed, whereas in the presence of this element, the promoter is tightly repressed (134), leading to minimal expression of the *lacZ* gene. We grew cultures of *B. subtilis* either expressing or repressed for transcription of the P_{xis}-*lacZ* to similar optical densities, spotted serial dilutions onto LB agar plates, and enumerated colony forming units (CFUs) arising under both conditions (**figure 2.1a**). We then quantified any survival defects associated with deleting genes for homologous recombination under both conditions. To determine the specific contribution of each protein of interest to survival of the highly expressed head-on gene, we took ratios (expressed as plating efficiency percentages) of CFUs arising in the presence of transcription of *lacZ* to CFUs in strains where the *lacZ* was repressed (**figure 2.1b**).

We found that cells expressing P_{xis}-*lacZ* in otherwise wild-type backgrounds did not display significant reductions in plating efficiency, consistent with the presence of multiple pathways to resolve the negative consequences of head-on collisions between replication and transcription in bacteria (1). Cells lacking *recA* displayed significant survival defects associated with transcription from the *lacZ* gene, leading to a five-fold reduction in plating efficiency (**figure 2.1a & b**). Neither deletion of *recO*, nor *addAB*, individually caused significant transcription-specific survival defects (**figure 2.1b**).

However, *lacZ* expression significantly impaired survival of cells lacking both *recO* and *addAB* (**figure 2.1a & b**). To confirm that the survival defects we observed were specific to head-on conflicts with transcription and not simply *lacZ* expression itself, we also performed plating efficiency assays in strains where the gene was oriented co-directionally with replication (i.e. on the leading strand) (**figure 2.2**). Although co-directional transcription does impede the replisome (8, 53), these encounters are much less disruptive compared to when replication and transcription meet head-on. When *lacZ* was expressed from the leading strand, deletion of *recA* or *recO* and *addAB* did not cause any significant reduction in plating efficiencies, arguing against transcription of *lacZ* alone causing survival defects. The orientation-specificity implies that head-on conflicts with replication upon *lacZ* transcription necessitate RecA and its loader proteins for efficient survival. The lack of a significant phenotype for either *addAB* or *recO* single mutants, combined with the synergistic effect of combining the two mutations indicate one of two possibilities: either each individual pathway alone is not important, or that one may compensate for the other's absence.

We wondered if AddAB and RecO contributed to survival by loading RecA at regions of head-on transcription. To determine whether *recO* and *addAB* act in the same pathway as *recA,* we constructed strains lacking all three genes. We found that combining deletions of *recA* with the *recO* and *addAB* deletions, did not lead to any additional survival defects as compared to the parent strains (**figure 2.1a & b**). The epistatic relationship of *recA, addAB,* and *recO* indicates that both RecO and AddAB facilitate survival of obstacles to replication through RecA.

Both RecO and AddAB load RecA at head-on conflict regions

Two pathways potentially load RecA onto DNA in *B. subtilis:* RecFOR at single-strand gaps, or AddAB at double stranded breaks. Previous studies indicated that RecO is required for RecA localization in response to treatment with DNA damaging agents in cells grown in minimal media (113). However, the potential contribution of either pathway to RecA loading in cells grown in rich medium, without exogenous damage is unclear. The plating efficiency data indicated that both pathways contribute to survival of head-on conflicts through a genetic interaction with RecA. We set out to determine if

RecFOR and/or AddAB load RecA onto the DNA at a specific replication-transcription conflict region in *B. subtilis*.

To determine which RecA loading pathway, RecFOR and/or AddAB, contributed to RecA localization in response to head-on transcription, we first quantified RecA-GFP focus formation using microscopy. In the absence of transcription we found that RecA-GFP formed foci in roughly 10% of cells, consistent with previous reports (135)(**figure 2.3a**). In the backgrounds where the *lacZ* reporter was not expressed, cells lacking RecO did not show a reduction in RecA-GFP focus formation, whereas cells without AddAB displayed reduced RecA-GFP localization. When the reporter gene was transcribed, we observed RecA-GFP foci in 20% of cells in otherwise wild-type backgrounds. Deletion of either *recO* or *addAB* led to reduced RecA-GFP localization in cells expressing the *lacZ* gene. Combining the two mutations (*recO addAB* double) had an additive effect (**figure 2.3a**).

The results of the microscopy experiments suggested that AddAB, and not RecO, is the major mediator required for RecA localization during growth in rich media. Cells lacking RecO did show a modest increase in RecA-GFP localization in the absence of transcription, relative to control cells. Therefore, RecO may, to some degree, also load RecA at single stranded gaps under these conditions. The increase observed in RecO deficient strains probably reflects unresolved single-stranded gaps being converted to double-stranded breaks due to run-off replication (136).

RecA-GFP focus formation cannot distinguish between general replication stress (i.e. RecA localizing to the DNA elsewhere in the genome) and the specific effects of transcription of the reporter gene. We set out to directly determine the contribution of each pathway to RecA localization specifically at the collision region. To do this, we used ChIPs to measure the relative association of RecA with the replication roadblock, compared to the control locus, *yhaX* (other control loci produce similar results to *yhaX* (18)), in the presence and absence of transcription. When the reporter gene was repressed, we did not detect preferential association of RecA with the region in any genetic background. In strains expressing the P_{xis}-*lacZ* construct, we observed a five-fold enrichment of RecA at the region relative to the control locus (**figure 2.3b**). RecA association with the region was reduced in strain backgrounds deficient for *recO* or *addAB*. Similar to the results of the microscopy experiments, combining the *recO* and *addAB* deletions had an additive effect; we no longer detected any

preferential RecA association at the replication roadblock when *lacZ* was expressed in cells lacking both RecO and AddAB (**figure 2.3b**). The RecA ChIPs suggest that, consistent with the results of the plating efficiency assays, at least RecO or AddAB is required for RecA to associate with regions of head-on transcription, but one pathway may compensate for the other's absence.

RecU acts in the same pathway as RecO, AddAB and RecA to help cells survive head-on conflicts

The localization of RecA at the conflict region and contribution to efficient survival suggested that strand invasion at the site of the conflict helped cells tolerate head-on encounters between replication and transcription. Strand invasion by RecA leads to the formation of four-way Holliday junction intermediates, which are cleaved by resolvases to separate the interlinked chromosomes (121, 122). We hypothesized that resolution of recombination intermediates would therefore contribute to cellular survival of replication-transcription conflicts. To test this, we extended our epistasis analysis to determine the effect of the *B. subtilis* Holliday junction resolvase, RecU, which binds Holliday junctions with sub-nanomolar affinity (128), to plating efficiencies for cells harboring the highly expressed head-on *lacZ* gene.

Deletion of *recU* alone led to reduced plating efficiencies, suggesting that cleavage of four-way junctions contributes to survival of head-on conflicts (**figure 2.4**). No additional reductions in plating efficiencies were observed when the *recU* and *recA* deletions were combined, indicating that these genes function in the same pathway, likely by forming and resolving the four way junction, respectively. However, because of the limited sensitivities of plating efficiency assays at saturated of repair capacities, we cannot exclude the possibility that RecU performs another RecA-independent DNA processing function at head-on conflict regions.

Because single deletions of *recO* or *addAB* did not cause significant survival defect phenotypes, the genetic interactions between *recU* and either individual pathway cannot be inferred from cellular survival data. However, deleting *recU* in backgrounds lacking both *recO* and *addAB* did not lead to any additional survival defects (**figure 2.4**). The observed epistasis suggests that RecU acts in the same pathway as both RecO and AddAB, as well as RecA, to promote survival of cells upon head-on collisions

between replication and transcription. These genetic interactions are consistent with RecU's classically defined function to resolve Holliday Junction intermediates arising from homologous recombination.

RecA loading is required for the restart protein, DnaD, to associate with head-on conflict regions

Based on the results presented above, we hypothesized that RecA aids replication reactivation in response to conflicts by facilitating the formation of D-loops, which can promote association of replication restart proteins with stalled forks (132, 133). To test this model, we used ChIP to measure the relative association of the restart protein, DnaD, with the roadblock-containing region compared to the control locus, *yhaX*, in the presence and absence of transcription in strains deficient for each recombination protein. During primosome assembly, PriA facilitates DnaD loading in order to reload the replicative helicase DnaC.

Expression of the reporter gene in recombination-proficient backgrounds led to significantly increased relative DnaD association with the *lacZ*-containing region. In strains lacking RecA, we no longer detected DnaD association with the region when the reporter gene was expressed (**figure 2.5**). Individual deletions of either *recO* or *addAB* reduced DnaD association with the *lacZ* locus (**figure 2.5**). Similar to the results obtained with the *recA* deletion, in backgrounds lacking both *recO* and *addAB* we no longer observed preferential DnaD association with the highly expressed head-on gene (**figure 2.5**). To rule out the possibility that altered transcription of the P_{xis}-*lacZ* construct in backgrounds lacking *recA* eliminated the obstacle to replication in these strains, we used ChIP to measure the association of RNA polymerase beta subunit (RpoB) with the *lacZ*-containing region with and without transcription from P_{xis}. We found that deletion of *recA* did not alter RNA polymerase occupancy as compared to wild-type backgrounds (**figure 2.6a**). Furthermore, to rule out the possibility that pleiotropic effects of the *recA* deletion on replication somehow reduced the severity of the effects of conflicts with transcription in these strains, we measured the relative association of the replicative helicase, DnaC, with the region in the presence and absence of transcription. The replicative helicase is expected to localize evenly along the chromosome in an asynchronous population of cells in the absence of specific replication stress. As such, preferential association relative to a control locus probably indicates replication stalling at that region. We found that, as we have reported previously, active transcription from P_{xis} led

to a significant increase in the relative association of DnaC with the region (18). This association however was not affected by deletion of *recA* (**figure 2.6b**), indicating that severe conflicts between transcription and replication still occur even in *recA* deletion backgrounds, and that the absence of RecA specifically inhibits replication restart.

Taken together, these results suggest that a RecA-dependent mechanism is needed for restart proteins to localize to head-on conflict regions. Though RecO and AddAB each partially contribute, at least one pathway is necessary for DnaD association with highly expressed head-on genes.

2.3 Discussion

Our results demonstrate that RecA contributes to efficient survival of head-on replication-transcription conflicts, likely by promoting replication restart. The plating efficiencies and ChIP data strongly suggest that both single-strand gaps and double-stranded breaks occur at head-on conflict regions, both of which are acted upon by RecA (**figure 2.7**). The survival defects we observe are modest, indicating that not every conflict causes catastrophic consequences for the cell. However, the requirement for RecA loading for replication restart proteins to associate with the conflict region suggests that when severe collisions do occur, replication requires RecA to proceed. Our results establish RecA-dependent recombination as a prerequisite for the reactivation of replication upon stalling at transcription units in *B. subtilis*.

RecO and AddAB-mediated conflict resolution

RecA binds single-stranded DNA with 3' overhangs (137). RecA rapidly polymerizes on DNA to form a nucleoprotein filament, the functional complex that catalyzes strand invasion (138). However, RecA monomers alone have low affinity for Ssb-coated DNA (139). In order for RecA filaments to form, monomers require nucleation onto DNA, a process called RecA loading (140). Two pathways exist for RecA loading in *B. subtilis* depending on the type of DNA break that occurs.

RecOR loads RecA onto single stranded DNA in *B. subtilis*. RecO is likely recruited to ssDNA by an interaction with the C-terminal tail of Ssb, as has been demonstrated in *E. coli* (141). AddAB processes double-stranded breaks to generate 3' overhangs (119). Previous work suggested that RecO is required to load RecA onto the DNA regardless of the type of damage, even at double-stranded breaks (113). These investigations used RecA-GFP focus formation as a readout for RecA loading in cells growing in minimal medium. We observed RecA-GFP focus formation and DNA-association by ChIP even in cells lacking RecO, suggesting that other pathways, such as AddAB, also facilitate loading of RecA at stalled replication forks. The differences between our results and the previous study could arise from the different growth conditions (minimal vs. rich medium) or some undetermined co-factor present at replication forks stalled due to transcription that is absent when forks stall in response to chemical DNA damage.

Our results also indicate that one pathway may act in the other's absence, which suggests that AddAB can activate RecA loading independently of RecO. The ability of AddAB to compensate for RecO likely arises because gaps formed at stalled replication forks are converted to double stranded breaks during subsequent rounds of replication (136). RecO may compensate for the absence of AddAB if other nucleases within the cell, such as RecJ in concert with RecQ or RecS (142), resect the double-stranded ends to generate 3' overhangs for RecA loading. Studies in *D. radiodurans* have demonstrated that RecFOR can play a predominant role in repairing double stranded breaks (143), suggesting that the absence of AddAB doesn't necessarily preclude this type of damage repair. Regardless of the mechanism or type of damage, however, in our system, both RecO and AddAB act through RecA.

Sources of single-stranded gaps and double-stranded breaks at head-on conflicts

The results of the plating efficiency assays and ChIPs showing that both RecO and AddAB contribute to RecA loading and restart suggest that both single-stranded gaps and double-stranded breaks occur at head-on conflict regions. Our assays lack the sensitivity to directly determine which type of damage predominates, however, it is clear that RecA is important for replication restart at both gaps and breaks. The initial encounter between replication and transcription likely generates a single stranded gap, which would be recognized by RecO. However, because both pathways contribute to tolerating collisions, some double stranded breaks may occur.

We envision three possible scenarios (as illustrated in **figure 2.7)**: direct RecA loading by RecO at single-stranded gaps, end-processing and RecA loading by AddAB at any breaks that do occur, and, occasionally, gaps being converted into double-stranded breaks due to run-off replication.

Single stranded gaps may arise because both RNA polymerase and DNA polymerase unzip the double-helical DNA template, causing negative supercoils and underwound DNA accumulation behind each of the machineries (144-146). Negative supercoiling behind RNAP contributes to R-Loop formation, which may maintain one strand of the DNA duplex in a single-stranded state (147). At the site of the collision itself, persistent gaps on the lagging strand template, such as observed in restart-deficient PriA mutants (148) could necessitate RecFOR-mediated repair.

It is unclear if head-on collisions between replication and transcription directly cause double-stranded breaks in the DNA. The torsional strain that accumulates due to excessive positive supercoiling between the two machineries might be sufficient to break the ester linkages of the sugar-phosphate backbone. However, evidence of chromosome breaks arising directly due to a collision (as opposed to endonucleoltyic processing of the DNA in the region) has not been demonstrated.

Any single stranded gaps in the template can be converted into double-stranded breaks through replication run-off as forks converge. Direct restart from these structures would likely result in aberrant chromosomes. Therefore, double-stranded break repair may be required even if the initial insult to the DNA only generated a single-stranded gap or nick.

The role of D-Loops in replication restart

PriA allows for origin-independent replication initiation by recruiting helicase loader proteins to the DNA in *B. subtilis*. Purified PriA binds DNA bubble structures with low affinity (K_d=150nm), yet displays 10-fold enhanced binding kinetics to D-loop structures (Kd=15 nm) (131). Previous studies of primosome assembly have relied extensively on reconstituted systems *in vitro* because of the essential nature of most replication restart proteins. Our system allowed us to investigate the requirements for restart *in vivo*. Because DnaD required RecA and RecA loading to associate with the conflict-containing region, we hypothesize that a RecA-mediated D-loop is required for restart at conflict regions. Additionally, the contribution of RecU to efficient survival suggests a requirement for Holliday junction resolution at head-on conflicts. Consistent with this observation, PriA binds four-way DNA junctions with extremely low affinity (131) and RecU does not cleave D-loops in *B. subtilis* (149). Therefore, although speculative, we hypothesize that RecU remodels interlinked chromosomes after strand invasion, generating a D-loop structure, which is recognized by PriA.

Differences between replication fork reactivation mechanisms in *E. coli* and *B. subtilis*

Head-on collisions between replication and transcription at inverted rDNA operons in *E. coli* are overcome through a RecA-independent replication fork reversal mechanism (RFR) (10, 44). During RFR, the nascent DNA strands dissociate from their templates, regress, and re-anneal to form a four-way

DNA junction with a double-stranded end. Processing of the regressed fork by RecBCD allows for replication restart subsequent to these encounters. Alternatively, the Holliday junction resolvase RuvC may cleave the reversed fork, causing chromosome breakage (150).

Our results suggest that at head-on conflicts, RecA-independent Replication Fork Reversal (RFR) does not occur, at least in *B. subtilis*. In RFR in *E. coli*, RecA is dispensable but RecBCD is essential for viability. Additionally, the RFR model predicts that cleavage of the reversed fork by RuvC (RecU in *Bacillus*) leads to chromosome breakage in RecBCD-deficient cells. Our results clearly demonstrate that RecA associates with head-on conflict regions and establish that RecA loading and Holliday junction resolution re-establish the fork at conflict regions. Additionally, RecA and RecU act in the same pathway to promote conflict survival, suggesting that RecU alone does not play a role in processing of the forks such as in the proposed RFR reaction. Nevertheless, we cannot rule out RFR at the conflict region by a RecA-dependent process. Indeed, RecA regresses forks *in vitro* (151) and studies in eukaryotic cells demonstrate that RAD51 promotes fork reversal in response to a broad array of genotoxic stresses (152).

The observed discrepancies between the RecA-independent RFR model and our results could arise either from the nature of the head-on conflict investigated or the model organism studied. The *E. coli* studies were based on rDNA inversions, whereas our results rely on a protein coding gene. rDNA operons are repetitive and significantly longer than the single head-on protein-coding gene we used in our study. The impact of gene length on conflict severity and cellular strategies for resolution remains an understudied question. Additionally, rDNA transcription is distinct from expression of a protein-coding gene. Anti-termination factors such as NusB stabilize RNA polymerase in the rDNA (63), and high expression from rDNA promoters maintains the DNA in these regions in a persistently underwound state behind the transcription apparatus, whereas positive supercoils accumulate ahead of RNAP. Potentially the DNA topology innate to ribosomal DNA contributes to catalyzing fork reversal when operons are inverted head-on to replication.

The fundamental differences between the γ-proteobacteria *E. coli* and the firmicute *B. subtilis* may also account for an absence of detectable RecA-independent RFR in our experiments. *B. subtilis* uses two polymerases for chromosomal replication, DnaE and PolC (23), whereas *E. coli* employs only

one (20). Additionally, *E. coli* and *B. subtilis* overcome replication-transcription conflicts by distinct strategies. *E. coli* harbors two accessory helicases, UvrD and Rep, which cooperate to facilitate replication progression across inverted rDNA operons, whereas *B. subtilis* appears to only harbor one essential accessory helicase (PcrA), which is probably also important for conflict resolution. *B. subtilis* displays a more significant leading strand bias (74%) in the organization of its genome than observed in *E. coli* (55%) (1). Finally, inverting the rDNA causes cell death in 10% of cells in *B. subtilis*, whereas these genomic rearrangements are not significantly detrimental to survival of *E. coli* cells (10, 43). The results we present here, together with the prior investigations on replication-transcription conflicts in bacteria, highlight the significant differences in resolution mechanisms used by these two model organisms. Therefore, a deep understanding of the mechanisms employed by cells to facilitate accurate and timely DNA replication requires investigations in more than one species.

2.4 Figures

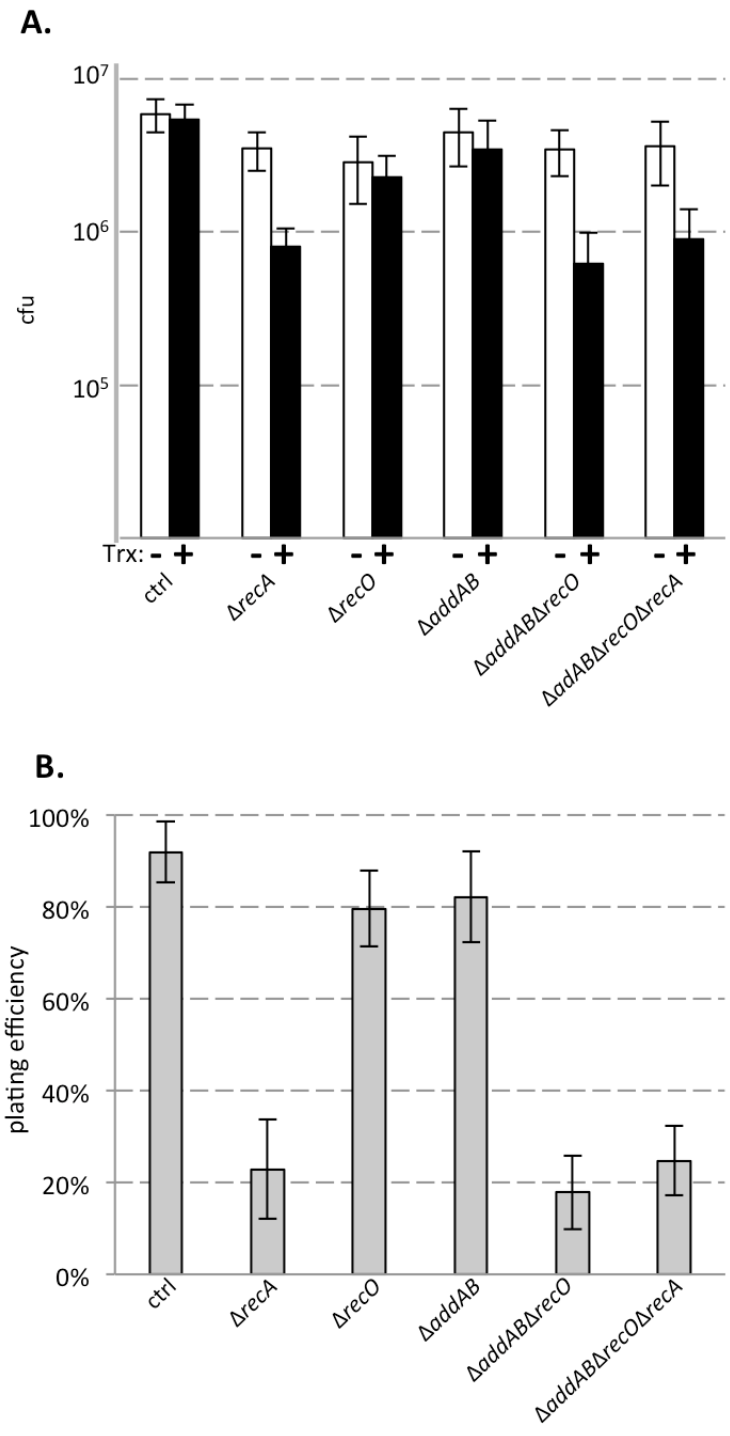


Figure 2.1 A) Colony forming units arising from exponentially growing cultures at OD_{600}=0.3 were enumerated in the presence (Trx +, black bars) and absence (Trx -, white bars) of transcription from P_{xis}-*lacZ*. **B)** Plating efficiencies (transcription-specific survival) were determined by enumerating the ratio of CFUs arising from cultures of strains expressing P_{xis}-*lacZ* to strains repressed for transcription of the construct in a given mutant background. Data shown are averages from between 8 to 12 biological replicates per strain. Error bars represent error of the mean (**A**) or standard deviations (**B**).

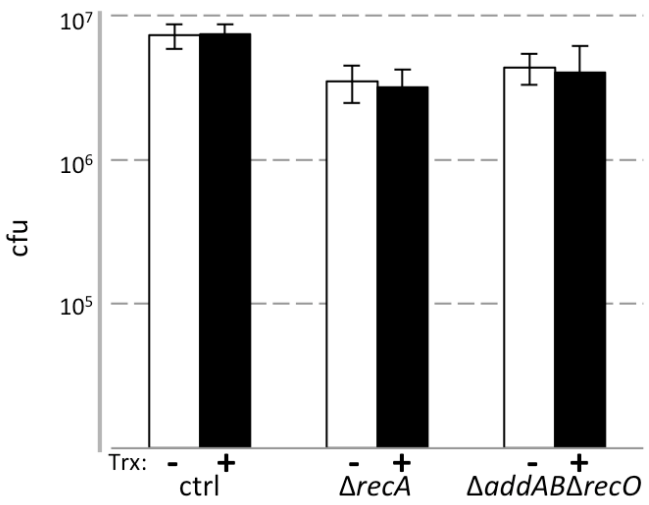


Figure 2.2 Colony forming units arising from exponentially growing cultures at OD_{600}=0.3 were enumerated in the presence (Trx +, black bars) and absence (Trx -, white bars) of transcription from P_{xis}-*lacZ* for strains harboring the construct in the co-directional orientation.

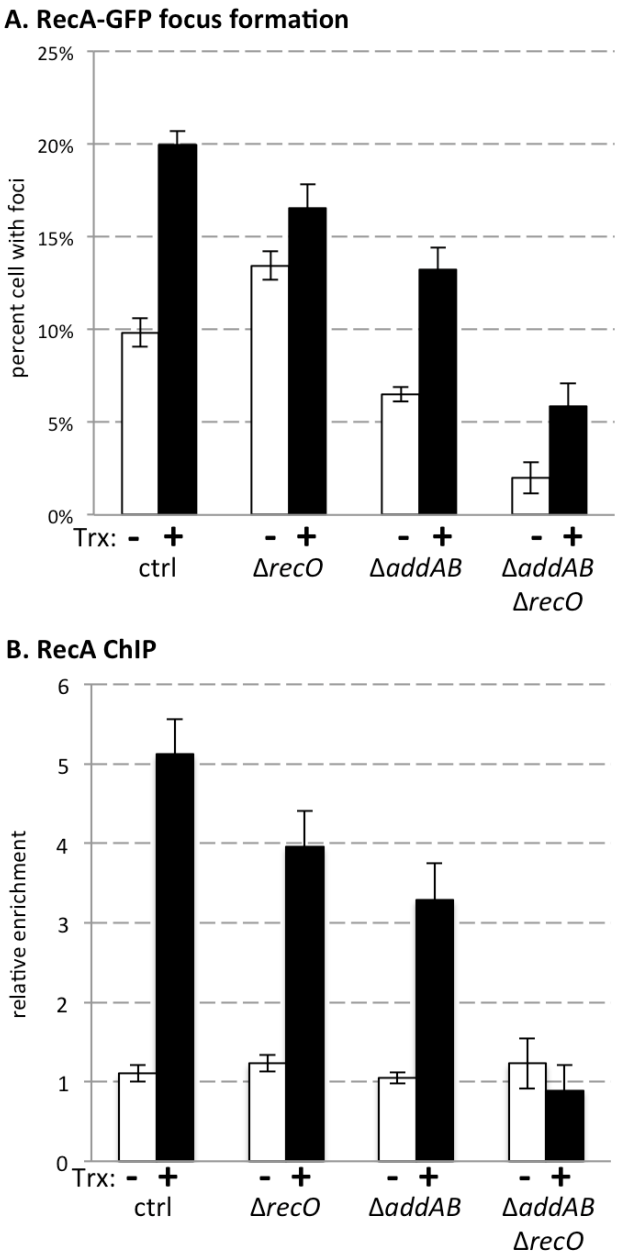


Figure 2.3 A) RecA-GFP focus formation was quantified by microscopy in the absence (Trx -, white bars) and presence (Trx +, black bars) of transcription of the P_{xis}-*lacZ* gene **B)** Relative association of RecA with the head-on gene-containing region was measured by ChIP-qPCR in the absence (Trx -, white bars) and presence (Trx +, black bars) of transcription of the P_{xis}-*lacZ* gene. Data shown represents averages of 6 to 16 biological replicates. Error bars represent standard error of the mean.

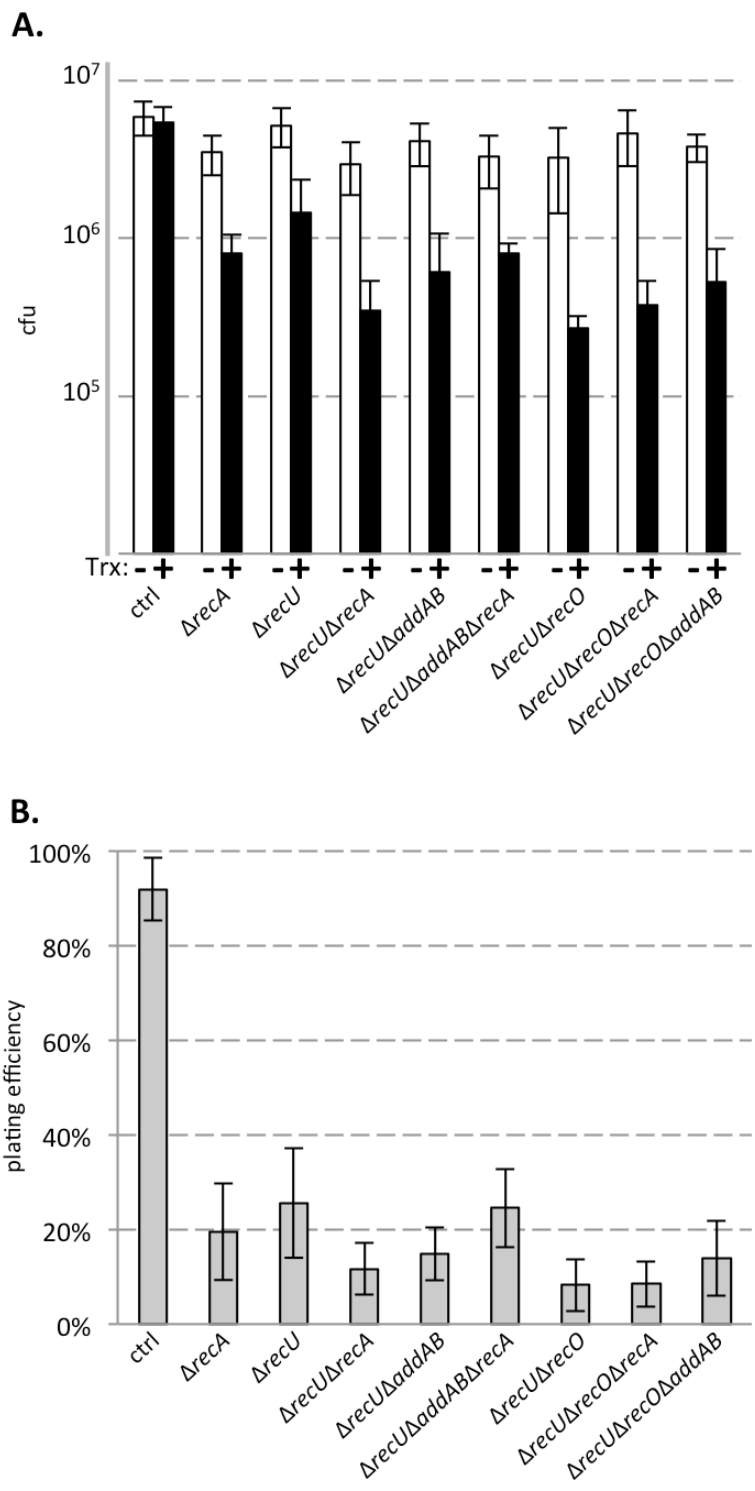


Figure 2.4 A) Colony forming units arising from exponentially growing cultures at OD_{600}=0.3 were enumerated in the presence (Trx +, black bars) and absence (Trx -, white bars) of transcription from P_{xis}-*lacZ*. **B)** Plating efficiencies (transcription-specific survival) were determined by enumerating the ratio of CFUs arising from cultures of strains expressing P_{xis}-*lacZ* to strains repressed for transcription of the construct in a given mutant background. Data shown are averages from between 8 and 14 biological replicates per strain. Error bars represent standard error of the mean (**A**) or standard deviation (**B**).

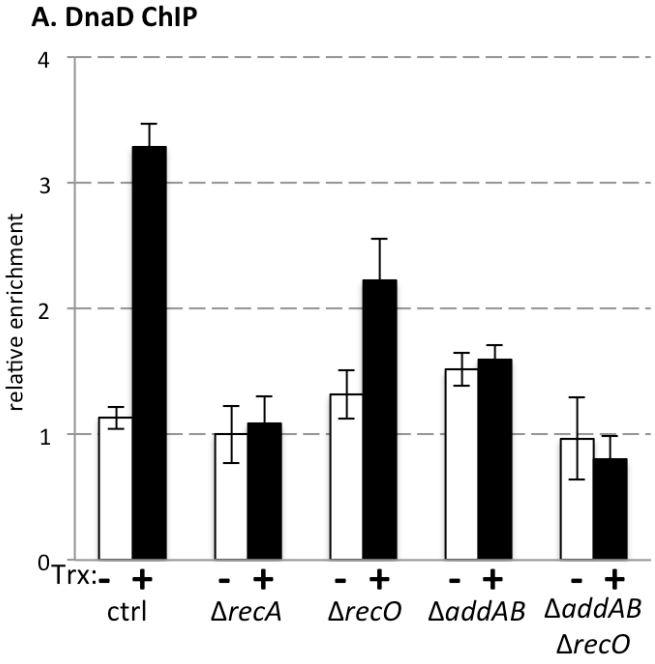


Figure 2.5 Relative association of DnaD with the head-on gene-containing region was measured by ChIP-qPCR in the absence (Trx -, gray bars) and presence (Trx +, white bars) of transcription of the P_{xis}-*lac*Z gene. Data shown represents averages of 12 biological replicates. Error bars represent standard error of the mean.

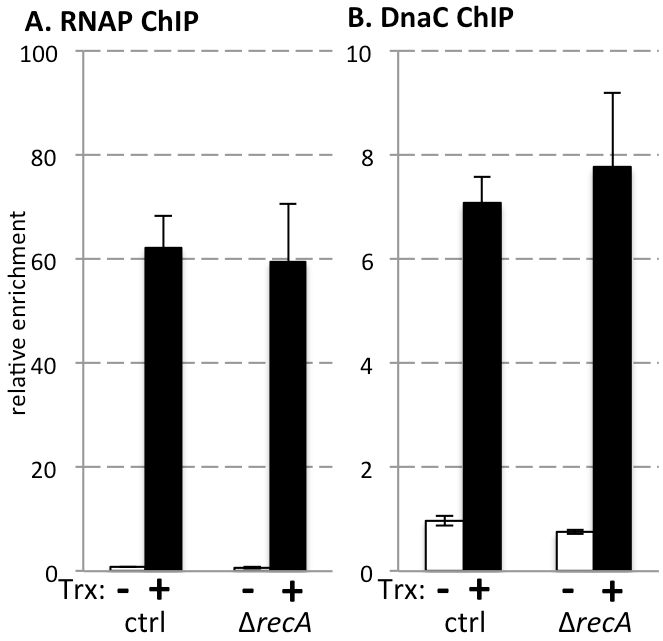


Figure 2.6 A) Relative association of RNA polymerase with the head-on gene-containing region. **B)** Relative association of the replicative helicase, DnaC, with the head-on gene-containing region. Data shown represents averages of 3 biological replicates. Error bars represent standard errors of the means.

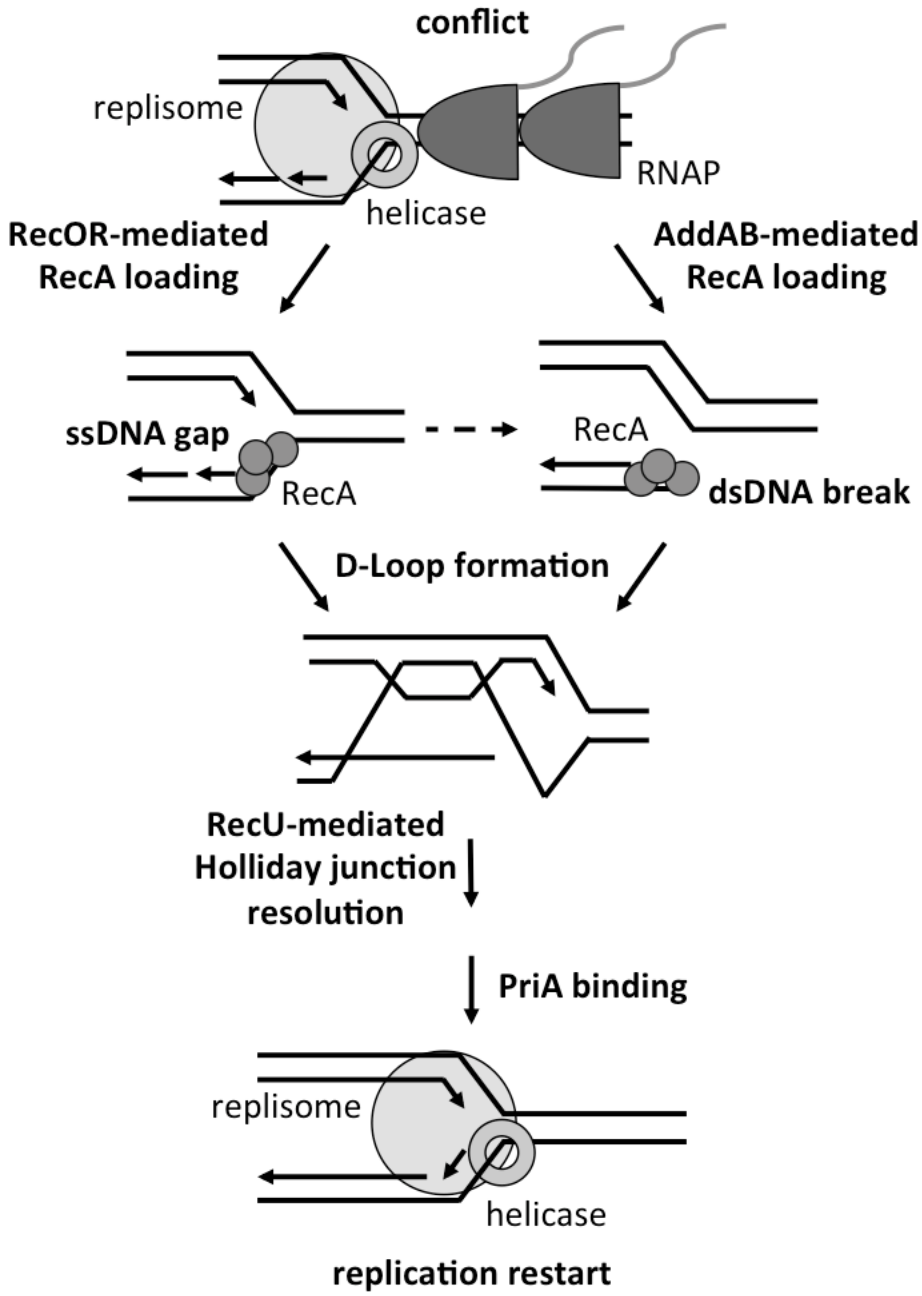


Figure 2.7 Model for restart upon replication-transcription collision in *B. subtilis*. RecFOR may recognize single stranded DNA at stalled forks, leading to RecA loading, recombination, and restart. Alternatively, double-stranded breaks in the DNA at stalled forks may be processed by AddAB, leading to RecA loading, recombination, and restart.

Chapter 3.

Transcription asymmetrically increases mutation rates specifically for lagging strand genes

A portion of the data contained in this chapter was previously published in *Nature*

Paul S, **Million-Weaver S**, Chattopadhyay S, Sokurenko E, Merrikh H. Accelerated gene evolution through replication-transcription conflicts. *Nature*. 2013 Mar 28;495(7442):512-5. doi: 10.1038/nature11989.

SP performed the bioinformatics analyses; SMW performed the mutation rate estimation experiments.

3.1 Introduction

Head-on conflicts between replication and transcription stall the replisome, necessitating restart and potentially breaking the DNA (1-3, 5-7, 18, 43). Presumably to minimize these encounters, the majority of genes in *B. subtilis* are encoded on the leading strand, co-directionally with replication. Despite the detrimental effects of lagging strand transcription, however, 26% of core genes and 6% of essential genes remain head-on to replication (16). We wondered if the lagging strand orientation of genes carried biological significance for *B. subtilis*.

Models to explain the co-directional biases observed in bacterial genomes invoke one of two selective pressures disfavoring transcription in the head-on orientation: interruptions to essential gene expression (67-69), or impediments to replication (43, 56, 60). Consistent with both of these possibilities, lagging strand genes are generally transcribed at low levels during laboratory growth conditions, minimizing the potential to disrupt either process.

Previous studies indicated that the lagging strand of the chromosome might experience increased mutagenesis (86, 153), though the underlying reason was unknown. Studies in *E. coli* and *P. putida* found that genes with changes in transcription orientation between strains experienced elevated mutation rates, although these differences were also associated with other large chromosomal rearrangements (90, 154). Work in *S. cerevesiae* identified altered mutation spectra in highly transcribed head-on reporter genes, though they did not observe any effect on gene orientation on mutation rates (19). While mutagenic repair mechanisms were known to show increased activity on the lagging strand (89, 153), the reason for this asymmetry historically evaded elucidation.

We set out to directly determine how gene orientation and expression influence mutation rates between the strands in *B. subtilis*. We identified the core and essential genes (155) from five available complete *B. subtilis* genome sequences and determined their orientation with respect to replication then calculated the mutation rates for genes on each strand. We found a significant increase in rates of non-synonymous substitutions for lagging strand genes, which correlated with gene length. We used classical reversion assays to measure mutation rates for an identical reporter gene in either orientation under both high and low transcription conditions. We found that transcription increased mutation rates for both orientations, yet to a two-fold higher degree when the gene was on the lagging strand. These results suggest a role for replication-transcription conflicts in increasing mutation rates for genes on the lagging strand.

3.2 Results

Highly conserved core genes on the lagging strand accumulate mutations at a higher rate than those on the leading strand

We analyzed mutation rates in core genes on the leading and lagging strand from five clonally related *B. subtilis* strains. Core genes were defined as loci present in all five strains with at least 95% identity and 95% length coverage. Of the 759 core genes identified, 132 were encoded on the lagging strand (**table 3.1).** Only six of the 148 identified essential core genes were encoded on the lagging strand, consistent with the described strong leading-strand bias for essential genes (16, 59, 66).

We compared the rates of synonymous (dS) and non-synonymous (dN) mutations for core genes in both orientations and found that the rate of amino-acid changing mutations was 42% higher on the lagging strand of replication (**figure 3.1**). The synonymous mutation rate was similar for each orientation, though lagging strand genes did accumulate silent changes at a 2% higher frequency. These results demonstrate that lagging strand genes accumulate mutations that alter their amino acid sequences at a higher rate than genes on the leading strand.

Gene length is positively correlated with non-synonymous mutation rates on the lagging strand

The increased non-synonymous mutation rate observed in head-on genes could arise because of the innate difference between how each strand is replicated (20, 30, 33, 87). If this were indeed the case and the differences in mutagenesis were caused by replication alone, mutation rates would be expected to be relatively constant among all genes on a given strand. If other factors, such as head-on conflicts with transcription, promote mutagenesis, then lagging-strand genes that experience these influences to an especially high degree should display further increased mutation rates. To determine whether conflicts with transcription contributed to mutagenesis, we analyzed the relationship between gene length and mutation rates for genes on both strands.

Gene length may be used as a surrogate for conflict severity because the prolonged time necessary to transcribe and replicate long genes increases the probability that the machineries encounter each other in these loci (156). Indeed, recent work in Eukaryotic cells demonstrated that collisions between replication and transcription potentially generate mutations leading to autism

spectrum disorders in exceptionally long genes whose transcription spans greater than a single cell cycle (15, 157). In prokaryotic genomes, long operons tend to localize to the leading strand of replication (16), presumably to minimize the likelihood of head-on collisions.

We found that lagging-strand genes are 48% shorter than those on the leading strand (459 vs. 681 base pairs, on average). Additionally, genes longer than 200 amino acids are overrepresented on the leading strand of replication, with only 26% of genes falling into this size category on the lagging strand (**figure 3.2a)**. These results are consistent with the previously noted biases against long genes localizing to the lagging strand. Strikingly, when we plotted the mutation rate per base pair as a function of translated protein length, we found that non-synonymous mutation rates were significantly positively correlated with length specifically for genes on the lagging strand of replication (**figure 3.2b**). The relationship between mutation rates and gene length was especially pronounced for genes encoding proteins greater than 200 amino acids.

Transcription asymmetrically increases mutation rates for head-on genes

The positive correlation with length for mutation rates of head-on genes suggested that replication alone could not account for the strand-dependent differences. While transcription in itself should be identical between the strands, the orientation with respect to replication distinguishes leading- and lagging-strand genes. To experimentally determine the contributions of transcription and gene-orientation to mutagenesis, we directly compared mutation rates of an inducible reporter gene integrated into the *B. subtilis* chromosome either head-on or co-directionally to replication under both high and low transcription conditions (**figure 3.3a)**.

When the reporter was expressed at low levels, mutation rates were identical for each orientation. High transcription conditions were associated with increased muatgenesis for both orientations, however the increase in mutation rates was two-fold higher when the gene was oriented head-on to replication, on the lagging strand **(figure 3.3b)**.

The increased mutagenesis of the head-on reporter gene could arise directly due to conflicts between the machineries. Alternatively, replication stress from the head-on collision itself at the mutation rate reporter gene could cause a shift in cell physiology to increase mutation rates genome-

wide, analogous to the SOS response to DNA-damaging conditions (99, 100, 104). To distinguish between these possibilities, I measured the spontaneous mutation frequency for the endogenous RNA polymerase beta subunit gene in cells harboring the mutation rate reporters in both orientations under high and low transcription conditions. I found that the rate of spontaneous mutations in *rpoB* conferring rifampicicn resistance was essentially identical for cells harboring the reporter gene in either orientation, under both high and low transcription conditions (**figure 3.4**). These results suggest that the transcription-dependent mutagenesis for the head-on orientation does not reflect a global increase in mutation rates.

3.2 Discussion

These results demonstrate that lagging-strand genes accumulate non-synonymous mutations at a higher rate than genes on the leading strand in *B. subtilis*. The positive correlation between mutation rates and gene length for lagging-strand genes suggest a role for conflicts with transcription in mutagenesis. Consistent with this model, the reversion assays demonstrate that expression asymmetrically increases mutagenesis specifically for genes in the head-on orientation. The underlying mechanism responsible for generating the mutations at regions of head-on transcription, however, remains unknown.

High-level reporter gene expression could cause metabolic imbalances, potentially inducing the stringent response (104, 158). Additionally, collisions between replication and transcription cause replication stress and single-stranded DNA accumulation, leading to a requirement for RecA for replication restart. Persistent RecA localization activates the SOS response (99, 159). Both the SOS and stringent responses lead to up-regulation of error prone polymerases and increased mutagenesis genome-wide (101, 102, 104, 105, 160, 161). However, because the frequency of mutations conferring spontaneous rifampicin resistance was neither influenced by expression nor orientation of the reporter genes, the observed asymmetric mutagenesis likely doesn't arise due to one of these global stress responses.

Transcription- and orientation- dependent mutagenesis on the lagging strand represents a mechanism for spatial modulation of mutation rates throughout the genome, distinct from other characterized mechanisms that promote variation in bacteria. The SOS and stringent responses increase spontaneous mutation rates genome-wide (101, 161-165), potentially causing detrimental mutations in essential genes (93, 166). Lagging strand genes, in general, are functionally enriched for involvement in stress response or interaction with the exterior of the cell; expression-dependent variation in these loci could play a role in allowing bacteria to adapt to changing environmental conditions (167).

An alternative explanation for the increased variability observed on the lagging strand of replication is simply that these genes, by nature, tolerate a higher degree of mutagenesis than those on the leading strand. Because lagging strand genes tend to only be expressed at high levels during times of stress (57, 58), selection may act infrequently at these loci. Under this argument, a simple

lack of selection against mutations in lagging strand genes drives the organization of bacterial genomes. However, because lagging strand genes play a role in the stress response, strong selection likely operates during conditions of their expression (92, 106, 168). Additionally, we found evidence for convergent substitutions and positive selection occurring in lagging strand genes (17). Although we cannot rule out the null hypothesis, increased variation in specific genes through conflict-mediated mutagenesis likely plays a role in shaping bacterial genomes.

Long-term evolution experiments measuring mutation rates between the strands in response to stress could help determine if conflict-mediated mutagenesis truly plays a role in adaptive evolution. The results outlined above clearly demonstrate increased mutation rates in lagging strand genes, and implicate transcription in causing the asymmetry. However, the specific footprint of mutations at lagging strand genes and the cellular factors responsible for increasing genetic variation at these regions are yet to be discovered.

Strain	Sequences Analyzed	All genes		Core genes	
		% Leading	% Lagging	# Genes Leading	# Genes Lagging
B. subtilis subsp. subtilis str. 168	4176	73	27	627	132
B. subtilis BSn5	4145	74	26		
B. subtilis subsp. subtilis str. RO-NN-1	4101	71	29		
B. subtilis subsp. spizizenii TU-B-10	4297	73	27		
B. subtilis subsp. spizizenii str. W23	4062	73	27		

Table 3.1 Distribution of genes between leading and lagging strand in 5 *Bacillus subtilis* strains used in the bioinformatics analysis.

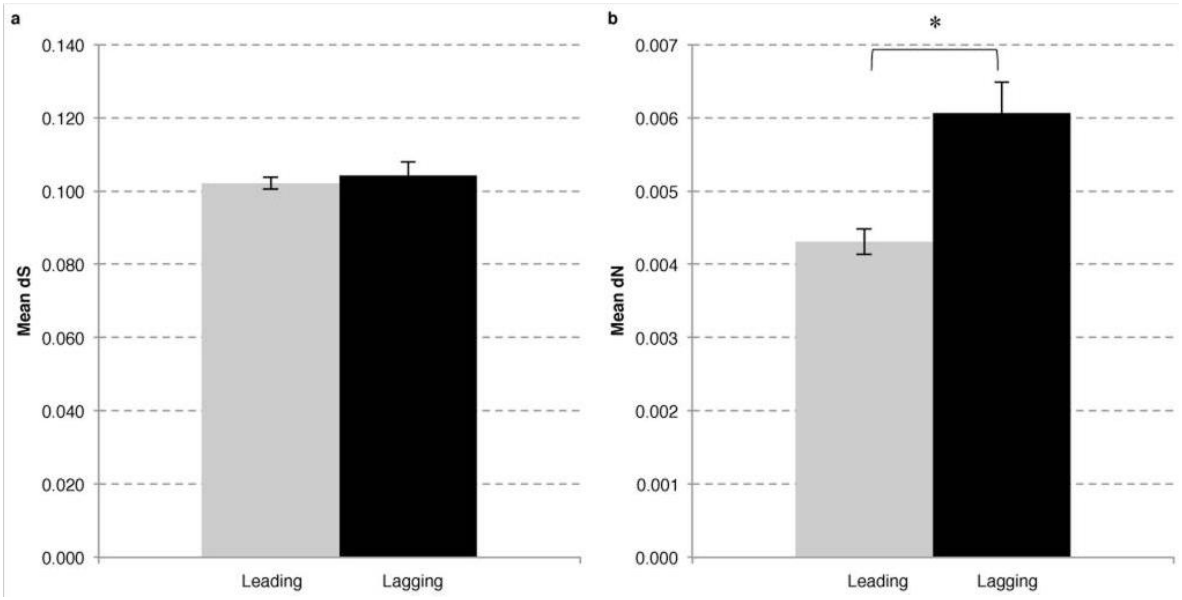


Figure 3.1 The mean values for the core genes in relation to diversity are presented. The dS (A) and dN (B) values are plotted for the leading (gray) and lagging (black) strands. Error bars represent the standard error of the mean. A statistically significant difference in dN between the two strands is detected with a $p<0.0001$ (*). Analysis of statistical significance was performed using the z-test for dS and dN values (see methods).

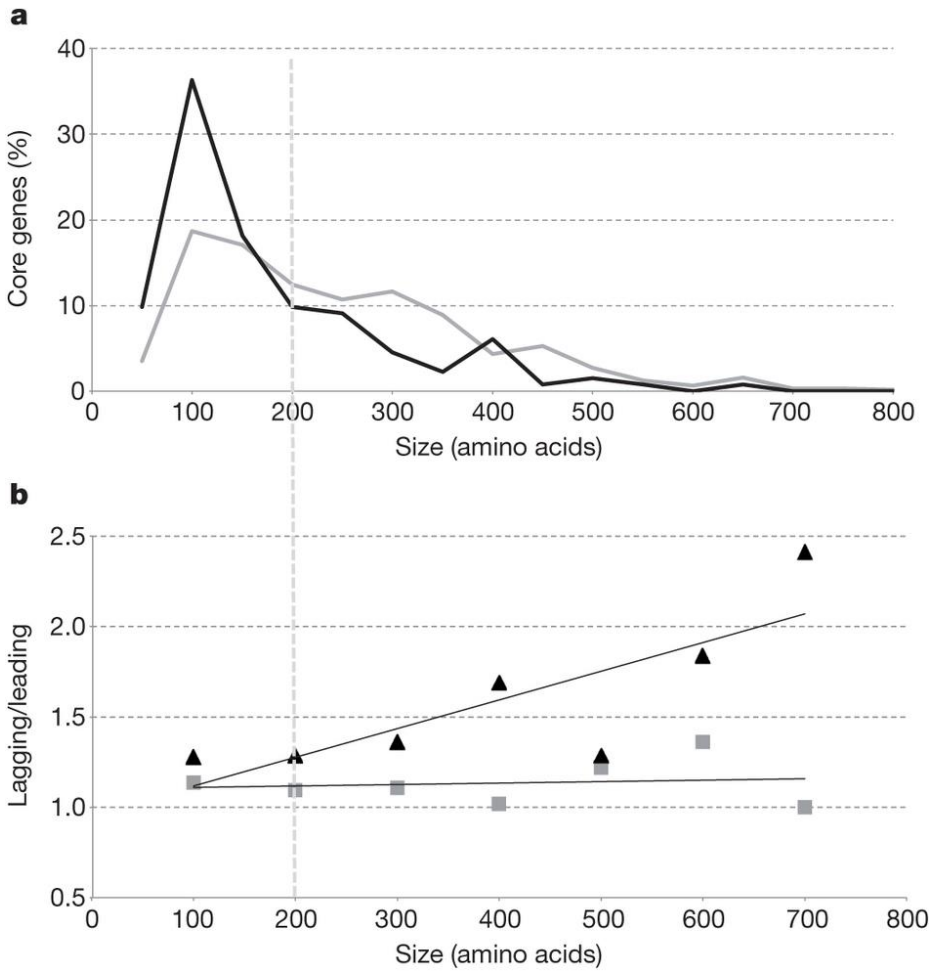


Figure 3.2 A) The percentage of genes in each size category for lagging (black) and leading (gray) strands within increasing windows of 50 aas. **B)** The fold difference between either dN (black) or dS (gray) for lagging compared to the leading strand genes in seven size categories (R2=0.65 for dN and 0.02 for dS, over length).

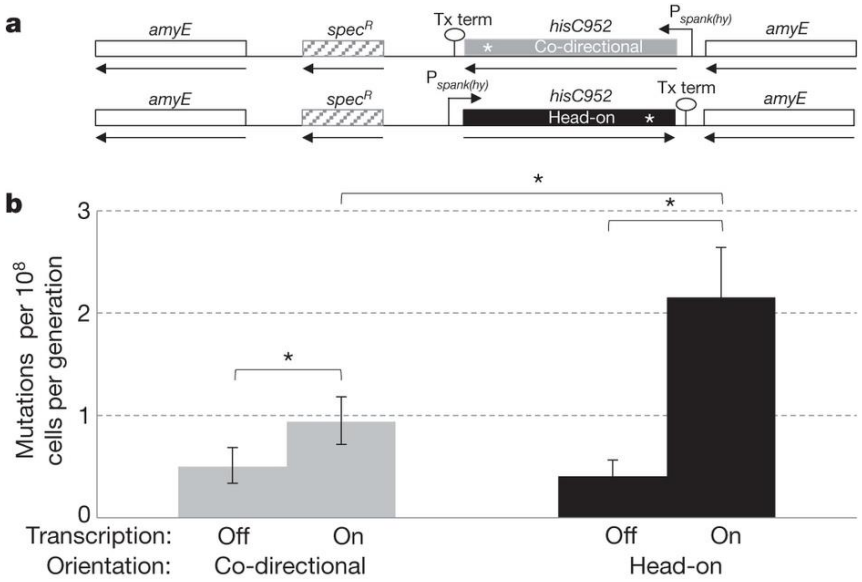


Figure 3.3 A) Diagram of the *amyE* locus containing reporter co-directionally (HM420) or head-on (HM419) to replication. **B)** Mutation rates based on reversion assays for *hisC952* when the gene was oriented head on or co-directionally to replication, in the presence (transcription on) or absence (transcription off) of IPTG (*p<0.05). Error bars represent 95% confidence intervals.

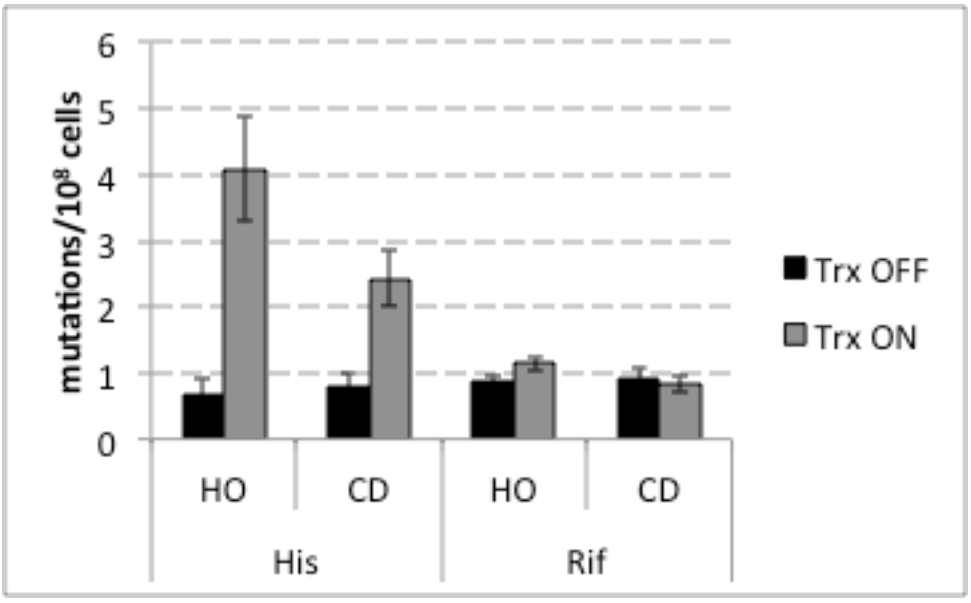


Figure 3.4 Head-on transcription does not increase the global spontaneous mutation frequency. Mutation rates based on reversion assays for *hisC952* (left, "His") and *rpoB* mutation frequencies based on Rifampicin resistance (right, "Rif") when the reporter gene was oriented head on or co-directionally to replication, in the presence (transcription on) or absence (transcription off) of IPTG. Error bars represent 95% confidence intervals, n=24.

Chapter 4.

An underlying mechanism for increased mutagenesis of lagging-strand genes in *Bacillus subtilis*

Originally published as an article in *Proceedings of the National Academy of Sciences*

Million-Weaver S, Samadpour AN, Moreno-Habel DA, Nugent P, Brittnacher MJ, Weiss E, Hayden HS, Miller SI, Liachko I, Merrikh H An underlying mechanism for increased mutagenesis of lagging-strand genes in *Bacillus subtilis*. *Proc Natl Acad Sci* U S A. 2015 Mar 10;112(10):E1096-105. doi: 10.1073/pnas.1416651112. Epub 2015 Feb 23.

4.1 Abstract

We previously reported that lagging-strand genes accumulate mutations faster than those encoded on the leading strand in *Bacillus subtilis*. Although we proposed that orientation-specific encounters between replication and transcription underlie this phenomenon, the mechanism leading to the increased mutagenesis of lagging-strand genes remained unknown. Here, we report that the transcription-dependent and orientation-specific differences in mutation rates of genes require the *B. subtilis* Y-family polymerase, PolY1 (*yqjH*). We find that without PolY1, association of the replicative helicase, DnaC, and the recombination protein, RecA, with lagging-strand genes increases in a transcription-dependent manner. These data suggest that PolY1 promotes efficient replisome progression through lagging-strand genes, thereby reducing potentially detrimental breaks and single-stranded DNA at these loci. Y-family polymerases can alleviate potential obstacles to replisome progression by facilitating DNA lesion bypass, extension of D-loops, or excision repair. We find that the nucleotide excision repair (NER) proteins UvrA, UvrB, and UvrC, but not RecA, are required for transcription-dependent asymmetry in mutation rates of genes in the two orientations. Furthermore, we find that the transcription-coupling repair factor Mfd functions in the same pathway as PolY1 and is also required for increased mutagenesis of lagging-strand genes. Experimental and SNP analyses of *B. subtilis* genomes show mutational footprints consistent with these findings. We propose that the interplay between replication and transcription increases lesion susceptibility of, specifically, lagging-strand genes, activating an Mfd-dependent error-prone NER mechanism. We propose that this process, at least partially, underlies the accelerated evolution of lagging-strand genes.

4.2 Introduction

Encounters between replication and transcription destabilize the genomes of organisms across all domains of life, causing DNA breaks, increased recombination, and mutagenesis. This genomic instability has been implicated in human fragile X syndrome, oncogene activation in cancer, and the generation of autism disorder-associated mutations (1, 3, 5, 14, 147, 157). Multiple strategies to modulate the effects of the interplay between replication and transcription have been identified, which are generally highly conserved across most life-forms.

Transcription occurs in two orientations with respect to replication: head-on, when a gene is coded for on the lagging strand, and codirectionally, when a gene is coded for on the leading strand of replication (**figure 1.2**). In either case, encounters between the two machineries can result in replication stress (8, 60). However, expression from the lagging strand is thought to be disfavored, because it can lead to increased replication stalling, replication restart, and genomic instability compared with codirectionally oriented genes (2, 4, 43, 60). Although head-on transcription is demonstrably more detrimental to the cell, the fundamental mechanism that underlies orientation-dependent severity remains a mystery. Additionally, in *Bacillus subtilis* (and other bacteria), a significant number of highly conserved genes (17%) and some essential genes (6%) remain on the lagging strand (17) and thus are transcribed head-on with respect to replication. Whether there are biological reasons for the head-on orientation of these genes is unclear.

We recently proposed that one reason for the maintenance of the head-on orientation of some genes may be to accelerate their evolution in a targeted manner. This model was based on evolutionary analyses of several different *B. subtilis* genomes, where we found that highly conserved head-on genes accumulate amino acid-changing mutations at a higher rate compared with codirectional genes (17). Furthermore, based on experimental evidence, we proposed that the differential rate of mutagenesis for the genes in the two orientations is due to the two types of replication–transcription collisions. How replication–transcription conflicts could lead to this orientation-specific mutational asymmetry is unknown.

Y-family polymerases are highly conserved across species and localize to stalled replication forks both in prokaryotes and eukaryotes (169-171). It is well known that Y-family polymerases

facilitate replication progression past bulky lesions on the DNA after damage (172-176). The lesion bypass activity of these polymerases is error-prone, leading to increased mutagenesis (105, 177, 178). Y-family polymerases lack proofreading and have large open active-site pockets. Y-family polymerases misincorporate nucleotides at a high rate and efficiently extend from mismatched primer termini, which cause them to copy undamaged DNA with low fidelity (179-181). These polymerases have been implicated in transcription-coupled nucleotide excision repair (TC-NER) in yeast, D-loop extension during recombinational repair, and stress-induced mutagenesis in stationary phase in *Escherichia coli* (101, 104, 160, 162, 182-184). Whether these polymerases impact mutation rates through any of these mechanisms in an orientation-dependent manner is unknown. *B. subtilis* has two Y-family error prone polymerases, PolY1 (*yqjH*) and PolY2 (*yqjW*) (185). It is unclear whether these polymerases play a role in DNA damage survival or in UV-induced mutagenesis in *B. subtilis* because the few reports on their functions under these conditions have been conflicting (185-188).

Here, we report that PolY1 is responsible for the higher mutation rates of genes on the lagging strand. Without PolY1, transcription is less mutagenic in the head-on orientation, and consequently, the mutation rate asymmetry of genes in the two orientations is alleviated. Furthermore, PolY1 reduces replication stalling and stress specifically in response to transcription of lagging-strand genes. We find that PolY1-dependent asymmetry in mutagenesis is epistatic with the transcription-coupled nucleotide excision repair pathway (TC-NER). In contrast, RecA and AddAB do not influence this mechanism. Furthermore, genomic SNP analyses are consistent with a Y-family polymerase contributing to the increased mutation rates of lagging-strand genes.

4.3 Results

PolY1 Is Required for Transcription and Orientation-Dependent Mutation Asymmetry.

Because the replication machinery is generally high fidelity (38), the majority of mutagenesis within cells, across the majority of species, is thought to arise from DNA damage and its subsequent error-prone bypass or repair (189-191). We wondered whether the observed transcription and orientation-dependent difference in mutation rates arises from differential error-prone mechanisms that deal with transcription-induced and possibly orientation-dependent DNA damage. Y-family polymerases are recruited to stalled replication forks and can facilitate both bypass and repair of DNA lesions in a mutagenic manner. *B. subtilis* harbors two error-prone polymerases, PolY1 and PolY2. We constructed and characterized deletion strains of both polymerases. We carried out survival assays in response to DNA-damaging agents and found that PolY1, but not PolY2, promoted cell survival in response to 4NQO and UV treatment, consistent with the Sung et al. report (186) (**figure 4.1**)

To investigate whether either PolY1 or PolY2 influence transcription-dependent mutation rates, we constructed three different sets of mutation rate reporters (*hisC952*, *metB5*, and *leuC27*) consisting of three different nonfunctional amino acid genes, which were inactivated by either a premature stop (*hisC952* and *metB5*) or missense codon (*leuC427*). Each reporter was placed under the control of the leaky isopropyl ß-D-1-thiogalactopyranoside (IPTG)-inducible promoter P*spank(hy)* (**figure 4.2a & b**). Ectopic copies of these nonfunctional alleles were introduced into strain backgrounds auxotrophic for each amino acid, encoded either on the lagging (head-on to replication) or leading (codirectional to replication) strand. Using these reporters, and the Ma–Sandri–Sarkar maximum-likelihood method, we estimated mutation rates for each reporter, at low and high transcription levels, in each orientation, with and without PolY1 or PolY2, and at two different locations on the chromosome [*amyE*, right arm of the chromosome; and *thrC*, left arm of the chromosome (**figure 4.3a & b**)

Induction of transcription led to increased reversion rates for all three constructs in both orientations (**figure 4.2c-e and figure 4.3b-d**). This orientation and transcription-dependent increase in mutagenesis does not stem from differences in growth rate of these strains under the different conditions examined (**figure 4.**4) The absolute degree of transcription-dependent increase in mutation rates varied depending on the reporter used, although the underlying reason is unclear and could be

due to the sequence context of the particular inactivating mutation. These differences are not due to altered levels of transcription as they are driven by the same promoter, and quantification of mRNA levels confirms that they are expressed to the same degree (**figure 4.**5) Despite the differences in absolute levels of mutagenesis, the transcription-dependent increase in mutation rates for all three constructs examined was about twofold higher when the gene was oriented head-on to replication (**figure 4.2c-e and figure 4.3b-d**).

At low transcription levels, deletion of PolY1 did not impact mutation rates for any of the reporters examined, regardless of orientation. However, in cells lacking PolY1, transcription-induced mutation rates of all three reporters in the head-on orientation were reduced by about 50%, leading to a loss of transcription-dependent mutation asymmetry (**figure 4.2c-e and figure 4.3b-d**). These effects were specific to PolY1: we did not observe significant effects of PolY2 on transcription-dependent mutation rates (**figure 4.6**). Furthermore, deletion of PolY1 had no effects on mRNA levels or RNA polymerase (RNAP) occupancy of the reporter genes, indicating that PolY1's impact on mutation rates is not due to alterations in gene expression or RNAP occupancy (**figure 4.7**). These results strongly suggest that the increased mutation rates of lagging-strand genes are due to a transcription and orientation-dependent activity of PolY1 on the lagging strand.

PolY1 Reduces Replication Stalling at Regions of Lagging-Strand Transcription.

Because transcription alone should be identical in both orientations, the interplay between transcription and replication is likely responsible for PolY1's orientation-dependent activity. We reasoned that, regardless of whether PolY1 facilitates lesion bypass, excision repair, or recombination, lesions or breaks will increase at regions of lagging-strand transcription without PolY1. The increased obstacles should lead to an increase in replication stalling at these regions without PolY1.

We tested this hypothesis using the P*spank(hy)-hisC* reporter described above (at *amyE*) and constructed an additional reporter system, consisting of a copy of the *lacZ* gene, under the control of a different promoter, P*xis*. This engineered construct was inserted at the *thrC* locus, either head-on (8) or co-directionally to replication. P*xis* is tightly repressed by the ImmR protein, which is coded for by the mobile genetic element ICE*Bs*1 (192). We introduced the P*xis-lacZ* constructs into strains that

either harbored (transcription off) or were cured of (transcription on) ICE*Bs1*. To determine whether PolY1 impacts replication stalling or restart, we used chromatin immunoprecipitations (ChIPs) to measure the relative association of the replicative helicase (DnaC) with these conflict regions, compared with several different control loci. Increased relative association of the replicative helicase, DnaC, with a particular chromosomal locus is likely indicative of replisome stalling or restart at that region (8). In an asynchronous population of cells, if replication progresses at a similar rate throughout the chromosome, then replisome proteins, such as DnaC, are not expected to preferentially associate with any given locus. However, the replisome spending an increased amount of time at a given region (i.e., stalling) can be detected as preferential association of replisome proteins with that locus, relative to control loci.

At low transcription levels, we did not detect preferential association of DnaC with any of the examined engineered conflict regions compared with the control locus, *yhaX*, or other control regions 10 kb upstream or downstream of the constructs (**figure 4.8a & b** and **figure 4.9a**). Although we find normalized values overall to be most consistent between experiments, to rule out potential artifacts of normalization, we also analyzed nonnormalized, raw immunoprecipitation (IP)/total values (**figure 4.9b & c**). As expected, induction of transcription led to increased relative association of DnaC with the engineered conflict regions, for both orientations, but to a higher degree when the reporters were oriented head-on to replication. In the absence of PolY1, relative association of DnaC with both *hisC* and *lacZ* increased about twofold, and to different degrees when the data were plotted without normalization as IP/total (**figure 4.8a & b** and **figure 4.9b & c**). This increase, i.e., the effect of PolY1, was specific to the head-on constructs. We also analyzed any impact of PolY1 with highly transcribed endogenous loci: the ribosomal protein-coding genes *rplGB* (codirectional) and *rpsD* (head-on). We found that, similar to the results from the engineered constructs, without PolY1, DnaC association increased at the head-on gene, *rpsD*, but not the codirectionally oriented *rplGB* gene (**figure 4.8c**). These data strongly suggest that PolY1 aids replisome progression through, specifically, regions of lagging-strand transcription.

To rule out the possibility that DnaC localization to the engineered conflict regions is due to a replication-independent event, such as DNA repair, we treated cells with the highly potent DNA

polymerase inhibitor 6(*p*-hydroxyphenylazo)-uracil (HPUra), which specifically arrests DNA replication (193). In cells inhibited for replication with HPUra, we no longer detected preferential association of DnaC with the highly transcribed P*xis-lacZ* gene (**figure 4.10**) irrespective of orientation or the presence of PolY1. The dependence of DnaC association with the conflict region on DNA replication argues against the possibility that the helicase is operating in a replication-independent context at highly expressed genes in our experiments.

PolY1 Reduces Genomic Instability Caused by Lagging-Strand Transcription.

Because stalled replication forks can lead to breaks, the results of the DnaC ChIPs presented above suggest that PolY1's activity at these regions should reduce genomic instability. Roadblocks to replication can lead to the accumulation of single-stranded and/or broken DNA that is bound by RecA, as the first step in recombinational repair. RecA-GFP focus formation in microscopy experiments (135, 194) is traditionally used as a measure of replication stress. We obtained a previously characterized RecA-GFP allele (135) and, as a control, determined that strains harboring the construct do not display significant growth defects under conditions where P*xis-lacZ* is highly expressed head-on to replication (**figure 4.11**). Given that deletions of *recA* are associated with severe growth defects under these conditions, the data from the growth curves of the RecA-GFP strains suggest that the fluorescent tag does not significantly impact RecA's function at forks stalled due to encounters with transcription. Using this construct, we quantified RecA-GFP focus formation in strains harboring either the P*xis-lacZ* or P*spank(hy)-hisC* reporters, at low and high transcription levels, for both orientations, with and without PolY1. In strains with repressed transcription from these constructs, RecA-GFP foci formed in 10–20% of cells, depending on the strain background, regardless of the orientation of the reporter gene (**figure 4.12a & b** and **figure 4.13**). This result aligns well with reported basal RecA localization levels in both *B. subtilis* and *E. coli* (135, 194), although the exact sources for these basal levels of replication stress in cells are unclear. The absolute numbers of cells containing RecA-GFP foci were different in strains containing the *lacZ* and *hisC* reporters (10% vs. 20%, respectively). This difference is probably due to the differences in the strain backgrounds used for each set of reporters. Under conditions where transcription was high from either the P*xis* or the P*spank(hy)* promoter, we observed

a statistically significant increase in RecA-GFP focus formation, only when the gene was expressed on the lagging strand, head-on to replication (**figure 4.12a & b** and **figure 4.13**). In isogenic strains lacking PolY1, we found that significantly more cells display RecA-GFP foci when P*xis-lacZ* [and modestly, P*spank(hy)-hisC*] was highly transcribed in the head-on, but not the codirectional orientation. The engineered constructs are expressed at significantly higher levels than most endogenous lagging-strand genes, creating conflicts that are much more severe than those that exist normally during fast growth. Most endogenous lagging-strand genes are only highly induced in response to different stresses and/or during spore outgrowth, when DNA replication is slow (195). Therefore, during fast growth, it is unsurprising that these artificial constructs cause a significant increase in genomic instability

To determine whether the increased RecA focus formation reflects replication stress emanating specifically from the engineered reporter regions, we used ChIPs to measure the relative enrichment of RecA with the two sets of engineered conflicts [P*spank(hy)-hisC*, and P*xis-lacZ*] compared with several different control loci. At low transcription levels, we did not detect preferential association of RecA with the reporter regions compared with control loci (**figure 4.12c & d** and **figure 4.14**). At low transcription levels, the relative association of RecA with the examined loci was not impacted by the absence of PolY1. Induction of transcription led to increased relative (and absolute, nonnormalized) association of RecA with these regions, for both orientations, to a higher degree when the reporters were expressed on the lagging strand (**figure 4.12c & d** and **figure 4.14**). The relative association of RecA with both P*spank(hy)-hisC*, and P*xis-lacZ*, compared with the control loci, in PolY1-deficient backgrounds further increased when transcription was induced from the lagging strand. We found similar effects of PolY1 on RecA association with highly transcribed endogenous loci: without PolY1, RecA association increased with the lagging-strand–encoded ribosomal protein gene, *rpsD*, but not with the codirectionally oriented *rplGB* gene (**figure 4.12e**). Altogether, these data indicate that PolY1 reduces genomic instability and/or replication stress at regions of lagging-strand transcription, most likely due to an effect of replication.

PolY1 Activity at Lagging-Strand Transcription Regions Is Independent of RecA and AddAB.

In *E. coli*, recombinational break repair by Y-family polymerases is proposed to be mutagenic (83, 184, 196). Therefore, it is possible that PolY1 activity during the repair of breaks generated by conflicts at lagging-strand genes causes the increased mutagenesis associated with head-on transcription. Alternatively, PolY1 could act upstream of break repair.

To determine whether PolY1 acts to prevent or repair breaks, we used the P*spank(hy)-hisC952* reporter to measure transcription-induced mutation rates, in both orientations, in strains lacking either RecA or AddAB. Similar to RecA, AddAB is required for double-stranded break repair in *B. subtilis* (119, 197). We determined that transcription was still mutagenic, in an asymmetric manner, even when AddAB or RecA were not present (**figure 4.15a & b**). Furthermore, in the *addAB* and *recA* deletion backgrounds, the asymmetric nature of transcription-dependent mutagenesis still depended on PolY1. These results suggest that PolY1 acts independent of RecA and AddAB, most likely upstream of D-loop formation and extension, as well as double-stranded break repair. Furthermore, because RecA is required for the SOS response, which can induce expression of Y-family polymerases in some organisms (198), these results indicate that the SOS response is not required for the effects of PolY1 on the orientation and transcription-dependent mutagenesis. This is consistent with the previous finding that, unlike the *E. coli dinB*, *polY1* transcription is not induced in response to DNA damage (185).

PolY1 Functions in the Same Pathway as TC-NER.

The eukaryotic error-prone polymerase Polκ, which is homologous to PolY1, has been shown to fill in gaps generated by the NER proteins (169). The first step of NER in prokaryotes is DNA binding by UvrA. UvrA subsequently recruits UvrB and UvrC, leading to excision of the damage-containing region and subsequent gap fill-in (199). Previous studies have connected the Y- and B-family polymerases to transcription-coupled repair (TCR), which can recruit excision repair proteins to stalled RNAPs (19, 183, 200). Although a study showing orientation-dependent effects of NER in yeast observed a change in mutation spectra (19), to our knowledge, gene orientation-dependent changes in mutation rates by alternative polymerases have not been reported in any organism. Putting all of these reports together, and considering the *recA* and *addAB* data presented above, we hypothesized that the orientation-dependent mutation asymmetry is due to PolY1 activity in NER.

To test this model, we obtained (from the *Bacillus* Genetic Stock Center) a deletion strain lacking *uvrA* and confirmed its genotype via PCR. Using the P*spank(hy)-hisC952* reporter to measure transcription-induced mutation rates, we found that similar to the *polY1* deletion strain, without UvrA, transcription-induced mutations were no longer asymmetric (**figure 4.16a**). As we found in the *polY1* deletion strain, the deletion of *uvrA* specifically reduced mutation rates when the reporter was expressed from the lagging strand, but not from the leading strand. To determine whether UvrA and PolY1 act in the same pathway, we constructed a double mutant lacking both proteins. We found that the two deletions did not produce an additive effect: they were epistatic to each other (**figure 4.16a**). To confirm that the effect of UvrA and PolY1 on mutagenesis in lagging-strand genes was through NER, we estimated mutation rates in backgrounds lacking UvrB and UvrC. Consistent with the well-established roles of these proteins downstream of UvrA in NER, we found that, similar to the *uvrA* and *polY1* deletion strains, mutations were no longer asymmetric without UvrB or UvrC (**figure 4.17**). These experiments demonstrate that UvrA and PolY1 act in the same pathway, most likely carrying out NER specifically in response to transcription from the lagging strand.

NER proteins can be recruited to transcribed regions through TCR. TCR is initiated when RNA polymerase stalls, typically at a bulky lesion (201). Stalled RNA polymerases are canonically recognized by the transcription–repair coupling factor, Mfd, which recruits UvrA, initiating NER (202) [studies in *E. coli* indicate that the accessory helicase UvrD may also facilitate this process (203)]. We hypothesized that the NER-dependent mutational asymmetry we observe in transcribed lagging-strand genes is due to TCR-mediated recruitment of excision repair proteins. To test this model, we constructed a deletion strain of *mfd* and estimated transcription-induced mutation rates for the P*spank(hy)-hisC952* reporters. We observed that, similar to the strains deficient in NER proteins (and PolY1), without Mfd, mutation rates were no longer asymmetric (**figure 4.16b**). Furthermore, the *mfd* deletion was epistatic to the *polY1* deletion, confirming that these proteins act in the same pathway (**figure 4.16b**). Altogether, these data strongly suggest that Mfd recruits NER proteins. This recruitment, through the function of PolY1, leads to increased mutagenesis of transcribed lagging-strand genes.

Genomic Mutational Footprints Are Consistent with Orientation-Dependent PolY1 Activity.

The mutational signature of bacterial, archeal, and eukaryotic Y-family polymerases has been examined both in vivo and in vitro. In general, Y-family polymerases predominantly make −1 frameshifts, T→C, T→G, and A→G mutations in all organisms examined (103, 179, 181, 204-210). Our analyses of the PolY1-dependent base substitutions leading to prototrophy [multiple substitutions are tolerated at this position (**figure 4.18**) in the mutation rate assays described above are consistent with the reported mutational signatures of these polymerases (**figure 4.18**). Consistent with reports of other Y-family polymerases and our mutation rate measurements, we find that transcription-dependent T→C mutations are partially PolY1 dependent. It is important to note that detectable mutations are limited to changes in the sequence of the reporters that would allow for functionality. For example, frameshifts, or for the *hisC* reporter, a G-to-A substitution (which would create another stop codon) cannot be detected in these assays.

To determine whether a conflict-related and PolY1-dependent mutational signature is present in *B. subtilis* genomes, we identified SNPs within highly conserved (90% similarity at the nucleotide level) "core" genes from the available *B. subtilis* genomes. We then determined the rate of accumulation of each type of SNP, in relation to protein length, for both strands. Our group, as well as others, has used gene length as a surrogate for conflict frequency (15, 17), because replication and transcription machineries both spend an increased length of time traversing longer genes, increasing the chances for collisions. Because the analysis compares SNPs within species, the ancestral identity of a given nucleotide is unknown; thus, it is not possible to determine the directionality of the mutations. For example, T→C and C→T mutations cannot be distinguished and must be grouped together. Nevertheless, the potential contribution of Y-family polymerases can, at least partially, be deduced from the observed patterns. The SNP analyses indicate that T↔C and A↔G are the most commonly found base substitutions, consistent with the well-known bias toward transitions in bacterial genomes (211-214). Remarkably, for all types of base substitutions, we found that both SNP counts, and mutation rates displayed a highly statistically significant nonlinear increase with gene length (*P* values of <0.001 and <0.007) (**figure 4.20**, **figure 4.21** and **table 4.1**). A linear increase with gene length would suggest that SNPs arose due to random chance. The nonlinearity of the correlation of mutation rates with gene length, however, indicates that active mechanisms influenced their occurrence. All

SNP groups were identically correlated with gene length between the two orientations, except for the TC/CT group, which showed a modest, but highly statistically significant, increased correlation for genes on the lagging strand (25% vs. 20% increase per kilobase; $P = 0.005$). The AG/GA group also displayed a slightly increased correlation with length for lagging-strand genes, although this difference was not statistically significant. The TC/CT group, followed second by the AG/GA group, includes the most preferred substitutions made by Y-family polymerases, in that order.

Although small, the stronger positive correlation of these particular SNPs with gene length in lagging vs. leading-strand genes is consistent with orientation- and transcription-dependent PolY1 activity. It is important to note that the observed differences may be small because our analysis can only measure the fixed SNPs, and not those that have been eliminated through purifying selection. Therefore, the difference observed is not necessarily representative of the actual orientation-dependent differences in the rates of these mutations. However, these results are significant in the context of our findings as they demonstrate an orientation-dependent difference in the relevant SNPs.

4.4 Discussion

Our results show that the Y-family polymerase, PolY1, aids replisome progression through transcription units on the lagging strand. Whether this occurs directly during head-on conflicts at these regions, or indirectly through reduction of lesions on the lagging strand left behind by the replication forks, is unclear. However, it is clear that this mechanism is mutagenic and is in the same pathway as both UvrA, a protein required for NER, and Mfd, the NER–transcription coupling factor. Our work brings previously unidentified insights into the consequences of transcription on the lagging strand, and different strategies for cells to overcome transcription-related problems. Overall, the results presented here identify an underlying mechanism that likely contributes to the accelerated evolution of lagging-strand–encoded genes.

Models of Orientation and Transcription-Dependent PolY1 Activity in NER.

Because the impacts of PolY1, UvrA, and Mfd are transcription- and orientation-dependent, our data indicate that the mechanism we uncovered involves both replication and transcription. We propose two different models of orientation-dependent PolY1 function in TC-NER (**figure 4.22**). These two models are not mutually exclusive.

In the first model, lagging-strand transcription leads to head-on conflicts with the replisome. This encounter would expose single-stranded DNA (ssDNA), either due to excess positive supercoils between the two machineries, and/or sustained transcription bubbles. ssDNA is known to be more susceptible to DNA damage and would therefore lead to increased lesions in these regions. Stalled RNAPs at the site of the conflict activate TCR through recognition by Mfd, recruiting NER proteins UvrABC. After excision of the lesion-containing patch by NER, the remaining gap is filled-in by PolY1.

An alternative model, which also depends on the activity of PolY1 in TC-NER, as well as both replication and transcription, is a mechanism where TCR occurs post-replication. The discontinuous nature of lagging-strand synthesis increases the lesion susceptibility of exposed ssDNA. When the orientation of a gene changes with respect to replication, the template strand of transcription also changes (**figure 1.2**). Hence, when a lagging-strand–encoded gene is expressed, the transcription machinery uses the lagging strand as the template, increasing the number of DNA lesions encountered

by RNAPs. Because TCR is thought to be more active on the template strand, NER would be recruited more frequently during lagging-strand transcription. As in the first model, NER would then be completed by PolY1-mediated gap filling.

Dealing with Replication–Transcription Conflicts.

Cells have multiple strategies to overcome replication–transcription conflicts. Recently, it has become clear that the varied consequences of these collisions necessitate different resolution mechanisms. *E. coli* accessory helicases, UvrD and Rep, promote survival of cells harboring severe head-on conflicts at inverted rDNA loci (10). DksA and other transcription-modulating factors (GreA/B) also ease severity of replication–transcription conflicts in *E. coli* (11, 12, 53). Helicases and transcriptional modulators presumably facilitate replication progression by removing the RNAP obstacle from the replisome's path. Whether these strategies are universal among bacteria is unclear. For instance, homologs of DksA or GreB do not exist in *B. subtilis*. The potential role of *B. subtilis* accessory helicase PcrA, the homolog of UvrD and Rep, in replication–transcription conflicts requires further investigation.

If the PolY1-dependent NER mechanism takes place during head-on conflicts, this represents a distinct strategy for relieving adverse consequences of replication–transcription conflicts. Rather than modulating the obstacle and clearing the path for replication, these polymerases are likely involved in replicating the DNA at the site of the conflict. Conflicts cause genomic instability both by physically hindering replisome progression, as well as by jeopardizing the genomic integrity of the DNA itself at the conflict region. To overcome conflicts, cells must remove the obstacle and either repair or bypass the damaged template. A failure to repair the lesions on the DNA by TC-NER would increase ssDNA and/or breaks at the site of the conflict. Our data suggest that PolY1 acts upstream of RecA-dependent repair, removing lesions, and consequently, either directly or indirectly reducing conflict-induced insults on the integrity of the DNA at these regions.

Sources of Transcription-Induced Mutagenesis.

Our results suggest that PolY1 is involved in NER-dependent transcription-induced mutagenesis in an orientation-dependent manner, unraveling an underlying reason for the increased mutagenesis of lagging-strand genes. However, even without PolY1, a significant amount of transcription-induced mutagenesis is detected in our assays, for both orientations, suggesting that other mechanisms are also at play. There are many proposed models that could explain the sources of transcription-dependent mutagenesis. Some of these models include the following: recombination, DNA supercoiling, R-loops (RNA:DNA hybrids), non-B DNA, DNA damage, and rNTP incorporation (3, 15, 45, 215-221). These mechanisms are not mutually exclusive and may be contributing to transcription-dependent mutations collectively, and under different conditions. In our system, RecA and AddAB do not significantly influence transcription-dependent mutagenesis, arguing against the recombination model. Rather, the impact of UvrA, UvrB, UvrC, and consequently NER, supports DNA lesions as at least one of the main sources of transcription-associated mutagenesis. This is consistent with most of the other models proposed because generally these mechanisms involve exposed ssDNA at some stage [recombination, supercoiling, R-loops (by sustaining or possibly stabilizing transcription bubbles), etc.], which is more susceptible to DNA lesions.

Potential Sources of DNA Lesions in Lagging-Strand Genes.

NER canonically recognizes the helix-distorting properties of bulky lesions on DNA. Bulky lesions can also stall RNA polymerase, leading to the activation of TCR. Our results demonstrate that NER, and the Y-family polymerase PolY1, increase transcription-dependent mutagenesis in lagging-strand genes, suggesting that these regions are particularly susceptible to DNA lesions. Consistent with our findings, the classic observation that a significant amount of DNA turnover (up to 0.02% of the genome) occurs in cells unexposed to exogenous DNA damage hints at repair of basal DNA damage (222). Although determining the exact chemical structure of the predicted DNA lesions at regions of conflict between replication and transcription is beyond the scope of this study, recent reports, as well as classic publications, suggest several different cellular sources of damage potentially leading to NER.

Evidence continues to mount that NER acts on multiple substrates in addition to UV-induced pyrimidine dimers. The spontaneous mutation frequency of *uvrA*-deficient *E. coli* strains shows a

significant reduction in T→C transitions (223), consistent with NER and gap fill-in by PolY1 at endogenously damaged DNA. NER may also play a role in removing misincorporated ribonucleotides from DNA, as has been observed in *E. coli* (224). The NER pathway in yeast efficiently processes *N*-6-methyladenine lesions (225, 226) and is also recruited to abasic sites in the transcribed strand through TCR (227). Additionally, NER and TCR play a role in protecting mitochondrial DNA from oxidative lesions, such as 8-oxo-dG (228) or bulky thymine glycol lesions, which distort the double helix and block transcription (229-231).

An alternative model for the mechanism causing excision repair-mediated mutagenesis at regions of lagging-strand transcription is the provocative possibility of "gratuitous TCR," or Mfd recruitment of NER proteins to RNA polymerases stalled independently of damage. The UvrABC complex (and its human homologs) does excise short patches of undamaged DNA in vitro (232, 233). It is not outside the realm of possibility that Mfd recruitment to RNA polymerases stalled subsequent to head-on collisions with the replisome could activate TCR at undamaged DNA in lagging-strand genes. However, we favor the model that increased NER substrates, such as bulky lesions on DNA or abasic sites, underlie lagging-strand–specific TC-NER. The sources of endogenous DNA damage, and cellular factors participating in this process, deserve further study and will likely yield interesting results as detection techniques improve.

The Role of Y-Family Polymerases in Accelerated Evolution of Lagging-Strand Genes.

An extensive body of work has connected Y-family polymerases to stress-induced mutagenesis (101, 102, 104, 162, 182). These studies have shown stress-dependent elevation of mutation rates in non-growing cells. PolY1's activity in replicating cells expands its role in generating genetic diversity beyond stationary phase and shows that Y-family polymerases are a versatile group of proteins driving potentially adaptive mutagenesis at all stages of life. Furthermore, although other studies have shown transcription-dependent increases in mutation rates, in both eukaryotes and prokaryotes (19, 234), the proposed mechanisms would impact all genes and therefore would not target variation to specific genes. We previously proposed that transcription together with gene orientation can increase diversity in certain genes, rather than impacting global mutation rates (17). This mechanism would allow cells to

maintain globally low mutation rates the majority of the time and for the majority of the genome, as proposed (91, 166). However, under conditions where adaptation can be beneficial, cells may increase the mutation rates of specific, head-on–oriented genes, merely by increasing their expression. The PolY1- and NER- dependent mutagenesis of, specifically head-on genes, reduces the chances of potentially deleterious mutations in highly expressed (i.e., rDNA and tRNA) and essential genes, which are strongly biased to be co-oriented with replication.

4.5 Figures

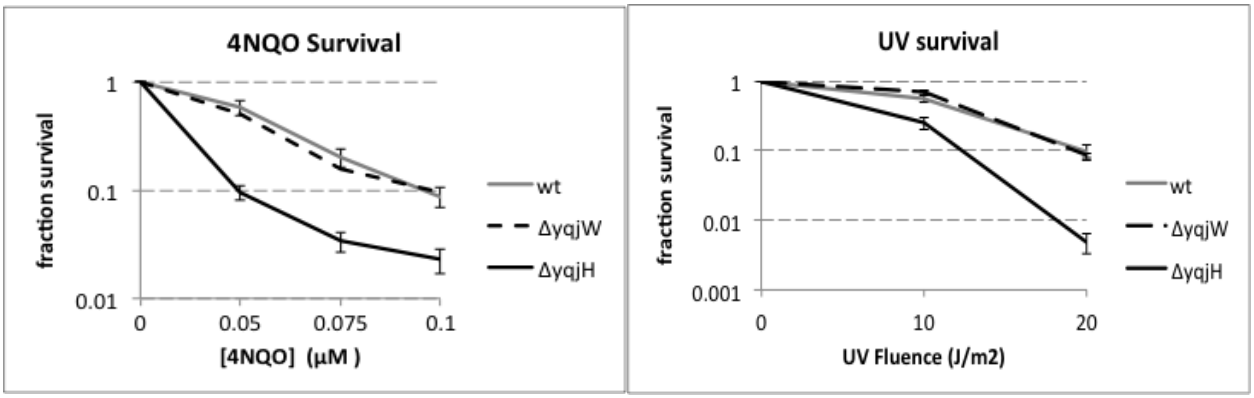


Figure 4.1 Cells lacking PolY1 are sensitized to UV and 4NQO. Survival assays were performed in wild type (HM1, gray lines), Δ*yqjH* (HM391, black lines), and Δ*yqjW* (HM425, dashed lines) strains as described in **supplementary methods**. Data shown is the average from 8 (UV) or 12 (4NQO) biological replicates, error bars represent standard deviations.

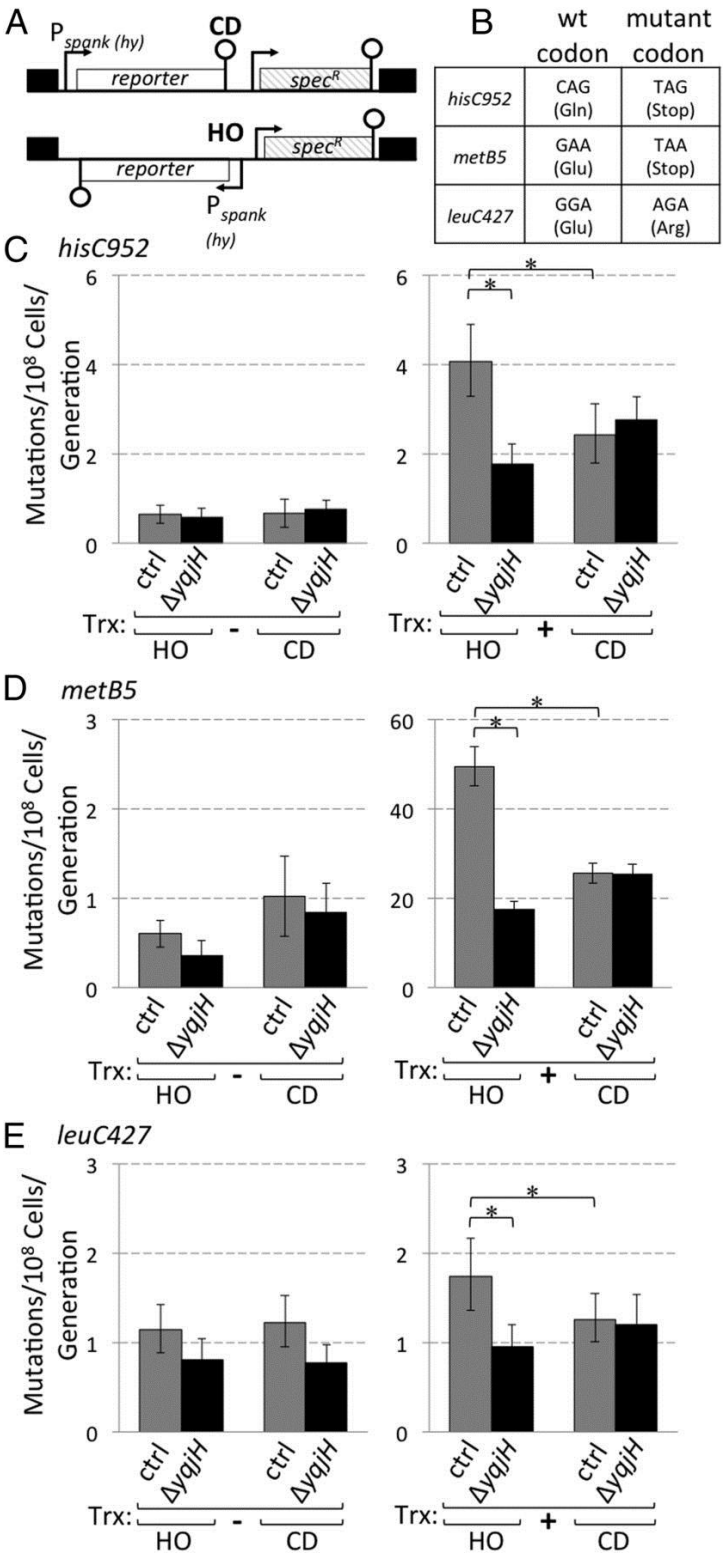


Figure 4.2 The transcription-dependent mutation asymmetry requires PolY1. (*A*) Schematic of the *amyE* locus harboring mutation reporters oriented head-on or codirectionally to replication. (*B*) Wild-type and mutant stop codons for the *hisC952*, *metB5*, and *leuC427* mutation reporters. (*C*) Mutation rates for the *hisC952* reporter in the presence (HM419 and HM420; gray) or absence of PolY1 (*ΔyqjH*) (HM421 and HM422; black), for the head-on and codirectional orientations, with (Trx+) and without (Trx−) IPTG. (*D*) Same as *C* but for the *metB5* reporter strains. (*E*) Same as *C* but for the *leuC427* reporter strains. Mutation rates were estimated based on C = 36–48 for each strain and condition. Error bars are 95% confidence intervals (**P* < 0.05). Significance was determined by Student's *t* test.

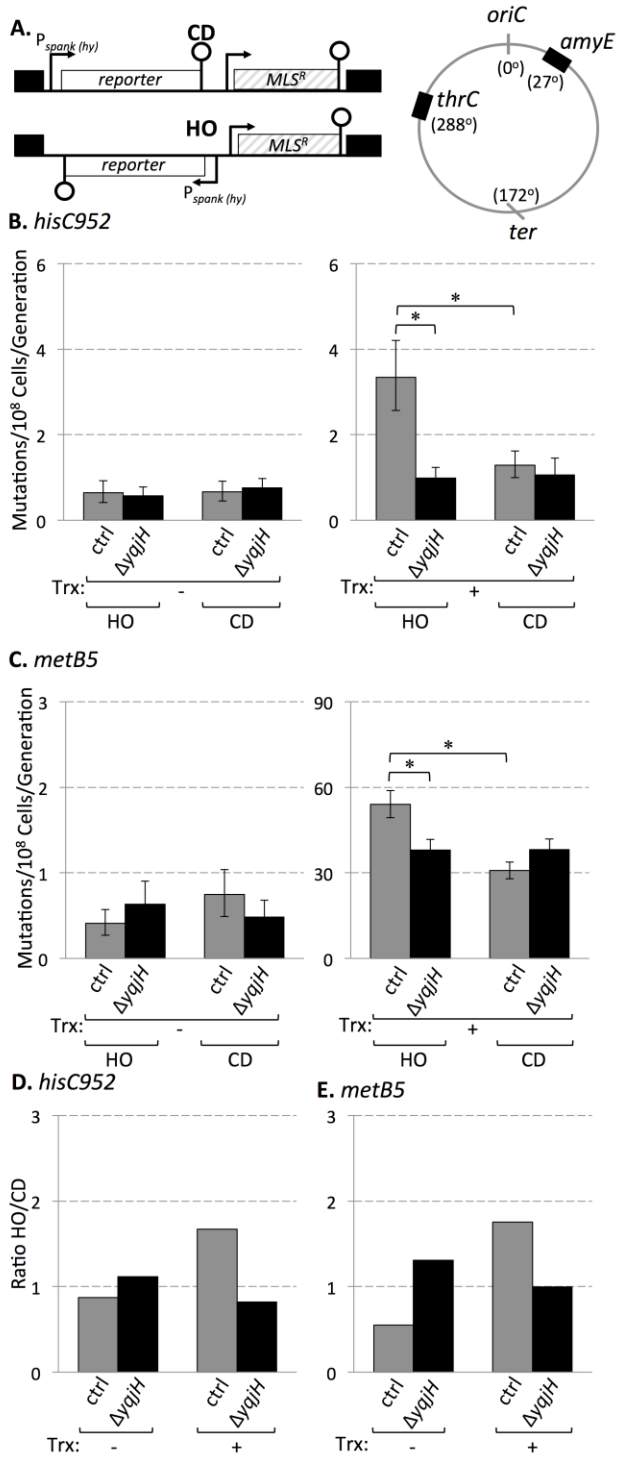


Figure 4.3 The transcription-dependent mutation asymmetry requires PolY1 regardless of chromosomal location of the reporter gene. **A.** Schematic of the *thrC* locus harboring mutation reporters oriented head-on (HM449 and HM672) or co-directionally (HM450 and HM673) to replication, as well as a chromosomal map depicting the positions of both the *amyE* and *thrC* loci. **B.** Mutation rates were estimated for the *hisC952* reporter (CAG➔TAG) at the *thrC* locus in the presence (HM449 and HM450; gray bars), or absence of *yqjH* (HM611 and HM612; black bars), for the head-on and co-directional orientations, with (Trx+) and without (Trx-) IPTG. **C.** Same as **B.** for the *metB5* (GAA➔TAA) reporter at the t*hrC* locus (HM449 and HM450, ctrl, gray bars; HM674 and HM675, Δ*yajH,* black bars. **D.** The ratio of mutation rates in the presence of transcription for the head-on to co-directional orientation is plotted for strains harboring the *hisC952* at the *thrC* locus, in both wild-type (gray bars) and *ΔyqjH* mutant backgrounds (black bars). **E.** Same as **D.** for the *metB5* reporters. Mutation rates were estimated based on C=36-48, for each strain and condition. Error bars are 95% confidence intervals, (*P<0.05).

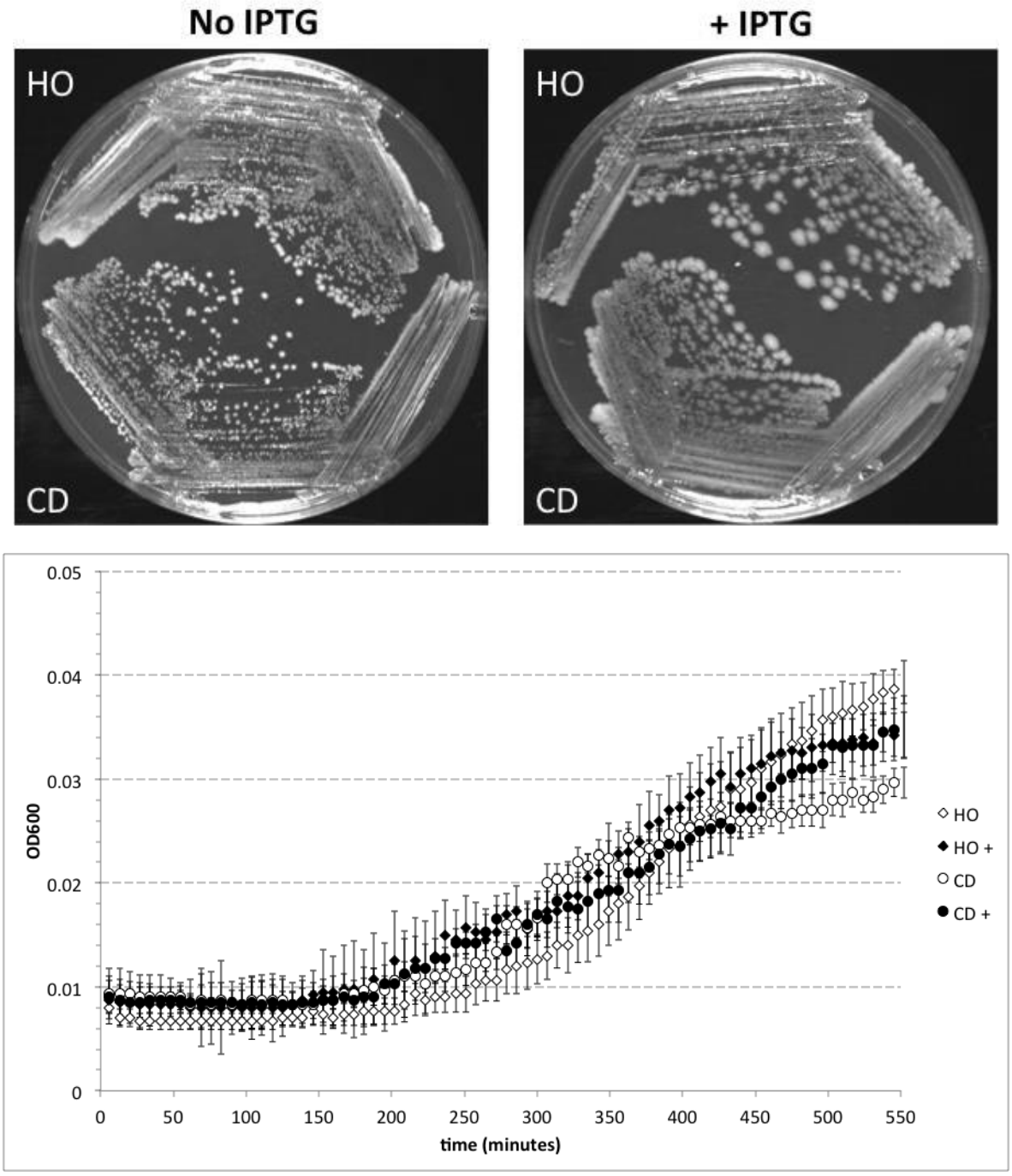


Figure 4.4 Asymmetric mutation rate measurements do not arise due to orientation-dependent differences in growth rate. Strains 1177 (HO) and 1178 (CD), which harbor wild type alleles of *hisC* under control of P*spank (hy)* were streaked from freezer stocks onto plates containing minimal medium lacking histidine with or without IPTG. Additionally, growth curves were obtained in minimal medium lacking histidine for each strain with (+) or without IPTG.

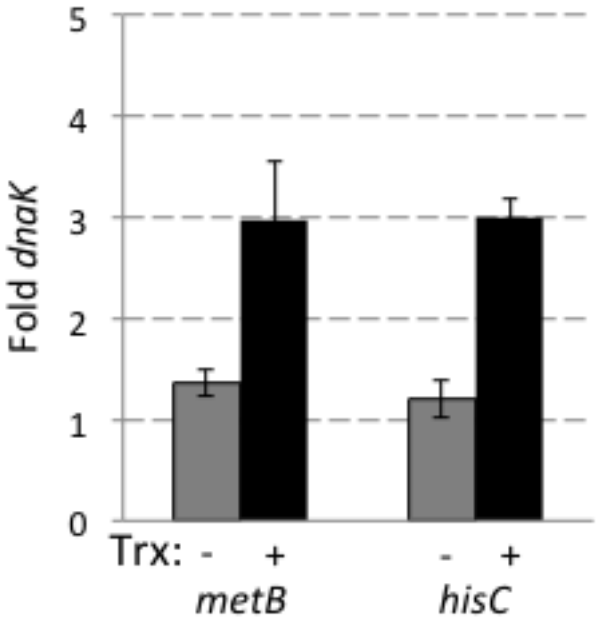


Figure 4.5 Mutation rate reporters are expressed to the same degree. RNA levels for the *metB5* (HM633) and *hisC952* (HM420) reporter strains, normalized to *dnaK*, for the CD orientation, with (Trx+, gray bars) and without (Trx-, black bars) IPTG. Data shown is the average from 6 biological replicates; error bars represent standard error of the mean.

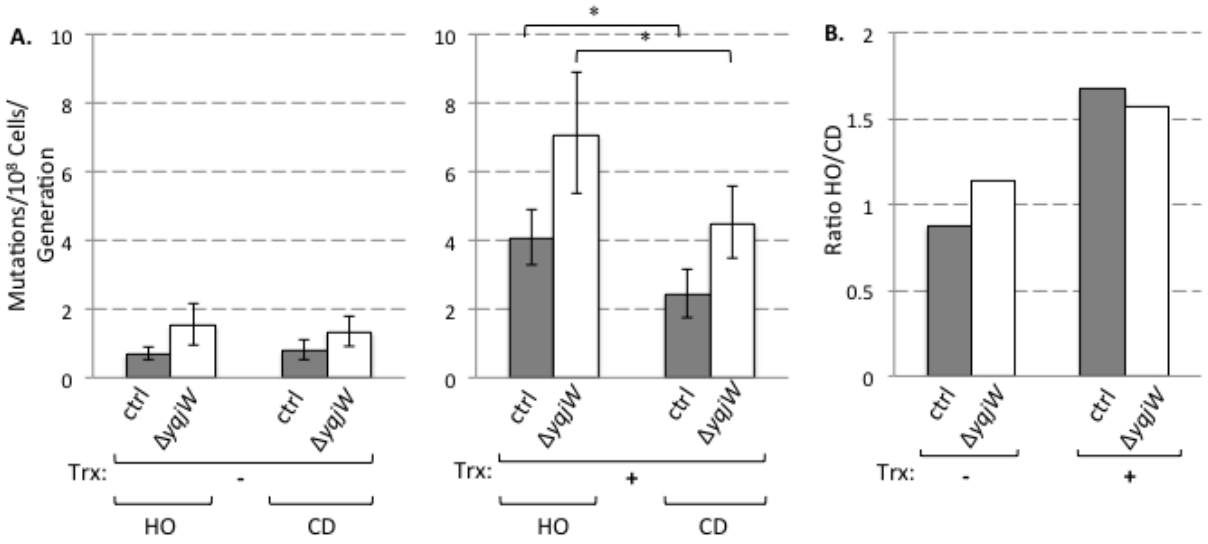


Figure 4.6 Transcription-dependent mutation asymmetry is independent of PolY2. A. Mutation rates for the *hisC952* reporter in the presence (HM419 and HM420; gray bars), or absence of PolY2 (*ΔyqjW*) (HM425 and HM426; black bars), for the head-on and co-directional orientations, with (Trx+) and without (Trx-) IPTG. **B.** The ratio of mutation rates for the head-on to co-directional orientations for strains HM419/HM420 ("ctrl,"gray bars) as well as for strains HM425/HM426 ("Δ*yqjW*," white bars) in the presence (Trx +) and absence (Trx -) of 1mM IPTG. Mutation rates were estimated based on C=36-48, for each strain and condition. Error bars are 95% confidence intervals, (*P<0.05).

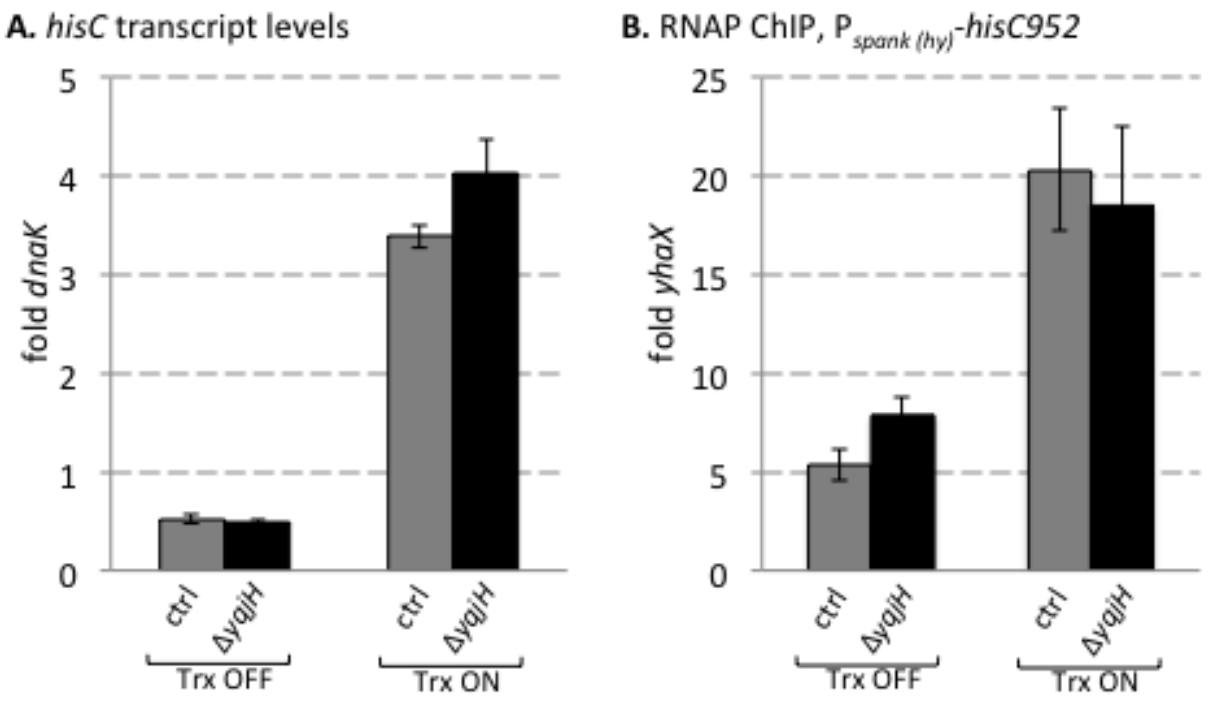


Figure 4.7 Transcript levels and RNA polymerase occupancy at the reporter gene are not altered in cells lacking PolY1 A. RNA levels for the *hisC952* reporter strains, normalized to *dnaK,* (HM419, ctrl, gray bars; HM421 and HM4222, Δ*yajH,* black bars) with (Trx+) and without (Trx-) IPTG. **B**. Relative association of RpoC by ChIP–qPCR with the region harboring *Pspank(hy)-hisC952* reporter gene, compared to the control region *yh*aX (HM419, ctrl, gray bars; HM421 and HM4222, Δ*yajH,* black bars), with (Trx+) and without (Trx-) IPTG. Error bars represent standard error. Data is from 3 independent experiments (n=9).

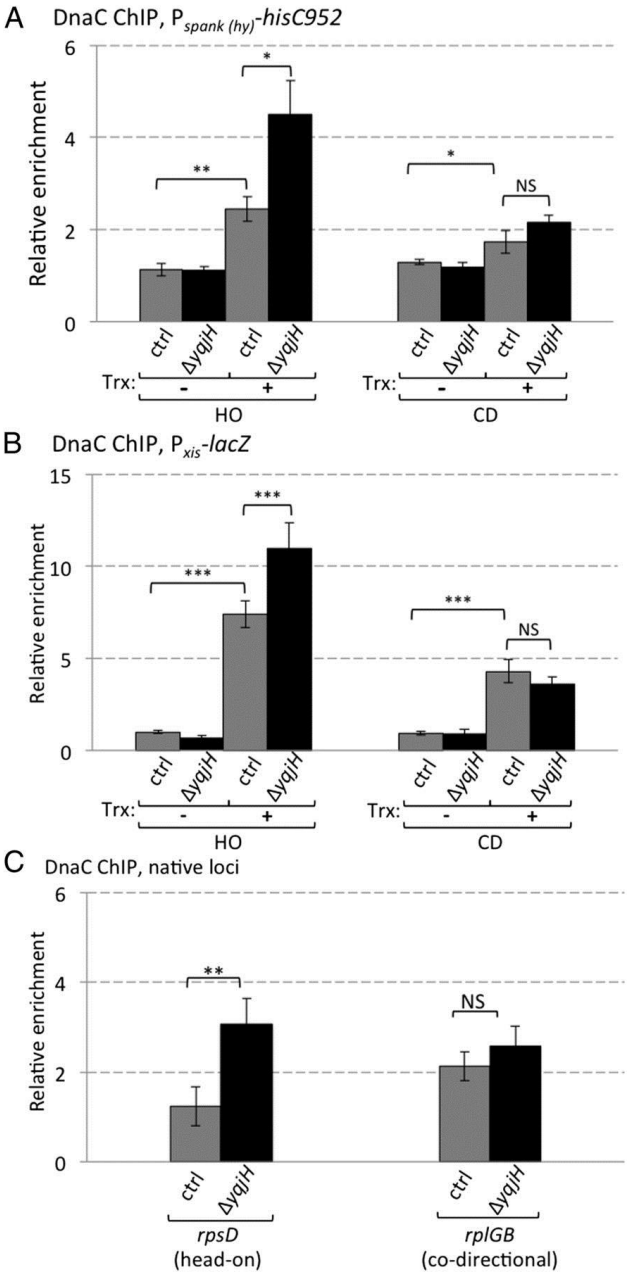


Figure 4.8 Association of the replicative helicase with lagging, but not leading-strand genes increases without PolY1 A. Relative association of DnaC was analyzed by ChIP–qPCR with the region harboring the P*spank(hy)-hisC952* reporter gene, compared with the control region *yh*aX, in the head-on (HO) or codirectional (CD) orientations (HM449 and HM450; gray bars), with and without PolY1 (HM611 and HM612; black bars), with (Trx+) and without (Trx−) IPTG. **B.** Relative association of DnaC by ChIP–qPCR with the region harboring the P*xis-lacZ* reporter gene, compared with the control region *yh*aX, in the HO orientation, with [HM352 (Trx+) and HM630 (Trx−); gray bars] and without PolY1 (*ΔyqjH*) [HM629 (Trx+) and HM631 (Trx−); black bars] as well as the relative association of DnaC with the P*xis-lacZ* reporter gene in the CD orientation, with [HM664 (Trx+) and HM663 (Trx−); gray bars] and without PolY1 (*ΔyqjH*) [HM685 (Trx+) and HM684 (Trx−); black bars]. **C.** Relative association of DnaC with the genomic loci *rpsD* and *rplGB* with (wt; gray bars) and without PolY1 (*ΔyqjH*; black bars). Error bars represent SE (n = 12) (**P < 0.01, ***P < 0.005). Significance was determined by Student's t test.

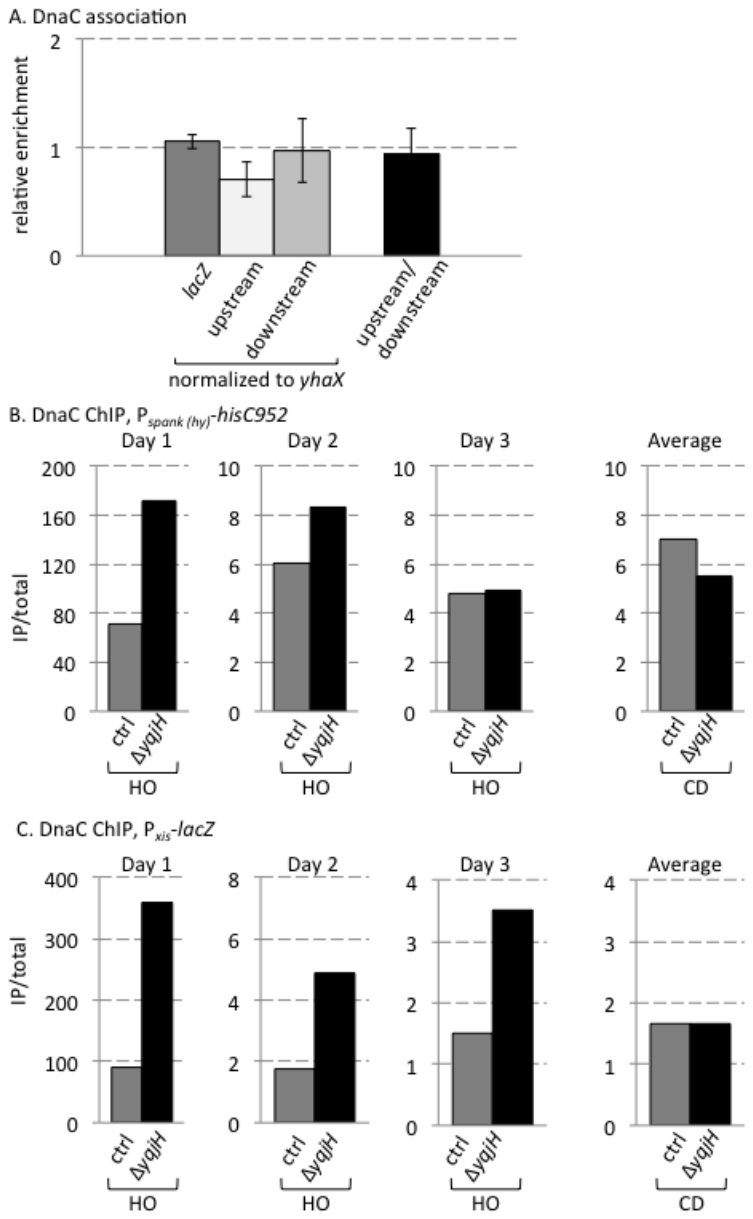
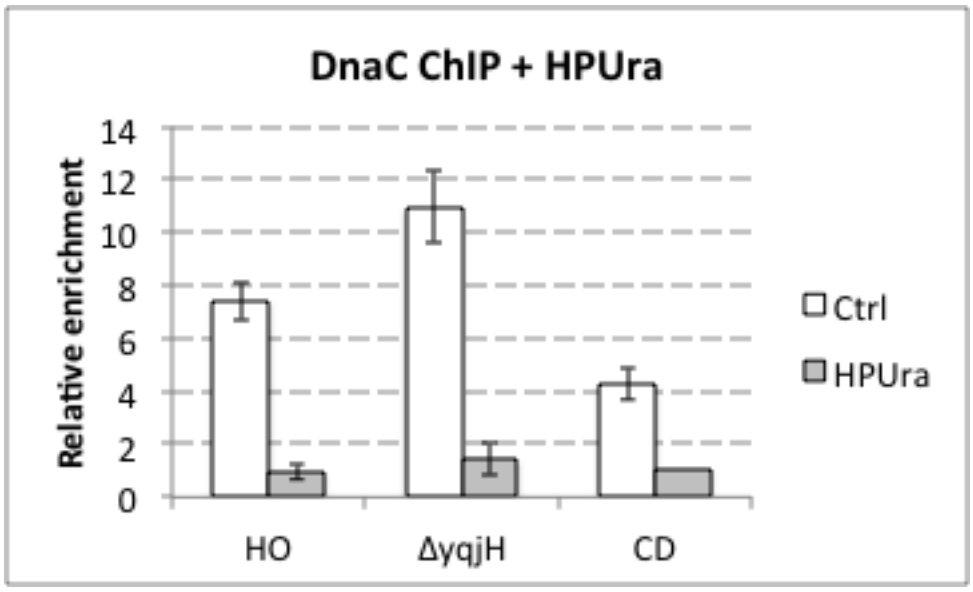
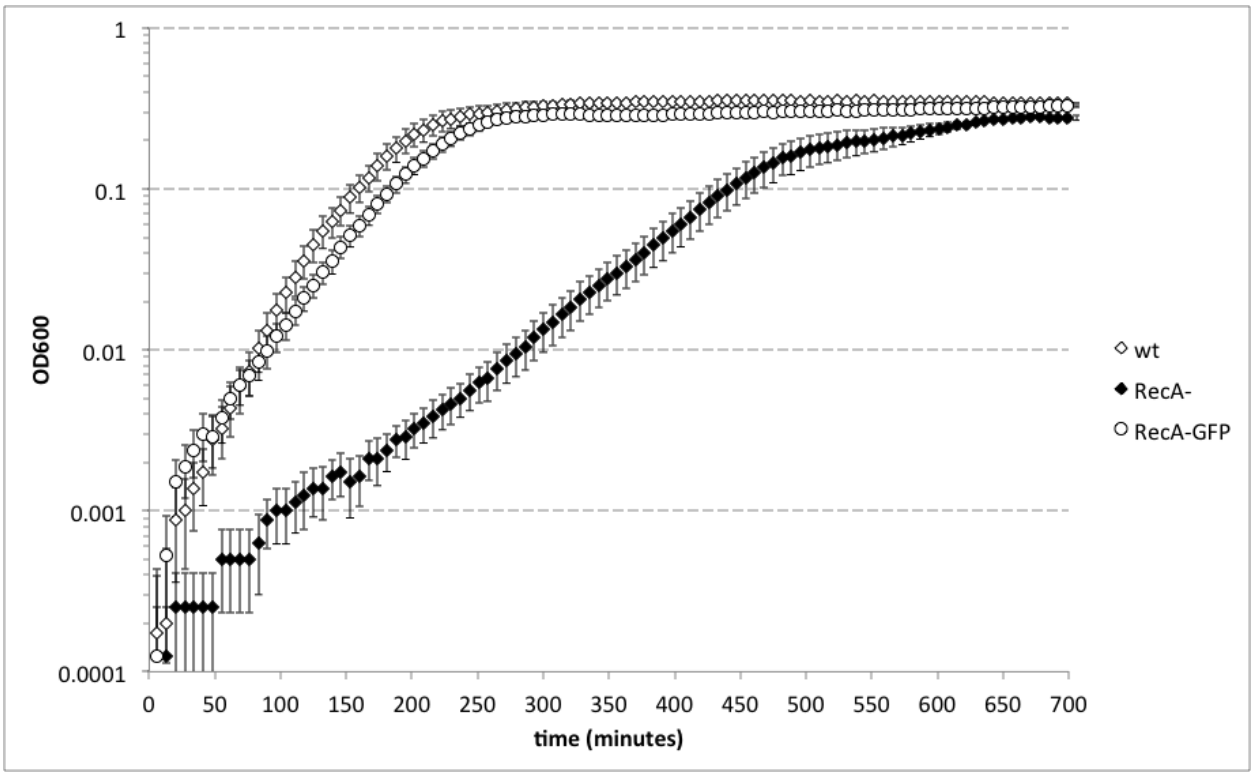


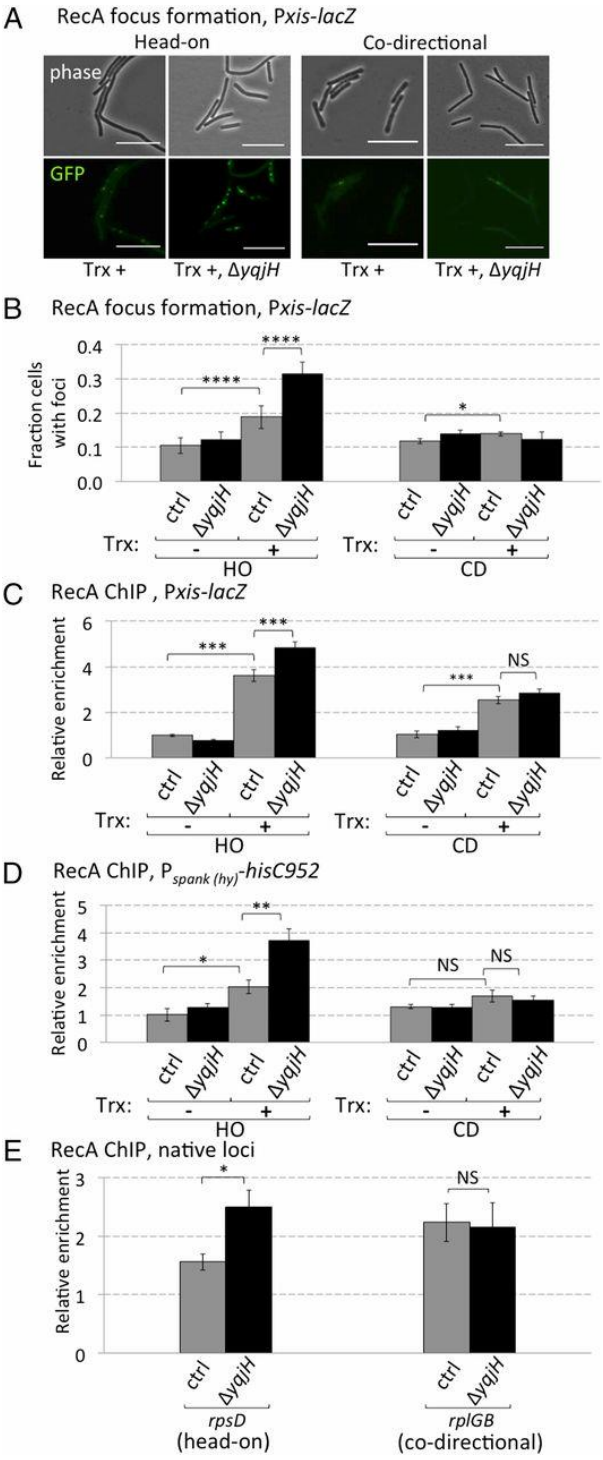
Figure 4.9 Alternate normalization loci and raw, non-normalized DnaC ChIP data A. Relative association of DnaC with regions 1 kb upstream (*guaC* locus) and 1 kb downstream (intragenic region) of *thrC* in strains harboring the P*xis-lacZ* reporter gene, compared to the control region *yha*X was analyzed by ChIP-qPCR in the absence of transcription. Data shown represents averages of nine biological replicates. **B.** Absolute ChIP signal for the enrichment of DnaC at the *hisC952* conflict region relative to total input samples for three independent experiments for the head on orientation, and the average of three experiments for the co-directional orientation is shown. **C.** Absolute ChIP signal for the enrichment of DnaC at the *lacZ* conflict region relative to total input samples for three independent experiments for the head on orientation, and the average of three experiments for the co-directional orientation is shown.



4.10 DnaC association with head-on conflict regions requires replication. Relative association of DnaC in strains expressing the P*xis-lacZ* reporter gene, compared to the control region *yh*aX was analyzed by ChIP–qPCR in the presence (HPUra, gray bars) and absence (ctrl, white bars) of the replication inhibitor HPUra at a concentration of 38.6 μg/ml. Data shown is the average of 3-9 biological replicates. Error bars represent standard deviations.

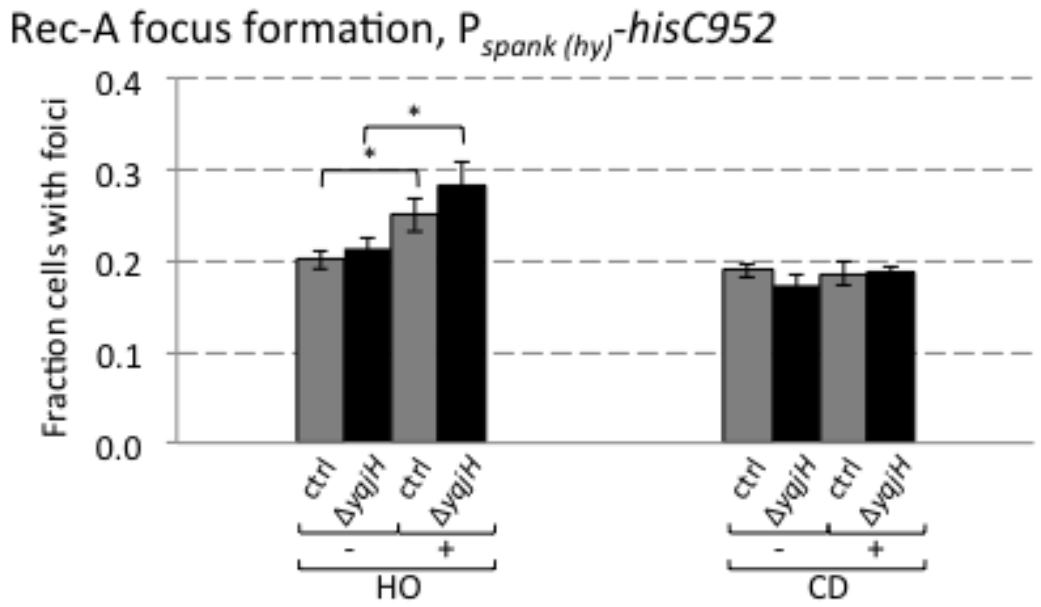


4.11 Cells harboring RecA-GFP do not display growth defects in response to severe head-on conflicts. Growth curves in LB medium were obtained for strains expressing the P*xis-lacZ* reporter in wt backgrounds (HM211, wt, white diamonds), backgrounds with *recA* deleted (HM824, RecA-, black diamonds), or backgrounds harboring RecA-GFP (HM352, RecA-GFP). Data shown is the average of eight biological replicates grown on two different days.

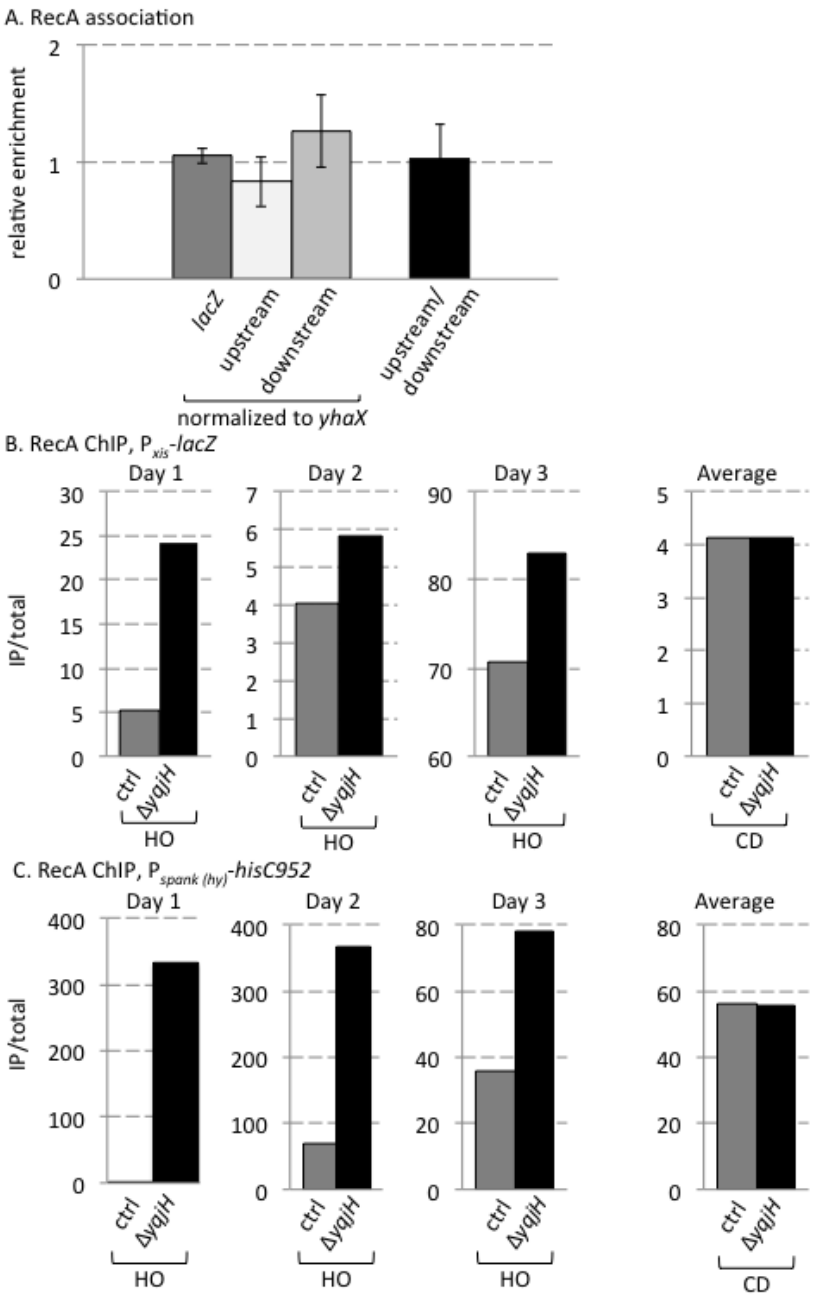


4.12 Without PolY1, RecA localization to lagging- but not leading-strand genes increases

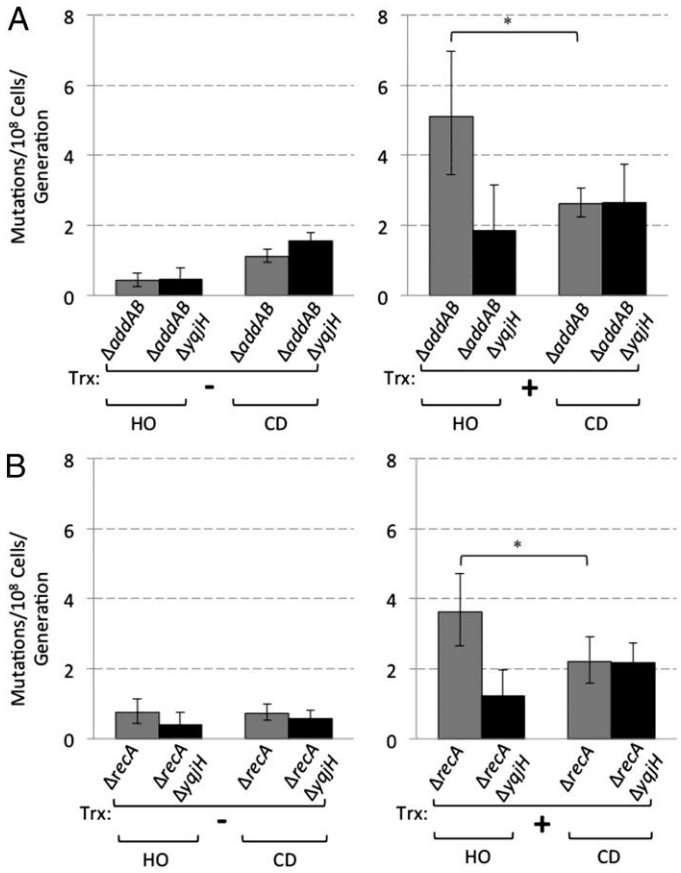
A. Representative images of cells expressing P*xis-lacZ* in the head-on orientation with (HM352) and without PolY1 (HM629), and the codirectional orientation with (HM664) and without PolY1 (HM685).**B.** Quantification of RecA-GFP focus formation in cells harboring the P*xis-lacZ* reporter in the HO orientation with [HM352 (Trx+) and HM630 (Trx−); gray] and without PolY1 [HM629 (Trx+) and HM631 (Trx−); black] and the CD orientation, with [HM664 (Trx+) and HM663 (Trx−); gray] and without PolY1 [HM685 (Trx+) and HM684 (Trx−); black]. **C.** Relative association of RecA by ChIP–qPCR with the region harboring the P*xis-lacZ* reporter gene, compared with the control region *yha*X, in the HO orientation, with [HM352 (Trx+) and HM630 (Trx−); gray] and without PolY1 (*ΔyqjH*) [HM629 (Trx+) and HM631 (Trx−); black] as well as the relative association of RecA with the P*xis-lacZ* reporter gene in the CD orientation, with [HM664 (Trx+) and HM663 (Trx−); gray] and without PolY1 (*ΔyqjH*) [HM685 (Trx+) and HM684 (Trx−); black]. **D.** Same as *C* but for strains harboring the P*spank(hy)-hisC952* reporter genes with (HM594 and HM595; gray) and without PolY1 (*ΔyqjH*) (HM596 and HM597; black), with (Trx+) or without (Trx−) IPTG. **E.** Relative association of RecA with the genomic loci *rpsD* and *rplGB* with (wt; gray) and without PolY1 (*ΔyqjH*; black). Error bars represent SE (n = 6–12) (**P < 0.01, ***P < 0.005, ****P < 0.001). Significance was determined by Student's t test.



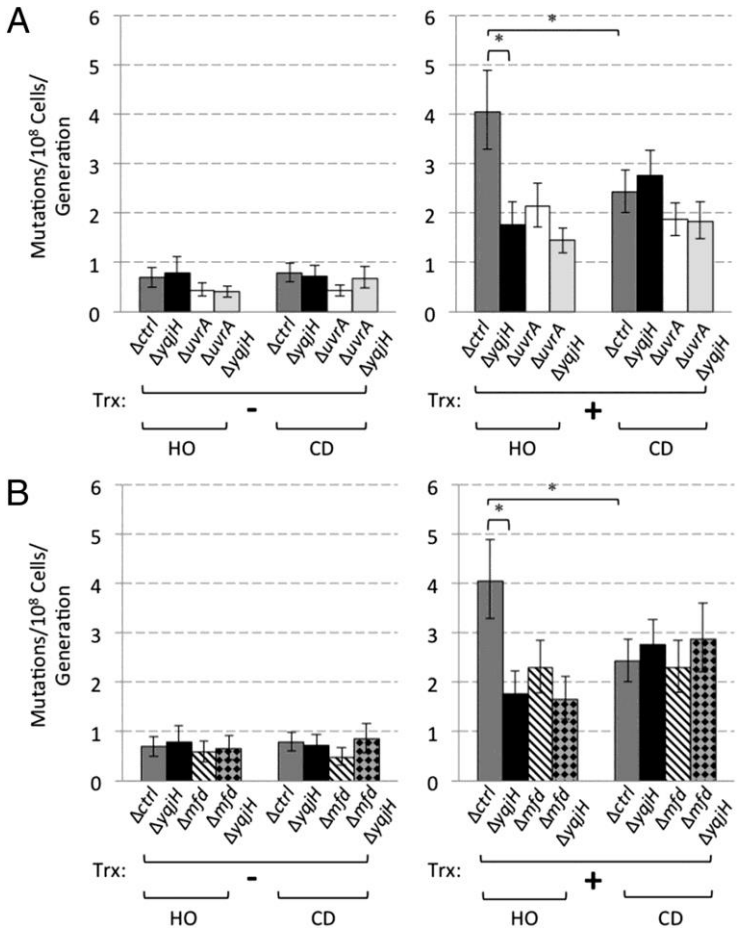
4.13 RecA-GFP localization in response to head-on *hisC* transcription increases without PolY1 A. Microscopy was performed and RecA-GFP focus formation quantified as described in **methods** for strains harboring the *hisC952* reporter in the presence (HM594 and HM595; gray bars), or absence of *yqjH* (HM596 and HM597; black bars), for the head-on and co-directional orientations, with (Trx+) and without (Trx-) IPTG. The increased RecA-GFP focus formation between the transcription off and transcription on condition is statistically significant (P=0.01).



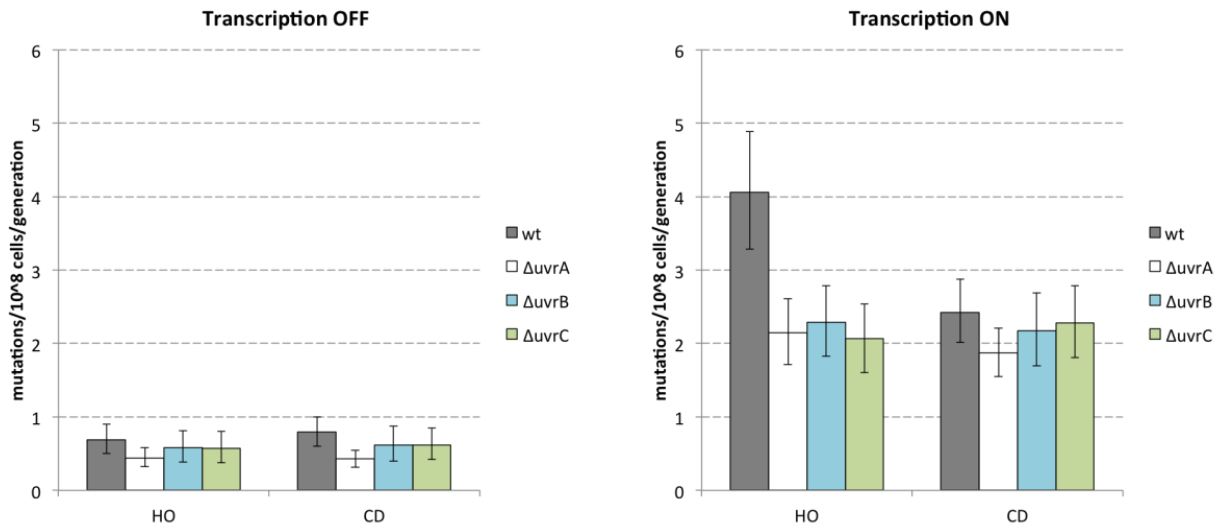
4.14 Alternate normalization loci and raw, non-normalized RecA ChIP data A. Relative association of RecA with regions 1 kb upstream (*guaC* locus) and 1 kb downstream (intragenic region) of *thrC* in strains harboring the P*xis*-*lacZ* reporter gene, compared to the control region *yh*aX was analyzed by ChIP–qPCR in the absence of transcription. Data shown represents averages of nine biological replicates. **B.** Absolute ChIP signal for the enrichment of RecA at the *lacZ* conflict region relative to total input samples for three independent experiments for the head on orientation, and the average of three experiments for the co-directional orientation is shown. **C.** Absolute ChIP signal for the enrichment of RecA at the *hisC* conflict region relative to total input samples for three independent experiments for the head on orientation, and the average of three experiments for the co-directional orientation is shown. Error bars represent standard error. (n=6-12) (**P<0.01, ***P<0.005, ****P<0.001).



4.15 PolY1 functions independently of RecA and AddAB A. Mutation rates were estimated for the *hisC952* reporter in strains lacking AddAB (HM720 and HM721, *ΔaddAB*; gray) or AddAB and PolY1 (HM740 and HM741, *ΔaddAB ΔyqjH*; black), for the head-on (HO) and codirectional (CD) orientations with (Trx+) and without (Trx−) IPTG. **B.** Same as *A* but for strains lacking either RecA (HM726 and HM727, *ΔrecA*; gray) or RecA and PolY1 (HM746 and HM747, *ΔrecA ΔyqjH*; black), with (Trx−; black bars) or without (Trx+; gray bars) IPTG. Mutation rates were estimated based on C = 30–36, for each strain and condition. Error bars represent 95% confidence intervals (*$P < 0.05$).



4.16 PolY1 acts in the same pathway as UvrA and Mfd A. Mutation rates were estimated for the *hisC952* reporter in control backgrounds (HM419 and 420, ctrl; gray), strains lacking PolY1 (HM421 and HM422, *ΔyqjH*; black), strains lacking UvrA (HM735 and HM736, *ΔuvrA*; white) or UvrA and PolY1 (HM1102 and HM1103, *ΔuvrA ΔyqjH*; light gray), for the head-on (HO) and codirectional (CD) orientations, with (Trx−) or without (Trx) IPTG. **B.** Same as *A* but for strains lacking either Mfd (HM444 and HM445, *Δmfd*; stripes) or Mfd and PolY1 (HM874 and HM875, *Δmfd ΔyqjH*; argyle), with (Trx−) or without (Trx+) IPTG. Mutation rates were estimated based on C = 30–48, for each strain and condition. Error bars represent 95% confidence intervals (*$P < 0.05$).



4.17 Transcription-dependent mutation asymmetry requires UvrB and UvrC A. Mutation rates for the *hisC952* reporter in the presence (HM419 and HM420; gray bars), or absence of UvrA (*ΔuvrA*) (HM735 and HM736; white bars), UvrB (*ΔuvrB*) (HM1281 and HM1282; teal bars), or UvrC (*ΔuvrC*) (HM1283 and HM1284; teal bars), for the head-on and co-directional orientations, with (Transcription ON, right panel) and without (Transcription OFF, left panel) IPTG. Mutation rates were estimated based on C=30-36, for each strain and condition. Error bars are 95% confidence intervals.

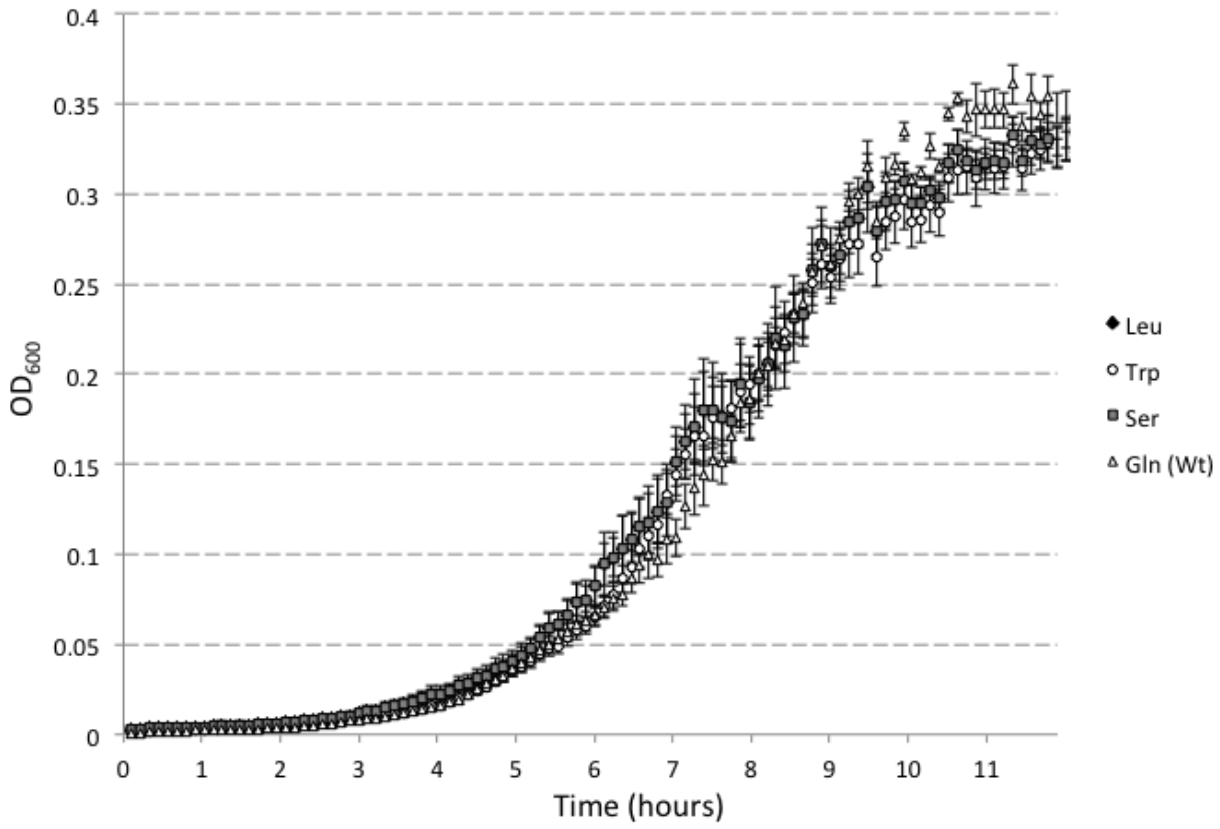


Figure 4.18 Multiple amino acids are tolerated at the stop codon position in the *hisC* reporter. Growth curves were obtained in minimal medium lacking histidine for strains harboring the *hisC* mutation reporter with different base pair mutations, and thus differing amino acids, at the position of the stop codon. Data shown is average optical density measurements taken over time for four independent clones of each *hisC* allele.

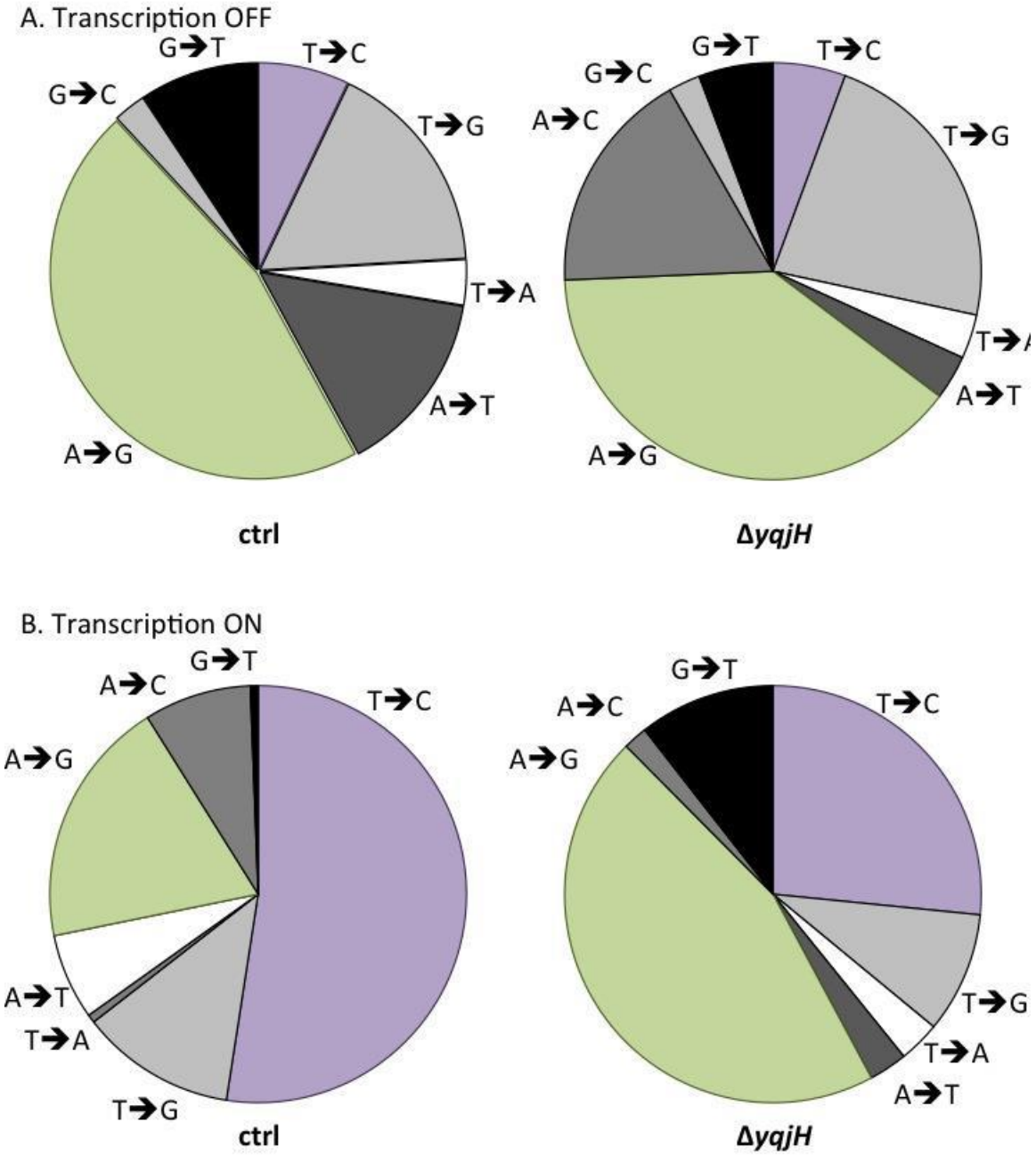


Figure 4.19 Transcription-dependent T➔C mutations are reduced in cells lacking PolY1 Frequencies for each observed point mutation leading to reversion of the *hisC952* reporter in the head-on orientation for wt cells (left chart) and cells lacking PolY1 (*ΔyqjH*, right chart) grown in **A.** the absence (transcription OFF) or **B.** in the presence of IPTG (transcription ON). Mutation frequency is calculated as the ratio of reads for a particular mutation divided by the total number of reads.

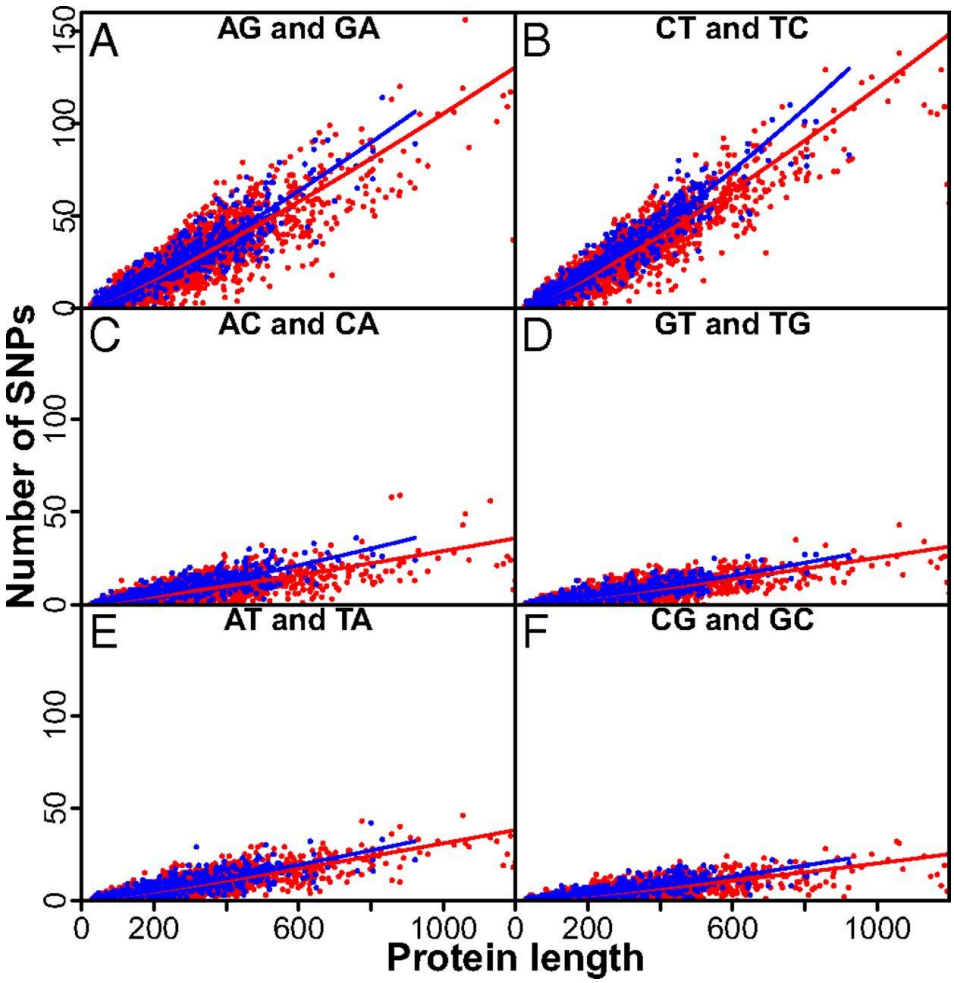


Figure 4.20 SNP counts show a stronger positive correlation with gene length on the lagging strand. The number of SNPs for each gene on the lagging strand (blue dots) and the leading strand (red dots) is plotted as a function of its (translated) protein length. The SNPs AG and GA (**A**), CT and TC (**B**), AC and CA (**C**), GT and TG (**D**), AT and TA (**E**), and CG and GC (**F**), were pooled because direction of the mutations was not distinguished. Negative binomial regression with a log link function was used to model the gene length dependence of the number of SNPs for the lagging strand (blue lines) and the leading strand (red lines). The mutation rate (slope) is greater for the GA/AG and TC/CT mutations but increases nonlinearly with gene length (with $P < 0.0001$) for all six mutations. The difference in the rate of increase on the lagging strand compared with the leading strand of 4.5% per kb is significantly greater ($P = 0.005$) for the TC/CT mutation.

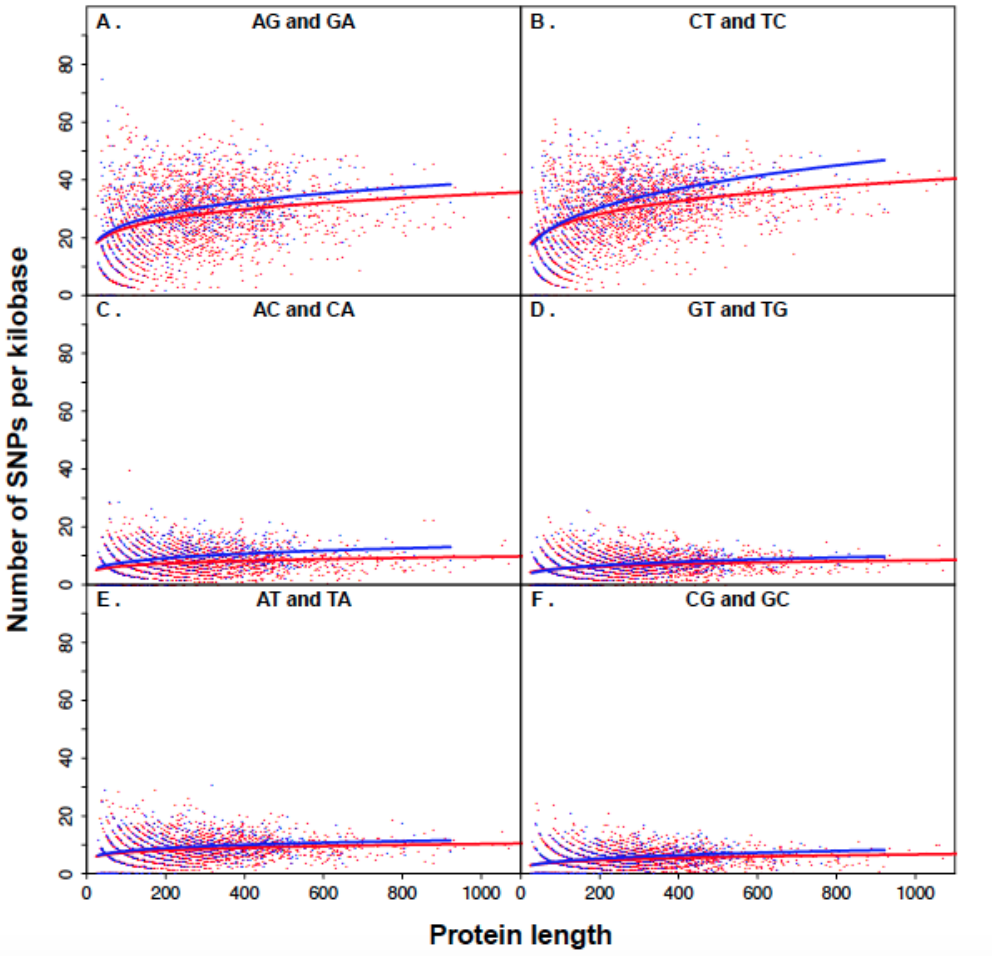


Figure 4.21 Only the TC/CT transitions show a significantly greater rate of mutation on the lagging versus the leading strand. The mutation rate (SNPs per kilobase) for each gene on the lagging strand (blue dots) and the leading strand (red dots) is shown as a function of its (translated) protein length. A large dispersion in mutation rates is evidence of the large range of gene conservation represented on both strands. The modeled mutation rates for the lagging strand (blue lines) and the leading strand (red lines) show the non-linear increase in the rates with gene length. The rates are higher for the GA/AG and TC/CT mutations. The difference in the rate of increase on the lagging strand compared to the leading strand of 4.5%/kb is significantly greater (p=0.005) for the TC/CT mutation (see **table 4.1**).

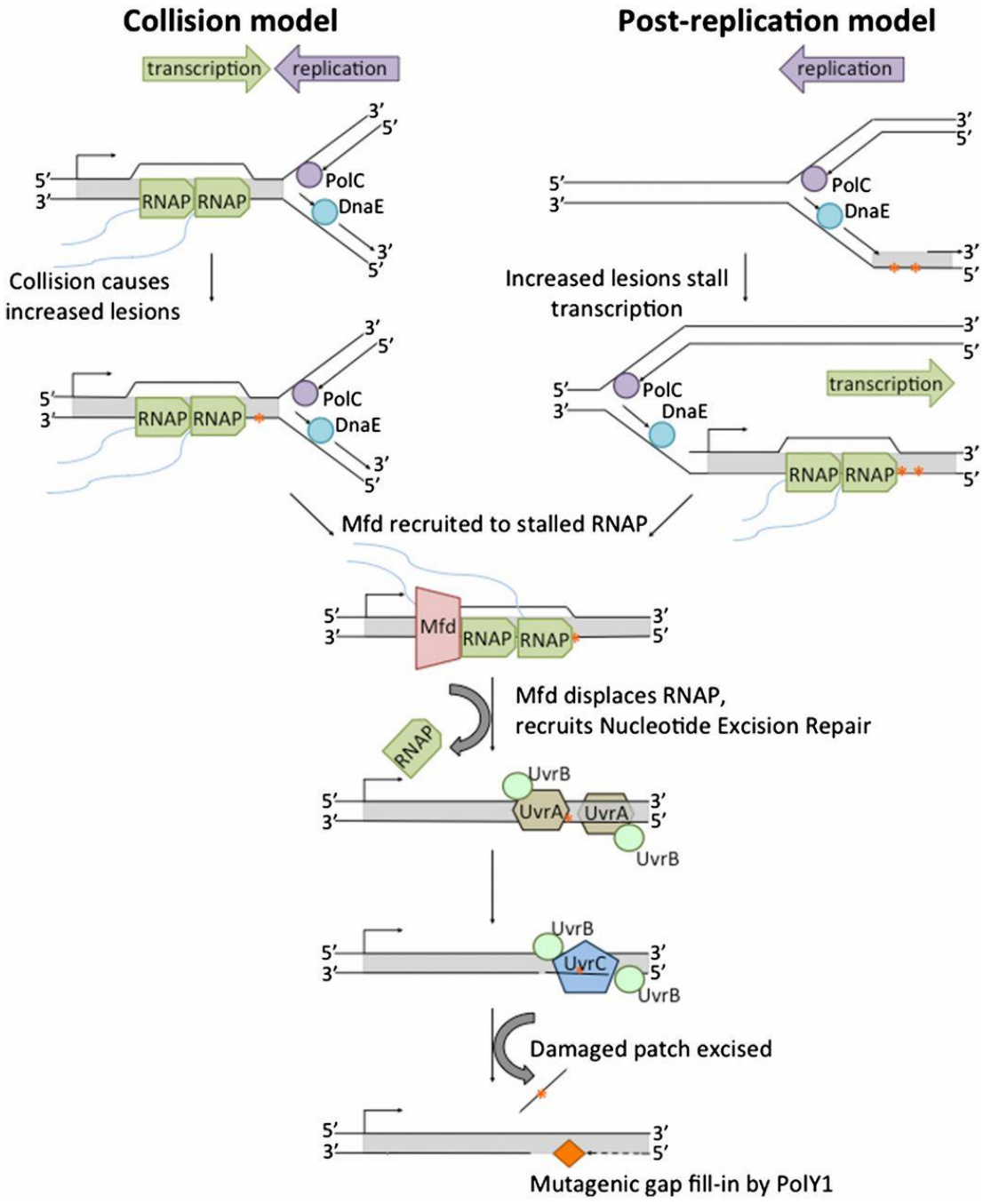


Figure 4.22 Models of PolY1 function at regions of lagging-strand transcription. Replication of the lagging-strand–encoded genes generates increased DNA lesions, either due to the discontinuous nature of its synthesis and/or head-on collisions between the replication and transcription machineries. RNA polymerases stalled at lesions (as in the postreplication repair model) or at the site of the collision itself (as in the collision model) trigger transcription-coupled recruitment of the NER machinery. Upon excision of the lesion-containing patch by NER, PolY1 fills in the resulting gap, causing mutations.

Mutation	%/kb increase (lagging)	%/kb increase (leading)	p-value
CA/AC	16.7	14.8	NS
GA/AG	14.9	14.9	NS
TA/AT	9.5	10.8	NS
GC/CG	24.7	19.7	NS
TC/CT	25.3	20.8	0.005
TG/GT	19.3	14.8	NS

Table 4.1 The modeled mutation rate increases non-linearly with gene length on the lagging and leading strands. The percent increase in the mutation rate per kilobase of gene length is shown for the leading and lagging strands for pooled mutations (since direction of mutation is unknown). This non-linear increase in the mutation rate, modeled using negative binomial regression, had statistical significance ($p<0.0001$) for all mutations on the lagging strand (not shown). For all mutations, the increase in mutation rate with gene length is greater on the lagging strand than the leading strand. For most mutation pairs, the statistical support for the difference in the increase between the leading and lagging strand (p-value) is not significant (NS). Only the pooled TC + CT mutations achieved statistical significance ($p=0.005$) in the difference in rate between the strands.

Chapter 5: Head-on transcription increases the susceptibility of genes to UV lesions

5.1 Introduction

The hallmarks of head-on conflicts: increased mutagenesis, replication stalling, and DNA breaks (1, 3-5, 8, 17-19, 156), could all stem from increased levels of DNA damage in expressed lagging-strand genes. The involvement of PolY1 acting with transcription coupled-nucleotide excision repair in promoting asymmetric mutagenesis at head-on genes indicated that the DNA in these regions might experience elevated levels of bulky lesions(18). Acute, high-level exposures to chemical insults may cause bulky lesions genome-wide; however, cells in the environment more likely experience sub-lethal doses of genotoxic stresses (235). Therefore, mechanisms that promote mutagenesis, such as transcription, could exert their effects by increasing the susceptibility of DNA in a particular region to damage.

The spontaneous mutation spectrum observed in yeast unexposed to DNA damaging agents suggests a major role for lesions in altering the DNA sequence (235). Additionally, previous studies examining transcription-associated mutagenesis in *S. cerevisiae* identified a mutational signature indicative of damage to the transcribed DNA (19). Transcription has a well-established role in promoting mutagenesis; work by our lab, as well as others, revealed that expression asymmetrically increases mutation rates specifically for lagging-strand genes (17-19, 90, 154). I wondered whether transcription and gene orientation could sensitize DNA to a common bulky lesion.

The principle lesions processed by the TC-NER pathway are pyrimidine dimers arising due to UV-light exposure (197, 199, 201, 202, 233). Dimers form in the DNA when high energy photons excites double-bonded pi electrons in pyrimidine bases, allowing them to react with neighboring molecules (**Figure 5.1**) (236, 237). Pyrimidine dimers distort the double helix, and present a strong obstacle to both replication and transcription (199). Distortion in the double helix activates the Nucleotide Excision Repair pathway (199); alternatively the repair machinery may be recruited to RNA polymerases stalled at the lesion by the transcription-repair coupling factor, Mfd (201)(discussed in detail in **chapter 4**). We demonstrated that Nucleotide Excision Repair increases mutation rates in transcribed lagging-strand genes through the action of PolY1. We speculated that increased lesions in these loci might

underlie this phenomenon, yet we were unable to resolve whether the lagging strand orientation alone or specifically transcription promoted lesion formation.

In order to determine whether transcription renders, specifically, head-on genes more susceptible to bulky lesions, I developed a DNA pull-down assay to detect thymidine dimers. I found that transcription in general increased the susceptibility of DNA to lesions. Importantly, however, I detected increased levels of pyrimidine dimers within genes expressed head-on to replication. These data lend support to the notion that head-on conflicts increase the susceptibility of the DNA to at least one type of DNA lesion.

5.2 Results

I wondered if the orientation and transcription-dependent mutation asymmetry we previously reported arises due to increased lesions at head-on conflict regions. To test this hypothesis, I used a modified DNA pull down method (DIP) (238) to directly measure UV-induced pyrimidine dimers on DNA in a specific region.

Using a commercially available, monoclonal antibody that recognizes pyrimidine dimers (GeneTex GTX10347) in DIPs followed by qPCRs, I quantified the relative amounts of lesions on the DNA. To determine if the antibody was specific to UV-induced damage, I first compared the DIP signal in the presence and absence of *in vitro* UV treatment for both single- and double-stranded DNA. I found that UV treatment increased the DIP signal 5-fold for double-stranded and 13 fold for single stranded DNA (**figure 5.2**).

The high-energy photons of UV light may cause pyrimidine dimers to form between adjacent thymidine or cytosine residues (237). However, this type of damage may also break the DNA or cause inter-strand crosslinks (236, 239). The UV-dependent increased amounts of DNA I detected in IPs could arise either from lesion recognition, or non-specific interactions between the antibody and an alternate damaged DNA product. To distinguish between these possibilities I compared the relative amounts of immunoprecipitated, UV-damaged DNA before and after two different *in vitro* reactions with enzymes that specifically process dimers: T4 Pyrimidine Dimer Glycosylase and *E. coli* photolyase.

T4 Pyrimidine Dimer Glycosylase's lyase activity flips out damaged bases, then the enzyme cleaves the DNA 3' of the lesion (240). Incubation of UV-treated DNA samples with T4 pyrimidine dimer glycosylase (NEB) caused a 50% reduction in dimer-containing DNA in the Ips (**figure 5.3a**). The *E. coli* photolyase enzyme, PhrB, harnesses the energy of visible wavelengths of light to catalyze the direct repair of pyrimidine dimers to the original bases (241). Treatment of UV-damaged DNA with *E. coli* photolyase reduced the DIP signal to background levels (**figure 5.3b**). These results indicate that the increased DIP signal observed with UV damage likely arose due to increased pyrimidine dimers in the DNA.

I next measured the effects of transcription- and orientation- on UV-damage sensitivity in an identical gene in the two orientations, at low and high transcription levels, inside living cells. For this,

I took advantage of three sets of engineered conflicts, where either the *hisC, metB,* or *lacZ* gene is placed under the control of the $P_{spank(hy)}$ or P_{xis} promoter, respectively (18). I grew cells harboring each reporter with or without transcription from the promoters, treated the cultures with UV light, isolated genomic DNA then performed DIPs, as described in **methods.**

At low transcription levels, the orientation of the gene had no effect on the DIP signal for any reporter. High transcription conditions were associated with increased DIP signals, which were more pronounced for genes in the head-on orientation (**figure 5.4**). Although the absolute amounts of dimer-containing DNA detected in the DIPs varied between reporters, upon induction of transcription, the levels were always two-fold higher when the gene was oriented head-on (**figure 5.4**). The increased detection of lesion-containing DNA in the DIPS for cells harboring head-on versus co-directional constructs indicates that transcription in general increases susceptibility of DNA to UV-induced lesions, but to a higher degree when the gene is oriented head-on to replication.

DNA harvested from cultures not treated with UV, under low transcription conditions displayed minimal enrichment of the reporter-gene containing regions in the immunoprecipitations. High transcription conditions in the absence of UV did not increase the DIP signal for the *metB* and *lacZ* reporters (**figure 5.5**). Surprisingly, for the *hisC* reporter only, transcription itself, without exogenous UV damage, led to an increased enrichment of dimer-containing DNA in the DIPs (**figure 5.5**). Similar to conditions of exogenous UV damage, this increase was roughly two-fold higher when the gene was in the head-on orientation. The increased signal in the absence of exogenous damage cannot be attributed to ambient UV-light because a similar orientation- and transcription-dependent increase in DIP signal was observed when cells were maintained in absolute darkness throughout the course of the experiment (**figure 5.6**). It is unclear what property intrinsic to the *hisC* reporter sequence causes a transcription- and orientation- dependent DNA structure independent of UV damage that is recognized by the anti-thymidine dimer antibody.

5.2 Discussion

The DNA immuno-precipitations demonstrate that transcription asymmetrically increases the susceptibility of head-on genes to at least one type of DNA lesion. Transcription could render, specifically, head-on genes more sensitive to damage through multiple mechanisms. While chemical photosensitizers can increase the susceptibility of DNA to pyrimidine dimer formation (236), the effects of these compounds should be identical for genes in either orientation. The increased susceptibility of lagging-strand genes to UV lesions more likely arises from a pronounced accumulation of single stranded DNA at regions of head-on transcription.

Exposed single stranded DNA is known to be extra susceptible to damage (242-244); consistent with these reports, I detected almost ten-fold greater levels of pyrimidine dimers in DNA that was single-stranded during UV treatment compared to double stranded DNA (**figure 5.2**). These *in vitro* results mirror the two-fold higher levels of pyrimidine dimers I detected in transcribed head-on as compared to co-directional genes *in vivo* (**figure 5.3**). Potentially the asymmetric sensitivity could arise from differential single-stranded DNA exposure at expressed lagging strand genes.

Multiple mechanisms may cause single-stranded DNA to accumulate at head-on genes. Short stretches of DNA may be exposed during discontinuous lagging strand synthesis (27, 29). However, Okazaki fragment synthesis itself should be independent of gene expression and any gaps in the template at the fork are rapidly bound and protected by SSB (116, 245). The increased lesion sensitivity observed in the DIPs arose due to a transcription-dependent process, arguing against replication alone as the underlying mechanism.

Transcription itself necessitates local DNA unwinding to generate the template for RNA synthesis, though the short 6-8 nucleotide stretch of single-stranded DNA within the transcription bubble is likely shielded from chemical damage by the RNA polymerase holoenzyme complex (246). Additionally, transcription alone should be identical for a gene in either orientation. However, underwound DNA accumulates behind RNA polymerase (145, 247). Head-on collisions between the replisome and RNA polymerase that stall both machineries would cause single stranded DNA to persist on both sides of the fork, increasing the sensitivity to chemical damage in these regions.

Strikingly, the results of the DIPs correlate with the mutation rate measurements presented in **chapter 4**. The two-fold greater transcription-dependent susceptibility to DNA damage could, in part, give rise to the two-fold higher mutation rates observed in expressed lagging strand genes. Additionally, the five-fold higher absolute levels of DIP signal for the *metB* as compared to the *hisC* reporter mirrors the discrepancies in absolute transcription-induced mutation rates between the two genes (18). The disparity in mutation rates and damage sensitivity between the two reporters likely arises from sequence-intrinsic properties. Local sequence context is well established to dramatically influence mutagenesis (208, 248, 249). The local proximity of pyrimidines or secondary structures that maintain nucleotides in energetically favorable conformations for lesion formation (250) could conspire to render specific genes or regions more sensitive to a variety of chemical insults.

Although the TT-dimer antibody was highly specific for UV-induced damage for two of the three reporters examined, surprisingly, I observed a transcription- and orientation-dependent increase in DIP signal for the *hisC* reporter in the absence of damage. The results of the T4 PDG and photolyase experiments demonstrate that *in vitro* treatments to reduce dimers on the DNA attenuate the DIP signal, suggesting that the antibody is, for the most part, specific. It is unclear why *in vivo*, for the *hisC* reporter only, head-on transcription alone caused a change in the DNA that was recognized by the anti-thymidine dimer antibody.

The pyrimidine dimer antibody was raised by isolating antisera from rabbits injected with UV-irradiated oligo-pyrimidines conjugated to BSA. The sequence of the oligomers used as antigens is proprietary, making it impossible to determine whether similar motifs may be found within the sequence of *hisC*, giving rise to non-specific recognition. However, because I did not detect UV-independent DIP signal when the gene was not expressed **(figure 5.4)**, sequence alone cannot account for the structure in the DNA recognized by the pyrimidine dimer antibody.

The classical lesion induced by ultraviolet light is the pyrimidine-pyrimidine dimer (237, 251, 252), which is recognized and repaired by TC-NER (201, 202). However, TC-NER repairs other insults to the DNA (253) and bulky lesions may arise via numerous mechanisms intrinsic to normal cellular metabolism (254). The thymidine-glycol adduct and cyclopurine products, bulky lesions formed by reaction with oxygen radicals (254-256), are readily removed by excision repair (223, 229, 230).

Potentially transcription renders the sequence of the DNA in the *hisC* reporter exquisitely susceptible to a bulky lesion with similar steric properties to a pyrimidine dimer. Highly sensitive mass-spectrometric analysis of the DNA at the single nucleotide level would be necessary to elucidate the nature of the lesion occurring in the *hisC* reporter under high transcription conditions. Regardless, it is clear that transcription and orientation conspire to render genes more susceptible to at least one type of lesion.

The effects of gene orientation on lesion-sensitivity represent a heretofore-unappreciated influence on genomic instability, potentially paving the way towards a deeper understanding of the processes responsible for chromosome pathologies.

5.4 Figures

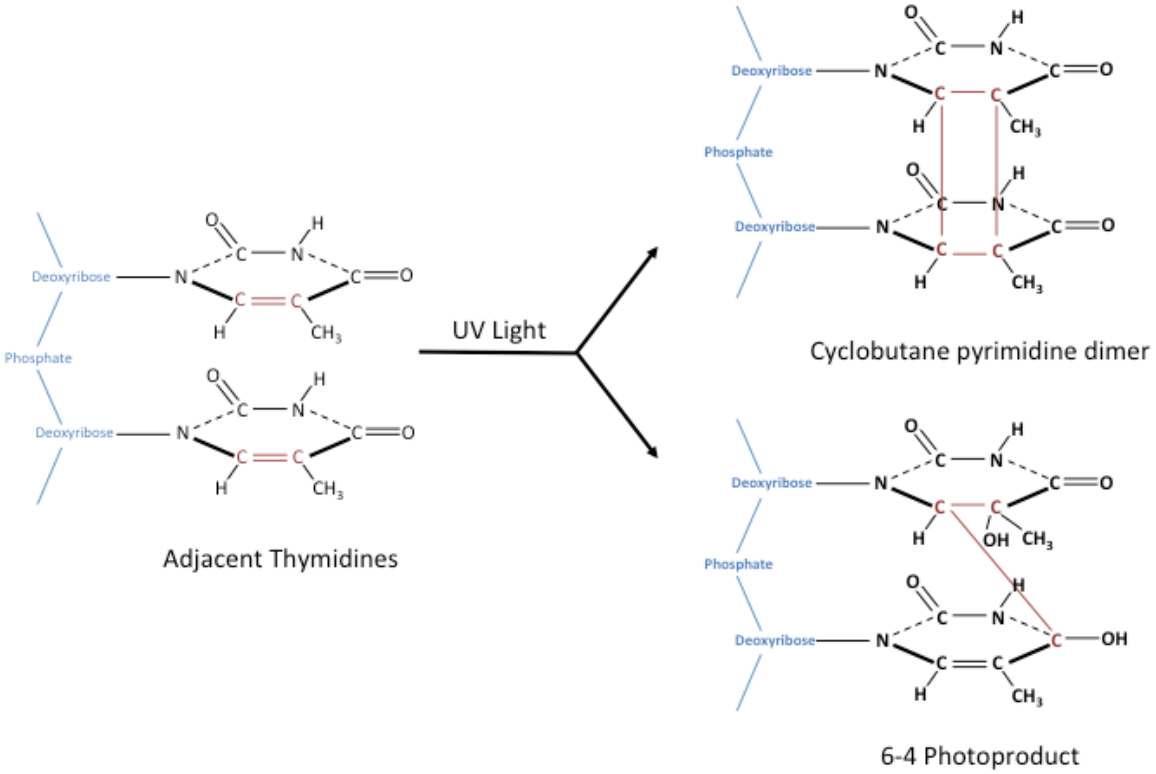


Figure 5.1 Schematic diagram of pyrimidine dimer formation in UV-irradiated DNA. High-energy photons from incoming UV light excites pi electrons in the C_4-C_5 double bond present in pyrimidine nucleotides. If the adjacent nucleotide is also a pyrimidine, excited electrons may react with C_5 of the nearby base to form a cyclobutane pyrimidine dimer (top). Alternatively, the excited electrons may react with C_6 of the adjacent base to for the 6-4 photoproduct. Both classes of pyrimidine dimers are substrates for excision repair due to their double helix distorting and transcription blocking tendencies.

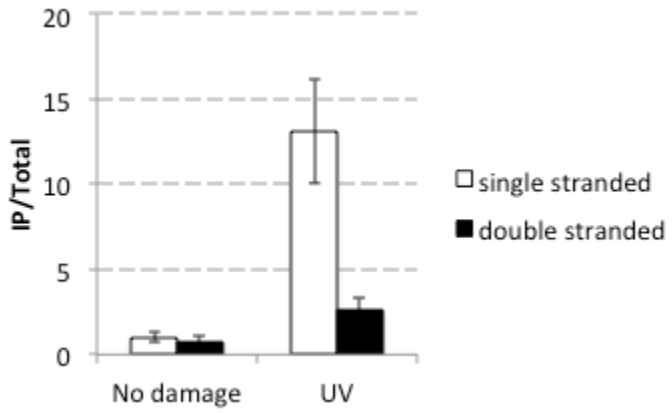


Figure 5.2 UV-irradiation increases DIP signal to a higher degree for signal stranded DNA. Single-stranded (white bars) and double-stranded (black bars) DNA was either added directly into IP buffer (No damage, left) or treated with 500 J/m^2 UV light (UV, right). DNA was immunoprecipitated using an anti-pyrimidine dimer antibody and the relative amount of *hisC* dimer containing DNA in the pull-downs, relative to total input controls was quantified by qPCR using primers HM495 & HM496.

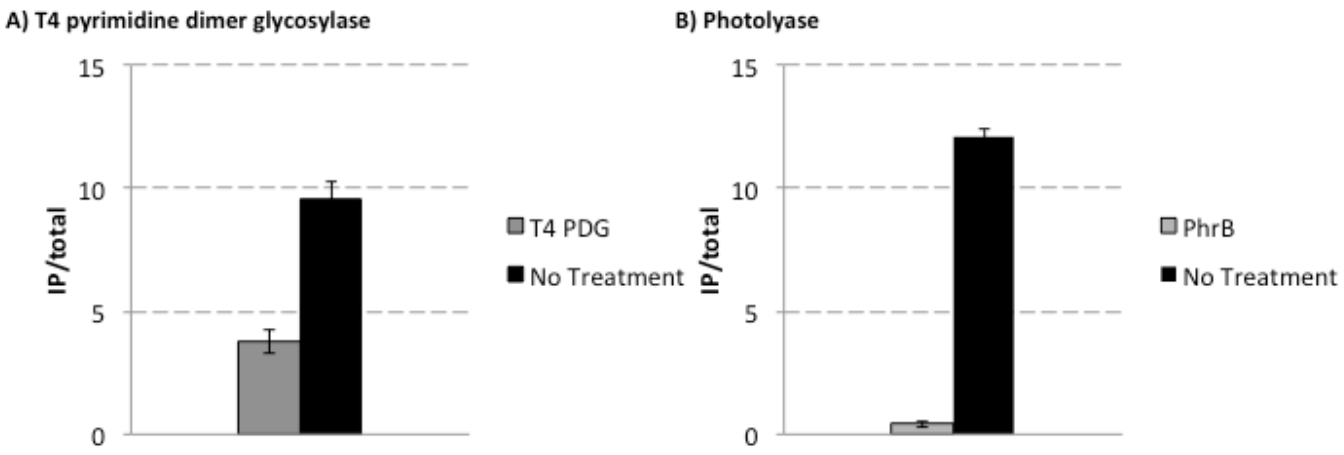


Figure 5.2 Enzymatic treatments of UV-irradiated DNA *in vitro* reduce pyrimidine dimer DIP signal

A) DNA was UV-irradiated, then either incubated with T4 pyrimidine dimer glycosylase (T4 PDG, black) or added to a mock reaction lacking the enzyme (No Treatment, gray). DNA was extracted then immunoprecipitated as described in **methods**. Relative amounts of *hisC* DNA in the pull-downs compared to total input controls was determined by qPCR using the primer pair HM495 & HM496. **B)** DNA was UV-irradiated, then either incubated with purified *E. coli* photolyase (PhrB, black) or added to a mock reaction lacking the enzyme (No Treatment, gray DNA was extracted then immunoprecipitated as described in **methods**. Relative amounts of *metB* DNA in the pulldowns compared to total input controls was determined by qPCR using the primer pair HM842 & HM843.

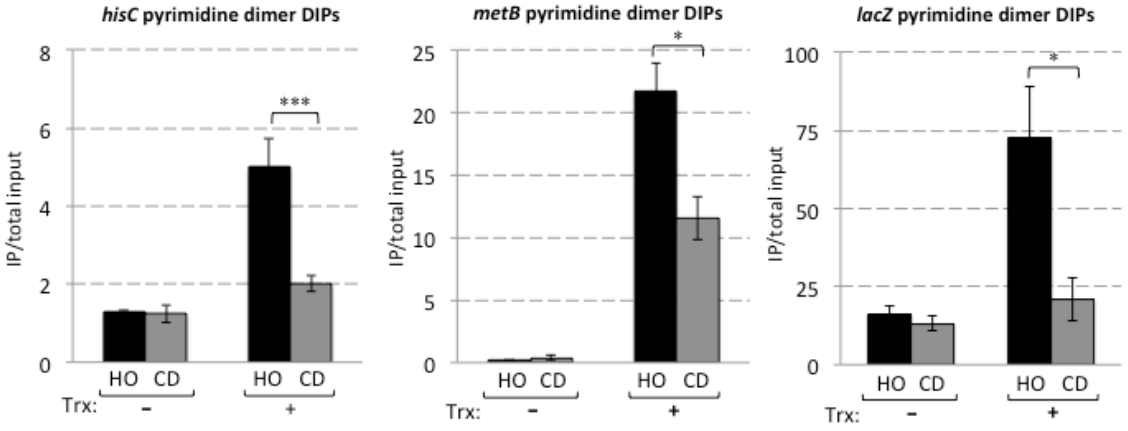


Figure 5.3 Transcription asymmetrically increases pyrimidine dimers in DNA harvested from cells exposed to UV damage. Cells harboring the *hisC*, *metB*, or *lacZ* reporters at the *amyE* (for *hisC* and *metB*) or *thrC* (for *lacZ*) loci head-on (HO, black bars) and co-directionally (CD, gray bars) to replication were grown under conditions of high (Trx +) and low transcription (Trx -) and treated with 500 J/m^2 UV light. DNA was extracted and immunoprecipitated using the anti-pyrimidine dimer antibody. Relative enrichment of *hisC*, *metB*, or *lacZ* DNA in the DIPs as compared to toal input controls was determined by qPCR using the primers HM495 & HM496, HM842 & HM843, and HM188 & HM189, respectively.

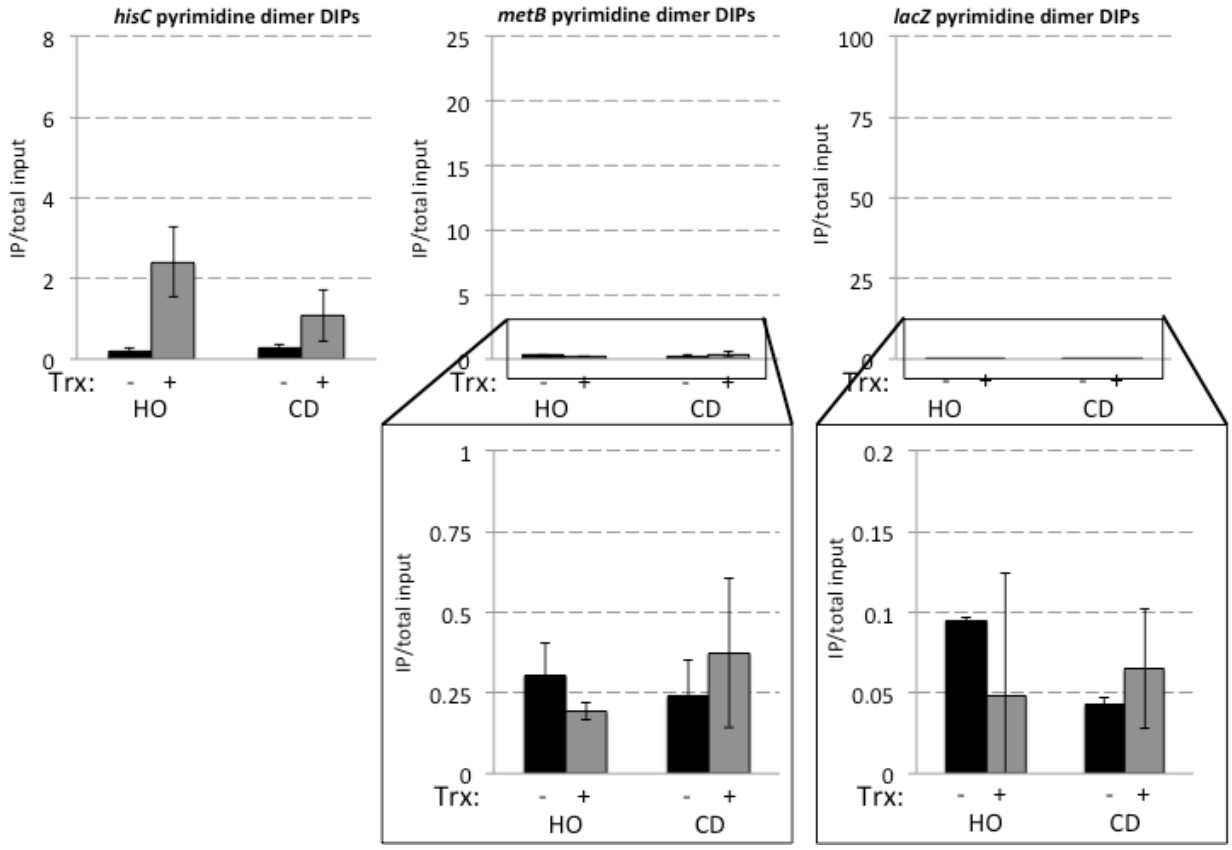


Figure 5.4 Transcription asymmetrically increases DIP signal in the absence of UV damage for the *hisC* reporter. Cells harboring the *hisC*, *metB*, or *lacZ* reporters head-on (HO, left) and co-directionally (CD, right) to replication were grown under conditions of high (Trx +, gray bars) and low transcription (Trx -, black bars) without exogenous UV damage. DNA was extracted and immunoprecipitated using the anti-pyrimidine dimer antibody. Relative enrichment of *hisC*, *metB*, or *lacZ* DNA in the DIPs as compared to toal input controls was determined by qPCR using the primers HM495 & HM496, HM842 & HM843, and HM188 & HM189, respectively. The inset graphs for the *metB* and *lacZ* enrichment values are plotted on a reduced scale to facilitate comparisons due to the lack of signal in the absence of UV damage for these reporters.

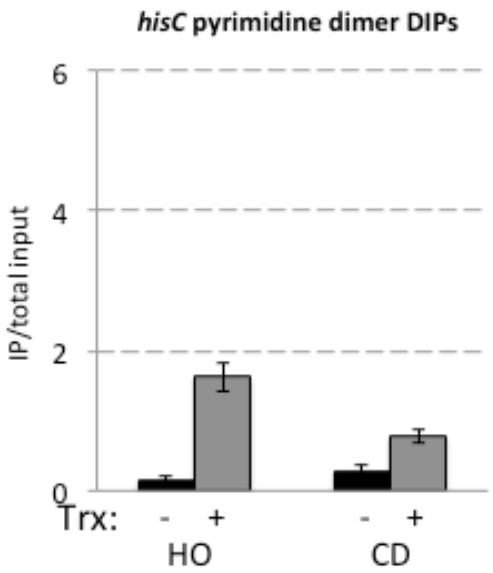


Figure 5.5 Cells grown in the dark display a transcription-dependent asymmetric increase in DIP signal for the *hisC* reporter. Cells harboring the *hisC* reporters head-on (HO, left) and co-directionally (CD, right) to replication were grown in the dark under conditions of high (Trx +, gray bars) and low transcription (Trx -, black bars) without exogenous UV damage. DNA was extracted and immunoprecipitated using the anti-pyrimidine dimer antibody. Relative enrichment of *hisC* DNA in the DIPs as compared to toal input controls was determined by qPCR using the primers HM495 & HM496.

Chapter 6. Thesis outlook

Replication-transcription conflicts are a double-edged sword. The aftermath of head-on collisions causes replication restart and mutagenic DNA repair, yet transcription-dependent mutations in lagging strand genes potentially confer adaptive advantages to bacteria. My work uncovered two distinct strategies cells use to cope with head-on conflicts: one of which allows replication to proceed in the immediate aftermath of conflicts, and another that generates genetic diversity while repairing collateral damage to the DNA.

The work described in this thesis reveals two mechanisms for managing the consequences of lagging-strand transcription, while opening the door to several further avenues of inquiry. The fine molecular details of collisions at the fork, as well as the long-term adaptive effects of mutagenesis in lagging strand genes remain to be elucidated. Further research into conflicts will yield important insights into fundamental mechanisms for genome maintenance and evolution in *Bacillus subtilis*, and potentially all life.

6.1 The role of conflicts in generating genetic diversity

According to classic evolutionary theory, bacteria decrease the potential of deleterious variations by maintaining mutation rates near a theoretical minimum threshold (97, 166, 257-259). Indeed, in the absence of selective pressure, the genomes *E. coli* and *B. subtilis* accumulate changes at a rate of roughly $1x10^{-3}$ mutations per genome per generation (211, 248). However, external growth conditions and local chromosomal context both can modulate mutation rates in cells (163, 260, 261). Asymmetric mutagenesis through head-on replication-transcription conflicts allows cells to spatially and temporally regulate mutation rates throughout the chromosome through the timing of expression lagging-strand oriented genes.

6.1 a) Spatial regulation of mutagenesis throughout the chromosome by conflicts

Mutagenesis is not uniform throughout the chromosome; local sequence context is well established to promote specific types of mutations in surrounding nucleotides. Mutation accumulation experiments in the absence of selection, followed by whole genome sequencing for both *E. coli* and *B.*

subtilis identified a role for purine tracks in promoting transitions(211, 248). Multiple mechanisms can increase sequence variation within short, specific motifs. Certain repeats promote mutagenesis through a template-switching mechanism (83). Context also influences polymerase fidelity (181, 205), and promotes the formation of multiple types of DNA lesions, including pyrimidine dimers (250). These studies of context-dependent mutagenesis, however, largely ignored gene-orientation in their analyses but rather focused on the effects of several nucleotides surrounding a specific polymorphism.

Short sequence motifs that promote mutagenesis are, necessarily, limited to specific close-range effects. Mutations are classically defined as evolutionarily silent if they do not change the amino acid sequence of the encoded product (262). However, selection is likely minimal on mutations that change amino acids in functionally unimportant regions of a protein such as non-catalytic, exterior loops(263). Therefore, close-range mechanisms for modulating mutation rates in a gene are limited by the architecture of the encoded protein. Additionally, the potential adaptive benefits conferred by motifs that promote local mutagenesis may be outweighed by functional constraints and purifying selection in surrounding regions of the gene (166). Gene-orientation, by contrast, is independent of the coding nucleotides and influences mutation rates across the entire coding sequence. Changing the orientation of a gene allows bacteria to explore a broad fitness landscape with every nucleotide becoming a potential target for variation (106).

It is formally possible that head-on collisions with transcription induce a heretofore-unidentified signal to promote mutagenesis genome-wide. Such a signal, if it exists, would likely arise through a small molecule generated at stalled forks so that cells could rapidly react to the collision, such as is observed in DisA-mediated inhibition of sporulation under conditions of DNA damage through cyclic-di-AMP (264). However, because I did not detect an increased frequency of rifampicin resistance mutations in cells experiencing head-on conflicts, asymmetric mutagenesis likely occurs in a gene-specific manner.

A key feature of conflict-mediated mutagenesis is increased variation in specific subsections of the genome. The possibility exists that lagging-strand genes simply tolerate variation, and thus a lack of negative selection underlies the head-on orientation of these loci. However, the functional enrichment for environmental response in lagging strand genes combined with the occurrence of

convergent substitutions hints at positive selection in favor of the head-on orientation of these genes (92, 167, 260). Additionally, the presence of a mechanism within cells that specifically increases mutation rates and promotes replication progression at regions of lagging strand transcription suggests that conflict-mediated mutagenesis may carry an adaptive benefit.

Investigations into how changes in gene-orientation influence mutation rates in laboratory evolution experiments under selective conditions could further uncover the relationship between genome organization and adaptation in bacteria. The work in this thesis adds gene-orientation as an additional variable that must be considered in the ongoing debate between the neutral drift and adaptation-selection models of bacterial evolution (259, 265).

6.1 b) Temporal regulation of mutagenesis through conflicts

Several well-characterized mechanisms exist in bacteria to increase genetic variation under specific conditions. DNA damage and nutrient-limitation promote mutagenesis through the upregulation of error-prone polymerases (99-101, 104, 160-163). Contingency loci, or preferential inactivation of the mismatch repair fidelity mechanisms (266), allow mutator strains of *Salmonella typhimirium, Yersinia pestis,* and *Helicobacter pylori* to emerge at a high frequency in high-stress environments , such as growth in a eukaryotic host (266, 267). Intriguingly, recent reports in *E. coli* have identified a role for social interactions in mutation rate plasticity, with mutagenesis promoted under low cell-density conditions (268). These mechanisms, however, globally increase mutation rates and thus may cause detrimental variation in essential genes necessary for growth under more favorable conditions (93).

The adaptive advantage conferred to mutator strains is finite and limited. Simulations and co-culture experiments demonstrate that while constitutive mutators initially outcompete wild-type populations, eventually the cost of the accumulated mutation burden overtakes any potentially beneficial variation (93, 269). Mutagenesis through head-on conflicts, by contrast, is a transient state for cells, contingent upon lagging strand gene expression.

Lagging strand genes tend to be expressed during times of stress and during lag phase growth (57, 58, 195). The 80-minute period between replication initiation during lag phase and the time-point when cells begin to increase in optical density is characterized by upregulation of over 100 lagging

strand genes in *B. subtilis* (195). Lagging strand genes show enrichment for nutrient uptake and stress response functions (270). This expression pattern may lend cells an additional layer of temporal regulation to conflict-mediated mutagenesis. Variation occurring in expressed proteins may be immediately acted upon by selection (106), meaning that any changes in lagging-strand genes advantageous for stress tolerance or nutrient uptake may rapidly allow mutants to overtake a population (95, 167, 271).

Robust transcription from the lagging strand is an obstacle to processive DNA synthesis. The conditions where lagging strand genes are expressed at high levels, i.e. during times of stress and during lag phase growth, are characterized by decreased replication rates. Exploring new sequence space through conflict-mediated mutagenesis during periods when replication is slow could serve as an adaptive bet-hedging strategy (92-94). The detrimental consequences of collisions, such as stalling and restart, are comparatively minor if replication is already attenuated. Additionally, periods of slow growth could allow for fidelity mechanisms such as mismatch repair to correct any severely deleterious mutations that would otherwise produce inviable progeny (166, 272).

An alternative, possibility is that lagging strand gene expression itself is the underlying mechanism for attenuated replication during lag phase growth or stress conditions. In this model, the coordinated expression of lagging strand genes does not occur because replication is attenuated, but rather is the causative agent of the impediment. Similar to the observed inhibition of replication elongation under low nutrient conditions in *E. coli* and *B. subtilis* through ppGpp(p)-dependent signaling pathways (273), a mechanism to slow replication via conflicts with transcription might serve two adaptive functions for cells. Decelerating replication could both increase the available time for biosynthesis before cell division and maximize expression of nutrient uptake genes from the lagging strand (270).

High-resolution transcriptomics and extensive high throughput whole genome sequencing of cells growing under multiple conditions will be necessary to untangle the temporal control of mutagenesis through replication-transcription conflicts. However, understanding mechanisms that vary mutation rates during different phases of the bacterial life cycle at different positions on the chromosome will yield important insight into bacterial adaptation.

6.2 The nature of the DNA at head-on conflict regions

My work raises fundamental questions about the nature of the DNA in regions where replication and transcription collide in bacteria. The recombination-dependent replication restart (**chapter 2**) occurring after head-on collisions between replication and transcription likely contributes to the increased mutagenesis in lagging strand genes (**chapters 3 &** 4) by exposing single-stranded DNA to lesion-causing agents (**chapter 5**). However, the specific types of DNA breaks inflicted by head-on conflicts, as well as the chemical DNA damage occurring at these regions remain to be determined.

Electron microscopy experiments have successfully visualized the DNA transactions occurring at Eukaryotic replication forks in response to treatment with a variety of genotoxic agents (152). Unfortunately, the rapid rate of bacterial replication makes these types of experiments infeasible for prokaryotic systems. *In vitro* experiments have yielded conflicting results concerning the effects of collisions on the DNA, depending upon which proteins are included in the reaction mixtures. However, recent advances in next-generation sequencing methodologies offer the tantalizing potential to untangle the uncertainty surrounding DNA dynamics at conflict regions within living cells (212, 274).

6.2 a) DNA lesions at head-on conflict regions

The results of the DNA immunoprecipitations indicate that head-on transcription increases the detectable levels of UV-induced pyrimidine dimers in these loci as compared to co-directional regions. The participation of PolY1 in TC-NER at regions of head-on transcription implies that lagging strand genes experienced increased levels of DNA lesions. However, the mutagenesis experiments that uncovered PolY1's role in preventing genomic instability while promoting genetic variation, at regions of head-on transcription were performed in the absence of exogenous DNA damaging agents. The exact nature of the lesions occurring at head-on conflict regions remains to be determined.

Although TC-NER classically processes bulky UV lesions, this pathway has the potential to act at undamaged DNA (232). The involvement of NER in mutagenesis without exogenous DNA damage has been demonstrated in *S. cerevesiae* (235) as well as *P. putida* (275). Studies in *E. coli* indicate that up to 0.02% of the genome is synthesized *de novo* due to repair every cell cycle even without exposure to

DNA damaging agents (202, 222). “Gratuitous” TCR at undamaged DNA may occur when the transcription-repair coupling factor, Mfd, inappropriately recruits the excision repair machinery to paused or backtracked RNA polymerases independent of DNA lesions (202). Persistently stalled RNA polymerases in lagging strand genes, arrested by an oncoming replisome, could activate transcription-coupled repair even in the absence of damage.

“Gratuitous TCR” could underlie the asymmetric mutagenesis in lagging-strand genes; however, the results of the DNA immunoprecipitations discussed in **chapter 5** indicate that head-on transcription increases the susceptibility of the DNA to at least one type of bulky lesion. While UV damage may have played a role in generating genetic diversity in enviromental *B. subtilis* over evolutionary time, other alterations to the chemical structure of the DNA probably promote mutagenesis at head-on conflicts, especially under indoor laboratory growth conditions.

Although the canonical substrate for TC-NER is the UV-induced pyrimidine dimer, the pathway may repair other types of DNA damage, some of which arise from normal cellular metabolism. TC-NER plays a role in removing misincorporated ribonucleotides from the DNA (224) that escape recognition and digestion by RNaseHII (276). A high local concentration of ribonucleotides at regions of replication-transcription collision increases the likelihood of these anomalous DNA structures (219). However, mutation rates in RNaseHII-deficient cells were indistinguishable from wild-type backgrounds (data not shown). Additionally, the rate of misincorporation should be independent of replication orientation, yet PolY1 specifically increases mutation rates at regions of head-on transcription. Therefore, ribonucleotides within the DNA do not likely significantly contribute to asymmetric, transcription-dependent mutagenesis in *Bacillus subtilis*.

Reactive oxygen species are a major source of chemical DNA damage within cells (189). While non-bulky 8-oxo-dG adducts caused by oxidative damage are typically recognized by base excision repair, some products, such as the bulky thymine-glycol lesion or 8,5 - cyclopurine-2 - deoxynucleosides (254), activate TC-NER because of their strong transcription-blocking effects. Additionally, the AP site generated as the first step of BER may itself be a substrate for NER (227, 229). Indeed, NER-deficient cell lines display sensitivity to hydrogen peroxide, suggesting that this pathway helps cells tolerate oxidative stress (256). Reactive oxygen species arise due to normal cellular

metabolism, and single stranded DNA at head-on conflict regions would be particularly susceptible to electron radical attack (189).

Elucidating the exact nature of the lesions necessitating TC-NER at head-on conflict regions would require highly sensitive mass spectrometric analysis. However, advances in next-generation sequencing have allowed for genome-wide mapping of some specific lesions in Eukaryotic cells. Recently the genome-wide localization of misincorporated uracil bases and two types of UV-photoproducts (the CPD and 6,4 TT dimer) in *S. pombe* were determined with high precision using a methodology termed "Excision-seq." The technique was limited, however, by a requirement for large amounts of lesions in the DNA; experiments were performed in mutant backgrounds, subjected to dramatically high levels of genotoxic stress (274).

Understanding the contribution of the interplay between transcription and replication to alter DNA susceptibility to exogenous and endogenous agents could uncover new insight into mechanisms of genomic instability, potentially applicable across all domains of life.

6.2 b) Direct contact between the replisome and RNA polymerase?

One provocative mystery is whether the replication and transcription machineries ever come into physical contact. During head-on gene expression, the replicative helicase, DnaC, and RNA Polymerase both translocate along the same lagging-strand template. However, because both complexes unwind the double helix positive supercoils accumulate ahead of both machineries (144-146, 247). Therefore, torsional strain in the DNA might halt the proteins before direct head-on collisions occur.

B. subtilis encodes two essential enzymes capable of relaxing positive supercoils: DNA Gyrase and Topoisomerase IV (277, 278). The topoisomerases are not required for *in vitro* replication reactions using linear DNA substrates (where torsional strain may be relieved by free rotation of the template) (23). However, positive supercoils arrest reconstituted replisomes lacking gyrase on circular plasmids (279). Gyrase translocates ahead of the fork *in vivo*, introducing negative supercoils (278). Overwound DNA in the region between the replication fork and RNA polymerase could overwhelm Gyrase's relaxation kinetics, preventing a physical encounter between the machineries.

However, because the replication fork does not present a barrier to lateral supercoil diffusion along the DNA, positive supercoils may disperse behind the replisome to form precatenanes (280)(which are relieved by Topoisomerase IV) (144). Supercoil diffusion may disperse torsional strain as the machineries advance, allowing for direct collisions between the helicase and RNA polymerase. In support of this model, studies in *E. coli* demonstrated that expression from a lagging-strand promoter only inhibited replication within the transcribed region (9). Additionally, mutagenesis studies indicated that the footprint of transcription-associated DNA damage is confined to the open reading frames of genes (281).

Cross-linking or *in vivo* FRET experiments between the helicase and RNA polymerase could address the question of whether the replisome and transcription proteins come into close proximity during lagging strand gene expression. Additionally, MNase or DNase protection assays, probing regions where the replisome and RNA polymerase approach each other head-on, such as at the 3' ends of lagging-strand genes, could reveal whether DNA in these loci is physically occupied by proteins.

6.2 c) Does DNA topology contribute to replication-transcription conflict phenotypes?

Regardless of whether the machineries come into direct physical contact, topological constraints likely play a significant role in the detrimental consequences of conflicts. Although supercoils freely diffuse behind replication forks, RNA polymerase presents a topological barrier on the DNA (144, 247). Therefore, local supercoiling effects may be magnified when the replisome encounters transcription head-on. Positive supercoils build up ahead of replication forks and RNA polymerases (146, 247), increasing transcription-associated recombination (45, 215), potentially breaking the DNA, or causing cruciform extrusion (70). Negative supercoils accumulate behind elongating RNA polymerases (145), promoting R-loop and G-quadruplex formation (282, 283). All of these structures present significant obstacles to replication.

In addition to impeding replication, supercoiling-associated structures may cause prolonged exposure of single-stranded DNA. R-Loops, secondary structures, and break repair all disrupt double helical base pairing (234, 284). The increased lesion susceptibility observed in transcribed head-on genes could arise because of a supercoiling induced DNA transition. Co-directional encounters between

the replisome and RNA polymerase likely rewind DNA within the transcription unit, whereas underwound DNA would accumulate in these regions during head-on collisions (45, 145, 146, 216), potentially setting the stage for chemical damage.

Investigations into DNA topology in bacteria have either focused on global chromosomal supercoiling domains or made use of plasmid systems. While the essential topoisomerases, DNA gyrase and Topoisomerase IV, play well-characterized roles in relaxing supercoils on both sides of the replication fork (144), other factors likely contribute to linking-number maintenance at conflict regions. Studies in yeast, for example, have begun to demonstrate crucial roles for THO/TREX in relieving topological strain and R-loop formation associated with head-on transcription (13). Further work is needed to understand how the architecture of the DNA template itself contributes to the effects of replication-transcription conflicts on bacterial physiology.

Thesis outlook

In the 20 years since Bonita Brewer's and Sarah French's landmark observations, studies of replication-transcription conflicts continuously reveal unappreciated aspects of these encounters. From accelerating evolution to impeding replication fork progression, it is clear that lagging strand gene expression carries conflicting consequences for cells. Understanding the interplay between the fundamental processes of the central dogma of molecular biology has far-reaching applications to all life.

7. Methods

7.1 *B. subtilis* strains.

Experiments were performed in *B. subtilis* 168, or the prophage-cured auxotrophic derivative YB955. Genetic manipulation of *B. subtilis* was performed as described (100). Knockout strains were constructed via allelic exchange, or obtained from the Bacillus Genetic Stock Center (Koo BM et al., *in preparation*). Markerless deletions were made by evicting Erythromycin^R cassettes via flanking *loxP* sites and a plasmid-encoded Cre recombinase (pDR224 carries the *cre* gene and a temperature-sensitive origin of replication), obtained from the BGSC. Strains harboring head-on and co-directional reporter constructs were constructed as detailed in (17, 18). A full list of *B. subtilis* strains used and generated in this work can be found in **supplementary table 8.1.**

7.2 Plasmid construction.

Plasmids were constructed using standard molecular biological techniques. All plasmids were propagated in *E. coli* DH5α under ampicillin selection. Plasmids were isolated using the 5-Prime Plasmid Mini kit. PCR reactions used Phusion Flash DNA polymerase (Thermo) or the Extend Long-Template PCR system (Roche). Linear DNA fragments were isolated using the 5-Prime Agarose Gel Extract kit. DNA was digested and ligated using Fast Digest Restriction Enzymes and Rapid Ligase with the appropriate buffers (Fermentas). A full list of plasmids may be found in **supplementary table 8.2.**

7.3 Growth Conditions.

Strains were streaked on LB agar plates supplemented with the appropriate antibiotic. Pre-cultures were inoculated from single colonies into 10 mL of LB medium supplemented with appropriate antibiotic and incubated at 37 °C, with shaking.

7.4 Detailed media compositions.

The concentrations for each antibiotic used were: spectinomycin (50 µg/ml), kanamycin (50 µg/ml), chloramphenicol (50 µg/ml), or lincomycin and erythromycin at 25 µg/ml each. *E. coli* was

grown in LB supplemented with ampicillin at 100 µg/ml. To determine viabilities for fluctuation analyses, *B. subtilis* cells were plated on solid agar plates containing Spizizen's Minimal Medium (0.2 mg/ml ammonium sulphate, 1.4 mg/ml monobasic potassium phosphate, 0.6 mg/ml dibasic potassium phosphate, 0.1 mg/ml sodium citrate dihydrate, 0.02 mg/ml magnesium sulphate heptahydrate, 100 µg/ml glutamic acid, 5 µg/ml glucose), supplemented with isoleucine, methionine, leucine and histidine at 50 µg/ml. To select for reversion to prototrophy in fluctuation analyses, cells were plated on solid agar plates containing essentially the above medium, lacking the amino acid corresponding to the reporter gene to be tested.

7.5 Plating efficiency assays

Cultures were grown in LB supplemented with the appropriate antibiotic at 37° with aeration (260 rpm) to exponential phase (OD = 0.3). Exponential phase cultures that grew to higher optical densities were diluted to OD 0.3. Serial dilutions were plated on LB-agar medium and total viable cells were quantified for each experimental condition. The plating efficiency ratio for a given genetic background was determined by dividing the viable colonies arising for strains expressing the *lacZ* gene to repressed for the reporter.

7.6 Chromatin Immuno-Precipitations.

Pre-cultures were initiated from isolated colonies growing on solid agar plates supplemented with the appropriate antibiotic and grown at 37ºC, with shaking, to OD=0.3-0.4. Starter cultures were diluted to OD=0.05 in 25 ml LB (supplemented with 1mM IPTG, where indicated to induce transcription of the reporters) and grown again at 30ºC, with shaking, to OD=0.3. Upon reaching OD=0.3, cultures were harvested into 0.1% formaldehyde and processed as described in (8). IPs were performed using rabbit polyclonal antibodies against DnaD, DnaC, RpoB, or GFP. The DnaD and DnaC antibodies were provided by the laboratory of Dr. Allan Grossman. The RpoB and GFP antibodies were purchased from AbCam (#Ab290 & 8RB13). IPs were rotated overnight at 4 °C for GFP, and DnaD ChIPs, and for 2 h at room temperature for RpoB and DnaC ChIPs.

7.7 Quantitative PCRs.

DNA was isolated by phenol-chloroform extraction. Quantitative PCRs (qPCRs) were performed using SsoAdvanced SYBR Green master mix and the CFX96 Touch Real-Time PCR system (Bio-Rad). Data were normalized to gene copy number by the ratios of total input to IP samples. Relative enrichment was determined by the ratio of gene copy number for the locus of interest to a control locus. A list of primer pairs and their amplification targets can be found in **supplementary table 8.4.**

7.8 Microscopy.

Precultures were set back to OD 0.05 and grown to exponential phase (OD 0.3–0.6). Cells were fixed with 4% (vol/vol) formaldehyde, DAPI-stained, and spotted onto 1% agarose pads. Images were taken using a Nikon Ti-E inverted microscope fitted with a 60× oil objective, automated focusing (Perfect Focus System; Nikon), a xenon light source (Sutter Instruments), and a CCD (Nikon). DAPI-stained nucleoids and RecA-GFP foci were counted in ImageJ. At least 1,000 cells were counted per 6 to 12 biological replicates per strain.

7.9 Mutation Rates.

Fluctuation analyses were performed as described previously (12, 101). Briefly, precultures were grown for ~14 h, and then diluted to an OD_{600} of 0.0005 in individual 18-mm culture tubes containing 2 mL of fresh LB either supplemented or not with 1 mM IPTG. Cultures were incubated at 37°C, with shaking, until reaching OD_{600} of 0.5. A volume of 1.5 mL of each culture was centrifuged for 2 minutes at 12,000 rpm, resuspended in 1× Spizizen's salts, and plated on minimal medium lacking the appropriate amino acid either with or without IPTG. The remainder of each culture was serially diluted and plated to quantify viable cells. Colonies were enumerated after 40- to 48-h incubation at 37 °C. Mutation rates were estimated using the Ma–Sandri–Sarkar maximum-likelihood method as described (Hall, et al., *Bioinformatics*, 2009)

7.10 Identification of leading and lagging strand genes.

We considered the origin of replication to be positioned between ribosomal protein L34 and chromosomal replication initiation protein (DnaA), whereas the terminus region before replication terminator protein (Rtp) for all 5 strains (Kunst et al. 1997). Depending on these locations we defined the genes present in leading strand and lagging strand for each strain. For analysis, we considered those genes which are present in the same strand (either leading or lagging) for all five strains. In total we found 759 core polymorphic genes, 627 in the leading strand and 132 in the lagging strand

7.11 Bioinformatics analysis.

For the mutation rate analysis, the sequence of eight complete genomes of *B. subtilis* were obtained from National Center for Biotechnology Information GenBank: *B. subtilis* 168, *B. subtilis* 6051-HGW, *B. subtilis* BSn5, *B. subtilis* BSP1, *B. subtilis* RO-NN-1, *B. subtilis* XF-1, *B. subtilis* subsp. *spizizenii* TU-B-10, and *B. subtilis* subsp. *spizizenii* W23. Core genes were identified as homologous genes found in all eight genomes, for which none were annotated as pseudogenes, using the Prokaryotic Genome Analysis Tool, PGAT (102). Further filtering of the set of homologs was performed using the Perl programming language (http://www.perl.org). Only single-copy genes were considered to avoid false SNP discovery in duplicated genes or paralogs. Highly diverse genes were excluded by selecting only homologs with at least 90% sequence identity and 90% length coverage. Gene sequences were then aligned using Muscle (103) and parsed to identify synonymous and nonsynonymous substitutions. Genes with sequence alignments having insertions or deletions were not considered to remove the possibility of false SNP identification due to potential ambiguous alignments. For the purpose of identification of leading- and lagging-strand genes, we considered the origin of replication to be located between ribosomal protein L34 and the chromosomal replication initiation protein (DnaA), and the terminus region to be upstream of the replication terminator protein (Rtp). This screen identified 2,662 core genes, of which 1,993 (75%) were encoded on the leading strand and 669 (25%) were encoded on the lagging strand.

7.12 Survival assays.

Exponentially growing pre-cultures (OD 0.3-0.6) were set back to OD 0.05 in LB, and grown to exponential phase again (OD 0.3-0.6). Cells were serially diluted $1:10^{-5}$ and 5 µl of each dilution was spotted onto LB agar plates either supplemented or not with 4-Nitro-Quinolone Oxide. To determine UV sensitivity cells were spotted onto LB agar plates and exposed to UV using a Mineralight XX 15V UV light source (UVP). Surviving colonies were enumerated after overnight incubation at 30°C.

7.13 Transcript levels.

Exponentially growing pre-cultures (OD 0.3-0.6) were set back to OD 0.05 in 7.5 ml LB, and grown to exponential phase again. Upon reaching OD=0.3, cultures were harvested directly into 7.5 ml ice-cold 100% methanol. Cells were collected by centrifugation, re-suspended in 100 µl TE pH 7.0 supplemented with 20 mg/ml lysozyme and then incubated at room temperature for 15 minutes. Total RNA was isolated using the Qiagen RNeasy RNA extraction kit. DNA was removed from the RNA preps by following the supplied on-column DNaseI treatment protocol. Total RNA was quantified, then reverse transcribed into cDNA using the Roche Transcriptor 1st Strand Synthesis Kit. cDNA was quantified, diluted to 2 ng/µl, then analyzed by quantitative PCR.

7.14 Growth curves.

Exponentially growing pre-cultures were diluted to OD=0.005 into minimal medium lacking histidine and supplemented with 1mM IPTG (to induce transcription of the particular *hisC* allele). Cultures were aliquoted into the wells of a clear-bottomed 96-well plate. ODs were measured at seven-minute intervals using a Biotek Synergy H1 plate reader, using a kinetic program with continuous shaking at 37oC.

7.15 Mutation footprint analysis.

Precultures were grown for ~14 h, and then diluted to an OD_{600} of 0.0005 into 24 individual 18-mm culture tubes per strain and condition containing 2 mL of fresh LB either supplemented or not with 1 mM IPTG, as for **mutation rate** measurements. Cultures were incubated at 37°C, with shaking, until

reaching OD_{600} of 0.5. Individual cultures were pooled, collected by centrifugation, then spread onto 150 mm petri dishes containing solid agar minimal medium lacking histidine. After 48 hours incubation at 37°C 100 prototrophic revertant colonies from each strain and condition were picked into 500 µl 1x spizizen's salts. Genomic DNA was prepared by phenol-chloroform extraction. The *hisC* locus was amplified by low-cycle PCR using custom primers with illumina adapters and unique indices for each strain and condition. PCR reactions were performed in 100 µl volumes for 15 cycles, to avoid amplification biases. PCR reactions were ethanol precipitated, rehydrated in 50 µl TE pH 7.6, analyzed by agarose gel electrophoresis. The band corresponding to the size of the expected amplification product was excised and DNA was purified using the 5' GelExtract column clean up kit. Libraries were quantified using a qBit fluorimeter, diluted to 2 picomolar, then loaded into an illumina miSeq 50 cycle v2 reagent cartridge along with custom index and sequencing primers. Libraries were sequenced on the illumina miSeq next-generation sequencing platform. Paired-end reads were concatenated and sequences were analyzed in 30 nucleotide sliding windows along the length of the *hisC* gene using custom perl scripts.

7.16 SNP Analysis.

To determine the dependence of the rate of mutation on gene length, counts for the pairs of substitutions AG and GA, AC and CA, AT and TA, etc., were pooled because the direction of the substitutions was not determined. Mutation counts were modeled as a function of gene nucleotide length (5,000 bases or less) and strand (leading vs. lagging) using negative binomial regression with a log link function. The negative binomial model accounts for zero inflation and dispersion in the frequency distribution of mutations counts. Zero inflation is a consequence of the presence of short genes and highly conserved genes in which not all possible mutations occur. Dispersion is a consequence of the large range in the level of gene conservation. All statistical analyses were performed using R (The R Project for Statistical Computing; http://www.R-project.org).

7.17 DNA Immunoprecipitations

Exponentially growing pre-cultures were diluted to OD600= 0.05 in 12 ml LB medium either in the presence or absence of 1 mM IPTG. Cultures were grown to OD_{600}= 0.3 centrifuged, and resuspended in 12 ml 1X spizizen's salts. Cultures were split into 6 ml aliquots and either and UV treated at 500 J/m^2 or harvested directly into 6 ml ice-cold methanol. DNA was isolated by phenol-chloroform extraction. 1 µg DNA was added to either to 1 ml IP buffer, or 49 µl TE pH 8.0 contiaing 2.5% SDS for total input samples. Samples were boiled 5 minutes, and chilled on ice. 1 µl anti-TT dimer antibody (Genetex) was added to IP samples and incubated overnight at 4°C.

7.18 *in vitro* photorepair

For T4 PDG treatment, 0.5 µg genomic DNA was denatured by boiling, UV irradiated to 100 J/m^2 and then added to 20 µl reaction volumes containing 100 mM NaCl,1 mM DTT, 1 mM EDTA, 25 mM Na_2HPO_4 , pH 7.2 supplemented with 100X Bovine Serum Albumin. 1 µl 10% glycerol or 10 units of T4 PDG (New England Biolabs) was added to control, "mock," reactions and treatment reactions, respectively. Reactions were incubated at room temperature for 20 minutes, then DNA was isolated by phenol-chloroform extraction, added to 1X IP buffer, fragmented by sonication, denatured by boiling, and immuno-precipitated as described above. The relative amounts of *hisC* containing DNA in IP and total samples was determined by qPCR.

For photolyase treatment, 0.5 µg genomic DNA was denatured by boiling, UV irradiated to 100 J/m^2 and then added to 200 µl reaction volumes containing 20 mM Phosphate buffer pH7.5, 2mM EDTA, 0.1 mM DTT and either 0.5 µg purified *E. coli* PhrB enzyme (Acris Antibodies Catalogue # AR50868PU-S) or an equivalent volume of sterile water ("mock" treatment controls). Reactions were incubated at room temperature subject to visible light illumination platform light-box for 20 minutes. DNA was isolated by phenol-chloroform extraction, added to 1X IP buffer, sonicated, boiled and immuno-precipitated as described above. The relative amounts of *metB* containing DNA in IP and total samples was determined by qPCR.

8. Supplementary tables

Supplementary table 8.1: *B. Subtilis strains*

Strain	Genotype	Reference
HM1	*trpC2, pheA1*	
HM209	*trpC2*, *pheA1*, Δ*recA*-MLS/CAT	Cheo, *Biochimie*, 1992
HM210	*trpC2, pheA1, thrC*::P_{xis}-*lacZ,*	Merrikh, *Nature*, 2011
HM211	*trpC2, pheA1, thrC*::P_{xis}-*lacZ* ICE*Bs1*(0)	Merrikh, *Nature*, 2011
HM253	*trpC2, pheA1*, ICE*Bs1*(0)	Merrikh, *Nature*, 2011
HM313	*trpC2, pheA1, recA-mGFP*mut2-$spec_R$	Simmons, *PNAS*, 2007
HM352	*trpC2, pheA1, thrC*::P_{xis}-*lacZ*, ICE*Bs1*(0), *recA-mGFP*mut2-$spec_R$	Merrikh, *Nature*, 2011
HM391	*trpC2, pheA1, yqjH::CAT*	Million-Weaver et al., PNAS, 2015
HM380 (YB955)	*hisC952, metB5, leuC427,* SPβ$_{sens}$, xin$_{-1}$	Yasbin, *Genetics of industrial microorganisms*, 1987
HM410	ΔrecO-CAT	Simmons, *Journal of Bacteriology*, 2009
HM415	*trpC2, pheA1, addAB::KAN*	Million-Weaver et al., PNAS, 2015
HM417	*hisC952, metB5, leuC427,* SPβ$_{sens}$, xin$_{-1}$, *amyE::*$P_{spank\ (hy)}$-*leuC4272* [head-on] $spec_R$	Million-Weaver et al., PNAS, 2015
HM418	*hisC952, metB5, leuC427,* SPβ$_{sens}$, xin$_{-1}$, *amyE::*$P_{spank\ (hy)}$-*leuC427* [co-directional] $spec_R$	Million-Weaver et al., PNAS, 2015
HM419	*hisC952, metB5, leuC427,* SPβ$_{sens}$, xin$_{-1}$, *amyE::*$P_{spank\ (hy)}$-*hisC952* [head-on] $spec_R$	Paul et al., Nature, 2013, 2013
HM420	*hisC952, metB5, leuC427,* SPβ$_{sens}$, xin$_{-1}$, *amyE::*$P_{spank\ (hy)}$-*hisC952* [co-directional] $spec_R$	Paul et al., Nature, 2013, 2013
HM421	*hisC952, metB5, leuC427,* SPβ$_{sens}$, xin$_{-1}$, *amyE::*$P_{spank\ (hy)}$-*hisC952* [head-on] $spec_R$, *yqjH::CAT*	Million-Weaver et al., PNAS, 2015
HM422	*hisC952, metB5, leuC427,* SPβ$_{sens}$, xin$_{-1}$, *amyE::*$P_{spank\ (hy)}$-*hisC952* [co-directional] $spec_R$, *yqjH::CAT*	Million-Weaver et al., PNAS, 2015
HM425	*hisC952, metB5, leuC427,* SPβ$_{sens}$, xin$_{-1}$, *amyE::*$P_{spank\ (hy)}$-*hisC952* [head-on] $spec_R$, *yqjW::CAT*	Million-Weaver et al., PNAS, 2015
HM426	*hisC952, metB5, leuC427, SPβsens, xin-1, amyE::Pspank (hy)-hisC952 [co-directional] specR, yqjW::CAT*	Million-Weaver et al., PNAS, 2015
HM430	*trpC2, pheA1*, *addAB*::Kan	Million-Weaver et al., PNAS, 2015
HM445	*hisC952, metB5, leuC427, SPβsens, xin-1, amyE::Pspank (hy)-hisC952 [head-on] specR, mfd::CmR*	Million-Weaver et al., PNAS, 2015
HM444	*hisC952, metB5, leuC427, SPβsens, xin-1, amyE::Pspank (hy)- hisC952 [co-directional] specR, mfd::CmR*	Million-Weaver et al., PNAS, 2015
HM449	*hisC952, metB5, leuC427, SPβsens, xin-1, thrC::Pspank (hy)-hisC952 [head-on] MLSR*	Million-Weaver et al., PNAS, 2015

HM450	*hisC952, metB5, leuC427, SPβsens, xin-1, thrC::Pspank (hy)-hisC952 [co-directional] MLSR*	Million-Weaver et al., PNAS, 2015
HM594	*hisC952, metB5, leuC427, SPβsens, xin-1, thrC::Pspank (hy)-hisC952 [head-on] MLSR, recA-GFP-specR*	Million-Weaver et al., PNAS, 2015
HM595	*hisC952, metB5, leuC427, SPβsens, xin-1, thrC::Pspank (hy)-hisC952 [co-directional] MLSR recA-GFP-specR*	Million-Weaver et al., PNAS, 2015
HM596	*hisC952, metB5, leuC427, SPβsens, xin-1, thrC::Pspank (hy)-hisC952 [head-on] MLSR, yqjH::CAT recA-GFP-specR*	Million-Weaver et al., PNAS, 2015
HM597	*hisC952, metB5, leuC427, SPβsens, xin-1, thrC::Pspank (hy)-hisC952 [co-directional] MLSR, yqjH::CAT recA-GFP-specR*	Million-Weaver et al., PNAS, 2015
HM611	*hisC952, metB5, leuC427, SPβsens, xin-1, thrC::Pspank (hy)-hisC952 [head-on] MLSR, yqjH::CAT*	Million-Weaver et al., PNAS, 2015
HM612	*hisC952, metB5, leuC427, SPβsens, xin-1, thrC::Pspank (hy)-hisC952 [co-directional] MLSR, yqjH::CAT*	Million-Weaver et al., PNAS, 2015
HM 624	trpC2, pheA1, thrC::Pxis-lacZ [head-on] MLS^R	Merrikh, Nature, 2011
HM630	*trpC2, pheA1, thrC::Pxis-lacZ [head-on] MLSR, recA-mGFPmut2-specR*	Million-Weaver et al., PNAS, 2015
HM631	*trpC2, pheA1, thrC::Pxis-lacZ [head-on] MLSR, yqjH::CAT, recA-mGFPmut2-specR*	Million-Weaver et al., PNAS, 2015
HM632	*hisC952, metB5, leuC427, SPβsens, xin-1, amyE::Pspank (hy)-metB5 [head-on] specR*	Million-Weaver et al., PNAS, 2015
HM633	*hisC952, metB5, leuC427, SPβsens, xin-1, amyE::Pspank (hy)-metB5 [co-directional] specR*	Million-Weaver et al., PNAS, 2015
HM634	*hisC952, metB5, leuC427, SPβsens, xin-1, amyE::Pspank (hy)-metB5 [head-on] specR, yqjH::CAT*	Million-Weaver et al., PNAS, 2015
HM635	*hisC952, metB5, leuC427, SPβsens, xin-1, amyE::Pspank (hy)-metB5 [co-directional] specR, yqjH::CAT*	Million-Weaver et al., PNAS, 2015
HM639	*trpC2, pheA1, thrC::Pxis-lacZ [co-directional] MLS^R*	Million-Weaver et al., PNAS, 2015
HM640	*trpC2, pheA1, thrC::Pxis-lacZ [co-directional] MLS^R, ICEBs1(0)*	Million-Weaver et al., PNAS, 2015
HM663	*trpC2, pheA1, thrC::Pxis-lacZ [co-directional] MLSR, recA-mGFPmut2-specR*	Million-Weaver et al., PNAS, 2015
HM664	*trpC2, pheA1, thrC::Pxis-lacZ [co-directional] MLSR, recA-mGFPmut2-specR, ICEBs1(0)*	Million-Weaver et al., PNAS, 2015
HM667	*trpC2*, *pheA1*, *thrC*::Pxis-*lacZ* [head-on] MLS^R , *recA-mGFPmut2*-spec, *addAB::*Kan	Million-Weaver et al., J Bact, 2015
HM672	*hisC952, metB5, leuC427, SPβsens, xin-1, thrC::Pspank (hy)-metB5 [head-on] MLSR*	Million-Weaver et al., PNAS, 2015
HM673	*hisC952, metB5, leuC427, SPβsens, xin-1, thrC::Pspank (hy)-metB5 [co-directional] MLSR*	Million-Weaver et al., PNAS, 2015

HM674	*hisC952, metB5, leuC427, SPβsens, xin-1, thrC::Pspank (hy)-metB5 [head-on] MLSR, yqjH::CAT*	Million-Weaver et al., PNAS, 2015
HM675	*hisC952, metB5, leuC427, SPβsens, xin-1, thrC::Pspank (hy)-metB5 [co-directional] MLSR, yqjH::CAT*	Million-Weaver et al., PNAS, 2015
HM676	*trpC2*, *pheA1*, *thrC*::Pxis-*lacZ* [head-on] , MLS^R, *recO::*CAT	Million-Weaver et al., J Bact, 2015
HM677	*trpC2*, *pheA1*, *thrC*::Pxis-*lacZ* [head-on] MLS^R, *recA-mGFPmut2*-spec, *recO::*CAT	Million-Weaver et al., J Bact, 2015
HM684	*thrC::Pxis-lacZ [co-directional] MLSR, recA-mGFPmut2-specR, yqjH::CAT*	Million-Weaver et al., PNAS, 2015
HM685	*trpC2, pheA1, thrC::Pxis-lacZ [co-directional] MLSR, recA-mGFPmut2-specR, ICEBs1(0), yqjH::CAT*	Million-Weaver et al., PNAS, 2015
HM706 (BKE23870)	*yqjH::MLS, trpC2*	Million-Weaver et al., PNAS, 2015
HM713 (BKE35160)	*uvrA::MLS, trpC2*	Bacillus Genetic Stock Center
HM720	*hisC952, metB5, leuC427, SPβsens, xin-1, thrC::Pspank (hy)-hisC952 [head-on] MLSR, addAB::KAN*	Koo BM et al. (In preparation)
HM721	*hisC952, metB5, leuC427, SPβsens, xin-1, thrC::Pspank (hy)-hisC952 [co-directional] MLSR, addAB::KAN*	Bacillus Genetic Stock Center
HM724	*hisC952, metB5, leuC427, SPβsens, xin-1, amyE::Pspank (hy)-leuC427 [head-on] specR, yqjH::CAT*	Koo BM et al. (In preparation)
HM725	*hisC952, metB5, leuC427, SPβsens, xin-1, amyE::Pspank (hy)-leuC427 [co-directional] specR, yqjH::CAT*	Million-Weaver et al., PNAS, 2015
HM726	*hisC952, metB5, leuC427, SPβsens, xin-1, thrC::Pspank (hy)-hisC952 [head-on] MLSR, ΔrecA-MLS/CAT*	Million-Weaver et al., PNAS, 2015
HM727	*hisC952, metB5, leuC427, SPβsens, xin-1, thrC::Pspank (hy)-hisC952 [co-directional] MLSR, ΔrecA-MLS/CAT*	Million-Weaver et al., PNAS, 2015
HM730	*hisC952, metB5, leuC427, SPβsens, xin-1, thrC::Pspank (hy)-leuC427 [head-on] MLSR*	Million-Weaver et al., PNAS, 2015
HM731	*hisC952, metB5, leuC427, SPβsens, xin-1, thrC::Pspank (hy)-leuC427 [co-directional] MLSR*	Million-Weaver et al., PNAS, 2015
HM735	*hisC952, metB5, leuC427, SPβsens, xin-1, amyE::Pspank (hy)-hisC952 [head-on] specR, uvrA::MLS*	Million-Weaver et al., PNAS, 2015
HM736	*hisC952, metB5, leuC427, SPβsens, xin-1, amyE::Pspank (hy)- hisC952 [co-directional] specR, uvrA::MLS*	Million-Weaver et al., PNAS, 2015
HM737	*hisC952, metB5, leuC427, SPβsens, xin-1, thrC::Pspank (hy)-leuC427 [head-on] MLSR, yqjH::CAT*	Million-Weaver et al., PNAS, 2015
HM738	*hisC952, metB5, leuC427, SPβsens, xin-1, thrC::Pspank (hy)-leuC427 [co-directional] MLSR, yqjH::CAT*	Million-Weaver et al., PNAS, 2015
HM739	*hisC952, metB5, leuC427, SPβsens, xin-1, ΔyqjH*	

HM740	*hisC952, metB5, leuC427, SPβsens, xin-1, thrC::Pspank (hy)-hisC952 [head-on] MLSR, addAB::KAN, yqjH::CAT*	Million-Weaver et al., PNAS, 2015
HM741	*hisC952, metB5, leuC427, SPβsens, xin-1, thrC::Pspank (hy)-hisC952 [co-directional] MLSR, addAB::KAN, yqjH::CAT*	Million-Weaver et al., PNAS, 2015
HM742	*hisC952, metB5, leuC427, SPβsens, xin-1, thrC::Pspank (hy)-hisC952 [head-on] MLSR, ΔyqjH*	Million-Weaver et al., PNAS, 2015
HM743	*hisC952, metB5, leuC427, SPβsens, xin-1, thrC::Pspank (hy)-hisC952 [co-directional] MLSR, ΔyqjH*	Million-Weaver et al., PNAS, 2015
HM746	*hisC952, metB5, leuC427, SPβsens, xin-1, thrC::Pspank (hy)-hisC952 [head-on] MLSR, ΔyqjH, ΔrecA-MLS/CAT*	Million-Weaver et al., PNAS, 2015
HM747	*hisC952, metB5, leuC427, SPβsens, xin-1, thrC::Pspank (hy)-hisC952 [co-directional] MLSR, ΔyqjH, ΔrecA-MLS/CAT*	Million-Weaver et al., PNAS, 2015
HM760	*trpC2*, *pheA1*, *thrC*::Pxis-*lacZ*, ICEBs1(0) [head-on] , MLS^R, *recA-mGFPmut2*-specR, *addAB*::Kan	Million-Weaver et al., J Bact, 2015
HM761	*trpC2, pheA1, thrC::Pxis-lacZ [co-directional] MLS^R, amyE::Pspank (hy)-hisC952 [co-directional] specR*	Million-Weaver et al., PNAS, 2015
HM762	*trpC2, pheA1, thrC::Pxis-lacZ [co-directional] MLS^R, amyE::Pspank (hy)-hisC952 [co-directional] specR, ICEBs1(0)*	Million-Weaver et al., PNAS, 2015
HM765	*trpC2*, *pheA1*, *thrC*::Pxis-*lacZ* ICEBs1(0) [head-on] , MLS^R, *recO::*CAT	Million-Weaver et al., J Bact, 2015
HM766	*trpC2*, *pheA1*, *thrC*::Pxis-*lacZ* [head-on] , MLS^R, *recO::*CAT	Million-Weaver et al., J Bact, 2015
HM770	*trpC2*, *pheA1*, *thrC*::Pxis-*lacZ* [head-on], MLS^R, *addAB*::Kan	Million-Weaver et al., J Bact, 2015
HM807	*trpC2*, *pheA1*, *thrC*::Pxis-*lacZ* ICEBs1(0) [head-on] MLS^R, r*ecO::*CAT, *addAB*::Kan	Million-Weaver et al., J Bact, 2015
HM808	*trpC2*, *pheA1*, *thrC*::Pxis-*lacZ* [head-on] MLS^R, *recO*::CAT, *addAB::Kan*	Million-Weaver et al., J Bact, 2015
HM817	*trpC2*, *pheA1*, *thrC*::Pxis-*lacZ* [head-on] MLS^R, *recA-mGFPmut2*-spec, *recO*::CAT, *addAB*::Kan	Million-Weaver et al., J Bact, 2015
HM824	*trpC2, pheA1, thrC::Pxis-lacZ ICEBs1(0), ΔrecA-MLS/CAT*	Million-Weaver et al., PNAS, 2015
HM825	*trpC2, pheA1, thrC*::Pxis-lacZ [head-on] MLS^R, Δ*recA*-MLS/CAT	Million-Weaver et al., J Bact, 2015
HM854	*trpC2, pheA1, thrC*::Pxis-lacZ, ICEBs1(0), MLS^R, *recA*-mGFPmut2-specR, *recO*::CAT, *addAB*::Kan	Million-Weaver et al., J Bact, 2015
HM857	*trpC2, pheA1, thrC*::Pxis-lacZ ICEBs1(0) [head-on], MLS^R, *addAB*::Kan	Million-Weaver et al., J Bact, 2015
HM871	*trpC2, pheA1, thrC*::Pxis-lacZ [co-directional] MLS^R, ICEBs1(0), *recO*::CAT, *addAB*::Kan	Million-Weaver et al., J Bact, 2015
HM872	*trpC2, pheA1, thrC*::Pxis-lacZ [co-directional] MLS^R, *recO*::CAT, *addAB*::Kan	Million-Weaver et al., J Bact, 2015

HM875	*hisC952, metB5, leuC427, SPβsens, xin-1, amyE::Pspank (hy)- hisC952 [co-directional] specR, mfd::CmR, yqjH::MLS*	Million-Weaver et al., PNAS, 2015
HM907	*trpC2*, *pheA1*, *thrC*::Pxis-*lacZ* [co-directional] MLS^R, ICEBs1(0), Δ*recA*-MLS/CAT	Million-Weaver et al., J Bact, 2015
HM908	*trpC2*, *pheA1*, *thrC*::Pxis-*lacZ* [co-directional] MLS^R, Δ*recA*-MLS/CAT	Million-Weaver et al., J Bact, 2015
HM1069	*trpC2*, *pheA1*, *thrC*::Pxis-*lacZ* ICEBs1(0) [head-on] , MLS^R, Δ*recU*	Million-Weaver et al., J Bact, 2015
HM1076	*trpC2*, *pheA1*, *thrC*::Pxis-*lacZ* [head-on] , MLS^R, Δ*recU*	Million-Weaver et al., J Bact, 2015
HM1085	*trpC2*, *pheA1*, *thrC*::Pxis-*lacZ* ICEBs1(0) [head-on] , MLS^R, *recO::*CAT, Δ*recU*	Million-Weaver et al., J Bact, 2015
HM1089	*trpC2, pheA1, thrC*::Pxis-*lacZ* [head-on] , MLS^R, *recO*::CAT, Δ*recU*	Million-Weaver et al., J Bact, 2015
HM1090	*trpC2, pheA1, thrC*::Pxis-*lacZ* [head-on], MLS^R, *addAB*::Kan, Δ*recU*	Million-Weaver et al., J Bact, 2015
HM1102	*hisC952, metB5, leuC427, SPβsens, xin-1, amyE::Pspank (hy)-hisC952 [head-on] specR, uvrA::MLS*	Million-Weaver et al., PNAS, 2015
HM1103	*hisC952, metB5, leuC427, SPβsens, xin-1, amyE::Pspank (hy)- hisC952 [co-directional] specR, uvrA::MLS*	Million-Weaver et al., PNAS, 2015
HM1138	*trpC2, pheA1, thrC*::Pxis-*lacZ* [head-on] , MLS^R, *recO*::CAT, Δ*recU*, *addAB*::Kan	Million-Weaver et al., J Bact, 2015
HM1146	*trpC2, pheA1, thrC*::Pxis-*lacZ* ICEBs1(0) [head-on], MLS^R, Δ*recU*, Δ*recA*-MLS/CAT	Million-Weaver et al., J Bact, 2015
HM1155	*trpC2, pheA1, thrC*::Pxis-*lacZ* [head-on] , MLS^R, Δ*recU*, Δ*recA*-MLS/CAT	Million-Weaver et al., J Bact, 2015
HM1177	*hisC952, metB5, leuC427, SPβsens, xin-1, amyE::Pspank (hy)-hisC952 [head-on] specR, uvrA::MLS, yqjH::CAT*	Million-Weaver et al., PNAS, 2015
HM1178	*hisC952, metB5, leuC427, SPβsens, xin-1, amyE::Pspank (hy)- hisC952 [co-directional] specR, uvrA::MLS, yqjH::CAT*	Million-Weaver et al., PNAS, 2015
HM1231	*trpC2, pheA1, thrC*::Pxis-*lacZ* ICEBs1(0) [head-on], MLS^R, *addAB*::Kan, Δ*recU*	Million-Weaver et al., J Bact, 2015
HM1246	*trpC2, pheA1, thrC*::Pxis-*lacZ* [head-on] MLSR, *recO::*CAT, *addAB::*Kan, Δ*recA*-MLS/CAT	Million-Weaver et al., J Bact, 2015
HM1247	*trpC2, pheA1, thrC*::Pxis-*lacZ* ICEBs1(0) [head-on], MLS^R, *recO*::CAT, *addAB*::Kan, Δ*recA*-MLS/CAT	Million-Weaver et al., J Bact, 2015
HM1279 (BKE35710)	*uvrB::MLS, trpC2*	Bacillus Genetic Stock Center Koo BM et al. (In preparation)
HM1280 (BKE28490)	*uvrC::MLS, trpC2*	Bacillus Genetic Stock Center Koo BM et al. (In preparation)
BKE22310	trpC2, ΔrecO-MLS	Bacillus Genetic Stock Center Koo BM et al. (In preparation)

HM1281	*hisC952, metB5, leuC427, SPβsens, xin-1, amyE::Pspank (hy)-hisC952 [head-on] specR, uvrB::MLS*	Million-Weaver et al., PNAS, 2015
HM1282	*hisC952, metB5, leuC427, SPβsens, xin-1, amyE::Pspank (hy)- hisC952 [co-directional] specR, uvrB::MLS*	Million-Weaver et al., PNAS, 2015
HM1283	*hisC952, metB5, leuC427, SPβsens, xin-1, amyE::Pspank (hy)-hisC952 [head-on] specR, uvrC::MLS*	Koo BM et al. (In preparation)
HM1284	*hisC952, metB5, leuC427, SPβsens, xin-1, amyE::Pspank (hy)- hisC952 [co-directional] specR, uvrC::MLS*	Million-Weaver et al., PNAS, 2015
HM1287	*trpC2, pheA1, thrC*::Pxis-*lacZ* [head-on], MLS^R, *addAB*::Kan, Δ*recU*, Δ*recA*-MLS/CAT	Million-Weaver et al., J Bact, 2015
HM1288	*trpC2, pheA1, thrC*::Pxis-*lacZ* ICEBs1(0) [head-on], MLS^R, *addAB*::Kan, Δ*recU*, Δ*recA*-MLS/CAT	Million-Weaver et al., J Bact, 2015
HM1302	*trpC2, pheA1, thrC*::Pxis-*lacZ* [head-on] , MLS^R, *recO::*CAT, Δ*recU*, Δ*recA*-MLS/CAT	Million-Weaver et al., J Bact, 2015
HM1304	*trpC2, pheA1, thrC*::Pxis-*lacZ* ICEBs1(0) [head-on] , MLS^R, *recO*::CAT, Δ*recU*, Δ*recA*-MLS/CAT	Million-Weaver et al., J Bact, 2015
HM1305	*trpC2, pheA1*, *thrC*::Pxis-*lacZ* ICEBs1(0) [head-on], MLS^R, *recO*::CAT, Δ*recU*, *addAB*::Kan	Million-Weaver et al., J Bact, 2015

Supplementary table 8.2: Plasmids

Plasmid Number	Description	*B. subtilis* Integration Locus	*B. subtilis* Resistance Marker	*E. coli* Resistance Marker	Strain Number
pDR111	Integration vector with MCS and Pspank (hy)	amyE	spectinomycin^R	Amp^R	HM244
pCAL838	Integration vector with MCS and Pspank (hy)	thrC	MLS^R	Amp^R	HM252
pKG1	thrC::Pxis-LacZ head on to replication	thrC	MLS^R	Amp^R	HM312
pDR224	*rep*/*mob* from *S. aureus*, Ts *ori*, *cre* recombinase	N/A	spectinomycin^R	CAT	HM707
pHM59	pCAL838 hisC952 Head-On	thrC	MLS^R	Amp^R	HM423
pHM60	pCAL838 hisC952 Co Directional	thrC	MLS^R	Amp^R	HM424
pHM85	pDR111 with leuC427 + Tx Term at amyE head-on	amyE	spectinomycin^R	Amp^R	HM392
pHM88	pDR111 with hisC952 + Tx Term amyE head-on	amyE	spectinomycin^R	Amp^R	HM395
pHM89	pDR111 with leuC427 + Tx Term at amyE co-directional	amyE	spectinomycin^R	Amp^R	HM396
pHM91	pDR111 with hisC952 + Tx Term amyE co-directional	amyE	spectinomycin^R	Amp^R	HM398
pHM95	pDR111 with metB5 to be inserted at amyE head-on	amyE	spectinomycin^R	Amp^R	HM562
pHM96	pDR111 with metB5 to be inserted at amyE co-directional	amyE	spectinomycin^R	Amp^R	HM563
pHM107	Pxis-lacZ - Co-directional	thrC	MLS^R	Amp^R	HM636
pHM111	pCAL838 Pspank)hy) metB5 head on (inserts at thrC)	thrC	MLS^R	Amp^R	HM671
pHM112	pCAL838 Pspank)hy) metB5 co-directional (inserts at thrC)	thrC	MLS^R	Amp^R	HM672
pHM115	pCAL838 Pspank(hy) leuC427 + TxTerm to be inserted at thrC head-on	thrC	MLS^R	Amp^R	HM726
pHM116	pCAL838 Pspank(hy) leuC427 + TxTerm to be inserted at thrC co-directional	thrC	MLS^R	Amp^R	HM727

Supplementary table 8.3: Oligonucleotides

Number	Sequence	Use
HM112	CGCGTCAACAGGTAACACTTCC	Confirm integration into *thrC* or insertion into pCAL838
HM115	GGATCTTTCAGTCCGTTTCC	Confirm integration into *thrC* or insertion into pCAL838
HM205	TCCAAACTGGACACATGGAA	Confirm integration into *amyE* or insertion into pDR111
HM207	AAAGAGGCGTACTGCCTGAA	Confirm integration into *amyE* or insertion into pDR111
HM261	ATCGATGTTG CCATTCTGCC GAAAATGGTT	Amplify 1 kb upstream of *yqjW* to construct allelic exchange cassette
HM262	TAATATGAGATAATGCCGACTGTACACGACC ATTTAAAAAGGG	Amplify 1 kb upstream of *yqjW* to construct allelic exchange cassette, homology to CAT resistance gene
HM263	TAATATGAGATAATGCCGACTGTACACGACC ATTTAAAAAGGG	Amplify 1 kb downstream of *yqjW* to construct allelic exchange cassette
HM264	ATTTCCCCCTCATATGTTTTTGACAAATAG	Amplify 1 kb downstream of *yqjW* to construct allelic exchange cassette, homology to CAT resistance gene
HM265	GCTCAGTTTTTGAGCGATATAGG	Amplify 1 kb upstream of *yqjH* to construct allelic exchange cassette
HM266	CGCACAGATGCGTAAGGAGAAACAAATCTCC TTTTCGGTCA	Amplify 1 kb upstream of *yqjH* to construct allelic exchange cassette, homology to CAT resistance gene
HM267	TAATATGAGATAATGCCGACTGTACATCGCTT GAAAAAAAGGG	Amplify 1 kb downstream of *yqjH* to construct allelic exchange cassette
HM268	TACACGCTCTTCCTTCATAGCTGAGATAAA	Amplify 1 kb downstream of *yqjH* to construct allelic exchange cassette, homology to CAT resistance gene
HM382	AAAAGTCGACAAGGAGGTATACATATGATGC CTCGAACAATCATCG	Clone *leuC* into pDR111, head-on, SphI site

HM394	AAAAGCTAGCTCCTTATGAATGGGACTTACAC AACTGTTTTTTCCTG	Use with HM382, adds transcription-terminator to 3' end of gene, NheI site
HM384	AAAAAAGCTTAAGGAGGTATACATttgCCTATT AATATACCAACACACC	Clone metB into pDR111, head-on, HindII site
HM385	AAAAGTCGACTTAGTCCCATTCATAAGG	Use with HM384, SalI site
HM386	AAAAGTCGACAAGGAGGTATACATTTGCGTA TCAAAGAACATTTAAAAC	Clone *hisC* into pDR111, head-on, SalI site
HM396	AAAGCATGCTCCTTATGAATGGGACTTATAAA ATTTCAGCTAAAATGG	Use with HM386, adds transcription-terminator to 3' end of gene, SphI site
HM388	AAAGAATTCTTACACAACTGTTTTTTCC	Flip *leuC* and Pspank co-directionally in pDR111, EcoRi site
HM389	AAAGCATGCTGGCAAGAACGTTGCTCGAGG	Use with HM388, SphI site
HM392	AAAGAATTCTAACTCACATTAATTGCGTTGC	Flip *hisC* and Pspank (and *lacI*) co-directionally in pDR111, EcoRI site
HM393	AAAGGATCCTGGCAAGAACGTTGCTCGAGG	Use with HM392, BamHI site
HM444	AAAAGCTAGCAAGGAGGTATACATATGATGC CTCGAACAATCATCG	Clone *leuC* into pCAL838, head-on, flip with HM388 & HM389, NheI site
HM445	AAAAGCTAGCTCCTTATGAATGGGACTTACAC AACTGTTTTTTCCTG	Use with HM445, SphI site
HM657	TGAGCGGATAACAATTAAAGGAGGTATACAT ttgCCTATTAATATACCAACACACC	Flip *metB* co-directional in pDR111 via gibson assembly
HM658	AGCTTGCATGCGGTTAGTCCCATTCATAAGG	Used with HM657
HM706	GCTGTTTTAAATGTTCTTTGA	sequencing primer for promoters of *hisC* reporters
HM792	5’AATGATACGGCGACCACCGAGATCTACA ACACTCTTTCCCTACACGACGCTCTTCCGATCT NNNNTTAAACGTCCTGCTGATGAAC	PCR amplification primer fwd for deep sequencing *hisC952*
HM793a)	CAAGCAGAAGACGGCATACGAGACTGTAGG CATCGTAACCAGATGAAGCACTCTTTCCACTA TCCCTACAGTGT	PCR amplification primer rev for deep sequencing *hisC952* with custom index for head-on OFF
HM793b)	CAAGCAGAAGACGGCATACGAGAATCACACT GTCGTAACCAGATGAAGCACTCTTTCCACTAT CCCTACAGTGT	PCR amplification primer rev for deep sequencing *hisC952* with custom index for head-on ON

HM793c)	CAAGCAGAAGACGGCATACGAGAAGGAACG TTTCGTAACCAGATGAAGCACTCTTTCCACTA TCCCTACAGTGT	PCR amplification primer rev for deep sequencing *hisC952* with custom index for co-directional OFF
HM793d)	CAAGCAGAAGACGGCATACGAGAATTGTGCG GTCGTAACCAGATGAAGCACTCTTTCCACTAT CCCTACAGTGT	PCR amplification primer rev for deep sequencing *hisC952* with custom index for co-directional ON
HM793e)	CAAGCAGAAGACGGCATACGAGATATTGAGC CTCGTAACCAGATGAAGCACTCTTTCCACTAT CCCTACAGTGT	PCR amplification primer rev for deep sequencing *hisC952* with custom index for head-on Δ*yqjH* OFF
HM793f)	CAAGCAGAAGACGGCATACGAGACCAGTCGT ATCGTAACCAGATGAAGCACTCTTTCCACTAT CCCTACAGTGT	PCR amplification primer rev for deep sequencing *hisC952* with custom index for head-on ΔyqjH ON
HM793g)	CAAGCAGAAGACGGCATACGAGATACAAGCT CTCGTAACCAGATGAAGCACTCTTTCCACTAT CCCTACAGTGT	PCR amplification primer rev for deep sequencing *hisC952* with custom index for co-directional ΔyqjH OFF
HM793h)	CAAGCAGAAGACGGCATACGAGACCTCTAAC GTCGTAACCAGATGAAGCACTCTTTCCACTAT CCCTACAGTGT	PCR amplification primer rev for deep sequencing *hisC952* with custom index for co-directional ΔyqjH ON
HM794	ACACTCTTTCCCTACACGACGCTCTTCCGATCT	seq 1 primer for deep sequencing *hisC952*
HM795	CACTGTAGGGATAGTGGAAAGAGTGCTTCAT CTGGTTACGA	index primer for deep sequencing *hisC952*
HM796	TGC GTT TCC TGA CCG GAC GAT ATA GCC TTT TTC CAG C	seq 2 primer for deep sequencing *hisC952*
HM805	AAAA GCGGCCGC ATCGAATTCGATAACCCTAAAGTTATG	PCR amplify plasmid backbone to generate co-directional Pxis-*lacZ* reporter
HM806	AAAA GCATGC TAATAACCGGGCAGGCCAT	Use with HM805
HM807	AAAA GCGGCCGC TTATTTTTGACACCAGACCAA	PCR amplify *lacZ* to generate co-directional Pxis-*lacZ* reporter
HM808	AAAA GCATGC AAGCTTCCGATATTAAGTTTCTCTG	Use with HM807
HM814	GAGCCGAGATTATCGCAAT	Sequencing primer to confirm proper integration of co-directional Pxis-*lacZ* reporter

HM815	CAAACACTCCTAAAAGGAGAATTTAG	Sequencing primer to confirm proper integration of co-directional Pxis-*lacZ* reporter
HM835	AGCACCTGTTTTGCCGGCAGGTG	Sequencing primer to confirm *metB* reporter promoter
HM836	TCACTGCTTCAATCGGCTTTCC	Sequencing primer to confirm *hisC* reporter promoter
HM917	GTACGTACGATCTTTCAGCCGAC	Sequencing primer for Pspank (hy) in pDR111 & pCAL838

Supplementary table 8.4: qPCR Primers

Primer Number	Sequence	Amplification Target
HM188	GGCTTTCGCTACCTGGAGAG	*lacZ*
HM189	GACGAAGCCGCCCTGTAAAC	*lacZ*
HM192	CCGTCTGACCCGATCTTTTA	*yhaX*
HM193	GTCATGCTGAATGTCGTGCT	*yhaX*
HM489	AGTCGGATACGGAATTGCTG	*hisC*, middle of gene
HM490	CCCAGACGGCTGGTATTAAA	*hisC*, middle of gene
HM495	ACGATCGGAACAAAAGAGCA	*hisC*, 3' end of gene
HM496	TTTCCACCGAATTAGCTTGC	*hisC*, 3' end of gene
HM770	TCTCCAGCTGTGATAAACGGTA	*dnaK*
HM771	AAAACGGCATTGATTTGTCA	*dnaK*
HM842	TTTGACGGCATGATTATCACC	*metB,* middle of gene
HM843	TGGTAATACAAACCTGCTTGAGC	*metB,* middle of gene
HM892	CATGAAAAAGCTCGGCAAAG	*rplGB*
HM893	CATGAAAAAGCTCGGCAAAG	*rplGB*
HM952	GGT GTA AAC GAA CGT CAA TTC CGC AC	*rpsD*
HM953	GAAGAAAAA GTGAATGAGC TGCTGAAGGA A	*rpsD*
HM974	AGC TTTGTTATTGCTGG AATGTCG	*guaC*
HM975	TACATACTTTTCCCG GTCCGATGC	*guaC*
HM976	CGA AGG ACA TAC ATT TCC CTG CAG C	*Intergenic region 10 kb downstream from thrC*
HM977	GCT TAC GAA GGA ACG ACC AGA GC	*Intergenic region 10 kb downstream from thrC*

References

1. Merrikh H, Zhang Y, Grossman AD, Wang JD (2012) Replication-transcription conflicts in bacteria. *Nat Rev Microbiol* 10(7):449–458.

2. Rudolph CJ, Dhillon P, Moore T, Lloyd RG (2007) Avoiding and resolving conflicts between DNA replication and transcription. *DNA Repair (Amst)* 6(7):981–993.

3. Kim N, Jinks-Robertson S (2012) Transcription as a source of genome instability. *Nat Rev Genet* 13(3):204–214.

4. Lin Y-L, Pasero P (2012) Interference between DNA replication and transcription as a cause of genomic instability. *Curr Genomics* 13(1):65–73.

5. Helmrich A, Ballarino M, Nudler E, Tora L (2013) Transcription-replication encounters, consequences and genomic instability. *Nat Struct Mol Biol* 20(4):412–418.

6. McGlynn P, Savery NJ, Dillingham MS (2012) The conflict between DNA replication and transcription. *Mol Microbiol* 85(1):12–20.

7. Soultanas P (2011) The replication-transcription conflict. *Transcription* 2(3):140–144.

8. Merrikh H, Machón C, Grainger WH, Grossman AD, Soultanas P (2011) Co-directional replication-transcription conflicts lead to replication restart. *Nature* 470(7335):554–557.

9. Mirkin EV, Mirkin SM (2005) Mechanisms of transcription-replication collisions in bacteria. *Mol Cell Biol* 25(3):888–895.

10. Boubakri H, de Septenville AL, Viguera E, Michel B (2010) The helicases DinG, Rep and UvrD cooperate to promote replication across transcription units in vivo. *EMBO J* 29(1):145–157.

11. Tehranchi AK, et al. (2010) The transcription factor DksA prevents conflicts between DNA replication and transcription machinery. *Cell* 141(4):595–605.

12. Trautinger BW, Jaktaji RP, Rusakova E, Lloyd RG (2005) RNA polymerase modulators and DNA repair activities resolve conflicts between DNA replication and transcription. *Mol Cell* 19(2):247–258.

13. Santos-Pereira JM, García-Rubio ML, González-Aguilera C, Luna R, Aguilera A (2014) A genome-wide function of THSC/TREX-2 at active genes prevents transcription-replication collisions. *Nucleic Acids Res* 42(19):12000–12014.

14. Tuduri S, Crabbé L, Tourrière H, Coquelle A, Pasero P (2010) Does interference between replication and transcription contribute to genomic instability in cancer cells? *Cell Cycle* 9(10):1886–1892.

15. Helmrich A, Ballarino M, Tora L (2011) Collisions between replication and transcription complexes cause common fragile site instability at the longest human genes. *Mol Cell* 44(6):966–977.

16. Rocha EPC (2008) The organization of the bacterial genome. *Annu Rev Genet* 42(1):211–233.

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

18. Million-Weaver S, et al. (2015) An underlying mechanism for the increased mutagenesis of lagging-strand genes in Bacillus subtilis. *Proc Natl Acad Sci USA* 112(10):E1096–105.

19. Kim N, Abdulovic AL, Gealy R, Lippert MJ, Jinks-Robertson S (2007) Transcription-associated mutagenesis in yeast is directly proportional to the level of gene expression and influenced by the direction of DNA replication. *DNA Repair (Amst)* 6(9):1285–1296.

20. O'Donnell M (2006) Replisome architecture and dynamics in Escherichia coli. *J Biol Chem* 281(16):10653–10656.

21. Reyes-Lamothe R, Sherratt DJ, Leake MC (2010) Stoichiometry and architecture of active DNA replication machinery in Escherichia coli. *Science* 328(5977):498–501.

22. Duderstadt KE, Reyes-Lamothe R, van Oijen AM, Sherratt DJ (2014) Replication-fork dynamics. *Cold Spring Harb Perspect Biol* 6(1):a010157–a010157.

23. Sanders GM, Dallmann HG, McHenry CS (2010) Reconstitution of the B. subtilis replisome with 13 proteins including two distinct replicases. *Mol Cell* 37(2):273–281.

24. Cooper S, Helmstetter CE (1968) Chromosome replication and the division cycle of Escherichia coli B/r. *J Mol Biol* 31(3):519–540.

25. Kaguni JM, Kornberg A (1984) Replication initiated at the origin (oriC) of the E. coli chromosome reconstituted with purified enzymes. *Cell* 38(1):183–190.

26. Nüsslein V, Henke S, Johnston LH (1976) Replication of E. coli duplex DNA in vitro. The separation of the DNA containing fractions of a lysate from the soluble enzymes and their complementation properties. *Mol Gen Genet* 145(2):183–190.

27. Ogawa T, Okazaki T (1980) Discontinuous DNA replication. *Annu Rev Biochem* 49(1):421–457.

28. Wu CA, Zechner EL, Marians KJ (1992) Coordinated leading- and lagging-strand synthesis at the Escherichia coli DNA replication fork. I. Multiple effectors act to modulate Okazaki fragment size. *J Biol Chem* 267(6):4030–4044.

29. Okazaki R, Okazaki T, Sakabe K, Sugimoto K, Sugino A (1968) Mechanism of DNA chain growth. I. Possible discontinuity and unusual secondary structure of newly synthesized chains. *Proc Natl Acad Sci USA* 59(2):598–605.

30. McHenry CS (2011) Bacterial replicases and related polymerases. *Curr Opin Chem Biol* 15(5):587–594.

31. Timinskas K, Balvočiūtė M, Timinskas A, Venclovas Č (2014) Comprehensive analysis of DNA polymerase III α subunits and their homologs in bacterial genomes. *Nucleic Acids Res* 42(3):1393–1413.

32. Barnes MH, Miller SD, Brown NC (2002) DNA polymerases of low-GC gram-positive eubacteria: identification of the replication-specific enzyme encoded by dnaE. *J Bacteriol* 184(14):3834–3838.

33. Dervyn E, et al. (2001) Two essential DNA polymerases at the bacterial replication fork. *Science* 294(5547):1716–1719.

34. Rannou O, et al. (2013) Functional interplay of DnaE polymerase, DnaG primase and DnaC helicase within a ternary complex, and primase to polymerase hand-off during lagging strand

DNA replication in Bacillus subtilis. *Nucleic Acids Res* 41(10):5303–5320.

35. Park AY, et al. (2010) A single subunit directs the assembly of the Escherichia coli DNA sliding clamp loader. *Structure* 18(3):285–292.

36. Tsuchihashi Z, Kornberg A (1990) Translational frameshifting generates the gamma subunit of DNA polymerase III holoenzyme. *Proc Natl Acad Sci USA* 87(7):2516–2520.

37. Sakamoto Y, Nakai S, Moriya S, Yoshikawa H, Ogasawara N (1995) The Bacillus subtilis dnaC gene encodes a protein homologous to the DnaB helicase of Escherichia coli. *Microbiology (Reading, Engl)* 141 (Pt 3):641–644.

38. Biswas EE, Biswas SB (1999) Mechanism of DnaB helicase of Escherichia coli: structural domains involved in ATP hydrolysis, DNA binding, and oligomerization. *Biochemistry* 38(34):10919–10928.

39. Biswas SB, Chen PH, Biswas EE (1994) Structure and function of Escherichia coli DnaB protein: role of the N-terminal domain in helicase activity. *Biochemistry* 33(37):11307–11314.

40. Gotta SL, Miller OL, French SL (1991) rRNA transcription rate in Escherichia coli. *J Bacteriol* 173(20):6647–6649.

41. Vogel U, Jensen KF (1994) The RNA chain elongation rate in Escherichia coli depends on the growth rate. *J Bacteriol* 176(10):2807–2813.

42. Arndt KM, Kane CM (2003) Running with RNA polymerase: eukaryotic transcript elongation. *Trends Genet* 19(10):543–550.

43. Srivatsan A, Tehranchi A, MacAlpine DM, Wang JD (2010) Co-orientation of replication and transcription preserves genome integrity. *PLoS Genet* 6(1):e1000810.

44. De Septenville AL, Duigou S, Boubakri H, Michel B (2012) Replication fork reversal after replication-transcription collision. *PLoS Genet* 8(4):e1002622.

45. Prado F, Aguilera A (2005) Impairment of replication fork progression mediates RNA polII transcription-associated recombination. *EMBO J* 24(6):1267–1276.

46. Liu B, Alberts BM (1995) Head-on collision between a DNA replication apparatus and RNA polymerase transcription complex. *Science* 267(5201):1131–1137.

47. Elías-Arnanz M, Salas M (1999) Resolution of head-on collisions between the transcription machinery and bacteriophage phi29 DNA polymerase is dependent on RNA polymerase translocation. *EMBO J* 18(20):5675–5682.

48. Bedinger P, Hochstrasser M, Jongeneel CV, Alberts BM (1983) Properties of the T4 bacteriophage DNA replication apparatus: the T4 dda DNA helicase is required to pass a bound RNA polymerase molecule. *Cell* 34(1):115–123.

49. Liu B, Wong ML, Tinker RL, Geiduschek EP, Alberts BM (1993) The DNA replication fork can pass RNA polymerase without displacing the nascent transcript. *Nature* 366(6450):33–39.

50. Liu B, Wong ML, Alberts B (1994) A transcribing RNA polymerase molecule survives DNA replication without aborting its growing RNA chain. *Proc Natl Acad Sci USA* 91(22):10660–10664.

51. Elías-Arnanz M, Salas M (1997) Bacteriophage phi29 DNA replication arrest caused by codirectional collisions with the transcription machinery. *EMBO J* 16(18):5775–5783.

52. Pomerantz RT, O'Donnell M (2010) Direct restart of a replication fork stalled by a head-on RNA polymerase. *Science* 327(5965):590–592.

53. Dutta D, Shatalin K, Epshtein V, Gottesman ME, Nudler E (2011) Linking RNA polymerase backtracking to genome instability in E. coli. *Cell* 146(4):533–543.

54. Geertsema HJ, van Oijen AM (2013) A single-molecule view of DNA replication: the dynamic nature of multi-protein complexes revealed. *Curr Opin Struct Biol* 23(5):788–793.

55. Ellwood M, Nomura M (1982) Chromosomal locations of the genes for rRNA in Escherichia coli K-12. *J Bacteriol* 149(2):458–468.

56. Brewer BJ (1988) When polymerases collide: replication and the transcriptional organization of the E. coli chromosome. *Cell* 53(5):679–686.

57. Nicolas P, et al. (2012) Condition-dependent transcriptome reveals high-level regulatory architecture in Bacillus subtilis. *Science* 335(6072):1103–1106.

58. Blom E-J, Ridder ANJA, Lulko AT, Roerdink JBTM, Kuipers OP (2011) Time-resolved transcriptomics and bioinformatic analyses reveal intrinsic stress responses during batch culture of Bacillus subtilis. *PLoS ONE* 6(11):e27160.

59. Rocha EP, Danchin A, Viari A (1999) Universal replication biases in bacteria. *Mol Microbiol* 32(1):11–16.

60. French S (1992) Consequences of replication fork movement through transcription units in vivo. *Science* 258(5086):1362–1365.

61. Noller HF, et al. (1995) Structure and function of ribosomal RNA. *Biochem Cell Biol* 73(11-12):997–1009.

62. Torres M, Condon C, Balada JM, Squires C, Squires CL (2001) Ribosomal protein S4 is a transcription factor with properties remarkably similar to NusA, a protein involved in both non-ribosomal and ribosomal RNA antitermination. *EMBO J* 20(14):3811–3820.

63. Torres M, Balada J-M, Zellars M, Squires C, Squires CL (2004) In vivo effect of NusB and NusG on rRNA transcription antitermination. *J Bacteriol* 186(5):1304–1310.

64. Zellars M, Squires CL (1999) Antiterminator-dependent modulation of transcription elongation rates by NusB and NusG. *Mol Microbiol* 32(6):1296–1304.

65. Gupta MK, et al. (2013) Protein-DNA complexes are the primary sources of replication fork pausing in Escherichia coli. *Proc Natl Acad Sci USA* 110(18):7252–7257.

66. Rocha EPC (2004) The replication-related organization of bacterial genomes. *Microbiology (Reading, Engl)* 150(Pt 6):1609–1627.

67. Rocha EPC (2004) Order and disorder in bacterial genomes. *Curr Opin Microbiol* 7(5):519–527.

68. Rocha EPC, Danchin A (2003) Essentiality, not expressiveness, drives gene-strand bias in bacteria. *Nat Genet* 34(4):377–378.

69. Rocha EPC, Danchin A (2003) Gene essentiality determines chromosome organisation in bacteria. *Nucleic Acids Res* 31(22):6570–6577.

70. Aguilera A, García-Muse T (2013) Causes of genome instability. *Annu Rev Genet* 47(1):1–32.

71. Price MN, Alm EJ, Arkin AP (2005) Interruptions in gene expression drive highly expressed operons to the leading strand of DNA replication. *Nucleic Acids Res* 33(10):3224–3234.

72. Brüning J-G, Howard JL, McGlynn P (2014) Accessory replicative helicases and the replication of protein-bound DNA. *J Mol Biol* 426(24):3917–3928.

73. Marians KJ (2004) Mechanisms of replication fork restart in Escherichia coli. *Philos Trans R Soc Lond, B, Biol Sci* 359(1441):71–77.

74. Polard P, et al. (2002) Restart of DNA replication in Gram-positive bacteria: functional characterisation of the Bacillus subtilis PriA initiator. *Nucleic Acids Res* 30(7):1593–1605.

75. Heller RC, Marians KJ (2005) The disposition of nascent strands at stalled replication forks dictates the pathway of replisome loading during restart. *Mol Cell* 17(5):733–743.

76. McGlynn P, Lloyd RG (2002) Recombinational repair and restart of damaged replication forks. *Nat Rev Mol Cell Biol* 3(11):859–870.

77. Yeeles JTP, Poli J, Marians KJ, Pasero P (2013) Rescuing stalled or damaged replication forks. *Cold Spring Harb Perspect Biol* 5(5):a012815–a012815.

78. Michel B, Boubakri H, Baharoglu Z, LeMasson M, Lestini R (2007) Recombination proteins and rescue of arrested replication forks. *DNA Repair (Amst)* 6(7):967–980.

79. Michel B, Grompone G, Florès M-J, Bidnenko V (2004) Multiple pathways process stalled replication forks. *Proc Natl Acad Sci USA* 101(35):12783–12788.

80. Liu J, Xu L, Sandler SJ, Marians KJ (1999) Replication fork assembly at recombination intermediates is required for bacterial growth. *Proc Natl Acad Sci USA* 96(7):3552–3555.

81. Kreuzer KN (2005) Interplay between DNA replication and recombination in prokaryotes. *Annu Rev Microbiol* 59(1):43–67.

82. Cox MM (2001) Historical overview: searching for replication help in all of the rec places. *Proc Natl Acad Sci USA* 98(15):8173–8180.

83. Lovett ST (2006) Replication arrest-stimulated recombination: Dependence on the RecA paralog, RadA/Sms and translesion polymerase, DinB. *DNA Repair (Amst)* 5(12):1421–1427.

84. Lestini R, Michel B (2007) UvrD controls the access of recombination proteins to blocked replication forks. *EMBO J* 26(16):3804–3814.

85. Fijalkowska IJ, Jonczyk P, Tkaczyk MM, Bialoskorska M, Schaaper RM (1998) Unequal fidelity of leading strand and lagging strand DNA replication on the Escherichia coli chromosome. *Proc Natl Acad Sci USA* 95(17):10020–10025.

86. Szczepanik D, et al. (2001) Evolution rates of genes on leading and lagging DNA strands. *J Mol Evol* 52(5):426–433.

87. Rocha E (2002) Is there a role for replication fork asymmetry in the distribution of genes in

bacterial genomes? *Trends Microbiol* 10(9):393–395.

88. Francino MP, Chao L, Riley MA, Ochman H (1996) Asymmetries generated by transcription-coupled repair in enterobacterial genes. *Science* 272(5258):107–109.

89. Maliszewska-Tkaczyk M, Jonczyk P, Bialoskorska M, Schaaper RM, Fijalkowska IJ (2000) SOS mutator activity: unequal mutagenesis on leading and lagging strands. *Proc Natl Acad Sci USA* 97(23):12678–12683.

90. Juurik T, et al. (2012) Mutation frequency and spectrum of mutations vary at different chromosomal positions of Pseudomonas putida. *PLoS ONE* 7(10):e48511.

91. Sniegowski PD, Murphy HA (2006) Evolvability. *Curr Biol* 16(19):R831–4.

92. Sniegowski PD, Gerrish PJ (2010) Beneficial mutations and the dynamics of adaptation in asexual populations. *Philos Trans R Soc Lond, B, Biol Sci* 365(1544):1255–1263.

93. Giraud A, et al. (2001) Costs and benefits of high mutation rates: adaptive evolution of bacteria in the mouse gut. *Science* 291(5513):2606–2608.

94. Rainey PB (1999) Evolutionary genetics: The economics of mutation. *Curr Biol* 9(10):R371–3.

95. Ferenci T (2008) The spread of a beneficial mutation in experimental bacterial populations: the influence of the environment and genotype on the fixation of rpoS mutations. *Heredity (Edinb)* 100(5):446–452.

96. Eyre-Walker A, Keightley PD (2007) The distribution of fitness effects of new mutations. *Nat Rev Genet* 8(8):610–618.

97. Drake JW, Charlesworth B, Charlesworth D, Crow JF (1998) Rates of spontaneous mutation. *Genetics* 148(4):1667–1686.

98. Denamur E, Matic I (2006) Evolution of mutation rates in bacteria. *Mol Microbiol* 60(4):820–827.

99. Janion C (2008) Inducible SOS response system of DNA repair and mutagenesis in Escherichia coli. *Int J Biol Sci* 4(6):338–344.

100. Erill I, Campoy S, Barbé J (2007) Aeons of distress: an evolutionary perspective on the bacterial SOS response. *FEMS Microbiol Rev* 31(6):637–656.

101. Galhardo RS, Hastings PJ, Rosenberg SM (2007) Mutation as a stress response and the regulation of evolvability. *Crit Rev Biochem Mol Biol* 42(5):399–435.

102. McKenzie GJ, Lee PL, Lombardo MJ, Hastings PJ, Rosenberg SM (2001) SOS mutator DNA polymerase IV functions in adaptive mutation and not adaptive amplification. *Mol Cell* 7(3):571–579.

103. Wagner J, Nohmi T (2000) Escherichia coli DNA polymerase IV mutator activity: genetic requirements and mutational specificity. *J Bacteriol* 182(16):4587–4595.

104. Corzett CH, Goodman MF, Finkel SE (2013) Competitive fitness during feast and famine: how SOS DNA polymerases influence physiology and evolution in Escherichia coli. *Genetics* 194(2):409–420.

105. Friedberg EC, Wagner R, Radman M (2002) Specialized DNA polymerases, cellular survival, and the genesis of mutations. *Science* 296(5573):1627–1630.

106. Kassen R, Bataillon T (2006) Distribution of fitness effects among beneficial mutations before selection in experimental populations of bacteria. *Nat Genet* 38(4):484–488.

107. Cox MM, et al. (2000) The importance of repairing stalled replication forks. *Nature* 404(6773):37–41.

108. Mirkin EV, Mirkin SM (2007) Replication fork stalling at natural impediments. *Microbiol Mol Biol Rev* 71(1):13–35.

109. Cox MM (1998) A broadening view of recombinational DNA repair in bacteria. *Genes Cells* 3(2):65–78.

110. Cox MM (1991) The RecA protein as a recombinational repair system. *Mol Microbiol* 5(6):1295–1299.

111. Morimatsu K, Kowalczykowski SC (2003) RecFOR proteins load RecA protein onto gapped DNA to accelerate DNA strand exchange: a universal step of recombinational repair. *Mol Cell* 11(5):1337–1347.

112. Arnold DA, Kowalczykowski SC (2000) Facilitated loading of RecA protein is essential to recombination by RecBCD enzyme. *J Biol Chem* 275(16):12261–12265.

113. Lenhart JS, et al. (2014) RecO and RecR are necessary for RecA loading in response to DNA damage and replication fork stress. *J Bacteriol* 196(15):2851–2860.

114. Alonso JC, Lüder G, Tailor RH (1991) Characterization of Bacillus subtilis recombinational pathways. *J Bacteriol* 173(13):3977–3980.

115. Fernández S, Kobayashi Y, Ogasawara N, Alonso JC (1999) Analysis of the Bacillus subtilis recO gene: RecO forms part of the RecFLOR function. *Mol Gen Genet* 261(3):567–573.

116. Manfredi C, Carrasco B, Ayora S, Alonso JC (2008) Bacillus subtilis RecO nucleates RecA onto SsbA-coated single-stranded DNA. *J Biol Chem* 283(36):24837–24847.

117. Yeeles JTP, Dillingham MS (2010) The processing of double-stranded DNA breaks for recombinational repair by helicase-nuclease complexes. *DNA Repair (Amst)* 9(3):276–285.

118. Dillingham MS, Kowalczykowski SC (2008) RecBCD enzyme and the repair of double-stranded DNA breaks. *Microbiol Mol Biol Rev* 72(4):642–71– Table of Contents.

119. Wigley DB (2013) Bacterial DNA repair: recent insights into the mechanism of RecBCD, AddAB and AdnAB. *Nat Rev Microbiol* 11(1):9–13.

120. Churchill JJ, Kowalczykowski SC (2000) Identification of the RecA protein-loading domain of RecBCD enzyme. *J Mol Biol* 297(3):537–542.

121. Kuzminov A (1999) Recombinational repair of DNA damage in Escherichia coli and bacteriophage lambda. *Microbiol Mol Biol Rev* 63(4):751–813– table of contents.

122. Courcelle J, Hanawalt PC (2003) RecA-dependent recovery of arrested DNA replication forks. *Annu Rev Genet* 37(1):611–646.

123. Grompone G, Sanchez N, Dusko Ehrlich S, Michel B (2004) Requirement for RecFOR-mediated recombination in priA mutant. *Mol Microbiol* 52(2):551–562.

124. Iwasaki H, Takahagi M, Shiba T, Nakata A, Shinagawa H (1991) Escherichia coli RuvC protein is an endonuclease that resolves the Holliday structure. *EMBO J* 10(13):4381–4389.

125. Dunderdale HJ, et al. (1991) Formation and resolution of recombination intermediates by E. coli RecA and RuvC proteins. *Nature* 354(6354):506–510.

126. Sanchez H, et al. (2005) The RuvAB branch migration translocase and RecU Holliday junction resolvase are required for double-stranded DNA break repair in Bacillus subtilis. *Genetics* 171(3):873–883.

127. Carrasco B, Cozar MC, Lurz R, Alonso JC, Ayora S (2004) Genetic recombination in Bacillus subtilis 168: contribution of Holliday junction processing functions in chromosome segregation. *J Bacteriol* 186(17):5557–5566.

128. Ayora S, et al. (2004) Bacillus subtilis RecU protein cleaves Holliday junctions and anneals single-stranded DNA. *Proc Natl Acad Sci USA* 101(2):452–457.

129. Smits WK, Goranov AI, Grossman AD (2010) Ordered association of helicase loader proteins with the Bacillus subtilis origin of replication in vivo. *Mol Microbiol* 75(2):452–461.

130. Bruand C, Farache M, McGovern S, Ehrlich SD, Polard P (2001) DnaB, DnaD and DnaI proteins are components of the Bacillus subtilis replication restart primosome. *Mol Microbiol* 42(1):245–255.

131. McGlynn P, Al-Deib AA, Liu J, Marians KJ, Lloyd RG (1997) The DNA replication protein PriA and the recombination protein RecG bind D-loops. *J Mol Biol* 270(2):212–221.

132. Gregg AV, McGlynn P, Jaktaji RP, Lloyd RG (2002) Direct rescue of stalled DNA replication forks via the combined action of PriA and RecG helicase activities. *Mol Cell* 9(2):241–251.

133. Xu L, Marians KJ (2003) PriA mediates DNA replication pathway choice at recombination intermediates. *Mol Cell* 11(3):817–826.

134. Auchtung JM, Lee CA, Garrison KL, Grossman AD (2007) Identification and characterization of the immunity repressor (ImmR) that controls the mobile genetic element ICEBs1 of Bacillus subtilis. *Mol Microbiol* 64(6):1515–1528.

135. Simmons LA, Grossman AD, Walker GC (2007) Replication is required for the RecA localization response to DNA damage in Bacillus subtilis. *Proc Natl Acad Sci USA* 104(4):1360–1365.

136. Kuzminov A (2001) Single-strand interruptions in replicating chromosomes cause double-strand breaks. *Proc Natl Acad Sci USA* 98(15):8241–8246.

137. Eggleston AK, Kowalczykowski SC (1991) An overview of homologous pairing and DNA strand exchange proteins. *Biochimie* 73(2-3):163–176.

138. Lusetti SL, Cox MM (2002) The bacterial RecA protein and the recombinational DNA repair of stalled replication forks. *Annu Rev Biochem* 71(1):71–100.

139. Kowalczykowski SC, Krupp RA (1987) Effects of Escherichia coli SSB protein on the single-stranded DNA-dependent ATPase activity of Escherichia coli RecA protein. Evidence that SSB protein facilitates the binding of RecA protein to regions of secondary structure within single-

stranded DNA. *J Mol Biol* 193(1):97–113.

140. Cox MM (2007) Regulation of bacterial RecA protein function. *Crit Rev Biochem Mol Biol* 42(1):41–63.

141. Ryzhikov M, Gupta R, Glickman M, Korolev S (2014) RecO protein initiates DNA recombination and strand annealing through two alternative DNA binding mechanisms. *J Biol Chem* 289(42):28846–28855.

142. Sanchez H, Kidane D, Castillo Cozar M, Graumann PL, Alonso JC (2006) Recruitment of Bacillus subtilis RecN to DNA double-strand breaks in the absence of DNA end processing. *J Bacteriol* 188(2):353–360.

143. Bentchikou E, Servant P, Coste G, Sommer S (2010) A major role of the RecFOR pathway in DNA double-strand-break repair through ESDSA in Deinococcus radiodurans. *PLoS Genet* 6(1):e1000774.

144. Champoux JJ (2001) DNA topoisomerases: structure, function, and mechanism. *Annu Rev Biochem* 70(1):369–413.

145. Booker BM, Deng S, Higgins NP (2010) DNA topology of highly transcribed operons in Salmonella enterica serovar Typhimurium. *Mol Microbiol* 78(6):1348–1364.

146. Krasilnikov AS, Podtelezhnikov A, Vologodskii A, Mirkin SM (1999) Large-scale effects of transcriptional DNA supercoiling in vivo. *J Mol Biol* 292(5):1149–1160.

147. Aguilera A, García-Muse T (2012) R loops: from transcription byproducts to threats to genome stability. *Mol Cell* 46(2):115–124.

148. Grompone G, Bidnenko V, Ehrlich SD, Michel B (2004) PriA is essential for viability of the Escherichia coli topoisomerase IV parE10(Ts) mutant. *J Bacteriol* 186(4):1197–1199.

149. Carrasco B, Ayora S, Lurz R, Alonso JC (2005) Bacillus subtilis RecU Holliday-junction resolvase modulates RecA activities. *Nucleic Acids Res* 33(12):3942–3952.

150. Grompone G, Ehrlich D, Michel B (2004) Cells defective for replication restart undergo replication fork reversal. *EMBO Rep* 5(6):607–612.

151. Robu ME, Inman RB, Cox MM (2001) RecA protein promotes the regression of stalled replication forks in vitro. *Proc Natl Acad Sci USA* 98(15):8211–8218.

152. Zellweger R, et al. (2015) Rad51-mediated replication fork reversal is a global response to genotoxic treatments in human cells. *J Cell Biol* 208(5):563–579.

153. Veaute X, Fuchs RP (1993) Greater susceptibility to mutations in lagging strand of DNA replication in Escherichia coli than in leading strand. *Science* 261(5121):598–600.

154. Lin C-H, Lian C-Y, Hsiung CA, Chen F-C (2011) Changes in transcriptional orientation are associated with increases in evolutionary rates of enterobacterial genes. *BMC Bioinformatics* 12 Suppl 9(Suppl 9):S19.

155. Kobayashi K, et al. (2003) Essential Bacillus subtilis genes. *Proc Natl Acad Sci USA* 100(8):4678–4683.

156. Tuduri S, et al. (2009) Topoisomerase I suppresses genomic instability by preventing

interference between replication and transcription. *Nat Cell Biol* 11(11):1315–1324.

157. Mortusewicz O, Herr P, Helleday T (2013) Early replication fragile sites: where replication-transcription collisions cause genetic instability. *EMBO J* 32(4):493–495.

158. Goranov AI, Kuester-Schoeck E, Wang JD, Grossman AD (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.

159. Love PE, Yasbin RE (1984) Genetic characterization of the inducible SOS-like system of Bacillus subtilis. *J Bacteriol* 160(3):910–920.

160. Hersh MN, Ponder RG, Hastings PJ, Rosenberg SM (2004) Adaptive mutation and amplification in Escherichia coli: two pathways of genome adaptation under stress. *Res Microbiol* 155(5):352–359.

161. Bjedov I, et al. (2003) Stress-induced mutagenesis in bacteria. *Science* 300(5624):1404–1409.

162. Foster PL (2007) Stress-induced mutagenesis in bacteria. *Crit Rev Biochem Mol Biol* 42(5):373–397.

163. Hilbert L (2013) Stress-induced hypermutation as a physical property of life, a force of natural selection and its role in four thought experiments. *Phys Biol* 10(2):026001.

164. Maclean RC, Torres-Barceló C, Moxon R (2013) Evaluating evolutionary models of stress-induced mutagenesis in bacteria. *Nat Rev Genet* 14(3):221–227.

165. Aertsen A, Michiels CW (2005) Diversify or die: generation of diversity in response to stress. *Crit Rev Microbiol* 31(2):69–78.

166. Kibota TT, Lynch M (1996) Estimate of the genomic mutation rate deleterious to overall fitness in E. coli. *Nature* 381(6584):694–696.

167. Fernández L, Hancock REW (2012) Adaptive and mutational resistance: role of porins and efflux pumps in drug resistance. *Clin Microbiol Rev* 25(4):661–681.

168. Gordo I, Perfeito L, Sousa A (2011) Fitness effects of mutations in bacteria. *J Mol Microbiol Biotechnol* 21(1-2):20–35.

169. Ogi T, Kannouche P, Lehmann AR (2005) Localisation of human Y-family DNA polymerase kappa: relationship to PCNA foci. *J Cell Sci* 118(Pt 1):129–136.

170. Heltzel JMH, Maul RW, Wolff DW, Sutton MD (2012) Escherichia coli DNA polymerase IV (Pol IV), but not Pol II, dynamically switches with a stalled Pol III* replicase. *J Bacteriol* 194(14):3589–3600.

171. Kannouche P, et al. (2001) Domain structure, localization, and function of DNA polymerase eta, defective in xeroderma pigmentosum variant cells. *Genes Dev* 15(2):158–172.

172. Fuchs RP, Fujii S (2013) Translesion DNA synthesis and mutagenesis in prokaryotes. *Cold Spring Harb Perspect Biol* 5(12):a012682–a012682.

173. Sale JE, Lehmann AR, Woodgate R (2012) Y-family DNA polymerases and their role in tolerance of cellular DNA damage. *Nat Rev Mol Cell Biol* 13(3):141–152.

174. Beuning PJ, Simon SM, Godoy VG, Jarosz DF, Walker GC (2006) Characterization of Escherichia coli translesion synthesis polymerases and their accessory factors. *Meth Enzymol* 408:318–340.

175. Fuchs RP, Fujii S, Wagner J (2004) Properties and functions of Escherichia coli: Pol IV and Pol V. *Adv Protein Chem* 69:229–264.

176. Jarosz DF, Beuning PJ, Cohen SE, Walker GC (2007) Y-family DNA polymerases in Escherichia coli. *Trends Microbiol* 15(2):70–77.

177. Goodman MF (2002) Error-prone repair DNA polymerases in prokaryotes and eukaryotes. *Annu Rev Biochem* 71(1):17–50.

178. Shcherbakova PV, Fijalkowska IJ (2006) Translesion synthesis DNA polymerases and control of genome stability. *Front Biosci* 11:2496–2517.

179. Kokoska RJ, Bebenek K, Boudsocq F, Woodgate R, Kunkel TA (2002) Low fidelity DNA synthesis by a y family DNA polymerase due to misalignment in the active site. *J Biol Chem* 277(22):19633–19638.

180. Pata JD (2010) Structural diversity of the Y-family DNA polymerases. *Biochim Biophys Acta* 1804(5):1124–1135.

181. Kobayashi S, Valentine MR, Pham P, O'Donnell M, Goodman MF (2002) Fidelity of Escherichia coli DNA polymerase IV. Preferential generation of small deletion mutations by dNTP-stabilized misalignment. *J Biol Chem* 277(37):34198–34207.

182. Rosenberg SM (2001) Evolving responsively: adaptive mutation. *Nat Rev Genet* 2(7):504–515.

183. Cohen SE, et al. (2010) Roles for the transcription elongation factor NusA in both DNA repair and damage tolerance pathways in Escherichia coli. *Proc Natl Acad Sci USA* 107(35):15517–15522.

184. Shee C, Gibson JL, Rosenberg SM (2012) Two mechanisms produce mutation hotspots at DNA breaks in Escherichia coli. *Cell Rep* 2(4):714–721.

185. Duigou S, Ehrlich SD, Noirot P, Noirot-Gros M-F (2004) Distinctive genetic features exhibited by the Y-family DNA polymerases in Bacillus subtilis. *Mol Microbiol* 54(2):439–451.

186. Sung H-M, Yeamans G, Ross CA, Yasbin RE (2003) Roles of YqjH and YqjW, homologs of the Escherichia coli UmuC/DinB or Y superfamily of DNA polymerases, in stationary-phase mutagenesis and UV-induced mutagenesis of Bacillus subtilis. *J Bacteriol* 185(7):2153–2160.

187. Rivas-Castillo AM, Yasbin RE, Robleto E, Nicholson WL, Pedraza-Reyes M (2010) Role of the Y-family DNA polymerases YqjH and YqjW in protecting sporulating Bacillus subtilis cells from DNA damage. *Curr Microbiol* 60(4):263–267.

188. Duigou S, Ehrlich SD, Noirot P, Noirot-Gros M-F (2005) DNA polymerase I acts in translesion synthesis mediated by the Y-polymerases in Bacillus subtilis. *Mol Microbiol* 57(3):678–690.

189. Jena NR (2012) DNA damage by reactive species: Mechanisms, mutation and repair. *J Biosci* 37(3):503–517.

190. Walker GC (1984) Mutagenesis and inducible responses to deoxyribonucleic acid damage in Escherichia coli. *Microbiol Rev* 48(1):60–93.

191. Lehmann AR (2002) Replication of damaged DNA in mammalian cells: new solutions to an old problem. *Mutat Res* 509(1-2):23–34.

192. Bose B, Grossman AD (2011) Regulation of horizontal gene transfer in Bacillus subtilis by activation of a conserved site-specific protease. *J Bacteriol* 193(1):22–29.

193. Brown NC, Handschumacher RE (1966) Inhibition of the synthesis of deoxyribonucleic acid in bacteria by 6-(p-hydroxyphenylazo)-2,4-dihydroxypyrimidine. I. Metabolic studies in Streptococcus fecalis. *J Biol Chem* 241(13):3083–3089.

194. Renzette N, et al. (2005) Localization of RecA in Escherichia coli K-12 using RecA-GFP. *Mol Microbiol* 57(4):1074–1085.

195. Keijser BJF, et al. (2007) Analysis of temporal gene expression during Bacillus subtilis spore germination and outgrowth. *J Bacteriol* 189(9):3624–3634.

196. Pomerantz RT, Kurth I, Goodman MF, O'Donnell ME (2013) Preferential D-loop extension by a translesion DNA polymerase underlies error-prone recombination. *Nat Struct Mol Biol* 20(6):748–755.

197. Lenhart JS, Schroeder JW, Walsh BW, Simmons LA (2012) DNA repair and genome maintenance in Bacillus subtilis. *Microbiol Mol Biol Rev* 76(3):530–564.

198. Walker GC (1995) SOS-regulated proteins in translesion DNA synthesis and mutagenesis. *Trends Biochem Sci* 20(10):416–420.

199. Kisker C, Kuper J, Van Houten B (2013) Prokaryotic nucleotide excision repair. *Cold Spring Harb Perspect Biol* 5(3):a012591–a012591.

200. Cohen SE, Walker GC (2011) New discoveries linking transcription to DNA repair and damage tolerance pathways. *Transcription* 2(1):37–40.

201. Ganesan A, Spivak G, Hanawalt PC (2012) Transcription-coupled DNA repair in prokaryotes. *Prog Mol Biol Transl Sci* 110:25–40.

202. Hanawalt PC, Spivak G (2008) Transcription-coupled DNA repair: two decades of progress and surprises. *Nat Rev Mol Cell Biol* 9(12):958–970.

203. Epshtein V, et al. (2014) UvrD facilitates DNA repair by pulling RNA polymerase backwards. *Nature* 505(7483):372–377.

204. Wolff E, Kim M, Hu K, Yang H, Miller JH (2004) Polymerases leave fingerprints: analysis of the mutational spectrum in Escherichia coli rpoB to assess the role of polymerase IV in spontaneous mutation. *J Bacteriol* 186(9):2900–2905.

205. Brenlla A, Markiewicz RP, Rueda D, Romano LJ (2014) Nucleotide selection by the Y-family DNA polymerase Dpo4 involves template translocation and misalignment. *Nucleic Acids Res* 42(4):2555–2563.

206. Potapova O, Grindley NDF, Joyce CM (2002) The mutational specificity of the Dbh lesion bypass polymerase and its implications. *J Biol Chem* 277(31):28157–28166.

207. Ohashi E, et al. (2000) Fidelity and processivity of DNA synthesis by DNA polymerase kappa, the product of the human DINB1 gene. *J Biol Chem* 275(50):39678–39684.

208. Mukherjee P, Lahiri I, Pata JD (2013) Human polymerase kappa uses a template-slippage deletion mechanism, but can realign the slipped strands to favour base substitution mutations over deletions. *Nucleic Acids Res* 41(9):5024–5035.

209. Tissier A, McDonald JP, Frank EG, Woodgate R (2000) poliota, a remarkably error-prone human DNA polymerase. *Genes Dev* 14(13):1642–1650.

210. Zhang Y, et al. (2000) Human DNA polymerase kappa synthesizes DNA with extraordinarily low fidelity. *Nucleic Acids Res* 28(21):4147–4156.

211. Lee H, Popodi E, Tang H, Foster PL (2012) Rate and molecular spectrum of spontaneous mutations in the bacterium Escherichia coli as determined by whole-genome sequencing. *Proc Natl Acad Sci USA* 109(41):E2774–83.

212. Kennedy SR, Salk JJ, Schmitt MW, Loeb LA (2013) Ultra-sensitive sequencing reveals an age-related increase in somatic mitochondrial mutations that are inconsistent with oxidative damage. *PLoS Genet* 9(9):e1003794.

213. Hershberg R, Petrov DA (2010) Evidence that mutation is universally biased towards AT in bacteria. *PLoS Genet* 6(9):e1001115.

214. Raghavan R, Kelkar YD, Ochman H (2012) A selective force favoring increased G+C content in bacterial genes. *Proc Natl Acad Sci USA* 109(36):14504–14507.

215. la Loza de MCD, Wellinger RE, Aguilera A (2009) Stimulation of direct-repeat recombination by RNA polymerase III transcription. *DNA Repair (Amst)* 8(5):620–626.

216. Takeuchi Y, Horiuchi T, Kobayashi T (2003) Transcription-dependent recombination and the role of fork collision in yeast rDNA. *Genes Dev* 17(12):1497–1506.

217. Christman MF, Dietrich FS, Fink GR (1988) Mitotic recombination in the rDNA of S. cerevisiae is suppressed by the combined action of DNA topoisomerases I and II. *Cell* 55(3):413–425.

218. Reaban ME, Griffin JA (1990) Induction of RNA-stabilized DNA conformers by transcription of an immunoglobulin switch region. *Nature* 348(6299):342–344.

219. Kim N, Jinks-Robertson S (2009) dUTP incorporation into genomic DNA is linked to transcription in yeast. *Nature* 459(7250):1150–1153.

220. Gaillard H, Herrera-Moyano E, Aguilera A (2013) Transcription-associated genome instability. *Chem Rev* 113(11):8638–8661.

221. Robleto EA, Martin HA, Pedraza-Reyes M (2012) Mfd and transcriptional derepression cause genetic diversity in Bacillus subtilis. *Front Biosci (Elite Ed)* 4:1246–1254.

222. Grivell AR, Grivell MB, Hanawalt PC (1975) Turnover in bacterial DNA containing thymine or 5-bromouracil. *J Mol Biol* 98(1):219–233.

223. Kuipers GK, et al. (2000) The role of nucleotide excision repair of Escherichia coli in repair of spontaneous and gamma-radiation-induced DNA damage in the lacZalpha gene. *Mutat Res* 460(2):117–125.

224. Vaisman A, et al. (2013) Removal of misincorporated ribonucleotides from prokaryotic genomes: an unexpected role for nucleotide excision repair. *PLoS Genet* 9(11):e1003878.

225. Guo X, Jinks-Robertson S (2013) Removal of N-6-methyladenine by the nucleotide excision repair pathway triggers the repair of mismatches in yeast gap-repair intermediates. *DNA Repair (Amst)* 12(12):1053–1061.

226. Hoekstra MF, Malone RE (1986) Excision repair functions in Saccharomyces cerevisiae recognize and repair methylation of adenine by the Escherichia coli dam gene. *Mol Cell Biol* 6(10):3555–3558.

227. Kim N, Jinks-Robertson S (2010) Abasic sites in the transcribed strand of yeast DNA are removed by transcription-coupled nucleotide excision repair. *Mol Cell Biol* 30(13):3206–3215.

228. Melis JPM, van Steeg H, Luijten M (2013) Oxidative DNA damage and nucleotide excision repair. *Antioxid Redox Signal* 18(18):2409–2419.

229. Lin JJ, Sancar A (1989) A new mechanism for repairing oxidative damage to DNA: (A)BC excinuclease removes AP sites and thymine glycols from DNA. *Biochemistry* 28(20):7979–7984.

230. Kow YW, Wallace SS, Van Houten B (1990) UvrABC nuclease complex repairs thymine glycol, an oxidative DNA base damage. *Mutat Res* 235(2):147–156.

231. Nabben FJ, et al. (1984) Repair of damage in double-stranded phi X174 (RF) DNA due to radiation-induced water radicals. *Int J Radiat Biol Relat Stud Phys Chem Med* 45(4):379–388.

232. Branum ME, Reardon JT, Sancar A (2001) DNA repair excision nuclease attacks undamaged DNA. A potential source of spontaneous mutations. *J Biol Chem* 276(27):25421–25426.

233. Grossman L, Caron PR, Mazur SJ, Oh EY (1988) Repair of DNA-containing pyrimidine dimers. *FASEB J* 2(11):2696–2701.

234. Wimberly H, et al. (2013) R-loops and nicks initiate DNA breakage and genome instability in non-growing Escherichia coli. *Nat Commun* 4:2115.

235. Lehner K, Jinks-Robertson S (2014) Shared genetic pathways contribute to the tolerance of endogenous and low-dose exogenous DNA damage in yeast. *Genetics* 198(2):519–530.

236. Schreier WJ, Gilch P, Zinth W (2015) Early events of DNA photodamage. *Annu Rev Phys Chem* 66(1):497–519.

237. Cadet J, Grand A, Douki T (2015) Solar UV radiation-induced DNA Bipyrimidine photoproducts: formation and mechanistic insights. *Top Curr Chem* 356(Chapter 553):249–275.

238. Gossett AJ, Lieb JD (2008) DNA Immunoprecipitation (DIP) for the Determination of DNA-Binding Specificity. *CSH Protoc* 2008(4):pdb.prot4972–pdb.prot4972.

239. Richa, Sinha RP, Häder D-P (2015) Physiological aspects of UV-excitation of DNA. *Top Curr Chem* 356(Chapter 531):203–248.

240. Iwai S, et al. (1995) Reaction mechanism of T4 endonuclease V determined by analysis using modified oligonucleotide duplexes. *Biochemistry* 34(14):4601–4609.

241. Jorns MS (1990) DNA photorepair: chromophore composition and function in two classes of DNA photolyases. *Biofactors* 2(4):207–211.

242. Roberts SA, et al. (2012) Clustered mutations in yeast and in human cancers can arise from damaged long single-strand DNA regions. *Mol Cell* 46(4):424–435.

243. Chan K, et al. (2012) Base damage within single-strand DNA underlies in vivo hypermutability induced by a ubiquitous environmental agent. *PLoS Genet* 8(12):e1003149.

244. Yang W (2008) Structure and mechanism for DNA lesion recognition. *Cell Res* 18(1):184–197.

245. Lecointe F, et al. (2007) Anticipating chromosomal replication fork arrest: SSB targets repair DNA helicases to active forks. *EMBO J* 26(19):4239–4251.

246. Rothman-Denes LB (2013) Structure of Escherichia coli RNA polymerase holoenzyme at last. *Proc Natl Acad Sci USA* 110(49):19662–19663.

247. Fulcrand G, Zhi X, Leng F (2013) Transcription-coupled DNA supercoiling in defined protein systems and in E. coli topA mutant strains. *IUBMB Life* 65(7):615–622.

248. Sung W, et al. (2015) Asymmetric Context-dependent Mutation Patterns Revealed through Mutation-accumulation Experiments. *Mol Biol Evol*:msv055.

249. Kunkel TA (2004) DNA replication fidelity. *J Biol Chem* 279(17):16895–16898.

250. Law YK, Forties RA, Liu X, Poirier MG, Kohler B (2013) Sequence-dependent thymine dimer formation and photoreversal rates in double-stranded DNA. *Photochem Photobiol Sci* 12(8):1431–1439.

251. Schuch AP, Menck CFM (2010) The genotoxic effects of DNA lesions induced by artificial UV-radiation and sunlight. *J Photochem Photobiol B, Biol* 99(3):111–116.

252. Sinha RP, Häder D-P (2002) UV-induced DNA damage and repair: a review. *Photochem Photobiol Sci* 1(4):225–236.

253. Brooks PJ (2013) Blinded by the UV light: how the focus on transcription-coupled NER has distracted from understanding the mechanisms of Cockayne syndrome neurologic disease. *DNA Repair (Amst)* 12(8):656–671.

254. Jaruga P, Dizdaroglu M (2008) 8,5“-Cyclopurine-2-”deoxynucleosides in DNA: mechanisms of formation, measurement, repair and biological effects. *DNA Repair (Amst)* 7(9):1413–1425.

255. Pednekar V, Weerasooriya S, Jasti VP, Basu AK (2014) Mutagenicity and genotoxicity of (5'S)-8,5'-cyclo-2'-deoxyadenosine in Escherichia coli and replication of (5'S)-8,5“-cyclopurine-2-”deoxynucleosides in vitro by DNA polymerase IV, exo-free Klenow fragment, and Dpo4. *Chem Res Toxicol* 27(2):200–210.

256. Gopalakrishnan K, et al. (2010) Hydrogen peroxide induced genomic instability in nucleotide excision repair-deficient lymphoblastoid cells. *Genome Integr* 1(1):16.

257. Drake JW (1991) A constant rate of spontaneous mutation in DNA-based microbes. *Proc Natl Acad Sci USA* 88(16):7160–7164.

258. Kaplan RW (1972) Evolutionary adjustment of spontaneous mutation rates. *Humangenetik* 16(1):39–42.

259. Sung W, Ackerman MS, Miller SF, Doak TG, Lynch M (2012) Drift-barrier hypothesis and mutation-rate evolution. *Proc Natl Acad Sci USA* 109(45):18488–18492.

260. Chattopadhyay S, et al. (2009) High frequency of hotspot mutations in core genes of Escherichia coli due to short-term positive selection. *Proc Natl Acad Sci USA* 106(30):12412–

12417.

261. Crow JF (1986) Population consequences of mutagenesis and antimutagenesis. *Basic Life Sci* 39:519–530.

262. Jeffares DC, Tomiczek B, Sojo V, Reis dos M (2015) A beginners guide to estimating the non-synonymous to synonymous rate ratio of all protein-coding genes in a genome. *Methods Mol Biol* 1201(Chapter 4):65–90.

263. Firnberg E, Labonte JW, Gray JJ, Ostermeier M (2014) A comprehensive, high-resolution map of a gene's fitness landscape. *Mol Biol Evol* 31(6):1581–1592.

264. Bejerano-Sagie M, et al. (2006) A checkpoint protein that scans the chromosome for damage at the start of sporulation in Bacillus subtilis. *Cell* 125(4):679–690.

265. Kimura M (1979) The neutral theory of molecular evolution. *Sci Am* 241(5):98–100– 102– 108 passim.

266. Moxon R, Bayliss C, Hood D (2006) Bacterial contingency loci: the role of simple sequence DNA repeats in bacterial adaptation. *Annu Rev Genet* 40(1):307–333.

267. Graves CJ, Ros VID, Stevenson B, Sniegowski PD, Brisson D (2013) Natural selection promotes antigenic evolvability. *PLoS Pathog* 9(11):e1003766.

268. Krašovec R, et al. (2014) Mutation rate plasticity in rifampicin resistance depends on Escherichia coli cell-cell interactions. *Nat Commun* 5:3742.

269. Chopra I, O'Neill AJ, Miller K (2003) The role of mutators in the emergence of antibiotic-resistant bacteria. *Drug Resist Updat* 6(3):137–145.

270. Rolfe MD, et al. (2012) Lag phase is a distinct growth phase that prepares bacteria for exponential growth and involves transient metal accumulation. *J Bacteriol* 194(3):686–701.

271. Imhof M, Schlotterer C (2001) Fitness effects of advantageous mutations in evolving Escherichia coli populations. *Proc Natl Acad Sci USA* 98(3):1113–1117.

272. Johnson T, Barton NH (2002) The effect of deleterious alleles on adaptation in asexual populations. *Genetics* 162(1):395–411.

273. Denapoli J, Tehranchi AK, Wang JD (2013) Dose-dependent reduction of replication elongation rate by (p)ppGpp in Escherichia coli and Bacillus subtilis. *Mol Microbiol* 88(1):93–104.

274. Bryan DS, Ransom M, Adane B, York K, Hesselberth JR (2014) High resolution mapping of modified DNA nucleobases using excision repair enzymes. *Genome Res* 24(9):1534–1542.

275. Sidorenko J, Ukkivi K, Kivisaar M (2015) NER enzymes maintain genome integrity and suppress homologous recombination in the absence of exogenously induced DNA damage in Pseudomonas putida. *DNA Repair (Amst)* 25:15–26.

276. Wahba L, Amon JD, Koshland D, Vuica-Ross M (2011) RNase H and multiple RNA biogenesis factors cooperate to prevent RNA:DNA hybrids from generating genome instability. *Mol Cell* 44(6):978–988.

277. Barnes MH, LaMarr WA, Foster KA (2003) DNA gyrase and DNA topoisomerase of Bacillus subtilis: expression and characterization of recombinant enzymes encoded by the gyrA, gyrB

and parC, parE genes. *Protein Expr Purif* 29(2):259–264.

278. Orr E, Staudenbauer WL (1982) Bacillus subtilis DNA gyrase: purification of subunits and reconstitution of supercoiling activity. *J Bacteriol* 151(1):524–527.

279. Gupta S, Yeeles JTP, Marians KJ (2014) Regression of replication forks stalled by leading-strand template damage: I. Both RecG and RuvAB catalyze regression, but RuvC cleaves the holliday junctions formed by RecG preferentially. *J Biol Chem* 289(41):28376–28387.

280. Cebrián J, et al. (2015) Direct Evidence for the Formation of Precatenanes During DNA Replication. *J Biol Chem*:jbc.M115.642272.

281. Datta A, Jinks-Robertson S (1995) Association of increased spontaneous mutation rates with high levels of transcription in yeast. *Science* 268(5217):1616–1619.

282. Kim N, Jinks-Robertson S (2011) Guanine repeat-containing sequences confer transcription-dependent instability in an orientation-specific manner in yeast. *DNA Repair (Amst)* 10(9):953–960.

283. Hamperl S, Cimprich KA (2014) The contribution of co-transcriptional RNA:DNA hybrid structures to DNA damage and genome instability. *DNA Repair (Amst)* 19:84–94.

284. Kim N, Cho J-E, Li YC, Jinks-Robertson S (2013) RNA:DNA hybrids initiate quasi-palindrome-associated mutations in highly transcribed yeast DNA. *PLoS Genet* 9(11):e1003924.