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

Samuel Million-Weaver^a, Ariana N. Samadpour^{a,1}, Daniela A. Moreno-Habel^{a,1}, Patrick Nugent^a, Mitchell J. Brittnacher^a, Eli Weiss^a, Hillary S. Hayden^a, Samuel I. Miller^a, Ivan Liachko^b, and Houra Merrikh^{a,2}

Departments of ^aMicrobiology and ^bGenome Sciences, University of Washington, Seattle, WA 98195

Edited by Philip C. Hanawalt, Stanford University, Stanford, CA, and approved January 30, 2015 (received for review November 19, 2014)

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.

replication conflicts | transcription orientation | Y-family polymerases | excision repair | TCR

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 (1–6). 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 (*SI Appendix, Fig. S1*). In either case, encounters between the two machineries can result in replication stress (7, 8). 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 (7, 9–11). 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 (12) 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 (12). 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 (13–15). It is well known that Y-family polymerases facilitate replication progression past bulky lesions on the DNA after damage (16–20). The lesion bypass activity of these polymerases is error-prone, leading to increased mutagenesis (21–23). 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 (24–26). 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 (27–33) in *Escherichia coli* (27–33). Whether these polymerases impact mutation rates

Significance

Replication and transcription can occur concurrently and use the same DNA template. This can cause genomic instability such as common fragile site instability in eukaryotes, and accelerated evolution of lagging-strand genes in bacteria. Here, we report an underlying mechanism that increases mutation rates of lagging-strand genes in *Bacillus subtilis*. We find that this process is mediated through the transcription-dependent activity of the Y-family polymerase, PolY1, in transcription-coupled nucleotide excision repair. We find that PolY1, likely through this mechanism, reduces the potentially problematic impacts of specifically lagging-strand transcription on replisome progression and genomic instability. This work shows that gene orientation, together with transcription, can cause differential activities of repair mechanisms, specifically increasing mutagenesis in lagging-strand genes.

Author contributions: S.M.W., D.A.M.H., M.J.B., and H.M. designed research; S.M.W., A.N.S., P.N., M.J.B., and E.W. performed research; E.W. and I.L. contributed new reagents/analytic tools; S.M.W., E.W., H.S.H., S.I.M., and H.M. analyzed data; and S.M.W. and H.M. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

¹A.N.S. and D.A.M.H. contributed equally to this work.

²To whom correspondence should be addressed. Email: merrikh@uw.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1416651112/-DCSupplemental.

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*) (34). 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 (34–37).

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.

Results

PolyY1 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 (39–41). 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 (35) (*SI Appendix, Fig. S2*).

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 (*leuC27*). Each reporter was placed under the control of the leaky isopropyl β -D-1-thiogalactopyranoside (IPTG)-inducible promoter $P_{\text{spank(hy)}}$ (Fig. 1 *A* and *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 (*SI Appendix, Fig. S1*). 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 (*SI Appendix, Fig. S34*)].

Induction of transcription led to increased reversion rates for all three constructs in both orientations (Fig. 1 *C–E* and *SI Appendix, Fig. S3 B–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 (*SI Appendix, Fig. S4*). 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 (*SI Appendix, Fig. S5*). Despite the differences in absolute levels of mutagenesis, the transcription-dependent increase in mutation rates for all three constructs examined was about twofold higher

Fig. 1. The transcription-dependent mutation asymmetry requires PolY1. (A) Schematic of the *amyE* locus harboring mutation reporters oriented head-on (HM419, HM421, HM632, HM634, HM417, HM724) or codirectionally (HM420, HM422, HM633, HM635, HM418, HM725) to replication. (B) Wild-type and mutant stop codons for the *hisC952* (HM419, HM420, HM421, HM422), *metB5* (HM632, HM633, HM634, HM635), and *leuC27* (HM417, HM418, HM724, HM725) mutation reporters. (C) Mutation rates for the *hisC952* reporter in the presence (HM419 and HM420; gray) or absence of PolY1 ($\Delta yqjH$) (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 (HM 632 and 633; gray) (HM634 and 635; black). (E) Same as C but for the *leuC27* reporter strains (HM417 and HM418; gray) (HM724 and HM725; black). 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.

when the gene was oriented head-on to replication (Fig. 1 C–E and *SI Appendix*, Fig. S3 B–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 (Fig. 1 C–E and *SI Appendix*, Fig. S3 B–D). These effects were specific to PolY1: we did not observe significant effects of PolY2 on transcription-dependent mutation rates (*SI Appendix*, Fig. S6). 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 (*SI Appendix*, Fig. S7). 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 codirectionally to replication. P_{xis} is tightly repressed by the ImmR protein, which is coded for by the mobile genetic element ICEBs1 (42). We introduced the $P_{xis}-lacZ$ constructs into strains that either harbored (transcription off) or were cured of (transcription on) ICEBs1. 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 (Fig. 2 A and B and *SI Appendix*, Fig. S8A). Although we find normalized values overall to be most consistent between experiments, to rule out potential artifacts of normalization, we also analyzed non-normalized, raw immunoprecipitation (IP)/total values (*SI Appendix*, Fig. S8 B and 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 (Fig. 2 A and B and *SI Appendix*, S8 B and 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

Fig. 2. 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 *yhaX*, 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 *yhaX*, in the HO orientation, with [HM352 (Trx+) and HM630 (Trx-); gray bars] and without PolY1 ($\Delta yajH$) [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 ($\Delta yajH$) [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 ($\Delta yajH$; black bars). Error bars represent SE ($n = 12$) (** $P < 0.01$, *** $P < 0.005$). Significance was determined by Student's *t* test.

the engineered constructs, without PolY1, DnaC association increased at the head-on gene, *rpsD*, but not the codirectionally oriented *rplGB* gene (Fig. 2C). 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 (43). In cells inhibited for replication with HPUra, we no longer detected preferential association of DnaC with the highly transcribed P_{xis} -*lacZ* gene (SI Appendix, Fig. S9), 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 (44, 45) is traditionally used as a measure of replication stress. We obtained a previously characterized RecA-GFP allele (44) 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 (SI Appendix, Fig. S10). 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 (Fig. 3A and B and SI Appendix, Fig. S11). This result aligns well with reported basal RecA localization levels in both *B. subtilis* and *E. coli* (44, 45), 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 (Fig. 3A and B and SI Appendix, Fig. S11). 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 (46). 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 pref-

erential association of RecA with the reporter regions compared with control loci (Fig. 3C and D and SI Appendix, Fig. S12). 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 (Fig. 3C and D and SI Appendix, Fig. S12). 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 (Fig. 3E). 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 (31, 47, 48). 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* (49, 50). We determined that transcription was still mutagenic, in an asymmetric manner, even when AddAB or RecA were not present (Fig. 4A and 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 (51), 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 (34).

PolY1 Functions in the Same Pathway as TC-NER. The eukaryotic error-prone polymerase Polk, which is homologous to PolY1, has been shown to fill in gaps generated by the NER proteins (13). 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 (52). Previous studies have connected the Y- and B-family polymerases to transcription-coupled repair (TCR), which can recruit excision repair proteins to stalled RNAPs (30, 53, 54). Although a study showing orientation-dependent effects of NER in yeast observed a change in mutation spectra (54), 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 (Fig. 5A). As we found in the *polY1* deletion strain, the deletion of *uvrA* specifically reduced

Fig. 3. 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

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 (Fig. 5A). 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 (SI Appendix, Fig. S13). 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 (55). Stalled RNA polymerases are canonically recognized by the transcription-repair coupling factor, Mfd, which recruits UvrA, initiating NER (56) [studies in *E. coli* indicate that the accessory helicase UvrD may also facilitate this process (57)]. 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 (Fig. 5B). Furthermore, the *mfd* deletion was epistatic to the *polY1* deletion, confirming that these proteins act in the same pathway (Fig. 5B). 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, archaeal, 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 (24, 26, 58–65). Our analyses of the PolY1-dependent base substitutions leading to prototrophy [multiple substitutions are tolerated at this position (SI Appendix, Fig. S14)] in the mutation rate assays described above are consistent with the reported mutational signatures of these polymerases (SI Appendix, Fig. S15). 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

[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 *yhaX*, in the HO orientation, with [HM352 (Trx+) and HM630 (Trx-); gray] and without PolY1 ($\Delta yqjH$) [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 ($\Delta yqjH$) [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 ($\Delta yqjH$) (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 ($\Delta yqjH$; 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.

Fig. 4. PolY1 functions independently of RecA and AddAB. (A) Mutation rates were estimated for the *hisC952* reporter in strains lacking AddAB (HM720 and HM721, $\Delta addAB$; gray) or AddAB and PolY1 (HM740 and HM741, $\Delta addAB \Delta yqjH$; 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, $\Delta recA$; gray) or RecA and PolY1 (HM746 and HM747, $\Delta recA \Delta yqjH$; black), with (Trx-) or without (Trx+) IPTG. Mutation rates were estimated based on $C = 30-36$, for each strain and condition. Error bars represent 95% confidence intervals (* $P < 0.05$).

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 (12, 66), 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 (67–70). 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) (Fig. 6 and *SI Appendix*, Figs. S1 and S16 and Table S1). 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.

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

Fig. 5. 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, $\Delta yqjH$; black), strains lacking UvrA (HM735 and HM736, $\Delta uvrA$; white) or UvrA and PolY1 (HM1102 and HM1103, $\Delta uvrA \Delta yqjH$; 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, $\Delta mfd \Delta yqjH$; 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 (* $P < 0.05$).

Fig. 6. 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 (*SI Appendix, Table S1 and Fig. S1*).

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 (Fig. 7). 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 postreplication. 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 (*SI Appendix, Fig. S1*). 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 (71). DksA and other transcription-modulating factors (GreA/B) also ease severity of replication–transcription conflicts in *E. coli* (72–74). 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*

Fig. 7. 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.

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 (6, 66, 75–83). 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 (84). 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 (85), 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* (86). The NER pathway in yeast efficiently processes *N*-6-methyladenine lesions (87, 88) and is also recruited to abasic sites in the transcribed strand through TCR (89). Additionally, NER and TCR play a role in protecting mitochondrial DNA from oxidative

lesions, such as 8-oxo-dG (90) or bulky thymine glycol lesions, which distort the double helix and block transcription (91–93).

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 (94, 95). 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 (27–29, 33, 96). These studies have shown stress-dependent elevation of mutation rates in nongrowing 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 (54, 97), 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 (12). 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 (98, 99). 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 cooriented with replication.

Materials and Methods

***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). Detailed strain construction strategies and a full list of strains, plasmids, and oligonucleotides used in this study can be found in [SI Appendix](#).

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

Mutation Rates. Fluctuation analyses were performed as described previously (12, 101). Precultures were grown for ~14 h, and then diluted to OD₆₀₀ 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 OD₆₀₀ of 0.5. A volume of 1.5 mL of each culture was centrifuged, 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 (12).

ChIPs. Precultures were diluted to OD₆₀₀ of 0.05, either in the presence or absence of 1 mM IPTG. Cultures were grown to OD₆₀₀ of 0.3, and processed for ChIPs as described (8). IPs were performed using rabbit polyclonal

antibodies against DnaC or GFP. IPs were rotated overnight at 4 °C for GFP ChIPs, and for 2 h at room temperature for DnaC ChIPs.

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 *yhaX* (primers HM192/193), *guaC* (HM958/959), or an intergenic region 1 kb downstream of *thrC* (HM960/961). *lacZ*, *hisC*, and *dnaK* were detected with primer pairs HM188/189, HM495/496, and HM770/771.

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 $\times$ 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.

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 also 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 con-

sidered 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.

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>).

ACKNOWLEDGMENTS. We thank members of the H.M. laboratory for helpful discussions, as well as Larry Loeb and Ashiwani Kamath for feedback on the manuscript. We thank Benedicte Michel for helpful discussions. We also thank Sonya L. Heltshe (Seattle Children's Research Institute) for assistance with statistical analysis of the SNP data, as well as Christopher Merrikkh and Yaroslava Polosina for help with mRNA quantification. We are grateful to the Grossman laboratory for the anti-DnaC and anti-GFP antibodies, and the Simmons laboratory for providing HPUra. We also thank the Mougous laboratory for providing access to their microscope and Michele LeRoux for technical help with the microscopy experiments. Research reported in this publication was supported by National Institute of General Medical Sciences Award DP2GM110773 (to H.M.), National Research Service Award T32 GM07270 (to S.M.W.), National Institute of Allergy and Infectious Diseases funding (U54AI057141) (to M.J.B. and H.S.H.), and National Science Foundation Grant 1243710 (to I.L.).

- Mortusewicz O, Herr P, Helleday T (2013) Early replication fragile sites: Where replication-transcription collisions cause genetic instability. *EMBO J* 32(4):493–495.
- Merrikkh H, Zhang Y, Grossman AD, Wang JD (2012) Replication-transcription conflicts in bacteria. *Nat Rev Microbiol* 10(7):449–458.
- Helmrich A, Ballarino M, Nudler E, Tora L (2013) Transcription-replication encounters, consequences and genomic instability. *Nat Struct Mol Biol* 20(4):412–418.
- Tuduri S, Crabbe 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.
- Aguilera A, García-Muse T (2012) R loops: From transcription byproducts to threats to genome stability. *Mol Cell* 46(2):115–124.
- Kim N, Jinks-Robertson S (2012) Transcription as a source of genome instability. *Nat Rev Genet* 13(3):204–214.
- French S (1992) Consequences of replication fork movement through transcription units in vivo. *Science* 258(5086):1362–1365.
- Merrikkh H, Machón C, Grainger WH, Grossman AD, Soultanas P (2011) Co-directional replication-transcription conflicts lead to replication restart. *Nature* 470(7335):554–557.
- Lin YL, Pasero P (2012) Interference between DNA replication and transcription as a cause of genomic instability. *Curr Genomics* 13(1):65–73.
- 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.
- Srivatsan A, Tehranchi A, MacAlpine DM, Wang JD (2010) Co-orientation of replication and transcription preserves genome integrity. *PLoS Genet* 6(1):e1000810.
- Paul S, Million-Weaver S, Chattopadhyay S, Sokurenko E, Merrikkh H (2013) Accelerated gene evolution through replication-transcription conflicts. *Nature* 495(7442):512–515.
- 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.
- Heltsel JM, 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.
- 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.
- Fuchs RP, Fujii S (2013) Translesion DNA synthesis and mutagenesis in prokaryotes. *Cold Spring Harb Perspect Biol* 5(12):a012682.
- 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.
- Beuning PJ, Simon SM, Godoy VG, Jarosz DF, Walker GC (2006) Characterization of *Escherichia coli* translesion synthesis polymerases and their accessory factors. *Methods Enzymol* 408:318–340.
- Fuchs RP, Fujii S, Wagner J (2004) Properties and functions of *Escherichia coli*: Pol IV and Pol V. *Adv Protein Chem* 69:229–264.
- Jarosz DF, Beuning PJ, Cohen SE, Walker GC (2007) Y-family DNA polymerases in *Escherichia coli*. *Trends Microbiol* 15(2):70–77.
- Friedberg EC, Wagner R, Radman M (2002) Specialized DNA polymerases, cellular survival, and the genesis of mutations. *Science* 296(5573):1627–1630.
- Goodman MF (2002) Error-prone repair DNA polymerases in prokaryotes and eukaryotes. *Annu Rev Biochem* 71:17–50.
- Shcherbakova PV, Fijalkowska JJ (2006) Translesion synthesis DNA polymerases and control of genome stability. *Front Biosci* 11:2496–2517.
- 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.
- Pata JD (2010) Structural diversity of the Y-family DNA polymerases. *Biochim Biophys Acta* 1804(5):1124–1135.
- 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.
- 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.
- Foster PL (2007) Stress-induced mutagenesis in bacteria. *Crit Rev Biochem Mol Biol* 42(5):373–397.
- Rosenberg SM (2001) Evolving responsively: Adaptive mutation. *Nat Rev Genet* 2(7):504–515.
- 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.
- Shee C, Gibson JL, Rosenberg SM (2012) Two mechanisms produce mutation hot-spots at DNA breaks in *Escherichia coli*. *Cell Rep* 2(4):714–721.
- 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.
- 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.
- Duigou S, Ehrlich SD, Noirot P, Noirot-Gros MF (2004) Distinctive genetic features exhibited by the Y-family DNA polymerases in *Bacillus subtilis*. *Mol Microbiol* 54(2):439–451.
- Sung HM, Yeaman 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.

36. Rivas-Castillo AM, Yasbin RE, Robledo 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.
37. Duigou S, Ehrlich SD, Noirot P, Noirot-Gros MF (2005) DNA polymerase I acts in translesion synthesis mediated by the Y-polymerases in *Bacillus subtilis*. *Mol Microbiol* 57(3):678–690.
38. Kunkel TA (2004) DNA replication fidelity. *J Biol Chem* 279(17):16895–16898.
39. Jena NR (2012) DNA damage by reactive species: Mechanisms, mutation and repair. *J Biosci* 37(3):503–517.
40. Walker GC (1984) Mutagenesis and inducible responses to deoxyribonucleic acid damage in *Escherichia coli*. *Microbiol Rev* 48(1):60–93.
41. Lehmann AR (2002) Replication of damaged DNA in mammalian cells: New solutions to an old problem. *Mutat Res* 509(1–2):23–34.
42. 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.
43. 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.
44. 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.
45. Renzette N, et al. (2005) Localization of RecA in *Escherichia coli* K-12 using RecA-GFP. *Mol Microbiol* 57(4):1074–1085.
46. Keijser BJ, et al. (2007) Analysis of temporal gene expression during *Bacillus subtilis* spore germination and outgrowth. *J Bacteriol* 189(9):3624–3634.
47. Lovett ST (2006) Replication arrest-stimulated recombination: Dependence on the RecA paralogs, RadA/Sms and translesion polymerase, DinB. *DNA Repair (Amst)* 5(12):1421–1427.
48. 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.
49. 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.
50. Wigley DB (2013) Bacterial DNA repair: Recent insights into the mechanism of RecBCD, AddAB and AdnAB. *Nat Rev Microbiol* 11(1):9–13.
51. Walker GC (1995) SOS-regulated proteins in translesion DNA synthesis and mutagenesis. *Trends Biochem Sci* 20(10):416–420.
52. Kisker C, Kuper J, Van Houten B (2013) Prokaryotic nucleotide excision repair. *Cold Spring Harb Perspect Biol* 5(3):a012591.
53. Cohen SE, Walker GC (2011) New discoveries linking transcription to DNA repair and damage tolerance pathways. *Transcription* 2(1):37–40.
54. 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.
55. Ganesan A, Spivak G, Hanawalt PC (2012) Transcription-coupled DNA repair in prokaryotes. *Prog Mol Biol Transl Sci* 110:25–40.
56. Hanawalt PC, Spivak G (2008) Transcription-coupled DNA repair: Two decades of progress and surprises. *Nat Rev Mol Cell Biol* 9(12):958–970.
57. Epshtein V, et al. (2014) UvrD facilitates DNA repair by pulling RNA polymerase backwards. *Nature* 505(7483):372–377.
58. Wagner J, Nohmi T (2000) *Escherichia coli* DNA polymerase IV mutator activity: Genetic requirements and mutational specificity. *J Bacteriol* 182(16):4587–4595.
59. 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.
60. 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.
61. Potapova O, Grindley ND, Joyce CM (2002) The mutational specificity of the Dbh lesion bypass polymerase and its implications. *J Biol Chem* 277(31):28157–28166.
62. 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.
63. 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.
64. Tissier A, McDonald JP, Frank EG, Woodgate R (2000) poliota, a remarkably error-prone human DNA polymerase. *Genes Dev* 14(13):1642–1650.
65. Zhang Y, et al. (2000) Human DNA polymerase kappa synthesizes DNA with extraordinarily low fidelity. *Nucleic Acids Res* 28(21):4147–4156.
66. 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.
67. 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–E2783.
68. 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.
69. Hersberg R, Petrov DA (2010) Evidence that mutation is universally biased towards AT in bacteria. *PLoS Genet* 6(9):e1001115.
70. 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.
71. 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.
72. 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.
73. Tehrani AK, et al. (2010) The transcription factor DksA prevents conflicts between DNA replication and transcription machinery. *Cell* 141(4):595–605.
74. 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.
75. Prado F, Aguilera A (2005) Impairment of replication fork progression mediates RNA polII transcription-associated recombination. *EMBO J* 24(6):1267–1276.
76. de la Loza MC, Wellinger RE, Aguilera A (2009) Stimulation of direct-repeat recombination by RNA polymerase III transcription. *DNA Repair (Amst)* 8(5):620–626.
77. 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.
78. 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.
79. Reaban ME, Griffin JA (1990) Induction of RNA-stabilized DNA conformers by transcription of an immunoglobulin switch region. *Nature* 348(6299):342–344.
80. Kim N, Jinks-Robertson S (2009) dUTP incorporation into genomic DNA is linked to transcription in yeast. *Nature* 459(7250):1150–1153.
81. Gaillard H, Herrera-Moyano E, Aguilera A (2013) Transcription-associated genome instability. *Chem Rev* 113(11):8638–8661.
82. Gaillard H, Aguilera A (2013) Transcription coupled repair at the interface between transcription elongation and mRNP biogenesis. *Biochim Biophys Acta* 1829(1):141–150.
83. Robledo EA, Martin HA, Pedraza-Reyes M (2012) Mfd and transcriptional derepression cause genetic diversity in *Bacillus subtilis*. *Front Biosci (Elite Ed)* 4:1246–1254.
84. Grivell AR, Grivell MB, Hanawalt PC (1975) Turnover in bacterial DNA containing thymine or 5-bromouracil. *J Mol Biol* 98(1):219–233.
85. 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.
86. Vaisman A, et al. (2013) Removal of misincorporated ribonucleotides from prokaryotic genomes: An unexpected role for nucleotide excision repair. *PLoS Genet* 9(11):e1003878.
87. 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.
88. 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.
89. 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.
90. Melis JP, van Steeg H, Luijten M (2013) Oxidative DNA damage and nucleotide excision repair. *Antioxid Redox Signal* 18(18):2409–2419.
91. 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.
92. 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.
93. 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.
94. 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.
95. Grossman L, Caron PR, Mazur SJ, Oh EY (1988) Repair of DNA-containing pyrimidine dimers. *FASEB J* 2(11):2696–2701.
96. 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.
97. Wimberly H, et al. (2013) R-loops and nicks initiate DNA breakage and genome instability in non-growing *Escherichia coli*. *Nat Commun* 4:2115.
98. Kibota TT, Lynch M (1996) Estimate of the genomic mutation rate deleterious to overall fitness in *E. coli*. *Nature* 381(6584):694–696.
99. Sniegowski PD, Murphy HA (2006) Evolvability. *Curr Biol* 16(19):R831–834.
100. Harwood CR, Cutting SM, eds (1990) *Molecular Biological Methods for Bacillus* (Wiley, Chichester, UK).
101. Foster PL (2006) Methods for determining spontaneous mutation rates. *Methods Enzymol* 409:195–213.
102. Brittnacher MJ, et al. (2011) PGAT: A multistep analysis resource for microbial genomes. *Bioinformatics* 27(17):2429–2430.
103. Edgar RC (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32(5):1792–1797.